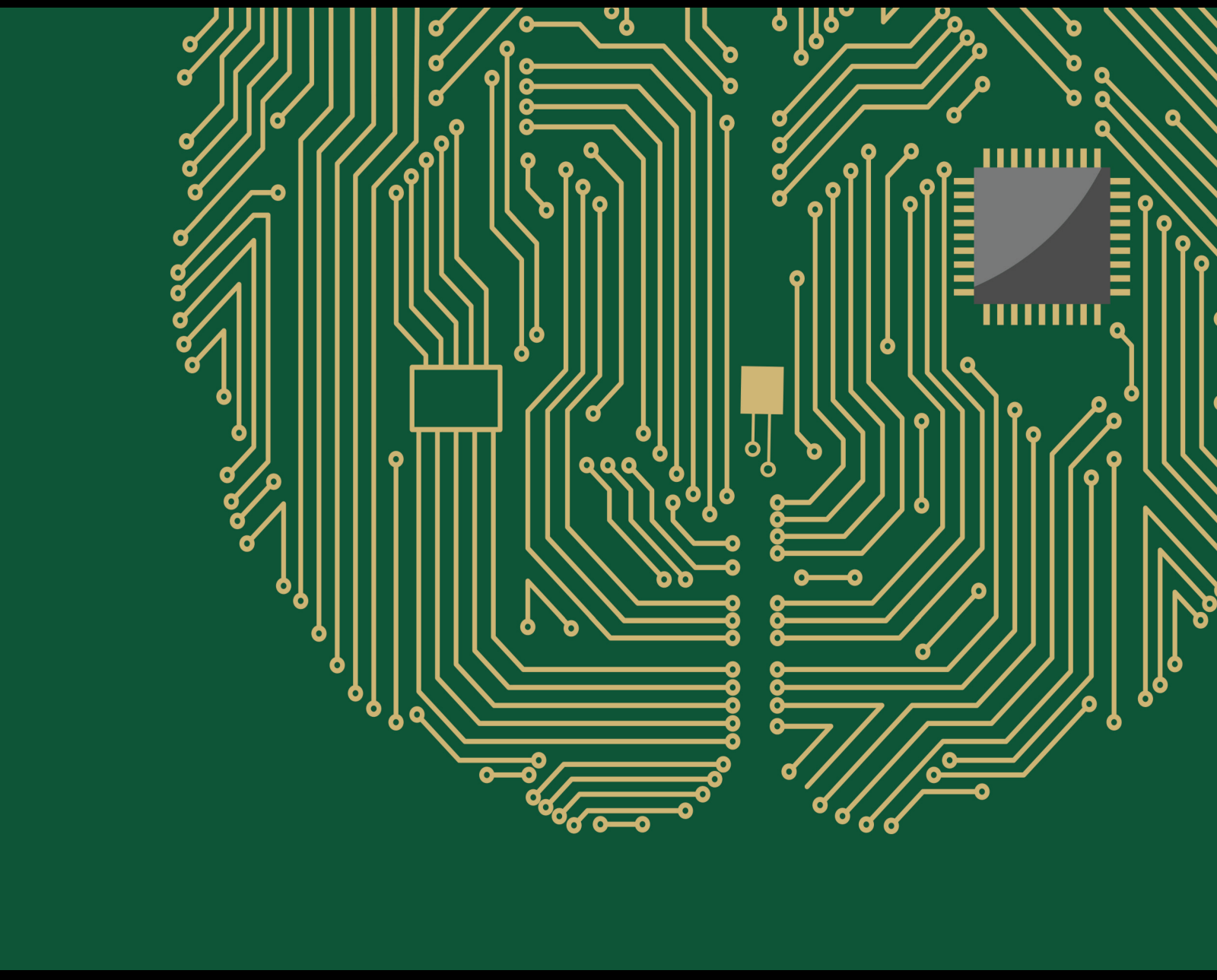
Computational Intelligence and Neuroscience

Artificial Intelligence and Machine Learning-Driven Decision-Making and Control 2022

Lead Guest Editor: Heng Liu
Guest Editors: Jun Zhang, Basil Mohammed Al–Hadithi, and Wei Xiang



Artificial Intelligence and Machine Learning-Driven Decision-Making and Control 2022

Computational Intelligence and Neuroscience

Artificial Intelligence and Machine Learning-Driven Decision-Making and Control 2022

Lead Guest Editor: Heng Liu
Guest Editors: Jun Zhang, Basil Mohammed Al–Hadithi, and Wei Xiang

Contents

Contents

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 3250499, 12 pages
https://doi.org/10.1155/2022/3250499

Research Article

Estimation of Parameters on Probability Density Function Using Enhanced GLUE Approach

Fuad S. Alduais[1] **and Neveen Sayed-Ahmed**[2]

[1]*Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia*
[2]*Statistics Department, Faculty of Commerce (Girl Branch), Al-Azhar University, Cairo, Egypt*

Correspondence should be addressed to Fuad S. Alduais; f.alduais@psau.edu.sa

Received 1 July 2022; Revised 6 September 2022; Accepted 10 September 2022; Published 14 October 2022

Academic Editor: Heng Liu

The most essential process in statistical image and signal processing is the parameter estimation of probability density functions (PDFs). The estimation of the probability density functions is a contentious issue in the domains of artificial intelligence and machine learning. The study examines challenges related to estimating density functions from random variables. Based on minimal predictions regarding densities, the study discusses a framework for evaluating probability density functions. During the Bayesian approach, which is to generate correct samplings that reflect the probability aspect of the variables, sampling is widely used to estimate as well as define the probabilistic model of unknown variables. Because of its effectiveness and extensive application, the generalized likelihood uncertainty estimation (GLUE) method has earned the most popularity among the various methodologies. The Bayesian technique allows parameters of the model to be estimated using prior expertise in the parameter results and experimental observations. The study uses a number of engineering issues that were lately looked into to illustrate the effectiveness of the upgraded GLUE. As the focus is on the examination of sampling effectiveness in view of engineering components, only a brief summary is provided to describe every challenge. The suggested GLUE method's outcomes are contrasted with those obtained using MCMC. Nevertheless, using the GLUE approach, the model's mean squared error of prediction is substantially higher than that of the previous algorithms. The methods' results are affected by the assumptions being made on parameter values in advance. The concepts of prediction accuracy, as well as the utility of geometric testing, are presented. Such notions are valuable in demonstrating that the GLUE approach defines an inconsistent and incoherent statistical inference process.

1. Introduction

Parameter estimate is described as the method for the determination of estimates for parameters that govern the response of the structure, provided that the system architecture is established. Parameter estimation is a field of statistics that involves estimating the parameters of a distribution utilizing data samples. For precise predicted results as well as effective model-based decision criteria, complete model parameter estimation is essential. Basic knowledge depended on expert information and ideas obtained through experimentation are still the two forms of knowledge commonly accessible for calculating the parameters of certain models. Just the information is utilized to estimate parameters of the model throughout most statistical methods, including the least squares method [1]. Throughout several engineering challenges, including such analysis of structures at the designing phase or the health maintenance of existing systems, parameter estimation is always a necessary step. Material parameters of the constitutive equations that have a substantial impact on the reliability of the simulation analysis must be accurately determined using data from direct or indirect observations during the design phase. In order to determine the remaining useful capacity of structures, degradation parameters of the underlying physical concept in decaying

frameworks must be estimated utilizing observed data throughout time.

In contrast to conventional techniques, the Bayesian methodology enables the estimation of parameter estimation to include not only observations but also preconceived notions regarding parameter values. Those approaches also have the benefit of being able to analyze the effects of parametric uncertainties on simulated data. The unidentified model parameters were represented as a stochastic process with a probability distribution that indicates uncertainty regarding model parameters in the Bayesian network [2]. The probability is adapted from previous information regarding parameter values proceeding to gather additional data. If no knowledge is accessible, an uninformative preceding distribution can be defined. Several of the parameters in ecological models, however, are immediately relevant, and minimum and maximum ranges can typically be specified for these values. In these circumstances, the previous parameter distributions could be determined by a homogenous distribution, for example. Bayes' theorem determines the parameter distribution prior to data collection. The previous distributions and the data both influence the posterior parameter distributions. As a result, that distribution comprises all of the knowledge on the design variables that are presently available [3].

The posterior parameter distributions cannot be computed and analyzed owing to the difficulty of certain prototypes (nonlinearity, large number of parameters). Nevertheless, as machines get more powerful and newer methodologies emerge, the Bayesian approach becomes extra manageable, also for complicated techniques. In recent years, the Bayesian network is becoming popular as a method for quantifying the uncertainties of variables in estimate processes. The following is a summary of the Bayesian approach: create a posterior probability for the uncertain variables derived from empirical data that indicates the level of confidence. Generate sampling that reflects the parameter range [4]. The variables of a framework are not considered particularly to be characterizations of physically constructive amounts with true (albeit unknown) principles in the traditional Bayesian approach, but instead, provisional "fake" or "suitable" amounts of uncertainties (on that every uncertainty is predictable) to be marginalized out by using one's posterior probability density that is acquired from observational data through the Bayesian inference procedure. If the goal is "parameter estimation," it is explicitly presumed that the variables have a true but uncertain value that could be approximated once the probability density was already computed, based on either the maximum probability value or the anticipated value [5].

The probability density function (PDF) is a probability function that indicates the distribution of a continuous probability distribution that falls among a set of values. In other terms, the probability density function calculates the probability of discrete random variable values [6]. A probability distribution function or simply a probability function is another name for it. However, several other reports indicate the function as a function across a wide range of values. It is also known as the cumulative distribution function or the probability mass function (PMF). PDF (probability density function) is established for continuous random variables, while PMF (probability mass function) is established for discrete random variables [7]. The probability density function is specified as the average of the variable density distribution over a certain range. Letter f stands for it (x). At any point on the graph, the variable is positive or non-negative, as well as the integral, more precisely the defined intrinsic of PDF over the entire facility, is always one. The likelihood of the occurrences is usually shown by a bell curve on the graphs of PDFs.

A continuous random variable's probability density function (PDF) could be easily estimated using the notion of power spectral density (PSD) estimation. A PDF varies from a PSD in that it has a space constraint, whereas a PSD does not. In pattern classification, imaging, and signal processing, the estimation method of probability density functions (PDFs) is essential [8]. For the identification of the fundamental signals, several actual duration signal processing applications demand autonomous, steady, and statistically developed management. Generating parameter estimates in the foundation of simple image processing tasks, including recognition and classification, is a typical difficulty that is encountered when using statistical methodologies. The system variable's probability density function form could be used as a better analytical technique to fully model the performance of a stochastic process. As a result, for management strategy creation with diverse design criteria, a PDF-based approach delivers accuracy and flexibility [9]. In general, the posterior distribution is expressed as a sophisticated or explicit formulation in terms of the dimensions, making sample generation difficult and preventing the use of typical probabilities functional approaches. Throughout this approach, numerous systems have been employed.

The generalized likelihood uncertainty estimation (GLUE) approach has recently been demonstrated as a statistically efficient method. Beven and Freer [10] proposed the generalized likelihood uncertainty estimating methodology based on notions that, while articulated in various terminologies, are quite comparable to Bayesian concepts. GLUE "represents the evolution of Bayesian or fuzzy averaging processes to the less formal likelihood or fuzzy measures," as per the researcher. A notable part of the GLUE technique is the idea of "the fewer formalized likelihood," which provides the fundamental point of variation through Bayesian inference. The GLUE approach is a Monte Carlo technique with the goal of identifying a variety of cognitive modeling from a range of model/parameter combinations. On the basis of various information and expertise, the phrase "behavioral" refers to concepts that are deemed "acceptable," that is, not ruled out. To construct GLUE, a significant amount of iterations are carried out for a model output with various parametric configurations picked at random using preceding parameter distributions. Every set of model parameters is allocated a probability value, which is a statistic that measures in what way that specific variable grouping (or model) replicates the systems, by evaluating expected and observed responses. Greater likelihood functional numbers

often propose a good match among predicted results as well as the clarifications. The complete collection of imitations is later divided into behavioral and nonbehavioral parameters depending on a cutoff threshold [11].

The idea behind this strategy is to discretize the number of variables by using the posterior probability to create a large number of model parameters. From the probability and probability density values, values are determined for each parameter value. Even though Beven [12] claim that GLUE has been created to address uncertainty related to several types of errors, including "error due to vaguely understood model parameters as well as input variables, and error associated with measurement techniques used during the calibration process, as well as error due to modeling process deficiency," in actuality, it only accounts for them implicitly through the stochastic nature of the model residuals. In fact, GLUE is dependent on the structural model, inputs mistakes, and initial values, which would not be considered unpredictable [13]. With the underlying understanding of cumulative mistakes, unpredictability is only believed to be caused by an inadequate knowledge about modeling attribute values, as well as perhaps by forecast and observational errors. The technique utilizes Monte Carlo (MC) analyses in combination having Bayesian estimates and uncertainty propagating to translate the variability in the model development onto the dimensional space. The GLUE methodology rejects the idea of a single global optimum input parameter inside a structural model, rather than acknowledging the acceptance of distinct parameter values that are as good at delivering fit forecasting accuracy within a model structure. The analysis of different model parameters inside pseudo-Bayesian MC architecture specifically addresses the idea, known as equifinality. The GLUE process produces parameter distributions including related uncertainty boundaries that are dependent on the available observational data [11].

The prevalence of GLUE could have been credited to its theoretical accessibility, comparative convenience of development and use, and capacity to control a variety of error models and theories without requiring large changes to the method itself. Although this advancement, GLUE has been criticized since it cannot be officially Bayesian in necessitating particular judgments on the likelihood function as well as cutoff threshold dividing behavioral and nonbehavioral techniques, and it will not be adopting a quantitatively reliable error method. Furthermore, in several of the GLUE implementations, a very simple MC sampling strategy is utilized to sample from the distributions of the prior parameters and identify a well-distributed variety of cognitive modeling including their corresponding prediction simulations uncertainty [14]. Practitioners of the GLUE technique typically use basic random selection or, in certain situations, the much more accurate Latin hypercube sampling (LHS) procedure to survey the prior parameter distributions. Despite their ease of implementation, random sampling techniques were uncertain to intensively illustrate the limited space close to the finest solution with a dense distribution of values. The hypothesis is that by utilizing an adaptive sampling technique that updates the search direction based on information from previous draws, significant gains in sampling can be made. Such a strategy would almost certainly result in more reliable parameters and prediction error estimations [15].

Numerous engineering issues that have lately been researched are used to highlight the effectiveness of the GLUE. Only a brief summary of every issue is provided because the analysis of sampled effectiveness in perspective of engineering fields is the main topic. The suggested GLUE method's outcomes are contrasted with those from the MCMC. In every one of the challenges, the procedure is carried out with $nl = 10{,}000$ LHS and $nm = 5000$ MCMC iterations. Several approaches to solve this problem were published in the statistical literature; however, satisfactory solutions in actuality are uncommon in engineering applications. The marginal PDF of every variable from the joint posterior distribution is used as a proposal distribution throughout this work, which results in a more robust technique. Because the numerical solution is computationally costly when constructing the marginal PDF, Latin hypercube sampling (LHS) is used to generate the PDF in a discrete manner. The contribution of the study is to illustrate the efficiency of the suggested GLUE strategy; multiple engineering problems with unknown parameters are solved using a Bayesian approach. The article is divided into six sections. In Section 2, existing techniques are briefly discussed. The problem statement is discussed in Section 3. Section 4 covers the GLUE approach, including the implementation of GLUE into PDF. Section 5 contains the results and comments, as well as tables and graphs. Finally, the article is concluded in Section 6.

2. Related Works

The subject of combined Bayesian model evaluation and parameter estimation for sinusoids in white Gaussian noise is investigated in this study. The authors provide a unique Bayesian framework that provides us to construct a posterior distribution on the parameter space. That distribution is then used for everything in Bayesian statistics. However, a direct assessment of such a distribution as well as its properties, such as posterior model probability, necessitates the assessment of certain complex high-dimensional integrals. To execute the Bayesian computations, researchers design an efficient stochastic system based on reversible jumping Markov chain Monte Carlo techniques. The algorithm's convergence outcome is verified. The effectiveness of identification based on posterior modeling probability tends to exceed standard detection systems in simulations. Numerical approaches are required to evaluate such posterior distribution and its aspects of interest. To determine this posterior distribution, an affective computing optimization technique on reversible jumping MCMC techniques was developed. The outcomes of large simulated research reveal that modeling selection based on posterior model probability outperforms the other traditional criteria. Whenever dealing with scenarios with low SNR, limited sample sizes, or tightly packed frequencies, such a strategy is quite useful. Of course, computationally efficient approaches

are a good option in more favorable scenarios. However, some approaches have significant flaws. First, because it might be huge, it is computationally expensive. Second, each parameter is assigned the same computing effort. In fact, several of the variables are irrelevant in practice since their posterior model probability is so low [16].

Analyzing chemically interacting circulating currents with probability density function (PDF) approaches has many benefits. This brings an exceptional and practical solution to the closing issues that result through averaging or filtering the highly nonlinear chemicals' central component, as well as conditions that approximate other one-point physical phenomena (e.g., radiative emission) in the immediate continuity equation. The study is restricted to transportable PDF approaches, in which a formula dictates the development of the one-point, one-time PDF for a collection of factors that influence the localized thermochemical and/or hydrodynamic condition of a reacting system. PDF-based methodologies as subfilter-scale modeling techniques for large-eddy computation (filtered density function methods), PDF-based simulations of thermoelectric radiation heat transfer as well as turbulence–radiation interrelations, PDF-based features for soot and liquid fuel splatters, and Eulerian survey strategies for fixing simulated PDF analytical solution are among the significant developments mentioned. To underline crucial ideas, instances of applicability to canonical processes and laboratory-scale flames, including real combustion devices, are offered. Throughout the book, an endeavor has been made to achieve a balance between rigor and readability, covering depth and breadth, including essential science and actual implications. The analysis is required to respond to increasing the availability of PDF approaches and debunking common misconceptions regarding them. Although PDF approaches have generally been used to react to the ideal-gas combinations utilizing single-turbulence-scale simulations, knowledge from various mechanics and scales is easily included. However, because this technique cannot easily generalize to any three-dimensional geometrical layout, many PDF techniques have disregarded it [17].

On the accelerating component, the probability distribution function has $R = 690$, which has provided observational evidence with probabilities fewer than 1007. That represents a significant improvement above previous observations, enabling us to assume that the fourth-moment convergence rate and the flattening are around 55. Researchers compared the probability distribution to that anticipated by numerous nonextensive statistically mechanics-inspired systems. They also discover that accelerator element stochastic models conditioning on a particle velocity with conditioning velocities up to 3 times the confidence interval are extremely non-Gaussian. The models based on log-normal statistical or multifractal analyses (Arimitsu) are quite comparable to the empirical pieces of evidence. Because the main distinction among the two models is the fundamental demographics (Tsallis or log-normal), the Tsallis stats assumption should be rejected in order to replicate the accelerated PDF's observable behavior. The appropriate reactive of the variable in the models to the diffusion equation and the use of log-normal statistical are in accordance with the improved Kolmogorov–Obukhov model of turbulent, which is recognized to capture well the properties of intermittent nature. The combined PDF of velocity and acceleration is likewise shown to be non-Gaussian. The velocity variability conditional on speed, on the other hand, is not consistent. Such findings have significant implications for the architecture of the stochastic equations necessary to simulate turbulent flow distribution [18].

Independent decision-makers' individuality is inextricably linked to group decision-making, therefore making attaining a group decision challenging. Several of the challenges is aggregating a small number of evaluations while accounting for individual identity or ambiguity using the probability theory. This tough problem is called probability distribution function aggregation (DFA). The study proposes a straightforward and effective solution to the DFA problem. The suggested approach's fundamental concept is to represent the DFA issue as a nonlinear system of a collection of probability distribution functions and to suggest a linear feedback iterative method to address the nonlinear model, resulting in a collective judgment or conclusion. A well-known DFA instance that was resolved using the Delphi method serves as an example of such a novel approach. The decision-making issue involves the DFA problem. As a result, the suggested technique is applicable to any decision-making task. The suggested terminology for methodically encoding the inaccurate group decision issue with the categorization of uncertainty into three forms, inadequate knowledge, ambiguous knowledge, and indeterminate details, would be another sign of this research. With the well-known Delphi technique, the recommended methodology was taken to a situation in the research. The method has been proved to be quite efficient; in the case of the example mentioned in this study, it took 3-4 iterations to attain convergence. The paper also came to the conclusion that the optimal aggregated dispersion functional must be limited and centralized. This discovery was used to create a system for calculating the starting weights in the suggested method. This broad notation is then applied to the specific DFA problem addressed in this study. DFA, on the other hand, has limited memory. It cannot hold any information of unlimited duration due to its finite memory [19].

The analysis of signal obtained on a number of sensors to locate the position of the transmitter is adequate relevance that it has been handled under a variety of special case assumptions. In a noise/interference situation including an arbitrary covariance matrix, the main difficulty involves sensors with random positions including directing properties (gain/phase/polarization). The study is focused on two aspects of the problem: the many emitters component and the universality of the solution. The multiple signal classification (MUSIC) procedures are given, which also offer asymptotically accurate estimations of the following: (1) the percentage of event wavefronts existent; (2) instructions of the entrance (or emitter locations); (3) abilities and traverse correlation coefficients between many occurrence waveforms; and (4) noise/interference resilience. Instead of

integration of an all-pole Fourier transform generated by white noise (i.e., autoregressive modeling, maximum entropy) or maximizing a probability underneath the premise that the vectors are zero average and Gaussian, MUSIC analyzes the information as the combination of specular reflection emission as well as noise (maximum likelihood for Gaussian data). Maximum likelihood minimizes a balanced aggregate of all component lengths, while MUSIC minimizes the length from the continuous to the signal subspace. There are demonstrations and correlations with maximum likelihood (ML) and maximal entropy (ME) approaches, as also traditional beamforming. Its application as a multiple frequency analyzer on data series is demonstrated. There are no assumptions concerning array geometry. The components can be placed in a regular or irregular manner, and their directed qualities (amplitude/phase) can change or be similar, as long as their polarization properties are all the same. The geometrical importance of certain vector space settings and the perception of a specific matrices eigenstructure, on the other hand, were overlooked [20].

Automatic modulation recognition is beneficial for modulation schemes, software-defined wireless, and traditional wireless telecommunication networks. In this study, researchers look at how to detect the modulation scheme of a linear modulation index using a phase-based linear discriminant (ML) approach. Researchers suggest two approximation ML alternatives, which can provide close-to-optimal performances using minimized cost because the optimum ML technique is computationally expensive. Then, for segmentation and classification kinds of modulating configurations, researchers give a general performance assessment. Researchers develop a set of upper constraints on PCC that represent a trade-off between precision and calculation cost. Also addressed is the asymptotic behavior of phase-based ML classification techniques. Researchers have developed 2 approximation phase PDFs that are described as arithmetic operations, by using the Gauss–Legendre quadrature method and the semi-infinite Gauss–Hermite computation method. The initial stage PDF, on the other hand, has a unique feature. These two approximation phase PDFs produce a PCC that is unrecognizable from the ideal ones provided by the initialization step PDF, according to the simulation solution. To the extent that they would operate well for any SNR values of practical interest, these outperformed the Tikhonov PDF and the Fourier series technique. It is important to note that the semi-infinite Gauss–Hermite quadrature rule provides enough precision. However, at a low SNR, the Tikhonov PDF loses precision, and at a high SNR, the Fourier series technique loses accuracy. The Gauss–Legendre quadrature approximation gets less precise as well [21].

Regarding large-eddy modeling of turbulent spray combustible, a coherent probability density function (PDF)-oriented ignition modeling technique is established. In the form of limited frequency LES equations, a Lagrangian Monte Carlo technique for resolving the PDF numerical scheme is devised. An innovative pilot-stabilized ethanol spray flame is simulated using the LES/PDF technique. Droplet evaporation happens distant from the flame front in of this flame, resulting in the distance between the two operations. All across the duration of the flame, a high-temperature preflame region with a stratification combination is discovered. The flame front cannot propagate through all this evaporating but well-mixed fuel/air mixture due to the high fluid velocity of the droplet-laden air. When compared to investigations, there were significant differences in droplet inflow parameters. Droplet-turbulence interaction in the injecting pipe, in this instance, greatly alters inflow disturbance. For application with the LES technique, a continuous PDF methodology for turbulent spray ignition has been developed. In the framework of a low-Mach number LES solver, a Lagrangian Monte Carlo technique that faithfully regenerates the higher-order moments of the PDF was been created. An ethanol spray flame was studied using the LES/PDF method. The LES/PDF technique accurately replicated the results of the experiment. This was also discovered that an absence of data regarding the intake circumstances resulted in significant differences between simulation and experimental results. However, the PDF equations provided in that paper do not achieve the correct vector moment numerical scheme that is critical for continuity. Furthermore, because of the inherent unpredictability of the computations, using the transported-PDF technique in LES necessitates special numerical constraints [22].

3. Problem Statement

The traditional parametric and nonparametric techniques for measuring density functions have a lot of shortcomings; parametric techniques generate poor results when dealing with unidentified distributions, whereas nonparametric methods necessitate a large number of construction measurements, storage capabilities, and computational power. As a result, the paper presented a method for assessing probability density functions since only minimal density estimates are given. Sampling is commonly used to estimate and describe the probabilistic model of unknown factors in the Bayesian approach, which is to provide accurate samplings that represent the probabilistic aspect of the parameters. The generalized likelihood uncertainty estimation (GLUE) approach has gained the most popular among the numerous methodologies due to its efficacy and wide application. The Bayesian method allows model parameters to be calculated based on prior knowledge of parameter outcomes and experimental measurements.

4. Proposed Methodology with Glue

4.1. Bayesian Technique. Bayesian theorem, the following equation provides the posterior parameter distribution [23]:

$$(\alpha \mid X) = \frac{\vartheta(X \mid \alpha)\vartheta(\alpha)}{\vartheta(X)}. \quad (1)$$

Here, X is the measurements vector, $\vartheta(\alpha)$ is the preceding parameter distribution, $\vartheta(\alpha \mid X)$ is the posterior parameter distribution, $\vartheta(X)$ is a proportional variable given by the condition that the integration of $\vartheta(\alpha|X)$across the

dimensional space approaches 1, and $\vartheta(X|\alpha)$ is a probability function. Given the parameters, $\vartheta$the probability is the value of the observation X. Its ratio is described by the inaccuracy probability distribution among modeled and measure actual data. It is clear that either the previous distribution or the emerged data have an impact on the posterior parameter distribution.

4.2. Enhanced Generalized Likelihood Uncertainty Estimation. The work proposes an enhanced robust GLUE sample technique for multidimensional characteristics that provides proper sampling by using a marginal PDF as a proposed distribution. The summation of the targeted joint PDF with consideration of many other parameters excluding it defines the marginal PDF of any random variable. The GLUE approach works on the assumption that the posterior parameter distributions $\vartheta(\alpha|X)$ may be approximated by a discrete probability distribution (α_i, r_i), where r_i is the likelihood connected with the parameter α_i. The procedure is as follows [24, 25]:

(i) Construct K vectors α_i, $i = 1, \ldots, K$ at arbitrary out from preceding parameter distribution $\vartheta(\alpha)$.

(ii) Compute the probability numbers $\vartheta(X|\alpha_i)$ and the probability density $\vartheta(\alpha_i)$, $i = 1, ..., K$, related to the various parameter vectors created.

(iii) Compute

$$r_i = \frac{\vartheta(X|\alpha_i)\vartheta(\alpha_i)}{\sum_{j=1}^{K} \vartheta(X|\alpha_j)\vartheta(\alpha_j)}, \quad j = 1, \ldots, K. \tag{2}$$

Where $\sum_{i=1}^{K} r_i = 1$.

The pairs (α_i, r_i), $i = 1, \ldots, K$, could be used to calculate different properties of the posterior distribution, such as the posterior means $\overline{\alpha} = \sum_{i=1}^{K} \alpha_i, r_i$. The GLUE method implementation is simpler than the others because it only involves the specification of the overall variable of created parameter variables K.

The marginal PDF of a single variable A is parameterized as the total of the targeted combined PDF having regard to many other variables without it that is determined by

$$s(A_i) = \int r(A_1 \ldots A_{i-1}, A_i, A_{i+1 \ldots .} A_n) f A_1 \ldots f A_{i-1}, f A_{i+1 \ldots .} f A_n. \tag{3}$$

The traditional method is to estimate a maximum range for every variable and partition the range because of an equal distance with the number n, as well as determine the joint PDF for all instances. The PDF has, therefore, generated at an arbitrary position A_i^K by the following equations:

$$s\left(A_i^K\right) \propto \sum_{x_1=1}^{nm} \ldots \sum_{x_1-1=1}^{nm} \sum_{x_1+1=1}^{nm} \ldots \sum_{x_n=1}^{nm} r\left(A_1^{x_1}, \ldots A_{1-1}^{x_1-1}, A_1^K, A_{1+1}^{x_1+1} \ldots . A_n^{x_n}\right), \quad K = 1, \ldots, nm. \tag{4}$$

The overall amount of PDF computations throughout this strategy is nm^n, which also gets computationally intensive as the multitude of variables increases. To optimize performance, a simplified approach based on Latin hypercube sampling (LHS) is suggested in this study. As an example, imagine a joint PDF with two variables A_1 and A_2. Combine these ranges by $nm = 8$ to get $8 * 8 = 64$ cells, and create points by the LHS such that every column and row of the square has just one sampling. Traditionally, the quantity is produced by adding the PDF numbers at each position all along a continuous line. The number is totaled at the spots that have relocated with the same A_2 from the LHS spots in the LHS approach. The motion is represented by arrows in the diagram. In this simple 2-D situation, both the traditional and LHS methods are similar, albeit taught differently. However, as demonstrated in the example of three variables, it does not apply when the variables are raised. The LHS elements are produced with an interval count of $nm = 8$. To calculate the marginal PDF value at any random A_1, the typical way is to add all the PDF readings at $8 = 64$ points on the $A_2 - A_3$ planes provided by the translucent grey. Nevertheless, just for performance, only 8 points have migrated from the initial LHS points with the same A_2, A_3 are employed in the suggested method. The marginal PDF of parameter A_i by the LHS technique is provided by the expression given as follows:

$$s\left(A_i^K\right) \propto \sum_{x_1=1}^{nm} r\left(A_1^x, \ldots A_{1-1}^x, A_i^K, A_{1+1}^x \ldots . A_n^x\right), \quad K = 1 \ldots, nm. \tag{5}$$

The nm calculations are at any randomA_i that must be continued across $K = 1, \ldots, nm$. For each variable A_i, nm^2 calculations will be performed and $nm^2 * n$ numbers will be determined for all the parameters, whereas the typical technique requires nm^2. When these two methods are compared, the LHS technique requires significantly less calculation. Though proven with a small proportion $nm = 8$, the amount in actuality is frequently in the tens of thousands. For n, however, just a few values fewer than ten are generally considered. Many computational environments could improve efficiency by calculating all nm PDFs at any arbitrary A_i^K in a single phase. Furthermore, in practice, the marginal PDF is not generated at each, $k = 1, \ldots, nm$, which is still too expensive. Rather than nm, the value of every variable is separated by a considerably smaller amount nm as low as so many tens, but

all nm PDFs were generated in a single step, as previously stated. After all, the overall amount of calculations is reduced to $1 * nm * n$, a far more solvable quantity in terms of processing performance.

The LHS technique's marginal PDF is then used as a proposition density to create an effective and resilient GLUE procedure. Consider the following target PDF with two parameters as an example:

$$r(a1, a2) \propto \prod_{n=1}^{4} Z_n^{Y_n} \left(1 - Z_n\right)^{5-Y_n}. \tag{6}$$

Here, we get

$$\vartheta(\alpha|X) = \frac{\exp[a1 + a2]}{1 + \exp[a1 + a2]}. \tag{7}$$

4.3. Assessment Predicted Values and Parameter Estimations. The vectors of the subsequent average $\overline{\alpha}$ are calculated for every posterior distribution. Therefore, $\overline{\alpha}$ is taken into account as an estimation of the mathematical model vectors α. For every variable excepting $x6$, $y6$, and $z6$, the average absolute difference between the parametric test and the genuine parameter number is determined. Hence, the absolute deviation reflects the actual value of the parameter. Since the real values for all three factors are zero, can be simply utilized the significant differences between the estimated parameters and the real values to evaluate the predicted values of $x6$, $y6$, and $z6$.

Because the input parameter estimation techniques were frequently included to forecast b, it is essential to assess the reliability of model predictions whenever the parameters were set to the parameter estimates. According to [26],the model predicts b values by employing $\overline{\alpha}$ as parameter estimation methods. The reliability of the predicted results is then assessed by computing the mean squared error of prediction (MSEP) values. The models were employed with the real parameters in order to produce 1000 observations; the mean squared error of prediction connected with every vector $\overline{\alpha}$ is predicted in the following equation:

$$\mathrm{MSEP}(\overline{\alpha}) = \frac{1}{1000} \sum_{i=1}^{1000} \left[b_i - f(\beta_i; \overline{\alpha})\right]^2. \tag{8}$$

Assuming parameters of the model were adjusted to the posterior means, $f(\beta_i; \overline{\alpha})$ is the modeling prediction. For every element of α in the posterior parameter distribution, this is the suggested estimating value for mean squared error. MSEP's posterior distribution could then be derived.

5. Results

5.1. Validation of the Enhanced GLUE Sampling Techniques. The effectiveness of the modified GLUE is demonstrated using many engineering challenges that scholars have recently researched. Because the focus is on sampling performance from the perspective of engineering applications, every topic is only briefly described. The suggested method's outcomes are determined by the standard GLUE. In all instances, the method is carried out with the amount of LHS $nm = 1000$ as well as the number of GLUE iterations $n = 500$. For the proposition density in traditional GLUE, a uniform distribution having finite duration is used.

5.1.1. Problem with Spring. Despite finite element analysis (FEA) being a useful method for examining structure elements' fatigue performance, it frequently fails to correctly forecast the existence owing to the intrinsic unpredictability of the fatigued variables. On the contrary, practically every organization conducts live tests at the end of the development process for quality assurance (QA). As a result, the information is collected on its own. Inspired by this, a Bayesian approach is created that uses these testing data to inversely predict the fatigue-life variables. The posterior distributions of the variables are computed based on the life data acquired from the QA tests on a regular basis. The level of uncertainty in the variables might well be decreased considerably when more data are collected. A high cycle fatigue concern of a car suspension coil spring was examined throughout this relation [27]. Life is projected utilizing the stress-life relationship and the Goodman model in the following equation:

$$G = \left(\frac{R_x / \left(1 - R_s / R_{ul}\right)}{y}\right)^{1/y}. \tag{9}$$

Where x, y are the stress-life coefficients, R_{ul} is the ultimate strength, while R_s and R_x are the average and alternation stresses, respectively. This happens during lifetime testing of the springs The variable y is considered to remain constant at -0.0725 for the sake of demonstration. The unknown variables are then x and R_{ul}. Assume that have 9 life test results for the three different springs of the same materials stated in Table 1 that have been standardized for ease of use. The following is the posterior distribution of the uncertain variables:

$$r(x, R_{ul}, \alpha) \propto \alpha^{-1} \exp\left[-\frac{1}{2\alpha^2} \sum_{c=1}^{9} \left(b_c^e - b_c^a\right)^2\right]. \tag{10}$$

The real lives acquired by the testing as well as the expected life by the FEA are denoted by b_c^e and b_c^a, respectively. The measurement error is multiplied by the unknown variables.

MCMC is used to incorporate the testing findings of the probability distribution function of the uncertain variables of enhanced GLUE. When the two outcomes are compared, the PDF forms are determined to be very similar. With the same structure, the association feature between x and R_{ul} is also visible. Table 1 shows the A1 and A10 life spans, which are in close proximity. Table 2 also includes proportions and confidence intervals (C.I.) for the parameters.

However, it should be noted that the MCMC result was acquired after dozens of trials and errors with regard to the size of proposition density, which is a uniform distribution,

Table 1: Testing surveillance of spring and estimated A1 and A10.

	Data testing			Estimated surveillance		
	1-Real surveillance	2-Real surveillance	3-Real surveillance	A1	A10	Techniques
1st type of spring	0.0327	1.7745	1.3124	1.2198	1.6277	MCMC
				1.1311	1.2585	GLUE
2nd type of spring	15.2246	8.8312	4.2734	6.1225	8.1337	MCMC
				4.0244	6.1157	GLUE
3rd type of spring	1.1780	1.1007	1.1845	1.1548	1.1812	MCMC
				1.1321	1.1514	GLUE

Table 2: Proportions and CI of variables.

Variables	Techniques	Proportion			
		r-10th	Average	r-90th	Confidence interval-95th
x	MCMC	2.2171	2.1418	2.2614	0.3511
	GLUE	1.2783	2.0598	2.2176	0.2869
R_{ul}	MCMC	1.7623	2.4355	2.1787	0.1752
	GLUE	1.3824	2.1245	2.0472	0.0638
α	MCMC	0.1083	0.1156	0.4932	0.1141
	GLUE	0.0769	0.1722	0.2346	0.1581

in terms of achieving convergence. It has taken a very long time. Improved GLUE, on the other hand, achieves the same outcome in just 50 seconds in a single effort. Furthermore, as shown in the following points, MCMC is only successful in this problem with three factors, and it loses whenever there have been more than this quantity. The enhanced GLUE, on the other hand, consistently produces convergence outcomes in just a few tries, despite the number of variables.

5.1.2. The Problem of Crack Growth. The framework utilizes the Paris approach to calculate damaged growth parameters depending on the calculated crack size across a number of cycles. The variables are quite often distributed widely among seemingly identical structures due to several uncertainties. The Bayesian approach is used to modify the distributions of such characteristics over time. The crack size is indicated in measures of the cycles C as per the Paris law shown in the following equation:

$$x(C) = \left[Cn\left(1 - \frac{p}{2}\right)\left(\alpha\vartheta^2\right)^2 + x_i^{1-(p/2)} \right]^{(2/2)-p}. \tag{11}$$

When n and p are the 2 damage growth factors that need to be predicted, x_i is the expected beginning crack size, and α is the stress variation due to fatigue stressing. Assume it has ten large datasets of crack size measurements from a variety of cycles. Based on the preceding data of two variables, the posterior distribution of the uncertain variables is calculated as follows:

$$r(p, n, y, \alpha) \propto \left(\frac{1}{\left(\alpha\vartheta^2\right)^2}\right)^{10} \exp\left[-\frac{1}{2\alpha^2}\sum_{c=1}^{10}\left(x_{mc} - x\left(C_c\right) - y\right)^2\right] r(p) * r(n), \tag{12}$$

Here,

$$\begin{aligned} r(p) &= V_{[2,2,3,2]}, \\ r(n) &= V_{\left[\log[6*10^{-12}], \log[6*10^{-13}]\right]}. \end{aligned} \tag{13}$$

Two variables $x$and y are appended to the uncertainties in (12), which have been the noise and bias measurements biases, respectively. As a result, there are four unknown parameters. Synthesis data are utilized in this problem to investigate the effects of noise and bias. The actual crack diameters are calculated as per (11) for a given C, and the true crack data are purposefully added with a deterministic bias and random noise.

These data will be used to determine the unknown variables. Despite numerous experiments, adequate sampling could not be achieved by using MCMC as depicted in Figure 1 The outcome of the improved GLUE, on the other hand, is better, and it may be reached in around two minutes with one effort.

Figure 2 depicts the enhanced GLUE measures, the produced PDF forms appear to be pretty credible, and the relationship between p and n is plainly visible. The solid blue curve represents genuine crack growth; the 90 percent prediction interval (PI) represents the red dashed line as shown in Figure 3. As predicted, MCMC failed to accurately estimate growth. The improved

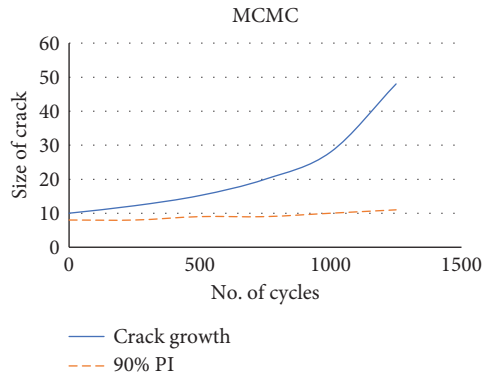


Figure 1: MCMC growth of crack estimation.

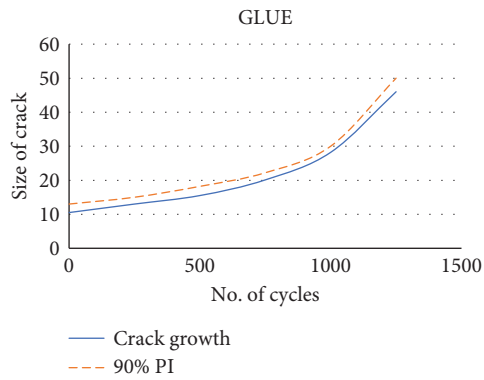


Figure 2: GLUE growth of crack estimation.

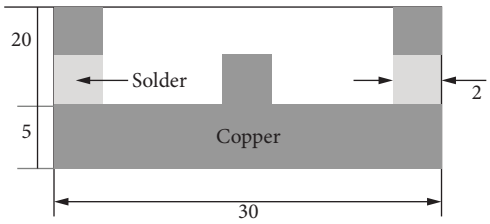


Figure 3: Solder joint sample.

GLUE, on the other hand, accurately forecasts crack growth by addressing the bias and adopting the genuine model.

5.1.3. The Problem of Solder Joint. The final instance is estimating the viscoplastic properties of materials of a solder junction in a microelectronics packaging in reverse. As shown in Figure 4, a solder joint experiment is developed so that the connection deforms similarly to the real packaging. Moire interferometry is used to quantify the distortion. The sample is subjected to viscoplastic FEA using the Anand model [28]. The parameters of the model are correspondingly set such that the anticipated distortion matches the experimental results during five different heating and cooling cycles.

Utilizing likelihood estimates, the approach addresses the uncertainties caused by sampling uncertainty and a lack of appropriate experimental results. Initially, the Anand concept required nine variables; however, after a sensitivity assessment, the set of parameters is decreased to the four that have the greatest influence, while the rest are left as fixed values. The posterior distribution, then again, is as follows:

$$r\left(A_0, \frac{C}{D}, X, \varepsilon, \alpha\right) \propto \left(\frac{1}{2\vartheta\alpha^2}\right)^{15} \exp\left[-\frac{1}{2\alpha^2}\sum_{c=1}^{12}\left(b_c^e - \hat{b}_c^f\right)^2\right]. \tag{14}$$

Here, A_0, (C/D), X, ε, and α are Anand parameters and systematic errors were applied according to the evaluated value. b^e. and $\hat{b}^f$ in the equations represent the solder movement acquired by experimental and analytical models, respectively. The cap symbol in $\hat{b}^f$ signifies the responsive surface model, which is provided by second-order polynomials in terms of the following 4 parameters to substitute the time-consuming FEA. The sample is put through the temperature cycle shown in Figure 4.

The PDFs and the sampled outcomes of the uncertain variables were generated using the enhanced GLUE. Also identified are the connections among A_0 and (C/D), as well as ε and X. Although the presence of two pairs of intricate connections between the variables could affect sample drawing challenges, the findings are produced in around eight minutes [29].

The MCMC solution for this issue could never be produced at all. Figure 5 shows the posterior predictive distributions of movement produced from the sampling data. The blue line containing diagonal marks in the pictures represents the MCMC experimental outcomes. The red line containing the square represents the average of the predictive distribution derived by the enhanced GLUE [30].

Figure 6 shows the posterior predictive distributions of pressure produced from the sampling data. At every occurrence, the 90 percent prediction boundaries are also represented utilizing vertical lines. The results of the FEA based on the existing literature are significantly varied from the experimental data, as seen in the figures. The average of the anticipated distributions gets significantly closer to the experimental values following inversely predicting the properties of materials [31].

5.2. Estimation of MSEP. Considering 1000 or 2000 generated parameter sets, two alternative beginning values, and 5 or 40 data samples, the posterior averages were calculated using the MCMC and GLUE methods. Table 3 illustrates the MSEP results calculated by fixing the parameters of the model to the posterior means. The performance was acquired for 100 samples of data, not for one specific sample of data points X. The vectors of the posterior variable distribution's averages$\overline{\alpha}$ were calculated for every sample by using the GLUE technique [32]. (8) was then used to determine the MSEP value assigned to every $\overline{\alpha}$ value. Lastly, the individual MSEP results were used to estimate the MSEP's predicted values over the 100 data samples.

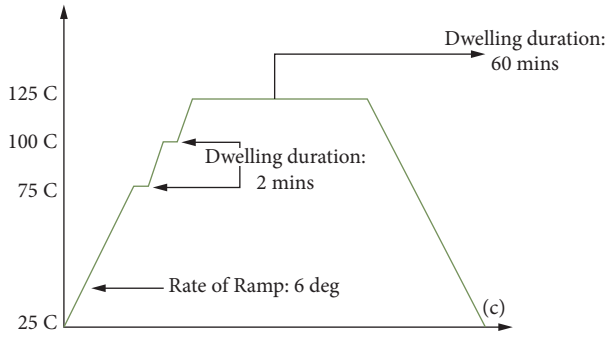


Figure 4: Temperature cycle.

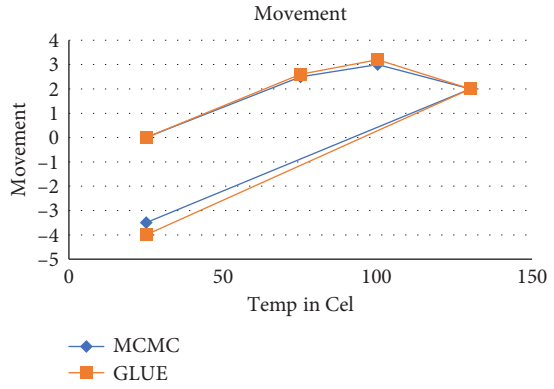


Figure 5: Preceding estimation of movement.

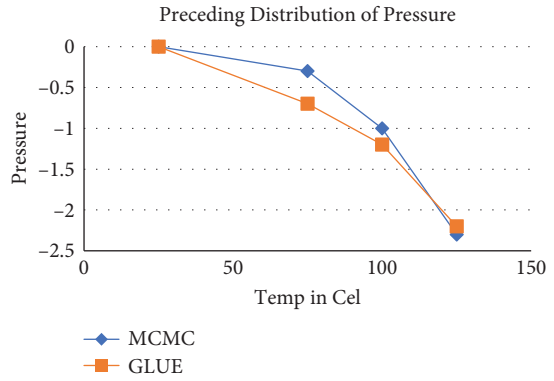


Figure 6: Preceding estimation of pressure.

Table 3: MSEP prediction analysis.

Amount of data	Techniques	Amount of iterations	Initial parameters	$\hat{e}(\mathrm{MSEP}(\bar{\alpha}))$
Beginning of the preceding parameter distribution				
0	—	—	—	15.32
6	MCMC	1000	2	4.91 (1.3)
6	GLUE	1000	—	2.52 (0.31)
6	MCMC	2000	1	3.72 (0.56)
6	GLUE	2000	—	4.21 (0.28)
50	MCMC	1000	2	1.33 (0.5)
50	GLUE	1000	—	1.33 (0.33)
50	MCMC	2000	2	1.17 (0.52)
50	GLUE	2000	—	1.20 (0.06)
Preceding parameters having lower variations				
6	MCMC	1000	2	4.53 (1.23)
50	MCMC	1000	2	2.92 (0.03)

The MSEP of the models is 17.17 once the requirements are identified to their preceding means. That value is significantly greater than the remaining error variance α^2, which is equivalent to 0.16 and indicates the model's smallest possible MSEP, so that it is the MSEP with the actual attribute values. Because when estimated values are set to the averages of the posterior parameter distribution obtained utilizing MCMC or the GLUE technique, the MSEP is significantly reduced. The MCMC and 2000 iterations get the least MSEP value (0.21). This result is quite close to the minimum MSEP value, which is 0.16. The MSEP has no meaningful effect on the beginning values. Once six data samples have been used, the computational time has a minor impact on the MSEP value, however, when fifty data samples are used, which has no impact. The Bayesian approach used to construct the posterior parameter distributions has little effect on the MSEP value. Whenever the number of data samples is 50, the predicted MSEP values produced using the GLUE approach are marginally greater [33]. The prior parameter distribution, on the other hand, has a significant impact on MSEP values. The anticipated MSEP results calculated with 6 and 50 data samples are 5.67 and 1.86, respectively, whenever the previous distribution having reduced variations is employed. Such MSEP is significantly greater than the actual preceding distribution values. That's because the true values of parameters are not included in the parameter space represented by the prior distribution having lower variations. As a result, even when a large set of data sets are employed, the posterior means of such variables generally remain distinct from the genuine parameter values.

6. Conclusion

Parameter estimation of probability density functions is one of the most significant processes in statistical image and signal processing (PDF). In the fields of artificial intelligence and machine learning, estimating the probability density function is a contentious topic. The research focuses on difficulties in estimating a density function from its random variable. When only minimal density forecasts are given, the paper presents a framework for evaluating probability density functions. Sampling is commonly used to estimate and describe the probabilistic model of unknown variables in the Bayesian approach, which is to provide correct samplings that reflect the probability aspect of the variables. The generalized likelihood uncertainty estimation (GLUE) approach has gained the most popular among the numerous methodologies due to its efficacy and wide application. The Bayesian method allows model parameters to be calculated based on prior knowledge of parameter outcomes and experimental observations. To address these difficulties, the investigation offers a strategy that employs the marginal probability density function (PDF) as a distribution to demonstrate the utility of the proposed method; specific engineering problems created using the Bayesian methodology are addressed. Despite this, the model's mean squared error of prediction is significantly larger when employing the GLUE technique than with the preceding algorithms. The results of the approaches are influenced by the assumptions made about parameter values in advance. The ideas of learning and prediction accuracy are discussed, as well as the value of a statistical experiment. These ideas are useful to demonstrate that the GLUE technique defines an incomplete statistical inference process. Despite the fact that numerous sampling techniques have been created, there is not a reliable, affordable solution that could address real-world issues in an engineering application. The most successful sampling technique, known as GLUE, has the drawback of becoming ineffective as the amount of parameters rises and the values become more closely connected. Prior to the primary GLUE technique, this method requires additional computation for marginal PDF creation. Instead, better resilience and efficiency are attained by such an approach since it always obtains convergent samples, that is, the identical samples when the technique is applied that were not achievable while using MCMC. In summary, a very useful technique to handle parameter estimation in the scenario of higher parameters as well as correlations is the GLUE combined with the marginal distribution, which is discussed. The numerous real-world engineering issues where sufficient samples are gathered for the posterior distribution serve as evidence for the method's viability.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] D. An and J.-H. Choi, "Improved MCMC method for parameter estimation based on marginal probability density function," *Journal of Mechanical Science and Technology*, vol. 27, no. 6, pp. 1771–1779, 2013.

[2] D. L. Kelly and C. L. Smith, "Bayesian inference in probabilistic risk assessment—the current state of the art," *Reliability Engineering & System Safety*, vol. 94, no. 2, pp. 628–643, 2009.

[3] H. G. C. Traven, "A neural network approach to statistical pattern classification by 'semiparametric' estimation of probability density functions," *IEEE Transactions on Neural Networks*, vol. 2, no. 3, pp. 366–377, 1991.

[4] I. Park, H. K. Amarchinta, and R. V. Grandhi, "A Bayesian approach for quantification of model uncertainty," *Reliability Engineering & System Safety*, vol. 95, no. 7, pp. 777–785, Jul. 2010.

[5] V. A. Krylov, G. Moser, S. B. Serpico, and J. Zerubia, "On the method of logarithmic cumulants for parametric probability density function estimation," *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 3791–3806, 2013.

[6] R. K. Jana and M. P. Biswal, "Stochastic simulation based genetic algorithm for chance constraint programming problems with some discrete random variables," *International Journal of Computer Mathematics*, vol. 81, no. 12, pp. 1455–1463, 2004.

[7] Z.-W. An, H.-Z. Huang, and Y. Liu, "A discrete stress–strength interference model based on universal generating

function," *Reliability Engineering & System Safety*, vol. 93, no. 10, pp. 1485–1490, 2008.

[8] S. Kay, "Model-based probability density function estimation," *IEEE Signal Processing Letters*, vol. 5, no. 12, pp. 318–320, Dec. 1998.

[9] M. Ren, Q. Zhang, and J. Zhang, "An introductory survey of probability density function control," *Systems Science & Control Engineering*, vol. 7, no. 1, pp. 158–170, 2019.

[10] K. Beven and J. Freer, "Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology," *Journal of Hydrology*, vol. 249, no. 1-4, pp. 11–29, 2001.

[11] R.-S. Blasone, J. A. Vrugt, H. Madsen, D. Rosbjerg, B. A. Robinson, and G. A. Zyvoloski, "Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov chain Monte Carlo sampling," *Advances in Water Resources*, vol. 31, no. 4, pp. 630–648, 2008.

[12] K. Beven, "How far can we go in distributed hydrological modelling?" *Hydrology and Earth System Sciences*, vol. 5, no. 1, pp. 1–12, 2001.

[13] J. Freer, K. Beven, and B. Ambroise, "Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach," *Water Resour. Res.*vol. 32, no. 7, pp. 2161–2173, 1996.

[14] G. Kuczera and E. Parent, "Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm," *Journal of Hydrology*, vol. 211, no. 1-4, pp. 69–85, 1998.

[15] X. Jin, C.-Y. Xu, Q. Zhang, and V. P. Singh, "Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model," *Journal of Hydrology*, vol. 383, no. 3-4, pp. 147–155, 2010.

[16] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC," *IEEE Transactions on Signal Processing*, vol. 47, no. 10, pp. 2667–2676, 1999.

[17] D. C. Haworth, "Progress in probability density function methods for turbulent reacting flows," *Progress in Energy and Combustion Science*, vol. 36, no. 2, pp. 168–259, 2010.

[18] N. Mordant, A. M. Crawford, and E. Bodenschatz, "Experimental Lagrangian acceleration probability density function measurement," *Physica D: Nonlinear Phenomena*, vol. 193, no. 1-4, pp. 245–251, 2004.

[19] G. Vachaud, A. Passerat De Silans, P. Balabanis, and M. Vauclin, "Temporal stability of spatially measured soil water probability density function," *Soil Science Society of America Journal*, vol. 49, no. 4, pp. 822–828, 1985.

[20] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[21] Q. Shi and Y. Karasawa, "Automatic modulation identification based on the probability density function of signal phase," *IEEE Transactions on Communications*, vol. 60, no. 4, pp. 1033–1044, 2012.

[22] C. Heye, V. Raman, and A. R. Masri, "LES/probability density function approach for the simulation of an ethanol spray flame," *Proceedings of the Combustion Institute*, vol. 34, no. 1, pp. 1633–1641, 2013.

[23] F. S. Al-Duais, "Bayesian estimations under the weighted LINEX loss function based on upper record values," *Complexity*, vol. 2021, Article ID 9982916, 7 pages, 2021.

[24] M. T. Muronda, S. Marofi, H. Nozari, and O. Babamiri, "Uncertainty analysis of reservoir operation based on stochastic optimization approach using the generalized likelihood uncertainty estimation method," *Water Resources Management*, vol. 35, no. 10, pp. 3179–3201, 2021.

[25] M. Mirzaei, Y. F. Huang, A. El-Shafie, and A. Shatirah, "Application of the generalized likelihood uncertainty estimation (GLUE) approach for assessing uncertainty in hydrological models: a review," *Stochastic Environmental Research and Risk Assessment*, vol. 29, no. 5, pp. 1265–1273, 2015.

[26] R. Harmon and P. Challenor, "A Markov chain Monte Carlo method for estimation and assimilation into models," *Ecological Modelling*, vol. 101, no. 1, pp. 41–59, 1997.

[27] D. An, J.-H. Choi, N. H. Kim, and S. Pattabhiraman, "Fatigue life prediction based on Bayesian approach to incorporate field data into probability model," *Structural Engineering & Mechanics*, vol. 37, no. 4, pp. 427–442, 2011.

[28] A. Rayan, A. I. Taloba, M. Rasha, and A. Amr, "IoT enabled secured fog based cloud server management using task prioritization strategies," *International Journal of Advanced Research in Engineering & Technology*, vol. 11, p. 9, 2020.

[29] G. Z. Wang, Z. N. Cheng, K. Becker, and J. Wilde, "Applying Anand model to represent the viscoplastic deformation behavior of solder alloys," *Journal of Electronic Packaging*, vol. 123, no. 3, pp. 247–253, 1998.

[30] A. I. Taloba and M. R. Riad, "Developing an efficient spectral clustering algorithm on large scale graphs in spark," in *Proceedings of the 2017 8th International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 292–298, IEEE, Cairo, Egypt, December 2017.

[31] A. I. Taloba, "An artificial neural network mechanism for optimizing the water treatment process and desalination process," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9287–9295, 2022.

[32] S. S. I. Ismail, R. F. Mansour, R. M. Abd El-Aziz, A. I. Taloba, and A. I. Taloba, "Efficient E-mail spam detection strategy using genetic decision tree processing with NLP features," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 7710005, 16 pages, 2022.

[33] M. Elloumi, M. A. Ahmad, A. H. Samak, A. M. Al-Sharafi, D. Kihara, and A. I. Taloba, "Error correction algorithms in non-null aspheric testing next generation sequencing data," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9819–9829, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 3089239, 12 pages
https://doi.org/10.1155/2022/3089239

Research Article

An Empirical Analysis of Social Interaction in Online Teaching in Open Universities Based on Flipped Classroom

Xinjie Wang[1] **and Ding Yuan**[2]

[1]*College of Humanities and Education, Beijing Open University, Beijing, China*
[2]*Beijing Mycourse Online Educational Technology Co. Ltd, Beijing, China*

Correspondence should be addressed to Xinjie Wang; wangxj@bjou.edu.cn

Received 31 August 2022; Accepted 24 September 2022; Published 11 October 2022

Academic Editor: Heng Liu

As a teaching model, flipped classroom can stimulate students' learning interests and promote students' positive social interactions. As an adult university relying on information technology to carry out online teaching, open universities need to pay attention to the online learning quality of adult students. This study attempts to apply the concept of flipped classroom to online teaching in open universities. While exploring the feasibility of the online flipped classroom teaching model, it adopts an equal group experiment method, content analysis method, and social network analysis method to compare the social interaction level difference between online teaching based on flipped classroom and conventional online teaching. The results show that online flipped classroom has a positive and significant impact on student's cognitive processing level and knowledge construction level and can promote groups to form complex knowledge network forms. This shows that network teaching based on flipped classroom can exercise students' high-order thinking ability, enhance students' cooperative consciousness and behavior, and stimulate students' subjectivity in learning. The purpose of this study is to provide references for promoting interactive behavior and quality in online teaching.

1. Study Background

In 2021, the outline of the National 14th Five-Year Plan once again proposed to "give full play to the advantages of online education and build a society in which people enjoy learning." As a new type of university that leverages information technology to carry out distance education for working adults, open universities are expected to innovate the approaches to facilitation of learning and improve the effectiveness of online teaching (JZC [2016] no.2). In an online learning environment, social interaction is a manifestation of mutual influence and mutual role playing between learners and resources, teachers, and peers. Quality, in-depth interactions are required to enhance the effectiveness of teaching and achieve in-depth learning [1]. However, currently, in online teaching at open universities, desired classroom behaviors such as active collaboration, focused exploration, and in-depth communication among students are seldom seen in the interaction sessions. How to break such a deadlock, promote interactive participation among students, and improve the quality of interaction? This is the vision and need for open universities to improve the quality of online education and teaching.

The flipped classroom originated in the United States. It is designed to establish a classroom centered on students' learning so that classroom-based learning can be upgraded from superficial to in-depth learning and develop the higher-order cognition capacity of students [2]. Studies have shown that flipped classroom can mobilize students to actively participate in classroom activities, facilitate students to learn actively in a collaborative way, and improve their academic performance [3]. Therefore, applying flipped classroom to online teaching at open universities can be expected to enhance the depth and breadth of student interactions, thus improving learning outcomes. In order to validate this hypothesis, this study attempted to design a flipped

classroom-based online teaching model and applied it to the course "Kindergarten Management" at Beijing Open University for an empirical study.

2. Relevant Studies

2.1. Definition of the Concept

2.1.1. Flipped Classroom. There are different interpretations of flipped classroom in Chinese and international academic circles, which highlight different characteristics of flipped classroom. The first interpretation sees flipped classroom as a student-centered teaching approach that emphasizes teachers' use of heuristic, exploratory, and discussion-based teaching methods in the classroom to motivate students to acquire knowledge through collaborative learning and active learning [4]. The second considers flipped classroom as a new learning model realized via information technology that extends conventional classroom activities outside the classroom [5]. The third focuses on the value of flipped classroom in terms of knowledge internalization, which increases the instances of knowledge internalization during active interactions between teachers and students and among students in the classroom based on time series reconstruction, in order to achieve in-depth learning [6]. The above studies, respectively, reveal the vision, methods, environment, and outcomes of teaching through flipped classroom. Generally speaking, the vision for teaching through flipped classroom is student-centered, which can leverage information technology to expand the learning environment and promote students' acquisition and application of in-depth knowledge through appropriate and orderly self-directed and collaborative learning. This study focuses on the design of teaching based on the concept and value of flipped classroom, i.e., the classroom is centered on students' questions, with diverse learning activities such as heuristic, exploratory learning activities, and teachers stimulate the intrinsic motivation for learning by guiding students to communicate and collaborate, share knowledge, and others, ultimately resulting in the teaching process composed of three stages: self-directed basic learning, exploratory in-depth learning, and reflective enhanced learning.

2.1.2. Social Interaction. Social interaction is a key element of distance education, which enables learners to achieve individualized interaction with learning resources and communication with teachers and other learners in the online learning environment [7], which is conducive to the joint development and sharing of learning knowledge, as well as cognitive enhancement [8]. Bates [9] first proposed that social interaction is the interaction of learners with teachers and other learners about learning [9]. Based on Bates' classification theory, Chen Li, a Chinese scholar, further pointed out that social interaction in the field of distance learning means teacher-student information exchange and student-student information exchange [10]. Only those interactions through which learners actively share, collaborate, and enhance their cognition capacity can facilitate effective learning.

Social interaction can be categorized into synchronous interaction and asynchronous interaction according to its timeliness. Asynchronous interaction is the popular form of social interaction for adult learners, which provides learners with a space for discussion that is relatively free in terms of time and space and records the content of discussion thoroughly through the online platform, thus triggering learners' in-depth thinking [11]. In summary, this study focuses on social interaction in an asynchronous interaction environment, which refers to the information exchange and discussion between students, teachers, and peers through asynchronous interaction in a classroom-based learning environment in order to construct knowledge structures relevant to course learning. Specifically, it includes interactive behaviors such as teachers and students expressing their views, asking questions, and giving feedback in a discussion forum on an online learning platform.

2.2. Analysis of the Effect of Flipped Classroom Teaching.

The core idea of flipped classroom, as a major revolution affecting classroom teaching, is to turn the conventional classroom into an interactive setting. Compared with the conventional teaching model, the model of flipped classroom teaching can have a positive impact on learners' learning performance, cognitive skills, and emotions/attitudes towards learning, which is conducive to improving learning outcomes; the stronger the learners' autonomy, initiative, willpower, and control, the better the learning outcomes [12]. Meanwhile, there are also studies concluding that small-scale teaching and humanities courses can be better taught using flipped classroom [13, 14]. Subsequently, more studies attempted to combine flipped classroom with other teaching models in order to take advantage of their respective strengths to jointly enhance teaching effectiveness. For example, the combination of microlectures and MOOCs with flipped classroom solves the problem of work-learning contradiction among adult learners, enabling them to learn anytime and anywhere, and effectively leverages face-to-face teaching sessions to organize student discussion, which increases teacher-student interactions and improves the efficiency of learning [15, 16]. Besides, the measurement and evaluation of the effectiveness of flipped classroom teaching mostly focus on three aspects, namely, students' learning attitude, learning process, and learning outcome. Learning attitude mainly examines whether students voluntarily and actively participate in learning activities, learning process mainly examines students' development of capability in communication and interaction, collaboration, thinking, and other cognitive skills in the classroom, and learning outcome refers to students' academic performance [1, 12, 17].

There have been studies testifying the applicability and effectiveness of flipped classroom and describing the advantages and effectiveness of flipped classroom. However, the majority of previous studies simply examined face-to-face teaching in primary and secondary schools and regular

higher education institutions and seldom focused on online teaching in open universities. Among the few studies that examined teaching in flipped classroom for adult students in open education settings, most simply focused on the design and evaluation of the teaching models, and there are limited empirical studies examining the teaching outcomes.

2.3. Associations between Classroom-Based Learning and Social Interaction. According to Vygotsky (1978), learners need to interact with people in their environment in order to awaken a variety of internal developmental processes [18]. Thus, social interaction is an important precondition for learning to take place. The learning theory of social constructionism views learning as an active process that occurs in concrete environments and interpersonal relationships, where knowledge is constructed through language and other social tools and where the meaning of knowledge is verified through discussion and application [19]. Therefore, in classroom-based learning, teacher-student interaction is mainly about students receiving guidance and assistance from the teacher, while student-student interaction is mainly about communication and collaboration in order to complete a certain task [20]. In this process, the more students proactively communicate and collaborate, initiate profound cognitive processing and involvement, and engage in meaningful construction, the more effective the learning will be.

Connectivism learning emphasizes the establishment of personal neural networks, conceptual networks, and social networks, and essentially only complicated connections that can yield knowledge innovations are true connectivism learning [13]. In connectionist classroom-based learning, the teacher, as a node in the overall learning network, plays the role of facilitating, enhancing, and maintaining the classroom environment and knowledge network. Student-student interaction is the process of social network formation in the course of classroom-based learning, followed by the process of collective knowledge innovation. Generally speaking, more complicated network construction and stronger connectivity mean a higher level of students' participation, more knowledge innovation based on such network, and better outcomes of collective learning.

2.4. Studies on the Methods for Evaluation of Social Interaction Quality. The evaluation of the quality of social interaction is mostly based on a combination of the content analysis method and the social network analysis method. The content analysis method involves qualitative analysis of the quality of social interaction content [21], which usually leverages a coding scale with multidimensional and multilevel gauges to analyze the quality of the content of interaction to reveal the pattern of interaction [22], including Henri's five-dimensional framework for cognitive learning, Stacy's collaborative learning framework, Gunawardena's interaction analysis model based on the five-phase social construction of knowledge, and others[23]. Bloom's taxonomy of learning objectives is also commonly used to examine the depth of instructional interactive textual content and the learners' cognition process patterns and levels. Social network analysis is a method to analyze the representation and characteristics of social networks by examining the interactive relationships among different actors [24]. UCINET, a social network analysis software with powerful matrix analysis functions and support for generating various visual maps, can effectively reveal the characteristics of the overall interaction network of learners and the roles played by individuals in the network [25].

3. Study Design and Methodology

3.1. Study Sample. The study sample was composed of 77 working adult students taking the preschool education course "Kindergarten Management" at Beijing Open University, as well as one instructor. All the students were newly enrolled by the university in the fall of 2018, and most were working at kindergartens with some experience in preschool education. The course "Kindergarten Management" was the first specialty course that the students have been exposed to since they were enrolled; relevant professional-level tests were conducted before the class, and the results showed that there was little difference in the student's prior knowledge and experience in the subject and familiarity with the operation of the learning platform. The instructor has been trained in the expertise related to the course and the learning platform and has been tutoring the course for five years.

3.2. Study Design

3.2.1. Design of Flipped Classroom Teaching. The course "Kindergarten Management" at Beijing Open University has been implemented with flipped classroom teaching design. The course was developed on the Moodle platform in the fall of 2013 and has been implemented several times since then, with a robust practice basis. The course is taught online for six weeks, with each week being an independent learning module. The learning resources include videos and texts, the learning activities are online discussions, and the learning appraisal approaches include tests, forums, and assignments. Since this study focuses on the quality of students' social interactions, the weeks when forum activities were organized, i.e., the third and sixth weeks, were selected for the implementation of flipped classroom. The design is based on the concept of "student-centered learning" with learning tasks throughout the teaching process and focuses on the internalization in the classroom through three phases of teaching, namely, "self-directed basic learning," "exploratory in-depth learning," and "reflective enhanced learning." In such a design, superficial knowledge precedes in-depth knowledge, and the in-depth knowledge is taught through heuristic dialogue activities and exploratory extension activities. The specific operating procedures are shown in Figure 1.

Phase 1: self-directed basic learning, focusing on students' acquisition of superficial knowledge. The teacher presents the learning tasks and timeline for the module before the start of the course module. On the course learning platform, the teacher designs and develops relevant learning

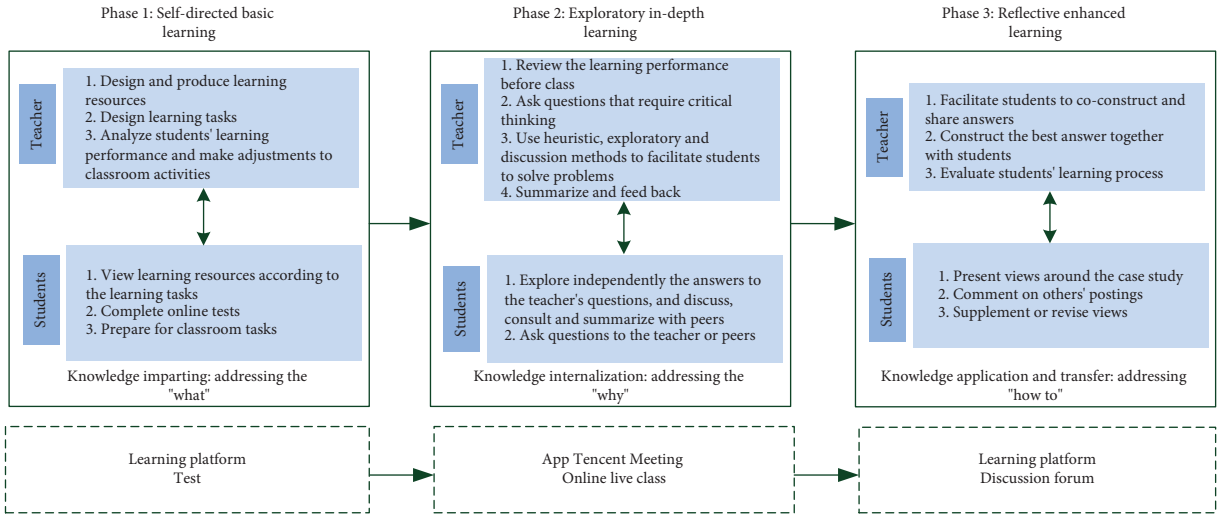


Figure 1: Procedures of online teaching based on flipped classroom.

resources, including video and text learning materials, posts test questions, observes results, identifies issues with students' learning, and adjusts the content of classroom teaching. This phase allows teachers to avoid a "one-size-fits-all" didactic teaching style in the process of knowledge imparting and to integrate the current stage of learning with the next stage of learning by understanding the current knowledge level of students so that the learning can be more targeted.

Phase 2: exploratory in-depth learning, focusing on students' acquisition of in-depth knowledge. Teacher and students enter the same video conference at the same time using the app "Tencent Meeting" and participate in at least one hour of intensive learning around the learning topic of the week. During such sessions, the teacher will make full use of conversation, scenarios, collaboration, and other elements to bring into play students' initiative and help them internalize their knowledge. The specific procedure is to analyze and review the common, complicated problems in the learning process based on the feedback from the prelesson test. The teacher then facilitates a heuristic dialogue session around the students' current problems and the week's learning topic. For example, based on the content of the learning topic, students are asked to think about why they want to learn about the topic and how the content is related to their work, and then, they are facilitated to discuss with other students the possible answers to questions raised based on their own work experience. In this phase of learning, students are familiar with the learning content and there are possibilities for them to explore, which can stimulate their initiative and motivation to participate in learning, as well as exercise their self-reflection and communication/expressing skills. It should be noted that the questions which the teacher gives to students at this stage are not related to the discussion in the forum at the next stage and students will not be encouraged to actively participate in the subsequent forum discussions.

Phase 3: reflective enhanced learning, focusing on students' acquisition of in-depth knowledge. In this phase, the teacher assigns forum discussion on the learning platform. The forum topics introduce case studies relevant to students' work through scenarios, and students are asked to leverage what they have learned in the week to analyze and solve the problems in the cases and propose recommendations for improvement. The forum activities require students to consolidate their knowledge through communication, collaboration, and sharing and learn to use their knowledge to solve actual problems. In this process, the teacher selects and edits the cases by referring to the knowledge covered in the previous phase. For example, if questions of "what? why?" are asked in the previous phase, the case should present the "what" and "why" in a story scenario so that students can apply their internalized knowledge. During the forum activities, teachers need to guide students to think and encourage their communication and collaboration.

3.2.2. Design of Equal Group Experiment. Using the natural experiment method, the sample of 77 students was randomly divided into two groups, with 39 students in the experimental group implementing a flipped classroom-based approach and 38 students in the control group following the conventional approach. The difference between the two groups was primarily articulated in the sequence of teaching phases. Students in the experimental group were required to complete the learning tasks in the order of the three phases depicted in Figure 1, while students in the control group were free to choose the sequence of learning tasks according to their individual needs. Secondly, the teaching process in the second phase was different, as the experimental group was taught through exploratory and collaborative learning activities, while the control group was mainly taught by a teacher who directly imparted knowledge.

Both models were implemented online through the same app and course platform. The course resources, quiz questions, forum activities, and assessment criteria were the same for both groups, and they were taught by the same instructor. The instructor refers to the same teaching

syllabus; that is, the teaching objectives and content of the two models are relatively consistent. The instructor used a "sandwich" approach to facilitate forum activities for both groups of students, first affirming students' responses, then proposing modifications, and finally encouraging students to browse other students' postings. The experimental environment was kept as consistent as possible. The experiment lasted for six weeks, i.e., from the beginning to the end of the course. The study was based on the following three hypotheses:

H1: implementing an online flipped classroom helps improve students' capacity in cognitive processing

H2: implementing an online flipped classroom helps to improve students' knowledge construction

H3: implementing an online flipped classroom helps establish complicated knowledge network patterns

3.3. Data Analysis Tools and Methods. Since this study primarily examines social interactions in the context of asynchronous interactions, changes in students' cognition and construction levels are manifested in the quality of their cognitive processing process and the quality of meaning negotiation and knowledge coconstruction processes in the course forum, which can be measured by content analysis method. The student group knowledge network is manifested in the process of establishing close connections with peers, and the network density in the course forum can be measured by the social network analysis method. In the study, data analysis was performed to evaluate the quality of forum interactions by using SPSS 20.0 and NVivo 11.0, and UCINET 6.0 was used to measure the density of the forum network.

4. Experiment Results

4.1. Students' Cognitive Engagement Levels. The taxonomy of learning objectives proposed by American educator Benjamin Bloom defines the educational objectives in the cognitive domain as memorization, comprehension, application, analysis, evaluation, and creation. This represents the process of cognitive engagement from superficial to in-depth level and the process of cognition capacity development from low to high level. This study analyzes the content depth of student-student interaction texts to understand students' in-depth knowledge cognition capacity based on Bloom's taxonomy of learning objectives and drawing on the cognitive processing behavior coding framework of Chen Beilei et al. [26], as shown in Table 1. By uniformly coding the content text data posted by students in the case study activities in the course forum, the cognitive processing behavior score of individual student was obtained, which represented the highest cognition capacity reached by the content of his/her post, e.g., a student whose post reached the levels of memorization, comprehension, and analysis would have a score of 4. Each student's score on each cognitive processing behavior was calculated.

After independent sample *t*-tests were conducted on the cognitive processing behavior level scores of all students in the experimental and control groups, the results of the analysis are shown in Table 2.

As shown in Table 2, there was a statistically significant difference in cognition capacity between the experimental group and the control group ($p = 0.000 < 0.01$). Specifically, under the influence of flipped classroom, the cognition capacity of the experimental group students was significantly higher than that of the control group students and H1 was verified.

Further coding of 501 post-based units at the cognitive structure level revealed that the proportion of the two superficial cognitive processing behaviors, i.e., memorization and comprehension, was 21.66%, and the proportion of the four in-depth cognitive processing behaviors, i.e., application, analysis, evaluation, and creation, was 56.95% among the students in the experimental group. In contrast, the proportions were 52.57% and 23.63%, respectively, among students in the control group. The proportion of interactive content not related to the topic in the experimental and control groups was 21.39% and 23.62%, respectively, as shown in Figure 2. This shows that the implementation of flipped classroom enables students not only to memorize and understand the factual attributes and conceptual information of knowledge but also to process it to a greater extent to solve real-world problems, develop a theoretical system by analyzing the problems and reflecting on the principles and values of knowledge, develop higher-order thinking skills such as critical thinking and creative thinking, and thus enhance the depth of cognition.

4.2. Students' Knowledge Construction Levels. The level of construction of student-student interaction is reflected by the phases of knowledge construction. In this study, the content text data of student postings in the case study activities in the course forum were coded uniformly based on Gunawardena's cue-based retrieval model [27], while considering the content meaning, i.e., there might be more than two codes for a posting. The codes were then matched with different interaction phases based on the content of the posting, the social construction level of each member was determined by the highest phase reached by the content of their interaction, and the construction levels were assigned the value of 1, 2, 3, 4, or 5 in a hierarchical order, from low to high, respectively [28]. An individual student's knowledge construction level score was the highest interaction phase he/she reached; e.g., a student whose posting content reached phase 1, phase 3, and phase 5 had a score of 5. The higher the student's score, the higher the quality of his/her meaning construction in that forum and hence the higher the level of construction. In this study, the content of interactions other than knowledge construction was uniformly coded as "other," specifically, meaningless "like," recognition, encouragement, and others towards others' posting, and was assigned a score of 0. See Table 3 for details.

Independent sample *t*-tests were conducted on all students' knowledge construction level scores in the

Table 1: Cognitive processing behavior coding framework.

Indicator for coding	Description	Score
Memorization	Retrieval of relevant knowledge or facts from long-term memory	1
Comprehension	Apprehension of knowledge, relating new knowledge to existing knowledge or previous experiences, and constructing the meaning	2
Application	Application in order to complete a task or solve a problem in a scenario	3
Analysis	Break down knowledge into its parts, analyze each part, and reveal the relationship between each part and the whole	4
Evaluation	Make judgments or comments based on certain principles and criteria	5
Creation	Combine different elements and create a complete, functional whole, and reorganize the elements into a new pattern or structure through the thinking process	6
Other	Interactive content not related to the topic	0

Table 2: Comparison of cognition capacity of the experimental group and the control group.

Dimension	Variable	Group	Sample size	Mean	Standard deviation	t
Cognition capacity	Flipped classroom	Experimental group	39	3.5385	1.3148	5.065**
		Control group	38	2.2105	0.9630	

Note: $^{*}p < 0.05$; $^{**}p < 0.001$.

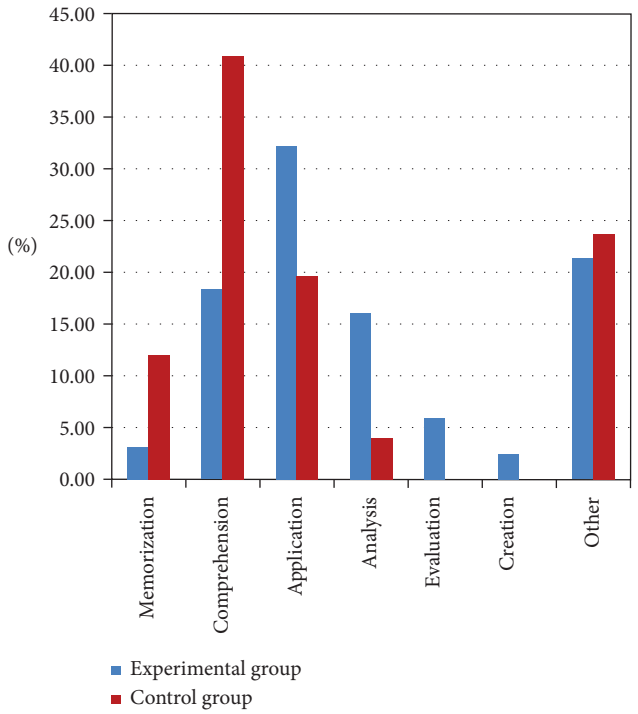


Figure 2: Characteristics of students' cognitive processing behaviors in the experimental and control groups.

Table 3: Coding based on Gunawardena's cue-based retrieval model.

Interaction phase	Indicator for interaction content coding	Score
Phase 1	Present and share views	1
Phase 2	Express disagreement	2
Phase 3	Discuss and present new views	3
Phase 4	Test and revise views	4
Phase 5	Members reach a consensus and apply new constructs	5
Other	Meaningless "like" or recognition	0

experimental and control groups. The results of the analysis are shown in Table 4.

As shown in Table 5, there was a statistically significant difference in the level of construction between the experimental and the control group students ($p = 0.000 < 0.01$). Specifically, under the influence of flipped classroom, the experimental group students had a significantly higher level of construction than the control group students and H2 was verified.

Further coding of 501 post-based units at the construction phase level revealed that 48.13% of the students in the experimental group posted in the first and second phases of the low-level construction, with interactive information involving observations or opinion statements and questioning others' views. In contrast, the percentage of postings in the third, fourth, and fifth phases of high-level construction was 30.48%, and the interactive information involved expressing their own views or modifying views about others' opinions and updating their own views, and others. In the control group, the proportion of postings in the first and second phases was 56.70% and the proportion of postings in the third to fifth phases was 19.69%. The percentages of "other" postings in the two groups were 21.39% and 23.62%, respectively, as shown in Figure 3. These show that flipped classroom can help students learn the learning method of communication and collaboration with peers. In the process of knowledge construction, students can not only express their own views but also pay attention to others' views, make judgment, analysis, and reflection based on others' knowledge and experience, and then adjust their own conceptual system and experience system to achieve a meaningful construction process.

Table 4: Comparison of the level of knowledge construction between the experimental group and the control group.

Dimension	Variable	Group	Sample size	Mean	Standard deviation	t
Construction level	Flipped classroom	Experimental group	39	3.0256	1.49538	4.526**
		Control group	38	1.7059	0.97014	

Note: $^{*}p < 0.05$; $^{**}p < 0.01$.

Table 5: Social network attributes of students in the experimental and control groups.

Teaching model	Network density		Number of nodes	Number of connections	Reciprocity (%)
	Mean	Standard deviation			
Experimental group	0.2744	0.7674	40	428	21.3
Control group	0.1066	0.4977	39	158	6.59

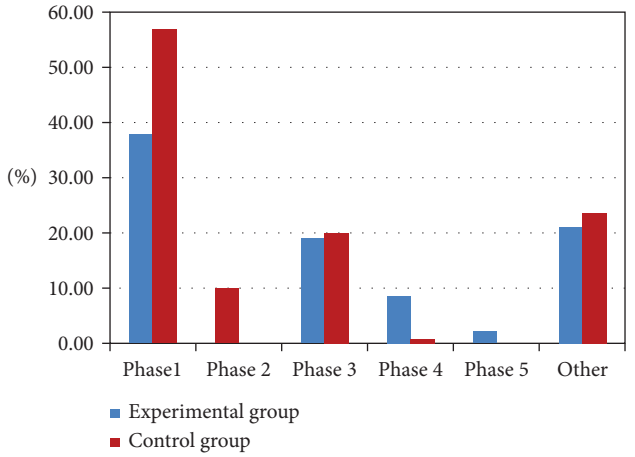


Figure 3: Characteristics of students' construction behaviors in the experimental and control groups.

4.3. Students' Knowledge Network Patterns. Density is commonly measured to understand the density of social relationships and the evolution trend in online social networks, which is the ratio of the number of relationships actually present in the network to the maximum number of relationships that can be accommodated; the greater the overall network density, the stronger the connections between network members, i.e., the more complex the pattern of interaction behaviors [28]. In this study, data on the echo relations between students' postings in the course forum in the experimental and control groups were generated, and a relationship matrix was drawn. Based on the two interaction adjacency matrices obtained, a knowledge network diagram of student interactions in different teaching models was drawn using NetDraw, a visualization tool in UCINET. It should be noted that in this study, only postings with explicit echo targets were counted in the statistics, and postings without explicit echo targets or announcement-type postings targeting all forum members were not counted. The results of the analysis are shown in Table 5.

The results show that there were 39 students and 1 instructor in the course forum of the experimental group, and there were 428 pairs of relationships in the interaction process; the network witnessed 27.44% of network connections, and 21.3% of the connections were two-way interactions. In the control group, there were 38 students and 1 teacher in the course forum, and 158 pairs of relationships existed in the interaction process; the network witnessed 10.66% of network connections, and only 6.59% of the connections were two-way interactions. This shows that under the influence of flipped classroom, the knowledge network patterns formed among the members are more complicated, with stronger connectivity, suggesting that there are more knowledge innovation and better collective learning based on such network and H3 was verified.

Further analysis of the community mapping and the individual degree centrality in the network revealed that the knowledge network pattern formed by the experimental group was more complicated, with stronger connectivity. The teacher and student nos. 32, 20, 7, 22, and 38 were at a more central position in the network and exhibited some initiative. The out-degree of the teacher was 140 and the in-degree was 16, indicating that the teacher gave feedback to almost all students' postings, and in turn, some students responded to the teacher's feedback. The means of individual out-degree and in-degree of other core members were 13 and 15, respectively, indicating that although the core members did not respond to each of their peers' postings as the teachers did, they were trying to follow and give feedback to peers and their behaviors increasingly resembled the "teacher role." In the control group, only the teacher was in the central position and was the dominant player in the network. The teacher's out-degree was 94 and the in-degree was 2, suggesting that the teacher responded to students' postings, while the students seldom responded to the teacher. Other students had a mean in-degree of about 2 and a mean out-degree of 4, indicating that the students seldom interacted with each other and were simply used to answering questions around the teacher's topic postings. This shows that flipped classroom can draw students' attention to issues in their own learning and those of their peers so that they can go beyond the conventional "teacher-led" model and help each other to solve problems on their own. See Figure 4 and Figure 5 for details.

5. Analysis and Discussion

The experimental results show that online flipped classroom can effectively improve students' in-depth cognition

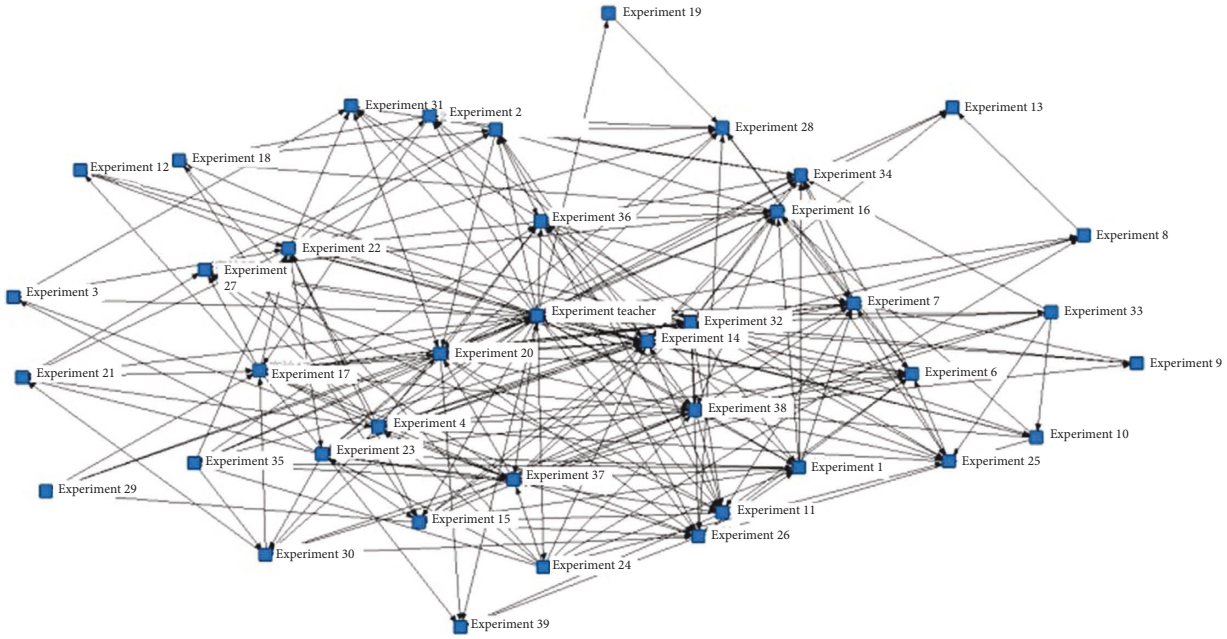


Figure 4: Social network diagram of the experimental group.

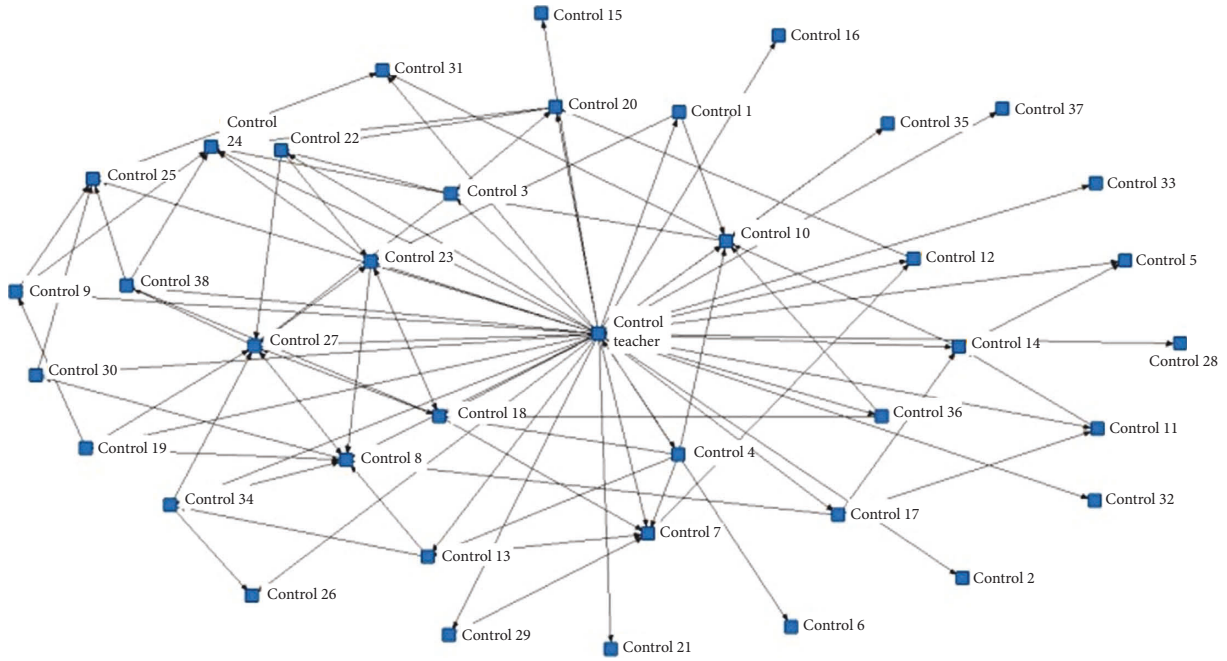


Figure 5: Social network diagram of the control group.

capacity and lead to quality construction and more diversified knowledge network patterns, which can be attributed to the following.

5.1. Online Flipped Classroom Helps Develop Students' Higher-Order Thinking Capability and Mobilize In-Depth Cognitive Engagement. The experiment results showed that the cognition capacity of the students in the experimental group was significantly higher than that of the control group, and their cognitive behaviors were mostly at the "application and analysis" phases, while a few of them reached the "evaluation and creation" phases. In contrast, the cognitive behaviors of students in the control group were concentrated in the two superficial cognition phases of memorization and comprehension, and only a few of them reached the phases of "analysis and application"; rational analysis and multiperspective evaluation were generally absent in their application of knowledge. This suggests that conventional online teaching in open universities can hardly stimulate students' engagement with in-depth cognition. Conventional online teaching generally sets up free and flexible

classroom sessions based on adult learning experiences, including viewing learning resources, participating in online live classes, completing discussion activities, and completing assignments. Teacher-student and student-student interactions usually occur in online live classes and discussion activities. However, the teaching during online live class is mostly composed of the teacher's lecture on knowledge. Moreover, in the process of such lecture, although the teacher will ask questions to mobilize students' thinking, the questions are mainly about factual knowledge and fail to mobilize students' higher-order capabilities such as critical thinking and creative thinking, so students' cognition remains at the phase of memorization and comprehension. This is the same as the cognitive processing the students implement in viewing learning resources, both of which simply involve the development of students' superficial cognition.

Online flipped classroom develops a three-phase process of "self-directed learning-exploratory learning-reflective learning" and designs classroom-based learning activities from superficial to in-depth levels at different phases. This helps students engage their thinking skills at each phase in a sequential manner. Firstly, the students view the learning resources to additionally remember and comprehend the knowledge and concepts they have already been exposed to, which stimulates the engagement of basic thinking. In an online live class, the instructor organizes exploratory thinking activities, facilitating students to retrieve the information they have viewed in the self-directed learning phase and to use higher-order thinking skills such as "application, analysis, and evaluation" to retrieve and summarize knowledge, thus establishing a unique knowledge system. Real-time interactive comments from peers and teachers also help students to constantly update their knowledge and experience and ensure cognitive validity. Such exploratory online live class provides a foundation for students to apply higher-order thinking skills in subsequent discussion forum activities. On the one hand, the students' cognition capacity has been raised to the "application, analysis, and even evaluation" phases in the online live class, so they can continue to think at the same level of cognitive engagement in the discussion forum. In addition, a few students are able to participate in the interactions with teachers and peers in the discussion forum to further improve and innovate their knowledge system and reach the highest cognition level. This step-by-step learning process helps adult students to logically sort out their fragmented knowledge and experiences through knowledge-imparting-style teaching, further enhance their construction in effective exploratory teaching, and eventually transfer and enhance them in interactive scenario-based teaching. This also confirms what has been mentioned in previous studies: students' higher-order thinking occurs after they exercise procedural knowledge and is enhanced through their self-directed exploration, thinking, and problem-solving under the guidance of the teacher [29]. The entire process of flipped classroom, from self-directed learning before class and exploratory learning based on learning activities during class to reflection and extension after class, can support students to achieve in-depth learning in a step-by-step manner, which is conducive to the development of students' higher-order thinking [29, 30].

5.2. Online Flipped Classroom Enhances Students' Sense and Behavior of Collaboration and Helps Achieve Collaborative Learning. According to the results of the experiment, after the implementation of flipped classroom, the level of social construction of the students in the experimental group was significantly higher than that of the students in the control group. In the low-level construction phase, 38% of the postings of the students in the experimental group were stating and sharing their own views and 10% were questioning others' views. In contrast, 57% of the control group's postings were stating and sharing their own views and no one challenged others' postings. In the high-level construction phase, 19% of the postings of the experimental group students were discussing with others to propose new views, 9% were revising their own views, and 2% applied the new views for construction. In contrast, 19% of the postings of control group students were proposing new views in response to others' postings, very few students revised their own views, and no students applied new views for construction. Respectively, 21.39% and 23.62% of the postings of the two groups were simply showing emotions. This shows that in conventional online teaching at open universities, students believe that the main tasks in discussion activities are to post and share their views and that collaboration simply involves telling others about their postings in the form of replies or replying to postings that show emotions. These further show that the students are not aware of what is a meaningful act of constructing or the method of constructing.

In online flipped classroom, students' collaborative learning is reflected in two teaching sessions: online live class and discussion forum. The online live class is a type of synchronous, interactive, exploratory learning process, in which the teacher first summarizes the issues with students' tests and discusses with the class the questions and challenges in learning, facilitating students to ask questions, analyze, integrate, reflect, and communicate by referring to their existing knowledge and experience. The exploratory process not only deepens students' understanding and views of the issues but also develops students' ability to express, communicate, and exchange and focuses on the methods they learn in the process of constructing knowledge. As a result, students learn to apply the relevant collaboration methods in subsequent forum activities and are able to revise and even reconstruct the questions through communication and reflection, thus improving their level of construction. This further confirms what has been mentioned in previous studies: flipped classroom can significantly improve the problem that students' cooperative learning is a mere formality, and the interactive, cooperative, and participatory nature of flipped classroom makes it necessary for students to collaborate with each other to complete the learning, which promotes the development of students' collaboration skills and thus ensures the quality of learning [31, 32].

5.3. Online Flipped Classroom Stimulates Students' Initiative and Contributes to the Generation of Knowledge Innovation. According to the experimental results, in flipped classroom model, members have closer relationships with each other, participate more interactively in a forum, and exhibit more personal contributions. In comparison, although conventional online teaching provides learners with open online learning resources and advocates that adult learners are "teachers" with experience, the process of teaching implementation remains at the level of a "teacher-centered" approach; i.e., the teacher leads the entire teaching process and students follow the "command" to complete the learning. Although such open forum activities give students space for independent exploration, the students simply follow the teacher's instructions to answer the test questions and attend online live classes, with limited self-efficacy and hence no desire to show themselves in the forum activities. These result in limited interactive participation of students, which is not conducive to the "collision" of ideas and generation of new knowledge. This is consistent with what has been mentioned in other studies: adult learners are influenced by conventional education and are accustomed to simply listening and receiving information in online learning, which leads to their low motivation for raising questions and sharing answers in the learning process and limited ability to explore knowledge in depth [33, 34]. In addition, because there is basically no communication among students prior to forum activities, the emotional linkages are weak, and therefore, they are reluctant to point out the shortfalls of others in the forum activities and lack two-way communication. This confirms what has been mentioned in other studies: students are reluctant to communicate with people they are not familiar with in online learning and therefore two-way interaction is limited [10].

Flipped classroom is consistently based on the "student-centered" approach, in which the teacher is merely the organizer and facilitator of teaching activities. Therefore, online flipped classroom organizes online live class activities targeting the issues exhibited by students in their self-directed learning and designs exploratory activities focusing on learners' initiative actions, which is conducive to stimulating students' interests and experiences to participate in the construction of knowledge. Firstly, online live class is centered on the problems that students face, so that students gain stronger ownership, showcase themselves through dialogue, reflection, and communication to solve problems in group activities, enhance their self-confidence, and establish emotional connections. As a result, in subsequent forum activities, students' learning potential and motivation will be stimulated, and they will be able to express their views courageously, identify issues proactively, and contribute their knowledge and power in order to better present themselves in an interpersonal environment more familiar to them. In this process, teaching and learning are no longer one-to-one but many-to-many; i.e., each student has the confidence, ability, and emotional basis to disagree with and revise others' postings as a "teacher," while accepting others' suggestions as a student and working together to improve and revise answers, thus realizing knowledge innovation. This also confirms what has been mentioned in other studies: the speech behavior of students in the traditional mode is mainly reflected in the passive response of students to teachers. In the flipped classroom mode, students use more time to share their knowledge achievements and play the role of teachers. It can stimulate students' learning motivation and potential, make them actively participate in learning, gain a better sense of self-efficacy through self-exploration, and finally enhance learning effectiveness [35, 36].

6. Study Limitations and Outlook

This study developed an online teaching model based on flipped classroom, which is designed to promote students' interactive participation in online learning and enhance the quality of interactions by leveraging the advantages of flipped classroom. In order to evaluate the actual effect of the teaching model, an empirical analysis was conducted from three perspectives: students' cognitive participation level, knowledge construction level, and knowledge network patterns. Although the effectiveness of online flipped classroom was confirmed, the study came with several limitations. First, the design of the online teaching model based on flipped classroom is still immature. Therefore, the researchers were prudent in defining the scope of the experiment and designed and studied only two weeks of course learning which covered a short period of time and a limited number of forum activities. Hence, the validity of the experimental results will still need to be established in the future through large-scale and long-term validation. Second, this study selected adult students majoring in preschool education as the data source, and the sample was not adequately representative; whether there are differences among students of different majors remains to be examined. Third, constrained by the length of the paper, although this study explored the changes in student interaction behaviors and interaction quality, it fails to address the influence of teachers in this process [37].

Therefore, the following recommendations are proposed for subsequent future studies: (1) continue to explore the cognitive characteristics of students in flipped classroom-based online teaching, and examine the impact of online flipped classroom on students' in-depth cognition from two dimensions: objectives and outcomes, (2) focus on the influence and role of teachers in flipped classroom, and analyze how interactive teaching activities are carried out between teachers and students, and (3) examine the interaction behaviors and quality of interactions among different types of learners in flipped classroom, and conduct an empirical study on improvement to the teaching effectiveness from the perspective of social division of labor.

Data Availability

The datasets used during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] F. Chen, "Analyzing teaching design and implementation effect of "economics" flipped classroom," *Computer Engineering & Science*, vol. 41, pp. 161–165, 2019.

[2] Z. Zhu, "New development of smart education: from Flipped Classroom to smart classroom and smart learning space," *Open Education Research*, vol. 49, no. 1, pp. 18–26, 2016.

[3] C. G. Prober, S. Khan, and Medical Education Reimagined, "Medical education Reimagined," *Academic Medicine*, vol. 88, no. 10, pp. 1407–1410, 2013.

[4] A. E. Chis, A. N. Moldovan, L. Murphy, P. Pathak, and C. H. Muntean, "Investigating flipped classroom and problem-based learning in a Programming module for Computing conversion course," *Educational Technology & Society*, vol. 21, no. 4, pp. 232–247, 2018.

[5] M. Graham, J. McLean, A. Read, S. Suchet-Pearson, and V. Viner, "Flipping and still learning: experiences of a flipped classroom approach for a third-year undergraduate human geography course," *Journal of Geography in Higher Education*, vol. 41, no. 3, pp. 403–417, 2017.

[6] P. Zhang, L. Ding, and W. Zhang, "Flipped classroom: theory, development history and effectiveness," *Journal of Educational Studies*, vol. 13, no. 01, pp. 47–52, 2017.

[7] Li. Chen and Y. Tong, "Social interaction strategies and methods in distance learning," *Distance Education in China*, vol. 2006, no. 08, pp. 14–17, 2006.

[8] K. M. Russ, "Relating reading attitude to reading achievement in an East Los Angeles Junior high school," *Reading Improvement*, vol. 26, no. 3, pp. 208–214, 1989.

[9] A. W. Bates, *Interactivity as a Criterion for media Selection in Distance Education*, , p. 16, Never Too Far, 1991.

[10] Li Chen, "Information exchange activities and student information exchange network in distance learning," *Distance Education in China*, vol. 2004, no. 5, pp. 15–19, 2004.

[11] L. Cao and Li Chen, "Behavioral styles and characteristics of asynchronous interaction of distance learners," *Distance Education in China*, vol. 2006, no. 1, pp. 15–19, 2006.

[12] K. Ning, X. Gu, and W. Wang, "The Meta-analysis of the teaching effect of the flipped classroom-based on the 70 related research Articles adopting random experiment or Quasi-experiment," *Modern Educational Technology*, vol. 28, pp. 39–45, 2018.

[13] Z. Wang and L. Chen, "Conceptual interaction and learning assessment in distance learning," *Distance Education in China*, vol. 12, pp. 12–19, 2017.

[14] K. He, "From the essence of "Flipped Classroom" to see the future development of "Flipped Classroom" in China," *E-Education Research*, vol. 35, pp. 5–16, 2014.

[15] Y. Jiang, "Research on the design of flipped classroom teaching mode of open education based on Microcourse," *Distance Education in China*, vol. 2014, no. 03, pp. 52–60, 2014.

[16] M. Huang and M. Shen, "Research on the effective combination of MOOCs and the flipped classroom in the open university: on the combination of the flipped classroom and MOOCs through Tai Chi Quartet nest model," *The Journal of Distance Education*, vol. 33, pp. 62–70, 2015.

[17] J. Hu and J. Jiang, "The application of flipped classroom in the teaching of software courses," *Distance Education in China*, vol. 2016, no. 03, pp. 45–51, 2016.

[18] L. Vygotsky, "Interaction between learning and development," *Readings on the development of children*, vol. 23, no. 3, pp. 34–41, 1978.

[19] Li Chen and Z. Wang, "Instructional interaction in three generations of distance learning," *Distance Education in China*, vol. 10, pp. 32–34, 2016.

[20] L. Lv and F. Gong, "A comparative study of in-depth learning and its influencing mechanism among undergraduate students in Chinese and American research universities-an empirical analysis based on SERU surveys in eight universities in China and the United States," *Educational Research*, vol. 2018, no. 4, pp. 111–120, 2018.

[21] Q. Liu, X. Wang, and H. Jiang, "Social network analysis of interactive learning behavior in teachers' Workshops," *Modern Distance Education*, vol. 03, pp. 22–29, 2019.

[22] H. Yan, C. Duan, and W. Wang, "Development and effect Verification of quality analysis tools of online discussion," *Modern Distance Education Research*, vol. 2018, no. 01, pp. 88–97, 2018.

[23] J. Zhao and J. Kong, "Coding Schemes of online discussion analysis: Categories and application," *The Journal of Distance Education*, vol. 2015, no. 02, pp. 11–20, 2015.

[24] B. Zhong and Li. Yi, "On social network analysis in the application of educational research-A Literature review based on core education Journals," *Educational Research*, vol. 2013, no. 09, pp. 27–34, 2013.

[25] J. Li and M. Wang, "Visualization analysis of patent Mining at Home and Abroad," *Library Work in Colleges and Universities*, vol. 39, no. 02, pp. 7–12, 2019.

[26] B. Chen and Yi Zhang, "A study on how teaching interaction strategies supported by technologies promote the depth of interactions," *China Educational Technology*, vol. 2019, no. 08, p. 103, 2019.

[27] C. N. Gunawardena, C. A. Lowe, and T. Anderson, "Analysis of a Global online Debate and the development of an interaction analysis model for examining social construction of knowledge in Computer conferencing," *Journal of Educational Computing Research*, vol. 17, no. 4, pp. 397–431, 1997.

[28] Li Mei and J. Yang, "The analysis of Interbehavior in peer to peer Reciprocal learning," *China Educational Technology*, vol. 2016, no. 05, p. 93, 2016.

[29] A. Dong, S. Zhuang, and J. Shang, "Achievement Pathway of higher-order thinking through classroom interactions in the flipped classroom," *Modern Educational Technology*, vol. 29, pp. 46–51, 2019.

[30] X. Wang, J. Zhang, H. Yang, and S. Zhang, "Research on flipped classroom teaching model in Colleges from perspective of Deep learning," *E-Education Research*, vol. 41, pp. 85–91+128, 2020.

[31] F. Huang, Li Cheng, Y. Ke, and Li Shang, "Practice of heterogeneous group cooperative learning in flipped classroom teaching for undergraduate nursing students[J]," *Journal of Nursing Science*, vol. 36, no. 07, pp. 1–3, 2021.

[32] Y. Hu, Y. Dong, and M. Yang, "Model construction of College Students'Problem solving ability in flipped classroom," *Modern Distance Education*, vol. 2018, no. 1, pp. 26–34, 2018.

[33] F. Yang and Li Ling, "Analysis of factors influencing asynchronous interaction between teachers and students in network assisted instruction," *Journal of Jiangxi Radio & TV University*, vol. 2011, no. 03, pp. 65–68, 2011.

[34] Li Chen, "The essence of the term "teaching interaction" and its related concepts," *Distance Education in China*, vol. 2004, no. 03, pp. 12–16, 2004b.

[35] H. Hu, C. He, X. Hu, and H. Li, "The characteristics of teacher-student interaction behaviors in the traditional class and the flipped class," *Chinese Journal of Nursing Education*, vol. 19, no. 05, pp. 389–393, 2022.

[36] J. Ding, "Analysis of learners' cognitive network and motivation Strategy in Deep learning: study on the effect of U-flipped teaching directed Deep learning," *The Journal of Distance Education*, vol. 37, pp. 32–40, 2019.

[37] Li Wang and Li Lin, "Analysis of effectiveness of flipped classrooms in universities," *Heilongjiang Researches on Higher Education*, vol. 2019, no. 09, pp. 157–160, 2019.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 9903645, 7 pages
https://doi.org/10.1155/2022/9903645

Research Article

Vibration Isolation and Noise Reduction Method Based on Phononic Crystal

Haiqing Li and Ping Sun

Liuzhou Vocational and Technical College, Liuzhou 545006, China

Correspondence should be addressed to Ping Sun; sunping@lzzy.edu.cn

Received 4 August 2022; Revised 20 September 2022; Accepted 26 September 2022; Published 10 October 2022

Academic Editor: Heng Liu

Phononic crystal is a new kind of sound insulation material, which has an elastic wave gap. When the elastic wave falls within the band gap, it will attenuate strongly in phononic crystal, and its attenuation degree is far greater than the predicted value of the mass density theorem. In this paper, the practical application of phononic crystals in low frequency sound insulation is taken as the breakthrough point. Firstly, the theory of phononic crystal band gap generation is calculated and analyzed, and the band structure of one-dimensional two-component Bragg scattering phononic crystals is calculated. Hypermesh is used to build the model, and the sound insulation performance of phononic crystals is simulated and analyzed through Nastran. A sound isolation test platform was built for the local resonant phononic crystal samples to verify its sound isolation ability.

1. Introduction

Vibration and noise reduction theory plays an important role in automobile, ship, aircraft, micromachinery, and other fields. It is also the basis of modern automobile comfort design, low-noise submarine design, aircraft instrument vibration isolation, micromachinery high-precision processing, and other high technology. In modern design and research, vibration and noise control of solid structures is a common concern [1–3]. The traditional theory of vibration and noise reduction has become mature after many years of development. It is urgent for the birth of new theories and methods to meet people's needs for low vibration and low-noise environment. Therefore, the concept of phononic crystal is proposed, and it is found that when an elastic wave propagates in periodic composite media, an elastic wave or acoustic band gap similar to a photon band gap will be generated, that is, the elastic wave or acoustic wave within the band gap frequency range will be prohibited to propagate in the medium.

In recent years, due to the bandgap characteristics of phononic crystals, which plays an important role in preventing the propagation of elastic waves, research studies on their practical application mostly focus on vibration control and noise prevention [4, 5]. Due to elastic wave propagation in the medium is the most direct impact vibration, by taking advantage of the band gap of phononic crystal characteristics can be used to reduce the spread of vibration energy. On the one hand, in mechanical manufacturing process can reduce the manufacturing error, improve the accuracy of the product, on the other hand can reduce the surrounding environment adverse effects of instruments and equipment, and improve its operating life [5–8]; as the propagation of noise obeys the elastic wave equation, phononic crystals also have the characteristics of bandgap, so this phenomenon can provide a new idea for the development of sound insulation materials.

2. Phononic Crystal Theory Method

Phononic crystals are periodic coincidence materials or structures with elastic wave band gaps and can be regarded as an extension of the crystal concept in solid state physics. In essence, the study of phononic crystals is to study elastic wave propagation in periodic inhomogeneous media [9, 10]. Therefore, the lattice and band theory in elastic dynamics

and solid state physics is the theoretical basis of phononic crystals.

2.1. Elastodynamics. In elastodynamics, the stress-strain relationship should be clarified first, and the generalized Hooke's law should be understood. On this basis, the basic equations of elastodynamics can be established, and three basic equations describing the relationship between particle force, displacement, and stress-strain can be established [11–13]. The equation of motion with displacement as an unknown function is called the Wiener equation.

$$\rho u_i = \rho f_i + \sum_{j=1}^{3} \left\{ \frac{\partial}{\partial x_i} \left(\lambda \frac{\partial u_j}{\partial x_j} \right) + \frac{\partial}{\partial x_j} \left[\mu \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right) \right] \right\}. \quad (1)$$

In the formula, $i, j = 1, 2, 3, x_1, x_2, x_3$ corresponds, respectively, to x, y, z, and u_1, u_2, u_3 corresponds, respectively, to u_x, u_y, u_z or u, v, w. On the basis of this equation, the propagation characteristics of elastic waves in periodic composites are obtained. It provides a numerical calculation for the study of phononic crystals.

2.2. Lattice and Band Theory. The effect of phononic crystals on elastic waves is similar to that of atomic periodic potential fields on electrons. Due to this similarity, phononic crystal research directly adopts the concepts, description methods, and main conclusions related to lattice theory and band theory in solid physics [14–16].

The concept of the lattice is abstracted from the internal structure of a crystal. The lattice is translational periodicity and spatial symmetry. Because these two properties of crystal make the eigenfrequency and eigenmode of the field have a certain symmetry [17], the problem can be simplified, and the Bloch theorem is often used to explain the characteristics and laws of the eigenfield in crystal.

Energy band theory is the main theoretical basis for studying electron motion in a solid. It abstracts a solid into an ideal crystal with translational periodicity and symmetry, and simplifies the motion of the electron in solid to the motion of a single electron in a periodic potential field, thus establishing a series of methods to calculate the energy band of the electron in solid. Many basic physical properties of solids can be explained and explained in principle by the solid energy band theory [18–20]. Due to the phonon crystal analogy in crystals of the periodic structure, therefore, to calculate the solid band structure, usually select a Bloch functions form a complete, elastic wave function with the function of the phononic crystal base to expand, and then plug in the elastic dynamics equation, to determine the coefficient of expansion must satisfy the secular equation, the eigenvalue of energy are obtained accordingly.

2.3. Calculation Method of Band Gap Characteristics of Phononic Crystals. The study of the band gap mechanism and properties of phononic crystals depends on effective calculation methods of band gap properties. At present, the main calculation methods include the transfer matrix method, plane wave expansion method, finite difference time domain method, multiple scattering method, and concentrated mass method.

In this project, the transfer matrix method is used to calculate the band gap of one-dimensional phononic crystals. The transfer matrix method is based on the basic equation of continuous state parameters (stress, particle displacement, etc.), combined with the interface continuity condition, the transfer matrix of a single period is obtained. The corresponding dispersion relation and band structure are obtained by introducing periodic boundary conditions. At the same time, the finite period transmission characteristics can be obtained by multiplying finite transfer matrices.

According to the propagation law of SH wave and one-dimensional wave equation in phononic crystals, the relation between the NTH protocell and n-1 protocell of ideal phononic crystals can be obtained, $\psi_{n2} = T\psi_{(n-1)2}$ where $T = K_1^{-1} H_1 H_2^{-1} K_2$ is the transfer matrix [21]. Due to the periodicity of x direction, it can be obtained by using the Bloch theorem, $\psi_{n2} = e^{ika}\psi_{(n-1)2}$ where k is one-dimensional Bloch wave vector. The standard matrix eigenvalue problem can be obtained by combining the above two equations:

$$\left| T - e^{\mathrm{ika}} I \right| = 0, \quad (2)$$

where I is the identity matrix.

By solving the eigenvalue of the matrix T, the dispersion relation between wave vector K and frequency W (perpendicular incidence) can be obtained.

$$\cos(ka) = \cos\left(\frac{w}{c_1}a_1\right)\cos\left(\frac{w}{c_2}a_2\right) - \frac{1}{2}\left(\frac{\rho_1 c_1}{\rho_2 c_2} + \frac{\rho_2 c_2}{\rho_1 c_1}\right)\sin\left(\frac{w}{c_1}a_1\right)\sin\left(\frac{w}{c_2}a_2\right), \quad (3)$$

where $a = a_1 + a_2$ are the lattice constant, and ρ_1, ρ_2, c_1, c_2 are the density and propagation speed of the two materials, respectively.

3. Calculation of Band Gap of One-Dimensional Phononic Crystals

In this study, phononic crystals composed of aluminum and epoxy resin mentioned in many literature were studied, and it was found that the band gaps were generally at a higher frequency. In order to explore the research method of phononic crystals, the band gap of ideal phononic crystals composed of these two materials was firstly calculated and analyzed [22, 23]. The specific structure is shown in Figure 1, and the materials are shown in Table 1. It is found that the band gap is at high frequency, the initial frequency of the first band gap is above several kilohertz, and the low frequency band gap is difficult to obtain.

Lattice periodic size $a = 0.01$ m, aluminum column diameter $d_1 = 0.020$ m, period height $a_1 = 0.005$ m; Silicon rubber diameter $d_2 = 0.012$ m, $a_2 = 0.005$ m. Specific material parameters are shown in Table 1.

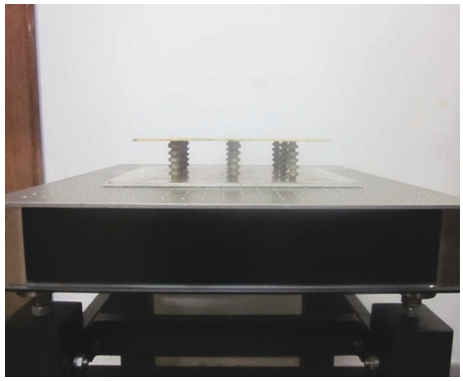


Figure 1: Phononic crystal composed of aluminum and epoxy resin.

Table 1: Material parameters.

Material	Density ($kg \cdot m^{-3}$)	Elasticity ($10^{10} Pa$)	Poisson's ratio
Aluminum	2730	7.76	0.3
Epoxy resin	1180	0.435	0.4

To make the phononic crystals start at a low frequency, the density of the metallic part should be as high as possible, especially in applications where both the band gap start frequency and the band gap width are required. The change of elastic modulus within a certain range does not affect the band gap characteristics, so steel is selected for the metal part. As the rubber density increases, both the initial frequency and cutoff frequency of the band gap decrease, but as the density increases, the cutoff frequency of the band gap decreases more, and the width of the band gap decreases until the band gap disappears [24]. Therefore, the density of rubber should be appropriate, according to the rubber as silicone rubber.

Based on the above-given principles, material size parameters are determined: lattice periodic size $a = 0.05m$, steel column diameter $d_1 = 0.028m$, periodic height $a_1 = 0.03m$; Silicone rubber diameter $d_2 = 0.025m$, $a_2 = 0.02m$. Specific material parameters are shown in Table 2.

The band gap characteristics of one-dimensional ideal phononic crystals calculated by the transfer matrix method based on equivalent parameters are shown in Figure 2. It can be seen that the first gap is generated at 290 Hz~1325 Hz.

4. Finite Element Simulation Verification

Based on the structure and parameters of the material mentioned above, the Hypermesh model was established and meshed. Nastran was used to calculate its direct frequency response and intercept part of the response cloud map, as shown in Figures 3–5. It can be concluded that the upper surface amplitude responses of phononic crystals are all less than 1 at 280 Hz~1360 Hz, while the amplitude is greater than 1 at 1380 Hz. It can be judged that the band gap ranges from 280 Hz to 1360 Hz. Considering that the transfer matrix method results in infinite cycles and the finite element method calculates three cycles and the error of finite element numerical calculation, the results are considered to be credible [25].

Table 2: Material parameters.

Material	Density ($kg \cdot m^{-3}$)	Elasticity ($10^{10} Pa$)	Poisson's ratio
Steel	7780	21.6	0.3
Silicone rubber	1481	4.2e–4	0.4

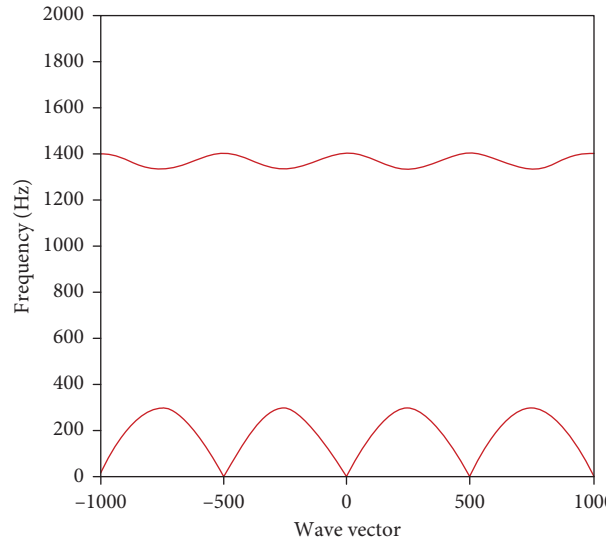


Figure 2: Band gap characteristics of one-dimensional ideal phononic crystals.

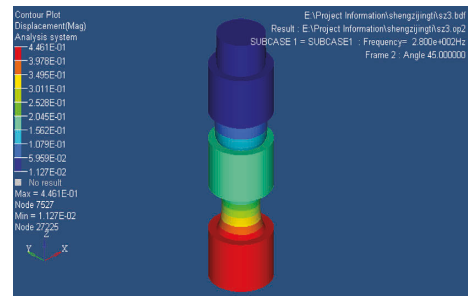


Figure 3: Cloud diagram of direct frequency response when frequency is 280 Hz.

5. Vibration Isolation and Noise Reduction Test Verification

5.1. Preparation of Phononic Crystals. Based on the above-given calculation results and the material size verified by the finite element method, the structural design is carried out, and the two materials are periodically bonded together. Due to the limitation of cost and other conditions, only phononic crystals with three periods were made in this project. For ease of measurement and installation, four phononic crystals were made and arranged as shown in Figures 6 and 7.

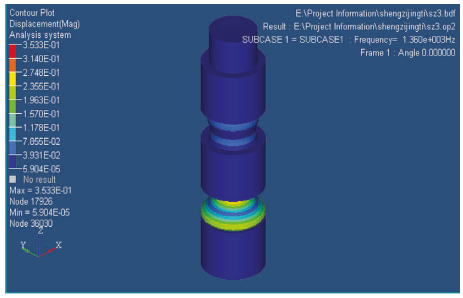


Figure 4: Cloud diagram of direct frequency response at 1360 Hz.

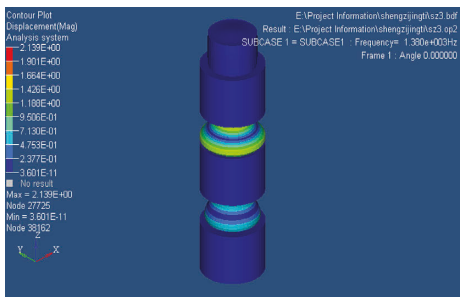


Figure 5: Cloud diagram of direct frequency response at 1380 Hz.

Figure 6: Phononic crystal is composed of steel and silicone rubber.

5.2. Test Verification Method. Sensors used in the noise test are shown in Table 3. The sampling frequency is set to 12800 Hz. The sampling time was 3 s and the test times were 10 for each.

(1) Use rubber rope to suspend phonon crystal assembly on a suspension bracket of body in White, as shown in Figure 8

(2) Install the three-way acceleration sensor, connect the sensor and computer to the signal acquisition front end, and start the test module of the test software, as shown in Figure 9

(3) Channel Setup, as shown in Figure 10

(4) Tracking Setup The test duration is 10 s, and the step length is 0.5 s

(5) Acquisition Setup Set bandwidth and sampling frequency; See Figure 11

(6) Start measuring

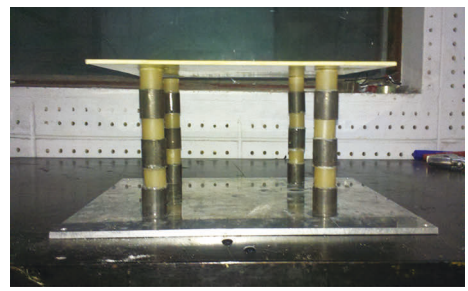


Figure 7: Array of four phononic crystals.

Table 3: Sensors used in noise test.

Sensor type	Microphone	Accelerometer
Model	PCB 378B02	PCB 356A02
Sensitivity	50 mV/Pa	10 mV/g
Frequency range	3.75 to 20000 Hz	1 to 20000 Hz
Mass loading	45.8 g	10.5 g

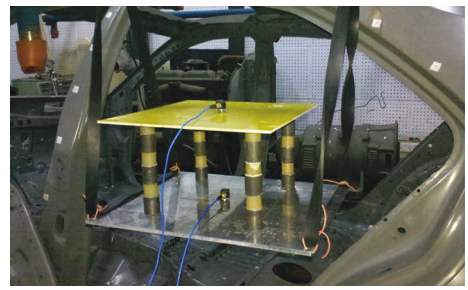


Figure 8: Suspension mode of phononic crystal.

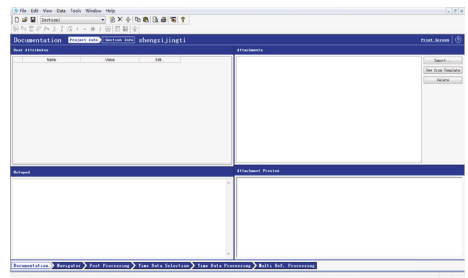


Figure 9: Test module in test software.

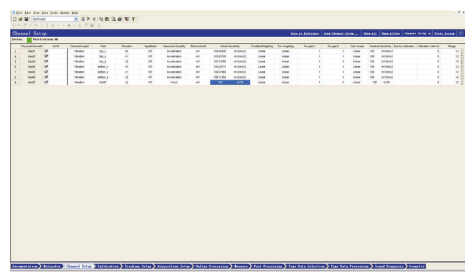


Figure 10: Channel Settings.

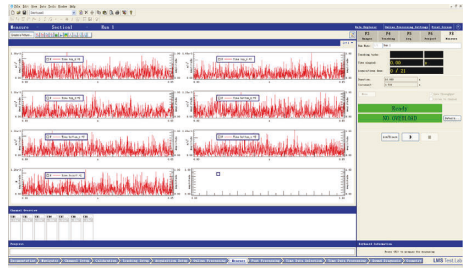


Figure 11: Setting and measurement interface.

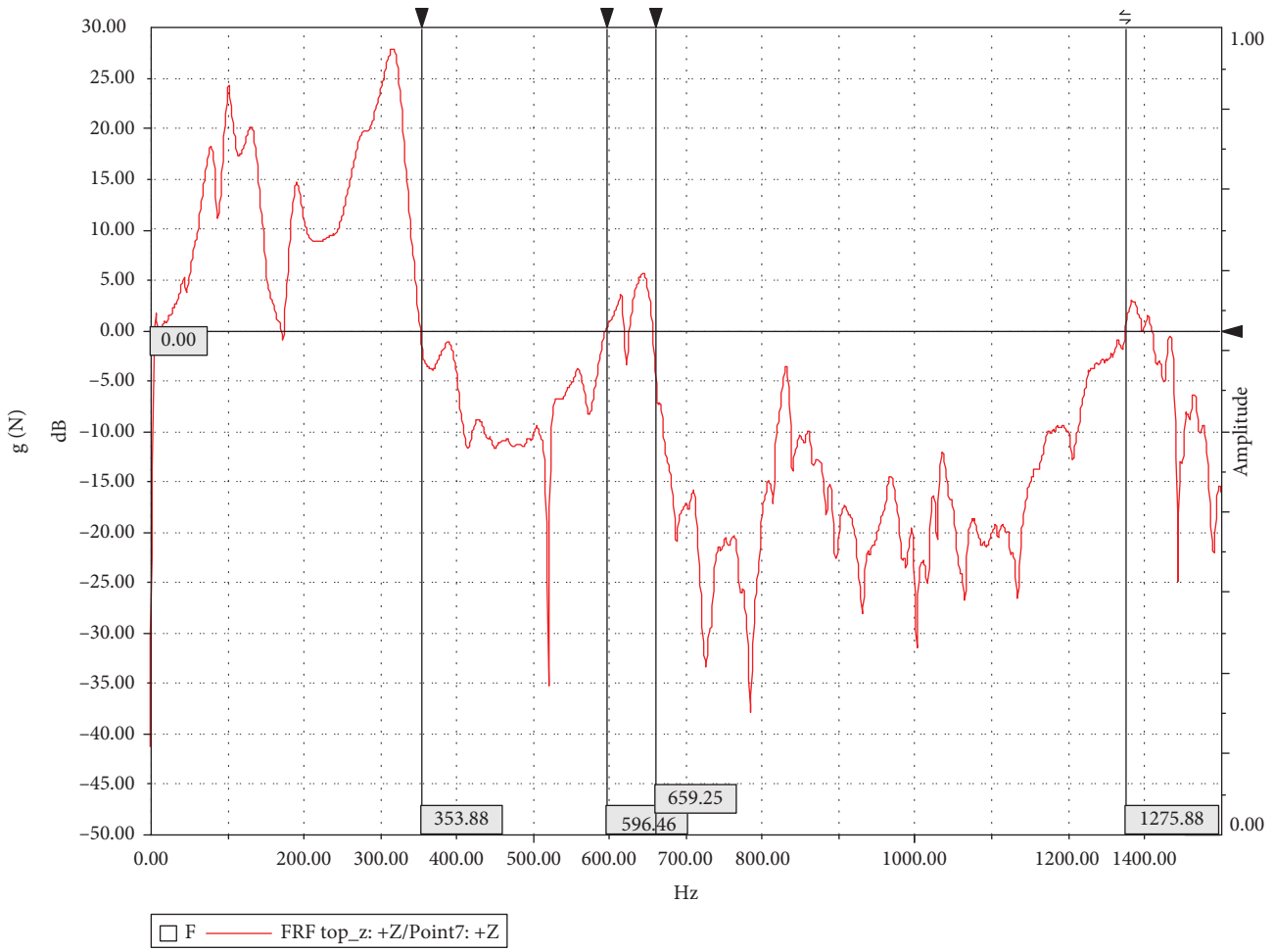


Figure 12: Frequency response function of the system.

5.3. Effect and Analysis of Vibration Isolation and Noise Reduction. Through the postprocessing function of the testing software, the frequency response function of the phonon crystal system is obtained, so as to clarify the band gap characteristics obtained by the test. The frequency response function obtained by processing is shown in Figure 12.

According to the frequency response function in Figure 12, the frequency response is less than 0 dB in the range of 353~596 Hz and 659–1275 hz, that is, band gaps are generated in these frequency ranges. The band gap frequency range is partly consistent with the band gap obtained by theoretical calculation and simulation analysis, but the initial frequency is quite different [26, 27], around 60 Hz, and the band gap is discontinuous. Based on the analysis of the production process and test scheme, the main reasons for this difference may be as follows:

(1) Due to the influence of manufacturing accuracy, there is a disalignment phenomenon between the two materials during bonding, which affects the vibration characteristics of phononic crystals;

(2) As the upper and lower plates of the system have certain vibration characteristics, they contribute to

the system during measurement. How to eliminate this contribution needs further research in the future;

(3) It is better to use a shaking table to measure the frequency response function, but this test refers to the modal test method to suspend the system for testing, so there are certain test errors, and the test scheme needs to be further improved in the future research.

6. Conclusion

In order to explore the application of phonon crystal in the field of low frequency sound insulation, based on the low frequency band gap theory of phonon crystal, a one-dimensional two-element Bragg scattering phonon crystal is designed, its band structure is calculated, and its sound insulation ability is verified by experiments. The main research results are as follows:

(1) The band results of the theoretical calculation method and finite element calculation are in good agreement with the experimental results, and the calculation method can be used for the band gap design of one-dimensional two-element Bragg scattering phonon crystal

(2) The designed one-dimensional two-element Bragg scattering phonon crystal structure can form a low-frequency band gap and effectively isolate low-frequency vibration and noise

Data Availability

If you need data support, please contact the contributor Haiqing Li, e-mail: lhq@lzzy.edu.cn.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the 2022 Research Basic Ability Improvement Project of Young and Middle-Aged College Teachers in Guangxi: Study on Band Gap Mechanism of Phononic Crystals and Its Application in NVH (Project no. 2022KY1049), 2022 Liuzhou Science and Technology Bureau "High-level Talents and Team Innovation and Entrepreneurship Industry" (Project no. 2022DAA0102), and the 2020 Annual Project of Liuzhou Vocational and Technical College (Project no. 2020KA02). HQL and PS appreciate the financial support from these grants.

References

[1] Du. Gong, *Acoustic Theory and Engineering application, Bei Jing: Acoustic Theory and Engineering Application*, Beijing, Science Press, 2006.

[2] T. Chen, S. Li, and H. Sun, "Metamaterials application in sensing," *Sensors*, vol. 12, no. 3, pp. 2742–2765, 2012.

[3] R. Zhu, X. N. Liu, G. K. Hu, C. T. Sun, and G. L. Huang, "A chiral elastic metamaterial beam for broadband vibration suppression," *Journal of Sound and Vibration*, vol. 333, no. 10, pp. 2759–2773, 2014.

[4] J. Lydon, G. Theocharis, and C. Daraio, "Nonlinear resonances and energy transfer in finite granular chains," *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, vol. 91, no. 2, Article ID 023208, 2015.

[5] L. Y. Zheng, G. Theocharis, V. Tournat, and V. Gusev, "Quasitopological rotational waves in mechanical granular graphene," *Physical Review B: Condensed Matter*, vol. 97, no. 6, Article ID 060101, 2018.

[6] A. Colombi, P. Roux, S. Guenneau, P. Gueguen, and R. V. Craster, "Forests as a natural seismic metamaterial: Rayleigh wave bandgaps induced by local resonances," *Scientific Reports*, vol. 6, no. 1, Article ID 19238, 2016.

[7] Y. Yan, Z. Cheng, F. Menq, Y. L. Mo, Y. Tang, and Z. Shi, "Three dimensional periodic foundations for base seismic isolation," *Smart Materials and Structures*, vol. 24, no. 7, Article ID 075006, 2015.

[8] L. Schwan, O. Umnova, C. Boutin, and J.-P. Groby, "Nonlocal boundary conditions for corrugated acoustic metasurface with strong near-field interactions," *Journal of Applied Physics*, vol. 123, Article ID 091712, 2018.

[9] M. I. Hussein and M. J. Frazier, "Band structure of phononic crystals with general damping," *Journal of Applied Physics*, vol. 108, no. 9, Article ID 093506, 2010.

[10] R. Pernas-Salomon, R. Perez-Alvarez, Z. Lazcano, and J. Arriaga, "The scattering matrix approach: a study of elastic waves propagation in one-dimensional disordered phononic crystals," *Journal of Applied Physics*, vol. 118, no. 23, Article ID 234302, 2015.

[11] W. Deng and H. Xiao, "Study on band gap of band edge phononic crystal structure based on finite element method," *Mechanical Engineer*, vol. 1, no. 7, pp. 105–108, 2021.

[12] Yi. Zhang, D. Y. Shi, and D. Ze He, "Vibration characteristics of one-dimensional phononic crystal corner beam with constant cross-section," *Science Technology and Engineering*, vol. 21, no. 17, pp. 7084–7091, 2021.

[13] J. Xu, L. Cui, H. Yu, and M. Hong, "Study on sound insulation performance of phononic crystal sandwich plate structure," *Vibration and shock*, vol. 40, no. 9, pp. 285–291, 2021.

[14] D. L. Yu, *Study on Vibration Band gap Characteristics of Beam Plate Periodic Structures Based on Phononic Crystal Theory*, Chang Sai:National University of Defense Science and technology, Chang Sai, China, 2006.

[15] F. Li and LvM. Lian, "Calculation of band gap of phononic crystals with imperfect interface triangular lattice based on boundary element method," *Journal of intraocular lens*, vol. 49, no. 01, pp. 27–32, 2020.

[16] H. Liu, Y. Pan, J. Cao, Y. Zhou, and H. Wang, "Positivity and stability analysis for fractional-order delayed systems: a T-S fuzzy model approach," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 4, pp. 927–939, 2021.

[17] Y. Zhu, C. Lu, Z. Liu, W. Song, L. Xie, and J. Shen, "Simulated operational path analysis method for the separation of intake and exhaust noises," *Acoustics Australia*, vol. 48, no. 2, pp. 261–270, 2020.

[18] Y. Ju, G. Parks, and C. Zhang, "A bisection-sampling-based support vector regression–high-dimensional model representation metamodeling technique for high-dimensional problems," *Proceedings of the Institution of Mechanical*

Engineers - Part C: Journal of Mechanical Engineering Science, vol. 231, no. 12, pp. 2173–2186, 2017.

[19] Y. Xu, "Spatial bandwidth enlargement and field enhancement of shear horizontal waves in finite graded piezoelectric layered media," *Physics Letters A*, vol. 379, no. 30-31, pp. 1752–1756, 2015.

[20] A. Arreola-Lucas, G. Báez, F. Cervera, A. Climente, R. A. Mendez-Sanchez, and J Sanchez-Dehesa, "Experimental evidence of rainbow trapping and Bloch oscillations of torsional waves in chirped metallic beams," *Scientific Reports*, vol. 9, no. 1, p. 1860, 2019.

[21] G. Xue, F. Lin, S. Li, and H. Liu, "Adaptive dynamic surface control for finite-time tracking of uncertain nonlinear systems with dead-zone inputs and actuator faults," *International Journal of Control, Automation and Systems*, vol. 19, no. 8, pp. 2797–2811, 2021.

[22] H. L. Zhao, J. B. Huang, and H. Gu, "Research on vibration isolation and noise reduction technology based on Michelson interferometer," *Ship Electronic Engineering*, vol. 41, no. 12, pp. 200–204, 2021.

[23] D. Ze He, D. Y. Shi, Q. S. Wang, and M. Chun Long, "Study on vibration characteristics of one-dimensional phononic crystal structures with elastic supports," *Vibration and shock*, vol. 41, no. 04, pp. 185–191, 2022.

[24] F. Y. He, Q. D. Hui, and J. Tu, "Study on band gap of thin plate structure of equal local resonance type phononic crystal," *Low temperature building technology*, vol. 42, no. 11, pp. 38–41+49, 2020.

[25] Lu. Tang, Lv Y. J. Jia, and L. Cheng, "The dynamic model and band gap of a phononic crystal beam with local resonance," *Vibration. Testing and diagnostics*, vol. 41, no. 06, pp. 1132–1137+1238, 2021.

[26] H. Liu, Y. Pan, J. Cao, H. Wang, and Y. Zhou, "Adaptive neural network backstepping control of fractional-order nonlinear systems with actuator faults," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 12, pp. 5166–5177, 2020.

[27] Z. Han, S. Li, and H. Liu, "Composite learning sliding mode synchronization of chaotic fractional-order neural networks," *Journal of Advanced Research*, vol. 25, pp. 87–96, 2020.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 9752003, 20 pages
https://doi.org/10.1155/2022/9752003

Research Article

A Modified Reptile Search Algorithm for Numerical Optimization Problems

Qihang Yuan,[1] **Yongde Zhang**,[1,2] **Xuesong Dai**,[1] **and Shu Zhang**[2]

[1]*Key Laboratory of Advanced Manufacturing and Intelligent Technology, Ministry of Education, Harbin University of Science and Technology, Harbin 150080, China*
[2]*Foshan Baikang Robot Technology Co., Ltd, Nanhai, Foshan, Guangdong 528225, China*

Correspondence should be addressed to Yongde Zhang; zhangyd@hrbust.edu.cn

Received 18 June 2022; Revised 13 August 2022; Accepted 20 August 2022; Published 10 October 2022

Academic Editor: Heng Liu

The reptile search algorithm (RSA) is a swarm-based metaheuristic algorithm inspired by the encirclement and hunt mechanisms of crocodiles. Compared with other algorithms, RSA is competitive but still suffers from low population diversity, unbalanced exploitation and exploration, and the tendency to fall into local optima. To overcome these shortcomings, a modified variant of RSA, named MRSA, is proposed in this paper. First, an adaptive chaotic reverse learning strategy is employed to enhance the population diversity. Second, an elite alternative pooling strategy is proposed to balance exploitation and exploration. Finally, a shifted distribution estimation strategy is used to correct the evolutionary direction and improve the algorithm performance. Subsequently, the superiority of MRSA is verified using 23 benchmark functions, IEEE CEC2017 benchmark functions, and robot path planning problems. The Friedman test, the Wilcoxon signed-rank test, and simulation results show that the proposed MRSA outperforms other comparative algorithms in terms of convergence accuracy, convergence speed, and stability.

1. Introduction

The rapid advancement of technology has generated a large number of optimization problems that require solving. These optimization problems arise in various fields, such as finance, chemicals, electronics, machinery, and materials. Real-world optimization problems are often mixed with various unknown factors and have very complex solution spaces. These problems frequently have substantial computational efforts, complex nonlinear constraints, and large numbers of variables and constraints [1–6]. Traditional optimization methods have difficulty solving these nonproductivity discontinuity problems effectively because they cannot strike a balance between accuracy and time cost [7]. Metaheuristic optimization algorithms have demonstrated better performance in balancing the solution quality and time cost [8]. Due to a simple structure and no requirement for a problem to be continuously derivable, metaheuristic optimization algorithms have been widely used to solve complex optimization problems in natural and engineering fields [9–13].

In recent decades, metaheuristic algorithms have made great progress in memetic computing manner, balance of exploitation and exploration, self-adaption of hyperparameters, population structure evolution, and theoretical analysis of the search dynamics [14]. Memetic computing manner improves algorithm performance through metaheuristic algorithms incorporated with local search operator. Charin et al. used particle swarm optimization (PSO) combined with levy flight optimization (LFO) to track the maximum power point of a photovoltaic system [15]. Yu et al. showed that the combination of chaotic local search (CLS) and brain storm optimization (BSO) can significantly improve the performance of BSO [16]. How to balance the exploration and exploitation of the algorithm to improve the performance is a research hotspot of the metaheuristic algorithms. Many researchers use various operators or change the algorithm parameters to balance [17]. Cai et al. proposed an alternate search pattern strategy to balance the exploration and exploitation of BSO [18]. In the optimization process, the search performance of some metaheuristic

algorithms is greatly affected by adjustable parameters such as crossover rate, mutation rate, and population size. In order to solve the problem of parameter value control at different stages in the optimization process, adaptive parameter control has been extensively studied by researchers [19]. Lei et al. proposed a variant of gravitational search algorithm (GSA) with a self-adaptive gravitational constant called ALGSA, which greatly improved the search performance of GSA [20]. Population structure evolution has great influence on the search performance of the metaheuristic algorithms. Zhong et al. proposed a variant to improve the performance of differential evolution (DE) algorithm called EHDE by incorporating elite elements into the hierarchical population structure [21]. Inspired by the two-layered structure GSA, Wang et al. proposed a four-layered GSA variant with stronger search capability called MLGSA [22]. In addition to the above factors, theoretical analysis of the search dynamics has recently attracted a great deal of attention from researchers [23].

In general, metaheuristic optimization algorithms can be classified into three categories [24]: evolutionary-based algorithms, physical-based algorithms, and swarm-based algorithms.

Evolutionary-based algorithms are inspired by the laws of natural evolution. Genetic algorithms are a typical example and their proposal was inspired by Darwinian evolutionary theory [25]. Genetic algorithms provide solutions through the concept of crossover and mutation of species in nature. In addition, other evolutionary-based algorithms have been proposed, including DE [26], evolutionary programming [27], and evolutionary strategies [28]. The second category is physics-based algorithms, which originate from natural physics laws. Simulated annealing [29] and GSA [30] are two common physics-based algorithms. They utilize the laws of thermodynamics and gravity for optimization. In addition, researchers have proposed other physics-based algorithms. Wei et al. proposed a nuclear reaction optimizer using the phenomenon of atomic nuclear reactions [31]. Inspired by the sine and cosine laws of mathematics, Mirjalili proposed the sine cosine algorithm [32]. Eskandar et al. proposed a water cycle algorithm based on the natural water cycle phenomena [33]. The third category is swarm-based algorithms, which build optimization models by emulating the social behavior of animal groups. PSO [34] and the ant colony algorithm [35] are two of the most common swarm-based algorithms. They provide solutions by sharing information about all individuals in the optimization process. Others include the grey wolf optimizer (GWO) [36], the whale optimization algorithm (WOA) [37], the butterfly optimization algorithm (BOA) [38], the firefly algorithm (FA) [39], the artificial bee colony (ABC) algorithm [40], the reptile search algorithm (RSA) [41], the Harris hawks optimizer (HHO) [42], the equilibrium optimizer (EO) [43], the tunicate swarm algorithm (TSA) [44], the salp swarm algorithm (SSA) [45], the Tasmanian devil optimization (TDO) [46], the arithmetic optimization algorithm (AOA) [47], and the pathfinder algorithm (PFA) [48].

The reptile search algorithm (RSA) is a novel swarm-based algorithm proposed by Abualigah. RSA is inspired by the encircling mechanism, hunting mechanism, and social behavior of crocodiles [41]. RSA has good performance but also has disadvantages such as diminished population diversity and unbalanced exploitation and exploration capabilities. To improve the performance of RSA and enhance the search capability, this paper proposes a modified variant of RSA, named MRSA. To improve the population diversity, an adaptive chaotic reverse learning strategy is proposed to optimize from the initialization and in each iteration update. To balance exploitation and exploration, an elite alternative pool strategy was developed. A shifted distribution estimation strategy was used to modify all the individuals and guide the evolutionary direction. To fully validate the performance of MRSA, 23 benchmark functions, IEEE CEC2017 benchmark functions, and robot path planning problems are used for testing. The superiority of the proposed algorithm is demonstrated by a convergence analysis, stability analysis, and statistical tests.

The rest of this paper is organized as follows. Section 2 provides a review of the basic RSA. The proposed MRSA is described in detail in Section 3. In Section 4, the effectiveness of the proposed improved strategy and the superiority of the modified algorithm are verified using classical test functions, IEEE CEC2017 benchmark functions, and robot path planning problems. Finally, Section 5 provides the conclusion and discusses future work.

2. Reptile Search Algorithm

In this section, the basic procedures of RSA are presented. RSA is a swarm-based metaheuristic algorithm inspired by the enveloping mechanism, hunting mechanism, and social behavior of crocodiles.

2.1. Initialization Phase. RSA is similar to other metaheuristics in that the initial solution is generated randomly in the solution space. The initialization formula is as follows:

$$X_i^1 = LB + \text{rand} \cdot (UB - LB), \tag{1}$$

where X_i^1 is the i^{th} initial individual and LB and UB are the upper and lower boundaries of the search space, respectively.

2.2. Encircling Phase (Exploration). Crocodiles perform high and sprawl walks during the global search phase. In RSA, the search strategy is determined by the number of current iterations. When $t \le 0.25T$, RSA performs a high walk. When $t \le 0.5T$ and $t > 0.25T$, the RSA performs a sprawl walk. The specific mathematical models of the mechanism are described as follows:

$$X_i^{t+1} = \begin{cases} X_{best}^t - \eta_i \times \beta - R_i^t \times rand, t \le \dfrac{T}{4}, \\ X_{best}^t \times X_{rand}^t \times ES \times rand, t \le \dfrac{T}{2} \ and \ t > \dfrac{T}{4}, \end{cases} \tag{2}$$

$$\eta_i = X_{best}^t \times P_i, \tag{3}$$

(1) Initialize RSA parameters and generate initial population randomly
(2) While $t < T$
(3) Calculate the Fitness of each solution
(4) Find the Best solution so far
(5) Update the ES using equation (5)
(6) For each crocodile X_i do
(7) Update the η, R, and values using equations (3), (4), and (6), respectively
(8) If $t < 0.25T$
(9) Calculate the new position X_i using equation (2)
(10) Else if $t \leq 0.5T$ and $t > 0.25T$
(11) Calculate the new position X_i using equation (2)
(12) Else if $t \leq 0.75T$ and $t > 0.5T$
(13) Calculate the new position X_i using equation (7)
(14) Else
(15) Calculate the new position X_i using equation (7)
(16) End if
(17) Calculate the fitness and select the better one
(18) End for
(19) $t = t + 1$
(20) End while
(21) Return the best position and fitness

Algorithm 1: Pseudocode of the reptile search algorithm (RSA).

Table 1: Set of chaotic functions.

ID	Type	Function
1	Chebyshev map	$x_{i+1} = \cos(i \cos^{-1}(x_i))$
2	Circle map	$x_{i+1} = \mathrm{mod}(x_i + b - (a/2\pi)\sin(2\pi x_i)), 1), a = 0.5 \, and \, b = 0.2$
3	Gauss map	$x_{i+1} = \begin{cases} 1 & x_i = 0 \\ 1/\mathrm{mod}(x_i, 1) & \text{otherwise} \end{cases}$
4	Iterative map	$x_{i+1} = \sin(a\pi/x_i), a = 0.7$
5	Logistic map	$x_{i+1} = a x_i (1 - x_i), a = 4$
6	Pricewise map	$x_{i+1} = \begin{cases} x_i/p & 0 \leq x_i < p \\ (x_i - p)/(0.5 - p) & p \leq x_i < 0.5 \\ (1 - x_i - p)/(0.5 - p) & 0.5 \leq x_i < 1 - p \\ (1 - x_i)/(p) & 1 - p \leq x_i < 1 \end{cases}$
7	Sine map	$x_{i+1} = a/4 \cdot \sin(\pi x_i), a = 4$
8	Singer map	$x_{i+1} = \mu(7.86x_i - 23.32x_i^2 + 28.75x_i^3 - 13.301875x_i^4), \mu = 1.07$
9	Sinusoidal map	$x_{i+1} = a x_i^2 \sin(\pi x_i), a = 2.3$
10	Tent map	$x_{i+1} = \begin{cases} x_i/0.7 & x_i < 0.7 \\ 10/3 \times (1 - x_i) & x_i \geq 0.7 \end{cases}$

$$R_i = \frac{X_{best}^t - X_i^t}{X_{best}^t + \varepsilon}, \tag{4}$$

$$ES = 2 \times r_1 \times \left(1 - \frac{1}{T}\right), \tag{5}$$

$$P_i = \alpha + \frac{X_i^t - M\left(X_i^t\right)}{X_{\mathrm{best}}^t \times (UB - LB) + \varepsilon}, \tag{6}$$

where X_{best}^t is the current best solution, t is the current number of iterations, T is the maximum number of iterations, β is a constant taking the value of 0.1 to control the speed of exploration, X_{rand}^t is a randomly chosen individual, ES is a random value decreasing in the interval $[-2, 2]$, ε is a minimal value to ensure that the denominator is not equal to 0, r_1 is a random number in the interval $[-1, 1]$, α is a constant taking the value of 0.1, and rand is a random number with values from 0 to 1.

2.3. Hunting Phase (Exploitation). In RSA, crocodiles use two strategies for foraging: hunting coordination and cooperation. When $t < 0.75T$ and $t \geq 0.5T$, the RSA performs hunting coordination. When $t < T$ and $t \geq 0.75T$, a hunting cooperation strategy is employed by the RSA. The formula for position updating in the hunting phase is as follows:

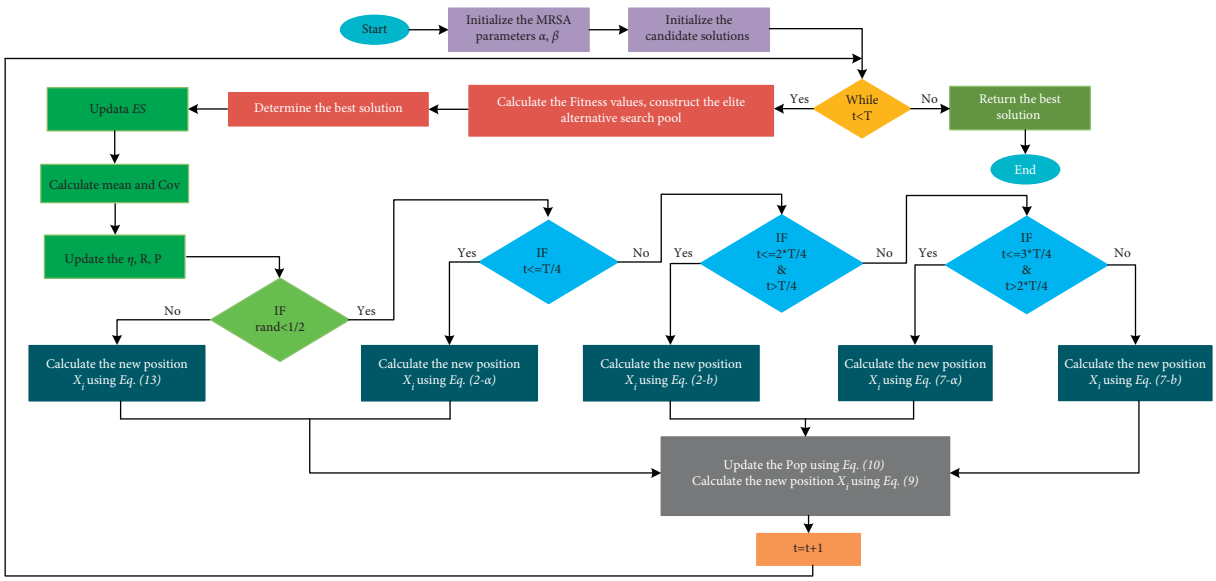


Figure 1: Flow chart of MRSA.

(1) Initialize RSA parameters and generate initial population randomly
(2) While $t < T$
(3) Calculate the Fitness of each solution, construct the elite alternative search pool based on equation (12)
(4) Find the Best solution so far
(5) Update the ES using equation (5)
(6) Calculate mean and Cov based on equation (13)
(7) For each crocodile X_i do
(8) Update the η, R, and values using equations (3), (4), and (6), respectively
(9) If rand < 0.5
(10) If $t \le 0.25T$
(11) Calculate the new position X_i using equation (2)
(12) Else if $t \le 0.5T$ and $t > 0.25T$
(13) Calculate the new position X_i using equation (2)
(14) Else if $t \le 0.75T$ and $t > 0.5T$
(15) Calculate the new position X_i using equation (7)
(16) Else
(17) Calculate the new position X_i using equation (7)
(18) End if
(19) Else
(20) Calculate the new position X_i using equations (13)
(21) End if
(22) Calculate the fitness and select the better one
(23) Update the Pop using equation (10)
(24) Calculate the new position X_i using equation (9)
(25) Calculate the fitness and select the better one
(26) $t = t + 1$
(27) End while
(28) Return the best position and fitness

Algorithm 2: Pseudocode of the MRSA.

$$X_i^{t+1} = \begin{cases} X_{\text{best}}^t \times P_i \times \text{rand}, t \le \dfrac{3T}{4} \text{ and } t > \dfrac{T}{2}, \\ X_{\text{best}}^t - \eta_i \times \varepsilon - R_i^t \times \text{rand}, t \le T \text{ and } t > \dfrac{3T}{4}. \end{cases} \quad (7)$$

RSA generates the initial population randomly in the search space first and then chooses different search strategies depending on the number of iterations. The pseudocode for the RSA is shown in Algorithm 1.

Table 2: The classic benchmark functions (M: multimodal, U: unimodal, S: separable, N: nonseparable, Dim: dimension, Range: limits of search space, Optimum: global optimal value) [55].

Test function	Name	Type	*Dim*	Range	Optimum
$f_{01}(x) = \sum_{i=1}^{D} x_i^2$	Sphere	US	30	$[-100, 100]$	0
$f_{02}(x) = \sum_{i=1}^{D} \lvert x_i \rvert + \prod_{i-1}^{D} \lvert x_i \rvert$	Schwefel 2.22	UN	30	$[-10, 10]$	0
$f_{03}(x) = \sum_{i=1}^{D} (\sum_{j-1}^{D} x_i)^2$	Schwefel 1.2	UN	30	$[-100, 100]$	0
$f_{04}(x) = \max_i \{\lvert x_i \rvert, 1 \le i \le D\}$	Schwefel 2.21	US	30	$[-100, 100]$	0
$f_{05}(x) = \sum_{i=1}^{D} 100(x_{i+1}^2 - x_i^2)^2 + (x_i - 1)^2$	Rosenbrock	UN	30	$[-30, 30]$	0
$f_{06}(x) = \sum_{i=1}^{D} (\lfloor x_i + 0.5 \rfloor)^2$	Step	US	30	$[-100, 100]$	0
$f_{07}(x) = \sum_{i=1}^{D} i x_i^4 + \text{random}[0, 1)$	Quartic	US	30	$[-1.28, 1.28]$	0
$f_{08}(x) = \sum_{i=1}^{D} -x_i \sin(\sqrt{\lvert x_i \rvert})$	Schwefel 2.26	MS	30	$[-500, 500]$	-418.9829D
$f_{09}(x) = \sum_{i=1}^{D} (x_i^2 - 10\cos(2\pi x_i) + 10)$	Rastrigin	MS	30	$[-5.12, 5.12]$	0
$f_{10}(x) = 20 + e - 20\exp(-0.2\sqrt{1/D \sum_{i=1}^{D} x_i^2}) - \exp(1/D \sum_{i=1}^{D} \cos(2\pi x_i))$	Ackley	MS	30	$[-32, 32]$	$8.8818e-16$
$f_{11}(x) = 1/4000 \sum_{i=1}^{D} (x_i^2) - (\prod_{i=1}^{D} \cos(x_i/\sqrt{i})) + 1$	Griewank	MN	30	$[-600, 600]$	0
$f_{12}(x) = \pi/D \left\{ 10\sin(\pi y_i) + \sum_{i-1}^{D-1} (y_i - 1)^2 [1 + 10\sin^2(\pi y_{i+1})] + (y_D - 1)^2 \right\} + \sum_{i-1}^{D} u(x_i, 10, 100, 4)$, where $y_i = 1 + x_i + u(x_i, a, k, m) = \begin{cases} k(x_i - a)^m & x_i > a \\ 0 & -a < x_i < a \\ k(-x_i - a)^m & x_i < a \end{cases}$	Penalized	MN	30	$[-50, 50]$	0
$f_{13}(x) = 0.1(\sin^2(3\pi x_i) + \sum_{i=1}^{D} (x_i - 1)^2 [1 + \sin^2(3\pi x_i)] + (x_D - 1)^2 1 + \sin^2(2\pi x_D))] + \sum_{i-1}^{D} u(x_i, 5, 100, 4)$	Penalized2	MN	30	$[-50, 50]$	0
$f_{14}(x) = (1/500 + \sum_{j=1}^{25} 1/j + \sum_{i=1}^{2} (x_i - a_{ij})^6)^{-1}$	Foxholes	MS	2	$[-65.53, 65.53]$	0.998004
$f_{15}(x) = \sum_{i=1}^{11} (a_i - x_1(b_i^2 + b_i x_2)/(b_i^2 + b_i x_3 + x_4))^{-1}$	Kowalik	MS	4	$[-5, 5]$	0.0003075
$f_{16}(x) = 4x_1^2 - 2.1x_1^4 + 1/3x_1^6 + x_1x_2 - 4x_2^2 + x_2^4$	Six-hump camel back	MN	2	$[-5, 5]$	-1.03163
$f_{17}(x) = (x_2 - 5.1/4\pi^2 * x_1^2 + 5/\pi * x_1 - 6)^2 + 10(1 - 1/8\pi)\cos x_1 + 10$	Branin	MS	2	$[-5, 10] \times [0, 15]$	0.398
$f_{18}(x) = [1 + (x_1 + x_2 + 1)^2 (19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)] \times [30 + (2x_1 - 3x_2)^2 (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)]$	Goldstein-Price	MN	2	$[-5, 5]$	3
$f_{19}(x) = -\sum_{i=1}^{4} (c_i \exp - (\sum_{j-1}^{3} a_{ij}(x_j - p_{ij}{}^2))$	Hartman 3	MN	3	$[0, 1]$	-3.8628
$f_{20}(x) = -\sum_{i=1}^{4} (c_i \exp - (\sum_{j-1}^{6} a_{ij}(x_j - p_{ij}{}^2))$	Hartman 6	MN	6	$[0, 1]$	-3.32
$f_{21}(x) = -\sum_{i=1}^{5} [(X - a_i)(X - a_i)^T + c_i]^{-1}$	Langermann 5	MN	4	$[0, 10]$	-10.1532
$f_{22}(x) = -\sum_{i=1}^{7} [(X - a_i)(X - a_i)^T + c_i]^{-1}$	Langermann 7	MN	4	$[0, 10]$	-10.4029
$f_{23}(x) = -\sum_{i=1}^{10} [(X - a_i)(X - a_i)^T + c_i]^{-1}$	Langermann 10	MN	4	$[0, 10]$	-10.5364

3. The Proposed RSA Variant

To enhance the performance of the basic RSA, three improvement strategies are proposed in this paper. An adaptive chaotic reverse learning strategy is first introduced to enhance the population diversity of RSA using the characteristics of chaotic mapping and reverse learning. Second, an elite alternative pooling strategy is used to balance the development and exploration of RSA. In addition, a distribution estimation strategy is used to modify the evolutionary direction. By sampling the dominant population information, the population direction is better guided, thus improving the algorithm's convergence efficiency. The three improvement strategies are described in detail in the following.

3.1. Adaptive Chaotic Reverse Learning Strategy. One of the shortcomings of the metaheuristic algorithm is that the diversity of the algorithmic population continues to diminish as the optimization proceeds. To enhance the diversity of the algorithms, the researchers employ different approaches. The reverse learning strategy is a new technique that is widely used to improve population diversity. The reason for the popularity of reverse learning is that extensive literature shows that the probability of a reverse solution

Table 3: Results of 10 chaotic maps on all benchmark functions.

Function	Index	MRSA-C1	MRSA-C2	MRSA-C3	MRSA-C4	MRSA-C5	MRSA-C6	MRSA-C7	MRSA-C8	MRSA-C9	MRSA-C10	RSA [41]
F1	Mean	2.51$E-155$	8.11$E-219$	5.13$E-226$	3.37$E-157$	4.35$E-219$	3.96$E-207$	1.73$E-223$	4.11$E-224$	6.09$E-226$	8.09$E-260$	2.48$E-65$
	Std	1.37$E-154$	0.00$E+00$	0.00$E+00$	1.84$E-156$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	1.36$E-64$
	Rank	10	7	2	9	6	8	5	4	3	**1**	11
F2	Mean	8.71$E-82$	2.94$E-111$	4.34$E-108$	5.97$E-78$	2.47$E-109$	7.33$E-109$	2.41$E-114$	5.55$E-111$	7.68$E-104$	3.33$E-136$	7.60$E-32$
	Std	3.52$E-81$	1.61$E-110$	1.66$E-107$	2.87$E-77$	1.29$E-108$	3.33$E-108$	1.04$E-113$	2.35$E-110$	4.21$E-103$	1.82$E-135$	2.36$E-31$
	Rank	9	3	7	10	5	6	2	4	8	**1**	11
F3	Mean	7.40$E-113$	9.15$E-192$	7.28$E-180$	2.82$E-117$	2.94$E-186$	7.00$E-187$	6.73$E-185$	1.23$E-182$	1.34$E-190$	1.54$E-247$	6.80$E-43$
	Std	4.06$E-112$	0.00$E+00$	0.00$E+00$	1.55$E-116$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	3.72$E-42$
	Rank	10	2	8	9	5	4	6	7	3	**1**	11
F4	Mean	5.72$E-80$	3.02$E-106$	1.24$E-104$	5.63$E-80$	4.65$E-109$	1.13$E-107$	2.23$E-108$	8.95$E-111$	9.60$E-111$	6.19$E-137$	5.88$E-35$
	Std	2.84$E-79$	1.65$E-105$	6.79$E-104$	2.18$E-79$	2.53$E-108$	5.94$E-107$	1.22$E-107$	3.20$E-110$	3.57$E-110$	3.32$E-136$	3.22$E-34$
	Rank	10	7	8	9	4	6	5	2	3	**1**	11
F5	Mean	1.80$E-07$	4.62$E-07$	3.64$E-07$	4.71$E-08$	4.50$E-07$	3.54$E-07$	2.61$E-07$	3.13$E-07$	5.97$E-07$	1.02$E-06$	4.34$E-05$
	Std	3.25$E-07$	1.16$E-06$	1.17$E-06$	9.59$E-08$	2.05$E-06$	1.47$E-06$	5.61$E-07$	5.42$E-07$	1.68$E-06$	2.49$E-06$	6.90$E-05$
	Rank	2	8	6	**1**	7	5	3	4	9	10	11
F6	Mean	4.18$E-12$	1.24$E-12$	1.28$E-12$	5.05$E-12$	4.65$E-12$	8.84$E-13$	6.78$E-13$	1.06$E-12$	1.07$E-12$	5.88$E-13$	1.54$E-09$
	Std	1.22$E-11$	2.70$E-12$	3.04$E-12$	2.36$E-11$	2.21$E-11$	1.42$E-12$	2.25$E-12$	1.90$E-12$	2.74$E-12$	1.04$E-12$	3.30$E-09$
	Rank	8	6	7	10	9	3	2	4	5	**1**	11
F7	Mean	3.84$E-04$	2.66$E-04$	1.92$E-04$	3.40$E-04$	2.71$E-04$	2.56$E-04$	2.21$E-04$	2.02$E-04$	2.54$E-04$	1.81$E-04$	5.54$E-04$
	Std	3.47$E-04$	1.86$E-04$	1.98$E-04$	2.65$E-04$	1.56$E-04$	2.01$E-04$	2.36$E-04$	1.47$E-04$	2.30$E-04$	1.55$E-04$	5.97$E-04$
	Rank	10	7	2	9	8	6	4	3	5	**1**	11
F8	Mean	$-7.63E+03$	$-8.97E+03$	$-8.67E+03$	$-8.00E+03$	$-8.39E+03$	$-8.22E+03$	$-8.96E+03$	$-9.29E+03$	$-8.99E+03$	$-8.44E+03$	$-8.29E+03$
	Std	7.50$E+02$	1.73$E+03$	1.64$E+03$	1.04$E+03$	9.94$E+02$	1.02$E+03$	1.45$E+03$	1.74$E+03$	1.47$E+03$	5.81$E+02$	5.78$E+02$
	Rank	11	3	5	10	7	9	4	**1**	2	6	8
F9	Mean	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$
	Std	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$
	Rank	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
F10	Mean	8.88$E-16$	8.88$E-16$	8.88$E-16$	8.88$E-16$	8.88$E-16$	8.88$E-16$	8.88$E-16$	8.88$E-16$	8.88$E-16$	8.88$E-16$	8.88$E-16$
	Std	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$
	Rank	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
F11	Mean	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$
	Std	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$
	Rank	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
F12	Mean	6.09$E-13$	4.86$E-13$	6.84$E-13$	2.35$E-13$	1.55$E-12$	3.62$E-13$	3.36$E-13$	2.93$E-13$	5.16$E-13$	6.21$E-13$	4.51$E-11$
	Std	2.06$E-12$	6.87$E-13$	1.87$E-12$	4.26$E-13$	5.95$E-12$	6.65$E-13$	6.92$E-13$	5.74$E-13$	1.16$E-12$	1.59$E-12$	1.95$E-10$
	Rank	7	5	9	**1**	10	4	3	2	6	8	11
F13	Mean	5.56$E-12$	2.44$E-11$	1.86$E-11$	7.55$E-12$	1.54$E-11$	1.41$E-11$	1.00$E-11$	8.05$E-12$	1.27$E-11$	2.79$E-11$	1.84$E-09$
	Std	1.56$E-11$	6.58$E-11$	4.90$E-11$	2.39$E-11$	7.35$E-11$	2.19$E-11$	2.03$E-11$	2.37$E-11$	2.74$E-11$	1.11$E-10$	6.87$E-09$
	Rank	**1**	9	8	2	7	6	4	3	5	10	11

Table 3: Continued.

Function	Index	MRSA-C1	MRSA-C2	MRSA-C3	MRSA-C4	MRSA-C5	MRSA-C6	MRSA-C7	MRSA-C8	MRSA-C9	MRSA-C10	RSA [41]
F14	Mean	1.06$E+00$	1.79$E+00$	1.59$E+00$	1.52$E+00$	1.62$E+00$	2.17$E+00$	1.46$E+00$	2.76$E+00$	1.20$E+00$	9.98$E-01$	8.98$E+00$
	Std	3.62$E-01$	1.90$E+00$	1.86$E+00$	1.85$E+00$	1.90$E+00$	2.97$E+00$	1.83$E+00$	3.68$E+00$	6.05$E-01$	1.54$E-16$	5.34$E+00$
	Rank	2	8	6	5	7	9	4	10	3	**1**	11
F15	Mean	3.07$E-04$	3.07$E-04$	3.07$E-04$	3.07$E-04$	3.07$E-04$	3.07$E-04$	3.07$E-04$	3.07$E-04$	3.07$E-04$	3.07$E-04$	3.35$E-04$
	Std	2.19$E-11$	3.38$E-11$	1.30$E-10$	2.78$E-10$	1.02$E-10$	7.23$E-11$	6.56$E-12$	2.16$E-12$	5.91$E-12$	7.98$E-12$	1.03$E-04$
	Rank	5	6	9	10	8	7	4	**1**	2	3	11
F16	Mean	−1.03$E+00$	−1.03$E+00$	−1.03$E+00$	−1.03$E+00$	−1.03$E+00$	−1.03$E+00$	−1.03$E+00$	−1.03$E+00$	−1.03$E+00$	−1.03$E+00$	−1.03$E+00$
	Std	5.98$E-16$	5.68$E-16$	6.12$E-16$	5.76$E-16$	5.61$E-16$	5.22$E-16$	5.53$E-16$	5.83$E-16$	5.68$E-16$	5.68$E-16$	5.68$E-16$
	Rank	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
F17	Mean	3.98$E-01$	3.98$E-01$	3.98$E-01$	3.98$E-01$	3.98$E-01$	3.98$E-01$	3.98$E-01$	3.98$E-01$	3.98$E-01$	3.98$E-01$	3.98$E-01$
	Std	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$	0.00$E+00$
	Rank	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
F18	Mean	4.80$E+00$	3.00$E+00$	3.00$E+00$	3.00$E+00$	3.00$E+00$	3.00$E+00$	3.00$E+00$	3.00$E+00$	3.00$E+00$	3.00$E+00$	6.60$E+00$
	Std	6.85$E+00$	1.62$E-15$	1.51$E-15$	1.79$E-15$	1.95$E-15$	1.74$E-15$	7.71$E-11$	2.00$E-15$	1.10$E-15$	1.28$E-15$	9.34$E+00$
	Rank	10	6	2	4	8	4	9	6	**1**	2	11
F19	Mean	−3.86$E+00$	−3.86$E+00$	−3.86$E+00$	−3.86$E+00$	−3.86$E+00$	−3.86$E+00$	−3.86$E+00$	−3.86$E+00$	−3.86$E+00$	−3.86$E+00$	−3.86$E+00$
	Std	2.37$E-15$	2.42$E-15$	2.43$E-15$	2.34$E-15$	2.46$E-15$	2.42$E-15$	2.45$E-15$	2.46$E-15$	2.46$E-15$	2.39$E-15$	2.42$E-15$
	Rank	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**
F20	Mean	−3.27$E+00$	−3.27$E+00$	−3.26$E+00$	−3.30$E+00$	−3.28$E+00$	−3.27$E+00$	−3.28$E+00$	−3.27$E+00$	−3.27$E+00$	−3.27$E+00$	−3.28$E+00$
	Std	5.92$E-02$	5.99$E-02$	6.03$E-02$	4.84$E-02$	5.70$E-02$	5.99$E-02$	5.83$E-02$	5.92$E-02$	5.99$E-02$	5.92$E-02$	5.83$E-02$
	Rank	5	8	11	**1**	2	8	3	5	8	5	3
F21	Mean	−1.02$E+01$	−1.02$E+01$	−1.02$E+01$	−1.02$E+01$	−1.02$E+01$	−1.02$E+01$	−1.02$E+01$	−1.02$E+01$	−1.02$E+01$	−1.02$E+01$	−8.96$E+00$
	Std	5.63$E-15$	5.56$E-15$	5.51$E-15$	5.44$E-15$	5.78$E-15$	5.56$E-15$	5.63$E-15$	5.46$E-15$	5.55$E-15$	5.46$E-15$	2.19$E+00$
	Rank	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	11
F22	Mean	−1.04$E+01$	−1.04$E+01$	−1.04$E+01$	−1.04$E+01$	−1.04$E+01$	−1.04$E+01$	−1.04$E+01$	−1.04$E+01$	−1.04$E+01$	−1.04$E+01$	−9.16$E+00$
	Std	8.73$E-16$	8.73$E-16$	8.08$E-16$	4.66$E-16$	8.08$E-16$	8.73$E-16$	7.38$E-16$	9.33$E-16$	7.38$E-16$	1.09$E-15$	2.29$E+00$
	Rank	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	11
F23	Mean	−1.05$E+01$	−1.05$E+01$	−1.05$E+01$	−1.05$E+01$	−1.05$E+01$	−1.05$E+01$	−1.05$E+01$	−1.05$E+01$	−1.05$E+01$	−1.05$E+01$	−9.45$E+00$
	Std	2.11$E-15$	2.24$E-15$	1.92$E-15$	1.78$E-15$	1.75$E-15$	2.03$E-15$	1.75$E-15$	2.09$E-15$	2.01$E-15$	1.78$E-15$	2.20$E+00$
	Rank	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	**1**	11

approximating the global optimum is approximately fifty percent higher than the current original solution, and reverse learning strategies have been used to improve other algorithms with success [49–51]. The mathematical model of the reverse learning strategy is described as follows:

$$X_i^o = LB + UB - X_i^t, \tag{8}$$

where X_i^o is the inverse solution corresponding to X_i^t. The population diversity is related to the distribution of individuals in the search space. The more uniform the distribution of the individuals, the better the diversity. Chaotic mappings are characterized by random selection and ergodicity, which can help RSA generate new solutions and avoid premature convergence, and chaotic mappings have been successfully used to improve other algorithms [52]. Therefore, this paper combines a reverse learning strategy with chaotic mappings, called the chaotic reverse learning strategy, and it is given as follows:

$$X_i^{co} = LB + UB - \lambda_i X_i^t, \tag{9}$$

where X_i^{co} denotes the solution generated by the chaotic reverse learning mechanism corresponding to the i^{th} individual in the population. λ_i is the corresponding chaotic mapping value. There are ten common chaotic mappings, with the formulas and numerical distributions shown in Table 1.

For swarm-based algorithms, the quality of the initial population has a significant impact on the algorithm's performance. Therefore, the initial population is first generated using COBL to improve the population quality and to increase the algorithm's convergence accuracy. Second, during each iteration, the corresponding reverse population is generated using COBL and evaluated separately to retain the dominant individuals in the next generation.

In addition, as the algorithm proceeds, there will be many useless searches using the chaotic inverse learning strategy for all the individuals, which increases the computational cost and is not conducive to the convergence of the algorithm, so this paper proposes using the linear decreasing population strategy. As the iteration proceeds, the number of individuals using the chaotic inverse learning strategy is gradually reduced, and the specific mathematical formula is as follows:

$$Pop = \text{round}\left(\frac{(pop_{\min} - pop_{\max}) \cdot t}{T} + pop_{\max}\right), \tag{10}$$

where Pop denotes the number of populations using the chaotic backward learning strategy and $pop_{\max}$ and $pop_{\min}$ denote the maximum number and minimum number of populations, respectively.

3.2. Elite Alternative Pool Strategy. RSA performs position updates by following the best individual. This facilitates a faster convergence of the algorithm but diminishes population diversity and tends to trap local optimums. To maintain a balance between the exploitation and exploration of the algorithm, an elite alternative pooling strategy is proposed in this section. We place the current best three individuals into a pool as shown in the following equation:

Table 4: Parameter setting for comparison algorithm.

Algorithm	Parameters
HHO [42]	$\beta = 1.5, E_0 \in [-1, 1]$
EO [43]	$a_1 = 2, a_2 = 1$
TSA [44]	$P_{\max} = 4, P_{\min} = 1$
GWO [36]	$a = 2$ (linearly decreased over iterations)
SSA [45]	$c_1 = rand$, $c_2 = rand$
WOA [37]	$a_1 = 2$ (linearly decreased over iterations)
RSA [41]	$\alpha = 0.1, \beta = 0.1$

$$X_{eap} = \left\{X_{eap1}, X_{eap2}, X_{eap3}\right\}, \tag{11}$$

where X_{eap1}, X_{eap2}, and X_{eap3} are the three best individuals in the population thus far. The food source is chosen randomly from these three individuals each time. By using the elite alternative pooling strategy, the position of the food source changes from the best individual to one of the best three individuals. This goes some way to avoiding the premature convergence of the algorithm due to the best individual falling into a local optimum. To better balance the development and exploration of the algorithm, we also put the globally optimal individuals into the elite alternative pool to ensure that each individual has the opportunity to move closer to the optimal individual and ensure the convergence efficiency of the algorithm. Thus, the final mathematical model of the elite alternative pooling strategy is described as follows:

$$\mathbf{X}_{eap} = \left\{\mathbf{X}_{eap1}, \mathbf{X}_{eap2}, \mathbf{X}_{eap3}, \mathbf{X}_{\text{best}}\right\}. \tag{12}$$

3.3. Shifted Distribution Estimation Strategy. RSA searches by following the optimal individuals, ignoring valid information from other individuals. To make full use of the position information of the dominant population, some scholars use a distribution estimation strategy for implementation [53, 54]. This strategy uses the current dominant population to calculate a probability distribution model, generates a new offspring population based on the sampling of the probability distribution model, and eventually obtains the optimal solution through continuous iteration. In addition to using the dominant population, this paper considers a modification of it by introducing information about the optimal individual and its own position and proposes a shifted distribution estimation strategy. The mathematical model is as follows:

$$\begin{aligned}
X_i^{t+1} &= \mathbf{mean} + \mathbf{y}, \mathbf{y} \sim N(0, \mathbf{Cov}),\\
\mathbf{mean} &= \frac{\left(X_{\text{best}}^t + X_{\text{mean}}^t + X_i^t\right)}{3},\\
\mathbf{Cov}(i) &= \frac{1}{NP/2}\sum_{i=1}^{NP/2}\left(X_i^{t+1} - X_{\text{mean}}^t\right) \times \left(X_i^t - X_{\text{mean}}^t\right)^T,\\
X_{\text{mean}}^t &= \sum_{i=1}^{NP/2} \omega_i \times X_i^t,\\
\omega_i &= \frac{\ln(0.5NP + 0.5) - \ln(i)}{\sum_{i=1}^{NP/2}(\ln(0.5NP + 0.5) - \ln i)}
\end{aligned} \tag{13}$$

Table 5: Test results of different algorithms for F1–F13 (Dim = 30).

30D	MRSA	HHO [42]	EO [43]	TSA [44]	GWO [36]	SSA [45]	WOA [37]	RSA [41]
F1	**0.00E + 00**	1.06$E-187$	1.68$E-72$	6.24$E-27$	2.09$E-34$	1.14$E-02$	2.32$E-166$	2.10$E-182$
F2	**1.29E − 248**	4.21$E-100$	2.03$E-42$	1.12$E-17$	6.77$E-21$	1.43$E+01$	5.11$E-107$	8.28$E-93$
F3	**0.00E + 00**	5.81$E-135$	5.43$E-06$	7.48$E+02$	7.43$E-01$	1.91$E+04$	7.02$E+05$	6.13$E-116$
F4	**7.53E − 256**	3.88$E-96$	1.87$E-11$	2.95$E+01$	6.86$E-05$	2.05$E+01$	7.27$E+01$	4.55$E-109$
F5	**3.89E − 07**	3.67$E-03$	9.41$E+01$	9.80$E+01$	9.72$E+01$	6.74$E+02$	9.73$E+01$	4.69$E-05$
F6	**4.26E − 10**	3.74$E-05$	9.21$E-02$	1.36$E+01$	7.86$E+00$	1.64$E-02$	5.45$E-01$	4.76$E-08$
F7	6.11$E-05$	**4.18E − 05**	7.93$E-04$	1.32$E-02$	1.76$E-03$	8.03$E-01$	8.11$E-04$	2.38$E-04$
F8	−3.60$E+04$	**−4.19E + 04**	−2.92$E+04$	−1.42$E+04$	−1.61$E+04$	−2.38$E+04$	−4.01$E+04$	−2.47$E+04$
F9	**0.00E + 00**	**0.00E + 00**	**0.00E + 00**	9.12$E+02$	2.23$E-01$	1.34$E+02$	3.79$E-15$	**0.00E + 00**
F10	**8.88E − 16**	**8.88E − 16**	7.99$E-15$	4.96$E-14$	6.97$E-14$	5.14$E+00$	4.91$E-15$	**8.88E − 16**
F11	**0.00E + 00**	**0.00E + 00**	**0.00E + 00**	2.04$E-03$	2.57$E-03$	1.01$E-01$	**0.00E + 00**	**0.00E + 00**
F12	**9.70E − 12**	3.36$E-07$	4.13$E-04$	1.01$E+01$	1.83$E-01$	1.15$E+01$	5.96$E-03$	3.51$E-10$
F13	**3.42E − 10**	1.51$E-05$	2.30$E+00$	1.17$E+01$	5.60$E+00$	1.61$E+02$	7.29$E-01$	2.92$E-08$

Table 6: Test results of different algorithms for F1–F13 (Dim = 100).

100D	MRSA	HHO [42]	EO [43]	TSA [44]	GWO [36]	SSA [45]	WOA [37]	RSA [41]
F1	**0.00E + 00**	3.63$E-193$	7.51$E-73$	1.25$E-27$	3.22$E-34$	1.03$E-02$	4.14$E-166$	9.57$E-183$
F2	**3.08E − 251**	4.29$E-100$	2.29$E-42$	1.10$E-17$	7.27$E-21$	1.23$E+01$	3.27$E-106$	2.05$E-101$
F3	**0.00E + 00**	1.47$E-142$	2.51$E-06$	1.03$E+03$	9.92$E-01$	2.04$E+04$	6.76$E+05$	1.78$E-122$
F4	**1.01E − 254**	4.63$E-93$	2.04$E-11$	2.98$E+01$	4.34$E-04$	2.07$E+01$	7.47$E+01$	8.33$E-93$
F5	**1.86E − 07**	4.44$E-03$	9.39$E+01$	9.81$E+01$	9.72$E+01$	5.52$E+02$	9.72$E+01$	3.67$E-05$
F6	**2.52E − 11**	2.76$E-05$	1.56$E-01$	1.38$E+01$	7.65$E+00$	6.61$E-03$	5.85$E-01$	4.85$E-08$
F7	7.56$E-05$	**3.24E − 05**	8.65$E-04$	1.63$E-02$	1.75$E-03$	7.38$E-01$	1.66$E-03$	2.54$E-04$
F8	−3.36$E+04$	**−4.19E + 04**	−2.92$E+04$	−1.45$E+04$	−1.62$E+04$	−2.38$E+04$	−3.72$E+04$	−2.50$E+04$
F9	**0.00E + 00**	**0.00E + 00**	**0.00E + 00**	9.28$E+02$	4.94$E-01$	1.29$E+02$	3.79$E-15$	**0.00E + 00**
F10	**8.88E − 16**	**8.88E − 16**	7.88$E-15$	8.02$E-02$	6.93$E-14$	5.04$E+00$	4.56$E-15$	**8.88E − 16**
F11	**0.00E + 00**	**0.00E + 00**	**0.00E + 00**	2.25$E-03$	1.61$E-03$	1.07$E-01$	2.41$E-03$	**0.00E + 00**
F12	**9.43E − 13**	2.55$E-07$	2.02$E-04$	9.78$E+00$	1.80$E-01$	1.13$E+01$	5.40$E-03$	7.39$E-10$
F13	**2.25E − 10**	1.12$E-05$	2.07$E+00$	1.19$E+01$	5.72$E+00$	1.57$E+02$	6.03$E-01$	1.13$E-08$

Table 7: Test results of different algorithms for F1–F13 (Dim = 500).

500D	MRSA	HHO [42]	EO [43]	TSA [44]	GWO [36]	SSA [45]	WOA [37]	RSA [41]
F1	**0.00E + 00**	2.61$E-195$	1.13$E-72$	7.72$E-28$	1.82$E-34$	6.70$E-03$	1.27$E-165$	7.20$E-175$
F2	**1.13E − 250**	3.07$E-99$	2.64$E-42$	9.40$E-18$	7.22$E-21$	1.31$E+01$	1.54$E-108$	7.77$E-103$
F3	**0.00E + 00**	3.88$E-121$	3.93$E-05$	5.72$E+02$	3.37$E-01$	1.67$E+04$	7.11$E+05$	5.41$E-136$
F4	**1.47E − 259**	3.18$E-98$	9.92$E-03$	3.19$E+01$	6.68$E-05$	2.01$E+01$	8.02$E+01$	2.27$E-105$
F5	**4.99E − 07**	3.65$E-03$	9.40$E+01$	9.83$E+01$	9.71$E+01$	5.06$E+02$	9.72$E+01$	3.68$E-05$
F6	**8.10E − 11**	6.80$E-05$	1.33$E-01$	1.36$E+01$	7.81$E+00$	7.64$E-03$	5.83$E-01$	5.17$E-08$
F7	7.12$E-05$	**5.01E − 05**	7.65$E-04$	1.54$E-02$	1.62$E-03$	7.27$E-01$	1.79$E-03$	2.23$E-04$
F8	−3.40$E+04$	**−4.19E + 04**	−2.90$E+04$	−1.44$E+04$	−1.63$E+04$	−2.45$E+04$	−3.99$E+04$	−2.46$E+04$
F9	**0.00E + 00**	**0.00E + 00**	**0.00E + 00**	8.69$E+02$	7.10$E-02$	1.37$E+02$	**0.00E + 00**	**0.00E + 00**
F10	**8.88E − 16**	**8.88E − 16**	7.99$E-15$	1.78$E-01$	6.98$E-14$	4.97$E+00$	3.85$E-15$	**8.88E − 16**
F11	**0.00E + 00**	**0.00E + 00**	**0.00E + 00**	3.16$E-03$	3.28$E-04$	1.10$E-01$	**0.00E + 00**	**0.00E + 00**
F12	**6.72E − 12**	3.29$E-07$	9.68$E-04$	1.03$E+01$	1.85$E-01$	1.15$E+01$	5.97$E-03$	2.03$E-10$
F13	**9.35 $E-11$**	8.81$E-06$	2.23$E+00$	1.17$E+01$	5.63$E+00$	1.55$E+02$	6.80$E-01$	3.90$E-08$

where X^t_{mean} denotes the weighted position of the dominant population and ω_i denotes the weight coefficient in the dominant population in descending order of fitness values. **Cov** is the weighted covariance matrix of the dominant populations. The flow chart of MRSA is shown in Figure 1 and its pseudocode is in Algorithm 2.

3.4. Time Complexity. The time complexity determines the operating efficiency of the algorithm. In RSA, the computational complexity of the initialization process is $O(N)$, where N is the population size. The computational complexity of the update process is $O(T \times N) + O(T \times N \times D)$, where D is the problem's dimensionality and T is the

Table 8: Test results of different algorithms for F1–F13 (Dim = 1000).

1000D	MRSA	HHO [42]	EO [43]	TSA [44]	GWO [36]	SSA [45]	WOA [37]	RSA [41]
F1	**0.00E + 00**	7.31$E-188$	6.27$E-56$	3.27$E-08$	5.24$E-10$	1.19$E+05$	8.86$E-163$	1.80$E-186$
F2	**2.83E − 236**	6.65$E-99$	2.70$E-33$	3.03$E-07$	2.64$E-05$	8.48$E+02$	2.51$E-105$	1.52$E-86$
F3	**0.00E + 00**	6.88$E-83$	2.98$E+04$	3.99$E+06$	5.80$E+05$	2.34$E+06$	1.03$E+08$	1.38$E-157$
F4	**4.64E − 266**	1.59$E-94$	8.56$E+01$	9.95$E+01$	7.27$E+01$	3.63$E+01$	8.08$E+01$	1.09$E-97$
F5	**5.37E − 05**	3.62$E-02$	9.96$E+02$	9.91$E+03$	9.97$E+02$	3.03$E+07$	9.91$E+02$	4.88$E-04$
F6	**4.95E − 07**	3.61$E-04$	1.69$E+02$	2.06$E+02$	2.02$E+02$	1.18$E+05$	2.32$E+01$	5.34$E-06$
F7	7.24$E-05$	**5.54E − 05**	1.59$E-03$	1.92$E+00$	1.30$E-02$	4.38$E+02$	1.67$E-03$	3.87$E-04$
F8	−3.82$E+05$	**−4.19E + 05**	−1.57$E+05$	−4.98$E+04$	−1.08$E+05$	−1.30$E+05$	−4.04$E+05$	−1.60$E+05$
F9	**0.00E + 00**	**0.00E + 00**	**0.00E + 00**	1.10$E+04$	1.09$E+01$	5.61$E+03$	1.21$E-13$	**0.00E + 00**
F10	**8.88E − 16**	**8.88E − 16**	1.07$E-14$	7.83$E-06$	7.54$E-07$	1.29$E+01$	3.97$E-15$	**8.88E − 16**
F11	**0.00E + 00**	**0.00E + 00**	8.88$E-17$	1.77$E-02$	3.16$E-03$	1.08$E+03$	3.70$E-18$	**0.00E + 00**
F12	**1.75E − 10**	1.50$E-07$	5.43$E-01$	3.82$E+06$	8.37$E-01$	6.96$E+04$	1.89$E-02$	1.82$E-09$
F13	**8.09E − 08**	4.45$E-05$	9.90$E+01$	4.51$E+05$	9.50$E+01$	1.41$E+07$	1.04$E+01$	7.07$E-07$

Table 9: Test results of different algorithms for F14–F23.

	MRSA	HHO [42]	EO [43]	TSA [44]	GWO [36]	SSA [45]	WOA [37]	RSA [41]
F14	1.13$E+00$	1.06$E+00$	**9.98E − 01**	8.28$E+00$	3.00$E+00$	**9.98E − 01**	1.78$E+00$	5.34$E+00$
F15	**3.07E − 04**	3.46$E-04$	3.68$E-03$	5.17$E-03$	3.04$E-03$	7.66$E-04$	6.61$E-04$	3.21$E-04$
F16	**−1.03E + 00**	**−1.03E + 00**	**−1.03E + 00**	**−1.03E + 00**	**−1.03E + 00**	**−1.03E + 00**	**−1.03E + 00**	**−1.03E + 00**
F17	**3.98E − 01**	**3.98E − 01**	**3.98E − 01**	**3.98E − 01**	**3.98E − 01**	**3.98E − 01**	**3.98E − 01**	**3.98E − 01**
F18	**3.00E + 00**	**3.00E + 00**	**3.00E + 00**	8.40$E+00$	**3.00E + 00**	**3.00E + 00**	**3.00E + 00**	5.70$E+00$
F19	**−3.86E + 00**	**−3.86E + 00**	**−3.86E + 00**	**−3.86E + 00**	**−3.86E + 00**	**−3.86E + 00**	**−3.86E + 00**	**−3.86E + 00**
F20	**−3.27E + 00**	−3.17$E+00$	−3.26$E+00$	−3.24$E+00$	**−3.27E + 00**	−3.22$E+00$	−3.22$E+00$	−3.26$E+00$
F21	**−1.02E + 01**	−5.22$E+00$	−9.05$E+00$	−6.23$E+00$	−9.48$E+00$	−8.47$E+00$	−9.56$E+00$	−9.81$E+00$
F22	**−1.04E + 01**	−5.44$E+00$	**−1.04E + 01**	−7.62$E+00$	−1.00$E+01$	−9.97$E+00$	−9.08$E+00$	**−1.04E + 01**
F23	**−1.05E + 01**	−5.67$E+00$	−9.91$E+00$	−6.05$E+00$	−1.04$E+01$	−1.04$E+01$	−9.64$E+00$	−1.04$E+01$

maximum number of iterations, so the computational complexity of RSA is $O(N \times (T \times D + T + 1))$.

The computational complexity of MRSA is determined by six main factors (initialization process, solution update, number of fitness evaluations, chaotic reverse learning strategy, elite alternative pooling strategy, and shifted distribution estimation strategy). The computational complexity of the MRSA initialization process is $O(N)$. The computational complexity of the update process is $O(T \times N) + O(2 \times T \times N \times D)$. Therefore, the computational complexity of MRSA is $O(N \times (2 \times T \times D + T + 1))$. The introduction of three improved strategies causes the computational complexity of MRSA to increase slightly compared to RSA. RSA and MRSA can be considered to have similar levels of operating efficiency.

4. Experimental Results and Discussion

In this section, we first evaluate various chaotic mapping combination algorithms using benchmark test functions and then determine which chaotic mapping sequence to use in combination with the adaptive reverse learning strategy. The performance of MRSA is then evaluated, and 23 benchmark functions, IEEE CEC2017 benchmark functions, and robot path planning problems are compared with other state-of-the-art algorithms.

4.1. Benchmark Test Functions. This section uses 23 benchmark test functions that are commonly found in the literature. These benchmark test functions include seven unimodal functions, six multimodal functions, and ten fixed dimensional functions [55]. Unimodal functions F1–F7 have only one global optimum and are primarily used to test the local exploitation capabilities of the algorithms. The multimodal functions have multiple local minimums and can be used to check the global exploration capability and local optimum avoidance capability of the algorithm. Details of the benchmark test functions are shown in Table 2.

4.2. Chaos Mapping Selection Test. The adaptive chaotic reverse learning strategy proposed in this paper combines a chaotic mapping and a reverse learning mechanism. To verify which chaotic mapping is employed, each of the 10 chaotic mappings is combined with a reverse learning mechanism. The MRSA using the chaotic mapping with ID 1 is called MRSA-C1. The rest of the MRSA algorithms using chaotic mappings are named similarly. For a fair comparison, the number of populations was set to 50, and the maximum number of iterations was set to 300 on the same experimental platform. All the algorithms were programmed using MATLAB R2016b, the computer operating system was Windows 10, and the processor was AMD R5 3600 × 16 GB. Table 3 shows the statistical results of each algorithm run

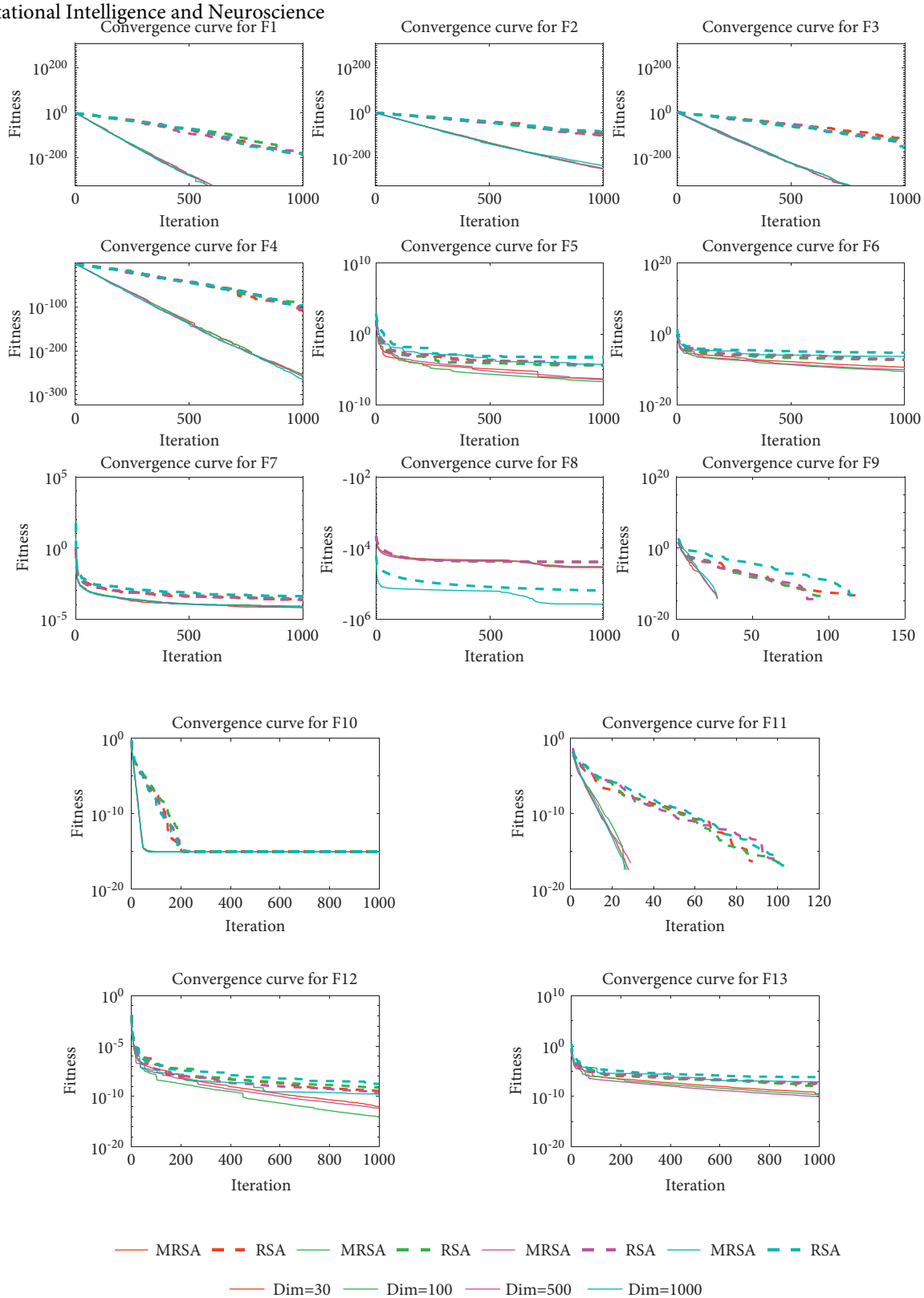


Figure 2: Convergence curves for different dimensional functions.

independently 30 times. In presenting the simulation results, "avg" is the average of the best candidate solutions obtained, and "std" is the standard deviation of these values.

As shown in Table 3, MRSA-C1 (Chebyshev map) and MRSA-C9 (Sinusoidal map) outperformed RSA on 10 out of 23 functions. MRSA-C2 (Circle map), MRSA-C3 (Gauss map), MRSA-C5 (Logistic map), MRSA-C6 (Pricewise map), and MRSA-C7 (Sine map) outperformed RSA in 9 functions. MRSA-C4 (Iterative map) outperformed RSA in 12 functions. MRSA-C8 (Singer map) performed better than

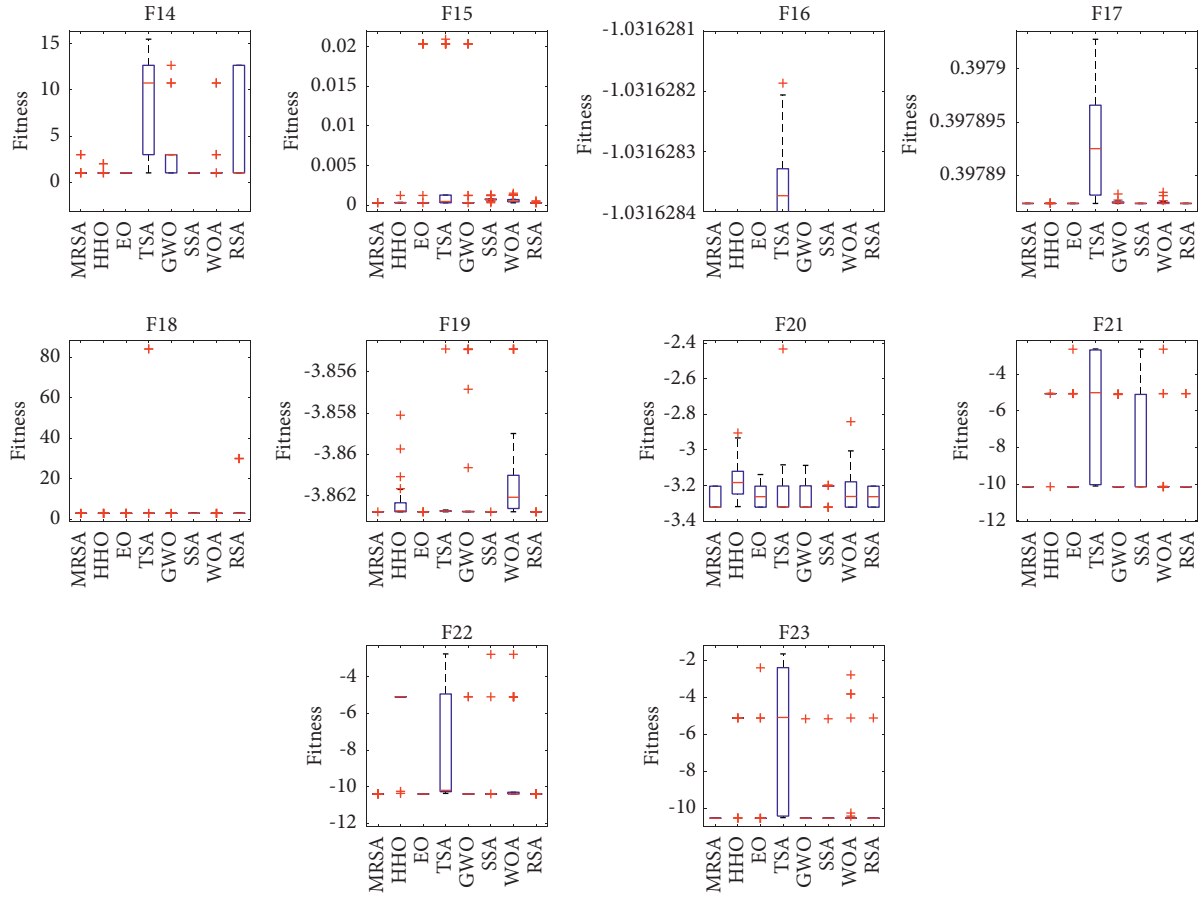


Figure 3: Box plot of different test functions.

Table 10: The Friedman test results for different algorithms.

Function	Type	MRSA	HHO [42]	EO [43]	TSA [44]	GWO [36]	SSA [45]	WOA [37]	RSA [41]
F1–F13	Dim = 30	1.58	2.42	4.19	7.00	6.08	7.23	4.77	2.73
	Dim = 100	1.54	2.38	4.15	7.08	5.85	7.23	5.08	2.69
	Dim = 500	1.62	2.62	4.31	7.15	5.77	7.23	4.69	2.62
	Dim = 1000	1.50	2.42	4.88	7.23	5.92	7.23	4.31	2.50
F14–F23	Fixed dim	1.75	5.80	3.05	7.10	5.20	4.20	5.50	3.40
All cases		**1.59**	3.00	4.17	7.11	5.79	6.74	4.84	2.76

RSA in 11 functions. MRSA-C10 (Tent map) outperformed RSA on 16 out of 23 functions. Remarkably, all the improved algorithms perform no worse than RSA in at least 22 functions, indicating that the improved strategies proposed in this paper effectively improve the algorithm's performance. Furthermore, the best of the ten chaotic mapping combination algorithms is MRSA-C10. Therefore, the MRSA-C10 algorithm was used to participate in the tests in the comparison that followed.

4.3. Performance Comparison Tests of MRSA with Other Advanced Algorithms on 23 Benchmark Functions. To verify the performance of the MRSA algorithm, the modified algorithm was compared with the original RSA [41], HHO [42], EO [43], TSA [44], GWO [36], SSA [45], and WOA [37]. The parameters of all the algorithms were set according to the original paper to ensure the performance of the comparison algorithms, as shown in Table 4. Given that F1–F13 are the multidimensional functions used in this section, the thirteen functions were solved under Dim = 30, 100, 500, and 1000. The means obtained by these algorithms are recorded, as shown in Tables 5–8.

The results in Tables 5–8 show that MRSA achieves better results in most of the functions. Specifically, MRSA obtains satisfactory results for the unimodal functions F1–F7 in both the low and high dimensions. MRSA achieves a stable optimal value of 0 when solving for F1 and F3 and remains so as the dimensionality increases. For the other unimodal functions, HHO outperforms MRSA in solving F7. The unimodal function results show that MRSA outperforms RSA in all the functions, and MRSA does not show

Table 11: The Wilcoxon signed-rank test results for different algorithms.

Functions type	Comparison	p value	$\alpha = 0.05$
F1–F13 (Dim = 30)	MRSA versus HHO [42]	**0.182338**	No
	MRSA versus EO [43]	0.001871	Yes
	MRSA versus TSA [44]	0.001306	Yes
	MRSA versus GWO [36]	0.001306	Yes
	MRSA versus SSA [45]	0.001306	Yes
	MRSA versus WOA [37]	0.02537	Yes
	MRSA versus RSA [41]	0.002873	Yes
F1–F13 (Dim = 100)	MRSA versus HHO [42]	**0.209427**	No
	MRSA versus EO [43]	0.001871	Yes
	MRSA versus TSA [44]	0.001306	Yes
	MRSA versus GWO [36]	0.001306	Yes
	MRSA versus SSA [45]	0.001306	Yes
	MRSA versus WOA [37]	0.017496	Yes
	MRSA versus RSA [41]	0.002873	Yes
F1–F13 (Dim = 500)	MRSA versus HHO [42]	**0.182338**	No
	MRSA versus EO [43]	0.001871	Yes
	MRSA versus TSA [44]	0.001306	Yes
	MRSA versus GWO [36]	0.001306	Yes
	MRSA versus SSA [45]	0.001306	Yes
	MRSA versus WOA [37]	0.02313	Yes
	MRSA versus RSA [41]	0.002873	Yes
F1–F13 (Dim = 1000)	MRSA versus HHO [42]	**0.157939**	No
	MRSA versus EO [43]	0.001944	Yes
	MRSA versus TSA [44]	0.001306	Yes
	MRSA versus GWO [36]	0.001306	Yes
	MRSA versus SSA [45]	0.001306	Yes
	MRSA versus WOA [37]	0.017496	Yes
	MRSA versus RSA [41]	0.002873	Yes
F14–F23	MRSA versus HHO [42]	**0.083131**	No
	MRSA versus EO [43]	**0.192518**	No
	MRSA versus TSA [44]	0.010862	Yes
	MRSA versus GWO [36]	0.032969	Yes
	MRSA versus SSA [45]	**0.126279**	No
	MRSA versus WOA [37]	0.019059	Yes
	MRSA versus RSA [41]	0.019059	Yes

a significant decrease in performance as the dimensionality increases, which indicates that the improvement strategy proposed in this paper greatly improves the development capability of RSA. For the variable dimensional multimodal functions F8–F13, MRSA, RSA, HHO, and EO consistently achieve their respective optimal solutions in different dimensions when solving F9–F11. HHO and WOA outperform MRSA in solving F8. MRSA achieves a stable optimal value of 0 when solving for F9 and F11 and remains so as the dimensionality increases. MRSA shows the best performance in solving F12 and F13, outperforming all the compared algorithms. It is worth noting that MRSA does not perform any less than RSA in all the multimodal functions in the different dimensions and has significant improvements in three of the six variable dimensional multimodal functions, which indicates that MRSA has a better global search capability, and the improvement strategy proposed in this paper is well suited to enhance the population diversity and to expand the search range of the population, thus improving the exploration capability of the algorithm.

Table 9 presents the test results when different algorithms solve the fixed dimensional multimodal function. The comparison shows that HHO, EO, and SSA outperform MRSA on F14. For F15–F23, MRSA performs best in all the tested functions. In particular, MRSA provides better solutions in all the test functions compared to RSA. Since fixed-dimension functions are usually used to test the ability of an algorithm to maintain a balance between development and exploration, the above analysis shows that the MRSA proposed in this paper is able to balance the development and exploration capabilities effectively and has a strong local optimum avoidance capability.

The convergence speed and convergence accuracy are important indicators of the performance of the algorithm. Figure 2 shows the mean convergence curves of MRSA and RSA when solving the test functions in different dimensions. It can be seen that MRSA has a faster convergence speed and a better convergence accuracy in different dimensions. Moreover, the convergence speed and convergence accuracy of MRSA do not decrease much with increasing

Table 12: Summary of CEC2017 benchmark functions [41].

Type	Number	Function name	$f_i(x^*)$
Unimodal	3	Shifted and rotated Zakharov function	300
Multimodal	4	Shifted and rotated Rosenbrock's function	400
	5	Shifted and rotated Rastrigin's function	500
	6	Shifted and rotated expanded Scaffer's F6 function	600
	7	Shifted and rotated Lunacek bi-Rastrigin function	700
	8	Shifted and rotated noncontinuous Rastrigin's function	800
	9	Shifted and rotated Levy function	900
	10	Shifted and rotated Schwefel's function	1000
Hybrid	11	Hybrid function 1 ($N=3$)	1100
	12	Hybrid function 2 ($N=3$)	1200
	13	Hybrid function 3 ($N=3$)	1300
	14	Hybrid function 4 ($N=4$)	1400
	15	Hybrid function 5 ($N=4$)	1500
	16	Hybrid function 6 ($N=4$)	1600
	17	Hybrid function 6 ($N=5$)	1700
	18	Hybrid function 6 ($N=5$)	1800
	19	Hybrid function 6 ($N=5$)	1900
	20	Hybrid function 6 ($N=6$)	2000
Composition	21	Composition function 1 ($N=3$)	2100
	22	Composition function 2 ($N=3$)	2200
	23	Composition function 3 ($N=4$)	2300
	24	Composition function 4 ($N=4$)	2400
	25	Composition function 5 ($N=5$)	2500
	26	Composition function 6 ($N=5$)	2600
	27	Composition function 7 ($N=6$)	2700
	28	Composition function 8 ($N=6$)	2800
	29	Composition function 9 ($N=3$)	2900
	30	Composition function 10 ($N=3$)	3000

dimensionality, which indicates that the improvement strategy proposed in this paper can effectively improve the convergence ability of RSA and thus obtain better optimization results.

To analyze the distribution characteristics of each algorithm in the fixed dimensional test function, box plots were drawn based on the results obtained by solving F14–F23, as shown in Figure 3. For each algorithm, the center mark of each box indicates the median of the results of 30 runs, and the bottom and top edges of each box indicate the trivial and quartiles, respectively. "The "+" sign indicates bad values that are not inside the box. As seen from Figure 3, MRSA has no bad values for F17 and F21–F23, which indicates that the distribution of the solutions obtained by MRSA is more concentrated and MRSA is more stable. For the other test functions with some bad values, MRSA outperforms the comparison algorithm in terms of maximum, minimum, and median values, and the distribution of the solutions obtained by MRSA is more concentrated compared to the comparison algorithm. Therefore, MRSA solves the test function with better stability compared to the other comparison algorithms.

Apart from the convergence and stability analysis, to further analyze the experimental results, the Friedman test and Wilcoxon's signed-rank test were used for multiple comparisons in this paper. Table 10 is the Friedman test showing the average ranking results of each algorithm. The overall ranking value of MRSA is 1.59, which ranks first among all the algorithms. The remaining seven algorithms are ranked as follows: RSA, HHO, EO, WOA, GWO, SSA, and TSA. In solving F1–F13 in different dimensions, MRSA is ranked first, and HHO and RSA are ranked second and third, respectively. For fixed dimensions F14–F23, MRSA, EO, and RSA ranked in the top three. In either case, MRSA ranks better than RSA. The results of Wilcoxon's signed-rank test are shown in Table 11. In the case of F1–F13 (Dim = 30, 100, 500, and 1000), MRSA outperformed EO, TSA, GWO, SSA, WOA, and RSA at the 0.05 significance level, but there was no significant difference between MRSA and HHO. In the case of F14–F23, MRSA outperformed TSA, GWO, WOA, and RSA at the 0.05 significance level, but there was no significant difference between MRSA and HHO, EO, or SSA, which statistically proves that the improvement strategy proposed in this paper can effectively help MRSA balance the exploitation and exploration capabilities and has a better local optimal avoidance ability.

Table 13: Parameters setting.

Algorithm	Parameters setting
BOA [38]	$a = 0.1, c = 0.01, p = 0.6$
HHO [42]	$\beta = 1.5, E_0 \in [-1, 1]$
AOA [47]	$Mop_{\max} = 1, Mop_{\min} = 0.2, C = 1, a = 5, Mu = 0.499$
SSA [45]	$c_1 = \text{rand}, c_2 = \text{rand}$
PFA [48]	$u_1 = -1 + 2\text{rand}, u_2 = -1 + 2\text{rand}$
TDO [46]	~

Table 14: Statistical results of seven algorithms in the CEC2017 test.

		BOA [38]	HHO [42]	AOA [47]	SSA [45]	PFA [48]	TDO [46]	MRSA
F3	Mean	3.82E+04	1.68E+03	6.91E+04	8.40E+04	4.66E+04	3.77E+04	3.01E−06
	Std	6.97E+03	7.95E+02	1.15E+04	6.59E+03	1.22E+04	3.65E+03	5.54E−07
	Rank	4	2	6	7	5	3	**1**
F4	Mean	9.33E+03	1.23E+02	7.61E+03	1.44E+03	9.80E+01	4.96E+02	7.24E+01
	Std	1.29E+03	3.33E+01	2.45E+03	1.09E+03	1.77E+01	2.16E+01	2.96E+01
	Rank	7	3	6	5	2	4	**1**
F5	Mean	3.49E+02	2.05E+02	2.95E+02	3.50E+02	1.14E+02	6.19E+02	8.24E+01
	Std	2.16E+01	3.62E+01	3.20E+01	4.40E+01	3.11E+01	1.56E+01	2.43E+01
	Rank	5	3	4	6	2	7	**1**
F6	Mean	6.63E+01	5.62E+01	6.21E+01	8.06E+01	1.47E+01	6.00E+02	2.03E+01
	Std	5.76E+00	5.92E+00	6.71E+00	8.84E+00	4.99E+00	3.05E−02	5.92E+00
	Rank	5	3	4	6	**1**	7	2
F7	Mean	5.57E+02	4.98E+02	6.00E+02	7.12E+02	1.34E+02	8.49E+02	1.46E+02
	Std	3.17E+01	6.57E+01	5.66E+01	6.85E+01	3.12E+01	1.47E+01	3.38E+01
	Rank	4	3	5	6	**1**	7	2
F8	Mean	2.93E+02	1.40E+02	2.25E+02	2.72E+02	9.97E+01	9.22E+02	7.56E+01
	Std	1.54E+01	2.13E+01	2.67E+01	4.31E+01	2.66E+01	1.58E+01	2.31E+01
	Rank	6	3	4	5	2	7	**1**
F9	Mean	6.82E+03	4.69E+03	4.50E+03	9.35E+03	2.28E+02	9.01E+02	4.00E+02
	Std	8.69E+02	8.28E+02	7.24E+02	1.85E+03	1.83E+02	8.01E−01	2.11E+02
	Rank	6	5	4	7	**1**	3	2
F10	Mean	7.33E+03	4.35E+03	5.51E+03	7.05E+03	4.98E+03	5.15E+03	3.90E+03
	Std	2.85E+02	7.25E+02	5.83E+02	7.45E+02	9.01E+02	3.62E+02	6.06E+02
	Rank	7	2	5	6	3	4	**1**
F11	Mean	2.19E+03	1.61E+02	1.72E+03	3.91E+03	1.91E+02	1.17E+03	5.60E+01
	Std	6.72E+02	4.86E+01	9.74E+02	1.64E+03	5.28E+01	2.47E+01	2.69E+01
	Rank	6	2	5	7	3	4	**1**
F12	Mean	2.08E+09	7.61E+06	6.27E+09	4.69E+08	1.88E+06	1.85E+05	8.83E+02
	Std	7.43E+08	4.21E+06	2.56E+09	3.76E+08	1.97E+06	9.78E+04	7.27E+02
	Rank	6	4	7	5	3	2	**1**
F13	Mean	3.15E+08	1.51E+05	3.80E+04	8.55E+07	7.54E+04	1.21E+04	2.24E+02
	Std	2.10E+08	9.05E+04	1.71E+04	4.66E+08	4.12E+04	5.52E+03	1.60E+02
	Rank	7	5	3	6	4	2	**1**
F14	Mean	1.19E+05	3.82E+04	5.72E+04	1.50E+06	3.00E+04	2.97E+03	4.32E+01
	Std	7.62E+04	4.25E+04	4.92E+04	1.21E+06	2.94E+04	6.93E+02	1.13E+01
	Rank	6	4	5	7	3	2	**1**
F15	Mean	1.82E+06	6.86E+04	2.35E+04	1.83E+07	3.35E+04	2.40E+03	2.98E+01
	Std	1.46E+06	4.86E+04	1.22E+04	2.37E+07	1.77E+04	5.08E+02	1.67E+01
	Rank	6	5	3	7	4	2	**1**
F16	Mean	3.18E+03	1.55E+03	1.98E+03	2.74E+03	1.00E+03	2.40E+03	1.07E+03
	Std	4.12E+02	3.56E+02	5.09E+02	5.38E+02	2.63E+02	1.34E+02	3.20E+02
	Rank	7	3	4	6	**1**	5	2
F17	Mean	1.22E+03	7.48E+02	9.12E+02	1.20E+03	3.77E+02	1.88E+03	4.41E+02
	Std	2.49E+02	2.19E+02	2.67E+02	3.85E+02	1.71E+02	3.61E+01	2.10E+02
	Rank	6	3	4	5	**1**	7	2
F18	Mean	9.60E+05	6.90E+05	1.29E+06	1.51E+07	2.75E+05	6.11E+04	3.11E+01
	Std	6.22E+05	8.77E+05	1.60E+06	1.51E+07	2.82E+05	2.10E+04	5.54E+00
	Rank	5	4	6	7	3	2	**1**
F19	Mean	4.61E+06	1.46E+05	1.08E+06	4.23E+07	4.45E+04	5.14E+03	2.29E+01
	Std	4.06E+06	1.42E+05	1.39E+05	1.23E+08	3.91E+04	9.69E+02	5.65E+00
	Rank	6	4	5	7	3	2	**1**
F20	Mean	7.29E+02	6.71E+02	6.94E+02	8.59E+02	4.61E+02	2.29E+03	5.30E+02
	Std	9.88E+01	2.01E+02	1.54E+02	2.42E+02	1.52E+02	4.33E+01	1.62E+02
	Rank	5	3	4	6	**1**	7	2

Table 14: Continued.

		BOA [38]	HHO [42]	AOA [47]	SSA [45]	PFA [48]	TDO [46]	MRSA
F21	Mean	1.97E + 02	4.06E + 02	4.87E + 02	5.06E + 02	2.90E + 02	2.41E + 03	2.83E + 02
	Std	3.01E + 01	3.51E + 01	5.23E + 01	5.36E + 01	2.62E + 01	9.32E + 00	2.52E + 01
	Rank	**1**	4	5	6	3	7	2
F22	Mean	4.71E + 02	2.39E + 03	5.13E + 03	4.18E + 03	2.08E + 02	2.30E + 03	1.02E + 02
	Std	7.76E + 01	2.37E + 03	1.21E + 03	1.88E + 03	7.61E + 02	8.09E + 00	2.05E + 00
	Rank	3	5	7	6	2	4	**1**
F23	Mean	6.97E + 02	7.05E + 02	9.68E + 02	8.60E + 02	4.85E + 02	2.72E + 03	5.11E + 02
	Std	5.59E + 01	7.35E + 01	9.10E + 01	1.00E + 02	4.09E + 01	1.33E + 01	4.21E + 01
	Rank	3	4	6	5	**1**	7	2
F24	Mean	1.10E + 03	8.26E + 02	1.14E + 03	8.99E + 02	5.24E + 02	2.88E + 03	5.83E + 02
	Std	1.68E + 02	7.42E + 01	1.09E + 02	1.37E + 02	3.72E + 01	1.03E + 01	5.29E + 01
	Rank	5	3	6	4	**1**	7	2
F25	Mean	1.75E + 03	4.11E + 02	1.67E + 03	7.84E + 02	3.97E + 02	2.89E + 03	3.93E + 02
	Std	2.01E + 02	1.87E + 01	4.55E + 02	1.30E + 02	1.72E + 01	1.09E + 01	1.35E + 01
	Rank	6	3	5	4	2	7	**1**
F26	Mean	5.21E + 03	3.94E + 03	6.40E + 03	6.30E + 03	2.12E + 03	3.07E + 03	2.38E + 03
	Std	1.49E + 03	1.10E + 03	7.22E + 02	1.11E + 03	7.10E + 02	4.75E + 02	4.59E + 02
	Rank	5	4	7	6	**1**	3	2
F27	Mean	8.14E + 02	6.05E + 02	1.34E + 03	9.56E + 02	5.46E + 02	3.21E + 03	5.98E + 02
	Std	9.81E + 01	4.00E + 01	2.14E + 02	1.65E + 02	2.85E + 01	6.74E + 00	3.98E + 01
	Rank	4	3	6	5	**1**	7	2
F28	Mean	3.28E + 03	4.62E + 02	2.95E + 03	1.06E + 03	4.31E + 02	3.23E + 03	4.09E + 02
	Std	3.99E + 02	2.60E + 01	6.15E + 02	3.22E + 02	1.90E + 01	2.10E + 01	3.06E + 01
	Rank	7	3	5	4	2	6	**1**
F29	Mean	3.04E + 03	1.32E + 03	2.43E + 03	2.64E + 03	1.03E + 03	3.67E + 03	9.64E + 02
	Std	4.72E + 02	2.56E + 02	5.22E + 02	6.35E + 02	2.35E + 02	6.51E + 01	2.18E + 02
	Rank	6	3	4	5	2	7	**1**
F30	Mean	3.98E + 07	1.01E + 06	1.47E + 07	4.85E + 07	4.09E + 05	6.59E + 03	2.14E + 03
	Std	2.31E + 07	6.08E + 05	1.01E + 07	3.75E + 07	4.56E + 05	3.95E + 02	1.05E + 02
	Rank	6	4	5	7	3	2	**1**

Table 15: The Friedman test results for different algorithms.

Algorithm	Ranking
MRSA	**1.3929**
PFA [48]	2.2143
HHO [42]	3.5714
TDO [46]	5.2857
AOA [47]	5.6429
BOA [38]	6
SSA [45]	6.6071

Table 16: The Wilcoxon signed-rank test results for different algorithms.

Comparison	p value	$\alpha = 0.05$
MRSA versus BOA [38]	0.000004	YES
MRSA versus HHO [42]	0.000004	YES
MRSA versus AOA [47]	0.000004	YES
MRSA versus SSA [45]	0.000004	YES
MRSA versus PFA [48]	0.039321	YES
MRSA versus TDO [46]	0.000004	YES

4.4. Performance Comparison Tests of MRSA with Other Advanced Algorithms on CEC2017. To further verify the superior performance of the MRSA algorithm, the algorithm was tested using the IEEE CEC2017 [41] single objective test function defined in Table 12. In this section, six recently proposed algorithms were evaluated for comparison with MRSA. These state-of-the-art algorithms are BOA [38], HHO [42], AOA [47], SSA [45], PFA [48], and TDO [46]. For a fair comparison, all the algorithm parameters are set the same as those used by the authors of the original literature, as shown in Table 13. The dimension of the CEC2017 benchmark functions was set to 30 on the same experimental platform. Table 14 shows the statistical results of each algorithm run independently 51 times.

From the analysis in Table 14, we know that, for the unimodal test function F3, MRSA outperformed all the comparison algorithms, and although MRSA could not stably obtain the optimal solution, it performed the best among all the comparison algorithms, indicating that MRSA has a stronger exploitation ability. For the multi-peaked test functions F4–F10, MRSA performs best among the four test functions (F4, F5, F8, and F10), while PFA achieves the best results on F6, F7, and F9, with MRSA ranking second in all cases. The performance of MRSA on

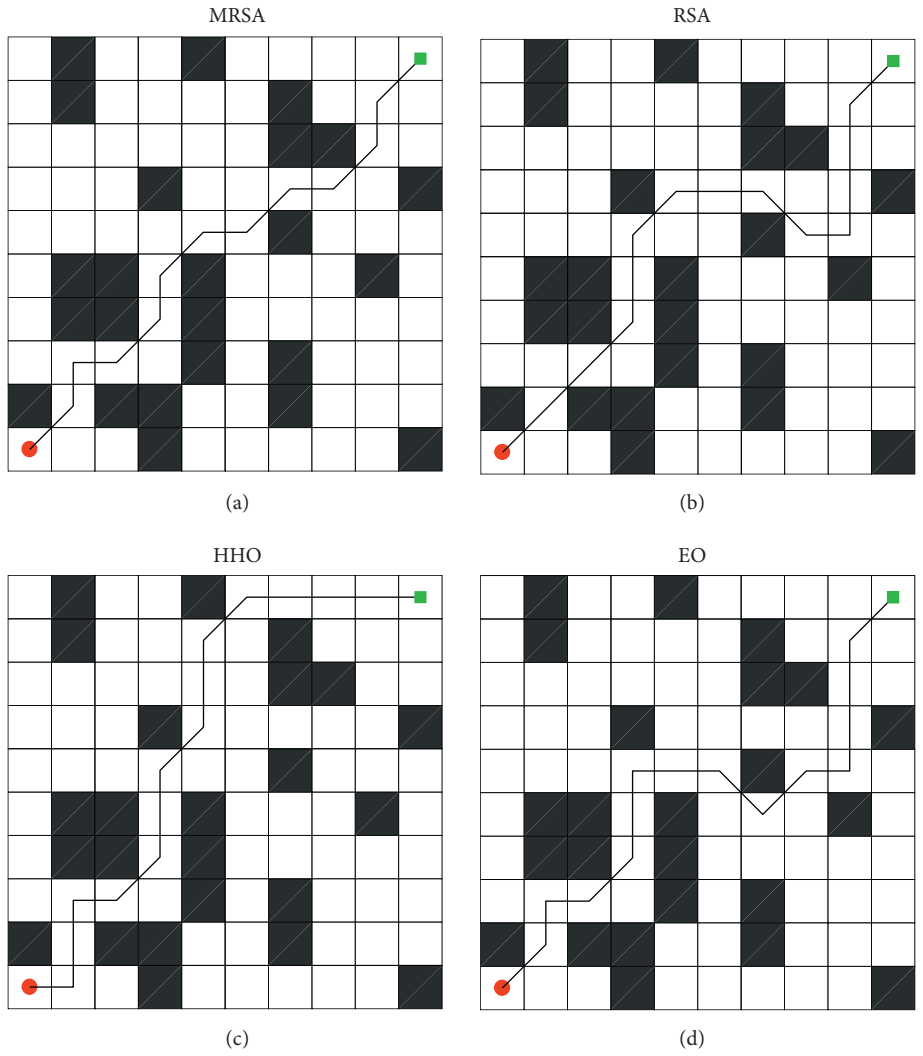


Figure 4: Path planning diagram: (a) MRSA. (b) RSA. (c) HHO. (d) EO.

the multipeaked functions indicates that the improved algorithm can maintain sufficient population diversity to avoid falling into local optima. For complex and combinatorial functions, each algorithm has its advantages and disadvantages. MRSA obtains optimal solutions on F11–F15, F18-F19, F22, F25, and F28–F30. PFA achieves better solutions on F16-F17, F20, F23-F24, and F26-F27. BOA outperforms the other comparison algorithms on F21. Overall, MRSA achieves the top two results in both the complex and combinatorial functions, better demonstrating the potential of MRSA to solve complex optimization problems in the real world.

To perform a statistical analysis on the performance of MRSA and the six competing algorithms, the Friedman test and Wilcoxon's signed-rank test were used for multiple comparisons in this paper. Table 15 is the Friedman test results showing the average ranking of each algorithm. The overall ranking value of MRSA is 1.3929, which ranks first among all the algorithms. The remaining six algorithms are ranked as follows: PFA, HHO, TDO, AOA, BOA, and SSA. The results of Wilcoxon's signed-rank test are shown in Table 16. MRSA outperformed BOA, AOA, SSA, PFA, and TDO at the 0.05 significance level, which statistically proves that the improvement strategy proposed in this paper can effectively help MRSA balance the exploitation and exploration capabilities and has a better local optimal avoidance ability.

4.5. Robot Path Planning Based on MRSA. To verify the performance of the improved strategy, MRSA is applied to solve the robot path planning in this paper. Each crocodile represents a possible path. It is assumed that there are N possible paths, and the dimension D is determined by the number of connections from the starting point to the destination point. The environment is modeled using the raster method, and the raster values are used to equate to the obstacles at the location. The robot's working environment is equated to a plane, similar to a lattice effect, and then the feasible and obstacle zones are determined based on the raster values. The grid number 0 is defined as the feasible area, and 1 is defined as the obstacle area. The robot can walk

Table 17: Robot root planning results.

Index	MRSA	RSA [41]	HHO [42]	EO [43]
Best	12.7279	15.5563	12.7279	15.5563
Mean	**14.4250**	16.4049	15.2735	17.8191
Worst	15.5563	21.2132	18.3848	21.2132
Std	1.4606	1.9090	1.6055	2.2311

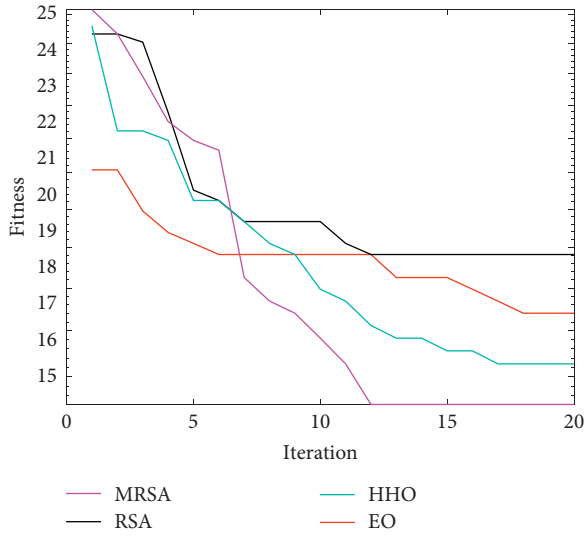


Figure 5: Convergence curves of four algorithms.

on the grid designated as 0. The cost function for the i^{th} crocodile is shown below:

$$\cos t_i = \sum_{j=1}^{D-1} \sqrt{\left(x_{j+1} + x_j\right)^2 + \left(y_{j+1} + y_j\right)^2}, \tag{14}$$

where j denotes the j^{th} dimension of each crocodile. In robot path planning, the population size is 100, and the number of iterations is 20. RSA [41], HHO [42], and EO [43] are used as competitors. Each algorithm works in a 10×10 model, and the optimal route is shown in Figure 4. To eliminate chance, each algorithm was run 10 times, and the mean, optimal, and worst values of each algorithm were recorded. The statistical results of each algorithm are shown in Table 17.

As shown in Figure 4, MRSA has the shortest route, followed by HHO, while EO and RSA are clearly trapped in a local optimum. As seen from Table 17, MRSA is the best among all the algorithms in terms of best cost, mean cost, and worst cost. This indicates that MRSA can consistently provide excellent solutions. Figure 5 shows the convergence curves of the four algorithms. MRSA has the fastest convergence speed and a higher convergence accuracy. Therefore, the introduction of multiple strategies makes the algorithm more comprehensive in its search, which greatly improves the search capability of MRSA and plans the least costly route.

5. Conclusion

This paper proposes a novel variant of the reptile search algorithm, called MRSA. First, the adaptive chaotic reverse learning strategy combines the advantages of the reverse learning mechanism and chaotic mapping to enhance the population diversity. Second, the elite alternative pool strategy balances the exploitation and exploration capabilities by controlling the reference points followed by the population. Finally, the shifted distribution estimation strategy makes full use of the dominant population information to guide the direction of individual evolution, thus improving the performance of RSA. The superiority of MRSA was verified in 23 benchmark functions, IEEE CEC2017 benchmark functions, and robot path planning problems. The experimental results show that the adaptive chaotic reverse learning strategy can effectively improve the population diversity, among which tent mapping is the most effective. The MRSA outperforms the comparison algorithm in terms of convergence accuracy, convergence speed, and stability. The results of the multimodal functions F8–F23 among the 23 benchmark functions show that the elite alternative pool strategy balances algorithm exploitation and exploration effectively and prevents the algorithm from falling into a local optimum. The adaptive chaotic reverse learning strategy enhances the population diversity. The shifted distribution estimation strategy enhances the convergence speed and convergence accuracy of the algorithm by learning information about the dominant populations. In addition, the test results were analyzed using the Friedman test and the Wilcoxon signed-rank test. The statistical results show that MRSA is significantly more effective than the comparison algorithm.

In a subsequent study, we plan to examine the following issues: First, the shifted distribution estimation strategy increases the computational cost of MRSA. Optimizing the algorithm structure and performance needs further investigation and discussion. Second, the capacity and composition of the elite replacement pool need to be further analyzed. Additionally, MRSA can be extended to multiobjective and binary versions. We will consider solving problems in image processing, industry, neural networks, text, and data mining as real-world optimization problems.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the Reserve Leader Funding Project of Leading Talent Echelon of Heilongjiang Province of China under Grant 2501050628 and in part by the Science and Technology Innovation Team Project of Foshan City of China under Grant 2018IT100302.

References

[1] H. Faris, A. M. Al-Zoubi, A. A. Heidari et al., "An intelligent system for spam detection and identification of the most relevant features based on evolutionary random weight networks," *Information Fusion*, vol. 48, pp. 67–83, 2019.

[2] R. Abbassi, A. Abbassi, A. A. Heidari, and S. Mirjalili, "An efficient salp swarm-inspired algorithm for parameters identification of photovoltaic cell models," *Energy Conversion and Management*, vol. 179, pp. 362–372, 2019.

[3] Y. Zhou, H. X. Wang, and H. Liu, "Generalized function projective synchronization of incommensurate fractional-order chaotic systems with inputs saturation," *International Journal of Fuzzy Systems*, vol. 21, no. 3, pp. 823–836, 2019.

[4] Y. Zhou, H. Liu, J. D. Cao, and S. G. Li, "Composite learning fuzzy synchronization for incommensurate fractional-order chaotic systems with time-varying delays," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 12, pp. 1739–1758, 2019.

[5] B. Jin, L. Cruz, and N. Goncalves, "Deep facial diagnosis: deep transfer learning from face recognition to facial diagnosis," *IEEE Access*, vol. 8, pp. 123649–123661, 2020.

[6] B. Jin, L. Cruz, and N. Goncalves, "Face depth prediction by the scene depth," in *Proceedings of the 2021 IEEE/ACIS 19th International Conference on Computer and Information Science (ICIS)*, Shanghai, China, June 2021.

[7] X. Y. Zhou, J. X. Lu, J. H. Huang, M. S. Zhong, and M. W. Wang, "Enhancing artificial bee colony algorithm with multi-elite guidance," *Information Sciences*, vol. 543, pp. 242–258, 2021.

[8] P. C. Wang, Y. Q. Zhou, Q. F. Luo, C. Han, Y. B. Niu, and M. Y. Lei, "Complex-valued encoding metaheuristic optimization algorithm: a comprehensive survey," *Neurocomputing*, vol. 407, pp. 313–342, 2020.

[9] M. Abd Elaziz, D. Yousri, M. A. A. Al-qaness, A. M. AbdelAty, A. G. Radwan, and A. A. Ewees, "A grunwald–letnikov based manta ray foraging optimizer for global optimization and image segmentation," *Engineering Applications of Artificial Intelligence*, vol. 98, 2021.

[10] M. Alweshah, "Solving feature selection problems by combining mutation and crossover operations with the monarch butterfly optimization algorithm," *Applied Intelligence*, vol. 51, no. 6, pp. 4058–4081, 2021.

[11] M. Alweshah, S. A. Khalaileh, B. B. Gupta, A. Almomani, A. I. Hammouri, and M. A. Al-Betar, "The monarch butterfly optimization algorithm for solving feature selection problems," *Neural Computing & Applications*, vol. 34, no. 14, pp. 11267–11281, 2020.

[12] X. K. Lin and Y. H. Wu, "Parameters identification of photovoltaic models using niche-based particle swarm optimization in parallel computing architecture," *Energy*, vol. 196, Article ID 117054, 2020.

[13] A. D. Tang, T. Han, H. Zhou, and L. Xie, "An improved equilibrium optimizer with application in unmanned aerial vehicle path planning," *Sensors*, vol. 21, no. 5, p. 1814, 2021.

[14] Z. Xu, H. C. Yang, J. Y. Li, X. Y. Zhang, B. Lu, and S. C. Gao, "Comparative study on single and multiple chaotic maps incorporated grey wolf optimization algorithms," *IEEE Access*, vol. 9, pp. 77416–77437, 2021.

[15] C. Charin, D. Ishak, M. A. A. Mohd Zainuri, B. Ismail, and M. K. Mohd Jamil, "A hybrid of bio-inspired algorithm based on levy flight and particle swarm optimizations for photovoltaic system under partial shading conditions," *Solar Energy*, vol. 217, pp. 1–14, 2021.

[16] Y. Yu, S. C. Gao, S. Cheng, Y. R. Wang, S. Y. Song, and F. G. Yuan, "CBSO: a memetic brain storm optimization with chaotic local search," *Memetic Computing*, vol. 10, no. 4, pp. 353–367, 2018.

[17] B. Morales-Castaneda, D. Zaldivar, E. Cuevas, F. Fausto, and A. Rodriguez, "A better balance in metaheuristic algorithms: does it exist," *Swarm and Evolutionary Computation*, vol. 54, 2020.

[18] Z. H. Cai, S. C. Gao, X. Yang, G. Yang, S. Cheng, and Y. H. Shi, "Alternate search pattern-based brain storm optimization," *Knowledge-Based Systems*, vol. 238, Article ID 107896.

[19] A. Aleti and I. Moser, "A systematic literature review of adaptive parameter control methods for evolutionary algorithms," *ACM Computing Surveys*, vol. 49, no. 3, pp. 1–35, 2017.

[20] Z. Y. Lei, S. C. Gao, S. Gupta, J. J. Cheng, and G. Yang, "An aggregative learning gravitational search algorithm with self-adaptive gravitational constants," *Expert Systems with Applications*, vol. 152, Article ID 113396, 2020.

[21] X. X. Zhong and P. Cheng, "An elite-guided hierarchical differential evolution algorithm," *Applied Intelligence*, vol. 51, no. 7, pp. 4962–4983, 2021.

[22] Y. R. Wang, S. C. Gao, M. C. Zhou, and Y. Yu, "A multi-layered gravitational search algorithm for function optimization and real-world problems," *IEEE-CAA Journal of Automatica Sinica*, vol. 8, no. 1, pp. 94–109, 2021.

[23] S. Gao, Y. R. Wang, J. H. Wang, and J. J. Cheng, "Understanding differential evolution: a Poisson law derived from population interaction network," *Journal of Computational Science*, vol. 21, pp. 140–149, 2017.

[24] L. Xie, T. Han, H. Zhou, Z. R. Zhang, B. Han, and A. D. Tang, "Tuna swarm optimization: a novel swarm-based metaheuristic algorithm for global optimization," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 9210050, 22 pages, 2021.

[25] A. Yaghoubzadeh-Bavandpour, O. Bozorg-Haddad, M. Rajabi, B. Zolghadr-Asli, and X. F. Chu, "Application of swarm intelligence and evolutionary computation algorithms for optimal reservoir operation," *Water Resources Management*, vol. 36, no. 7, pp. 2275–2292, 2022.

[26] X. W. Xia, L. Gui, Y. L. Zhang et al., "A fitness-based adaptive differential evolution algorithm," *Information Sciences*, vol. 549, pp. 116–141, 2021.

[27] S. Kumar and A. Sikander, "Optimum mobile robot path planning using improved artificial bee colony algorithm and evolutionary programming," *Arabian Journal for Science and Engineering*, vol. 47, no. 3, pp. 3519–3539, 2022.

[28] M. M. Rosso, R. Cucuzza, A. Aloisio, and G. C. Marano, "Enhanced multi-strategy particle swarm optimization for constrained problems with an evolutionary-strategies-based unfeasible local search operator," *Applied Sciences*, vol. 12, no. 5, p. 2285, 2022.

[29] K. P. Wang, X. Y. Li, L. Gao, P. G. Li, and S. M. Gupta, "A genetic simulated annealing algorithm for parallel partial disassembly line balancing problem," *Applied Soft Computing*, vol. 107, Article ID 107404.

[30] D. Pelusi, R. Mascella, L. Tallini, J. Nayak, B. Naik, and Y. Deng, "Improving exploration and exploitation via a hyperbolic gravitational search algorithm," *Knowledge-Based Systems*, vol. 193, Article ID 105404, 2020.

[31] Z. L. Wei, C. Q. Huang, X. F. Wang, T. Han, and Y. T. Li, "Nuclear reaction optimization: a novel and powerful physics-based algorithm for global optimization," *IEEE Access*, vol. 7, pp. 66084–66109, 2019.

[32] S. Mirjalili, "SCA: a sine cosine algorithm for solving optimization problems," *Knowledge-Based Systems*, vol. 96, pp. 120–133, 2016.

[33] H. Eskandar, A. Sadollah, A. Bahreininejad, and M. Hamdi, "Water cycle algorithm-A novel metaheuristic optimization method for solving constrained engineering optimization problems," *Computers & Structures*, vol. 110-111, pp. 151–166, 2012.

[34] W. Li, X. Meng, Y. Huang, and Z. H. Fu, "Multipopulation cooperative particle swarm optimization with a mixed mutation strategy," *Information Sciences*, vol. 529, pp. 179–196, 2020.

[35] Q. Luo, H. B. Wang, Y. Zheng, and J. C. He, "Research on path planning of mobile robot based on improved ant colony algorithm," *Neural Computing & Applications*, vol. 32, no. 6, pp. 1555–1566, 2020.

[36] M. H. Nadimi-Shahraki, S. Taghian, and S. Mirjalili, "An improved grey wolf optimizer for solving engineering problems," *Expert Systems with Applications*, vol. 166, 2021.

[37] S. Mirjalili and A. Lewis, "The whale optimization algorithm," *Advances in Engineering Software*, vol. 95, pp. 51–67, 2016.

[38] S. Arora and S. Singh, "Butterfly optimization algorithm: a novel approach for global optimization," *Soft Computing*, vol. 23, no. 3, pp. 715–734, 2019.

[39] A. M. Altabeeb, A. M. Mohsen, L. Abualigah, and A. Ghallab, "Solving capacitated vehicle routing problem using cooperative firefly algorithm," *Applied Soft Computing*, vol. 108, Article ID 107403.

[40] H. Wang, W. J. Wang, X. Y. Zhou et al., "Artificial bee colony algorithm based on knowledge fusion," *Complex & Intelligent Systems*, vol. 7, no. 3, pp. 1139–1152, 2021.

[41] L. Abualigah, M. A. Elaziz, P. Sumari, Z. W. Geem, and A. H. Gandomi, "Reptile search algorithm (RSA): a nature-inspired meta-heuristic optimizer," *Expert Systems with Applications*, vol. 191, Article ID 116158, 2022.

[42] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. L. Chen, "Harris hawks optimization: algorithm and applications," *Future Generation Computer Systems*, vol. 97, pp. 849–872, 2019.

[43] A. Faramarzi, M. Heidarinejad, B. Stephens, and S. Mirjalili, "Equilibrium optimizer: a novel optimization algorithm," *Knowledge-Based Systems*, vol. 191, Article ID 105190, 2020.

[44] S. Kaur, L. K. Awasthi, A. L. Sangal, and G. Dhiman, "Tunicate swarm algorithm: a new bio-inspired based metaheuristic paradigm for global optimization," *Engineering Applications of Artificial Intelligence*, vol. 90, 2020.

[45] M. Tubishat, S. Ja'afar, M. Alswaitti et al., "Dynamic salp swarm algorithm for feature selection," *Expert Systems with Applications*, vol. 164, Article ID 113873.

[46] M. Dehghani, S. Hubalovsky, and P. Trojovsky, "Tasmanian devil optimization: a new bio-inspired optimization algorithm for solving optimization algorithm," *IEEE Access*, vol. 10, pp. 19599–19620, 2022.

[47] L. Abualigah, A. Diabat, S. Mirjalili, M. Abd Elaziz, and A. H. Gandomi, "The arithmetic optimization algorithm," *Computer Methods in Applied Mechanics and Engineering*, vol. 376, Article ID 113609.

[48] H. Yapici and N. Cetinkaya, "A new meta-heuristic optimizer: pathfinder algorithm," *Applied Soft Computing*, vol. 78, pp. 545–568, 2019.

[49] A. D. Tang, S. Q. Tang, T. Han, H. Zhou, and L. Xie, "A modified slime mould algorithm for global optimization," *Computational Intelligence and Neuroscience*, vol. 2021, 22 pages, 2021.

[50] T. Y. Zheng and W. L. Luo, "An enhanced lightning attachment procedure optimization with quasi-opposition-based learning and dimensional search strategies," *Computational Intelligence and Neuroscience*, vol. 2019, pp. 1–24, Article ID 1589303, 2019.

[51] S. Wang, Q. X. Liu, Y. X. Liu et al., "A hybrid ssa and sma with mutation opposition-based learning for constrained engineering problems," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–21, Article ID 6379469, 2021.

[52] L. Yang, S. C. Gao, H. C. Yang, Z. H. Cai, Z. Y. Lei, and Y. Todo, "Adaptive chaotic spherical evolution algorithm," *Memetic Computing*, vol. 13, no. 3, pp. 383–411, 2021.

[53] A. D. Tang, H. Zhou, T. Han, and L. Xie, "A modified manta ray foraging optimization for global optimization problems," *IEEE Access*, vol. 9, pp. 128702–128721, 2021.

[54] X. F. Wang, H. Zhao, T. Han, H. Zhou, and C. Li, "A grey wolf optimizer using Gaussian estimation of distribution and its application in the multi-UAV multi-target urban tracking problem," *Applied Soft Computing*, vol. 78, pp. 240–260, 2019.

[55] X. X. Han, Q. X. Xu, L. Yue, Y. C. Dong, G. Xie, and X. Y. Xu, "An improved crow search algorithm based on spiral search mechanism for solving numerical and engineering optimization problems," *IEEE Access*, vol. 8, pp. 1–92382, 2020.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 1779582, 16 pages
https://doi.org/10.1155/2022/1779582

Research Article

Bifurcations of a Fractional-Order Four-Neuron Recurrent Neural Network with Multiple Delays

Yu Fei,[1] Rongli Li,[1] Xiaofang Meng,[1] and Zhouhong Li[1,2]

[1]*School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming, Yunnan 650221, China*
[2]*Department of Mathematics, Yuxi Normal University, Yuxi, Yunnan 653100, China*

Correspondence should be addressed to Zhouhong Li; zhouhli@yeah.net

Received 31 May 2022; Accepted 22 August 2022; Published 29 September 2022

Academic Editor: Heng Liu

This paper investigates the bifurcation issue of fractional-order four-neuron recurrent neural network with multiple delays. First, the stability and Hopf bifurcation of the system are studied by analyzing the associated characteristic equations. It is shown that the dynamics of delayed fractional-order neural networks not only depend heavily on the communication delay but also significantly affects the applications with different delays. Second, we numerically demonstrate the effect of the order on the Hopf bifurcation. Two numerical examples illustrate the validity of the theoretical results at the end.

1. Introduction

Recurrent neural network (RNN) is a type of recursive neural network that takes sequence data as input, recurses in the evolution direction of the sequence, and all nodes (recurrent units) are connected in a chain. Till now, several recurrent neural networks (RNNs) have been widely considered in various fields such as signal processing, optimizations control, image processing, robotics, pattern recognitions, and automatic control, so they have attracted extensive attention of researchers in recent years [1–7]. Since the applications of RNNs depend more heavily on dynamical neural networks, quite a few efforts have been undertaken to study their dynamical properties and a large number of useful results have been investigated, including oscillation, stability, bifurcation, synchronization, and chaos of various RNNs [8–14].

As the matter of fact, for some applications of nonlinear dynamical models, time delay has a significant impact, and in addition to affecting stability, it causes oscillations and other unstable phenomena, such as chaos [15]. Communication delays and the response times of neurons are considered key factors in the performance of neural networks, and this is caused by the finite switching speed of amplifiers and the noninstantaneous signal transmission between neurons [16]. In recently years, many scholars have been interested in studying the dynamics of neural networks with such time delays [17–19]. It must be pointed out that exponential stabilization of memristor-based RNNs with disturbance and mixed time delays by periodically intermittent control has been considered by Wang et al. [20]. Using the appropriate Lyapunov–Krasovski functionals and applying matrix inequality approach methods, Zhou [21] discussed the passivity of a class of recurrent neural networks with impulse and multiproportional delays. Zhou and Zhao [22] investigated the exponential synchronization and polynomial synchronization of recurrent neural networks with and without proportional delays. Robust stability analysis of recurrent neural networks is studied in Refs. [23, 24]. Furthermore, time delays are ubiquitous and unavoidable in the real world. Due to the existence of delays, the system can become unstable, and the dynamic behavior of nonlinear systems becomes more difficult. Moreover, since the solution space of the delay dynamical is infinite, it makes the systems more complex and bifurcation occurs. Hence, it is necessary to consider the properties and dynamics of neural networks via delays, such as time delay [25, 26], multiple delays [27, 28] time-varying delays [29, 30], and so on. In 2013, Zhang and Yang [31] studied a four-neuron recurrent neural network with multiple delays, described as follows:

$$\begin{cases} \cdot x_1(t) = -x_1(t) + f(x_2(t-\tau_1)), \\ \cdot x_2(t) = -x_2(t) + f(x_3(t-\tau_1)), \\ \cdot x_3(t) = -x_3(t) + f(x_4(t-\tau_1)), \\ \cdot x_4(t) = -x_4(t) + \omega_1 f(x_1(t-\tau_2)) + \omega_2 f(x_2(t-\tau_2)) + \omega_3 f(x_3(t-\tau_2)), \end{cases} \tag{1}$$

C5 $(\alpha_1\beta_1 + \alpha_2\beta_2)/(\alpha_1^2 + \alpha_2^2) \neq 0$where $x_i(t)(i = 1, 2, 3, 4)$ stand for state of the ith neuron at time t, $\omega_k \in R(k = 1, 2, 3)$ are the network parameters or weight, $f(\cdot)$ is the connection function between neurons, and $\tau_j \geq 0(j = 1, 2)$ are the communication time delay. By using the distribution of the solutions of the associated characteristic equation, the Hopf bifurcation and local stability of the four-dimensional RNNs with two delays are studied. For more recurrent neural network research results, see references [5, 10, 12, 20].

In more than three centuries, fractional calculus has developed into a classical mathematical concept. Nonlinear dynamics systems have shown that it has an exceptionally important role in generalizing ordinary differentiation and integration to arbitrary noninteger order. Therefore, if we study the effects of the memory and genetics factors, fractional neural network is sometimes more realistic and more general than integer neural networks. In recent years, the application of fractional order neural networks has developed rapidly, and the complex dynamical behaviors of fractional neural networks has become a very important research hot points, such as stability or multistability, Hopf bifurcation, synchronization, chaos, and so on. For instance, in Ref. [32], the multistability of a fractional-order competitive neural networks with delay is investigated by using the fractional calculus and partitioning of state space. In Lu and Xue [33] study, adaptive synchronization is investigated for fractional delayed stochastic neural networks. Yuan and Huang [34] considered the quantitative analysis of fractional-order neural networks with time delay. Udhayakumar and Rajan [35] discussed Hopf bifurcation of a delayed fractional-order octonion-valued neural networks.

We also know that Hopf bifurcations, which include subcritical and supercritical ones, can be used to efficiently design biochemical oscillators. Furthermore, fractional order neural networks with same delay cannot accurately describe the dynamical properties of real world neural networks compared with the ones with different delays. In recent years, some researchers have considered the dynamical behavior of fractional models with time delay [36–48]. In 2019 [49], we also investigated the existence of Hopf bifurcation for four-neuron fractional neural networks with leakage delays. To the best of our knowledge, so far there are few results on the Hopf bifurcation of four-dimensional fractional-order recurrent neural network with multiple delays are reported, and therefore, the study of Hopf bifurcation of fractional-order dynamical systems with multiple delays remains an open problem.

Based on the above motivations, we are dedicated to presenting a theoretical exploration of stability and Hopf bifurcation for a four-neuron fractional-order recurrent neural network with multiple delays in this work. The main contributions can be highlighted as follows:

(i) A novel delayed fractional-order recurrent neural network with four-neuron and two different delays is studied

(ii) Double main dynamical properties of the fractional-order recurrent neural network with two delays are investigated: stability and oscillation

(iii) The Hopf bifurcation is discussed in terms of delays and order

In the article, we shall give some some lemmas and definitions of fractional-order calculus in Section 2, and models description in Section 3. In Section 4, the local stability of the trivial steady state of delayed fractional-order RNNs is examined by applying the associated characteristic equation. In addition, the authors will care about the Hopf bifurcation of fractional-order RNNs with multiple delays. In Section 5, two numerical examples are provided to demonstrate the theoretical results. The last section gives some conclusions.

2. Preliminaries

This section we will give some Caputo definitions and lemma for fractional calculus as a basis for the theoretical analysis and simulation proofs.

Definition 1 (see [50]). The fractional integral of order ϕ for a function f(x) is defined as follows:

$$I^{\phi} f(x) = \frac{1}{\Gamma(\phi)} \int_{x_0}^{x} (x-s)^{\phi-1} f(s) ds, \tag{2}$$

where $\phi > 0$, and $\Gamma(\cdot)$ is the Gamma function satisfying $\Gamma(s) = \int_0^{\infty} x^{s-1} e^{-x} dx$.

Definition 2 (see [50]). Caputo fractional derivative of order ϕ for a function $\psi(x) \in C^k[x_0, \infty), R)$ is defined by

$$D^{\phi} \psi(x) = \frac{1}{\Gamma(n-\phi)} \int_{x_0}^{x} \frac{\psi^{(k)}(s)}{(x-s)^{\phi-k+1}} ds, \tag{3}$$

where $x \geq x_0$ and $k - 1 \leq \phi < k$, $k \in N^+$.

Moreover, when $\phi \in (0, 1)$, then

$$D^{\phi} \psi(x) = \frac{1}{\Gamma(1-\phi)} \int_{x_0}^{x} \frac{\psi'(s)}{(x-s)^{\phi}} ds. \tag{4}$$

Lemma 1 (see [51]). *Consider the following fractional order autonomous model.*

$$D^{\phi}u = Ju, u(0) = u_0, \tag{5}$$

in which $0 < \phi \leq 1$, $u \in R^k$, *and* $J \in R^{k\times k}$. *Then the zero solution of the system (5) is asymptotically stable in the Lyapunov sense if all roots* λ_i *are the system (5) of character equation satisfy* $|\arg(\lambda_i)| > \phi\pi/2\,(i = 1, 2, \ldots, k)$, *and then each component of the states decays towards 0 like* $t^{-\phi}$. *In addition, this model is stable if and only if* $|\arg(\lambda_i)| \geq \phi\pi/2$ *and those critical eigenvalues that satisfy* $|\arg(\lambda_i)| = \phi\pi/2$ *have geometric multiplicity one.*

3. Mathematics Model Elaboration

This article considers the following four-neuron fractional-order recurrent neural network with two delays:

$$\begin{cases} D^{\phi}x_1(t) = -x_1(t) + f(x_2(t-\tau_1)), \\ D^{\phi}x_2(t) = -x_2(t) + f(x_3(t-\tau_1)), \\ D^{\phi}x_3(t) = -x_3(t) + f(x_4(t-\tau_1)), \\ D^{\phi}x_4(t) = -x_4(t) + \omega_1 f(x_1(t-\tau_2)) + \omega_2 f(x_2(t-\tau_2)) + \omega_3 f(x_3(t-\tau_2)), \end{cases} \tag{6}$$

where $\phi \in (0, 1]$ are fractional order; $x_i(t)\,(i = 1, 2, 3, 4)$ stand for state variables; $\omega_i\,(i = 1, 2, 3)$ denote the connection weights; the function of connecting neurons is denoted by $f(x(\cdot))$; and τ_1 and τ_2 are the communication time delays.

Remark 1. In fact, if $\phi = 1$, the fractional delayed neural networks (6) changes into the general neural network (1).

Accordingly, the main purpose of this article is to investigate the stability and the application of Hopf bifurcations of the neural networks (6) taking different time delays τ_1 and τ_2 as the bifurcation parameters by the method of stability analysis [52]. In addition, the effects of the order on the creation of the Hopf bifurcation for the proposed fractional order neural network with multiple delays are also numerically discussed.

Throughout of this paper, assume that the following condition holds true:

(**C1**) $f(\cdot) \in C(R, R)$, $f(0) = 0$, $xf(x) > 0$, for $x \neq 0$.

4. Main Results

This section chooses τ_1 or τ_2 as a bifurcation parameter to study the stability analysis and Hopf bifurcation for the fractional order RNNs (6) and to study the bifurcation points accurately.

4.1. Bifurcation Depending on τ_1 *in Equation (6).* In this subsection, we first study the effects of τ_1 on bifurcations of system (6) by establishing τ_2.

Applying Taylor series formula, the following form of equation (6) at the origin is

$$\begin{cases} D^{\phi}x_1(t) = -x_1(t) + m_1x_2(t-\tau_1), \\ D^{\phi}x_2(t) = -x_2(t) + m_2x_3(t-\tau_1), \\ D^{\phi}x_3(t) = -x_3(t) + m_3x_4(t-\tau_1), \\ D^{\phi}x_4(t) = -x_4(t) + m_4x_1(t-\tau_2) + m_5x_2(t-\tau_2) + m_6x_3(t-\tau_2). \end{cases} \tag{7}$$

By applying Laplace transformation, its characteristic equation is given as

$$\det\begin{pmatrix} s^{\phi}+1 & -m_1e^{s\tau_1} & 0 & 0 \\ 0 & s^{\phi}+1 & -m_2e^{-s\tau_1} & 0 \\ 0 & 0 & s^{\phi}+1 & -m_3e^{-s\tau_1} \\ -m_4e^{-s\tau_2} & -m_5e^{-s\tau_2} & -m_6e^{-s\tau_2} & s^{\phi}+1 \end{pmatrix} = 0, \tag{8}$$

where $m_k = f'(0)\,(k = 1, 2, 3), m_k = \omega_j f'(0)\,(j = 1, 2, 3, k = 4, 5, 6)$.

From (8), we have

$$K_1(s) + K_2(s)e^{-s\tau_1} + K_3(s)e^{-2s\tau_1} + K_4(s)e^{-3s\tau_1} = 0, \tag{9}$$

where

$$\begin{aligned} K_1(s) &= s^{4\phi} + 4s^{3\phi} + 6s^{2\phi} + 4s^{\phi} + 1, \\ K_2(s) &= -m_3m_6\left(s^{2\phi} + 2s^{\phi} + 1\right)e^{-s\tau_2}, \\ K_3(s) &= -m_2m_3m_5\left(s^{\phi} + 1\right)e^{-s\tau_2}, \\ K_4(s) &= -m_1m_2m_3m_4e^{-s\tau_2}. \end{aligned} \tag{10}$$

Multiplying $e^{s\tau_1}$ and $e^{2s\tau_1}$ on both sides of equation (9), respectively, we can obtain

$$\begin{cases} K_1(s)e^{2s\tau_1} & +K_2(s)e^{s\tau_1} + K_3(s) + K_4(s)e^{-s\tau_1} = 0, \\ K_1(s)e^{s\tau_1} & +K_2(s) + K_3(s)e^{-s\tau_1} + K_4(s)e^{-2s\tau_1} = 0. \end{cases} \tag{11}$$

Let $K_1(s) = A_1 + iB_1$, $K_2(s) = A_2 + iB_2$, $K_3(s) = A_3 + iB_3$, $K_4(s) = A_4 + iB_4$, and from equation (9), we have

$$\begin{cases} (A_1 + iB_1)e^{2s\tau_1} + (A_2 + iB_2)e^{s\tau_1} + (A_3 + iB_3) + (A_4 + iB_4)e^{-s\tau_1} = 0, \\ (A_1 + iB_1)e^{s\tau_1} + (A_2 + iB_2) + (A_3 + iB_3)e^{-s\tau_1} + (A_4 + iB_4)e^{-2s\tau_1} = 0. \end{cases} \tag{12}$$

Take $s = iw = w(\cos \pi/2 + i \sin \pi/2)(\omega > 0)$ be a purely imaginary root of equation (11). Apply inserting s into equation (11) and separating the imaginary and real parts yields the following equations:

$$\begin{cases} A_1 \cos(2\omega\tau_1) - B_1 \sin(2\omega\tau_1) + (A_2 + A_4)\cos(\omega\tau_1) + (B_4 - B_2)\sin(\omega\tau_1) = -A_3, \\ B_1 \cos(2\omega\tau_1) + A_1 \sin(2\omega\tau_1) + (B_2 + B_4)\cos(\omega\tau_1) + (A_2 - A_4)\sin(\omega\tau_1) = -B_3, \\ A_1 \cos(2\omega\tau_1) - B_1 \sin(2\omega\tau_1) + (A_2 + A_4)\cos(\omega\tau_1) + (B_4 - B_2)\sin(\omega\tau_1) = -A_3, \\ B_1 \cos(2\omega\tau_1) + A_1 \sin(2\omega\tau_1) + (B_2 + B_4)\cos(\omega\tau_1) + (A_2 - A_4)\sin(\omega\tau_1) = -B_3. \end{cases} \tag{13}$$

Evidently,

$$\begin{cases} \cos(\omega\tau_1) = \dfrac{F_{12}(\omega)}{F_{11}(\omega)} = F_{c1}(\omega), \\ \sin(\omega\tau_1) = \dfrac{F_{22}(\omega)}{F_{21}(\omega)} = F_{s1}(\omega), \end{cases} \tag{14}$$

where $A_1, A_2, A_3, A_4, B_1, B_2, B_3, B_4, F_{11}, F_{12}, F_{21,}$ and F_{22} are given Appendix A. Obviously, from fist to second equation of system (14), it can be implied that

$$F_{c1}^2(\omega) + F_{s1}^2(\omega) = 1. \tag{15}$$

From equation (13), one can obtain

$$\tau_1^{(l)} = \frac{1}{w}\left[\arccos\frac{F_{12}(w)}{F_{11}(w)} + 2l\pi\right], l = 0, 1, 2, \ldots. \tag{16}$$

Remark 2. This is an inhomogeneous system of linear equations (13), and the independent variables are $\cos(2\omega\tau_1)$, $\sin(2\omega\tau_1)$, $\cos(\omega\tau_1)$, $\sin(\omega\tau_1)$, respectively. According Cramer's rule of linear equation, if the coefficient determinant of the system of linear equations is not equal to 0, we can easily solve solutions of linear equations (13). That is, we can obtain $\cos(\omega\tau_1)$ and $\sin(\omega\tau_1)$ or $\cos(\omega 2\tau_1)$ and $\sin(2\omega\tau_1)$.

Define the bifurcation point of fractional neural network with multiple delays (6) as

$$\tau_{10}^* = \min\left\{\tau_1^{(l)}\right\}, l = 0, 1, 2, \ldots. \tag{17}$$

If τ_1 vanishes, then equation (9) becomes

$$H_1(s) + H_2(s)e^{-s\tau_2} = 0, \tag{18}$$

where

$$\begin{aligned} H_1(s) &= s^{4\phi} + 4s^{3\phi} + 6s^{2\phi} + 4s^{\phi} + 1, \\ H_2(s) &= -m_3m_6s^{2\phi} - 2m_3m_6s^{\phi} - m_3m_6 - m_2m_3m_5s^{\phi} - m_2m_3m_5 - m_1m_2m_3m_4. \end{aligned} \tag{19}$$

If $\tau_2 = 0$, then the equation (18) becomes

$$\begin{aligned} 0 = {} & s^{4\phi} + 4s^{3\phi} + 6s^{2\phi} + 4s^{\phi} + 1 - m_3m_6s^{2\phi} - 2m_3m_6s^{\phi} - m_3m_6 \\ & - m_2m_3m_5s^{\phi} - m_2m_3m_5 - m_1m_2m_3m_4. \end{aligned} \tag{20}$$

Suppose that all roots s of the equation (18) obey Lemma 1, then we get that both roots λ_i in equation (18) have negative real parts.

The imaginary and real parts of $H_j(s)(j = 1, 2)$ can be denoted by H_j^I and H_j^R, respectively. Multiplying $e^{2s\tau_2}$ on both sides of equation (18), we can obtain

$$H_1(s)e^{s\tau_2} + H_2(s) = 0. \tag{21}$$

Also, let $s = iv = v(\cos \pi/2 + i\, sin\pi/2)(v > 0)$ be a purely imaginary root of equation (11) if and only if

$$\begin{cases} H_1^R \cos(v\tau_2) - H_1^I \sin(v\tau_2) = -H_2^R, \\ H_1^I \cos(v\tau_2) + H_1^R \sin(v\tau_2) = -H_2^I. \end{cases} \tag{22}$$

This leads to form

$$\begin{cases} \cos(\nu\tau_2) = -\dfrac{H_2^R H_2^R + H_1^I H_2^I}{H_1^R 2 + H_1^I 2} = f_{c1}(\nu), \\ \sin(\nu\tau_1) = -\dfrac{-H_2^R H_1^I + H_1^R H_2^I}{H_1^R 2 + H_1^R 2} = f_{s1}(\nu). \end{cases} \tag{23}$$

It is not difficult to see that

$$f_{c1}^2(w) + f_{s1}^2(w) = 1. \tag{24}$$

Additionally, we will give the following assumptions which hold true.

(**C2**) The equation (24) has at least a positive real root.

From equation (24), the values ofν can be obtained according to Mathematics software Mathematica 10.0, and then the Hopf bifurcation point τ_{20} of fractional order recurrent neural network (6) with $\tau_1 = 0$ can be derived.To demonstrate our main results, we further present the following hypothesis: (**C3**)$\Upsilon_1\Omega_1 + \Upsilon_2\Omega_2/\Omega_1^2 + \Omega_2^2 \neq 0$, where

$$\begin{aligned} \Upsilon_1 &= w_0\Big[A_2\sin w_0\tau_{10} - B_2\cos w_0\tau_{10} + 2\left(A_3\sin 2w_0\tau_{10} - B_2\cos 2w_0\tau_{10}\right) \\ &\quad + 3\left(A_4\cos 3w_0\tau_{10} + B_4\sin 3w_0\tau_{10}\right)\Big], \\ \Upsilon_2 &= w_0\Big[A_2\cos w_0\tau_{10} + B_2\sin w_0\tau_{10} + 2\left(A_3\cos 2w_0\tau_{10} + B_2\sin 2w_0\tau_{10}\right) \\ &\quad + 3\left(A_4\cos 3w_0\tau_{10} + B_4\sin 3w_0\tau_{10}\right)\Big], \\ \Omega_1 &= A_1' + \left[A_2' - \tau_1 A_2\right]\cos w_0\tau_{10} + \left[B_2' - \tau_1 B_2\right]\sin w_0\tau_{10} + \left[A_3' - 2\tau_1 A_3\right]\cos 2w_0\tau_{10} + \left[B_2' - \tau_1 B_3\right]\sin 2w_0\tau_{10} \\ &\quad + \left[A_4' - 3\tau_1 A_4\right]\cos 3w_0\tau_{10} + \left[B_4' - 3\tau_1 B_4\right]\sin 3w_0\tau_{10}, \\ \Omega_2 &= B_1' + \left[B_2' - \tau_1 B_2\right]\cos w_0\tau_{10} - \left[A_2' - \tau_1 A_2\right]\sin w_0\tau_{10} + \left[B_3' - 2\tau_1 B_3\right]\cos 2w_0\tau_{10} - \left[A_2' - \tau_1 A_3\right]\sin 2w_0\tau_{10} \\ &\quad + \left[B_4' - 3\tau_1 B_4\right]\cos 3w_0\tau_{10} - \left[A_4' - 3\tau_1 A_4\right]\sin 3w_0\tau_{10}. \end{aligned} \tag{25}$$

Lemma 2. *Let* $s(\tau_1) = \nu(\tau_{10}) + iw(\tau_1)$ *be a root of equation (9) near* $\tau_1 = \tau_{1j}$ *satisfying* $\nu(\tau_{1j}) = 0$, $w(\tau_{1j}) = w_0$, *then the following transversality condition is satisfied.*

$$\mathrm{Re}\left[\frac{ds}{d\tau_1}\right]\Big|_{(w=w_0,\tau_1=\tau_{10})} \neq 0. \tag{26}$$

Proof. With implicit function theorem, we can differentiate equation (9) with respect to τ_1, and thus we get

$$\begin{aligned} 0 &= K_1'(s)\frac{ds}{d\tau_1} + K_2'(s)e^{-s\tau_1}\frac{ds}{d\tau_1} + K_2(s)e^{-s\tau_1}\left(-\tau_1\frac{ds}{d\tau_1} - s\right) + K_3'(s)e^{-2s\tau_1}\frac{ds}{d\tau_1} \\ &\quad + K_3(s)e^{-2s\tau_1}\left(-2\tau_1\frac{ds}{d\tau_1} - 2s\right) + K_4'(s)e^{-3s\tau_1}\frac{ds}{d\tau_1} + K_4(s)e^{-3s\tau-1}\left(-3\tau\frac{ds}{d\tau_1} - 3s\right). \end{aligned} \tag{27}$$

$$\frac{ds}{d\tau_1} = \frac{\Upsilon(s)}{\Omega(s)},$$

where

$$\begin{aligned} \Upsilon(s) &= s\left[K_2(s)e^{-s\tau_1} + 2K_3(s)e^{2-s\tau_1} + 3K_4(s)e^{-3s\tau_1}\right], \\ \Omega(s) &= K_1'(s) + \left[K_2'(s) - \tau_1 K_2(s)\right]e^{-2s\tau_1} + \left[K_3'(s) - 2\tau_1 K_3(s)\right]e^{-2s\tau_1} \\ &\quad + \left[K_4'(s) - 3\tau_1 K_4(s)\right]e^{-3s\tau_1}. \end{aligned} \tag{28}$$

We further suppose that Υ_1 and Υ_2 are the real and imaginary parts of $\Upsilon(s)$, respectively, and Ω_1 and Ω_2 are the real and imaginary parts of $\Omega(s)$, respectively, then

$$\mathrm{Re}\left[\frac{ds}{d\tau}\right]\Big|_{(\tau=\tau_0^*,w=w_0^*)} = \frac{\Upsilon_1\Omega_1 + \Upsilon_2\Omega_2}{\Omega_1^2 + \Omega_2^2}. \tag{29}$$

From (**C3**), we conclude that the transversality condition holds true. This completes the proof of Lemma 2.

From the above investigation, we can obtain the following results. □

Theorem 1. *assumptions* (**C1**)–(**C3**) *hold true, then the following results can be given:*

(i) The zero equilibrium point of fractional order four-neuron recurrent neural network with multiple delays (6) is asymptotically stable when $\tau_1 \in [0, \tau_{10}^*)$.

(ii) If $\tau_1 \in [0, \tau_{10}^*)$*, then fractional order four neurons recurrent neural network with multiple delays (6) causes Hopf bifurcation at the origin when* $\tau_1 = \tau_{10}^*$*. That is, a branch of periodic solutions can bifurcate from the zero equilibrium point at* $\tau_1 = \tau_{10}^*$.

4.2. Bifurcation Depending on τ_2 in Equation (6). As in the previous subsection, next we change another delay τ_2 to the bifurcation parameter to account for the bifurcation of the model (6). It is hard to point out that equation (8) changes as follows:

$$q_1(s) + q_2(s)e^{-s\tau_2} = 0, \tag{30}$$

where

$$\begin{aligned} q_1(s) &= 1 + 4s^{2\phi} + 6s^{2\phi} + 4s^{3\phi} + s^{4\phi}, \\ q_2(s) &= -m_3m_6\left(1 + 2s^{\phi} + s^{2\phi}\right)e^{-s\tau_1} - m_2m_3m_5\left(1 + s^{\phi}\right)e^{-2s\tau_1} \\ &\quad - m_1m_2m_3m_4e^{-3s\tau_1}. \end{aligned} \tag{31}$$

Multiplying $e^{2s\tau_2}$ on both sides of equation (30), we can obtain

$$q_1(s)e^{s\tau_2} + q_2(s) = 0. \tag{32}$$

Suppose $q_1(s) = a_1 + ib_1$ and $q_2(s) = a_2 + ib_2$, and from equation (32), we have

$$(a_1 + ib_1)e^{s\tau_2} + a_2 + ib_2 = 0, \tag{33}$$

where a_1, a_2, b_1, b_2 are given in Appendix B.

Take $s = i\widetilde{w} = \widetilde{w}(\cos\pi/2 + i\sin\pi/2)(\widetilde{\omega} > 0)$ as a root of equation (33) if and only if

$$\begin{cases} a_1\cos(\widetilde{w}\tau_2) - b_1\sin(\widetilde{w}\tau_2) = -a_2, \\ b_1\cos(\widetilde{w}\tau_2) + a_1\sin(\widetilde{w}\tau_2) = -b_2, \end{cases} \tag{34}$$

that is,

$$\begin{cases} \cos\widetilde{w}\tau_2 = -\dfrac{a_1a_2 + b_1b_2}{a_1^2 + b_1^2} = \rho(\widetilde{w}), \\ \sin\widetilde{w}\tau_2 = -\dfrac{-a_2b_1 + a_1b_2}{a_1^2 + b_1^2} = \varrho(\widetilde{w}). \end{cases} \tag{35}$$

It is simple to derive the following equation.

$$\rho^2(\widetilde{w}) + \varrho^2(\widetilde{w}) = 1. \tag{36}$$

From (35), one can obtain

$$\tau_2^{(l)} = \frac{1}{\widetilde{w}}[\arccos\varrho(\widetilde{w}) + 2l\pi], l = 0, 1, 2, \ldots. \tag{37}$$

The bifurcation point is defined by $\omega_k(k = 1, 2, 3)$ (**C3**) $(\Upsilon_1\Omega_1 + \Upsilon_2\Omega_2)(\Omega_1^2 + \Omega_2^2) \neq 0$

$$\tau_{20}^* = \min\left\{\widetilde{\tau}_2^{(l)}\right\}, l = 0, 1, 2, \ldots. \tag{38}$$

C5 $(\alpha_1\beta_1 + \alpha_2\beta_2)/(\alpha_1^2 + \alpha_2^2) \neq 0$ here $\tau_2^l$is defined by equation (38)

If $\tau_2 = 0$, then the equation (32) becomes

$$M_1(s) + M_2(s)e^{-s\tau_1} + M_3(s)e^{-2s\tau_1} + M_4(s)e^{-3s\tau_1} = 0, \tag{39}$$

where

$$\begin{aligned} M_1(s) &= 1 + 4s^{\phi} + 6s^{2\phi} + 4s^{3\phi} + s^{4\phi}, \\ M_2(s) &= -m_3m_6\left(1 + 2s^{\phi} + s^{2\phi}\right) \\ M_3(s) &= -m_2m_3m_5\left(1 + s^{\phi}\right), \\ M_4(s) &= -m_1m_2m_3m_4. \end{aligned} \tag{40}$$

Assume that all roots s of equation (39) observe Lemma 1, then we get that both roots of equation (39) have negative real parts.

The imaginary and real parts of $M_i(s)(i = 1, 2, 3, 4)$ can be expressed as M_i^I an d M_i^R, respectively. Multiplying both sides of the equation (39) by $e^{2s\tau_1}$ and $e^{s\tau_1}$ yields

$$\begin{cases} M_1(s)e^{2s\tau_1} + M_2(s)e^{s\tau_1} + M_3(s) + M_4(s)e^{-s\tau_1} = 0, \\ M_1(s)e^{s\tau_1} + M_2(s) + M_3(s)e^{-s\tau_1} + M_4(s)e^{-2s\tau_1} = 0. \end{cases} \tag{41}$$

Let $s = i\widetilde{\nu} = \widetilde{\nu}(\cos\pi/2 + i\sin\pi/2)(\widetilde{\nu} > 0)$ be a solution of equation (41). Substituting s into equation (41) and separating the imaginary and real units yields the following equations:

$$\begin{cases} M_1^R\cos(2\widetilde{\nu}\tau_1) - M_1^I\sin(2\widetilde{\nu}\tau_1) + \left(M_2^R + M_4^R\right)\cos(\widetilde{\nu}\tau_1) + \left(M_4^I - P_2^I\right)\sin(\widetilde{\nu}\tau_1) = -M_3^R, \\ M_1^I\cos(2\widetilde{\nu}\tau_1) + M_1^R\sin(2\widetilde{\nu}\tau_1) + \left(M_2^I + M_4^I\right)\cos(\widetilde{\nu}\tau_1) + \left(M_2^R - P_4^R\right)\sin(\widetilde{\nu}\tau_1) = -M_3^I, \\ P_1^R\cos(2\widetilde{\nu}\tau_1) - M_1^I\sin(2\widetilde{\nu}\tau_1) + \left(M_2^R + P_4^R\right)\cos(\widetilde{\nu}\tau_1) + \left(M_4^I - P_2^I\right)\sin(\widetilde{\nu}\tau_1) = -M_3^R, \\ P_1^I\cos(2\widetilde{\nu}\tau_1) + M_1^R\sin(2\widetilde{\nu}\tau_1) + \left(M_2^I + P_4^R\right)\cos(\widetilde{\nu}\tau_1) + \left(M_2^R - P_4^R\right)\sin(\widetilde{\nu}\tau_1) = -M_3^I, \end{cases} \tag{42}$$

which lead to

$$\begin{cases} \cos\widetilde{\nu}\tau_1 = \dfrac{E_{12}(\widetilde{\nu})}{E_{11}(\widetilde{\nu})} = \mathfrak{C}(\widetilde{\nu})^2, \\ \sin\widetilde{\nu}\tau_1 = \dfrac{E_{22}(\widetilde{\nu})}{E_{21}(\widetilde{\nu})} = \mathfrak{S}(\widetilde{\nu})^2. \end{cases} \tag{43}$$

Obviously, from first and second equation of system (43), we get

$$\mathfrak{C}(\widetilde{\nu})^2 + \mathfrak{S}(\widetilde{\nu})^2 = 1. \tag{44}$$

To theoretically gain the sufficient conditions for the Hopf bifurcation, we assume that the following assumptions hold true:

(**C4**) Equation (36) has at least a positive real root.

By means of equation (36), the values of $\widetilde{\omega}$ can be obtained according to mathematical software Mathematica 10.0, and then the bifurcation point τ_{10} of recullrent fractional four-neuron neural networks (6) with $\tau_2 = 0$ can be derived.As a summary of our main results, we provide the following assumption: (**C5**)$\alpha_1\beta_1 + \alpha_2\beta_2/\alpha_1^2 + \alpha_2^2 \neq 0$, where

$$\begin{aligned} \alpha_1 &= a_1' + (a_2' - \tau_{20}a_2)\cos\widetilde{\omega}_0\tau_{20} + (b_2' - \tau_2 b_2)\sin\widetilde{\omega}_0\tau_{20}, \\ \alpha_2 &= b_1' + (b_2' - \tau_{20}b_2)\cos\widetilde{\omega}_0\tau_{20} - (a_2' - \tau_2 a_2)\sin\widetilde{\omega}_0\tau_{20}, \\ \beta_1 &= \widetilde{\omega}_0(a_2\sin\widetilde{\omega}_0\tau_{20} - b_2\cos\widetilde{\omega}_0\tau_{20}), \\ \beta_2 &= \widetilde{\omega}_0(a_2\cos\widetilde{\omega}_0\tau_{20} + b_2\sin\widetilde{\omega}_0\tau_{20}). \end{aligned} \tag{45}$$

Lemma 3. *Let* $s(\tau_2) = \eta(\tau_2) + i\widetilde{w}(\tau_2)$ *be a root of equation (9) near* $\tau_2 = \tau_{2j}$ *satisfying* $\eta(\tau_{2j}) = 0$, $w(\tau_{2j}) = \widetilde{w}_0$, *then we get the following transversality condition*

$$\mathrm{Re}\left[\frac{ds}{d\tau_2}\right]|_{(\widetilde{w}=\widetilde{w}_0,\tau_2=\tau_{20})} \neq 0. \tag{46}$$

Proof. Similar to Lemma 2, by utilizing the implicit function theorem and differentiating (9) with respect to τ_2, we get

$$\begin{aligned} 0 &= q_1'(s)\frac{ds}{d\tau_2} + q_2'(s)e^{-s\tau_2}\frac{ds}{d\tau_2} + q_2(s)e^{-s\tau_2}\left(-\tau_2\frac{ds}{d\tau_2} - s\right), \\ \frac{ds}{d\tau_2} &= \frac{\beta(s)}{\alpha(s)}, \end{aligned} \tag{47}$$

where

$$\begin{aligned} \beta(s) &= sq_2(s)e^{-s\tau_2}, \\ \alpha(s) &= q_1'(s) + q_2'(s)e^{-s\tau_2} - \tau_2 q_2(s)e^{-s\tau_2}. \end{aligned} \tag{48}$$

We further suppose that α_1 and α_2 are the real and imaginary units of $\alpha(s)$, respectively, and β_1 and β_2 are the real and imaginary parts of $\beta(s)$, respectively, then we get

$$\mathrm{Re}\left[\frac{ds}{d\tau_2}\right]|_{(\widetilde{w}=\widetilde{w}_0,\tau_2=\tau_{20})} = \frac{\alpha_1\beta_1 + \alpha_2\beta_2}{\alpha_1^2 + \alpha_2^2}. \tag{49}$$

As a direct consequence of (**C5**), we can conclude that the transversality condition is satisfied. Then the proof of Lemma 3 is complete.

Based on the above analysis, the following conclusions can be drawn. □

Theorem 2. *By assuming that assumptions* (**C1**), (**C4**), *and* (**C5**) *are valid, the following conditions can be inferred:*

(i) The zero equilibrium point of fractional order four-neuron recurrent neural network with multiple delays (6) is asymptotically stable when $\tau_2 \in [0, \tau_{20}^*)$*)*

(ii) The fractional order four-neuron recurrent neural network with multiple delays (6) experiences a Hopf bifurcation at its origin when $\tau_2 = \tau_{20}^*$*; that is, a family of periodic solutions can bifurcate from the zero equilibrium point near* $\tau_2 = \tau_{20}^*$

5. Numerical Examples

To demonstrate the validity and feasibility of the conclusions reached in this paper, we provide two examples. The simulations were based on a prediction and correction scheme [53] of Adama–Bashforth–Moulton and step-size $h = 0.01$.

5.1. Example 1. Consider the four-neuron fractional recurrent neural networks with multiple delays as

$$\begin{cases} D^{\phi}x_1(t) = -x_1(t) + f(x_2(t-\tau_1)), \\ D^{\phi}x_2(t) = -x_2(t) + f(x_3(t-\tau_1)), \\ D^{\phi}x_3(t) = -x_3(t) + f(x_4(t-\tau_1)), \\ D^{\phi}x_4(t) = -x_4(t) + \omega_1 f(x_1(t-\tau_2)) + \omega_2 f(x_2(t-\tau_2)) + \omega_3 f(x_3(t-\tau_2)). \end{cases} \tag{50}$$

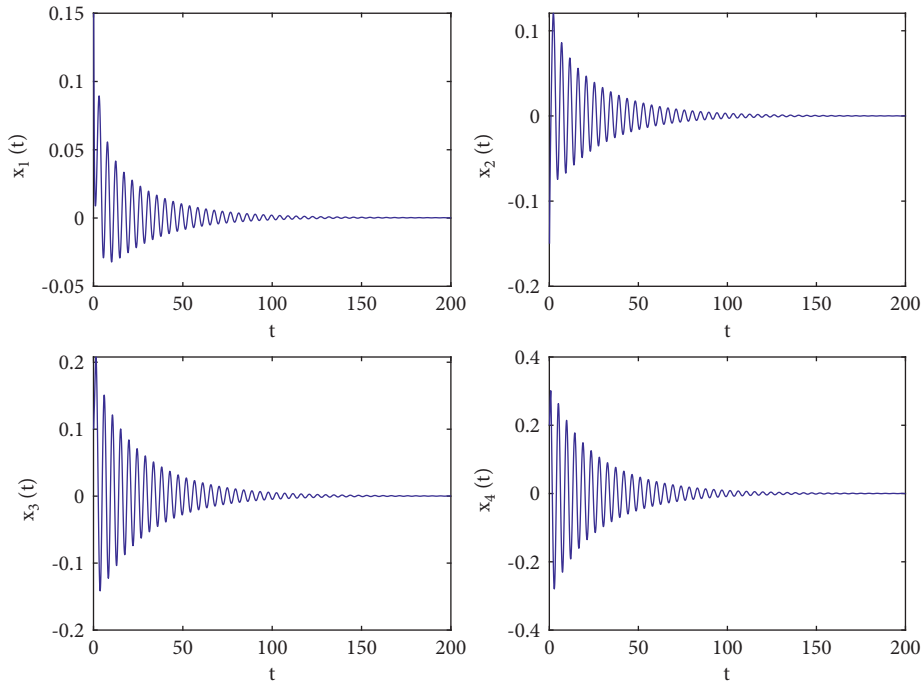


Figure 1: Time responses of system (50) with $\phi = 0.9$, $\tau_1 = 0.25 < \tau_{10} = 0.312709$.

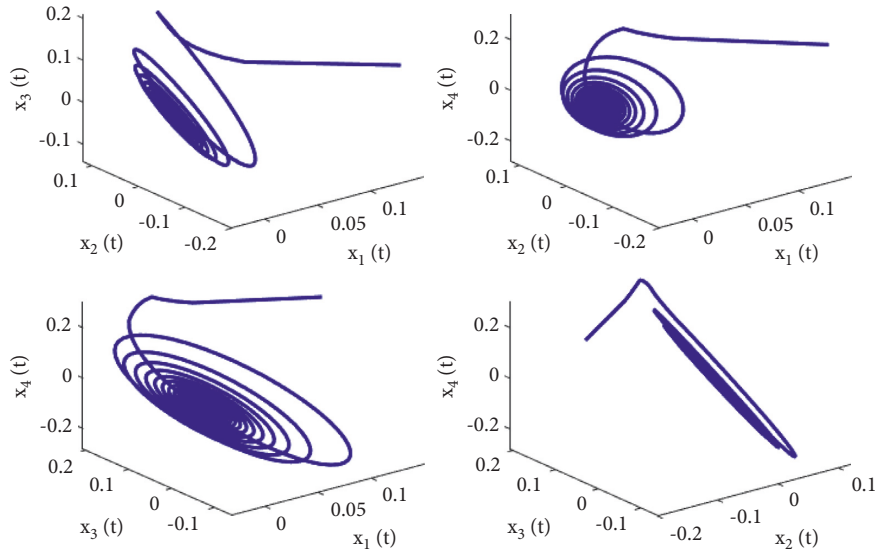


Figure 2: Phase diagrams of system (50) with $\phi = 0.9$, $\tau_1 = 0.25 < \tau_{10} = 0.312709$.

Choose parameters $\phi = 0.9, \omega_1 = 2, \omega_2 = \omega_3 = -2$, action function $f(\cdot) = \tanh(\cdot)$; therefore, $f(0) = \tanh(0) = 0$, $f'(0) = 1$.

Let the initial values be selected as $(x_1(0), x_2(0), x_3(0), x_4(0)) = (0.15, -0.14, 0.1, 0.2)$ for the system (50). First, taking fixed τ_2 such that $\tau_2 = 0.6$ by complex computing, we get $\omega_{10} = 5.23599$, and then $\tau_{10} = 0.312709$. Obviously, it is easy to verify that the conditions in Theorem 1 are satisfied. The numerical simulations in Figures 1 and 2 that the zero equilibrium point of system (50) is locally asymptotically stable when $\tau_1 = 0.25 < \tau_{10} = 0.312709$. Moreover, Figures 3 and 4 simulates that the zero equilibrium point of system (50) is unstable, and Hopf bifurcation occurs when $\tau_1 = 0.35 > \tau_{10} = 0.312709$. The bifurcation diagrams are plotted in Figure 5, which illustrates the theoretical results.

5.2. Example 2. The same as example 1, let $\phi = 0.95$, and now we consider the following four-neurons fractional current network with double different delays:

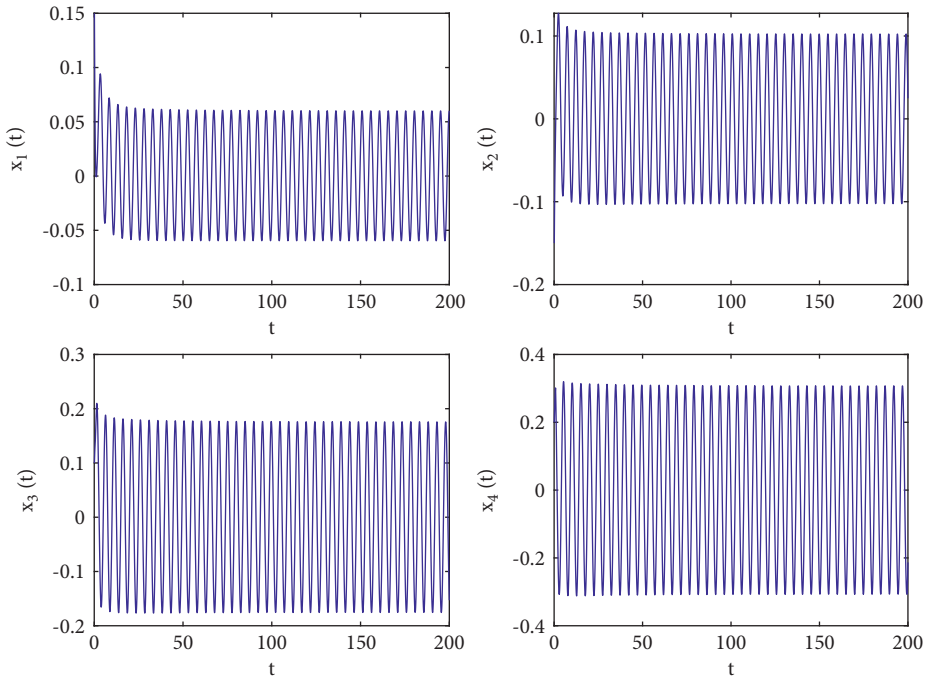


Figure 3: Time responses of system (50) with $\phi = 0.9$, $\tau_1 = 0.36 > \tau_{10} = 0.312709$.

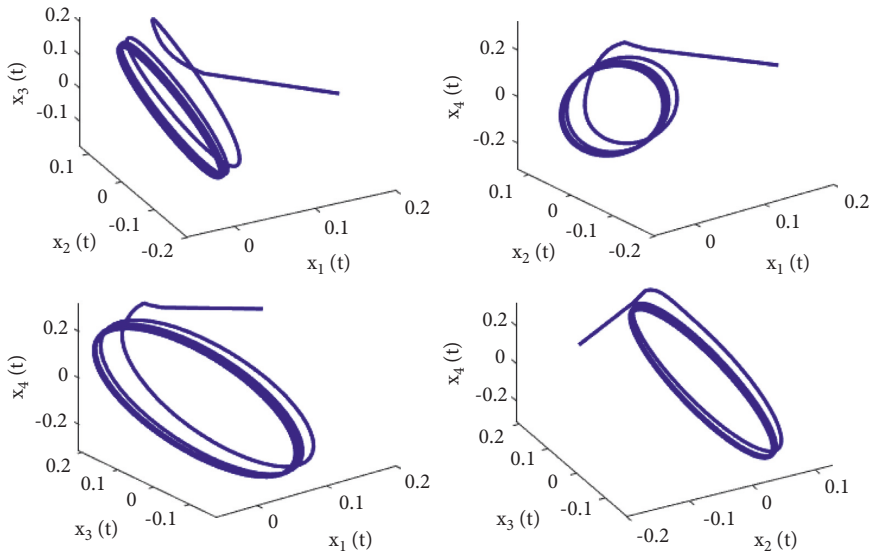


Figure 4: Phase diagrams of system (50) with $\phi = 0.9$, $\tau_1 = 0.36 > \tau_{10} = 0.312709$.

$$\begin{cases} D^{0.95}x_1(t) = -x_1(t) + f(x_2(t-\tau_1)), \\ D^{0.95}x_2(t) = -x_2(t) + f(x_3(t-\tau_1)), \\ D^{0.95}x_3(t) = -x_3(t) + f(x_4(t-\tau_1)), \\ D^{0.95}x_4(t) = -x_4(t) + \omega_1 f(x_1(t-\tau_2)) + \omega_2 f(x_2(t-\tau_2)) + \omega_3 f(x_3(t-\tau_2)). \end{cases} \tag{51}$$

Taking $\omega_1 = 1, \omega_2 = \omega_3 = -1.5, \phi = 0.95$, action function $f(\cdot) = \tanh(\cdot)$, then $f(0) = \tanh(0) = 0$, $f'(0) = 1$, and we first also set $\tau_1 = 0.8$, in the next step, we apply a complex calculation, and it obtains a $\widetilde{\omega}_{20} = 1.02089$ and $\tau_{20} = 0.329454$. Thus, Theorem 2 yields that the zero solution $(0, 0, 0, 0)$ of the system (51) is locally asymptotically stable when $\tau_2 = 0.22 < \tau_{20}$, which is simulated in Figures 6 and 7 which describes the impact of fractional order on τ_{20}. In

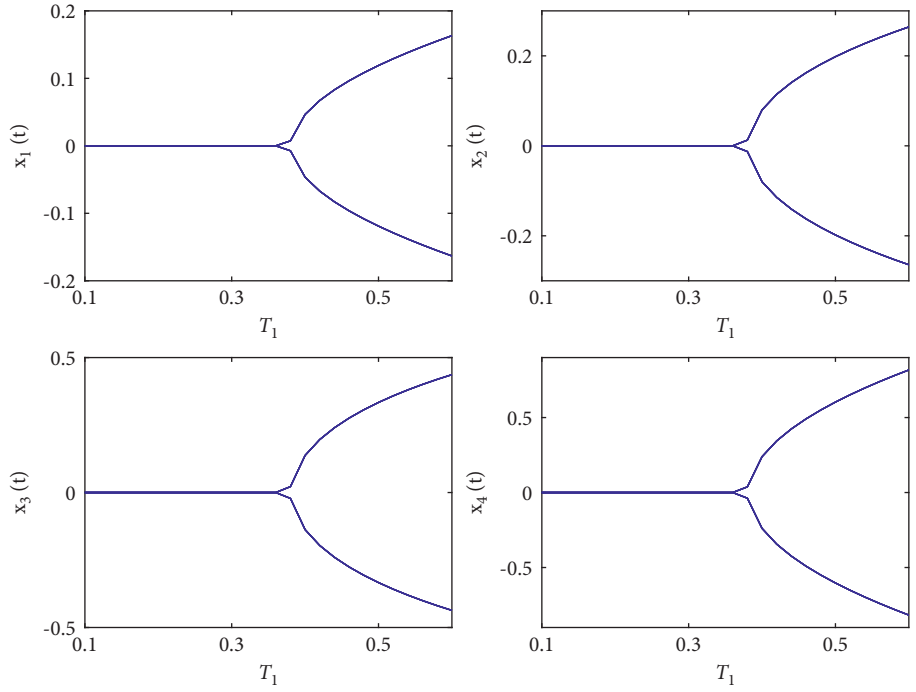


Figure 5: Bifurcation diagram of system (50) with $\phi = 0.9$, $\tau_1 = 0.36 > \tau_{10} = 0.312709$.

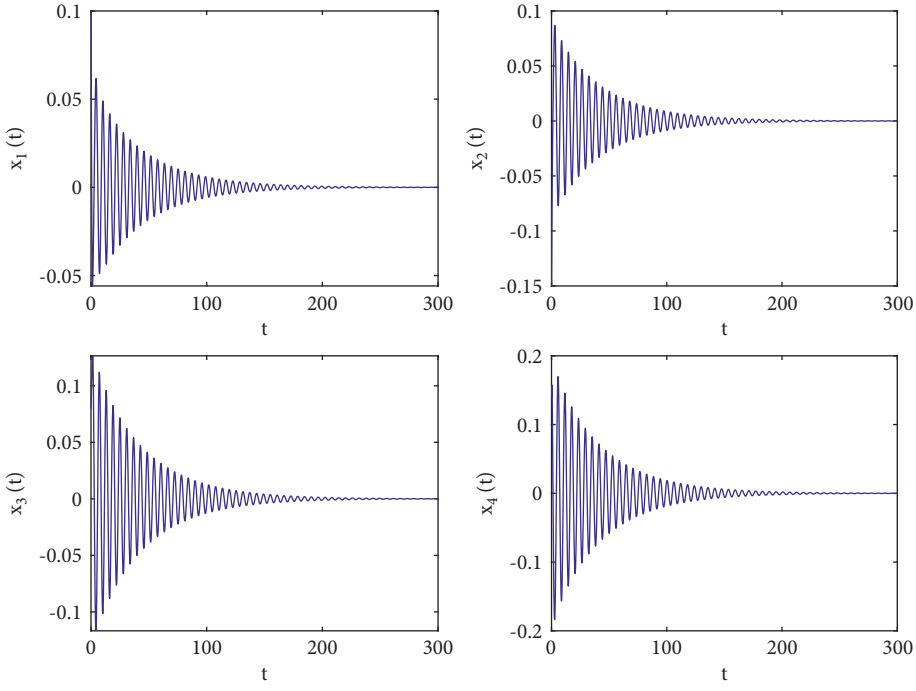


Figure 6: Time responses of system (51) with $\phi = 0.95$, $\tau_1 = 0.22 < \tau_{20} = 0.329454$.

addition, the zero equilibrium point of the system (51) is unstable, and Hopf bifurcation occurs when $\tau_2 = 0.38 > \tau_{20}$, as shown in Figures 8 and 9. Moreover, the bifurcation diagrams are plotted in Figure 10, which illustrates the theoretical results.

Remark 3. In fact, in order to better reflect the influence of different time delays at the bifurcation point of the systems (50) and (51), the corresponding bifurcation point τ_{10} and τ_{20} and τ_{10}^* and τ_{20}^* can be determined by changing the order of ϕ. This means that systems (50) and (51) involving

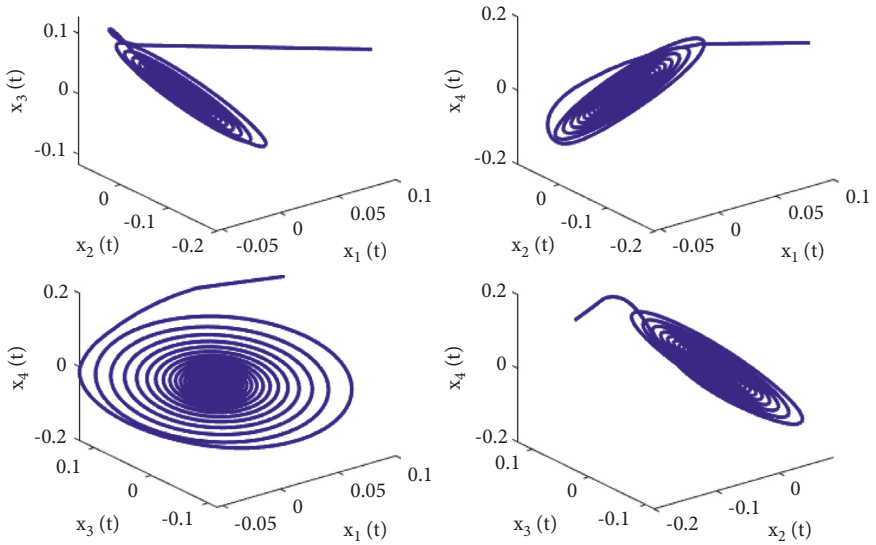


Figure 7: Phase diagrams of system (51) with $\phi = 0.95$, $\tau_1 = 0.22 < \tau_{20} = 0.329454$.

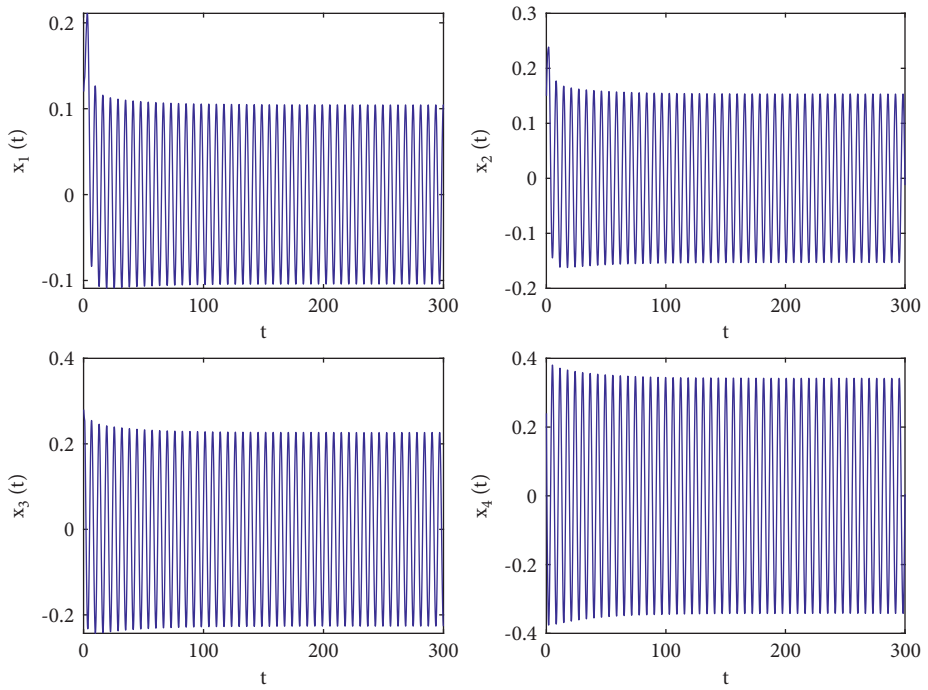


Figure 8: Time responses of system (51) with $\phi = 0.95$, $\tau_1 = 0.38 > \tau_{20} = 0.329454$.

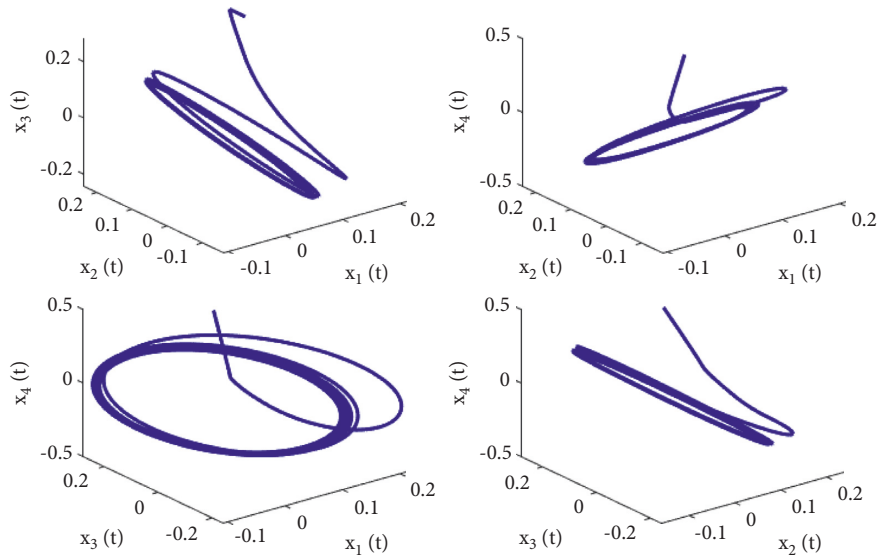


Figure 9: Phase diagrams of system (51) with $\phi = 0.95$, $\tau_1 = 0.38 > \tau_{20} = 0.329454$.

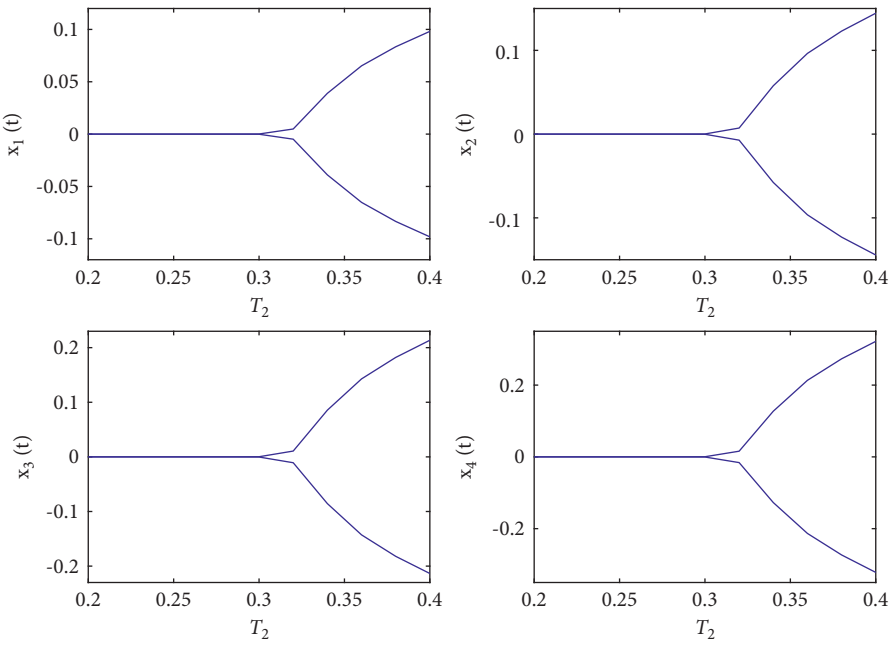


Figure 10: Bifurcation diagrams of system (51) with $\phi = 0.95$, $\tau_1 = 0.38 > \tau_{20} = 0.329454$.

different two delays are prone to earlier Hopf bifurcation for some fixed fractional order ϕ.

6. Conclusion

This paper examines the Hopf bifurcation problem of fractional recurrent neural networks with four neurons and two delays. Using time delay as the bifurcation parameter, several criteria are destabilized in order to ensure the Hopf bifurcation for the fractional four-neuron of recurrent neural networks. Based on our analysis, different communication time delays and order effects have quantitatively changed the dynamic behavior of the system (6). These results can contribute to our understanding of delayed fractional recurrent neural networks as a continuation of the previous work. The results of the simulations are illustrated by two numerical examples.

Appendix

A

$$
\begin{aligned}
A_1 &= 6\omega^{2\phi}\cos(\pi\phi) + 4\omega^{3\phi}\cos\left(\frac{3\pi\phi}{2}\right) + \omega^{4\phi}\cos(2\pi\phi) + 4\omega^{\phi}\cos\left(\frac{\pi\phi}{2}\right) + 1,\\
B_1 &= 6\omega^{2\phi}\sin(\pi\phi) + 4\omega^{3\phi}\sin\left(\frac{3\pi\phi}{2}\right) + \omega^{4\phi}\sin(2\pi\phi) + 4\omega^{\phi}\sin\left(\frac{\pi\phi}{2}\right),\\
A_2 &= -m_3m_6\omega^{2\phi}\sin(\pi\phi)\sin(\omega\tau_2) - m_3m_6\omega^{2\phi}\cos(\pi\phi)\cos(\omega\tau_2)\\
&\quad - 2m_3m_6\omega^{\phi}\sin\left(\frac{\pi\phi}{2}\right)\sin(\omega\tau_2) - 2m_3m_6\omega^{\phi}\cos\left(\frac{\pi\phi}{2}\right)\cos(\omega\tau_2) - m_3m_6\cos(\omega\tau_2),\\
B_2 &= -m_3m_6\omega^{2\phi}\sin(\pi\phi)\cos(\omega\tau_2) + m_3m_6\omega^{2\phi}\cos(\pi\phi)\sin(\omega\tau_2)\\
&\quad - 2m_3m_6\omega^{\phi}\sin\left(\frac{\pi\phi}{2}\right)\cos(\omega\tau_2) + 2m_3m_6\omega^{\phi}\cos\left(\frac{\pi\phi}{2}\right)\sin(\omega\tau_2) + m_3m_6\sin(\omega\tau_2),\\
A_3 &= -m_2m_3m_5\omega^{\phi}\sin\left(\frac{\pi\phi}{2}\right)\sin(\omega\tau_2) - m_2m_3m_5\omega^{\phi}\cos\left(\frac{\pi\phi}{2}\right)\cos(\omega\tau_2) - m_2m_3m_5\cos(\omega\tau_2),\\
B_3 &= -m_2m_3m_5\omega^{\phi}\sin\left(\frac{\pi\phi}{2}\right)\cos(\omega\tau_2) + m_2m_3m_5\omega^{\phi}\cos\left(\frac{\pi\phi}{2}\right)\sin(\omega\tau_2) + m_2m_3m_5\sin(\omega\tau_2),\\
A_4 &= -m_1m_2m_3m_4\cos(\omega\tau_2),\\
B_4 &= m_1m_2m_3m_4\sin(\omega\tau_2),\\
F_{11} &= -B_4^2(A_1B_4 + A_4B_1)\left(A_1^4 - A_1^2\left(A_3^2 + 2A_4^2 - 2B_1^2 + B_3^2 + 2B_4^2\right) + 2A_1(A_2A_3A_4\right.\\
&\quad \left. + A_2B_3B_4 + A_3B_2B_4 - A_4B_2B_3) - A_2^2\left(A_4^2 + B_4^2\right) + 2A_2B_1(A_4B_3 - A_3B_4)\right.\\
&\quad \left. - A_3^2B_1^2 + 2A_3A_4B_1B_2 + A_4^4 - 2A_4^2B_1^2 - A_4^2B_2^2 + 2A_4^2B_4^2 + B_1^4 - B_1^2B_3^2\right.\\
&\quad \left. - 2B_1^2B_4^2 + 2B_1B_2B_3B_4 - B_2^2B_4^2 + B_4^4\right),\\
F_{12} &= B_4\left(-\left(\left(A_4^2 + B_4^2\right)(A_2B_1 + A_3B_4) - (A_1B_4 + A_4B_1)(A_2A_4 + B_2B_4)\right)\left(A_1^2(B_1 - B_3)\right.\right.\\
&\quad \left. + A_1(A_2B_4 - A_4B_2) + B_1\left(A_2A_4 - A_4^2 + B_1^2 - B_1B_3 + B_2B_4 - B_4^2\right)\right) - B_4\left(-A_1^2A_2\right.\\
&\quad \left.\left. + A_1(A_3A_4 + B_3B_4) - B_1(A_2B_1 + A_3B_4 - A_4B_3)\right)\left(B_4\left(A_1^2 + B_1^2 - B_1B_3\right)\right.\right.\\
&\quad \left.\left. + B_4(B_2 - B_4)\right) - A_3(A_1B_4 + A_4B_1) + A_1A_4B_3 + A_4^2(B_2 - B_4)\right)\right),\\
F_{21} &= A_1^4 - A_1^2\left(A_3^2 + 2A_4^2 - 2B_1^2 + B_3^2 + 2B_4^2\right) + 2A_1(A_2A_3A_4 + A_2B_3B_4 + A_3B_2B_4\\
&\quad - A_4B_2B_3) - A_2^2\left(A_4^2 + B_4^2\right) + 2A_2B_1(A_4B_3 - A_3B_4) - A_3^2B_1^2 + 2A_3A_4B_1B_2\\
&\quad + A_4^4 - 2A_4^2B_1^2 - A_4^2B_2^2 + 2A_4^2B_4^2 + B_1^4 - B_1^2B_3^2 - 2B_1^2B_4^2 + 2B_1B_2B_3B_4 - B_2^2B_4^2 + B_4^4,\\
F_{22} &= A_1^3(-B_2) + A_1^2(A_2(B_1 + B_3) + A_3(B_4 - B_2) - A_4B_3) + A_1\left(A_2^2(-B_4) + 2A_2A_4B_2\right.\\
&\quad \left. + A_3^2B_4 - 2A_3A_4B_3 + A_4^2B_2 - B_1^2B_2 + B_2^2B_4 + B_2B_4^2 - B_3^2B_4\right) - A_2^2A_4B_1\\
&\quad + A_2\left(A_4^2(B_3 - B_1) + B_1^3 + B_1^2B_3 - B_1B_4(2B_2 + B_4) + B_3B_4^2\right) + A_3^2A_4B_1\\
&\quad - A_3A_4^2B_2 - A_3A_4^2B_4 - A_3B_1^2B_2 + A_3B_1^2B_4 + 2A_3B_1B_3B_4 - A_3B_2B_4^2 - A_3B_4^3\\
&\quad + A_4^3B_3 - A_4B_1^2B_3 + A_4B_1B_2^2 - A_4B_1B_3^2 + A_4B_3B_4^2.
\end{aligned}
\tag{A.1}
$$

B

$$
\begin{aligned}
a_1 &= 6\tilde{\omega}^{2\phi}\cos(\pi\phi) + 4\tilde{\omega}^{3\phi}\cos\left(\frac{3\pi\phi}{2}\right) + \tilde{\omega}^{4\phi}\cos(2\pi\phi) + 4\tilde{\omega}^{\phi}\cos\left(\frac{\pi\phi}{2}\right) + 1,\\
b_1 &= 6\tilde{\omega}^{2\phi}\sin(\pi\phi) + 4\tilde{\omega}^{3\phi}\sin\left(\frac{3\pi\phi}{2}\right) + \tilde{\omega}^{4\phi}\sin(2\pi\phi) + 4\tilde{\omega}^{\phi}\sin\left(\frac{\pi\phi}{2}\right),\\
a_2 &= -m_2m_3\left(m_1m_4\cos(3\tilde{\omega}\tau_1) + m_5\tilde{\omega}^{\phi}\sin\left(\frac{\pi\phi}{2}\right)\sin(2\tilde{\omega}\tau_1)\right)\\
&\quad - m_2m_3m_5\left(\tilde{\omega}^{\phi}\cos\left(\frac{\pi\phi}{2}\right)\cos(2\tilde{\omega}\tau_1) + \cos(2\tilde{\omega}\tau_1)\right)\\
&\quad - m_3m_6\left(\tilde{\omega}^{2\phi}\sin(\pi\phi)\sin(\tilde{\omega}\tau_1) + \tilde{\omega}^{2\phi}\cos(\pi\phi)\cos(\tilde{\omega}\tau_1)\right.\\
&\quad \left. +2\tilde{\omega}^{\phi}\sin\left(\frac{\pi\phi}{2}\right)\sin(\tilde{\omega}\tau_1) + 2\tilde{\omega}^{\phi}\cos\left(\frac{\pi\phi}{2}\right)\cos(\tilde{\omega}\tau_1) + \cos(\tilde{\omega}\tau_1)\right),\\
b_2 &= m_2m_3\left(m_1m_4\sin(3\tilde{\omega}\tau_1) + m_5\tilde{\omega}^{\phi}\sin\left(\frac{\pi\phi}{2}\right)\cos(2\tilde{\omega}\tau_1)\right) + m_2m_3m_5\left(\tilde{\omega}^{\phi}\cos\left(\frac{\pi\phi}{2}\right)\right.\\
&\quad \left.\times\sin(2\tilde{\omega}\tau_1) + \sin(2\tilde{\omega}\tau_1)\right) - m_3m_6\left(\tilde{\omega}^{2\phi}\sin(\pi\phi)\cos(\tilde{\omega}\tau_1) - \tilde{\omega}^{2\phi}\cos(\pi\phi)\sin(\tilde{\omega}\tau_1)\right.\\
&\quad \left. +2\tilde{\omega}^{\phi}\sin\left(\frac{\pi\phi}{2}\right)\cos(\tilde{\omega}\tau_1) - 2\tilde{\omega}^{\phi}\cos\left(\frac{\pi\phi}{2}\right)\sin(\tilde{\omega}\tau_1) - \sin(\tilde{\omega}\tau_1)\right).
\end{aligned}
\tag{B.1}
$$

Data Availability

Data sharing not is applicable in this article as no datasets were generated or analysed during the current paper.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 11971421, and Yunnan Fundamental Research Projects under Grant 202201AU070170, and Yunnan Provincial Department of Education Science Research Fund Project under Grants 2022J0480, 2022J0477 and 2022Y489.

References

[1] J. J. Hopfield, "Neurons with graded response have collective computational properties like those of two-state neurons," *Proceedings of the National Academy of Sciences*, vol. 81, no. 10, pp. 3088–3092, 1984.

[2] R. J. Williams and D. Zipser, "A Learning algorithm for continually running fully recurrent neural networks," *Neural Computation*, vol. 1, no. 2, pp. 270–280, 1989.

[3] P. J. Angeline, G. M. Saunders, and J. B. Pollack, "An evolutionary algorithm that constructs recurrent neural networks," *IEEE Transactions on Neural Networks*, vol. 5, no. 1, pp. 54–65, 1994.

[4] K. F. Wong, "A recurrent network mechanism of time integration in perceptual decisions," *Journal of Neuroscience*, vol. 26, no. 4, pp. 1314–1328, 2006.

[5] Z. D. Wang, Y. R. Liu, L. Yu, and X. H. Liu, "Exponential stability of delayed recurrent neural networks with Markovian jumping parameters," *Physics Letters A*, vol. 356, no. 4-5, pp. 346–352, 2006.

[6] M. Kobayashi, "Hyperbolic Hopfield neural networks with directional multistate activation function," *Neurocomputing*, vol. 275, pp. 2217–2226, 2018.

[7] P. P. Civalleri, M. Gilli, and L. Pandolfi, "On stability of cellular neural networks with delay," *IEEE Transactions on Circuits and Systems I Fundamental Theory and Applications*, vol. 40, no. 3, pp. 157–165, 1993.

[8] L. Wan, Q. H. Zhou, and J. Liu, "Delay-dependent attractor analysis of hopfield neural networks with time-varying delays," *Chaos, Solitons & Fractals*, vol. 101, pp. 68–72, 2017.

[9] P. C. Rech, "Chaos and hyperchaos in a Hopfield neural network," *Neurocomputing*, vol. 74, no. 17, pp. 3361–3364, 2011.

[10] J. D. Cao and J. Wang, "Global asymptotic stability of a general class of recurrent neural networks with time-varying delays," *IEEE Transactions on Circuits and Systems I*

Fundamental Theory and Applications, vol. 50, no. 1, pp. 34–44, 2003.

[11] C. J. Cheng, T. L. Liao, J. J. Yan, and C. C. Hwang, "Exponential synchronization of a class of neural networks with time-varying delays," *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 36, no. 1, pp. 209–215, 2006.

[12] H. Zhang and Z. G. Zeng, "Synchronization of recurrent neural networks with unbounded delays and time-varying coefficients via generalized differential inequalities," *Neural Networks*, vol. 143, pp. 161–170, 2021.

[13] A. C. C. Coolen and V. Del Prete, "Statistical mechanics beyond the Hopfield model: solvable problems in neural network theory," *Reviews in the Neurosciences*, vol. 14, no. 1-2, pp. 181–193, 2003.

[14] X. F. Liao, K. W. Wong, and Z. Wu, "Bifurcation analysis on a two-neuron system with distributed delays," *Physica D: Nonlinear Phenomena*, vol. 149, no. 1-2, pp. 123–141, 2001.

[15] K. Gu, V. L. Kharitonov, and C. Jie, *Stability of Time-Delay Systems*, Birkhuser Boston, Switzerland, 2003.

[16] M. Biggio, M. Storace, and M. Mattia, "Non-instantaneous synaptic transmission in spiking neuron networks and equivalence with delay distribution," *BMC Neuroscience*, vol. 14, no. S1, pp. 2677–P271, 2013.

[17] C. Aouiti, "Oscillation of impulsive neutral delay generalized high-order Hopfield neural networks," *Neural Computing & Applications*, vol. 29, no. 9, pp. 477–495, 2018.

[18] K. Smith and L. Wang, "Chaos in the discretized analog Hopfield neural network and potential applications to optimization," *Protein Science: A Publication of the Protein Society*, vol. 2, no. 2, pp. 1224–1231, 1998.

[19] Y. H. Xia and V. G. Romanovski, "Bifurcation analysis of a population dynamics in a critical state," *Bulletin of the Malaysian Mathematical Sciences Society*, vol. 38, no. 2, pp. 499–527, 2015.

[20] J. M. Wang, F. Q. Liu, and S. T. Qin, "Exponential stabilization of memristor-based recurrent neural networks with disturbance and mixed time delays via periodically intermittent control," *International Journal of Control, Automation and Systems*, vol. 19, no. 6, pp. 2284–2296, 2021.

[21] L. Zhou, "Delay-dependent and delay-independent passivity of a class of recurrent neural networks with impulse and multi-proportional delays," *Neurocomputing*, vol. 308, no. 25, pp. 235–244, 2021.

[22] L. Q. Zhou and Z. X. Zhao, "Exponential synchronization and polynomial synchronization of recurrent neural networks with and without proportional delays," *Neurocomputing*, vol. 372, no. 1, pp. 109–116, 2020.

[23] J. D. Cao and J. Wang, "Global Asymptotic and robust stability of recurrent neural networks with time delays," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 2, pp. 417–426, 2005.

[24] W. Zhang, C. D. Li, and T. W. Huang, "Global robust stability of complex-valued recurrent neural networks with time-delays and uncertainties," *International Journal of Biomathematics*, vol. 07, no. 02, Article ID 1450016, 2014.

[25] L. Berezansky and E. Braverman, "Exponential stability for systems of delay differential equations with block matrices," *Applied Mathematics Letters*, vol. 121, no. 2, Article ID 107364, 2021.

[26] Y. Y. Lv, L. J. Chen, F. D. Chen, and Z. Li, "Stability and bifurcation in an SI epidemic model with additive Allee effect and time delay," *International Journal of Bifurcation and Chaos*, vol. 31, no. 04, Article ID 2150060, 2021.

[27] W. H. Chen, Z. H. Guan, and X. Lu, "Delay-dependent exponential stability of uncertain stochastic systems with multiple delays: an LMI approach," *Systems & Control Letters*, vol. 54, no. 6, pp. 547–555, 2005.

[28] G. D. Zhang, Y. Shen, Q. Yin, and J. W. Sun, "Global exponential periodicity and stability of a class of memristor-based recurrent neural networks with multiple delays," *Information Sciences*, vol. 232, pp. 386–396, 2013.

[29] F. Qiu, B. T. Cui, and W. Wu, "Global exponential stability of high order recurrent neural network with time-varying delays," *Applied Mathematical Modelling*, vol. 33, no. 1, pp. 198–210, 2009.

[30] Y. J. Wang, C. L. Yang, and Z. Q. Zuo, "On exponential stability analysis for neural networks with time-varying delays and general activation functions," *Communications in Nonlinear Science and Numerical Simulation*, vol. 17, no. 3, pp. 1447–1459, 2012.

[31] Z. Z. Zhang and H. Z. Yang, "Hopf bifurcation analysis for a four-dimensional recurrent neural network with two delays," *Journal of Applied Mathematics*, vol. 2013, pp. 1–13, Article ID 436254, 2013.

[32] F. H. Zhang, T. W. Huang, Q. J. Wu, and Z. G. Zeng, "Multistability of delayed fractional-order competitive neural networks," *Neural Networks*, vol. 140, pp. 325–335, 2021.

[33] J. X. Lu and H. Xue, "Adaptive synchronization for fractional stochastic neural network with delay," *Advances in Difference Equations*, vol. 2021, pp. 1–12, 2021.

[34] J. Yuan and C. D. Huang, "Quantitative analysis in delayed fractional-order neural networks," *Neural Processing Letters*, vol. 51, no. 2, pp. 1631–1651, 2019.

[35] K. Udhayakumar and R. Rajan, "Hopf bifurcation of a fractional-order octonion-valued neural networks with time delays," *Discrete & Continuous Dynamical Systems - S*, vol. 13, no. 9, pp. 2537–2559, 2020.

[36] M. Xiao, W. X. Zheng, G. Jiang, and J. D. Cao, "Stability and bifurcation of delayed fractional-order dual congestion control algorithms," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, pp. 4819–4826, 2017.

[37] M. Xiao, W. X. Zheng, G. P. Jiang, and J. D. Cao, "Undamped oscillations generated by Hopf bifurcations in fractional-order recurrent neural networks with caputo derivative," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 12, pp. 3201–3214, 2015.

[38] C. D. Huang, Y. Meng, J. D. Cao, A. Alsaedi, and F. E. Alsaadi, "New bifurcation results for fractional BAM neural network with leakage delay," *Chaos, Solitons & Fractals*, vol. 100, pp. 31–44, 2017.

[39] M. Xiao, G. Jiang, J. D. Cao, and W. Zheng, "Local bifurcation analysis of a delayed fractional-order dynamic model of dual congestion control algorithms," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 2, pp. 361–369, 2017.

[40] Y. Cao, Y. Li, W. Ren, and Y. Chen, "Distributed coordination of networked fractional-order systems," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 2, pp. 362–370, 2010.

[41] H. Wang, Y. Yu, G. Wen, and S. Zhang, "Stability analysis of fractional-order neural networks with time delay," *Neural Processing Letters*, vol. 42, no. 2, pp. 479–500, 2015.

[42] C. D. Huang, J. D. Cao, and M. Xiao, "Hybrid control on bifurcation for a delayed fractional gene regulatory network," *Chaos, Solitons & Fractals*, vol. 87, pp. 19–29, 2016.

[43] Q. Sun, M. Xiao, B. Tao et al., "Hopf bifurcation analysis in a fractional-order survival red blood cells model and PD^{α}

control," *Advances in Differential Equations*, vol. 2018, pp. 1–10, 2018.

[44] C. J. Xu, Z. X. Liu, M. X. Liao, P. H. Li, Q. M. Xiao, and S. Yuan, "Fractional-order bidirectional associate memory (BAM) neural networks with multiple delays: the case of Hopf bifurcation," *Mathematics and Computers in Simulation*, vol. 182, pp. 471–494, 2021.

[45] H. Cao, D. Yan, and X. Xu, "Hopf bifurcation for an SIR model with age structure," *Mathematical Modelling of Natural Phenomena*, vol. 16, p. 7, 2021.

[46] E. Kaslik and I. R. Rădulescu, "Stability and bifurcations in fractional-order gene regulatory networks," *Applied Mathematics and Computation*, vol. 421, Article ID 126916, 2022.

[47] F. Yu and Y. S. Wang, "Hopf bifurcation and Bautin bifurcation in a prey-predator model with prey's fear cost and variable predator search speed," *Mathematics and Computers in Simulation*, vol. 196, pp. 192–209, 2022.

[48] C. Soresina, "Hopf Bifurcations in the Full SKT Model and where to Find Them," 2022, https://arxiv.org/abs/2202.04168.

[49] Z. H. Li, C. D. Huang, and Y. Zhang, "Comparative analysis on bifurcation of four-neuron fractional ring networks without or with leakage delays," *Advances in Differential Equations*, vol. 179, pp. 1–22, 2019.

[50] I. Podlubny, *Fractional Differential Equations*, Academic Press, New York, NY, USA, 1999.

[51] D. Matignon, "Stability results for fractional differential equations with applications to control processing," *IEEE-SMC Proceedings, Lille, France*, vol. 2, pp. 963–968, 1996.

[52] W. Deng, C. Li, and J. Lü, "Stability analysis of linear fractional differential system with multiple time delays," *Nonlinear Dynamics*, vol. 48, no. 4, pp. 409–416, 2007.

[53] S. Bhalekar and D. Varsha, "A predictor-corrector scheme for solving nonlinear delay differential equations of fractional order," *Fractional Calculus and Applied Analysis*, vol. 1, pp. 1–9, 2011.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 4728063, 10 pages
https://doi.org/10.1155/2022/4728063

Research Article

MTEDS: Multivariant Time Series-Based Encoder-Decoder System for Anomaly Detection

A. Reyana,[1] **Sandeep Kautish**,[2] **I. S. Yahia**,[3,4,5] **and Ali Wagdy Mohamed**[6,7]

[1]*Department of Computer Science and Engineering, Karunya Institute of Technology and Sciences, Coimbatore, Tamilnadu, India*
[2]*Department of Computer Science and Engineering, LBEF Campus, Kathmandu, Nepal*
[3]*Department of Physics, College of Science, King Khalid University, P.O. Box 9004, Abha 61413, Saudi Arabia*
[4]*Research Center for Advanced Materials Science (RCAMS), King Khalid University, P.O. Box 9004, Abha 61413, Saudi Arabia*
[5]*Department of Physics, Faculty of Education, Ain Shams University, Roxy, Cairo 11757, Egypt*
[6]*Operations Research Department, Faculty of Graduate Studies for Statistical Research, Cairo University, Giza 12613, Egypt*
[7]*Department of Mathematics and Actuarial Science School of Sciences Engineering, The American University in Cairo, Cairo 11835, Egypt*

Correspondence should be addressed to Sandeep Kautish; sandeepkautish@yahoo.com and Ali Wagdy Mohamed; aliwagdy@gmail.com

Received 19 July 2022; Accepted 27 August 2022; Published 28 September 2022

Academic Editor: Wei Xiang

Intrusion detection systems examine the computer or network for potential security vulnerabilities. Time series data is real-valued. The nature of the data influences the type of anomaly detection. As a result, network anomalies are operations that deviate from the norm. These anomalies can cause a wide range of device malfunctions, overloads, and network intrusions. As a result of this, the network's normal operation and services will be disrupted. The paper proposes a new multi-variant time series-based encoder-decoder system for dealing with anomalies in time series data with multiple variables. As a result, to update network weights via backpropagation, a radical loss function is defined. Anomaly scores are used to evaluate performance. The anomaly score, according to the findings, is more stable and traceable, with fewer false positives and negatives. The proposed system's efficiency is compared to three existing approaches: Multiscaling Convolutional Recurrent Encoder-Decoder, Autoregressive Moving Average, and Long Short Term Medium-Encoder-Decoder. The results show that the proposed technique has the highest precision of 1 for a noise level of 0.2. Thus, it demonstrates greater precision for noise factors of 0.25, 0.3, 0.35, and 0.4, and its effectiveness.

1. Introduction

Anomaly detection is a widely researched topic in machine learning and is crucial in many areas such as fraud detection, cyber security, and complex system health monitoring. As a result, it continues to be an active and challenging research area. The industrial multivariate time series data is the focus of this paper. Industrial systems, such as power plants, wind turbines, and engines, generate large amounts of time series data during their normal operations thanks to an increasing number of sensors and cost-effective data transmission and storage solutions. It is critical to monitor these systems for abnormal behavior, which, if not detected early, can have serious reliability consequences. Such relationships will not adhere to the learned representation in an abnormal situation, resulting in deviations. Today, new business opportunities are established through the use of the Internet and corporate networks. Numerous businesses and governments have established complex networks [1]. These networks are equipped with techniques for dispersed data storage, encryption, decryption, and authentication. Corporate networks are more accessible and utilize virtual private networks to enable employees to connect [2], thereby increasing the vulnerability of today's network to attackers. Network-based attacks are increasing in frequency and severity, resulting in increased financial losses. Firewalls have

ceased to provide appropriate protection [3]. According to a recent survey, forty new vulnerabilities are uncovered each month. This unsecured environment has necessitated the development of intrusion detection and prevention systems.

Intrusion detection systems do an analysis of the computer or network for potential security vulnerabilities. It identifies actions that jeopardize the integrity, reliability, and confidentiality of the system [2]. Time-series data are real-valued. Time series data are frequently used in financial, medicinal, and meteorological applications [4]. Routers and switches are used to facilitate network communication. Each of these entities contributes to the network's behavior. Due to the dynamic nature of the Internet protocol, it is difficult to comprehend the system. To comprehend network behavior, it is vital to understand how network probes are used [5]. This is a characteristic of uncontrolled networks. With an understanding of the network's layout, it is simple to identify network irregularities and performance bottlenecks. The type of anomaly detection varies according to the nature of the data [5]. Thus, network anomalies are operations that depart from the usual. These abnormalities can result in a variety of device malfunctions, overloads, and network incursions, among other things. This will cause a disruption in the network's normal operation and services. The normal behavior of a network is determined by a variety of factors, including the volume of traffic, the type of application, and the type of network data [5].

Today's commercially available network management systems monitor network activity in real-time and alert users to irregularities. The network manager monitors the environment and generates alarms when an unusual one occurs. Anomalies in networks are defined by changes in the measured data. Anomalies in networks are classified into two categories [5]. The first category results in network failures and performance problems such as file server errors and congestion. This abnormality will affect all users who share the network's available capacity. The second type results in security difficulties such as denial of service assaults, intrusions, and so forth. The correct type of network information is critical for any anomaly detection system. The more precise the traffic behavior, the more accurate the detection of anomalies.

1.1. Our Contribution

(i) To develop a novel Multi-variant Time-series- based Encoder-Decoder System (MTEDS) for detecting anomalies.

(ii) To construct multi-scale signature matrices to characterize multiple levels of the system statuses in different time steps.

(iii) To define a radical loss function to update the network weights through backpropagation.

(iv) To propose a technique that is noise-resistant and delivers a variety of anomaly scores based on the severity of the incident.

(v) To evaluate the performance of the proposed system in terms of anomaly score, noise factor, attention weight, anomaly count, precision, and recall. Finally, validate the proposed system against three existing approaches: Multiscaling Convolutional Recurrent Encoder-Decoder (MSCRED), Autoregressive Moving Average (ARMA), and Long Short Term Medium-Encoder-Decoder.

The remainder of the article's section is organized as follows: Section 2 discusses the various studies on time series-based anomaly detection that have been addressed. The proposed research methodology is illustrated in Section 3, followed by the results and discussion in Section 4, and finally the conclusion in Section 5.

2. Related Works

Anomaly detection in time series data has attracted the interest of a growing number of researchers in recent years. Numerous accessible techniques have difficulty in resolving these issues. Wang et al. in [6] created an online self-learning anomaly detection method to detect the abnormality automatically. Time series data are generated as a result of a succession of events, equipment sensors, or other sources. Anomalies can be structural, anomalous, or point in nature. In this case, the data points diverge. Structural does not follow the expected pattern, and anomalous time series contain deviations. In real-world applications, the difficulty is the occurrence of new anomalies with unknown behavior and duration. The study proposes techniques for detecting anomalies in univariate and multivariate time series. The algorithm is capable of detecting time series, locations, and sensors associated with them. Two algorithms were developed: SLADE TS and MTS. Initially, the algorithm splits the time series automatically into variable-length windows. Objects' behavioral patterns are not identified in all dimensions. However, accuracy and the determination of precise measurements and locations continue to be a challenge.

Memarzadeh et al. in [7] concentrated on identifying the National Airspace System's vulnerabilities. It is difficult to detect an anomaly in a high-dimensional heterogeneous time series. To capture the complicated patterns, the authors trained a convolutional variational auto-encoder. The model is composed of two components: an encoder that converts the original data space to a low-dimensional latent space, and a decoder that reconstructs the original data. The model optimizes mapping and quantifies the reconstruction error. This method is used to detect anomalies. Although the model performs well with high-dimensional data, dealing with heterogeneous multivariate time series data remains a barrier to vulnerability discovery. In [8] the authors surveyed the most advanced anomaly detection algorithms in the aviation domains. Anomaly detection is critical in a variety of industries. Due to the complexity of aviation's traffic circumstances, current anomaly detection approaches are insufficient. They evaluated current improvements in deep learning algorithms for scaling high-dimensional data in light of the availability of a vast amount of sensor data. Thus, it is clear that detecting anomalies in real-world applications continue to be a challenge.

Anomaly detection is the process of identifying anomalous behavior within a specified period. Li et al. in [9] proposed a method for detecting anomalies in multivariate time series using clustering. Due to the temporal link between points and individual time series, multivariate time series detection remains difficult. The authors classified aberrant signals into two categories based on their amplitude and shape. The fixed-length sliding window is initially constructed from multivariate subsequences. Subsequently, expanded Fuzzy-C-Means clustering is used. Thus, the anomaly score is identified, which quantifies the deviation from normal data. To deal with the initial feature space, the clustering procedure employs the Euclidean distance. Correlation of time series encapsulates shape information and identifies time series-based dissimilarity. Based on simulated datasets, the technique was tested. A confidence index was used to determine the cluster numbers and the interval length. As a result, the proposed approach recognized shape irregularities. While the approach performs admirably, it is limited in that it does not reveal the dimension of multivariate time series. Additionally, heavy computation generates overhead and is a time-consuming operation.

Khan et al. in [10] concentrated on anomaly detection in the aerospace domain, with a particular emphasis on machine learning techniques. Unsupervised anomaly detection is critical for detecting unknown cases in multivariate time series. The authors explored a variety of issues in their attempt to make a credible prognosis of the scenario. These include data availability, unsupervised learning for processing vast amounts of heterogeneous data, measurement error, and the complexity of learning real-time patterns. Despite the algorithms' strong results, determining the appropriate threshold level for isolating the anomaly remains a challenge. In [11] Han proposed an unsupervised method for detecting medical anomalies through the use of generative adversarial networks. The innovative method consists of two steps and is capable of detecting a variety of ailments. This phase covers the rebuilding of unhealthy and aberrant scans. The diagnosis phase then establishes the ground truth. The receiver operating characteristic (ROC) and area under the curve (AUC) were calculated. The study made use of the OASIS 3 clinical dementia rating dataset. Thus, the disease was diagnosed using many MRI sequence images. In [12] Fan and others examined the autoencoder's performance in detecting anomalies and constructing energy data. Autoencoders are a kind of neural network that is capable of doing unsupervised learning. There is no label variable in the learning process. The encoder takes identical inputs and converts them to high-level characteristics. The auto encoders deal with a variety of issues and are composed of fully connected layers. The autoencoder ensemble approach analyses energy data for abnormalities. The system was able to detect anomalies related to abnormal occurrences, operational errors, and inefficiency in the control strategy, among other things. However, developing unsupervised anomaly detection autoencoders remains a difficulty. The authors [13] concentrated on developing a secure and trustworthy system through the use of an anomaly detection technique. They deployed a DeepLog, a deep neural network with Long Short Term Memory that was used to represent the system log. DeepLog automatically discovers log patterns based on the log data. The approach makes use of the massive amount of system logs. The reason for this is that log data is unstructured and varies by system. It's difficult to diagnose unstructured log files. The recurrent neural network loops back from the previous state to the present input and maintains a record of previous predictions. The long-term repercussions are recalled. In a streaming approach, DeepLog trains on small data sets and recognizes the standard log sequence. However, the scientists noted that employing a recurrent neural network to diagnose a system remained challenging.

Al Osman and other experts in [14] place a premium on stress management through the use of ubiquitous biofeedback serious games. The device continuously monitors the subject's mental stress by monitoring their heart rate, respiration rate, and degree of activity. As with a biological sensory paradigm, the approach operates as a loop between the body and the brain. It takes physiological data throughout game sessions. Mental stress is monitored using electrodermal sensors. The two-way communication between the game and the player is established. As the player becomes immersed in the game, biological signals are captured and the time required to assess relaxation is recorded. The u-biofeedback system achieves continuous monitoring ubiquity. The algorithm has two modes of operation: limited and indefinite. The limited play can establish their monitoring period for a short time, whereas with indefinite, monitoring continues for the duration of the game. As a result, the technique aids individuals in determining their stress levels. With the proliferation of interconnected devices and sensors in the industrial system, it is critical to monitor for threats. This is especially crucial in power grids, water treatment facilities, and communication networks because as complexity develops, it becomes more difficult to spot anomalies. In [15], they introduced a novel graph deviation network to learn the sensor relationship graph. Four components comprise the approach: sensor embedding, graph structure learning, attention-based forecasting, and deviation score. The trials used datasets from water treatment plants to demonstrate the accuracy of the anomaly detection algorithm. While the model comprehends the anomaly, the authors' issue about improving anomaly detection with hyperparameters remains unresolved. The authors in [16] introduced a novel GAN model for anomaly detection and a time series imaging localization framework. The sequence of distance images from one to the next is memorized. The encoder and decoder convert multivariate time series to two-dimensional pictures. The encoder's pointwise convolution ensures that temporal information is encoded. The experiment is conducted using the smart power plant dataset. As a result, the generator generates a sequence of distance images. However, detecting anomalies in image localization continues to be a challenge.

Zhang et al. [17] noted that multivariate time series data are becoming more prevalent in the actual world. For instance, power plants and wearable technology. When dealing with

multivariate time series, the anomaly is identified, and the problem is capturing the temporal dependency inherent in each time series. Additionally, the system should be noise-resistant. Although the number of unsupervised anomaly detection algorithms continues to grow. The authors proposed the use of a Multiscale Convolutional Recurrent Encoder-Decoder (MSCRED) for anomaly detection in multivariate time series data. To begin, the technique generates a signature matrix that is used to characterize multi-level system states. Following that, the time-series correlation is encoded using the signature matrices in a convolutional encoder. Precision, recall, and $F1$ scores were determined during the examination. Thus, when applied to synthetic datasets, the model outperformed state-of-the-art approaches. In [18] the authors suggested a one-step-ahead load forecasting algorithm for estimating electricity requirements. The framework incorporates both the linear autoregressive and artificial neural network techniques. The Bayesian information criterion is used to counteract human-induced random events. The experimental results showed that the study is effective at selecting features from nonstationary time series data. Utilizing sophisticated approaches for anomaly prediction in the future remains a challenge. The Temporal Hierarchical One-Class (THOC) network was proposed in [19]. Initially, it used a recurrent neural network to extract multi-scale temporal properties from time series. The approach performed hierarchical clustering using the intermediate layer characteristics. This is because the upper layers include the characteristics with the lowest resolution. Multiple hyperspheres define the normal behaviors for each resolution. The intricate properties of real-world systems are captured through the use of multiscale support vector data descriptions. The one-class objectives are stated, and the end-to-end training of multiscale features and hypersphere centers is permitted. The deviation score is used to detect anomalies, and the approach outperforms contemporary state-of-the-art approaches. In [20] developed an improved deep model for detecting time series anomalies. The method was applied to a collection of hyperspheres with a hierarchical structure. To have access to all of time's events. The representation of features was enhanced to account for orthogonality and temporal self-supervision loss. Each component's effectiveness was evaluated.

Li et al. in [21] offered two solutions to the challenge of segmenting moving objects. The authors began with a transposed convolutional neural network and then added a Features Pooling Module. The first stage involved mapping the features to a pixel-level foreground probability map using the multiscale feature encoder and decoder. Then the binary segmentation labels were created. The proposed approaches are straightforward and operate with networks with numerous inputs. The review was conducted using publicly available datasets from CDnet2014, SBI2015, and UCSD. In [22] discussed the difficulties associated with anomaly detection in the Internet of Things. The Internet of Things facilitates data communication without the direct involvement of human agents. The gadgets interact with the environment via sensors, actuators, computers, and smart objects. The authors have highlighted several significant obstacles, including (i) Real-time processing, i.e., anomaly detection, which requires real-time decision-making. When the detector takes an extended period, the system fails. (ii) Incremental technique, i.e., maintaining the complete dataset for analysis is challenging, and using an incremental approach reduces memory needs. (iii) Online adaptive learning necessitates system reconfiguration.

Multivariate data on the environment and other sensor information must be addressed [23]. Thus, the author's point on historical anomaly detection for univariate exists. Yet multivariate data stream-based anomaly detection continues to be a difficult problem, which motivated this research. Although intrusion detection systems have evolved, next-generation anomaly detection systems face numerous obstacles. These include legacy intrusion detection systems that are incapable of adapting to emerging network paradigms such as wireless and mobile networks. They have not been scaled to meet the requirements of high-speed networks. Noise, constantly shifting traffic patterns, and massive networks go unaddressed [2]. False alarms must be suppressed. Despite the multiple methodologies used, there are currently no universally accepted criteria for evaluating intrusion detection systems. As a result, evaluating an intrusion detection system fairly using systematic criteria continues to be a challenge. Additionally, in today's actual network environment, having the necessary data for IDS evaluation is crucial [24]. Apart from this, another significant obstacle is the proliferation of attacks in various forms. The majority of intrusion detection systems are ineffective at detecting intrusions on time. Thus, the most difficult problem is multivariant time-series detection. While considerable work has been done in this area, additional research is necessary to discover realistically effective solutions [25].

3. Materials and Methods

Today's real-world systems generate an increasing amount of data. Identifying anomalous states and isolating the underlying reasons is challenging. The purpose of this study is to detect anomalies in multivariate time series data using a synthetic dataset of real power plants. In modern industrial facilities, complex systems are used. Numerous sensors are used to monitor the activity of these systems. The sensors generate multivariate time series data, which must be analyzed for anomalies [26]. In today's industrial world, multivariate time series data are generated by monitoring the behavior of complex systems, for instance, temperature and pressure sensor readings in a power plant. These systems must be repaired immediately to address the difficulties mentioned. Managing this procedure and undertaking exact detection would come at a significant financial and human cost [23]. In the long run, small perturbations in a system cannot be viewed as system failures. As a result, several different anomaly scores must be offered to aid system operations. This enables operators to take remedial action and rectify any difficulties that may arise. The discovered anomaly score shall indicate the system failure based on the sensor data. Accurate and prompt judgments are required to avert catastrophic losses in the case of a power plant. Additionally, it is required to locate and repair the sensors

responsible for the anomaly. Anomaly detection at various severity levels will finally reveal the robustness of the systems [27]. Systems that self-detect abnormalities are constructed using supervised algorithms. Although unsupervised algorithms exist, they face significant difficulties. In the event of an abnormality, multivariate time series data is critical for determining the aberrant condition and establishing the core causes [28, 29]. Constructing time series-based systems is challenging because it requires the intercorrelation of numerous time-series data. Thus, to achieve the abnormal status of identifying and diagnosing anomalies based on time series, this research study proposes a Multi-variant Time series-based Encoder-Decoder System (MTEDS), in which the system utilizes time-series inter-correlation with the help of the state vector and signature matrices, which aids in analyzing the anomaly score based on the client matrix series variation (residual matrices) in our dataset. Following an adequate number of training sessions, the network parameters are used to infer the reconstructed signature matrices for the validation and test data. Then, using latent vector characteristics, the anticipated number of anomalies may be checked against the enormous variety of Anomalies identified using this reconstruction [6].

The work introduces a novel MTEDS for handling anomalies in time series data with multiple variables. It begins by constructing multi-scale signature matrices to detect anomalies in the system's condition in a variety of time series. This value represents the degree to which the system is anomalous. Following that, the variational autoencoder is used to decode the time series data's correlation patterns. A convolutional long-short-term memory is used to capture the temporal patterns. Thus, the correlation between sensors and temporal information is sent to the Variational auto decoder for the reconstruction of the signature and residual matrices. The signature matrices are labeled S_{normal} and $S_{malicious}$, indicating which actions are considered normal and deleterious. Thus, the MTEDS was designed for anomaly detection and problem diagnosis. To evaluate the proposed MaVES system's performance, it is compared to three existing approaches: Multiscaling Convolutional Recurrent Encoder-Decoder (MSCRED), Autoregressive Moving Average (ARMA), and Long Short Term Medium-Encoder-Decoder (LSTM-ED). The proposed MTEDS framework is represented in Figure 1.

Initially, the system generates multi-scale signature matrices. The input is fed into the encoding vector, and each dimension represents a learned attribute. For each dimension, the suggested encoder employs a single value. Similarly, the decoder recreates the input's original value. The variation autoencoder depicts the observation in latent space, with the properties of the observation being used to calculate the probability distribution. Normally, the latent state for the input vector creation is chosen at random during the decoding phase. It is the feasible range of values as output feed into the decoder in the proposed model. Each time series' temporal dependency is correlated, and the system is noise-resistant. The various components of the proposed MTEDS framework are herewith described below.

3.1. Signature Matrices. The system status is characterized using different time series. Here for a multivariate time series of $m - w$ to m, we construct $n \times n$ signature matrices S^m. For a two-give time series as mentioned in

$$p_i^w = \left(p_i^{m-w}, p_i^{m-w-1}, \cdots p_i^w\right), \tag{1}$$

$$q_j^w = \left(q_j^{m-w}, q_j^{m-w-1}, \cdots q_j^w\right). \tag{2}$$

The correlation $S_{ij}^m \in S^m$ is calculated using

$$s_{ij}^m = \frac{\sum_{\delta=0}^{w} p_i^{m-\delta} p_j^{m-\delta}}{k}. \tag{3}$$

Here the rescaling factor is k and $k = w$. The signature matrix is assumed to be robust to noise and it does not capture the shape similarities. The segment interval is assumed as 10 and the signature matrix for $s = 3$ is constructed for lengths $w = 10$, 30, and 60.

3.2. Variational Auto Encoder. The variational autoencoder is a latent model that uses the regularized version of the autoencoder. The z, hidden codes enforce to draw samples using regular sampling. Here $\varphi(p)$ is the deterministic function, and the recognition model $t(z/p)$. The model employs a Gaussian disturbance over z with a neural network condition on p. The calculation is expressed in [21].

$$V(\theta, p) = -KV\left(b_\theta\left(\frac{z}{p}\right)a(z)\right) + Eb_\theta\left(\frac{z}{p}\right)\left[loga_\theta\left(\frac{p}{z}\right)\right] \leq loga(p). \tag{4}$$

This decodes every point in the latent space and has a reasonable probability.

3.3. Convolutional LSTM. The convolutional LSTM adapts hidden states across various time series. The feature maps $p^{m,l}$ from the lth layer of the convolutional neural network and $H^{m-1,l}$ the hidden state belongs to $R^{n1\times n1\times d1}$. Therefore, the calculations are expressed in

$$\widehat{H^{m,l}} = \sum_{i\in(m-h,m)} \alpha^i, H^{i,l}, \tag{5}$$

where α^i the previous weight is expressed in

$$\alpha^i = \frac{\exp\ \left\{vec\left(H^{m,l}\right)T\ vec\left(H^{i,l}\right)/k\right\}}{\sum_{i\in(m-h,m)} \exp\ \left\{vec\left(H^{m,l}\right)T\ vec\left(H^{i,l}\right)/k\right\}}, \tag{6}$$

where $vec()$ denotes the vector, k the rescaling factor. Thus the spatial patterns of the signature matrix are mapped with the temporal information at the convolutional layer.

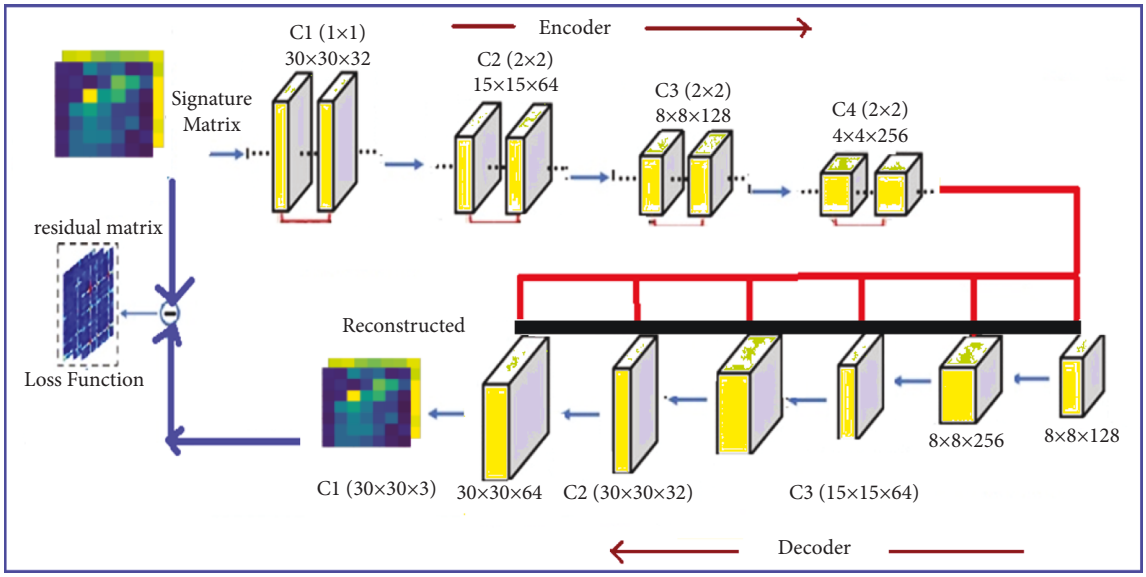


Figure 1: Framework of the proposed MTED.

3.4. Loss Function. The loss function defines the reconstruction errors over the signature matrix and is expressed in

$$L_f = \sum_{m} \sum_{c=1}^{s} \left\| p_{:::c}^{m,0} - \widehat{p}_{:::c}^{m,0} \right\|_F^2, \tag{7}$$

where $c \in R^{n \times n}$. The Adam optimization minimizes the loss function by employing stochastic gradient descent. The number of training epochs is sufficiently obtained the signature matrix is validated using the learned neural network. Based on the obtained residual matrix the anomaly is detected. Further, the performance of the proposed technique was evaluated using metrics such as anomaly score, noise factor, attention weight, anomaly count, precision, and recall.

4. Results and Discussion

This section describes the experimental setup, datasets, and various evaluation metrics.

4.1. Datasets. The synthetic dataset of a power plant was used. For the dataset the time series was formulated as follows:

$$\text{For } S_{\text{rand}} = 0; S(m) = \{\sin [m - m_0/w] + \lambda \cdot \epsilon\}$$
$$\text{For } S_{\text{rand}} = 1; S(m) = \{\cos [m - m_0/w] + \lambda \cdot \epsilon\}$$

The $S(m)$ value captures three sets of multivariate time series attributes. $C1$ is the temporal pattern simulation using trigonometric function; $C2$ is the time delay of the periodic cycles i.e. $m_0 \epsilon$ [50, 100] and $w \epsilon$ [40, 50]; $C3$ is the random Gaussian noise $n(0, 1)$ and the scaling factor λ is set as 0.3. The random selection of pairs of frequency and time at certain times leads to low and high correlations. Thus, the synthetic dataset considered consists of 30-time series, 20000 points, and 4 anomalies.

4.2. Experimental Setup. The experiment was conducted on a PC with a 3.6 GHz CPU, 12 GB of RAM, and Windows 10 OS. All codes are implemented in python language respectively.

4.3. Performance Evaluation Criteria and Parameters. The experiments were conducted to evaluate the various performance metrics such as anomaly score, noise factor, attention weight, anomaly count, precision, and recall. Finally, validate the proposed system against three existing approaches: Multiscaling Convolutional Recurrent Encoder-Decoder (MSCRED), Autoregressive Moving Average (ARMA), and Long Short Term Medium-Encoder-Decoder. Figures 2 and 3 depict the anomaly score results of the considered synthetic dataset.

Multiple levels of system status in various periods must be characterized. Following that, given the signature matrices, a convolutional encoder is used to encode the inter-time-series correlations based on the Client metric and changes, resulting in an alert and an increase in the anomaly score. The anomaly scores are used to assess the proposed MTEDS performance. The results in Figures 2 and 3 show that the suggested model is effective in terms of inter-sensor correlation and temporal patterns. The results reveal that the anomaly score is more consistent and detectable, with fewer false positives and negatives.

The anomaly periods are shown in blue in Figure 4, and the cutting threshold is shown in white. The results demonstrate the efficacy of the proposed model.

The matrix generated is shown in Figure 5 and Table 1. The matrices obtained from the residual and signature time series are validated based on the large variety of anomalies detected. Based on this, the reconstruction takes place, which is a latent vector where the predicted amount of the variation series is increased, and the reconstruction is enabled with the next set of matrices. The network weights are updated using

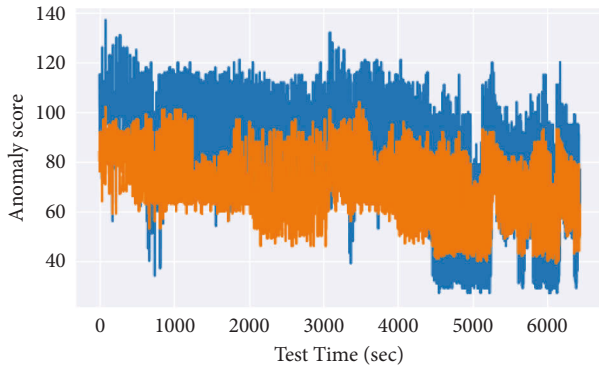


Figure 2: Anomaly score 1 results.

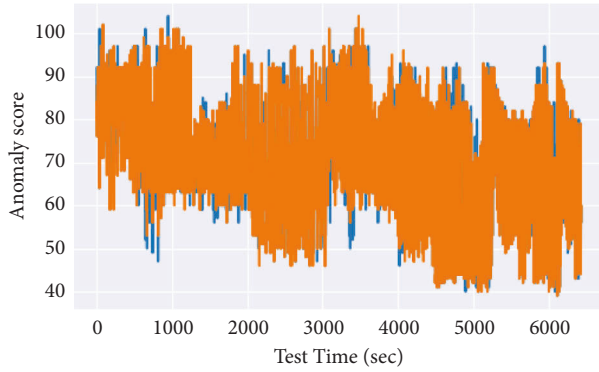


Figure 3: Anomaly score 2 results.

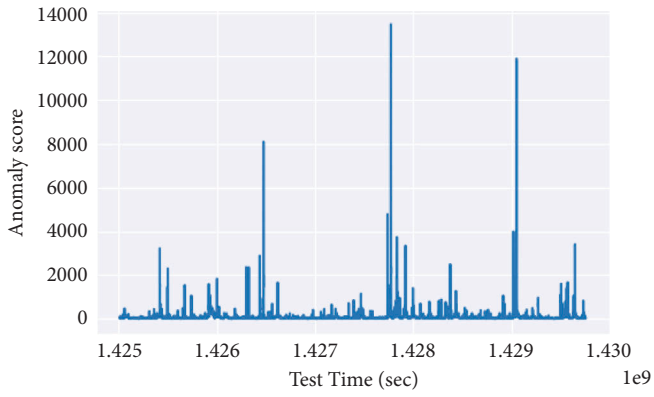


Figure 4: Mapping of anomaly detected against the threshold.

backpropagation after defining a differentiable loss function. Further variational autoencoders reduce the reconstruction error between the network's input and output by minimizing the generative model parameters m simultaneously. With the use of residual matrices, the evaluation of the vector time series sequence is subdivided into several parts with the help of the residual matrices by this the differentiation of how the anomaly variates are determined.

The distribution of attention weights over the previous timestamps is shown in Table 2 and Figure 6. Both the normal and abnormal attention weights are displayed separately. Even minor changes in the system status cause the attention to become more sensitive, which is useful for anomaly detection. The fourth and fifth-time steps, for instance, have an abnormal status of 0.22 and 0.23, respectively, whereas the first and second-time steps appear to be normal, and therefore the weights are assigned less than the normal segments.

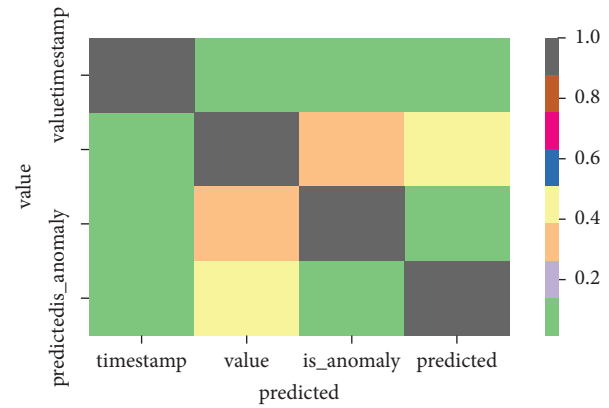


Figure 5: Generated matrix.

Table 1: Generated matrix.

Time stamp	1	0.1	0.1	0.1
Value	0.1	1	0.3	0.4
Anomaly	0.1	3	1	0.1
Predict	0.1	0.4	0.1	1
	Timestamp	*Value*	*is anomaly*	*Predicted*

Table 2: Distribution of attention weights.

Step of conv	Attention weight	
	Normal	Abnormal
	0.18	0.17
	0.19	0.18
	0.2	0.2
	0.21	0.22
	0.22	0.23

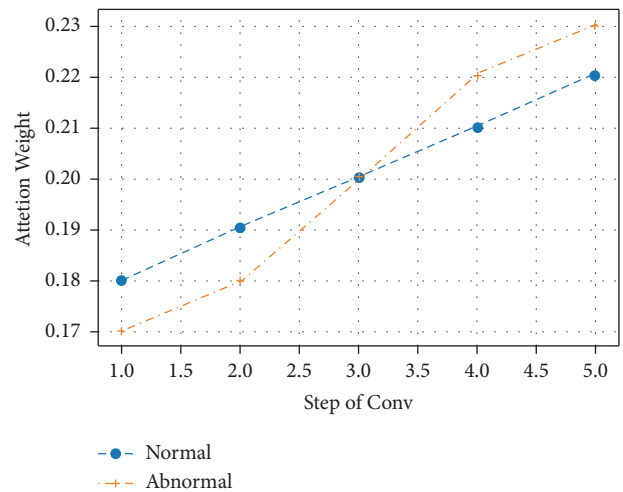


Figure 6: Distribution of attention weights.

Table 3: Performance evaluation of MTEDS.

Proposed MTEDS	
Noise factor	Precision
0.2	**1**
0.25	**1**
0.3	**1**
0.35	**0.98**
0.4	**0.95**
0.45	**0.8**

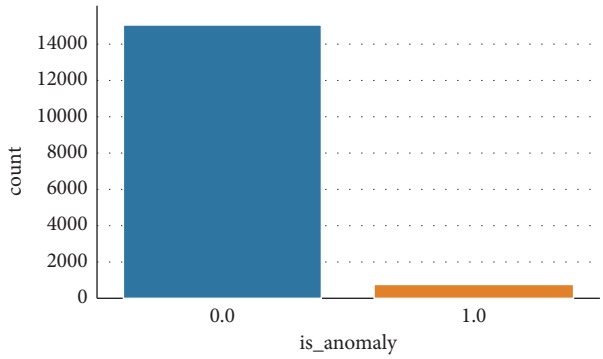


Figure 7: Performance evaluation of proposed MTEDS.

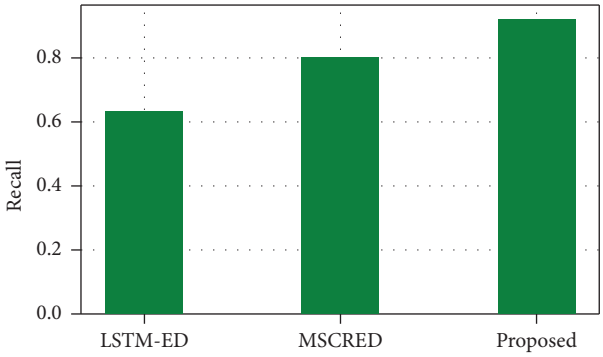


Figure 8: Recall measure comparison of the proposed MTEDS with the existing one.

Table 4: Precision measure comparison of the proposed with the existing system.

	Precision			
Noise factor	ARMA	LSTM-ED	MSCRED	*Proposed*
0.2	1	1	1	**1**
0.25	0.9	0.95	1	**1**
0.3	0.9	0.95	1	**1**
0.35	0.2	0.75	0.9	**0.98**
0.4	0.2	0.5	0.78	**0.95**
0.45	0.1	0.2	0.6	**0.8**

Table 3 and Figure 7 illustrate the performance results of the proposed MTEDS in terms of noise factors, precision, and is anomaly. The noise factor found appears to be relatively low, ranging between 0.2 and 0.45. The high precision value of 1 can be found in the majority of timestamps. As a result, the proposed system's efficiency is demonstrated. Further, the efficiency of the proposed MTEDS is compared with the three existing approaches: Multiscaling Convolutional Recurrent Encoder-Decoder (MSCRED), Autoregressive Moving Average (ARMA), and Long Short Term Medium-Encoder-Decoder.

Figure 8 shows the recall measure in comparison to the two previous systems, LSTM-ED and MSCRED. When

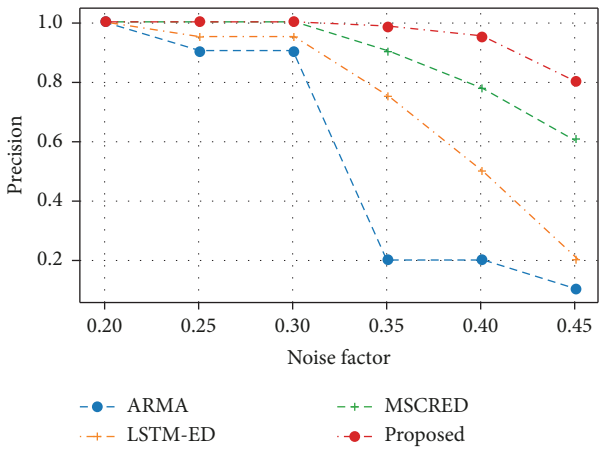


Figure 9: Precision measure comparison of the proposed MaVES with the existing.

compared to the MSCRED and LSTM-ED, which have recall measures of 0.81 and 0.63, respectively, the proposed MTEDS has the highest recall measure of 0.81. As a result, the proposed system's efficiency is demonstrated.

The precision measure comparison of the proposed MTEDS with the existing system is presented in Table 4 and Figure 9. The proposed approach had a precision of 1 for a noise level of 0.2. The precision value decreases as the noise level rises. However, for noise factors of 0.25, 0.3, 0.35, and 0.4, the proposed system had greater precision of 1, 1, 0.98, and 0.95, respectively. When compared to the proposed, the MSCRED has a precision of 1, 0.90, and 0.378. The least efficient systems were LSTM-ED and ARMA, demonstrating the efficacy of the proposed system.

5. Conclusion

To detect anomalies in multivariate time series data, the proposed research study MTEDS was implemented. It creates multi-scale signature matrices to describe many levels of system status in various time steps. As a result, a radical loss function is defined to update network weights via backpropagation. Multiple levels of system status were assessed over time. The suggested MTEDS performance is evaluated using anomaly scores. According to the findings, the anomaly score is more stable and traceable, with fewer false positives and negatives. The proposed technique provides the highest precision of 1 for a noise level of 0.2, according to the results. As the noise level increases, the precision value lowers. The proposed system, however, showed greater precision for noise factors of 0.25, 0.3, 0.35, and 0.4, proving its effectiveness. Although the proposed system performed better in terms of anomaly detection, the notion of anomaly would change as the dataset changes. It could be extended in the future to improve the method by applying it to different real-time datasets. Scaling and deploying the algorithm in complex operations might also be addressed while diagnosing the core cause of the problem. Other variables that could be examined for performance evaluation include latency and power analysis.

Data Availability

The data used to support the findings of this study are available on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this project under grant number R.G.P. 2/238/43.

References

[1] P. Esling and C. Agon, "Time-series data mining," *ACM Computing Surveys*, vol. 45, no. 1, pp. 1–34, 2012.

[2] A. Patcha and J. M. Park, "An overview of anomaly detection techniques: existing solutions and latest technological trends," *Computer Networks*, vol. 51, no. 12, pp. 3448–3470, 2007.

[3] A. T. Cemgil, S. Ghaisas, K. Dvijotham, S. Gowal, and P. Kohli, "Autoencoding variational autoencoder," 2020, https://arxiv.org/abs/2012.03715.

[4] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*, pp. 2–11, San Diego, CA, U.S.A, June 2003.

[5] M. Thottan and C. Ji, "Anomaly detection in IP networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2191–2204, 2003.

[6] X. Wang, J. Lin, N. Patel, and M. Braun, "Exact variable-length anomaly detection algorithm for univariate and multivariate time series," *Data Mining and Knowledge Discovery*, vol. 32, no. 6, pp. 1806–1844, 2018.

[7] M. Memarzadeh, B. Matthews, and I. Avrekh, "Unsupervised anomaly detection in-flight data using convolutional variational auto-encoder," *Aerospace*, vol. 7, no. 8, p. 115, 2020.

[8] T. Barbariol, E. Feltresi, and G. A. Susto, "Self-diagnosis of multiphase flow meters through machine learning-based anomaly detection," *Energies*, vol. 13, no. 12, p. 3136, 2020.

[9] J. Li, H. Izakian, W. Pedrycz, and I. Jamal, "Clustering-based anomaly detection in multivariate time series data," *Applied Soft Computing*, vol. 100, Article ID 106919, 2021.

[10] S. Khan, C. F. Liew, T. Yairi, and R. McWilliam, "Unsupervised anomaly detection in unmanned aerial vehicles," *Applied Soft Computing*, vol. 83, Article ID 105650, 2019.

[11] C. Han, L. Rundo, K. Murao et al., "MADGAN: unsupervised medical anomaly detection GAN using multiple adjacent brain MRI slice reconstruction," *BMC Bioinformatics*, vol. 22, no. S2, pp. 31–20, 2021.

[12] C. Fan, F. Xiao, Y. Zhao, and J. Wang, "Analytical investigation of autoencoder-based methods for unsupervised anomaly detection in building energy data," *Applied Energy*, vol. 211, pp. 1123–1135, 2018.

[13] M. Du, F. Li, G. Zheng, and V. Srikumar, "Deeplog: anomaly detection and diagnosis from system logs through deep learning," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1285–1298, Dallas, TX, U.S.A, October 2017.

[14] H. Al Osman, H. Dong, and A. El Saddik, "Ubiquitous biofeedback serious game for stress management," *IEEE Access*, vol. 4, pp. 1274–1286, 2016.

[15] A. Deng and B. Hooi, "Graph neural network-based anomaly detection in multivariate time series," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 5, pp. 4027–4035, Singapore, February 2021.

[16] Y. Choi, H. Lim, H. Choi, and I. J. Kim, "Gan-based anomaly detection and localization of multivariate time series data for power plant," in *Proceedings of the 2020 IEEE international conference on big data and smart computing (BigComp)*, pp. 71–74, IEEE, South Korea, February 2020.

[17] C. Zhang, D. Song, Y. Chen et al., "A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time series data," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, pp. 1409–1416, Honolulu, HI, USA, July 2019.

[18] J. An and S. Cho, "Variational autoencoder based anomaly detection using reconstruction probability," *Special Lecture on IE*, vol. 2, no. 1, pp. 1–18, 2015.

[19] J. Qiu, Q. Du, W. Wang, K. Yin, and L. Chen, "Short-term performance metrics forecasting for virtual machine to support anomaly detection using hybrid ARIMA-WNN model,"vol. 2, pp. 330–335, in *Proceedings of the 2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 2, pp. 330–335, IEEE, Milwaukee, WI, USA, July 2019.

[20] L. Shen, Z. Li, and J. Kwok, "Timeseries anomaly detection using temporal hierarchical one-class network," *Advances in Neural Information Processing Systems*, vol. 33, pp. 13016–13026, 2020.

[21] D. Li, D. Chen, J. Goh, and S. K. Ng, "Anomaly detection with generative adversarial networks for multivariate time series," 2018, https://arxiv.org/abs/1809.04758.

[22] A. A. Cook, G. Mısırlı, and Z. Fan, "Anomaly detection for IoT time-series data: a survey," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6481–6494, 2020.

[23] C. Lu, M. J. G. van de Molengraft, and G. Dubbelman, "Monocular semantic occupancy grid mapping with convolutional variational encoder–decoder networks," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 445–452, 2019.

[24] L. Basora, X. Olive, and T. Dubot, "Recent advances in anomaly detection methods applied to aviation," *Aerospace*, vol. 6, no. 11, p. 117, 2019.

[25] C. Zhou and G. Neubig, "Morphological inflection generation with multi-space variational encoder-decoders," in *Proceedings of the CoNLL SIGMORPHON 2017 Shared Task: Universal Morphological Reinflection*, pp. 58–65, Canada, August 2017.

[26] Z. Niu, K. Yu, and X. Wu, "LSTM-based VAE-GAN for time-series anomaly detection," *Sensors*, vol. 20, no. 13, p. 3738, 2020.

[27] L. Liu, X. Song, K. Chen, B. Hou, X. Chai, and H. Ning, "An enhanced encoder–decoder framework for bearing remaining useful life prediction," *Measurement*, vol. 170, Article ID 108753, 2021.

[28] A. I. Taloba, A. Elhadad, R. M. A. El-Aziz, and O. R. Shahin, "Prediction of data threats over web medium using advanced blockchain based information security with crypto strategies," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–17, 2021.

[29] A. Ben Aissa, A. I. Taloba, R. M. Abd-Aziz, I. Abdalla Mohamed, and L. F. Hussein, "Empirical analysis of security enabled cloud computing strategy using attribute based encryption scheme," *Materials Today Proceedings*, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 8157794, 10 pages
https://doi.org/10.1155/2022/8157794

Research Article

New Criteria for Synchronization of Multilayer Neural Networks via Aperiodically Intermittent Control

Taiyan Jing,[1] **Daoyuan Zhang**,[2] **and Xiaohua Zhang**[3]

[1]*School of Mathematics and Information Science, Henan Polytechnic University, Jiaozuo 454000, China*
[2]*Department of Artificial Intelligence and Data Science, Guangzhou Xinhua University, Guangzhou 523133, Guangdong Province, China*
[3]*Department of Mathematics, Jiangsu Vocational Institute of Commerce, Nanjing 210016, China*

Correspondence should be addressed to Taiyan Jing; taiyan825@163.com

Received 18 June 2022; Revised 3 August 2022; Accepted 8 August 2022; Published 27 September 2022

Academic Editor: Wei Xiang

In this paper, the globally asymptotic synchronization of multi-layer neural networks is studied via aperiodically intermittent control. Due to the property of intermittent control, it is very hard to deal with the effect of time-varying delays and ascertain the control and rest widths for intermittent control. A new lemma with generalized Halanay-type inequalities are proposed first. Then, by constructing a new Lyapunov–Krasovskii functional and utilizing linear programming methods, several useful criteria are derived to ensure the multilayer neural networks achieve asymptotic synchronization. Moreover, an aperiodically intermittent control is designed, which has no direct relationship with control widths and rest widths and extends existing aperiodically intermittent control techniques, the control gains are designed by solving the linear programming. Finally, a numerical example is provided to confirm the effectiveness of the proposed theoretical results.

1. Introduction

In the past few decades, coupled neural networks have drawn much attention because of their inherent characteristics and wide applications, such as secure communication [1], image encryption [2], and information processing [3]. As one of the fundamental research areas, synchronization is used to better understand the self-organization phenomena among coupled systems, which can be existed in many physical, social, and biological systems with various applications [4–6]. From the viewpoint of practical applications, investigating globally asymptotic synchronization of coupled neural networks is meaningful [7, 8].

It is well known that time delays are unavoidable for coupled neural networks due to the limited bit rate of communication channels and the limited bandwidth. Therefore, much attention has been attracted to studying the synchronization problem of coupled neural networks with time delays [9]. Yang et al. [10] investigated the synchronization problem of coupled time-delay neural networks with mode-dependent average dwell time switching. In [11], synchronization of memristive neural networks with mixed delays via quantized intermittent control was considered. However, the aforementioned results of delayed neural networks are based on one or two layers networks. In fact, multilayer neural networks with more than two layers can be seen as some subnetworks distributed in different layers. For example, there exist three-layers networks about information transmission in smart grids [12] and Kuramoto-oscillator networks [13–15]. Therefore, it is necessary to study the globally asymptotic synchronization for multilayer dynamic networks with time delays.

To drive dynamic networks to achieve synchronization, suitable controllers should be designed and added to the nodes of dynamic networks. In practical applications, the transmitted information is inevitably affected by external perturbations, which make the information weak or interrupted intermittently. In this case, the continuous-time control is not suitable. Hence, intermittent control schemes were proposed [16, 17]. Moreover, intermittent control can greatly reduce control cost and the amount of transmitted signals. Considering the fact that the structural limitation of

periodically intermittent control is not applicable in reality [18], aperiodically intermittent control was considered in [19, 20], which is characterized by nonfixed control time and rest time in a nonfixed time span. In [21], Liu et al. investigated the exponential synchronization problem for linearly coupled networks with delay by using aperiodically intermittent control. The authors of [22, 23] considered finite-time synchronization of delayed dynamic networks via aperiodically intermittent control. However, the above-given existing aperiodically intermittent control is complex and some strict restrictions are imposed on the control width and noncontrol width, which make it difficult to be implemented in practice. It is demonstrated that new aperiodically intermittent control methods [24] are proposed to improve the above existing control. Thus, adopting a new control strategy to study the asymptotic synchronization of multilayer delayed networks is necessary.

Motivated by the above-given analysis, this paper is devoted to studying the asymptotic synchronization of multilayer delayed networks by aperiodically intermittent control. The main contributions are summarized as follows:

(1) An original lemma is an extended form of many general Halanay-type differential inequalities [21, 25–27]. The lemma is proposed for the asymptotic stability with an intermittent divergence of system's state and is applicable to the asymptotic synchronization of delayed networks with intermittent control.

(2) Aperiodically intermittent controllers without any information of time delays are designed, which need less restrictive conditions and make the controllers more economic and practical than those controllers in [21–23].

(3) Novel Lyapunov–Krasovskii functional is designed, which can reduce the conservativeness of the results. Sufficient conditions derived by linear programming methods are acquired to ensure the asymptotic synchronization of delayed networks with intermittent control, where the effects of time delays are well dealt with.

The rest of this paper is organized as follows: in Section 2, some necessary assumptions and lemmas are given. In Section 3, the asymptotic synchronization of multilayer delayed neural networks is studied via aperiodically intermittent control. In Section 4, a numerical example is given to verify the effectiveness of the proposed theoretical schemes. Conclusions are drawn in Section 5.

Notations: let $\mathbb{R}^n$ denote the n-dimensional Euclidean space and $\mathbb{R}^{n\times m}$ denote the set of $n \times m$ real matrix. $\otimes$ stands for Kronecker product. $a_i(t), b_i(t)\,(i = 1, 2, 3, 4)\,(t \in \mathbb{R})$ are continuous and bounded functions. $a_{i*} = \inf_{t\in\mathbb{R}} a_i(t), a_i^* = \sup_{t\in\mathbb{R}} a_i(t), b_{i*} = \inf_{t\in\mathbb{R}} b_i(t), b_i^* = \sup_{t\in\mathbb{R}} b_i(t), i = 1, 2, 3, 4$.

2. Preliminaries

In this paper, we consider the following dynamic networks with p layers and N nonlinearly identical nodes:

$$\begin{aligned}\dot{x}_i(t) = {} & f_i(x_i(t)) + \sum_{j=1}^{N} a_{ij}^0 g_{0j}\big(x_j(t)\big) \\ & + \sum_{r=1}^{p-1}\sum_{j=1}^{N} a_{ij}^r g_{rj}\big(x_j(t-\tau_r(t))\big),\end{aligned} \tag{1}$$

where $x_i(t) = [x_{i1}(t), x_{i2}(t), \ldots, x_{in}(t)]^T \in \mathbb{R}^n$ denotes the state vector of the i th node. The function $f_i(\cdot)$: $\mathbb{R}^n \longrightarrow \mathbb{R}^n$ and $g_{rj}(\cdot)$: $\mathbb{R}^n \longrightarrow \mathbb{R}^n$ are continuous.

To simply the notations, let $x(t) = [x_1(t)^T, \ldots, x_N(t)^T]^T$, $f(x) = [f_1(x_1), \ldots, f_N(x_N)]^T$ and $g_r(x) = [g_{r1}^T, \ldots, g_{rN}^T]^T\,(r = 0, 1, \ldots, p-1)$, $g_{rj}(\cdot)$: $\mathbb{R}^n \longrightarrow \mathbb{R}^n$ are continuous. The functions $\tau_r(t) > 0\,(r = 1, \ldots, p-1)$ denote the bounded and continuously differentiable coupling delays, which means there exist positive constants $\tilde{\tau}$ and ϖ such that $0 < \tau_r(t) \le \tilde{\tau}$ and $0 \le \dot{\tau}_r(t) \le \varpi < 1\,(r = 1, \ldots, p-1)$. $A_r = (a_{ij}^r) \in \mathbb{R}^{N\times N}\,(r = 0, 1, \ldots, p-1)$ are the weight configuration matrices. If there is a link between nodes i and $j\,(i \ne j)$, then $a_{ij}^r = a_{ji}^r > 0$; otherwise, $a_{ij}^r = a_{ji}^r = 0$, and the diagonal elements of matrices A_r are represented by the following equation:

$$a_{ii}^r = -\sum_{j=1, j\ne i}^{N} a_{ij}^r, r = 0, 1, \ldots, p-1. \tag{2}$$

For simplicity, the drive system (5) can be written in the Kronecker product form:

$$\begin{aligned}\dot{x}(t) = {} & f(x(t)) + \big(A_0 \otimes I_n\big) g_0(x(t)) \\ & + \sum_{r=1}^{p-1} \big(A_r \otimes I_n\big) g_r\big(x\big(t-\tau_r(t)\big)\big),\end{aligned} \tag{3}$$

The corresponding response systems are written as follows:

$$\begin{aligned}\dot{y}(t) = {} & f(y(t)) + \big(A_0 \otimes I_n\big) g_0(y(t)) \\ & + \sum_{r=1}^{p-1} \big(A_r \otimes I_n\big) g_r\big(y\big(t-\tau_r(t)\big)\big) + u(t),\end{aligned} \tag{4}$$

where $y(t) = (y_1^T(t), \ldots, y_N^T(t))^T$, $y_i(t) = [y_{i1}(t), y_{i2}(t), \ldots, y_{in}(t)]^T \in \mathbb{R}^n$ denotes the response output vector of the i th node. $u(t) = [u_1^T(t), \ldots, u_N^T(t)]^T$, $u_i(t)$ denotes the input control of the i th node.

Remark 1. When the multi-layer parameter $p = 2$, the same models (3) degenerate into that considered in [23], even many similar models to (3) were discussed, e.g., [21, 22]. Therefore, the models of systems (3) in this paper are broader model forms. In addition, most of the practical neural networks are interrelated and interact with each other such that they generate more complicated structures and unpredictable behaviors than that with one layer. That is, general models with multilayer structures can simulate the real network world better, please see, e.g., [28, 29]. Thus, the models are worthy to be further discussed.

In this paper, the structure of aperiodically intermittent control is briefly described as follows: each time span

$[t_{2k}, t_{2k+1})$ and $[t_{2k+1}, t_{2k+2})$ $(t_0 = t_{-1} = t_{-2} = 0, k = -1, 0, 1, 2, \ldots)$ denote the control time and the rest time, respectively. The aperiodically intermittent control becomes a periodic one, when $t_{2k+2} - t_{2k} \equiv T$, $t_{2k+1} - t_{2k} \equiv \delta$, where T, δ are positive constants and $0 < \delta < T$.

The main objective is to apply the aperiodically intermittent control to force the states of networks (4) to be asymptotically synchronized with the ones of (3), i.e., $\lim_{t \longrightarrow \infty} \|y(t) - x(t)\| = 0$. The multilayer error systems $e(t) = y(t) - x(t)$ are obtained as follows:

$$\begin{aligned} \dot{e}(t) = {} & f(y(t)) - f(x(t)) + (A_0 \otimes I_n) g_0(y(t)) \\ & - (A_0 \otimes I_n) g_0(x(t)) + \sum_{r=1}^{p-1} (A_r \otimes I_n) g_r(y(t - \tau_r(t))) \\ & - \sum_{r=1}^{p-1} (A_r \otimes I_n) g_r(x(t - \tau_r(t))) + u(t). \end{aligned} \tag{5}$$

To obtain our main results, the following assumptions and lemmas are given as follows.

Assumption 1 (see [30]). If there exists a positive definite diagonal matrix $P = \mathrm{diag}\{P_1, P_2, \ldots, P_N\} \in \mathbb{R}^{nN \times nN}$, a diagonal matrix $\Delta = \mathrm{diag}\{\Delta_1, \Delta_2, \ldots, \Delta_N\} \in \mathbb{R}^{nN \times nN}$, and a positive scalar ξ such that

$$\begin{aligned} & (y(t) - x(t))^T P [f(y(t)) - f(x(t)) - \Delta(y(t) - x(t))] \\ & \quad \leq -\xi (y(t) - x(t))^T (y(t) - x(t)), \end{aligned} \tag{6}$$

holds for any $x(t), y(t) \in \mathbb{R}^{nN}$, where $P_i, \Delta_i \in \mathbb{R}^{n \times n}$ are diagonal matrices.

Lemma 1 (see [31]). *Given any real matrices X, Y and K of appropriate dimensions and a scalar $\varepsilon > 0$ such that $K = K^T > 0$. Then, the following inequality holds:*

$$X^T Y + Y^T X \leq \varepsilon X^T K X + \varepsilon^{-1} Y^T K^{-1} Y. \tag{7}$$

Obviously, when $K = I$ (I is an identity matrix), the inequality is transformed into $X^T Y + Y^T X \leq \varepsilon X^T X + \varepsilon^{-1} Y^T Y$.

Lemma 2 (see [32]). *Let $\tilde{x}(\cdot)$ be a nonnegative function satisfying the following equation:*

$$\begin{cases} \dot{\tilde{x}}(t) \leq -a(t)\tilde{x}(t) + b(t) \left(\sup_{t - \tau(t) \leq s \leq t} \tilde{x}(s), \right) t > t_0, \\ \tilde{x}(s) = |\varphi(s)|, \forall s \in [t_0 - \tau^*, t_0], \end{cases} \tag{8}$$

where $\tau(t)$ denotes a nonnegative, continuous and bounded function defined for $t \in \mathbb{R}$ and $\tau^* = \sup_{t \in \mathbb{R}} \tau(s)$; $\varphi(s)$ is continuous and defined for $s \in [t_0 - \tau^*, t_0]$; $a(t)$ and $b(t)$ $(t \in \mathbb{R})$ denote nonnegative, continuous, and bounded functions. Suppose

$$a(t) - b(t) \geq \sigma, t \in \mathbb{R}, \tag{9}$$

where $\sigma = \inf_{t \in \mathbb{R}} (a(t) - b(t)) > 0$. Let $0 < \tilde{\sigma} < \sigma$, there exists a positive number $\tilde{\mu}$ satisfying the following inequality:

$$-a(t) + \tilde{\mu} + b(t) \exp\{\tilde{\mu} \tau^*\} \leq -\tilde{\sigma} < 0, \text{for all} t \in \mathbb{R}. \tag{10}$$

Then,

$$\tilde{x}(t) \leq \left(\sup_{s \in [t_0 - \tau^*, t_0]} \tilde{x}(s) \right) \exp\{-\tilde{\mu}(t - t_0)\}, t > t_0. \tag{11}$$

Lemma 3 (see [33]). *Let $w(\cdot)$: $[t_0 - \overline{\tau}, +\infty) \longrightarrow [0, +\infty)$ be a continuous function such that $\dot{w}(t) \leq a(t) w(t) + b(t) \overline{w}(t)$ holds for $t \geq t_0$, where $\overline{w}(t) = \sup_{-\overline{\tau} \leq k \leq 0} (w(t + k))$. If $b(t) > 0$ and $a(t) + b(t) \geq m^*$, we have the following equation:*

$$w(t) \leq \overline{w}(t_0) \exp\{m^* (t - t_0)\}, t \geq t_0. \tag{12}$$

It should be noted that realizing the asymptotic synchronization is generally difficult due to the use of intermittent control and the effect of time-varying delay. the intermittent control shows some difficulties to be handled, including the fact that the values of error state $e(t) = y(t) - x(t)$ increase on all rest intervals. The time-varying delay brings several uncertain factors in tending towards the process of asymptotic synchronization. However, these difficulties will be well dealt with by studying new analytical methods.

Before proceeding with our research, a lemma is given in the following, with which the difficulty induced by intermittent control is surmounted.

Lemma 4. *Assume that a function $\tilde{y}(t)$ is continuous and nonnegative when $t \in (-\tilde{\tau}, +\infty)$, and satisfies the following condition:*

$$\begin{cases} \dot{\tilde{y}}(t) \leq -a_1(t)\tilde{y}(t) + a_2(t) \sup_{0 < \tau_r(t) \leq \tilde{\tau}, 1 \leq r \leq p-1} \tilde{y}(t - \tau_r(t)), t_{2k} \leq t < t_{2k+1}, \\ \dot{\tilde{y}}(t) \leq a_3(t)\tilde{y}(t) + a_4(t) \sup_{0 < \tau_r(t) \leq \tilde{\tau}, 1 \leq r \leq p-1} \tilde{y}(t - \tau_r(t)), t_{2k+1} \leq t < t_{2k+2}, \end{cases} \tag{13}$$

where $k \in \mathbb{N}$ and $p > 1$ is a positive integer. Assume further that the following inequalities hold, i.e.,

$$\sup_{t\in\mathbb{R}}\{a_1(t)-a_2(t)\}=\lambda>0, a_3(t)+a_4(t)\geq\varepsilon>0,$$
$$\frac{\lambda(t_{2k+1}-t_{2k})}{\varepsilon(t_{2k+2}-t_{2k+1})}=\chi_k>1. \tag{14}$$

Then, $\widetilde{y}(t)\longrightarrow 0$ *as* $t\longrightarrow+\infty$.

Proof 1. Denote $\overline{y}=\sup_{s\in[t_0-\widetilde{\tau},t_0]}\widetilde{y}(s)$, then from Lemma 2, when $t\in[t_0,t_1)$, it is obtained from that

$$\begin{aligned}\widetilde{y}(t)&\leq\overline{y}\exp\{-\lambda(t-t_0)\},\\ \widetilde{y}(t_1)&\leq\overline{y}\exp\{-\lambda(t_1-t_0)\},\end{aligned} \tag{15}$$

where $t_0=0$.

From Lemma 3, when $t\in[t_1,t_2)$, the second inequality of (13) leads to the following inequality.

$$\widetilde{y}(t)\leq\widetilde{y}(t_1)\exp\{\varepsilon(t-t_1)\}, \tag{16}$$

which means that

$$\widetilde{y}(t_2)\leq\widetilde{y}(t_1)\exp\{\varepsilon(t_2-t_1)\}. \tag{17}$$

Similarly, when $t\in[t_2,t_3)$, one has

$$\begin{aligned}\widetilde{y}(t)&\leq\widetilde{y}(t_2)\exp\{-\lambda(t-t_2)\}\leq\overline{y}\exp\{-\lambda(t_1-t_0)\}\\ &\exp\{\varepsilon(t_2-t_1)\}\exp\{-\lambda(t-t_2)\},\end{aligned} \tag{18}$$

and

$$\begin{aligned}\widetilde{y}(t_3)&\leq\overline{y}\exp\{-\lambda(t_1-t_0)\}\exp\{\varepsilon(t_2-t_1)\}\\ &\exp\{-\lambda(t_3-t_2)\}=\overline{y}\exp\{-\lambda[(t_1-t_0)+(t_3-t_2)]\}\\ &\exp\{\varepsilon(t_2-t_1)\}.\end{aligned} \tag{19}$$

By induction, when $t\in[t_{2k},t_{2k+1}), k\in\mathbb{N}^+$, it follows that

$$\begin{aligned}\widetilde{y}(t)&\leq\overline{y}\exp\left\{-\lambda\sum_{m=-1}^{k-1}(t_{2m+1}-t_{2m})\right\}\\ &\exp\left\{\varepsilon\sum_{m=0}^{k}(t_{2m}-t_{2m-1})\right\}\exp\{-\lambda(t-t_{2k})\},\end{aligned} \tag{20}$$

where $t_{-1}=t_{-2}=0$.

When $t\in[t_{2k+1},t_{2k+2}), k\in\mathbb{N}^+$, it follows that

$$\begin{aligned}\widetilde{y}(t)&\leq\overline{y}\exp\left\{-\lambda\sum_{m=0}^{k}(t_{2m+1}-t_{2m})\right\}\\ &\exp\left\{\varepsilon\sum_{m=0}^{k}(t_{2m}-t_{2m-1})\right\}\exp\{\varepsilon(t-t_{2k+1})\}.\end{aligned} \tag{21}$$

Then, for any $k\geq1$, one has from

$$\begin{aligned}&\widetilde{y}(t)\leq\overline{y}\exp\left\{-\lambda\sum_{m=-1}^{k-1}(t_{2m+1}-t_{2m})\right\}\exp\left\{\varepsilon\sum_{m=0}^{k}(t_{2m}-t_{2m-1})\right\}\exp\{-\lambda(t-t_{2k})\}=\overline{y}\exp\{-\lambda(t-t_{2k})\}\exp\\ &\left\{-\lambda\sum_{m=0}^{k-1}(t_{2m+1}-t_{2m})+\varepsilon\sum_{m=0}^{k-1}(t_{2m+2}-t_{2m+1})\right\}=\overline{y}\exp\{-\lambda(t-t_{2k})\}\exp\left\{\int\sum_{m=0}^{k-1}(1-\chi_k)(t_{2m+2}-t_{2m+1})\right\}, t\in[t_{2k},t_{2k+1}),\\ &\widetilde{y}(t)\leq\overline{y}\exp\left\{-\lambda\sum_{m=0}^{k}(t_{2m+1}-t_{2m})\right\}\exp\left\{\varepsilon\sum_{m=0}^{k}(t_{2m}-t_{2m-1})\right\}\exp\{\varepsilon(t-t_{2k+1})\}\\ &\leq\overline{y}\exp\{-\lambda(t_{2k+1}-t_{2k})\}\exp\left\{-\lambda\sum_{m=0}^{k-1}(t_{2m+1}-t_{2m})\right\}\exp\left\{\varepsilon\sum_{m=0}^{k}(t_{2m}-t_{2m-1})+\varepsilon(t-t_{2k+1})\right\}\\ &\leq\overline{y}\exp\{-\lambda(t_{2k+1}-t_{2k})+\varepsilon(t_{2k+2}-t_{2k+1})\}\exp\left\{-\lambda\sum_{m=0}^{k-1}(t_{2m+1}-t_{2m})\leq\overline{y}\exp\{-\lambda(t_{2k+1}-t_{2k})\right.\\ &\left.+\varepsilon(t_{2k+2}-t_{2k+1})\}\exp\ +\varepsilon\sum_{m=0}^{k-1}(t_{2m+2}-t_{2m+1})\left\{-\lambda\sum_{m=0}^{k-1}(t_{2m+1}-t_{2m})\varepsilon\sum_{m=1}^{k}(t_{2m}-t_{2m-1})\right\}\right\}\\ &=\overline{y}\exp\left\{\varepsilon\sum_{m=0}^{k}(1-\chi_k)(t_{2m+2}-t_{2m+1})\right\}, t\in[t_{2k+1},t_{2k+2}).\end{aligned} \tag{22}$$

For any $t\geq t_2$, there is a $\overline{k}\in\mathbb{N}_+$ such that $t\in[t_{2\overline{k}},t_{2\overline{k}+1})$ or $t\in[t_{2\overline{k}+1},t_{2\overline{k}+2})$. When $t\longrightarrow+\infty$, it follows that $\sum_{m=0}^{\overline{k}-1}(t_{2m+2}-t_{2m+1})\longrightarrow+\infty$. Since $\chi_k>1$, one has $\exp\{\varepsilon\sum_{m=0}^{\overline{k}-1}(1-\chi_k)(t_{2m+2}-t_{2m+1})\}\longrightarrow0$ and $\exp\{\varepsilon\sum_{m=0}^{\overline{k}}(1-\chi_k)(t_{2m+2}-t_{2m+1})\}\longrightarrow0$. To sum up, $\widetilde{y}(t)\longrightarrow0$ as $t\longrightarrow+\infty$. The proof is completed. □

Remark 2. When the coefficients of differential inequalities (13) become constants (i.e., $a_i(t)=a_i, i=1,2,3, a_4(t)=a_2$) and the layers become one layer ($p=2$), this lemma degenerates into the one studied in [21]. Lemma 4 relaxes the limiting conditions of the coefficients in the inequalities and generalizes the differential form of the inequalities. Moreover, without the existence of intermittent, Lemma 4 degenerates into Lemma 2 in [32] and Lemma 3 in [33]. Thus, Lemma 4 is a more general form and can be applied to the case of both nonintermittent and intermittent.

Remark 3. Lemma 4 can be applied to asymptotic synchronization or asymptotic stabilization via aperiodically

intermittent control. Compared with the conditions of the designed aperiodically intermittent control in [21, 22], Lemma 4 relaxes some harsh conditions. For example, in [23], the lower bound of control interval (i.e., $\inf_{k\in\mathbb{N}}\{t_{2k+1}-t_{2k}\}=\theta$) and the upper bound of rest width (i.e., $\inf_{k\in\mathbb{N}}\{t_{2k+2}-t_{2k+1}\}=\omega$) have to be previously assumed and satisfied $\theta<\omega$, where θ and ω are two positive constants. It presents from the condition $(\lambda(t_{2k+1}-t_{2k})/\varepsilon(t_{2k+2}-t_{2k+1}))=\chi_k$ that the intervals of control and rest can be flexibly adjusted. That is, both the control interval and the rest interval can be arbitrarily large. Therefore, the aperiodically intermittent controller shows its superiority and operability in practice.

3. Synchronization of Multilayer Neural Networks with Time-Varying Delays via Aperiodically Intermittent Control

This section is aimed to investigate the asymptotic synchronization problem of multi-layer neural networks with time-varying delays via aperiodically intermittent control. With the help of designing strictly aperiodically intermittent control and applying Lemma 4, Theorem 1 is obtained to ensure that multilayer neural networks (4) asymptotically synchronize on (3).

Design a mode-dependent controller as follows:

$$u(t)=\begin{cases}-De(t), & t_{2k}\le t<t_{2k+1},\\ 0, & t_{2k+1}\le t<t_{2k+2},\end{cases}\tag{23}$$

where $e(t)=[e_1^T(t),\ldots,e_N^T(t)]^T$, $e_i(t)=y_i(t)-x_i(t)$ $(i=1,2,\ldots,n)$, $D=\mathrm{diag}(d_1,d_2,\ldots,d_{nN})$, and $d_i>0$ are the control gains.

Remark 4. To eliminate the effects of time delays, some complex terms, such as integrals and time delays, are added to the intermittent controller in [34]. However, in many cases, it is difficult to obtain detailed information of time delays; hence, intermittent controller with time delays is usually difficult to be implemented in practice. Therefore, an intermittent controller without any information of delays is designed, which makes the controller more practical than the state-feedback controller in [35].

For the above-designed controller, we will reveal how to design suitable control gains $d_i(i=1,2,\ldots,nN)$ such that the multilayer error systems (5) can achieve asymptotic synchronization. The main results are elaborated as follows.

Theorem 1. *Assume that the function $f(t)$ satisfies Assumption 1 and the function $g(t)$ satisfies the Lipschitz condition, there exist positive constants $\xi,\xi_1,\xi_2,l,b_1^*,b_{3*}$, $\varpi,\varsigma,d_i(i=1,2,\ldots,nN)$ such that*

(i) $(b_1^*/2)P-\xi I_{nN}+P\Delta+\xi_1(A_0\otimes I_n)^T P(A_0\otimes I_n)+(l^2/\xi_1)P+\sum_{r=1}^{p-1}\xi_2(A_r\otimes I_n)^T P(A_r\otimes I_n)-D+\exp\{b_1^*\tilde{\tau}\}/2(p-1)P<0$,

(ii) $-(1/2)b_{3*}P-\xi I_{nN}+P\Delta+\xi_1(A_0\otimes I_n)^T P(A_0\otimes I_n)+(l^2/\xi_1)I_{nN}+\sum_{r=1}^{p-1}\xi_2(A_r\otimes I_n)^T P(A_r\otimes I_n)+(\exp\{b_1^*\tilde{\tau}\}/2)(p-1)P<0$,

(iii) $b_2(t)=2(p-1)((l^2/\xi_2)-((1-\varpi)\exp\{b_1^*(\tilde{\tau}-\tau(t))\}/2))\ge 0$,

where $p>1$ is a positive integer, the matrixes $P,\Delta\in\mathbb{R}^{nN}$ are defined in Assumption 1. Assume further that the inequalities hold, i.e.,

$$\sup_{t\in\mathbb{R}}\{b_1(t)-b_2(t)\}=\mu>0, b_3(t)+b_4(t)\ge\nu>0,\qquad \frac{\mu(t_{2k+1}-t_{2k})}{\nu(t_{2k+2}-t_{2k+1})}=\chi_k>1.\tag{24}$$

Then, the multilayer error systems (5) are globally asymptotically synchronized.

Proof 2. Consider the following Lyapunov–Krasovskii functional:

$$V(t)=V_1(t)+V_2(t),\tag{25}$$

where $V_1(t)=1/2e^T(t)Pe(t)$, $V_2(t)=\exp\{b_1^*\tilde{\tau}\}/2\sum_{r=1}^{p-1}\int_{t-\tau_r(t)}^{t}\exp\{b_1^*(s-t)\}e^T(s)Pe(s)ds$.

Computing the derivative of $V(t)$, we have the following equation:

$$\dot{V}_1(t)=-b_1(t)V_1(t)+\frac{b_1(t)}{2}e^T(t)Pe(t)+e^T(t)P\dot{e}(t),\tag{26}$$

$$\begin{aligned}\dot{V}_2(t)&=-b_1^*V_2(t)+\frac{\exp\{b_1^*\tilde{\tau}\}}{2}\Big[(p-1)e^T(t)Pe(t)\\&\quad-\sum_{r=1}^{p-1}\exp\{-b_1^*\tau_r(t)\}\times(1-\tau_r(t))e^T\\&\quad(t-\tau_r(t))Pe(t-\tau_r(t))\Big]\\&\le-b_1(t)V_2(t)+\frac{\exp\{b_1^*\tilde{\tau}\}}{2}\Big[(p-1)e^T(t)Pe(t)\\&\quad-\sum_{r=1}^{p-1}\exp\{-b_1^*\tilde{\tau}\}(1-\varpi)\times e^T\\&\quad(t-\tau_r(t))Pe(t-\tau_r(t))\Big].\end{aligned}\tag{27}$$

From (26) and (27), along the trajectories of (5) with controllers (23), we deduce two cases as follows. □

Case 1. When $t_{2k}\le t<t_{2k+1}$, for $k\in\mathbb{N}$

$$\begin{aligned}\dot{V}(t)\le&-b_1(t)V(t)+\frac{b_1(t)}{2}e^T(t)Pe(t)+e^T\\&(t)P\{[f(y(t))-f(x(t))]+(A_0\otimes I_n)[g_0(y(t))\\&-g_0(x(t))]+\sum_{r=1}^{p-1}(A_r\otimes I_n)[g_r(y(t-\tau_r(t)))\\&-g_r(x(t-\tau_r(t)))]-De(t)\}\\&+\frac{\exp\{b_1^*\tilde{\tau}\}}{2}\Big[(p-1)e^T(t)Pe(t)-\sum_{r=1}^{p-1}\exp\{-b_1^*\tau_r(t)\}\\&(1-\varpi)e^T(t-\tau_r(t))Pe(t-\tau_r(t))\Big].\end{aligned}\tag{28}$$

Since the function $f(t)$ satisfies Assumption 1 and the function $g(t)$ satisfies the Lipschitz condition, and by Lemma 2, we get the following equation:

$$\begin{aligned}
&e^T(t)P[f(y(t))-f(x(t))-\Delta e(t)] \le -\xi e^T(t)e(t), e^T(t)\\
&P(A_0\otimes I_n)[g_0(y(t))-g_0(x(t))] \le \xi_1 e^T(t)(A_0\otimes I_n)^T\\
&P(A_0\otimes I_n)e(t)+\frac{l^2}{\xi_1}e^T(t)Pe(t),\\
&e^T(t)P(A_r\otimes I_n)[g_r(y(t-\tau_r(t)))-g_r(x(t-\tau_r(t)))]\\
&\le \xi_2 e^T(t)(A_r\otimes I_n)^T P(A_r\otimes I_n)e(t)\\
&+\frac{l^2}{\xi_2}e^T(t-\tau_r(t))Pe(t-\tau_r(t)),
\end{aligned} \tag{29}$$

where $l, \xi, \xi_1, \xi_2 > 0$ are positive constants.

Then, according to the conditions (i) and (iii), it follows from (28) that

$$\begin{aligned}
&\dot V(t) \le -b_1(t)V(t)+e^T(t)\left[\frac{b_1^*}{2}P-\xi I_{nN}+P\Delta\right.\\
&+\xi_1(A_0\otimes I_n)^T P(A_0\otimes I_n)+\frac{l^2}{\xi_1}P+\sum_{r=1}^{p-1}\xi_2(A_r\otimes I_n)^T\\
&\left.P(A_r\otimes I_n)-D+\frac{\exp\{b_1^*\tilde\tau\}}{2}(p-1)P\right]\\
&e(t)+\sum_{r=1}^{p-1}e^T(t-\tau_r(t))\left[\frac{l^2}{\xi_2}-\frac{\exp\{b_1^*\tilde\tau\}}{2}\exp\{-b_1^*\tau_r(t)\}\right.\\
&\left.(1-\varpi)\right]\times Pe(t-\tau_r(t)) \le -b_1(t)V(t)\\
&+b_2(t)\sup_{0<\tau_r(t)\le\tilde\tau,1\le r\le p-1}V(t-\tau_r(t)),
\end{aligned} \tag{30}$$

where $b_2(t) = 2(p-1)((l^2/\xi_2) - ((1-\varpi)\exp\{b_1^*(\tilde\tau - \tau(t))\}/2))$.

Case 2. when $t_{2k+1} \le t < t_{2k+2}, k \in \mathbb{N}$, using the conditions (ii) and (iii), we have the following equation:

$$\begin{aligned}
&\dot V(t) \le -b_1(t)V(t)+(b_1(t)+b_3(t))V(t)+e^T(t)\\
&\left[-\frac{1}{2}b_{3*}P-\xi I_{nN}+P\Delta+\xi_1(A_0\otimes I_n)^T P(A_0\otimes I_n)+\frac{l^2}{\xi_1}P\right.\\
&\left.+\sum_{r=1}^{p-1}\xi_2(A_r\otimes I_n)^T P(A_r\otimes I_n)+\frac{\exp(b_1^*\tilde\tau)}{2}(p-1)P\right]\\
&e(t)+\sum_{r=1}^{p-1}e^T(t-\tau_r(t))\left[\frac{l^2}{\xi_2}-\frac{\exp\{b_1^*\tilde\tau\}}{2}\right.\\
&\left.\times\exp\{-b_1^*\tau_r(t)\}(1-\varpi)+\varsigma\right]Pe(t-\tau_r(t))\\
&\le b_3(t)V(t)+b_4(t)\sup_{0<\tau_r(t)\le\tilde\tau,1\le r\le p-1}V(t-\tau_r(t)),
\end{aligned} \tag{31}$$

where $b_4(t) = (p-1)(b_2(t)+\varsigma)$.

Thus, we get the following equation:

$$\begin{cases}
\dot V(t) \le -b_1(t)V(t)+b_2(t)\sup_{0<\tau_r(t)\le\tilde\tau,1\le r\le p-1}V(t-\tau_r(t)), t_{2k}\le t<t_{2k+1},\\
\dot V(t) \le b_3(t)V(t)+b_4(t)\sup_{0<\tau_r(t)\le\tilde\tau,1\le r\le p-1}V(t-\tau_r(t)), t_{2k+1}\le t<t_{2k+2}.
\end{cases} \tag{32}$$

By Lemma 4 and condition (24), we obtain $V(t) \longrightarrow 0$ as $t \longrightarrow +\infty$. Therefore, $\lim_{t\longrightarrow+\infty}\|e(t)\| = 0$. That is, (4) is globally asymptotically synchronized with (3). The proof is completed.

Remark 5. From condition (24) one can see that, for fixed value $\nu(t_{2k+2} - t_{2k+1})$, the value of the control interval $t_{2k+1} - t_{2k}$ can be decreased when the value μ is large as long as $\chi_k > 1$ is satisfied. That is, by tuning the value of the control gain D such that the value of μ increase or decrease. It can be seen that the conditions of Theorem 1 fully reveal the constraint relationship of parameters D, μ and ν.

Remark 6. Due to the existence of time delay, it will affect the convergence of the error system. By utilizing the inequality (13) in Lemma 4, a new Lyapunov–Krasovskii functional is designed in the proof of Theorem 1, which makes the effect of time delays to be well handled. With the help of Lemma 4, sufficient criteria by linear programming methods for asymptotic synchronization of the drive-response networks (3) and (4) are derived and the conservativeness of the obtained results can be reduced greatly.

4. Numerical Simulations

In this section, an example is given to verify the effectiveness of the proposed results in this paper.

Example 1. We consider the following multilayer dynamic networks with 10 identical nodes:

$$\dot{x}(t) = f(x(t)) + (A_0 \otimes I_3) g_0(x(t)) + \sum_{r=1}^{2} (A_r \otimes I_3) g_r(x(t - \tau_r(t))), \tag{33}$$

where

$$f_i(x_i(t)) = \begin{pmatrix} -a & a & 0 \\ b & -1 & 0 \\ 0 & 0 & -c \end{pmatrix} \begin{pmatrix} x_{i1}(t) \\ x_{i2}(t) \\ x_{i3}(t) \end{pmatrix} + \begin{pmatrix} 0 \\ -x_{i1}(t) x_{i3}(t) \\ x_{i1}(t) x_{i2}(t) \end{pmatrix},$$

$$g_0(x(t)) = \sin(x(t)),$$

$$g_1(x(t - \tau_1(t))) = \sin(x(t - \tau_1(t))),$$

$$g_2(x(t - \tau_2(t))) = 0.3 x(t) \cos(x(t - \tau_2(t))),$$

$$A_0 = \begin{pmatrix} -4 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & -4 & 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -5 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & -6 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & -3 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 & -7 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & -3 & 0 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 & -6 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & 1 & -5 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & -3 \end{pmatrix},$$

$$A_1 = \begin{pmatrix} -7 & 1 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 1 \\ 1 & -5 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & -5 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -3 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & -6 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & -5 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & -4 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & -4 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & -3 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & -4 \end{pmatrix},$$

$$A_2 = \begin{pmatrix} -6 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 1 \\ 1 & -4 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & -6 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & -3 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & -4 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 & -7 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & -5 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 & 0 & -5 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & -4 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 & -4 \end{pmatrix}, \tag{34}$$

where the parameters are set as $a = 10$, $b = 30$, $c = 8/3$, the time-varying delays are selected as $\tau_1(t) = 0.08 \exp(t)/(1 + \exp(t))$ and $\tau_2(t) = 0.04 \exp(t)/(1 + \exp(t))$. Note that, $A_i (i = 0, 1, 2)$ are randomly generated. The control gains in the controllers (23) are selected as $d_i = 100$. We choose $P_i = diag\{0.5, 0.4, 0.2\}$ and $\Delta_i = \mathrm{diag}\{50, 50, 50\}$ as [34], it is easy to verify that satisfies Assumption 1 holds with the parameter $\xi = 43.48$. Moreover, the initial values of the multilayer systems are given as follows: $x(0) = (3 + i, 5 + 2i, 7 + 2i)^T$, $y(0) = (-2 + 7i, -5 + 6i, -7 + 8i)^T$ $(i = 1, \ldots, 10)$.

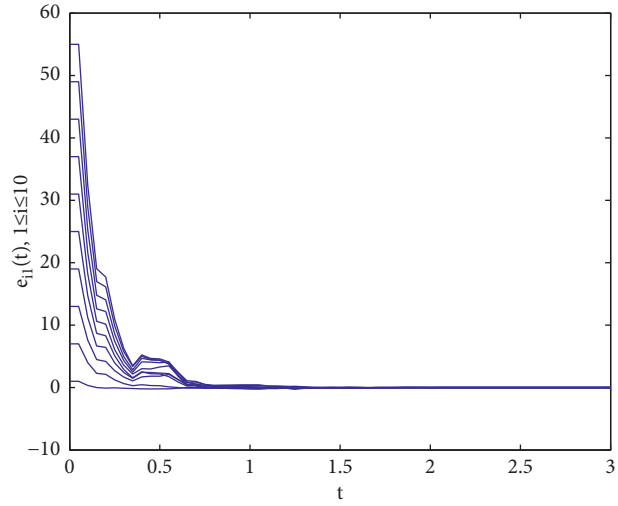


FIGURE 1: Trajectories of the synchronization errors e_{i1} roman for number with control gains $d_i = 100$.

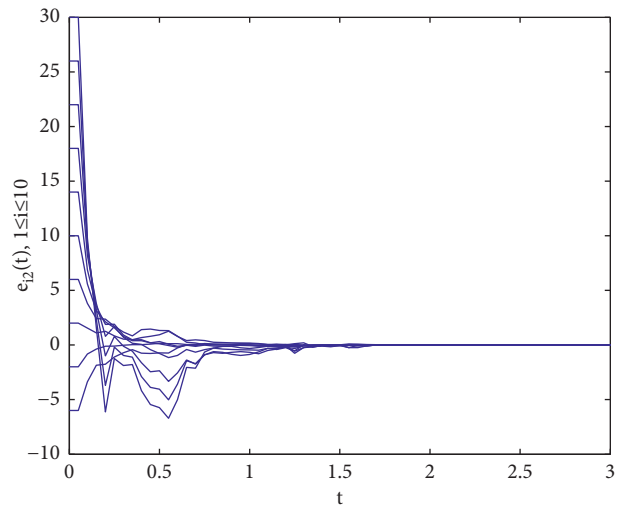


FIGURE 2: Trajectories of the synchronization errors e_{i2} with control gains $d_i = 100$.

As shown in Figure 1, the Lorenz system $\dot{x}_i(t) = f_i(x_i(t))$ has a chaotic attractor with the initial value $x(0) = (1, 2, -2)^T$. Each time span $t_{2k+2} - t_{2k} (k \geq 0)$ is randomly generated in the interval $[0.3s, 0.5s]$, and the ratio of the control time $t_{2k+1} - t_{2k}$ is randomly generated in the interval $[0.5s, 0.8s]$, the trajectories of the multilayer system errors with aperiodically intermittent controllers (23) are demonstrated in Figures 2–4.

Remark 7. To achieve the synchronization effect, we need to set appropriate control parameters $d_i (i = 1, 2, \ldots, 10)$ in the controllers (23), which are required to satisfy the conditions

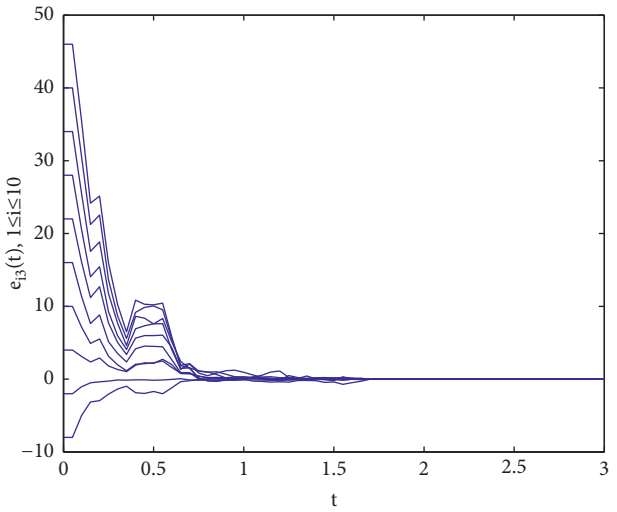


Figure 3: Trajectories of the synchronization errors e_{i3} with control gains $d_i = 100$.

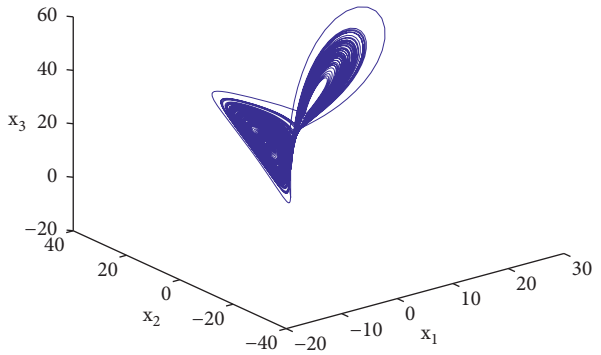


Figure 4: Chaotic attractor of the Lorenz system.

in Theorem 1. From these conditions, we can easily see that the upper and lower bounds of each continuous function $b_i(t)\,(i = 1, 2, 3, 4)$ exist, but the concrete values cannot be given. However, we give the sufficiently large parameters $d_i = 100\,(i = 1, 2, \ldots, 10)$ to satisfy condition 1 in Theorem 1.

5. Conclusions

In this article, asymptotic synchronization of multilayer neural networks with delays has been studied. By designing strict aperiodically intermittent controllers and establishing a set of novel Halanay-type inequalities, several criteria formulated by linear progressing methods are to ensure asymptotic synchronization. Our results can not only realize asymptotic synchronization of multilayer neural networks with delays by designing strict aperiodically intermittent controllers but also are less conservative in determining the control interval and designing the control gains. Numerical simulation verified the validity of the results obtained here.

At present, by using the aperiodically intermittent control with improved conditions $(\lambda(t_{2k+1} - t_{2k})/\varepsilon(t_{2k+2} - t_{2k+1}) = \chi_k > 1)$, many excellent research results on finite-time tracking of uncertain nonlinear systems [36] and finite-time synchronization of delayed neural networks have emerged [11, 19, 20, 24]. Note that, time delay or distributed delay, as one of the vital factors affecting the dynamic behaviors of neural networks, cannot be neglected [8]. In addition, the singularity [37, 38] and fractional-order [39] that constantly appear in practical engineering are the focus of current research. For example, singularly perturbed complex networks with cyberattacks were been considered in [40] and fractional-order nonlinear systems were been discussed in [41]. Therefore, how to realize finite-time synchronization of singularly or fractional-order complex networks with delays or distributed delays is our further research interest, which is also challenging.

Data Availability

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by the Natural Science Foundation of Henan Province (Grant no. 222300420171), Key Research Funds for the Universities of Henan Province (Grant no. 21A120005), Fundamental Research Funds for the Universities of Henan Provience (Grant no. NSFRF220420), and "Qinglan" Project of Jiangsu Province (Grant no. 2022004).

References

[1] L. Shanmugam, P. Mani, R. Rajan, and Y. H. Joo, "Adaptive synchronization of reaction–diffusion neural networks and its application to secure communication," *IEEE Transactions on Cybernetics*, vol. 50, no. 3, pp. 911–922, 2020.

[2] W. He, T. Luo, Y. Tang, W. Du, Y. C. Tian, and F. Qian, "Secure communication based on quantized synchronization of chaotic neural networks under an event-triggered strategy," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3334–3345, 2020.

[3] M. Abeles, Y. Prut, H. Bergman, and E. Vaadia, "Synchronization in neuronal transmission and its importance for information processing," *Progress in Brain Research*, vol. 102, pp. 395–404, 1994.

[4] G. Korats, S. Le Cam, R. Ranta, and V. Louis-Dorr, "A space-time-frequency dictionary for sparse cortical source localization," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 9, pp. 1966–1973, 2016.

[5] D. Wang, N. Zhang, J. L. Wang, and W. Wang, "Cooperative containment control of multiagent systems based on follower observers with time delay," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, pp. 1–11, 2016.

[6] D. Wang, N. Zhang, J. L. Wang, and W. Wang, "A PD-like protocol with a time delay to average consensus control for multi-agent systems under an arbitrarily fast switching topology," *IEEE Transactions on Cybernetics*, vol. 47, no. 4, pp. 898–907, 2017.

[7] X. S. Yang, X. X. Wan, C. Zunshui, J. D. Cao, Y. Liu, and L. Rutkowski, "Synchronization of switched discrete-time neural networks via quantized output control with actuator fault," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 9, pp. 4191–4201, 2021.

[8] X. S. Yang, Y. M. Feng, K. F. C. Yiu, Q. Song, and F. E. Alsaadi, "Synchronization of coupled neural networks with infinite-time distributed delays via quantized intermittent pinning control," *Nonlinear Dynamics*, vol. 94, no. 3, pp. 2289–2303, 2018.

[9] T. Y. Jing, F. Q. Chen, and X. H. Zhang, "Finite-time lag synchronization of time-varying delayed complex networks via periodically intermittent control and sliding mode control," *Neurocomputing*, vol. 199, pp. 178–184, 2016.

[10] X. S. Yang, Y. Liu, J. D. Cao, and L. Rutkowski, "Synchronization of coupled time-delay neural networks with mode-dependent average dwell time switching," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 12, pp. 5483–5496, 2020.

[11] Y. M. Feng, X. S. Yang, Q. Song, and J. D. Cao, "Synchronization of memristive neural networks with mixed delays via quantized intermittent control," *Applied Mathematics and Computation*, vol. 339, pp. 874–887, 2018.

[12] P. H. Nardelli, N. Rubido, C. Wang et al., "Models for the modern power grid," *The European Physical Journal - Special Topics*, vol. 223, no. 12, pp. 2423–2437, 2014.

[13] D. Yuan, H. Cui, J. Tian, Y. Xiao, and Y. Zhang, "Dynamics in the Kuramoto model with a bi-harmonic coupling function," *Communications in Nonlinear Science and Numerical Simulation*, vol. 38, pp. 23–29, 2016.

[14] P. Seliger, S. C. Young, and L. S. Tsimring, "Plasticity and learning in a network of coupled phase oscillators," *Physical Review*, vol. 65, no. 4, Article ID 041906, 2002.

[15] Y. H. Wen, M. R. E. Lamont, S. H. Strogatz, and A. L. Gaeta, "Self-organization in kerr-cavity-soliton formation in parametric frequency combs," *Physical Review A*, vol. 94, no. 6, Article ID 063843, 2016.

[16] J. Cheng, L. D. Liang, J. H. Park, H. C. Yan, and K. Z. Li, "A dynamic event-triggered approach to state estimation for switched memristive neural networks with nonhomogeneous sojourn probabilities," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 68, no. 12, pp. 4924–4934, 2021.

[17] Y. B. Wu, C. D. Wang, and W. X. Li, "Generalized quantized intermittent control with adaptive strategy on finite-time synchronization of delayed coupled systems and applications," *Nonlinear Dynamics*, vol. 95, no. 2, pp. 1361–1377, 2019.

[18] J. Feng, P. Yang, and Y. Zhao, "Cluster synchronization for nonlinearly time-varying delayed coupling complex networks with stochastic perturbation via periodically intermittent pinning control," *Applied Mathematics and Computation*, vol. 291, pp. 52–68, 2016.

[19] C. Xu, X. S. Yang, J. Q. Lu, J. W. Feng, F. E. Alsaadi, and T. Hayat, "Finite-time synchronization of networks via quantized intermittent pinning control," *IEEE Transactions on Cybernetics*, vol. 48, no. 10, pp. 3021–3027, 2018.

[20] X. S. Yang and J. D. Cao, "Stochastic synchronization of coupled neural networks with intermittent control," *Physics Letters A*, vol. 373, no. 36, pp. 3259–3272, 2009.

[21] X. W. Liu and T. P. Chen, "Synchronization of linearly coupled networks with delays via aperiodically intermittent pinning control," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 10, pp. 2396–2407, 2015.

[22] M. Liu, H. J. Jiang, and C. Hu, "Finite-time synchronization of delayed dynamical networks via aperiodically intermittent control," *Journal of the Franklin Institute*, vol. 354, no. 13, pp. 5374–5397, 2017.

[23] T. Y. Jing, D. Y. Zhang, and T. L. Jing, "Finite-time synchronization of hybrid-coupled delayed dynamic networks via aperiodically intermittent control," *Neural Processing Letters*, vol. 52, no. 1, pp. 291–311, 2020.

[24] R. Q. Tang, H. S. Su, Y. Zou, and X. S. Yang, "Finite-time synchronization of Markovian coupled neural networks with delays via intermittent quantized control: linear programming approach," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–11, 2021.

[25] A. Halanay, *Differential Equations: Stability, Oscillations, Time Lags*, Academic Press, New York, 1966.

[26] R. J. Yang, B. Wu, and Y. Liu, "A halanay-type inequality approach to the stability analysis of discrete-time neural networks with delays," *Applied Mathematics and Computation*, vol. 265, pp. 696–707, 2015.

[27] E. Liz, A. Ivanov, and J. B. Ferreiro, "Discrete Halanay-type inequalities and applications," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 55, no. 6, pp. 669–678, 2003.

[28] M. Y. Ge, Y. Jia, J. B. Kirunda et al., "Propagation of firing rate by synchronization in a feed-forward multilayer hindmarsh-rose neural network," *Neurocomputing*, vol. 320, pp. 60–68, 2018.

[29] J. L. Wang, Z. Qin, H. N. Wu, T. W. Huang, and P. C. Wei, "Analysis and pinning control for output synchronization and output synchronization of multiweighted complex networks," *IEEE Transactions on Cybernetics*, vol. 4, pp. 1314–1326, 2019.

[30] F. Yang, J. Mei, and Z. Wu, "Finite-time synchronisation of neural networks with discrete and distributed delays via periodically intermittent memory feedback control," *IET Control Theory & Applications*, vol. 10, no. 14, pp. 1630–1640, 2016.

[31] S. Boyd, L. Ghaoui, E. E. I. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, SIAM, Philadephia, 1994.

[32] S. Mohamad and K. Gopalsamy, "Continuous and discrete Halanay-type inequalities," *Bulletin of the Australian Mathematical Society*, vol. 61, no. 3, pp. 371–385, 2000.

[33] B. Liu, W. L. Lu, and T. P. Chen, "New criterion of asymptotic stability for delay systems with time-varying structures and delays," *Neural Networks*, vol. 54, pp. 103–111, 2014.

[34] D. Y. Zhang, Y. J. Shen, and J. Mei, "Finite-time synchronization of multi-layer nonlinear coupled complex networks via intermittent feedback control," *Neurocomputing*, vol. 225, pp. 129–138, 2017.

[35] F. F. Li, J. N. Li, and L. J. Shen, "State feedback controller design for the synchronization of Boolean networks with time delays," *Physica A: Statistical Mechanics and Its Applications*, vol. 490, pp. 1267–1276, 2018.

[36] G. M. Xue, F. N. Lin, S. G. Li, and H. Liu, "Adaptive dynamic surface control for finite-time tracking of uncertain nonlinear systems with dead-zone inputs and actuator faults," *International Journal of Control, Automation and Systems*, vol. 19, no. 8, pp. 2797–2811, 2021.

[37] X. F. Han and Z. B. Cheng, "Positive periodic solutions to a second-order singular differential equation with indefinite weights," *Qualitative Theory of Dynamical Systems*, vol. 21, no. 2, p. 53, 2022.

[38] Z. B. Cheng and L. L. Gu, "Positive periodic solution to a second-order differential equation with attractive-repulsive

singularities," *Rocky Mountain Journal of Mathematics*, vol. 52, no. 1, pp. 77–85, 2022.

[39] S. M. Ha, L. Y. Chen, H. Liu, and S. Y. Zhang, "Command filtered adaptive fuzzy control of fractional-order nonlinear systems," *European Journal of Control*, vol. 63, pp. 48–60, 2022.

[40] J. Cheng, L. D. Liang, H. C. Yan, J. D. Cao, S. D. Tang, and K. B. Shi, "Proportional-integral observer-based state estimation for Markov memristive neural networks with sensor saturations," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–12, 2022.

[41] Y. Xu, S. Gao, and W. X. Li, "Exponential stability of fractional-order complex multi-links networks with aperiodically intermittent control," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 9, pp. 4063–4074, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 9776776, 11 pages
https://doi.org/10.1155/2022/9776776

Research Article

Blockchain-Based Supply Chain System for Olive Fields Using WSNs

Oussama Ghorbel,[1] **Tarek Frikha**,[2] **Abir hajji**,[3] **Raed Alabdali**,[1] **Rami Ayadi**,[1] **and Mohamed Abbas Elmasry**[4]

[1]*Department of Computer Science, Jouf University, Sakakah, AlQurayat Region, Saudi Arabia*
[2]*CES-Lab, Sfax University, Sfax, Tunisia*
[3]*Gabes University, Gabes, Tunisia*
[4]*National Egyptian E-Leaning University (EELU), Giza, Egypt*

Correspondence should be addressed to Oussama Ghorbel; oaghorbel@ju.edu.sa and Rami Ayadi; rayadi@ju.edu.sa

Received 28 June 2022; Accepted 22 August 2022; Published 23 September 2022

Academic Editor: Heng Liu

The agricultural domain in developing countries is mostly dictated by archaic rules based on traditions and inherited practices. With the evolution of digitalization and technology, it seems essential to apply new technologies to the agricultural field. Among the technologies to be exploited in agriculture, we mention sensors, IoT, WSN, cloud, blockchain, etc. We talk about smart agriculture in this case. In this paper, we propose a platform secured by blockchain for monitoring and securing production. This platform uses IoT connected sensors to track and save data. Our system is used to monitor the production process of olive trees. The goal is to track everything that enters and leaves our olive tree production from fertilizers, insecticides, and fortifiers to olives, trimming etc. The blockchain via its decentralized system allow a secure, irreversible, and clear monitoring. A dashboard allow us to highlight the changes while facilitating the work of farmers. Our prototype will be embedded via a Raspberry Pi 4 platform.

1. Introduction

Using deep learning, machine learning, cloud, 4G, etc., it is becoming a common practice not only in technology fields such as IT, security, surveillance but also industry, transportation, e-health, smart cities, etc. Among the fields that need this evolution, we can also mention agriculture. The use of data and information has become increasingly crucial for the agricultural sector to improve productivity and sustainability. Internet of Things (IoT) technologies [1, 2] significantly increase the effectiveness and efficiency of data collection, storage, analysis, and use in agriculture. It allows farmers especially and the agricultural community more generally to easily obtain updated information and thus make better decisions in their daily farming.

For example, remote sensing data [3] on soil conditions can help farmers manage their crops, and the collected data can be accessed via the web or cell phones. This reduces the cost of information and thus facilitates farmers' access to markets and financial assistance. The development of the Global Positioning System (GPS) facilitates file mapping, machine guidance, and crop tracking. To enable the supply chain and minimize the risk of errors, fraud, and theft, blockchain technology is used. This technology is based on distributed registers [4, 5]. The use of blockchain is not only related to cryptocurrency (Bitcoin, Ethereum, etc.) but also other sectors that are starting to work on concrete use cases include industry via Industry 4.0 & 5.0 [6], e-health [7–9] such as medical records tracking and EHR [7, 10], paramedical and sports applications [2], and smart cities [11, 12].

In this paper, we propose an embedded agricultural system for monitoring an olive field based on IoT managed by a decentralized blockchain-based system guaranteeing the reliability of data provided by various sensors. We will highlight the practical case of smart and water efficient

irrigation system based on wireless sensor networks and secured by blockchain.

This paper is organized as follows: We present a state of the art with a definition of blockchain;

In part 2, we present our application scenario and system model; in part 3, we present our simulation; finally, we conclude this paper with a conclusion and future work.

2. Related Works

In this section, we discuss four classes of applications in the agricultural sector.

Agri-food production and supply chains [13–16] have been the subject of many studies. Indeed, using the technological evolution has allowed to fully improve the productivity. Li et al. [17] developed a dynamic planning method for the agri-food supply chain. As a result, they were able to maximize production, increase profiles, and decrease any waste.

Dey et al. [18] developed a Food-SQRBlock (Food Safety Quick Response Block) blockchain using blockchain and QR codes. They also proposed a large-scale cloud integration of the developed system to demonstrate the practicality and scalability of the framework along with supporting experimental evaluation.

While the majority of the work is used to track the supply chain, some have used the blockchain as an electronic money system and particularly as an electronic currency. In Ref. [19], Foroglou and Tsilidou have used the blockchain not only to implement a payment platform but also as a system for managing contracts, voting, property rights etc.

In Ref. [20], Tian proposed a theoretical real-time food traceability system using HACCP system, blockchain, and IoT. The authors claimed to be able to achieve transparency, reliability and security with the proposed model but did not provide an experimental implementation and evaluation of the system. Leng et al. [21] introduced a dual-chain based agricultural supply chain architecture using a public blockchain. They also studied storage mode, resource rent seeking, and consensus algorithms but did not assess the speed and skill of consensus algorithms considering the case of a large number of nodes and resources on the platform. In addition, access management for the user should be further investigated.

Surasak et al. [22] presented an IoT-based blockchain traceability system particularly designed for Thai agricultural products. Blockchain was used to create a distributed ledger to increase data integrity and transparency and a Structured Query Language (SQL) database was used to make the platform user-friendly. Further research on the integration of blockchain into the proposed system is needed to improve the efficiency of the system. Blockchain was used to create a distributed ledger to increase data integrity and transparency and a Structured Query Language (SQL) database was used to make the platform user-friendly. Further research on the integration of blockchain into the proposed system is needed to improve the efficiency of the system. Blockchain was used to create a distributed ledger to increase data integrity and transparency, and a Structured Query Language (SQL) database was used to make the platform user-friendly. Further research on the integration of blockchain into the proposed system is needed to improve the efficiency of the system.

Dakshayini and Prabhu. [23] proposed a blockchain, big data, and cloud-integrated crop monitoring system that attempts to realize effective demand-based decision support and achieve a simple, verifiable, and efficient system. The authors also proposed a crop exchange platform to sell the agri-products at different stages. However, they did not implement the IoT-based blockchain architecture to collect the actual field parameter data and then model the Big Data model.

The use of blockchain is not only applicable to agriculture (vegetables, fruits, etc.) [24] but private blockchain has also been used to track and secure dairy products. Rouaghi uses Hyperledger Fabric blockchain for product tracking [25].

In this paper, we present a system allowing not only to make the supply chain but also to manipulate all the data related to the olive trees such as the inputs (plants, fertilizers, insecticides, water, irrigation etc) but also the harvest (olive, trimming, etc). In this paper, we will propose a platform that uses different sensors, wireless sensor networks, blockchain, cloud etc.

3. Materials and Methods

3.1. Blockchain Choice. The concept of blockchain can be defined as a decentralized and distributed ledger to store time stamped transactions between many computers in a peer-to-peer network [2]. Thus, any record involved cannot be altered retroactively. This allows blockchain users to audit and verify transactions independently and transparently. Thus, the blockchain consists of blocks, which are connected using cryptographic techniques [26]. Each block must have a hash code of the previous block; a timestamp is a set of confirmed transactions. Figure 1 illustrates the blockchain.

We can subdivide our blockchain into three types: public, private and permissioned.

3.1.1. Public Blockchain. Public blockchain is an open blockchain. Everyone can access the private blockchain. Bitcoin blockchain operated continuously since its inception. All operate with the support of its public participants [27]. Thus, Bitcoin is the quintessential example of a public blockchain. Anyone can join and leave at their own will.

The various blocks of transactions and the blockchain are public and observable even if the participants are anonymous.

3.1.2. Private Blockchain. Moving on to a private blockchain, access to the blockchain is restricted to selected participants such as participants within an organization.

This restriction helps to simplify normal operations such as block creation and contingency model.

3.1.3. Permissioned Blockchain. The third classification of blockchain is permissioned blockchain, also known as consortium blockchain. It is intended for a consortium of

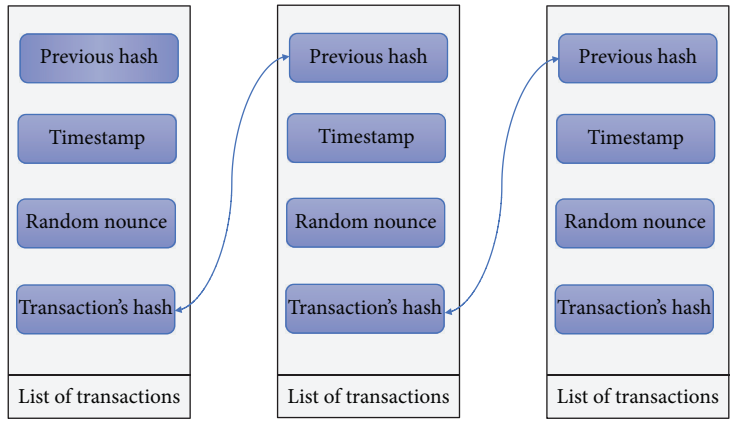


Figure 1: Blockchain illustration.

collaborating parties to conduct transactions on blockchain to facilitate governance, provenance, and accountability [28].

Examples include a consortium of all automotive companies, healthcare organizations, industry 4.0, smart agriculture, etc.

Permissioned blockchain has the advantages of a public blockchain by only allowing users with permission to collaborate and transact.

In this paper, we designed an integrated IoT-blockchain architecture, and we choose Ethereum as a type of blockchain for a decentralized agricultural system. We chose Ethereum because it is a permissioned blockchain that can be private or public. It can also execute smart contracts.

To create and maintain our Blockchain Ethereum network, Raspberry Pi 4 and 4 Go are used as it offers low-power consumption and simplicity of their interfaces.

The objective is to set up a platform to monitor the agricultural field, thus, appealing to the IoT on the basis of a blockchain platform. The reasons for setting up this platform are as follows:

(i) The importance of data confidentiality and the security aspect of agricultural information is justified by blockchain use.

(ii) The need for large volumes of data shows the need to use big data. PoW maintains the security layer of data.

(iii) The need to use data collected in real time from IoT sensors allows information to be updated in real time.

To improve the production result without making our solution complex and with highly consumption, we use the wireless sensor network which will be described in the next part. This WSNs field is considered the best solution to have data from different sensors to make our solution better.

4. Wireless Sensor Network

Wireless sensor network has been a growing interest in the scientific and industrial communities, thanks to innovations that have occurred during the last decades in the domains of microelectronics, MEMS (Microelectromechanical system) design, energy harvesting, and wireless technologies [29]. These wirelessly connected sensor networks consist of a large number of sensing nodes densely deployed in the wanted region. These sensing embedded elements are connected to each other through wireless links and work together to collect large amounts of high-fidelity information about different locations, processing them, and transmitting data to gateway nodes also known as sink points. Recent deployments have demonstrated their utility in various domains as described in Figure 2. WSNs are usually used in military operations [30]. Recently, a new set of possible applications has been an active subject of research, such as structural health monitoring [31–33], environmental monitoring [34, 35], agriculture [36], and industrial applications [37–39]. Recent experimentations are currently exploding in terms of usage and performances to improve the way of working in many contexts like automation smart cars to reduce the number of crashes and home automation. The quality of data collected by WSNs has been often unreliable and inaccurate because of the WSN's imperfect nature. Nonetheless, sensor nodes have stringent resource constraints such as memory capacity, computational complexity, and communication bandwidth, and energy consumption. These limited resources make the data generated by the sensor contaminated by noise, obvious error, missing data, duplicated values, and conflicting information. Furthermore, WSNs frequently utilize a large number of sensor nodes in harsh and hostile environments where sensor nodes are vulnerable to malicious attacks; hence, data generated and processed will be controlled by enemies.

4.1. Development of the Smart Water Irrigation System with Moisture Monitoring. In our work, we have developed a water-saving irrigation control system that is based on wireless sensor networks [40], whereby the system comprises low-power wireless sensor nodes that communicate through an adhoc ZigBee network. We will monitor soil moisture information parameters such as soil water degree, temperature, and relative humidity that can all be used to measure moisture potential. A four-channel temperature and humidity transmitter will be used to collect this data.

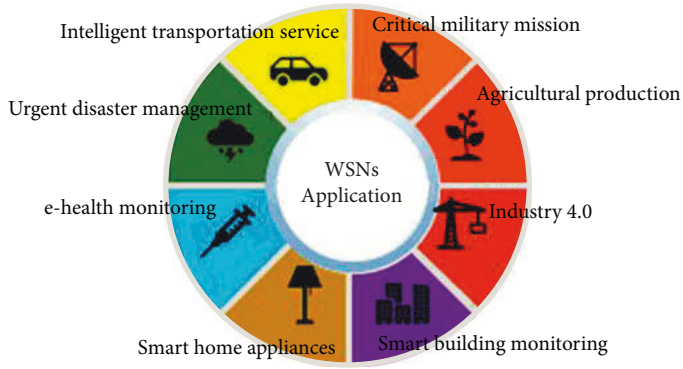


Figure 2: Application area of wireless sensor networks.

The information details are determined by using the various component of the sensor. In the following subsections, we will describe in detail the proposed design.

4.2. Smart Water Irrigation System for Farmers. As shown in Figure 3, our system of agriculture irrigation is based on famous technology entitled WSN is made up of four components: an irrigation controller, receiving sensors, a set of sensors, and a network of irrigation pipes. To construct an irrigation node group, sensor nodes that bear soil moisture are disseminated in accordance with the planting and irrigation status of farming. Each node is in charge of keeping track of the soil moisture in a certain area.

A standard WSN, utilizing ZigBee technology based on transmitted wireless data, consists of irrigation region and receiving sensors. Wireless multihop is used to transfer sensor data to the receiving node. Our discipline is designed to install a network for an irrigation pipe on the farming area in the irrigation region and an electrical control valve on the pipe to create an automated water-saving irrigation system. If the control of smart water irrigation is adaptable, the total system will be more versatile.

A water-saving irrigation system may be modified based on the original irrigation pipe network. For greater deployment of the irrigation system pipe network and to save money, an electronic control valve can be fitted. The irrigation controller in the WSN coverage region may spray irrigation in specified locations based on sensor data. This system contains a specific module taking in charge of network supervision. The proposed Smart Water Irrigation System is mainly dependent on wireless sensors networks and water pipeline.

4.3. System Hardware Structure. In this part, the hardware structure of the sensor node implemented in the proposed architecture is addressed and illustrated in Figure 4. The controller module, sensor module, ZigBee protocol communication module, and solar self-powered module make up the majority of the hardware structure.

The irrigation controller is built using an embedded system development board as the mainboard. The receiving node receives information through a serial connection and processes the control data. The system is very scalable. The WSNs measure of humidity is realized to be five times in

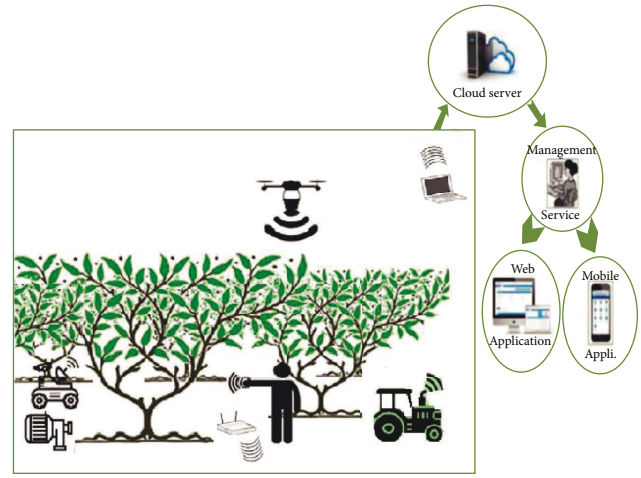


Figure 3: Smart water irrigation system based on wireless sensor networks.

minute (on a 12-second cycle) and send the data to the irrigation controller. When the irrigation controller detects that the humidity sensed by the WSN nodes in a specific location is lower than the prescribed value, it activates the irrigation network's electric control valve. The system will start irrigation and close the electric control valve of the pipe network in this region when the soil humidity in the area reaches a particular level.

4.4. Proposed Architecture. In this work, we propose the architecture as shown in Figure 5. Thus, the different actors of our agricultural system (farmers, vendors, distributors, etc.) are connected via a wireless sensor network. The data are saved on a database keeping the different traces of each transaction on blockchain. This traceability allows protecting not only all actors but also the plants of our farm.

It is important to point out that the data in the blockchain is encrypted using the Keccak 256 algorithm. This data can only be accessed using the public key of the sender and the private key of the receiver. The data are therefore encrypted and protected.

Each piece of data is encrypted using the receiver's public key and the sender's private key. Thus, only the person who has his private key can open a data encrypted with the public key.

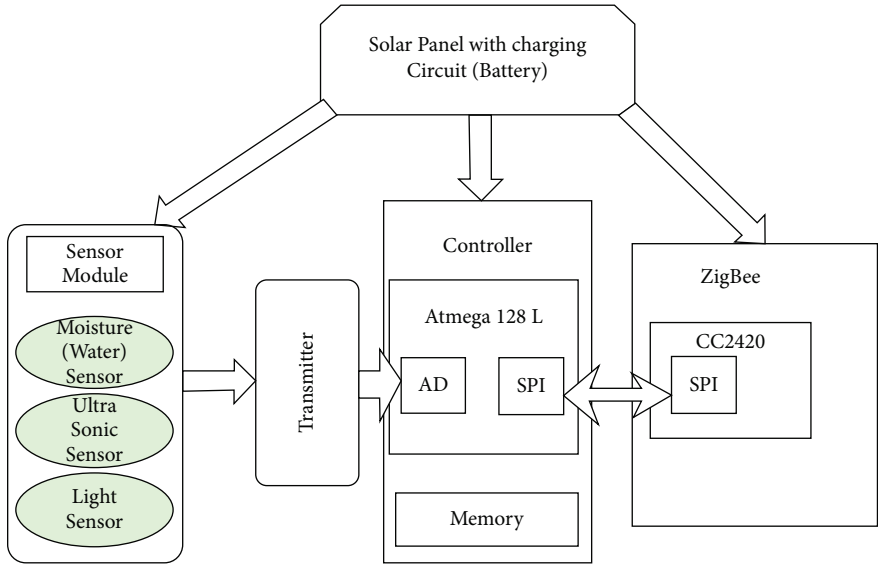


Figure 4: Hardware structure of the sensor node used in the proposed architecture.

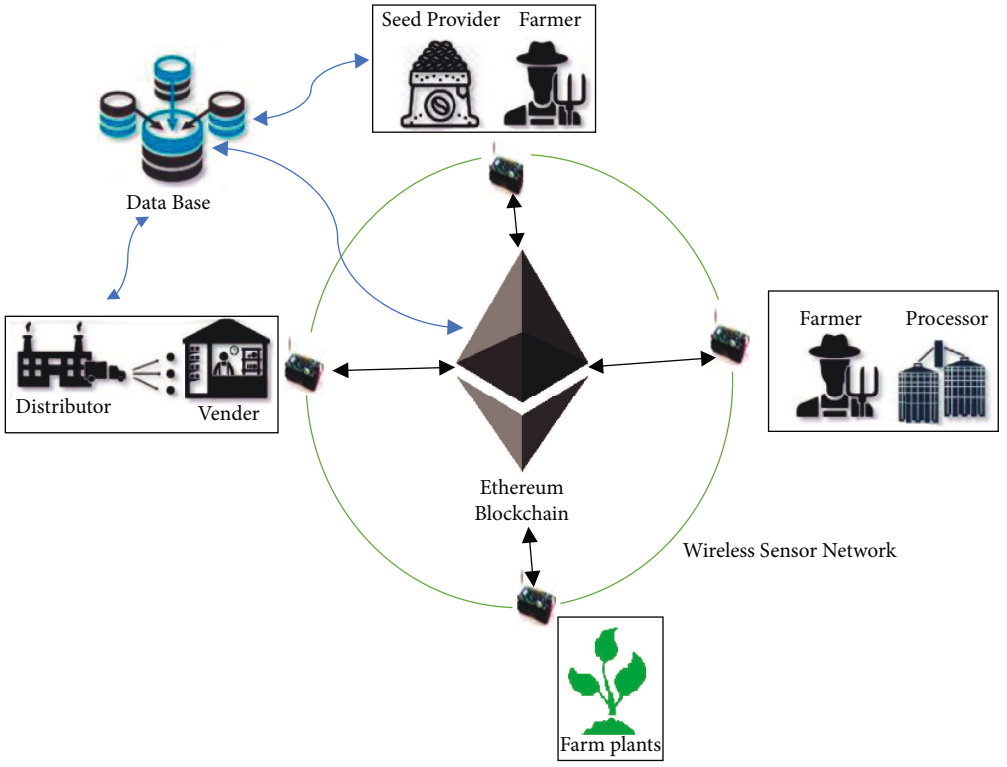


Figure 5: Proposed architecture.

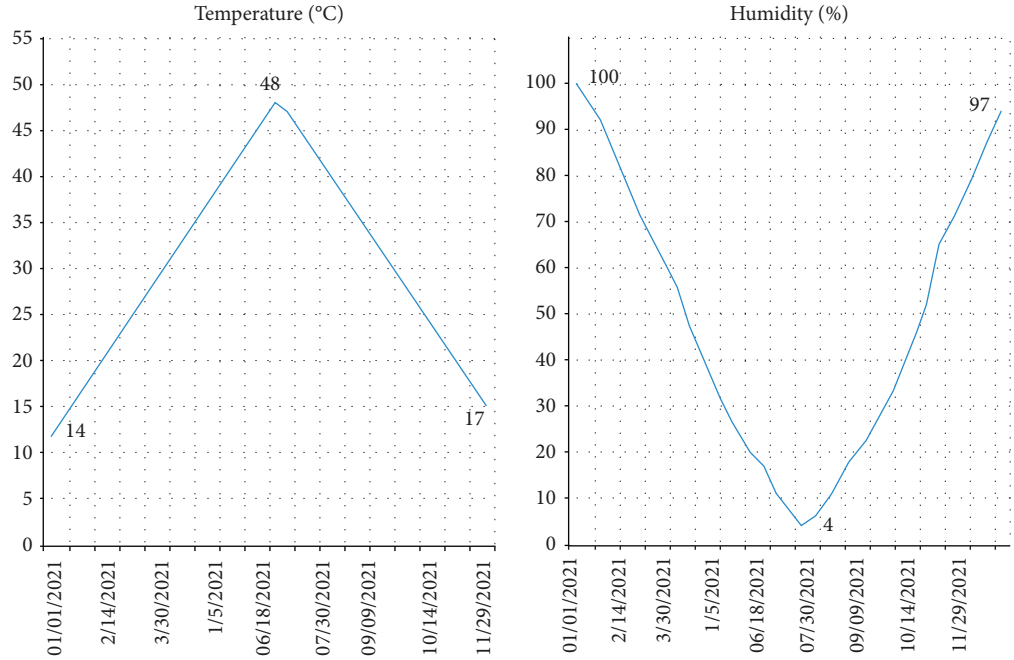


Figure 6: Temperature and humidity curves.

The system uses the PoW as a consensus. Since our blockchain does not use crypto-currencies and does not offer crypto-currencies as an offering for mining, it is therefore difficult to find malicious people who would like to decrypt the transactions or try to steal them. Thus, the security layer used in the blockchain allows having an efficient system for our farming application.

5. Obtained Results

In this section, we propose different results obtained after presenting the architecture, as shown in Figure 6. We present two different parts as follows:

(i) First, the results obtained by the different humidity, temperature, light, and wind sensors as well as the interfaces realized and connected to the blockchain

(ii) Second, a synthesis of the work on irrigation using wireless sensor networks

For example, the rise in temperature can also have contradictory consequences on plants. Thus, the increase in temperature decreases the yield. The use of IoT sensors (temperature, humidity, photoresistor, and wind) in the agricultural field which collects climatic data controls the field and helps producers make the right decisions to improve production.

The temperature and humidity results are shown in Figure 6, while the wind and light curves are shown in Figure 7. The results of our research show that the temperature varies between 0 and 65. Thus, relative humidity (RH) ranges from 0 to 100%. The air is dry when the relative humidity is below 35%. The air is moderately humid between 35% and 65%, and the air is humid at more than 65% relative humidity. Within the same space, the RH varies according to temperature changes; that is, it increases if the temperature drops and decreases if it rises.

For light, it varies between 0 and 100 lux. The light rate varies with temperature. It increases if the temperature increases and vice versa.

Wind speed affects plants. Frequent winds slow down plant growth and cause malformations: inclination of stems, nondevelopment of organs (leaves, buds, etc.).

We propose to break the supply chain management ecosystem using the Ethereum blockchain as follows: Farmers can control the field via a dashboard too and sell directly to consumers, so it can maintain the freshness of the food product, and its prices will be stable.

The data in the block of each transaction is divided into several parts of the block as follows:

(i) The system where the blockchain works is a server like a database.

(ii) Any valid transaction has been validated by a block regulated by the protocol.

(iii) Each blockchain block contains those as follows:

Data (timestamp and transaction information),

Hash (fingerprints of encrypted mathematical transactions),

Hash is the unique identifier of the block,

Hash of the previous block.

The Ethereum blockchain offers consumers who can see where the food product comes from, and then farmers can also view and verify the product until it is delivered to the customer based on location, farmer name, delivery date, number of purchases (kg), product type, and price. Here are the actors involved in the Ethereum blockchain.

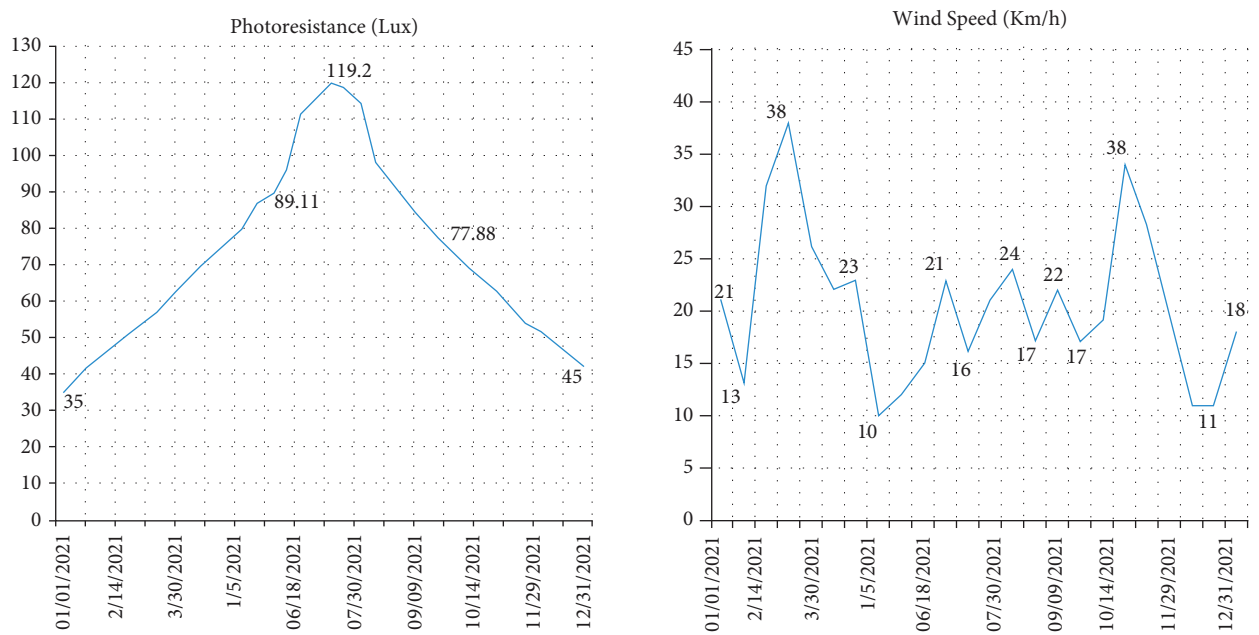


Figure 7: Light and wind curves.

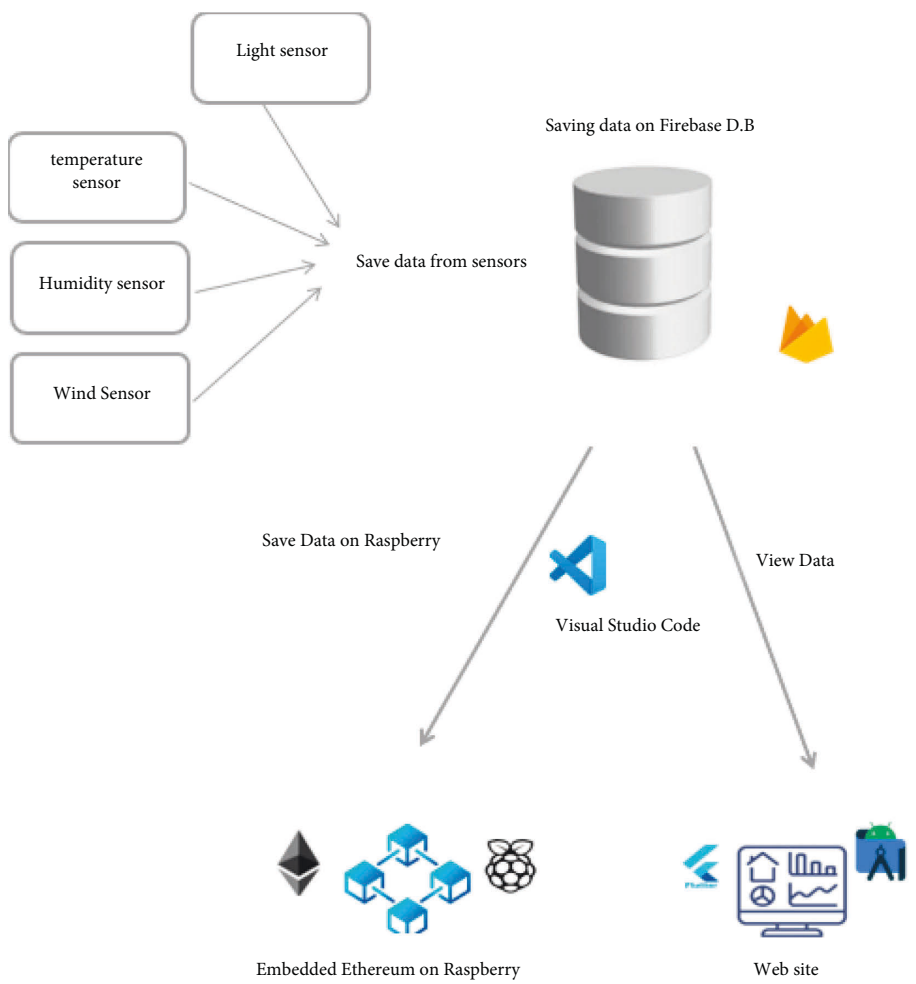


Figure 8: Final architecture of the agricultural application.

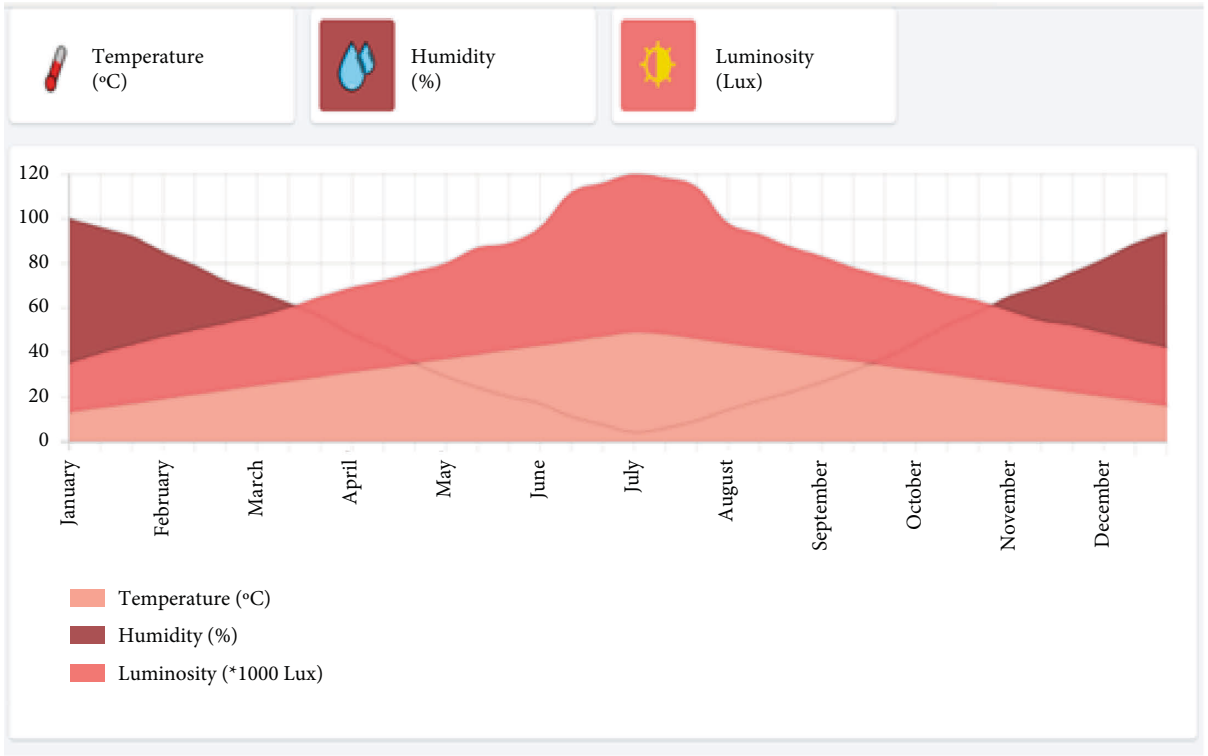


Figure 9: Dashboard.

5.1. Farmers. Those who control the field from the IoT sensors on-site produce the food product harvest and sell it in quantity (kilograms) to consumers.

5.2. Consumer. The party that buys product from farmers will be consumed or resold.

Here is a comparison between blockchain and the traditional database in all its aspects.

As described earlier, our application is presented in Figure 8 and is based on blockchain technology. We used Ethereum which was installed on the smart platform Raspberry Pi 4. The data captured by the three sensors must be recorded in Raspberry Pi 4 and in the blockchain. The result is displayed in a web page.

The developed application displays a web page identified by the web address 127.0.0.1, which presents a dashboard, as shown in Figure 9. This dashboard allows farmers to control and monitor the temperature values with their humidity and luminosity values for all months of a year. All collected data are stored in the database.

(i) The temperature is presented by the curve in light pink, where values are set between 13°C and 48°C.

(ii) The humidity is presented by the dark pink curve, where values vary from 4% to 100%.

(iii) The brightness is presented by the dark red color, where values vary from 35,000 Lux up to 120,000 Lux. Since their values are very large, we have divided the brightness values by 1000. In our dashboard, the brightness values are set between 35 Lux and 120 Lux.

6. Discussion

Many applications in various industries have used the new technology. As in our work, we focus on the agriculture field to compare related work. For example, Ref. [41] proposed a smart agriculture system, but some have used the sensors with fuzzy logic to present an intelligent sprinkler system, and others have based their approach only on sensors to monitor the condition of the paper. For precision agriculture, the authors in Ref. [40] used sensors and Raspberry Pi to assess soil fertility and productivity. This approach is different to our approach since we used Ethereum. In addition, in our work, we focus on collecting relevant data to control agricultural conditions mainly temperature, humidity, and light. In Ref. [42], authors added consensus mechanisms in electronic agriculture. In Ref. [43], the authors used smart contracts in the green IoT to create agreements. Green IoT helps realize the vision of a green ambient intelligence. There are several concepts of a sustainable green ecosystem with the use of blockchain such as green energy, green IT, and green finance. All can use smart contracts in a green ecosystem. It is an approach that is different from our approach, where they used smart contracts with IoT and our agricultural sensor-based approach with IoT. In addition, the authors in Ref. [44] used smart contracts for the hyperledger blockchain, where our approach is done on the Ethereum blockchain which can be private and/or public. The integration of IoT and blockchain in Ref. [45] offers more decentralization to make information and systems more reliable and stable. In our approach, we are interested in the IoT with a

Table 1: Comparison with related works.

Reference	Author	IoT	Fuzzy Logic	Sensors	Consensus mechanisms	Rasp. Pi	Smart contracts
[29]	A.H.Gilson (2020)	✗	✗	✓	✗	✓	✗
[30]	Y.Chen (2020)	✗	✗	✓	✓	✗	✗
[31]	M.S.Munir (2019)	✓	✓	✓	✗	✗	✗
[32]	A.N.Putri (2021)	✓	✗	✓	✗	✗	✗
[34]	P.S.Kumar (2020)	✓	✗	✗	✗	✗	✓
[35]	H.K.Ali (2020)	✗	✗	✗	✗	✗	✓
[36]	D.X.Li (2021)	✓	✗	✗	✗	✗	✓
[37]	T.Noshina (2019)	✓	✗	✗	✓	✗	✓
Our approach	(2022)	✗	✗	✓	✓	✓	✓

Blockchain is used in different application areas such as smart cities (smart parking [12, 48], water meter [49, 50], etc.).

decentralized system based on the blockchain, and we have added sensors for data collection. In Ref. [46], the blockchain is added to solve a lot of the problems, and above all, the trust and security. For our work, our goal is to improve trust between users without a third party and also to get a lot of information remotely without wasting time from sensors [47].

To summarize, our approach is based on a decentralized system with the use of the Ethereum blockchain in the agricultural sector. Various sensors were used to detect data and facilitate work for farmers. Table 1 below compares the aforementioned related works to our solution [51, 52].

7. Conclusions

The objective of this work focuses on the implementation of an intelligent agricultural framework managed by a decentralized system based on blockchain, as it is a new technology characterized by disintermediation in its use, traceability, and transparency of transactions, infeasibility, distribution, resiliency, security of all data, and greater trust. We have used Ethereum as a decentralized type of blockchain. This paper presents the installation of Ethereum on the Raspberry Pi 4; two nodes were created to have a connected and functional multinode Ethereum network. Sensors are used to obtain data for temperature, humidity, wind, and light; all data are stored in the database. Sensors play an essential role in agriculture. They are the key to collecting data more efficiently in order to make the most appropriate decisions.

The connection between the sensors is done using wireless sensor networks. One of the most challenging aspects of wireless sensor networks is their energy efficiency. That is why we propose an efficient way of utilizing the energy of wireless sensor networks for agriculture production. We are particularly interested in various monitoring techniques suitable our supply chain system for the olive field. To provide WSN with an information platform that can support a better role in agricultural production, the proposed model aims to improve on the shortcomings of existing WSN in the context of energy efficiency. So agriculture is a key factor impacting the global economy; therefore, WSN is important.

At the end, based on our solution, we conclude that our system has linked wireless sensor networks and their low-power system with blockchain technology that allows for secure data and tracking to avoid any attempt to steal or usurp data related to products or agriculture.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author(s) declare(s) that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was funded by the Deanship of Scientific Research at Jouf University under grant No (DSR-2021-02-0373)

References

[1] S. Ketu and P. K. Mishra, "A contemporary survey on IoT based smart cities: architecture, applications, and open issues," *Wireless Personal Communications*, vol. 125, no. 3, pp. 2319–2367, 2022.

[2] T. Frikha, A. Chaari, F. Chaabane, O. Cheikhrouhou, and A. Zaguia, "Healthcare and fitness data management using the IoT-based blockchain platform," *Journal of Healthcare Engineering*, pp. 1–12, 2021.

[3] A. Mukherjee, S. Misra, N. S. Raghuwanshi, and S. Mitra, "Blind entity identification for agricultural iot deployments," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3156–3163, 2019.

[4] M. Allouche, T. Frikha, M. Mitrea, G. Memmi, and F. Chaabane, "Lightweight blockchain processing. Case study: scanned document tracking on tezos blockchain," *Applied Sciences*, vol. 11, no. 15, p. 7169, 2021.

[5] T. Frikha, F. Chaabane, N. Aouinti, O. Cheikhrouhou, N. Ben Amor, and A. Kerrouche, "Implementation of blockchain consensus algorithm on embedded architecture," *Security and Communication Networks*, vol. 2021, Article ID 9918697, 11 pages, 2021.

[6] X. Xu, Y. Lu, B. Vogel-Heuser, and L. Wang, "Industry 4.0 and industry 5.0—inception, conception and perception," *Journal of Manufacturing Systems*, vol. 61, pp. 530–535, 2021.

[7] Y. Xiao, B. Xu, W. Jiang, and Y. Wu, "The HealthChain blockchain for electronic health records: development study,"

Journal of Medical Internet Research, vol. 23, no. 1, Article ID e13556, 2021.

[8] T. Frikha, N. Abdennour, F. Chaabane et al., "Source localization of EEG brainwaves activities via mother wavelets families for SWT decomposition," *Journal of Healthcare Engineering*, vol. 2021, pp. 1–11, Article ID 9938646, 2021.

[9] H. Choura, F. Chaabane, T. Frikha, and M. Baklouti, "Robust and secure watermarking technique for E-health applications," in *Proceedings of the 12th International Conference on Soft Computing and Pattern Recognition (SoCPaR 2020). SoCPaR 2020*vol. 1383, Cham, Switzerland, Springer, 2021.

[10] U. Chelladurai and S. Pandian, "A novel blockchain based electronic health record automation system for healthcare," *Journal of Ambient Intelligence and Humanized Computing*, vol. 13, no. 1, pp. 693–703, 2022, https://doi.org/10.1007/s12652-021-03163-3.

[11] P. Chinnasamy, C. Vinothini, S. Arun Kumar, A. Allwyn Sundarraj, S. V. Annlin Jeba, and V. Praveena, "Blockchain technology in smart-cities," in *Blockchain Technology: Applications and Challenges*, S. K. Panda, A. K. Jena, S. K. Swain, and S. C. Satapathy, Eds., vol. 203, Cham, Switzerland, Springer, 2021.

[12] M. Dhouioui and T. Frikha, "Design and implementation of a radar and camera-based obstacle classification system using machine-learning techniques," *J Real-Time Image Proc*, vol. 18, no. 6, pp. 2403–2415, 2021.

[13] G. Zhao, S. Liu, C. Lopez et al., "Blockchain technology in agri-food value chain management: a synthesis of applications, challenges and future research directions," *Computers in Industry*, vol. 109, pp. 83–99, 2019.

[14] F. Tian, "An agri-food supply chain traceability system for China based on RFID & blockchain technology," in *Proceedings of the 2016 13th International Conference on Service Systems and Service Management (ICSSSM)*, pp. 1–6, Kunming, China, June 2016.

[15] J. Astill, R. A. Dara, M. Campbell et al., "Transparency in food supply chains: a review of enabling technology solutions," *Trends in Food Science & Technology*, vol. 91, pp. 240–247, 2019.

[16] S. S. Kamble, A. Gunasekaran, and R. Sharma, "Modeling the blockchain enabled traceability in agriculture supply chain," *International Journal of Information Management*, vol. 52, Article ID 101967, 2020.

[17] D. Li, D. Kehoe, and P. Drake, "Dynamic planning with a wireless product identification technology in food supply chains," *International Journal of Advanced Manufacturing Technology*, vol. 30, no. 9-10, pp. 938–944, 2006.

[18] S. Dey, S. Saha, A. K. Singh, and K. McDonald-Maier, "FoodSQRBlock: digitizing food production and the supply chain with blockchain and QR code in the cloud," *Sustainability*, vol. 13, no. 6, p. 3486, 2021.

[19] G. Foroglou and A. L. Tsilidou, "Further applications of the blockchain," in *Proceedings of the 12th Student Conference on Managerial Science and Technology*, Athens, Greece, May 2015.

[20] F. Tian, "A supply chain traceability system for food safety based on HACCP, blockchain & Internet of things," in *Proceedings of the 2017 International Conference on Service Systems and Service Management*, Dalian, China, June 2017.

[21] K. Leng, Y. Bi, L. Jing, H. C. Fu, and I. Van Nieuwenhuyse, "Research on agricultural supply chain system with double chain architecture based on blockchain technology," *Future Generation Computer Systems*, vol. 86, pp. 641–649, 2018.

[22] T. Surasak, N. Wattanavichean, C. Preuksakarn, and Huang, "SCH Thai agriculture products traceability system using blockchain and internet of things," *System*, vol. 14, p. 15, 2019.

[23] M. Dakshayini and Prabhu, "BB an effective big data and blockchain (BD-BC) based decision support model for sustainable agriculture system," in *Proceedings of the EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing*Springer, Berlin, Germany, December 2020.

[24] I. Ehsan, M. Irfan Khalid, L. Ricci et al., "A conceptual model for blockchain-based agriculture food supply chain system," *Scientific Programming*, vol. 2022, Article ID 7358354, 15 pages, 2022.

[25] M. H. Ronaghi, "A blockchain maturity model in agricultural supply chain," *Information Processing in Agriculture*, vol. 8, no. 3, pp. 398–408, 2021.

[26] A. I. Taloba, M. R. Riad, and T. H. A. Soliman, "Developing an efficient spectral clustering algorithm on large scale graphs in spark," in *Proceedings of the 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 292–298, IEEE, Cairo, Egypt, December 2017.

[27] O. Ghorbel, A. Ayadi, R. Ayadi, M. Aseeri, and M. Abid, "Combined methods based outlier detection for water pipeline in wireless sensor networks," *AINA*, pp. 214–224, Springer, Berlin, Germany, 2019.

[28] A. I. Taloba and S. I. I. Safaa, "An intelligent hybrid technique of decision tree and genetic algorithm for e-mail spam detection," in *Proceedings of the 2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 99–104, IEEE, Cairo, Egypt, December 2019.

[29] "Missing reference,".

[30] M. P. Durisic, Z. Tafa, G. Dimic, and V. Milutinovic, "A survey of military applications of wireless sensor networks," in *Proceedings of the 2012 Mediterranean Conference on Embed-Ded Computing*, pp. 196–199, MECO, Bar, Montenegro, June 2012.

[31] J. J. Hao, "Application of wireless sensor network in modern agriculture in Guanzhong region," *Information and Computer*, vol. 32, no. 19, pp. 128-129, 2020.

[32] A. Ayadi, O. Ghorbel, A. Obeid, M. S. Ben Saleh, and M. Abid, "Leak detection in water pipeline by means of pressure measurements for WSN," in *Proceedings of the International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 1–6, IEEE, Fez, Morocco, May 2017.

[33] J. Ktari, T. Frikha, N. Ben Amor, L. Louraidh, H. Elmannai, and M. Hamdi, "IoMT-based platform for E-health monitoring based on the blockchain," *Electronics*, vol. 11, no. 15, p. 2314, 2022.

[34] P. Kułakowski, E. Calle, and J. L. Marzo, "Performance study of wireless sensor and actuator networks in forest fire scenarios," *International Journal of Communication Systems*, vol. 26, no. 4, pp. 515–529, Avr, 2013.

[35] M. Chen, X. Wang, and H. Zhang, "Wireless visual intelligent agricultural management system based on WSNs," *Agricultural Mechanization Research*, vol. 10, no. 7, 2021.

[36] A. Ayadi, O. Ghorbel, M. BenSalah, M. Abid, and M. Abid, "A framework of monitoring water pipeline techniques based on sensors technologies," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 2, pp. 47–57, 2022.

[37] A. Ayadi, O. Ghorbel, M. BenSalah, and M. Abid, "Spatio-temporal correlations for damages identification and localization in water pipeline systems based on WSNs," *Computer Networks*, vol. 171, no. 1, pp. 107134–134, 2020.

[38] O. Ghorbel, R. Ayadi, A. S. Alanazi, R. Alanazi, R. N. Alabdali, and M. Abid, "Adaptive KPCA for outlier detection in wireless sensor networks: water pipeline case," in *Proceedings of the ISNCC*, pp. 1–5, Dubai, UAE, November 2021.

[39] T. Frikha, N. Ben Amor, J. P. Diguet, and M. Abid, "A novel Xilinx-based architecture for 3D-graphics," *Multimedia Tools and Applications*, vol. 78, no. 11, pp. 14947–14970, 2019, https://doi.org/10.1007/s11042-018-6886-4.

[40] A. Manas and Y. Bosc-Haddad, "La (ou les) blockchain (s), une réponse technologique à la crise de confiance," *Ann. Mines - Réal. Ind. Août*, vol. 3, p. 102, 2017.

[41] F. Leal, A. E. Chis, S. Caton et al., "Smart pharmaceutical manufacturing: ensuring end-to-end traceability and data integrity in medicine production," *Big Data Research*, vol. 24, Article ID 100172, May 2021.

[42] D. Minoli and B. Occhiogrosso, "Blockchain mechanisms for IoT security," *Internet of Things*, vol. 1–2, pp. 1–13, 2018.

[43] M. S. Munir, I. S. Bajwa, and S. M. Cheema, "An intelligent and secure smart watering system using fuzzy logic and blockchain," *Computers & Electrical Engineering*, vol. 77, pp. 109–119, 2019.

[44] M. Pournader, Y. Shi, S. Seuring, and S. L. Koh, "Blockchain applications in supply chains, transport and logistics: a systematic review of the literature," *International Journal of Production Research*, vol. 58, no. 7, pp. 2063–2081, 2020.

[45] A. N. Putri, M. Hariadi, and A. D. Wibawa, "Smart agriculture using supply chain management based on hyperledger blockchain," *IOP Conference Series: Earth and Environmental Science*, vol. 466, no. 1, Article ID 12007, 2020.

[46] P. Schallum and I. Olson, "Les systèmes de paiement mobile à l'ère de la Covid-19: sécurité, vie privée et confiance numérique," *Technologie et innovation*, vol. 6, no. 1, 2021.

[47] M. Elloumi, M. A. Ahmad, A. H. Samak, A. M. Al-Sharafi, D. Kihara, and A. I. Taloba, "Error correction algorithms in non-null aspheric testing next generation sequencing data," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9819–9829, 2022.

[48] T. Frikha, H. Choura, N. Abdennour, O. Ghorbel, and M. Abid, "ESP2: embedded smart parking prototype," *Advances in Science, Technology and Engineering Systems Journal*, vol. 5, no. 6, pp. 1569–1576, 2020.

[49] J. Ktari, T. Frikha, M. Hamdi, H. Elmannai, and H. Hmam, "Lightweight AI framework for industry 4.0 case study: water meter recognition," *Big Data and Cognitive Computing*, vol. 6, no. 3, p. 72, 2022.

[50] J. Ktari, T. Frikha, M. A. Yousfi, M. K. Belghith, and N. Sanei, "Embedded Keccak implementation on FPGA," in *Proceedings of the 2022 IEEE International Conference on Design & Test of Integrated Micro & Nano-Systems (DTS)*, pp. 1–5, Cairo, Egypt, June 2022.

[51] G. Xue, F. Lin, S. Li, and H. Liu, "Adaptive dynamic surface control for finite-time tracking of uncertain nonlinear systems with dead-zone inputs and actuator faults," *International Journal of Control, Automation and Systems*, vol. 19, no. 8, pp. 2797–2811, 2021.

[52] S. Ha, L. Chen, H. Liu, and S. Zhang, "Command filtered adaptive fuzzy control of fractional-order nonlinear systems," *European Journal of Control*, vol. 63, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 5903514, 12 pages
https://doi.org/10.1155/2022/5903514

Research Article

Spatio-Frequency Decoupled Weak-Supervision for Face Reconstruction

Yanyan Li, Weilong Peng, Keke Tang, and Meie Fang

Guangzhou University, Guangzhou, China

Correspondence should be addressed to Weilong Peng; wlpeng@gzhu.edu.cn and Meie Fang; fme@gzhu.edu.cn

Received 13 May 2022; Accepted 12 July 2022; Published 22 September 2022

Academic Editor: Heng Liu

3D face reconstruction has witnessed considerable progress in recovering 3D face shapes and textures from in-the-wild images. However, due to a lack of texture detail information, the reconstructed shape and texture based on deep learning could not be used to re-render a photorealistic facial image since it does not work in harmony with weak supervision only from the spatial domain. In the paper, we propose a method of spatio-frequency decoupled weak-supervision for face reconstruction, which applies the losses from not only the spatial domain but also the frequency domain to learn the reconstruction process that approaches photorealistic effect based on the output shape and texture. In detail, the spatial domain losses cover image-level and perceptual-level supervision. Moreover, the frequency domain information is separated from the input and rendered images, respectively, and is then used to build the frequency-based loss. In particular, we devise a spectrum-wise weighted Wing loss to implement balanced attention on different spectrums. Through the spatio-frequency decoupled weak-supervision, the reconstruction process can be learned in harmony and generate detailed texture and high-quality shape only with labels of landmarks. The experiments on several benchmarks show that our method can generate high-quality results and outperform state-of-the-art methods in qualitative and quantitative comparisons.

1. Introduction

3D face reconstruction, which aims to recover 3D face shapes from a single image or multiple-view images, has been widely applied to face recognition [1], face animation [2], and artistic editing [3]. Traditional methods involve complex and costly optimization for accurate reconstruction [4–7]. Since deep learning has significant advantages of nonlinear fitting ability on complex tasks [8–12], there is an increasing interest in reconstructing 3D faces from a single image using deep convolutional neural networks [13–17]. However, the reconstruction accuracy is seriously affected by the challenging cases, e.g., various illumination poses, occlusions, etc.

Generally, deep learning-based methods could be roughly divided into families of supervised learning [16, 18, 19], unsupervised learning [20–22], and weakly supervised learning [13, 23]. For supervised learning, 3D ground-truth face data are needed as supervision information, but a large amount of label data are not easily accessible. For compromise, existing methods usually use 3DMM parameters [18] or traditional methods [19, 24] to synthesize 3D shapes as ground-truth face data, which limits the precision of reconstruction. Unsupervised and weakly-supervised learning overcome the weakness of relying on 3D ground-truth data and learning the reconstruction process based on image data with only labeled landmarks if necessary. Classically, based on the 3DMM model prior, Deng et al. [13] devised a robust loss function combining image-level and perception-level information as weakly supervised information to improve 3D face reconstruction. However, it could not handle the wrong texture when the face is occluded. Feng et al. [18] abandoned the 3DMM model and regressed the 3D shape from the network straightly, but their supervision data are still based on the 3DMM fitting, which has limitations.

In our opinion, a key reason for the lack of high reconstruction accuracy is that the commonly used CNNs

approach only considers spatial loss [13, 25, 26], e.g., landmark loss and pixel loss in the spatial domain, while ignoring the impact of frequency. Some studies have shown that DNNs tend to synthesize frequencies in order from low to high [27–29]. So it is hard to urge neural networks to learn the inconspicuous frequency of images and recover them with merely spatial loss [16] since spatial loss focuses on point-wise value and spatial associations but does not pay enough attention to harmony in the frequency domain [30].

Based on the abovementioned points, we proposed a spatio-frequency decoupled weak-supervision approach for 3D face reconstruction to address the unreality issue. We first use a convolutional neural network (ResNet-50) to regress 3DMM coefficients and render parameters. And then, we build the weakly supervision between the input and the re-rendered face image. Not limited to spatial domain loss covering image-level and perceptual-level loss, frequency spectrums are also separated from image pairs to measure the gap in the frequency domain based on differentiable discrete Fourier transformation. We devise the patch-level frequency loss based on spectrum-wise weighted Wing loss to capture further inconspicuous frequency affecting reality. In particular, the loss motivates the network to learn detailed textures and avoids the adverse effects of occlusion. Experiments show that our method can generate high-quality results and outperform several state-of-the-art methods in qualitative and quantitative comparisons on several benchmarks. To summarize, this paper makes the following contributions:

(i) We propose a spatio-frequency decoupled weak supervision method for 3D reconstruction with high-fidelity textures from a single in-the-wild image.

(ii) We propose a patch-based spectrum-wise weighted Wing loss in the frequency domain to improve the robustness of texture reconstruction to occlusion and the reality of the re-rendered image.

2. Related Work

2.1. 3D Face Detail Reconstruction. ***Geometry Reconstruction***. The 3DMM [31] makes it possible to recover the facial shape from a single image by regressing 3DMM face shape parameters. Some studies [14, 32] reconstructed a rough shape using the 3DMM in the first stage and then refined the shape by imposing some spatial domain constraints, e.g., asymmetric Euclidean loss [32] and identity consistency loss [14]. The other methods [26, 33] used a collaborative approach by employing a synergy process between 3DMM coefficients and 3D face landmarks [33] or an occlusion segmentation network [26]. These approaches narrow the error in the spatial domain to synthesize more realistic facial geometry with 3DMM. But 3DMM works well in the low-frequency domain, neglecting the critical frequency information that determines the realism. In contrast, we aim to capture the key frequency in the frequency domain.

3D Re-Renderable Modeling. 3D Re-renderable modeling makes the process of mapping a 3D face model to a 2D portrait image [21, 34–37]. These methods decompose a single face image into reflectance, geometry, and lighting and then render the face image by changing the lighting and fixing the geometry and reflectance [38]. Yamaguchi et al. [36] developed a deep learning method to estimate high-resolution facial reflectance and normal. However, they could not re-render the whole face image while leaving out the eye, teeth, and hair regions. Dib et al. [34] introduced ray tracing for face reconstruction within an optimization-based framework to make the re-rendered faces robust to lighting conditions. But the quality of their reconstruction is still influenced by the initialization landmarks. Yang et al. [37] proposed a novel, detailed illumination representation to disentangle facial texture and lighting, resulting in high-fidelity textures even with in-the-wild images. Their results are good but also decoupled in the spatial domain. Different from them, our method decouples illumination and albedo in the frequency domain to obtain an anti-occlusion, anti-illumination, and re-renderable face image.

2.2. Frequency Domain Studies of Neural Networks. Several studies [27, 28] have begun to use Fourier analysis to explore the neural network training process and found a learning bias of neural networks towards low-frequency components. Moreover, F-Principle [29] showed that the frequency fitting priority is different throughout the training process, usually from low to high. Therefore, when using CNN for reconstructing a 3D face shape, the network always avoids high-frequency components, which will cause the reconstructed 3D face to be too smooth, and some details cannot be reconstructed.

Recently, Jiang et al. [30] introduced frequency domain information into image synthesis to improve the effect of image synthesis by guiding the network to synthesize hard frequencies that are difficult to synthesize. Although the paper demonstrated the influence of frequency domain information on image synthesis, few studies have explored the effect of frequency in 3D face reconstruction. Wang et al. [39] are the first to introduce the concept of frequency domain into 3D face reconstruction. It enhances self-supervised learning by adding low-frequency albedo information to guide the network for generating intact albedos. However, the albedo model is still a linear subspace model that concentrates on low-frequency, failing to synthesize high-frequency information during training and address the frequency bias problem of DNN training. Our method aims to narrow the frequency gap during the training, i.e., by transforming the image from the spatial domain to the frequency domain based on a differentiable 2D Fourier transform and then reconstructing more detailed 3D faces and albedo.

2.3. Wing Loss. Wing loss is a supervised function for face landmark alignment proposed by Feng et al. [40]. After analyzing L1 loss, L2 loss, and smooth L1 loss function empirically and theoretically, they found that large errors easily dominate the step size of these loss functions so that some outliers may mislead the network during training. So

they proposed Wing loss to improve the resistance to large errors and the ability to amplify small and medium-scale errors during neural network training.

In 3D face reconstruction work, the importance of high-frequency and low-frequency components are different in an image, and then it is also different in the difficulty of fitting them via neural networks. In the early stage of training, the frequency gap between the input and the re-rendered image is large and becomes small in the middle and later stages of training. However, the error in pixel level may be large when occlusion occurs in the image, even though the frequency error can be very small. To narrow the gap further and improve the reconstructed face's accuracy, we use Wing loss to solve the problem. Inspired by the Wing loss's variant [41], we adjust the spectrum-wise weighting experimentally so that the differences could be suppressed even at the tiny frequency error by amplifying the spectrum-wise error. In this way, the effect brought by occlusion frequencies can be significantly alleviated.

3. Method

3.1. Preliminaries. Our approach regresses the shape and texture coefficients of the 3DMM model to reconstruct the 3D face shape, which is then rendered onto a 2D plane, using spatial and frequency domain information as weak supervision signals to assist the network training. We will introduce several foundation works involved in the procedure, including the 3DMM, illumination, and camera models.

Face prior model. 3DMM [31] is our face prior model for reconstructing face shape and texture based on principal component analysis (PCA). As the original 3DMM could not express facial expressions, we improved the 3DMM model by fusing the expression bases $\mathbf{A}_{\text{exp}}$ built from FaceWarehouse [42]. At last, the model is defined as:

$$\begin{aligned} \mathbf{S}(\boldsymbol{\alpha}, \boldsymbol{\beta}) &= \overline{\mathbf{S}} + \mathbf{A}_{\mathbf{id}}\boldsymbol{\alpha} + \mathbf{A}_{\mathbf{exp}}\boldsymbol{\beta}, \\ \mathbf{T}(\boldsymbol{\delta}) &= \overline{\mathbf{T}} + \mathbf{A}_{\mathbf{t}}\boldsymbol{\delta}, \end{aligned} \tag{1}$$

where $\overline{\mathbf{S}}$ and $\overline{\mathbf{T}}$ represent the mean shape and texture, $\mathbf{A}_{id}$ and $\mathbf{A}_{\mathbf{t}}$ are the PCA bases with a neutral expression. $\boldsymbol{\alpha} \in \mathbb{R}^{80}$, $\boldsymbol{\beta} \in \mathbb{R}^{64}$ and $\boldsymbol{\delta} \in \mathbb{R}^{80}$ are the shape, expression, and texture parameters to be regressed in our model.

Camera model. We use a perspective model as the camera model. It first converts any vertex $\mathbf{v}$ on $\mathbf{S}$ to a new position $\hat{\mathbf{v}}$ under the camera coordinate system with orthogonal rotation $\mathbf{R} \in SO(3)$ and translation $\mathbf{t} \in \mathbb{R}^3$. And then $\hat{\mathbf{v}}_i$ is projected to point $\mathbf{u}$ in an image plane. In particular, we set an empirical focal in the camera to display the 2D face. On the whole, there are six parameters in the perspective model.

Illumination model. Assuming the human face is a Lambert surface, we use the spherical harmonic (SH) function to represent scene illumination and then compute the radiosity of the vertex [43]. With the surface normal $\mathbf{n}_i$ at $\hat{\mathbf{v}}_i$, the radiosity I_i related to the pixel can be represented by the SH illumination model with three bands:

$$I_i(\mathbf{l}) = t_i \sum_{j=1}^{9} l_j H_j(\mathbf{n}_i), \tag{2}$$

where t_i is one channel of texture at $\mathbf{v}_i$ on $\mathbf{T}$, and $\mathbf{l}$ is channel-wise control coefficients, $\{H_j\}$ is orthogonal bases in spherical harmonic function. Generally, the SH model can accurately estimate the illumination information in different environments without estimating the direction of the light source, which greatly simplifies the illumination estimation.

Unsupervised learning reconstruction. Under an unsupervised schema, all the unknown parameters are predicted as $\boldsymbol{\Theta} \in \mathbb{R}^{257}$ that consists $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, $\mathbf{R}$, $\mathbf{t}$ and $\{\mathbf{l}_i\}_{i\in\{r,g,b\}}$ by a neural network for a given face image I, firstly. And then $\boldsymbol{\Theta}$ is applied to a differentiable image formation layer to generate a new rendered image I'. The shape and texture could be learned by supervising I' with input I:

$$\min \text{Dist}(I, I'(\boldsymbol{\Theta})), \text{with } \boldsymbol{\Theta} = \text{NeurlNet}(I). \tag{3}$$

Under the formulation, skin masks [44, 45], and weak supervision with landmarks [46, 47] could be introduced to learn high-quality face.

3.2. Framework of Spatio-Frequency Decoupled Weak-Supervision. We will introduce spatio-frequency decoupled weak-supervision into equation (3). In our framework, the learning process is applied with supervision in both spatial domain and frequency domain, seeing in Figure 1. Firstly, a convolutional neural network (ResNet-50) is used to regress the parameters of shape, texture, pose, and illumination from face image I. Then it outputs rendered image I' according to differentiable analytic synthesis. The spatial and frequency-domain losses are applied during the training stage to learn high-quality shapes and textures.

3.2.1. Spatial Domain Loss

Landmark-level. The alignment of facial landmarks is the alignment of high-level semantics between pixels of face images. To supervise the network, we usually project the shape we get abovementioned into the 2D image and minimize the difference between its 68 landmarks K_i^p and the ground-truth 68 landmarks K_i^g. Moreover, we assign different weights w_i to different face parts. The landmark loss is defined as:

$$L_{lmk} = \frac{1}{N}\sum_{i=1}^{N} w_i \left\| K_i^g - K_i^p \right\|^2. \tag{4}$$

Image-level. Based on equation (3), we build the image-level loss according to the photometric discrepancy between the original image I and the reconstructed image I'. To weaken the harmful effect brought by hair and face decoration, a skin mask is introduced to guide the loss as follows:

$$L_{\text{img}} = \frac{\left\| M \odot I - M \odot I' \right\|}{\|M\|}. \tag{5}$$

Perceptual-level. Some traditional methods use low-level information as the supervision information of the network, which results in smooth output images, so the appropriate

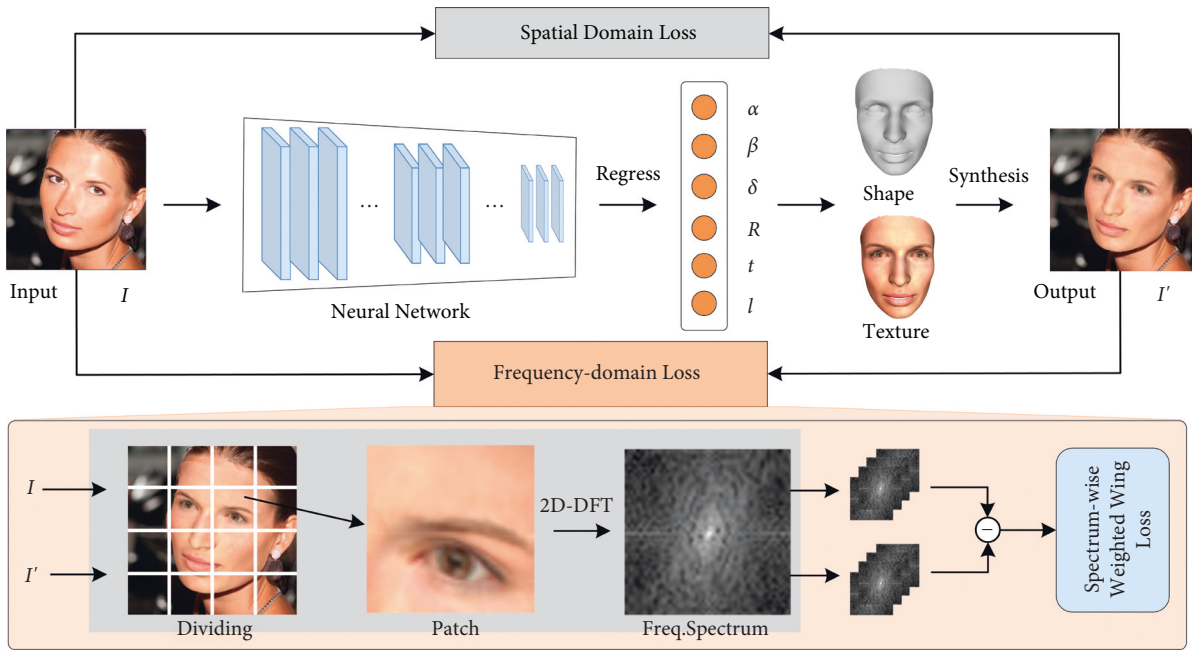


Figure 1: Overview of our approach. Our network is a weak-supervision network that considers both spatial and frequency-domain loss. The entire architecture feeds a single 2D image into the convolutional neural network (ResNet-50) to regress the 3DMM coefficients **α**, **β**, **δ** and rendering parameters I, p. With the parameters, we can reconstruct the 3D shape and texture, and synthesize the re-rendered image. A spectrum-wise weighted Wing loss is devised for fine fitting in the frequency domain.

Figure 2: The shadow problem brought by using only spatial domain loss: in the mask map (middle), we found that when the occlusion color is complex, the mask is correspondingly not good, so it will lead to the phenomenon of "under-eye dark circle" (right).

selection of a layer of output features input perceptual loss function can enhance the details. Influenced by recent work [13], we also use a pretrained face recognition network to fit this deep level of information during training. Perceptual loss is defined as:

$$L_{\text{per}} = 1 - \frac{f(I)f(I')}{\|f(I)\|\|f(I')\|}, \tag{6}$$

where $f(\cdot)$ denotes deep feature encoding.

The problem brought by spatial loss. Image-level loss learns uncertain texture when severe occlusions exist on the face. Figure 2 shows the output texture has black eyes when wearing glasses on the face. The reason is that DNN learns weights from high frequency to low frequency during the process of fitting images, but it is challenging to work in harmony without explicit guidance on the frequency domain [27].

3.2.2. Frequency-Domain Loss. Since the spatial domain loss could not handle the issue of facial occlusion well, we propose to use the frequency domain loss to alleviate it. Inspired by [30], we convert the input image and the rendered output image into their frequency representations and model the supervision between them.

Frequency representation. The representation in the frequency domain can be implemented by differentiable discrete Fourier transformation (DFT) [30], being formulated by:

$$F(u,v) = \sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\cdot e^{-i2\pi(ux/M+vy/N)}. \tag{7}$$

Figure 3 shows that there is a certain gap between the frequency spectrums with and without frequency supervision. The frequency difference between Figures 3(a) and 3(b)

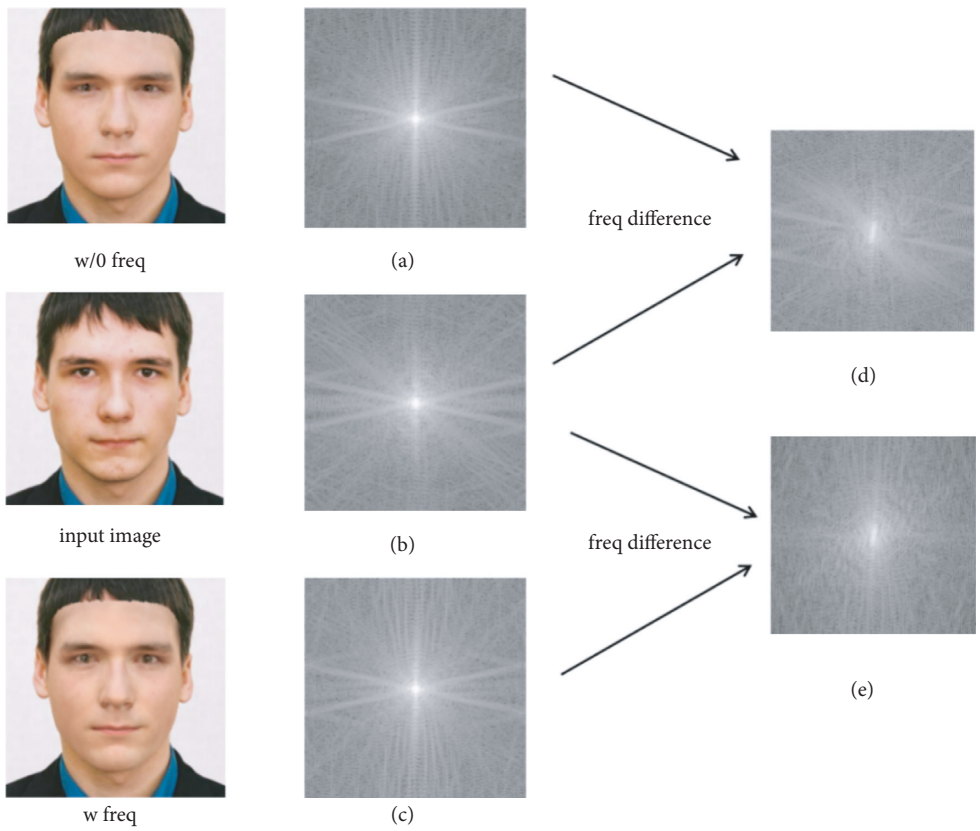


Figure 3: The difference in frequency spectrum with and without frequency supervision: (b) is the input image's spectrum. (a) and (c) are the frequency spectrums of the re-rendered images without and with frequency supervision; (d) and (e) are the difference between the two re-rendered spectrums, respectively.

and the frequency difference between Figures 3(c) and 3(b) are reflected in Figures 3(d) and 3(e), respectively. It is not difficult to find that after performing the differential calculation in the frequency domain, the generated frequency spectrum by our frequency domain supervision is closer to the original input image. Therefore, using the frequency domain loss, a supervision signal to assist the reconstruction of 3D faces, the network can synthesize frequencies that are not easy to synthesize effectively.

Frequency-based wing loss. We devise a loss function based on frequency representation for retrieving the missing frequencies in the re-rendered image.

Moreover, to learn more subtle changes in the frequency domain, the Wing loss [40] is adopted to design the frequency loss based on local patches divided from images:

$$L_{freq} = \frac{\sum_{u=0}^{M-1}\sum_{v=0}^{N-1}\sum_{p=0}^{P-1} Wing\left(F^{(p)}(u,v) - F'^{(p)}(u,v)\right)}{MNP}, \tag{8}$$

where M and N are height and width of image, and P is the number of patches. $F(u,v)$ is the spatial frequency value at the spectrum coordinate (u,v) of the input image I, and $F'(u,v)$ is that of re-rendered image I'. The advantage of Wing loss is that the gradient keeps high even at a minimal error. Thus, the low frequency that determines the realism of rendering could be enlarged to improve the reconstruction quality.

Spectrum-wise weighting. Under original Wing loss, the weights for different frequencies are equal and constant. In our design, we hope to pay more attention to the high-frequency part and less to the low-frequency. Therefore, we propose spectrum-wise weights for the frequency-based Wing loss, defined as:

$$\text{Wing}(y,u,v) = \begin{cases} w(u,v)\ln\left(\dfrac{1+|y|}{\epsilon}\right), & \text{if } |y| < w(u,v), \\ |y| - C, & \text{otherwise}, \end{cases} \tag{9}$$

where $y = \Delta a + \Delta b \cdot i$, and $C = w(u,v) - w(u,v)\ln(1 + w(u,v)/\epsilon)$. Δa and Δb are the difference of real and imaginary parts, respectively, between $F(u,v)$ and $F'(u,v)$. And $w(u,v)$ is also the spectrum-wise threshold between a linear and nonlinear part of the wing curve, which is learned adaptively during training. As Figure 4 shows, spectral weighted Wing loss gets more saturated and closer to the actual face texture. What's more, different from Figure 2, the neural network no longer only uses simple pixel-level supervision information but also the supervision in the frequency domain. It has a specific resistance to the phenomenon of dark circles under the occlusion of sunglasses.

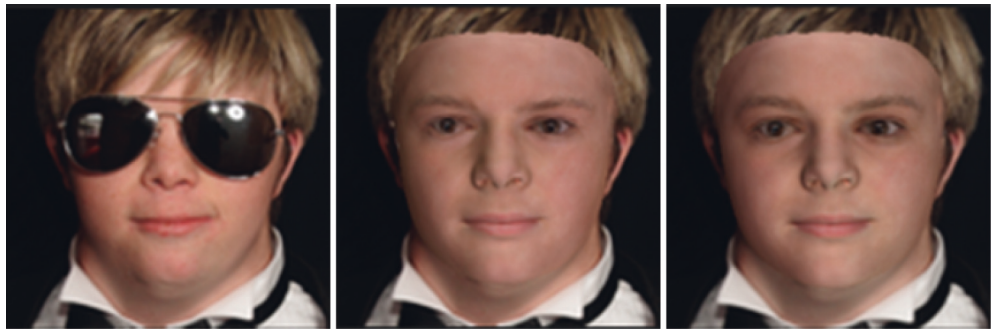


Figure 4: From left to right: the input image, the image rendered using the constant weighting, and the image rendered using the spectrum-wise weight. The right image is much more colorful than the middle one.

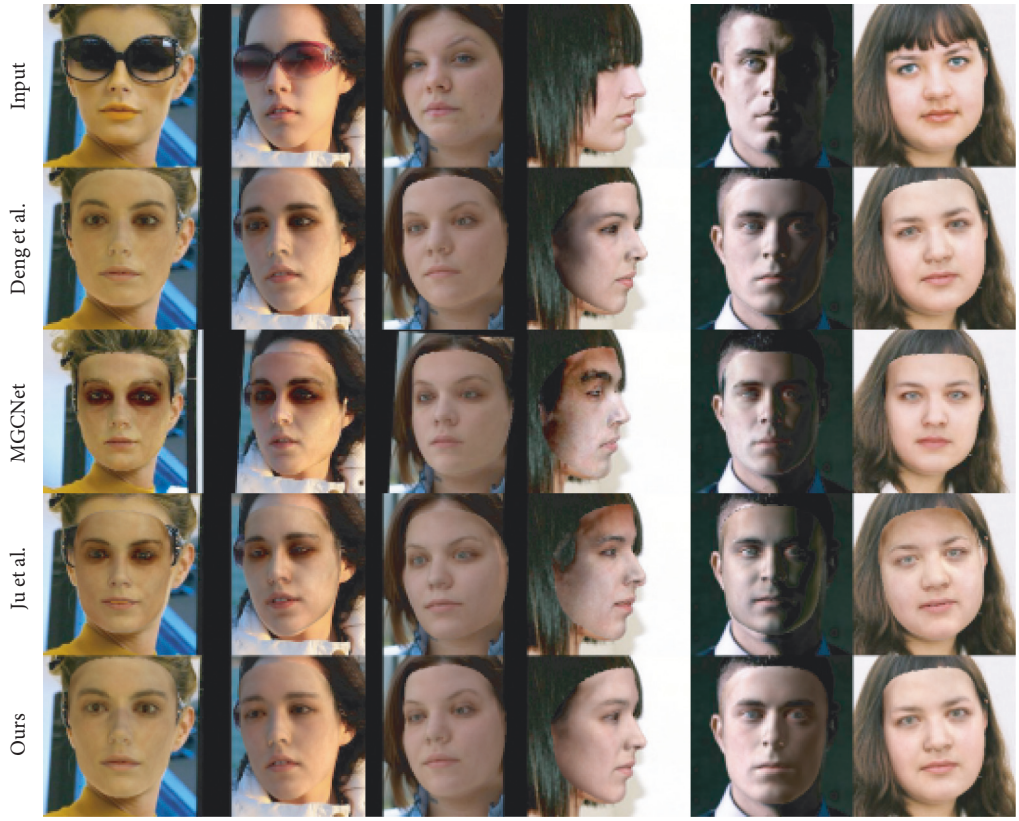


Figure 5: Comparison with Deng et al. [13], MGCNet [54]. and Ju et al. [55] Our re-rendered images are better in the details and are more consistent with the input image. The images are from AFLW2000 [17].

4. Experimental Results

Training data pipeline. In terms of training set, we get ~ 320 k face images from CelebA [48], FFHQ [49] and Multi-PIE [50]. Then we use the method of [51] to align and crop facial images for model input.

Detailed setting. We follow the method of [13] which trained a naïve Bayes classifier with Gaussian mixture model on a skin image dataset from [52] to generate the mask used in image-level loss, and then preprocess the training set. We use the Adam optimizer for ResNet-50 [53] that predicts $\boldsymbol{\Theta}$ and its initial learning rate is set to $1e-4$. The total loss converges after about 200 K iterations.

4.1. Qualitative Evaluation. Figures 5 and 6 shows the re-rendered images and textures overlayed on original images, respectively, by comparing the methods [13, 54, 55] on the AFLW2000 dataset [17]. Ju et al. [55] used GAN to repair the occlusion images after obtaining the textures from 3DMM model, which did not use the image-level loss but the adversarial loss. Deng et al. [13] used a robust loss including pixel loss, for 3D face reconstruction. MGCNet [54] is a multi-view-based 3D face reconstruction method. It shows that our texture

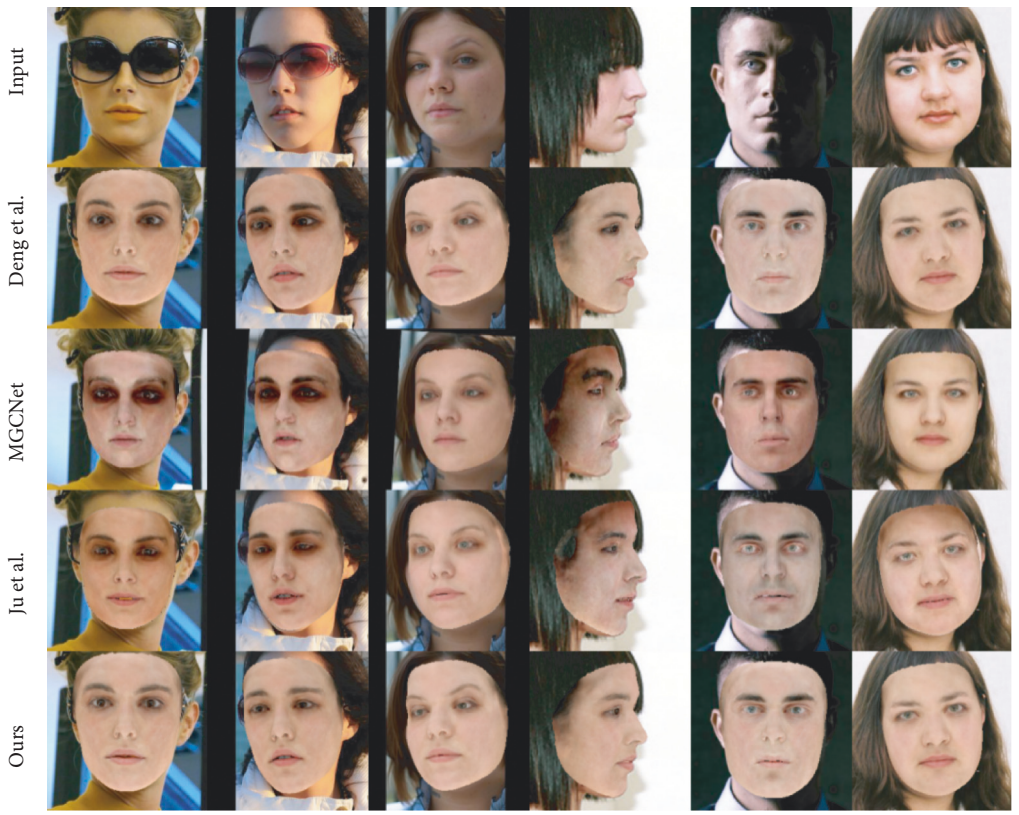


Figure 6: Comparison with Deng et al. [13], MGCNet [54] and Ju et al. [55]. Without illumination, the textures synthesized by our method more closely match the original images and are resistant to occlusion colors. The images are from AFLW2000 [17].

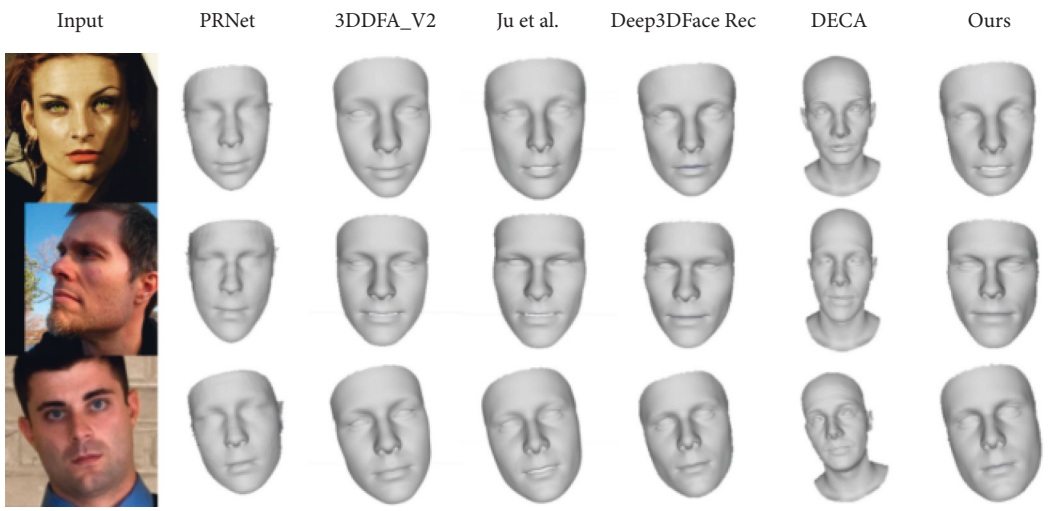


Figure 7: Comparison to other recent reconstruction methods, from left to right: PRNet [18], 3DDFA_V2 [56], Ju et al. [55], Deep3-DFaceRec [13], DECA [25] and Our method. The images are from AFLW2000 [17].

does not have black shadows in the presence of occlusions like hair, glasses, and poor lighting. Moreover, our method can also help the network reconstruct more detailed faces, such as the reconstruction of the eyes in the third column of Figure 5.

Figure 7 shows our results in shape compared to recent methods [13, 18, 25, 55, 56]. The 3D face shape reconstructed by our method is very close to the input image in the case of poor illumination and large occlusions. And we could find that our results are more finely synthesized on the eyes and mouth relative to Deep3DFaceRec [13], with a slight advantage.

4.2. Quantitative Comparison

4.2.1. FaceScape Benchmark. FaceScape benchmark [16] is an all-sided evaluation method that considers various poses,

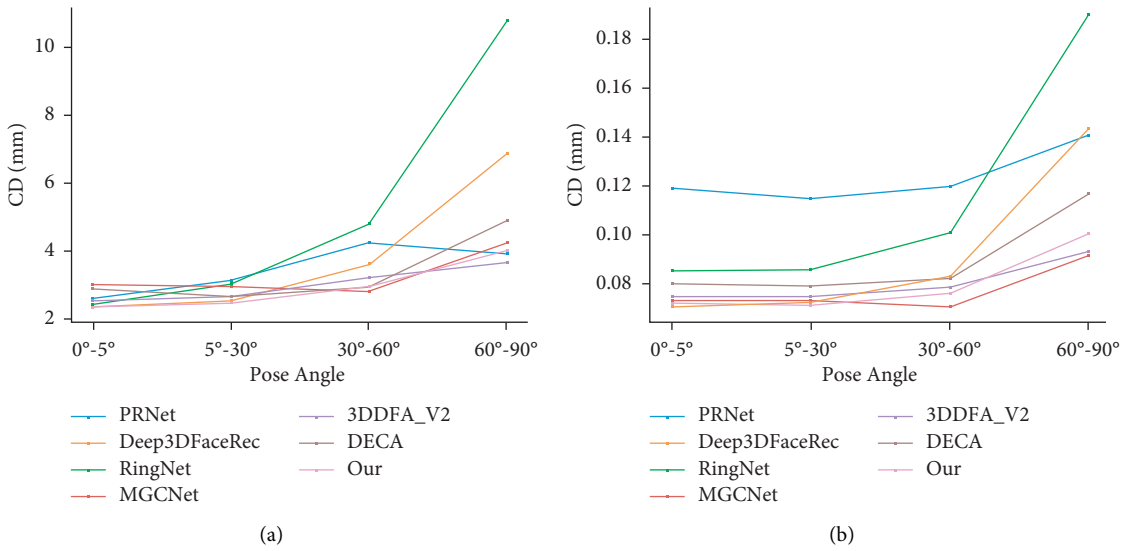


Figure 8: Charts to visualize the quantitative evaluation on FS-wild dataset: (a) Chamfer distance and (b) Mean normal error.

Table 1: Quantitative evaluation on FS-lab benchmark categorized by pose angle.

Pose angle	0°			30°			60°		
Method	CD	MNE	CR	CD	MNE	CR	CD	MNE	CR
Deep3DFaceRec [13]	2.30	0.080	91.8	3.95	0.092	87.3	4.80	0.122	79.5
MGCNet [48]	3.45	0.085	92.7	3.91	0.092	90.1	3.65	0.090	83.2
DECA [25]	3.30	0.093	99.8	4.14	0.100	97.4	4.20	0.106	97.1
3DDFA_V2 [56]	3.05	0.093	95.2	3.41	0.096	93.8	3.82	0.096	88.1
PRNet [18]	2.94	0.132	92.5	3.40	0.125	90.1	3.74	0.121	85.1
FaceScape_deep [16]	2.40	0.086	96.7	7.27	0.124	87.7	3.87	0.108	90.5
UDL [20]	2.21	0.092	79.5	5.33	0.122	71.3	5.63	0.167	62.0
Ours	2.12	0.077	92.1	2.30	0.079	89.8	3.28	0.109	85.2

expressions, environments, and focal lengths to evaluate the accuracy of single-view face 3D reconstruction. It includes two parts of data: FS-Wild data and FS-Lab data. The FS-Wild data consists of 400 face images of 400 synthesized subjects, each with a reference 3D face model, and is divided into four groups according to the camera orientation and the face orientation (0°–5°, 5°–30°, 30°–60°, and 60°–90°). The FS-Lab renders 330 images using the 20 detailed 3D models with three different focal lengths: 1200 (long focal), 600 (middle focal), 300 (short focal), and eleven different camera locations, which one camera at exact front 0°, eight cameras deflecting 30° and two cameras deflecting 60°.

We compared our methods with publicly available methods, i.e., Deep3DFaceRec [13], MGCNet [54], DECA [25], 3DDFA_V2 [56], PRNet [18], FaceScape_deep [16], and UDL [20]. Since the FaceScape benchmark has 3D ground-truth data, Chamfer Distance (CD) measures the error between the predicted and ground-truth mesh. Mean normal error (MNE) measures the intersection of the valid region distance between the predicted normal map and ground-truth normal map, which are obtained from the corresponding mesh rendered in the cylindrical coordinate. The complete rate (CR) measures the completeness of the reconstruction results.

4.2.2. Comparison on FS-Wild Datasets. Figure 8 shows the values of CD and MNE under different pose angles in the FS-Wild datasets. The Chamfer distance measured shows the overall error distance in Figure 8(a). Our method performs well in frontal and side views, especially for the frontal and small pose angle views. The results of MNE are shown in Figure 8(b), although, we are not as good as Deep3DFaceRecon [13] at a small pose angle, our effect is much better than as the face angle increases. Furthermore, in the case of large pose angle, our performance is third only to MGCNet [54] which used 3D-ground truth supervision, and 3DDFA_V2 [56] which, we exceed its performance on small pose angle.

4.2.3. Comparison on FS-Lab Datasets. This section reports the values of CD, MNE, and CR of several methods at different pose angles on FS-Lab datasets.

In Table 1, We can see that most methods perform well in the frontal view but severely degrade in the side view. And our method is not only relatively stable for side view but also has the best performance results.

In addition, it is worth noting that CR measures the completeness of the reconstruction results, which is defined as: $\eta = S(P_p \cap P_g)/S(P_g)$. The position map P_p and P_g are

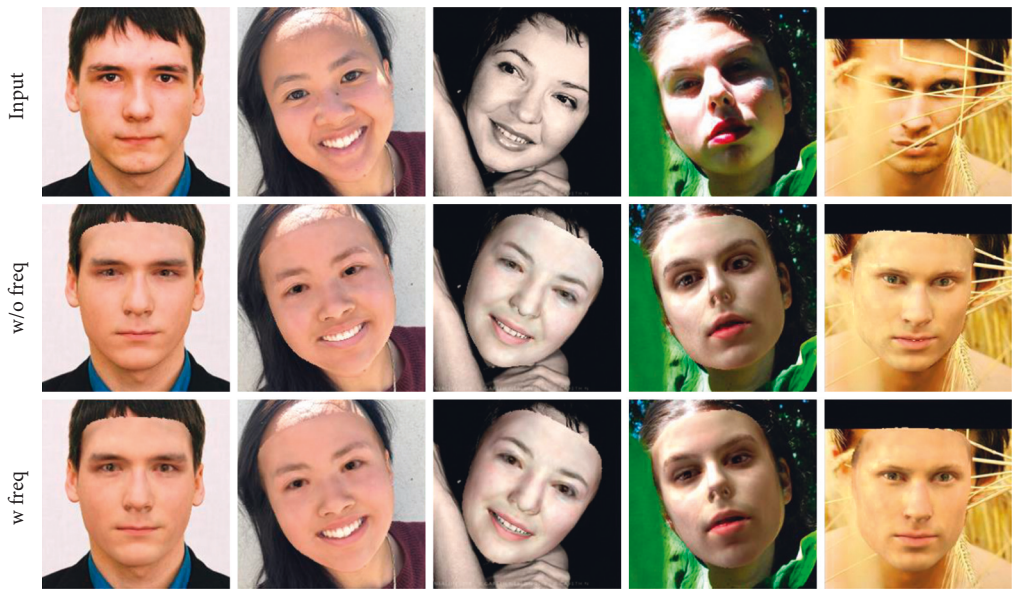


Figure 9: Ablation study on the proposed frequency-domain loss. The frequency-domain method has a better synthesis effect on the eyes, lips, etc. From up to down: the input, the result without frequency loss, and that with frequency loss.

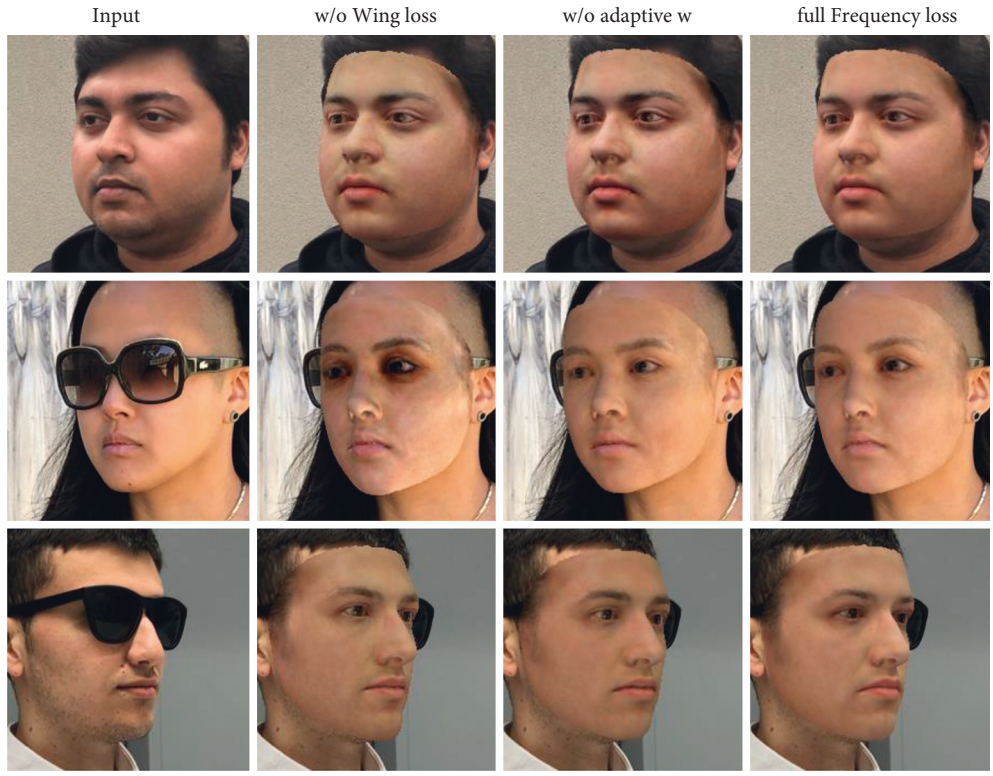


Figure 10: Ablation study of each key component for the frequency loss, i.e., Wing loss and spectrum-wise weight. From left to right: the input, with l_2 loss (w/o Wing loss), the re-rendered results with constant weighting (w/o adaptive w), and with full loss. The input samples are from [58].

the predicted mesh and the ground-truth mesh rendering in the cylindrical coordinate. $S(P)$ is the function that returns the area of the position map P. Limited by the 3DMM model, our method uses the BFM model, excluding the ear and neck region, to reconstruct the actual face area as much as possible. However, 3DDFA_V2 [56] used the MFF [57]

Table 2: Ablation studies of different patch sizes are important for the frequency loss.

Patch size	Indicator					
	AFLW2000			Now dataset		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
1×1	0.744	12.008	0.232	0.869	23.245	0.126
2×2	0.743	11.977	0.226	0.870	23.241	0.128
4×4	0.762	12.558	0.221	0.871	23.088	0.122
8×8	0.753	12.161	0.226	0.868	23.392	0.126

algorithm to fit 3DMM parameters and further completed the model to a complete head model with ears and neck. And the DECA [25] reconstructed the entire head with the FLAME model. Obviously, our 3D reconstruction is comparable with other methods.

4.3. Ablation Study. To verify the effectiveness of our frequency-domain loss, we perform ablation experiments on Now datasets [58] and AFLW2000-3D [17] datasets.

Frequency domain loss. To show the importance of our frequency-domain loss, we train our model with and without frequency-domain loss and compare the results. Figure 9 shows that the frequency loss can assist the convolutional neural network in synthesizing some details that are not easy to synthesize. More detailed face features can be captured in the areas of the eyes, mouth, etc. Moreover, the reconstruction is also very accurate when the face is occluded.

Wing loss and spectrum-wise weighting. Figure 10 shows that the full patch-based spectrum-wise weighted Wing loss achieves the best performance. If we use l_2 loss instead of Wing loss, it could not amplify some smaller frequencies error, resulting in underfitting the reconstructed frequency for face synthesizing. Thus, the facial texture is not uniform enough on the whole face.

Moreover, it is noteworthy that the occlusion part will be overfitted when the face is occluded. Wing loss can remove shadows caused by occlusion for two reasons. On the one hand, we use the generated mask to make the network pay little attention to the occlusion part. On the other hand, we use spectrum-wise weighted Wing loss to amplify the error of the high-frequency part and suppress the large frequency difference. Generally, the mask could not perfectly cover some complex, occluded parts of the face. If we only used the pixel-level loss, the color of the occluder would still be fitted. Actually, in the later stage of network training, the frequency gap of occluded parts between the input and reconstructed image will be much larger than that of the unoccluded parts. Spectrum-wise weighted Wing loss guides the network to synthesize frequencies that are not easy to synthesize rather than the shadow parts. Thereby, the reconstruction can be learned in harmony, and the shadow effect caused by occlusion is alleviated to a certain extent.

On the contrary, if we use Wing loss with fixed weighting, it ignores that different parts of the face have different frequency compositions. In that case, some face parts' frequency domain synthesis is insufficient, resulting in the facial texture appearance with spots. Moreover, the reconstructed face is not very fine for some details like the eyes.

Patch size for DFT. We also explored the effect of different patch sizes on the reconstruction results. We show this effect by rendering the reconstruction results on a 2D plane and comparing the similarity between the rendered and input images. Structural similarity (SSIM) [59] is an indicator proposed to measure images' similarity, which can be applied to luminance, contrast, and structure. Peak Signal-to-Noise Ratio (PSNR) is defined as: $PSNR = 10 * \log_{10}(MAX_I^2/MSE_{\langle I_i, I_r\rangle})$, where MAX_I^2 is the maximum pixel value of the picture and $\mathrm{MSE}_{\langle I_i, I_r\rangle}$ is the mean square error of the input image I_i and the rendered image I_r. Learned perceptual image patch similarity (LPIPS) metric [60] uses the deep feature to measure the similarity of images. We also report SSIM, PSNR, and LPIPS between re-rendered images and original images under four patch sizes on the AFLW2000 dataset [17] and Now dataset [58] in Table 2. According to the result, we can see that the patch size of 4×4 shows the best performance.

5. Conclusion

We propose a spatio-frequency decoupled weak-supervision for 3D face reconstruction and build the weakly supervision by applying both spatial domain loss and frequency domain loss to enhance the reality of re-rendered facial images based on the reconstructed shape and texture. The key contribution is the designed spectrum-wise weighted Wing loss based on frequency loss on image patches, which narrows the gap between input and output in the frequency domain and captures inconspicuous frequency affecting reality. Experiments show the effectiveness of our method and comparable results with several state-of-the-art methods.

Data Availability

Any data used to support the findings of this study are from previously reported studies and datasets, which have been cited.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (62072126, 62102105), the Guangdong Basic and Applied Basic Research Foundation (2020A1515110997, 2022A1515010138, 2022A1515011501), the Science and Technology Program of Guangzhou (202201020229, 202102010419, 202002030263), the Open Project Program of the State Key Lab of CAD and CG (Grant no. A2218), Zhejiang University.

References

[1] M. Wang and W. Deng, "Deep face recognition: a survey," *Neurocomputing*, vol. 429, pp. 215–244, 2021.

[2] A. Richard, M. Zollhöfer, and Y. Wen, "Meshtalk: 3d face animation from speech using cross-modality disentanglement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1173–1182, Montreal, QC, Canada, October 2021.

[3] J. Philip, *Multi-view image-based editing and rendering through deep learning and optimization*, PhD Thesis, Univeristé Nice Sophia Antipolis, Nice, France, 2021.

[4] T. F. Cootes, M. C. Ionita, and C. Lindner, "Robust and accurate shape model fitting using random forest regression voting," in *European Conference on Computer Vision*, vol. 7578, pp. 278–291, Springer, 2012.

[5] C. M. Grewe and S. Zachow, "Fully automated and highly accurate dense correspondence for facial surfaces," in *European Conference on Computer Vision*, vol. 9914, pp. 552–568, Springer, 2016.

[6] W. Peng, Z. Feng, and C. Xu, "Parametric t-spline face morphable model for detailed fitting in shape subspace," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5515–5523, HI, USA, July 2017.

[7] W. Peng, Y. Su, K. Tang, C. Xu, Z. Feng, and M. Fang, "Geometrical consistency modeling on b-spline parameter domain for 3d face reconstruction from limited number of wild images," *Frontiers in Neurorobotics*, vol. 15, Article ID 652562, 2021.

[8] Y. Chen, W. Peng, K. Tang, A. Khan, G. Wei, and M. Fang, "Pyrapvconv: efficient 3d point cloud perception with pyramid voxel convolution and sharable attention," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 2286818, 9 pages, 2022.

[9] H. Liu, Y. Pan, J. Cao, H. Wang, and Y. Zhou, "Adaptive neural network backstepping control of fractional-order nonlinear systems with actuator faults," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 12, pp. 5166–5177, 2020.

[10] A. Morales, G. Piella, and F. M. Sukno, "Survey on 3d face reconstruction from uncalibrated images," *Computer Science Review*, vol. 40, no. 100, Article ID 100400, 2021.

[11] G. Pang, C. Shen, L. Cao, and A. V. D Hengel, "Deep learning for anomaly detection: a review," *ACM Computing Surveys*, vol. 54, no. 2, pp. 1–38, 2022.

[12] C. Tian, L. Fei, W. Zheng, Y. Xu, W. Zuo, and C. W. Lin, "Deep learning on image denoising: an overview," *Neural Networks*, vol. 131, pp. 251–275, 2020.

[13] Y. Deng, J. Yang, and S. Xu, "Accurate 3d face reconstruction with weakly-supervised learning: from single image to image set," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, CA, USA, June 2019.

[14] B. Gecer, S. Ploumpis, and I. Kotsia, "Ganfit: generative adversarial network fitting for high fidelity 3d face reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1155–1164, CA, USA, June 2019.

[15] J. Lin, Y. Yuan, and T. Shao, "Towards high-fidelity 3d face reconstruction from in-the-wild images using graph convolutional networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5891–5900, WA, USA, June 2020.

[16] H. Yang, H. Zhu, and Y. Wang, "Facescape: a large-scale high quality 3d face dataset and detailed riggable 3d face prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 601–610, WA, USA, June 2020.

[17] X. Zhu, Z. Lei, and X. Liu, "Face alignment across large poses: a 3d solution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 146–155, NV, USA, June 2016.

[18] Y. Feng, F. Wu, and X. Shao, "Joint 3d face reconstruction and dense alignment with position map regression network," in *Proceedings of the European Conference on Computer Vision*, pp. 534–551, ECCV), Munich, Germany, September 2018.

[19] E. Richardson, M. Sela, and R. Kimmel, "3d face reconstruction by learning from synthetic data," in *Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV)*, pp. 460–469, IEEE, CA, USA, October 2016.

[20] Y. Chen, F. Wu, Z. Wang, Y. Song, Y. Ling, and L. Bao, "Self-supervised learning of detailed 3d face reconstruction," *IEEE Transactions on Image Processing*, vol. 29, pp. 8696–8705, 2020.

[21] K. Genova, F. Cole, and A. Maschinot, "Unsupervised training for 3d morphable model regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8377–8386, Salt Lake City, UT, USA, June 2018.

[22] H. Zhou, J. Liu, and Z. Liu, "Rotate-and-render: unsupervised photorealistic face rotation from single-view images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5911–5920, Seattle, WA, USA, June 2020.

[23] J. Zhang, L. Lin, and J. Zhu, "Weakly-supervised multi-face 3d reconstruction," 2021, https://www.arxiv-vanity.com/papers/2101.02000/.

[24] M. Sela, E. Richardson, and R. Kimmel, "Unrestricted facial geometry reconstruction using image-to-image translation," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1576–1585, MA, USA, June 2017.

[25] Y. Feng, H. Feng, M. J. Black, and T. Bolkart, "Learning an animatable detailed 3d face model from in-the-wild images," *ACM Transactions on Graphics*, vol. 40, no. 4, pp. 1–13, 2021.

[26] C. Li, A. Morel-Forster, and T. Vetter, "To Fit or Not to Fit: Model-Based Face Reconstruction and Occlusion Segmentation from Weak Supervision," 2021, https://arxiv.org/abs/2106.09614arXiv preprint arXiv:210609614.

[27] N. Rahaman, A. Baratin, and D. Arpit, "On the spectral bias of neural networks," in *International Conference on Machine Learning*, vol. 97, pp. 5301–5310, PMLR, 2019.

[28] M. Tancik, P. Srinivasan, and B. Mildenhall, "Fourier features let networks learn high frequency functions in low dimensional domains," *Advances in Neural Information Processing Systems*, vol. 33, pp. 7537–7547, 2020.

[29] Z. Q. J. Xu, Y. Zhang, and T. Luo, "Frequency Principle: Fourier Analysis Sheds Light on Deep Neural Networks," 2019, https://aps.arxiv.org/abs/1901.06523v3arXiv preprint arXiv:190106523.

[30] L. Jiang, B. Dai, and W. Wu, "Focal frequency loss for image reconstruction and synthesis," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, vol. 13, pp. 919–929, QC, Canada, October 2021.

[31] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3d faces," in *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 187–194, NY, USA, July 1999.

[32] A. Tuan Tran, T. Hassner, and I. Masi, "Regressing robust and discriminative 3d morphable models with a very deep neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5163–5172, PR, USA, June 2017.

[33] C. Y. Wu, Q. Xu, and U. Neumann, "Synergy between 3dmm and 3d landmarks for accurate 3d facial geometry," in *Proceedings of the 2021 International Conference on 3D Vision (3DV)*, pp. 453–463, IEEE, London, UK, December 2021.

[34] A. Dib, G. Bharaj, and J. Ahn, "Practical face reconstruction via differentiable ray tracing," in *Computer Graphics Forum*, pp. 153–164, Wiley Online Library, NJ, USA, 2021.

[35] L. Tran and X. Liu, "Nonlinear 3d face morphable model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7346–7355, PR, USA, June 2018.

[36] S. Yamaguchi, S. Saito, K. Nagano et al., "High-fidelity facial reflectance and geometry inference from an unconstrained image," *ACM Transactions on Graphics*, vol. 37, no. 4, pp. 1–14, 2018.

[37] M. Yang, J. Guo, and Z. Cheng, "Self-supervised High-Fidelity and Re-renderable 3d Facial Reconstruction from a Single Image," 2021, https://arxiv.org/abs/2111.08282arXiv preprint arXiv:211108282.

[38] H. Zhou, S. Hadap, and K. Sunkavalli, "Deep single-image portrait relighting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7194–7202, Seoul, Korea, June 2019.

[39] P. Wang, C. Lin, and B. Xu, "Low-frequency guided self-supervised learning for high-fidelity 3d face reconstruction in the wild," in *Proceedings of the 2020 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, London, UK, July 2020.

[40] Z. H. Feng, J. Kittler, and M. Awais, "Wing loss for robust facial landmark localisation with convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2235–2245, PR, USA, June 2018.

[41] X. Wang, L. Bo, and L. Fuxin, "Adaptive wing loss for robust face alignment via heatmap regression," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 6971–6981, Seoul, Korea, June 2019.

[42] C. Cao, Y. Weng, and S. Zhou, "Facewarehouse: a 3d facial expression database for visual computing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 3, pp. 413–425, 2014.

[43] R. Ramamoorthi and P. Hanrahan, "An efficient representation for irradiance environment maps," in *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 497–500, NY, USA, June 2001.

[44] Y. Nirkin, I. Masi, and A. T. Tuan, "On face segmentation, face swapping, and face perception," in *Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, pp. 98–105, IEEE, Xi'an, China, May 2018.

[45] M. P. P. Segundo, L. Silva, O. R. P. Bellon, and C. C. Queirolo, "Automatic face segmentation and facial landmark detection in range images," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 5, pp. 1319–1330, 2010.

[46] A. Tewari, M. Zollhöfer, and P. Garrido, "Self-supervised multi-level face model learning for monocular reconstruction at over 250 hz," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2549–2559, PR, USA, June 2018.

[47] A. Tewari, M. Zollhofer, and H. Kim, "Mofa: model-based deep convolutional face autoencoder for unsupervised monocular reconstruction," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 1274–1283, Montreal, BC, Canada, October 2017.

[48] Z. Liu, P. Luo, and X. Wang, "Deep learning face attributes in the wild," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3730–3738, MA, USA, June 2015.

[49] T. Karras, S. Laine, and M. Aittala, "Analyzing and improving the image quality of stylegan," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8110–8119, WA, USA, June 2020.

[50] J. Heo and M. Savvides, "Face recognition across pose using view based active appearance models (vbaams) on cmu multi-pie dataset," in *International Conference on Computer Vision Systems*vol. 459, , pp. 527–535, Springer, 2008.

[51] D. Chen, G. Hua, and F. Wen, "Supervised transformer network for efficient face detection," in *Proceedings of the European Conference on Computer Vision*, pp. 122–138, Springer, Amsterdam, The Netherlands, October 2016.

[52] M. J. Jones and J. M. Rehg, "Statistical color models with application to skin detection," *International Journal of Computer Vision*, vol. 46, no. 1, pp. 81–96, 2002.

[53] K. He, X. Zhang, and S. Ren, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, PR, USA, June 2016.

[54] J. Shang, T. Shen, and S. Li, "Self-supervised monocular 3d face reconstruction by occlusion-aware multi-view geometry consistency," in *European Conference on Computer Vision*, pp. 53–70, Springer, Berlin, Germany, 2020.

[55] Y. J. Ju, G. H. Lee, and J. H. Hong, "Complete face recovery gan: unsupervised joint face rotation and de-occlusion from a single-view image," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3711–3721, HI, USA, January 2022.

[56] J. Guo, X. Zhu, and Y. Yang, "Towards fast, accurate and stable 3d dense face alignment," in *European Conference on Computer Vision*, pp. 152–168, Springer, Berlin, Germany, 2020.

[57] S. Romdhani and T. Vetter, "Estimating 3d shape and texture using pixel intensity, edges, specular highlights, texture constraints and a prior," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 2, pp. 986–993, CA, USA, June 2005.

[58] S. Sanyal, T. Bolkart, and H. Feng, "Learning to regress 3d face shape and expression from an image without 3d supervision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7763–7772, CA, USA, June 2019.

[59] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[60] R. Zhang, P. Isola, and A. A. Efros, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586–595, PR, USA, June 2018.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 6316140, 14 pages
https://doi.org/10.1155/2022/6316140

Research Article

Ensemble Dilated Convolutional Neural Network and Its Application in Rotating Machinery Fault Diagnosis

Yuxiang Cai, Zhenya Wang, Ligang Yao, Tangxin Lin, and Jun Zhang

School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350108, China

Correspondence should be addressed to Zhenya Wang; wzyaxj@163.com and Ligang Yao; ylgyao@fzu.edu.cn

Received 23 June 2022; Revised 19 August 2022; Accepted 23 August 2022; Published 21 September 2022

Academic Editor: Wei Xiang

Fault diagnosis of rotating machinery is an attractive yet challenging task. This paper presents a novel intelligent fault diagnosis scheme for rotating machinery based on ensemble dilated convolutional neural networks. The novel fault diagnosis framework employs a model training strategy based on early stopping optimization to ensemble several one-dimensional dilated convolutional neural networks (1D-DCNNs). By varying the dilation rate of the 1D-DCNN, different receptive fields can be obtained to extract different vibration signal features. The early stopping strategy is used as a model update threshold to prevent overfitting and save computational resources. Ensemble learning uses a weighted mechanism to combine the outputs of multiple 1D-DCNN subclassifiers with different dilation rates to obtain the final fault diagnosis. The proposed method outperforms existing state-of-the-art classical machine learning and deep learning methods in simulation studies and diagnostic experiments, demonstrating that it can thoroughly mine fault features in vibration signals. The classification results further show that the EDCNN model can effectively and accurately identify multiple faults and outperform existing fault detection techniques.

1. Introduction

Rotating machinery is widely used in manufacturing, transportation, aerospace, and other industries [1, 2]. However, rotating machinery systems frequently operate in high-speed, heavy-duty environments, inevitably resulting in internal components (such as bearings and gears) that are susceptible to damage. While the efficiency of rotating machinery can be reduced by minor failures, the consequences of serious failures can be catastrophic. Furthermore, vibration signals monitored in harsh industrial environments are subject to significant noise interference, which poses a significant challenge for robust fault diagnosis. Fortunately, with the rapid development and integration of sensor technology in the modern industry, condition monitoring and fault diagnosis have become the most effective methods to avoid damage using the measured monitoring vibration signals [3, 4]. As a result, prognostics and health management (PHM) of rotating machinery under changeable working circumstances has emerged as a critical technique for economic efficiency and a hot topic of various research studies [5].

1.1. Problems and Motivation. The diagnosis of rotating machinery faults is essentially a pattern recognition issue related to the health condition. Traditional fault diagnosis techniques, such as the wavelet transform [6, 7], variable modal decomposition [8, 9], and empirical modal decomposition [10–13], are challenging to extract fault discriminative features from vibration signals with nonstationary and nonlinear characteristics and demand excessive expertise and expert knowledge, limiting their practical application. Furthermore, the development of artificial intelligence technologies has increased their application in a variety of industries, such as mechanical fault diagnostics. Intelligent fault diagnosis has two main forms: machine learning combined with manual feature extraction [14, 15] or deep learning with automated feature extraction [16–18]. Deep learning-based approaches have gained a lot of attention and popularity as a result of their ability to achieve

good end-to-end fault diagnosis and automated fault feature extraction. Traditional fault diagnosis methods or a combination of manual feature extraction and machine learning cannot accomplish the task [19, 20].

While deep learning-based mathematical frames decrease the requirement for expert knowledge and manual feature engineering, it is an effective tool for mechanical fault identification. Artificial neural networks (ANNs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs) are the most common deep learning techniques. For example, Moosavi et al. [21] used a multilayer ANN for fault detection and diagnosis of electric motors. Mao et al. [22] proposed a semirandom subspace method with a bidirectional gate recurrent unit (a modified RNN algorithm) to take full advantage of fusion features for bearing fault diagnosis. Wu and Ma [23] proposed an improved RNN method for wind turbine fault diagnosis based on long short-term memory and Kullback–Leibler divergence. The abovementioned deep learning-based research approaches produced good fault diagnostic conclusions. However, when compared to the other two deep learning approaches, the ANN-based diagnostic method suffers from weak nonlinear fitting ability. Furthermore, the RNN-based diagnostic technique suffers from gradient dispersion and gradient explosion conundrum in model training, as well as containing too many model parameters.

In this work, CNN was chosen over the other approaches because of its superior region feature extraction capabilities and unique model parameter sharing mechanism [24]. Many experts and researchers have conducted extensive research on CNN models. For example, Chen et al. [25] suggested a rolling element-bearing fault approach based on cyclic spectrum consistency and CNN to achieve high diagnostic accuracy. Plakias and Boutalis [26] proposed an attention-intensive CNN with improved generalization capabilities for recognizing rolling element-bearing faults. Guo et al. [27] developed a fault diagnosis model capable of reliable and quick fault identification of multichannel data utilizing multilinear principal component analysis and CNN. Han et al. [28] suggested a CNN-support vector machine system with high robustness in diagnosing bearing faults.

1.2. Proposed Methods. However, the abovementioned CNN-based fault diagnosis method achieves advanced diagnostic performance due to its robust local feature extraction and flexible structure. The abovementioned CNN-based fault diagnosis research, on the other hand, has such limitations as follows:

(1) The above CNN models are constrained by the classic convolution process, which is incapable of accurately diagnosing faults in complicated industrial diagnostic situations.

(2) In the case of a single receptive field (RF), fault diagnosis of the CNN frequently relies on a few feature maps to create unreliable judgments, posing a significant risk to decision-making.

Therefore, the purpose of this study is to investigate a mechanical health monitoring method with strong robustness in order to reduce the negative noise impact under various complex operating situations. To address the aforementioned limitations of classic CNN, this paper proposes an intelligent rotating machinery fault diagnosis model based on the ensemble dilated convolutional neural network (EDCNN) and early stopping optimization. Dilated convolutional neural network (DCNN) not only has a large RF but can also maintain the size of the model. EDCNN takes the concept of ensemble learning and applies it to fault classification by ensembling multiple weak classifiers to jointly consider multiple feature maps for decision making.

1.3. Contributions and Structure of This Paper. The main contributions of this work are as follows:

(1) A novel deep learning algorithm called EDCNN is proposed, which ensembles multiple dilated convolutional neural networks with different dilation rates to extract features effectively.

(2) An intelligent model training approach based on early stopping optimization is implemented. This technique conserves computing resources while minimizing overfitting and performance degradation.

(3) A novel EDCNN-based fault diagnosis framework applied to rotating machinery is proposed. The effectiveness and superiority of the proposed method are confirmed by the benchmark rolling bearing dataset and the wind turbine simulator dataset.

The rest of this paper can be summarized as follows. The proposed EDCNN and the suggested EDCNN-based intelligent fault diagnostic method for rotating machinery are described in Section 2. The proposed fault diagnostic model is validated using the rolling bearing and wind turbine datasets in Section 3. Finally, the main conclusions are summarized in Section 4.

2. Intelligent Fault Diagnosis Method for Rotating Machinery Based on the EDCNN

In this section, the basic theory of the proposed EDCNN method is first discussed. Subsequently, the proposed framework for intelligent fault diagnosis is presented.

2.1. Mathematical Model of the Proposed EDCNN

2.1.1. One-Dimensional Dilated Convolutional Neural Network (1D-DCNN). Deep learning-based fault diagnosis techniques have attracted widespread attention and have been extensively studied and applied. ANN is a mathematical model that simulates the activity mechanism of the human nervous system by computing the weight of each neuron on all neurons connected layer to layer. When neurons are overstacked, however, the computing resources are too enormous and the capacity to extract features is

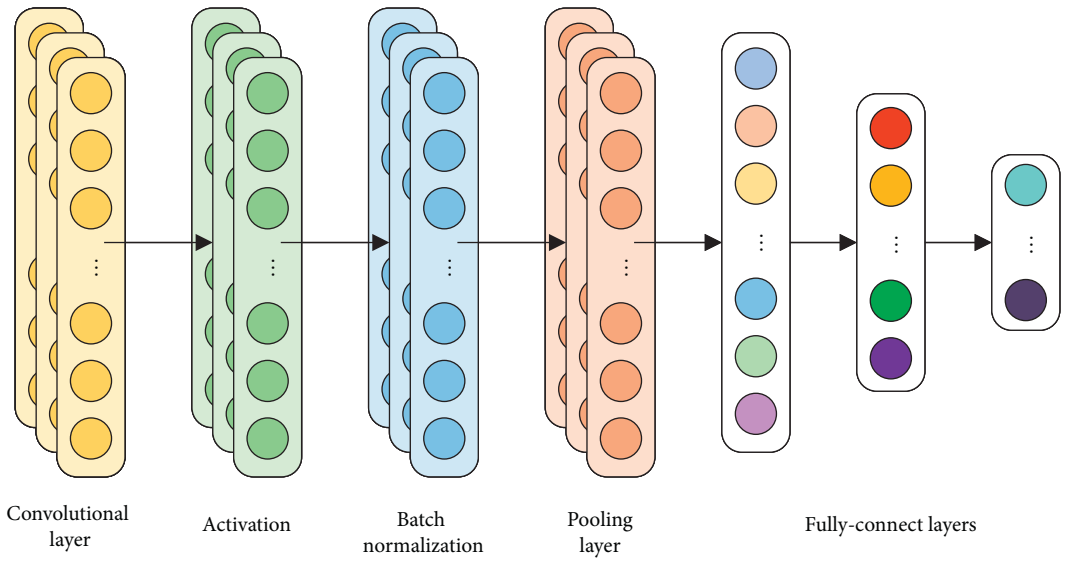


Figure 1: Structure of the one-dimensional convolutional neural network.

extremely limited. The RNN can extract temporal information more efficiently than ANN, but its nonparallel computing strategy will give training an appropriate diagnostic model harder. As a result, CNN was chosen by the authors for the research of deep learning in fault diagnosis.

Dilated convolutional neural networks are modified convolutional neural networks that are used for multipattern identification and sensitive feature extraction in complicated tasks. The same model volume can be captured efficiently with a more comprehensive range of RFs. In this work, the time-series signals are fed into a deep learning model, the diagnostic model extracts the characteristics of the input signals adaptively, and the final output is utilized to make the final conclusion.

Similar to the CNN model, the DCNN model consists of convolutional layers, pooling layers, activations, batch normalizations, and fully connected layers [29–31] as shown in Figure 1. Convolutional layers could extract features by producing highly focused and continuous information. The dilated convolution kernel (DCK) has a hyperparameter called the dilation rate (DR) that primarily indicates the dilation scale when compared to the normal convolution kernel. With DCK, RF can be dilated to capture different feature components without increasing the size of the convolution kernel. The ensemble model in this study is composed of subclassifiers 1, 2, 3, and 4, which use dilated convolution kernels with dilation rates of 1, 2, 3, and 4, respectively. The following equation expresses the dilation convolution process:

$$C_j^n = \sum_{i \in M_j} X_i^{n-1} \bullet W_{ij}^n + b_j^n, \tag{1}$$

where C_j^n is the j th element of the n th convolutional layer, M_j is the convolution region of the input signal, which varies with DR, as shown in Figure 2, X_i^{n-1} is the previous layer output inside M_j, W_{ij}^n is the weight matrix of the corresponding convolution kernel, and b_j^n is the bias. The activation follows convolutional layers, and the exponential linear unit (ELU) activation function is chosen and denoted as follows:

$$ELU(x) = \begin{cases} x, & if\ x > 0, \\ \alpha * (\exp(x) - 1), & if\ x \le 0', \end{cases} \tag{2}$$

where x is the input of neural network model. The activation is a nonlinear function that transforms input values and enhances the ability of the network to express nonlinearity. Lastly, α is a hyperparameter taken as 1 in this paper.

Pooling layers are used to accomplish sparse processing while assuring a low number of neurons and comprehensive feature representation. Max pooling, mean pooling, and stochastic pooling are all standard pooling methods. In this paper, the max pooling method is used and calculated as follows:

$$M_{m,n,k}^l = max\ \ pooling_{(m,n) \in R_{i,j}^l}^l \left(x_{m,n,k}^{l-1} \right), \tag{3}$$

where $M_{m,n,k}^l$ is the computed value of location (i, j) in the k th feature map of the l th layer after the pooling operation, $R_{i,j}^l$ is the pooling area around the location (i, j), and $x_{m,n,k}^{l-1}$ is the node at the location (m, n) in the pooling domain.

Batch normalization is used to normalize the input data into the network model in order to speed up the training process while preserving as much expressiveness as possible. The following is a description of the batch normalization:

$$\begin{aligned} \mu &= \frac{1}{N_{batch}} \sum_{s=1}^{N_{batch}} x_s, \\ \sigma^2 &= \frac{1}{N_{batch}} \sum_{s=1}^{N_{batch}} \\ (x_s - \mu)^2, \widehat{x}_s &= \frac{x_s - \mu}{\sqrt{\sigma^2 + \varepsilon}}, \\ y_s &= \gamma \widehat{x} + \beta, \end{aligned} \tag{4}$$

where N_{batch} represents the number of small batches of data, x_s represents the s th input, μ and σ^2 represent the mean and variance of small batches of data, respectively, ε represents a

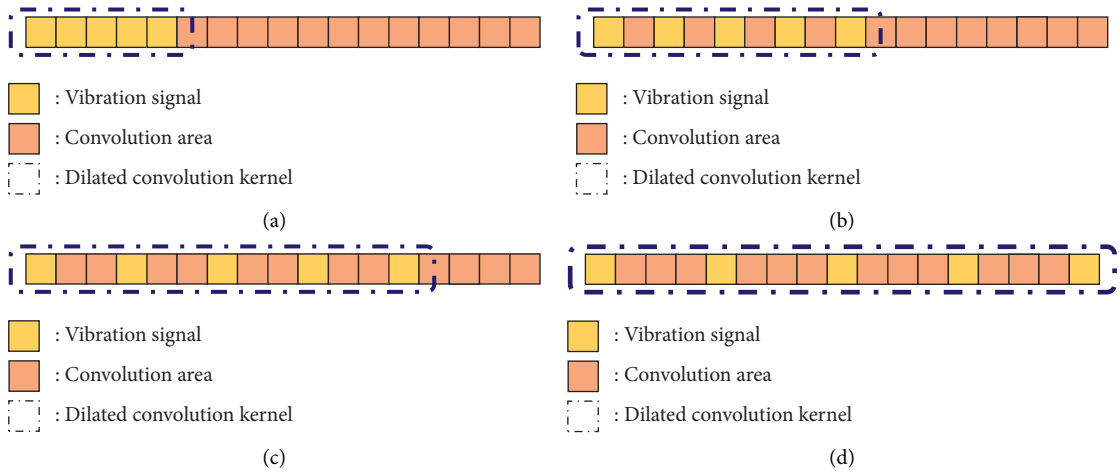


Figure 2: Convolution region with different dilation rates. (a) DR = 1, (b) DR = 2, (c) DR = 3, and (d) DR = 4.

constant close to but greater than 0, $\hat{x}_s$ represents the result of normalizing the data, γ and β define the parameters that can be learned by the network, and y_s represents the s th output of the data after batch normalization.

The fully connected layer performs feature categorization after numerous layered convolutional blocks. It takes place on the utterly connected layer and is used to forecast category labels in the output layer. The following is the equation for the fully connected layer:

$$y^l = w^l x^{l-1} + b^l, \tag{5}$$

where y^l is the output of the l th fully connected layer, x^{l-1} is the one-dimensional feature vector after flattening, w^l is the weight matrix, and b^l is the bias.

2.1.2. Ensemble Learning. Ensemble learning combines several 1D-DCNN subclassifiers into a single prediction model to reduce variance and bias and improve accuracy [32–34]. This study proposes an ensemble 1D-DCNN model approach based on a weighted mechanism as shown in Figure 3. Subclassifiers with different dilation rates initially have the same weights, and the weights are continuously updated based on the outputs of the proposed model. The way of the weighed procedure is shown in the following equation:

$$\hat{y} = \operatorname{argmax} \sum_{j=1}^{n} \left(w_j p_j \right), \tag{6}$$

where w_j is the weights of subclassifiers and p_j is the prediction of subclassifiers, and $\hat{y}$ is the final fault diagnosis decision. Forward and backward propagation mechanisms are present in ensemble model training. Forward propagation is performed by calculating model parameters (subclassifier weights and model weights) and vibration signals to make diagnostic decisions. According to the diagnostic objective, backward propagation finds the most appropriate weights for each neuron and subclassifiers as shown in Figure 4. The cross-entropy loss function [35] and

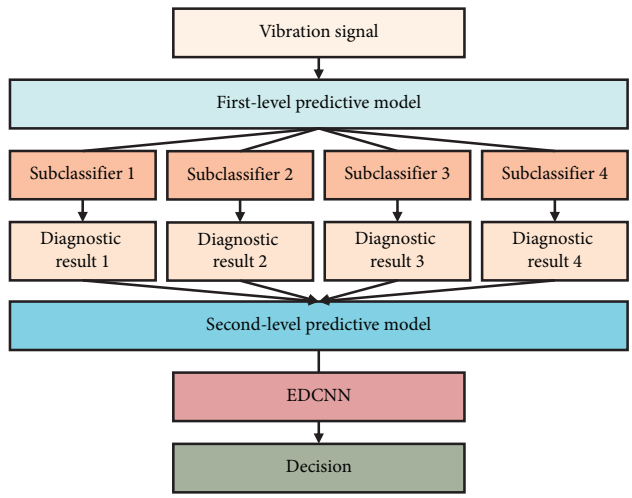


Figure 3: Ensemble learning.

the Adam optimization algorithm [36] play an important role in the backward propagation parameters. The former is a widely used loss function in multiclassification tasks, and the latter effectively minimizes the loss function. With the ensemble learning process, even if a subclassified incorrectly misclassifies faults, associating it with extremely low model weights yields the correct outcome in the final diagnosis of the ensemble model.

2.1.3. Early Stopping Optimization. An optimal diagnostic model with the best generalization performance is generally expected in model training. However, neural network architectures are prone to overfitting. The model may improve as the training and validation subset loss function simultaneously decrease. However, at a certain point in the training process, the loss function of the training subset will continue to decrease while the loss function on the validation subset starts to increase. This is known as overfitting.

To avoid overfitting, early stopping optimization can be used to stop the model training process depending on model

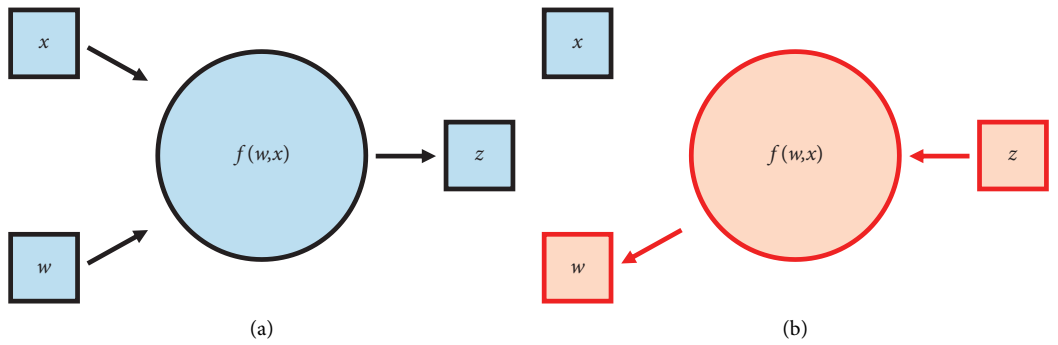


Figure 4: Mechanism of forward and backward propagation. (a) Forward propagation and (b) backward propagation.

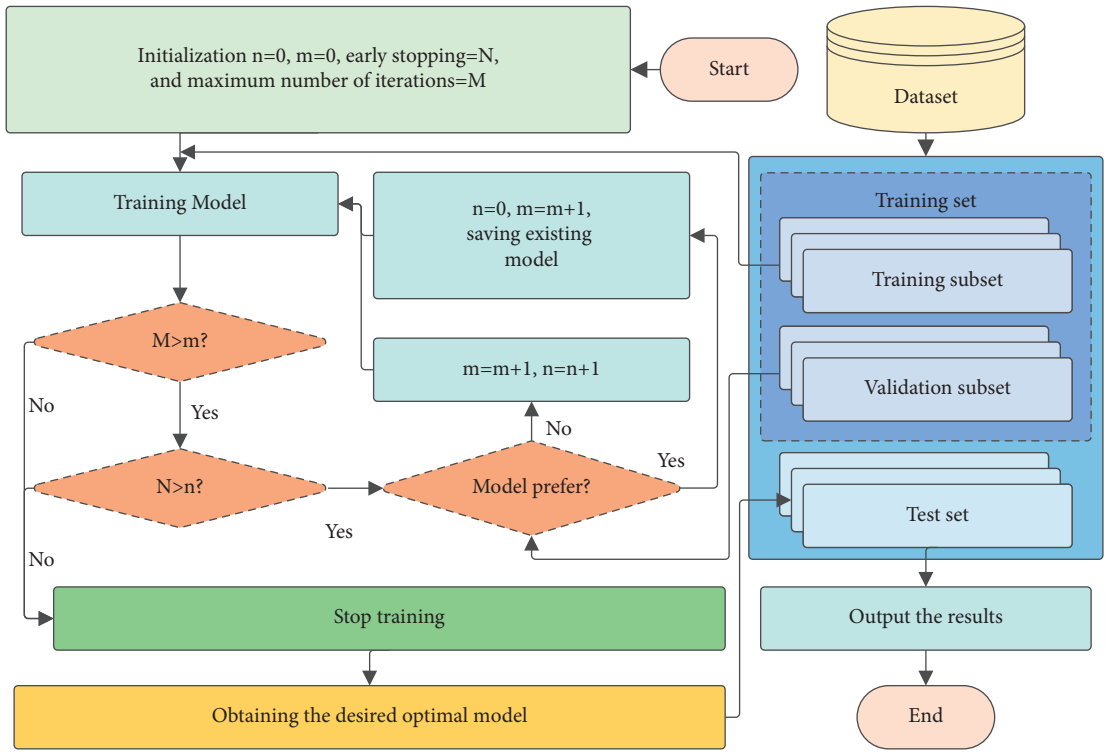


Figure 5: Flowchart of early stopping optimization.

updates as shown in Figure 5. A validating subset loss function-based early stopping optimization is proposed. During each iteration, the model is saved when the loss function of the validation subset decreases. The training process is stopped when the evaluation metric of the model no longer improves, and the number of iterations is within the early stopping optimization. Previous experiments have shown that the results obtained with early stopping do not significantly differ from those obtained with a high number of iterations. However, the computational cost may be several times lower. Early stopping optimization is used in all of the deep learning methods in this research, which is denoted as follows:

$$L_{obt}(t) = \min_{t' < t} L_{va}(t'), \tag{7}$$

where t is the number of iterations, $L_{obt}(t)$ is the validation subset loss function of the obtained validation subset, and $L_{va}(t')$ is the corresponding validation subset loss function at the moment t'.

2.2. A Detailed Structure of the Intelligent Model. EDCNN consists of a collection of four 1D-DCNN subclassifiers. The structural and parameters of the mathematical model were determined by referring to the paper [37, 38]. Apart from the dilation rate, the hyperparameters of each subclassifier in the proposed model are the same as shown in Figure 6 and listed in Table 1. Blocks 1, 2, and 3 of the subclassifiers serve as feature extractors, while Block 4 serves as the decision maker. Block 1, Block 2, and Block 3 are four-layer dilation CNNs, each containing a dilation convolution layer, a pooling layer, activation, and batch normalization. There are four channels in the first dilation convolution and pooling layer, eight channels in the second dilation convolution and pooling layer, and 16 channels in the third dilation

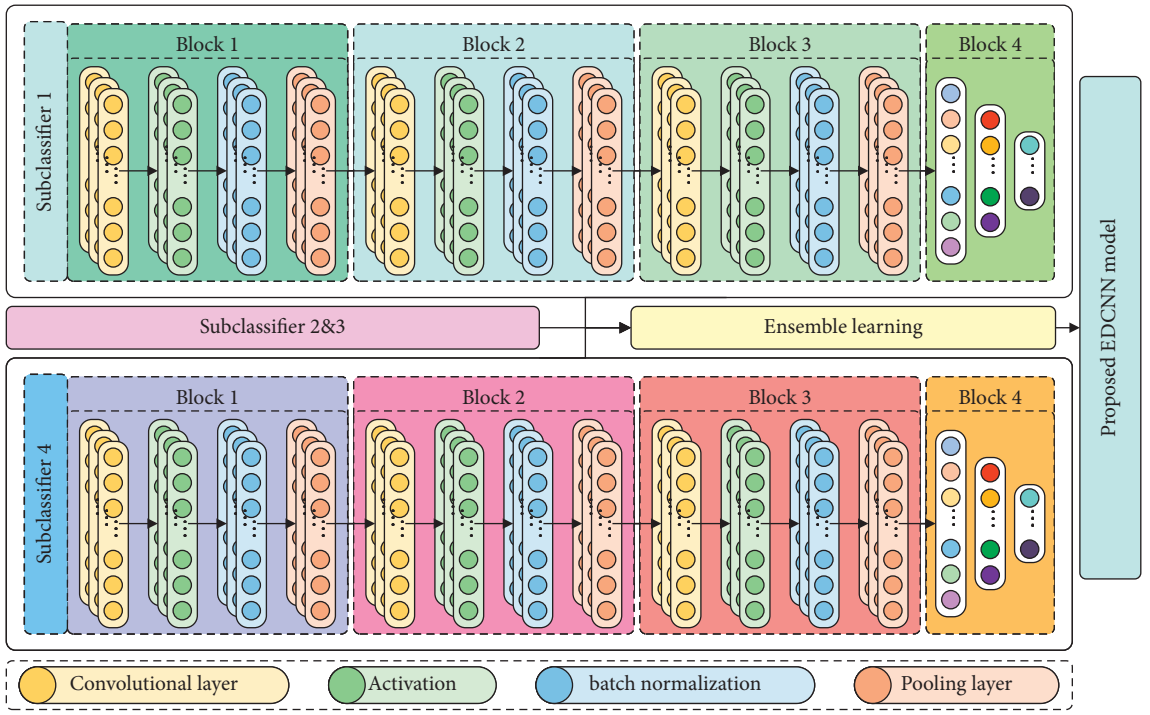


Figure 6: Ensemble dilated convolutional neural network model.

Table 1: Parameters of the dilated convolutional neural network model.

Object			Hyperparameter settings
Feature extractor	Block 1	Convolutional layer #1	Number of channels: 4, kernel width: 5, stride: 2
		Pooling layer #1	Kernel width: 2, stride: 2
	Block 2	Convolutional layer #2	Number of channels: 8, kernel width: 5, stride: 2
		Pooling layer #2	Kernel width: 2, stride: 2
	Block 3	Convolutional layer #3	Number of channels: 16, kernel width: 5, stride: 2
		Pooling layer #3	Kernel width: 2, stride: 2
Decision maker	Block 4	Fully connected layer #1	Network width: input dimension
		Fully connected layer #2	Network width: 128
		Fully connected layer #3	Network width: number of fault category
Early stopping			5
Maximum number of iterations			100
Learning rate			10^{-4}
Small batch size			100

convolution and pooling layer. The convolution kernel has a valid length of 5 with a stride of 1, and the pooling kernel has a valid length of 2 with a stride of 2. Block 4 is a three-layer fully connected neural network with the first layer (input layer) dimension as a flattened input dimension, the second layer (hidden layer) dimension as 128, and the third layer (output layer) dimension as a fault category. In addition, to save model training time and model convergence performance, this study sets the early stop to 5, the maximum number of iterations to 100, the learning rate to 10^{-4}, and the small batch size to 100.

2.3. Proposed EDCNN-Based Intelligent Fault Diagnosis Scheme. A new adaptive deep learning fault diagnosis scheme is proposed based on the advantages of the proposed EDCNN method. The flowchart of this scheme is shown in Figure 7, and the specific steps are as follows:

Step 1: Signal acquisition. Acceleration sensors are used to collect vibration acceleration signals from rotating machinery and divide them into a training set (which includes a training subset and a validation subset) and a test set.

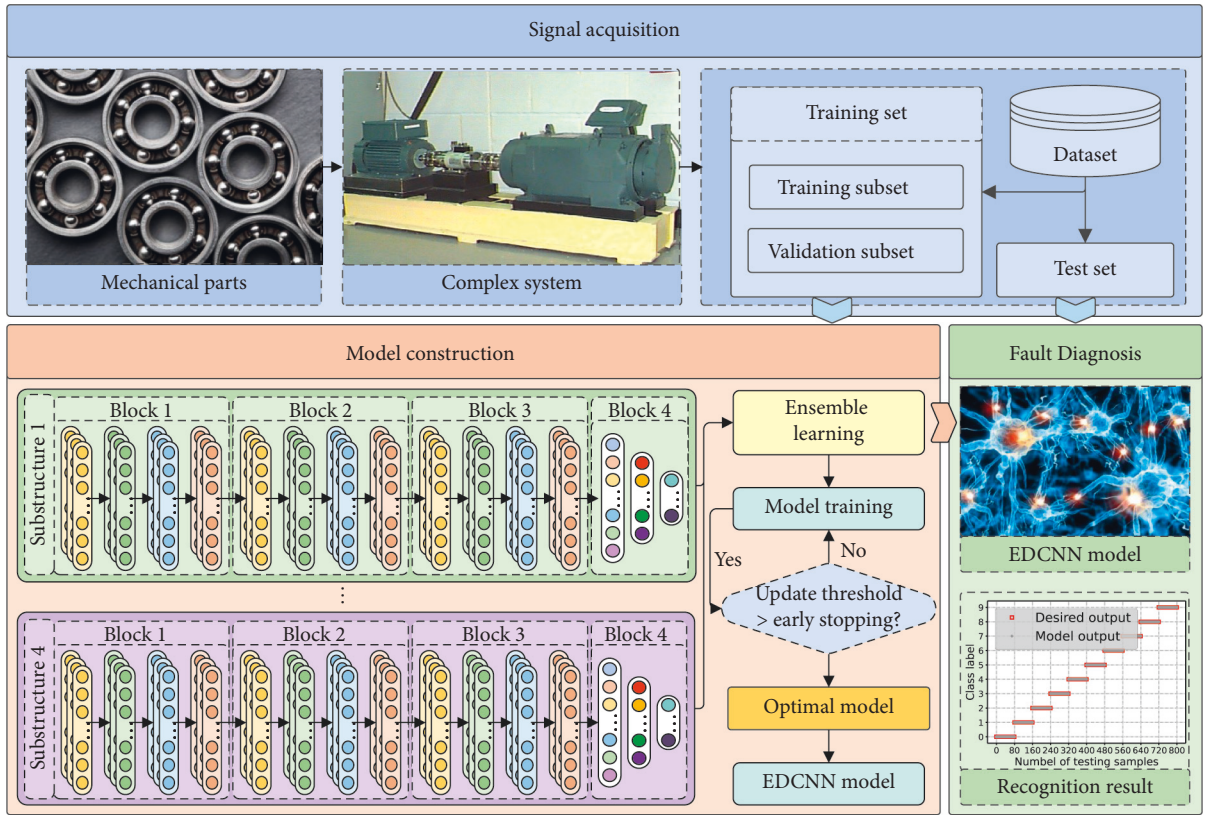


Figure 7: Proposed intelligent fault diagnosis scheme.

Step 2: Model construction. The EDCNN model is built using the training set as the input. The training subset is utilized for initial prediction model training, and the validation subset is used to stop model training at the proper moment in conjunction with early stop optimization.

Step 3: Fault diagnosis. The testing set is input to the prediction model for achieving end-to-end intelligent fault diagnosis.

3. Experimental Study

In this section, to evaluate the advanced diagnostic performance of the proposed fault diagnosis model, several experiments are carried out on Case Western Reserve University (CWRU) [39] rolling bearing and wind turbine dataset. Data acquisition is performed through a sliding time window, where the window size and overlapping structure are 1024 and 128, respectively. The dataset is composed of 800 sets of signals for each fault category. The training and test sets are divided by the dataset by a ratio of 0.8 : 0.2, and the training and validation subsets are divided by the training set by a ratio of 0.8 : 0.2. Moreover, several most advanced methods are selected for comparative analysis. Finally, for this experiment, the deep learning library PyTorch (version 1.9) was utilized, the suggested model was evaluated and implemented in Python (version 3.7), and the experiment was repeated ten times to eliminate random effects.

3.1. Comparative Methods. The following diagnosis methods are implemented for comparison to verify the superiority of the proposed model in fault diagnosis (the proposed EDCNN fault diagnosis method is abbreviated as FD-6):

FD-1: FD-1 is a fault diagnosis method based on the modified support vector machine, which employs the multiscale permutation entropy, linear local tangent space alignment, and least square support vector machine algorithms. According to reference [40], the settings are configured.

FD-2: FD-2 is a fault diagnosis method based on an artificial neural network. The ANN simulates the structure and function of neural networks in the brain, using mathematical models to model the activity of neurons. In this study, a three-layer ANN was used.

FD-3: FD-3 is a fault diagnosis method based on an improved recurrent neural network. The improved method, called the gated recurrent unit, can extract time-series features automatically. The training efficiency is significantly higher than that of long short-term memory due to the unique individual gate mechanism.

FD-4: FD-4 is a fault diagnosis method based on the CNN, which can classify input data according to its hierarchical structure in terms of shifted variables using representational learning. The model used is the sub-classified 1 mentioned above.

Total sample = P + N		Predicted condition	
		Positive (P)	Negative (N)
Actual condition	Positive (P)	True positive (TP)	True negative (TN)
	Negative (N)	False positive (FP)	False negative (FN)

Figure 8: Meaning of TP, TN, FP, and FN.

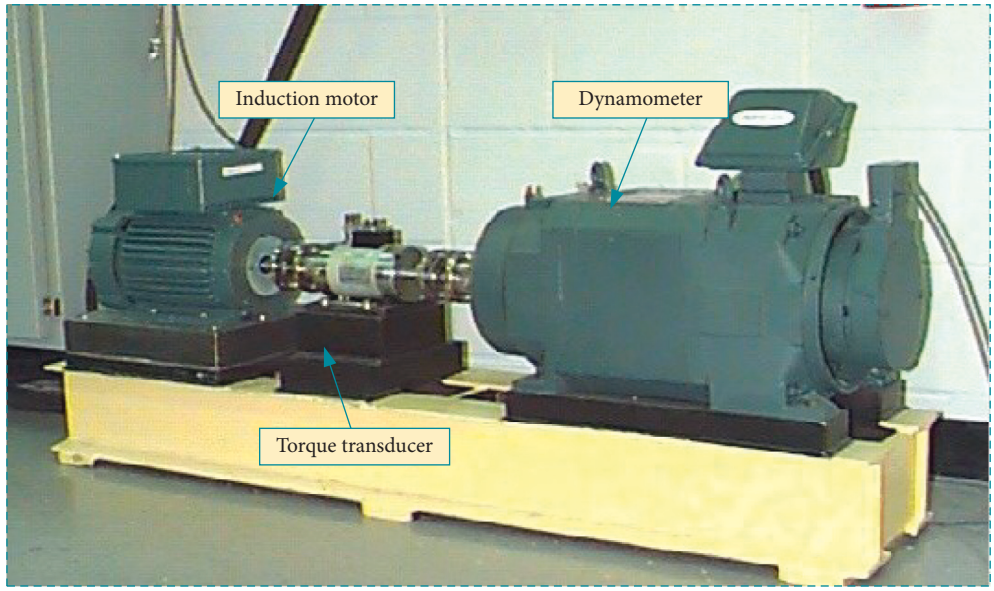


Figure 9: Experimental platform used to obtain the CWRU bearing data.

FD-5: FD-5 is a fault diagnosis method based on the DCNN. The DCNN has a wider receptive field than the normal convolutional neural network and may capture longer dependencies. The model utilized is subclassifier 2 from the previous section.

3.2. Evaluation Metric. To quantify the performance of the suggested intelligent fault diagnosis scheme, evaluation metrics were devised. The F score [41], a composite metric that combines precision and recall, is used as the evaluation criterion as follows:

$$\begin{aligned} \text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ \text{Recall} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{F score} &= \frac{\left(1 + \beta^2\right) * \text{Precision} * \text{Recall}}{\beta^2 * \text{Precision} + \text{Recall}}, \end{aligned} \tag{8}$$

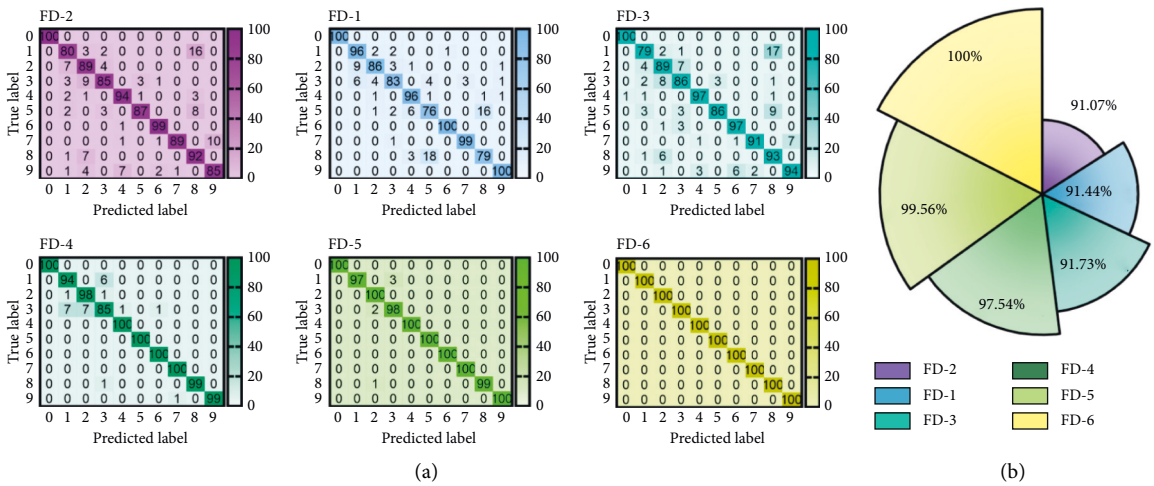
where TP denotes true positive, TN denotes true negative, FP indicates false positive, and FN indicates false negative. Their respective roles are shown in Figure 8; β^2 denotes the weights of precision and recall in the evaluation metrics. Here, β^2 is taken as 1, indicating that equal importance is given to precision and recall. Therefore, it is called the F1 score.

3.3. Case Study 1: Experimental Analysis with the CWRU Dataset

3.3.1. Dataset Description. The CWRU dataset is a remarkable and representative rolling bearing fault diagnostic dataset that has been utilized in many studies to validate condition monitoring and fault diagnosis methods for rotating motors [42, 44]. It is used as a benchmark in this work for experimental investigations to verify the advantages of the proposed EDCNN-based fault diagnosis method. The CWRU experimental platform is shown in Figure 9, which mainly consists of an induction motor, a torque transducer, a dynamometer, and an electronic controller. The vibration signals were collected from a faulty bearing mounted at the

Table 2: Description of ten working states of the CWRU experimental platform.

Status	Fault diameter	Abbreviation	Fault type	Label	Sample length	Dataset (training subset/validation subset/test set)
Normal	—	NOR	—	0	1024	512/128/160
Ball fault	0.007	B007	Single	1	1024	512/128/160
	0.014	B014	Single	2	1024	512/128/160
	0.021	B021	Single	3	1024	512/128/160
Inner race fault	0.007	IR007	Single	4	1024	512/128/160
	0.014	IR014	Single	5	1024	512/128/160
	0.021	IR021	Single	6	1024	512/128/160
Outer race fault	0.007	OR007	Single	7	1024	512/128/160
	0.014	OR014	Single	8	1024	512/128/160
	0.021	OR021	Single	9	1024	512/128/160



(a) (b)

Figure 10: Diagnostic performance of different methods for the CWRU dataset: (a) confusion matrix and (b) the mean F1 scores.

end motor fan and sampled at 12 kHz. This dataset studied ten single fault conditions corresponding to standard and different fault diameters for ball faults, inner race faults, and outer race faults as shown in Table 2.

3.3.2. Experimental Validation. The diagnostic performance in the CWRU experimental platform test set is depicted in Figure 10. In the first set of experiments, only the FD-6 indicated correctly recognized all fault kinds, whereas the remainder of the FD-1, FD-2, FD-3, FD-4, and FD-5 were misdiagnosed, as seen in Figure 10(a). In the repeated experiments, the average F1 scores of each diagnostic model are represented in Figure 10(b). The F1 scores of FD-1, FD-2, and FD-3 are inferior to CNN-based approaches (FD-4, FD-5, and FD-6). The reason for this is that, as compared to FD-1, FD-2, and FD-3, CNN-based fault detection methods have more powerful feature extraction capabilities for identifying various types of faults. Furthermore, the F1 scores of FD-6 are 2.46% and 0.44% higher in the CNN-based fault diagnosis method than those of FD-4 and FD-5, respectively. In comparison to the limited pattern recognition capability of other methods, the suggested FD-6 diagnostic model correctly identifies all health states in the benchmark experiments.

3.3.3. Robustness Analysis. To simulate fault diagnosis scenarios under complex operating scenarios, Gaussian white noise of 4 dB, 2 dB, 0 dB, −2 dB, and −4 dB is added to the original signals, respectively. The comparison approaches and the suggested EDCNN method were tested for robustness in the presence of additional noise.

The robustness analysis results of six fault diagnosis methods are shown in Figure 11. It can be concluded that in the presence of additive noise, the classification performance of the diagnostic model deteriorates as the signal-to-noise ratio decreases. The CNN-based fault diagnosis model still outperforms the FD-1, FD-2, and FD-3 diagnostic approaches. The F1 score of FD-6 in the CNN-based diagnostic model is 98.27%, which is better than the F1 scores of FD-4 and FD-5, which are 90.71% and 95.17%, respectively. Furthermore, the presence of a dilated convolution mechanism improves the accuracy and robustness of the fault identification effect. In both sets of CWRU experiments, the suggested FD-6 model obtained the best diagnostic results, ensemble multiple dilated convolutional neural networks, and improved diagnostic performance and robustness under multifeature map comprehensive decision, demonstrating the improved diagnostic performance, and robustness of the proposed model based on the addition of multifeature maps.

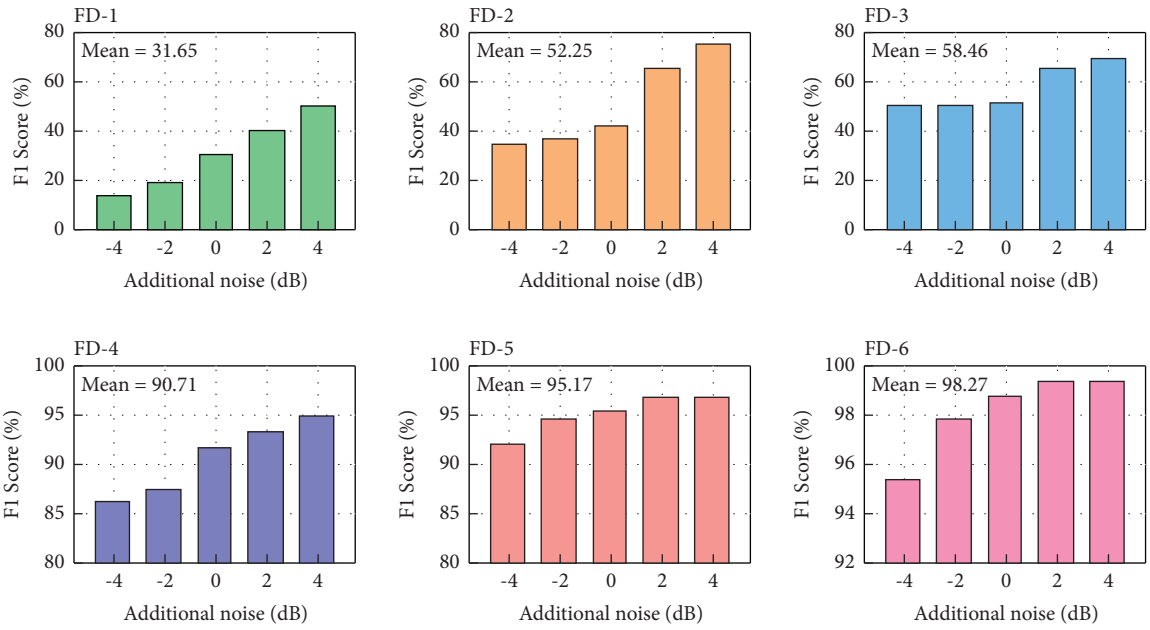
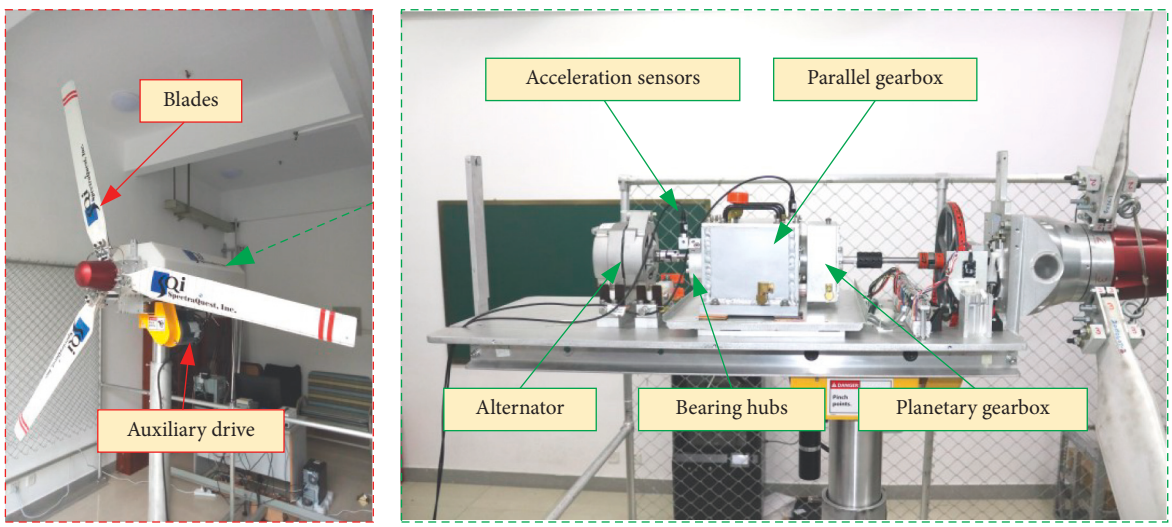


Figure 11: Comparative learning in CWRU tasks with different additional noise levels.



(a) (b)

Figure 12: Experimental platform used to obtain wind turbine simulator data.

3.4. Case Study 2: Experimental Analysis with the Wind Turbine Dataset

3.4.1. Dataset Description. Wind energy is widely considered to be the most commercially promising and environmentally friendly energy source; however, different harsh working conditions make wind turbines more susceptible to failure. An experimental platform of a wind turbine simulator was created and wind turbine datasets were collected in this work for the aim of wind turbine fault diagnostics. The wind turbine simulation experimental platform consisting of three fan blades, an auxiliary drive, a planetary gearbox, bearing hubs, and an alternator is shown in Figure 12. The vibration signals are sampled from the faulty bearing and the faulty gear phone at one end of the gearbox at 12.8 kHz. Nine health states were investigated using the data collected, including normal, several single fault types, and several compound fault types, as shown in Table 3. Compound faults include mutual interaction between several single fault pulses, degrading diagnostic method identification performance.

3.4.2. Experimental Validation. Compound faults in wind turbines occur concurrently and are coupled by multiple types of faults, posing a significant difficulty for feature

Table 3: Description of nine working states of the wind turbine simulator experimental platform.

Status	Abbreviation	Fault type	Label	Sample length	Dataset (training subset/validation subset/test set)
Normal	NOR	—	0	1024	512/128/160
Inner race fault	IRF	Single	1	1024	512/128/160
Outer race fault	ORF	Single	2	1024	512/128/160
Ball fault	BF	Single	3	1024	512/128/160
Gear wear fault	GWF	Single	4	1024	512/128/160
Tooth broken fault	TBF	Single	5	1024	512/128/160
Inner race outer race fault	IOF	Compound	6	1024	512/128/160
Outer race gear wear fault	OGF	Compound	7	1024	512/128/160
Outer race tooth broke fault	OTF	Compound	8	1024	512/128/160

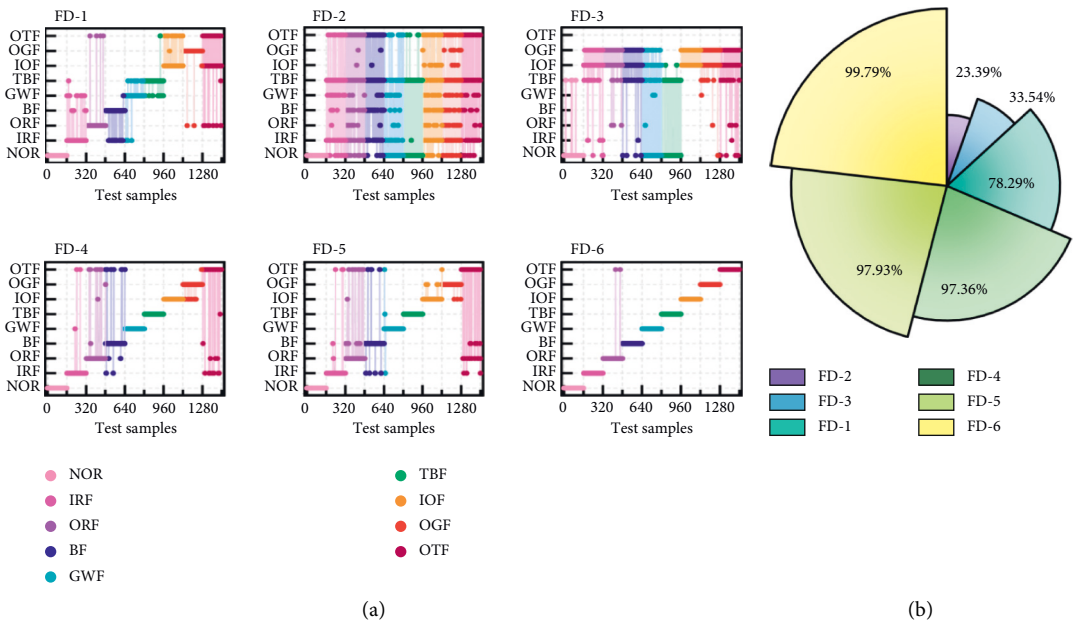


Figure 13: Diagnostic performance of the wind turbine simulation experimental platform: (a) recognition results for a given experiment and (b) the mean F1 scores of different methods.

extraction, and pattern recognition in fault diagnosis. The fault diagnosis results of the wind turbine experimental platform test set are illustrated in Figure 13. All diagnostic approaches were misdiagnosed in the initial set of wind turbine experiments, as shown in Figure 13(a). The identification results of the remaining approaches demonstrate large-scale misclassification, with the exception of the proposed FD-6 diagnostic method, which misidentifies two fault types. The comprehensive performance of each diagnostic method for repeated experiments on the wind turbine dataset is shown in Figure 13(b). None of the non-CNN-based mathematical models are adequate for diagnosing compound faults. Due to its great feature capability capacity, the suggested FD-6 model is able to retain good recognition performance while dealing with compound fault diagnostic scenarios and is 2.43% and 1.86% ahead of the relatively decent FD-4 and FD-5 in terms of F1 scores.

3.4.3. Robustness Analysis. Likewise, the additional noise was applied to the wind turbine dataset. For different health conditions, 4 dB, 2 dB, 0 dB, −2 dB, and −4 dB additive noise is applied to the vibration signal. The recognition performance of the comparative learning is shown in Figure 14. Obviously, diagnosing compound faults in wind turbines is more difficult than diagnosing single faults in bearings. The performance of the comparison method is still inadequate. Due to CNN's outstanding feature extraction capacity, FD-4, FD-5, and FD-6 outperform FD-1, FD-2, and FD-3. The wider RF of FD-5 and FD-6 feature extractors, which can extract correlation features between longer signals, results in 0.66% and 6.69% better mean performance than FD-4. For the advantage of ensemble learning, FD-6 may gain more effective fault discrimination information in multi-feature maps, resulting in superior classification performance.

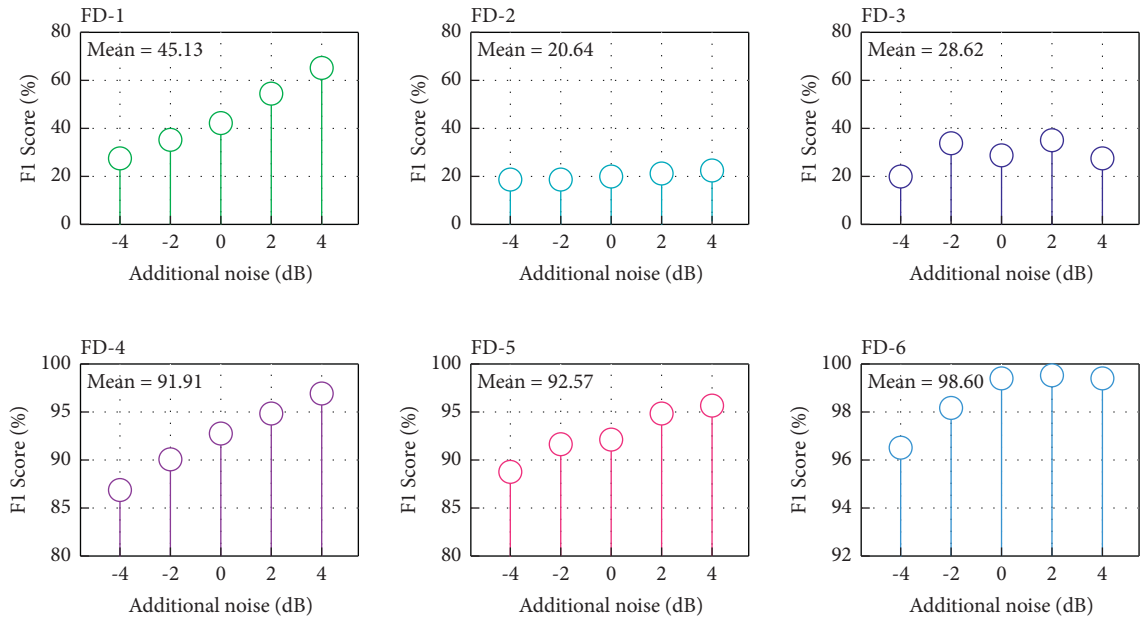


Figure 14: Comparative learning in wind turbine tasks with different additional noise levels.

4. Conclusions

In this paper, an intelligent fault diagnosis approach for rotating machinery is proposed using ensemble dilated convolutional neural networks (EDCNN). On the CWRU bearing dataset and the wind turbine dataset, the proposed approach is examined and validated. The following conclusions can be drawn:

(1) In both the bearing and wind turbine datasets, the proposed EDCNN adaptive fault diagnostic approach accurately identifies all single and compound faults.

(2) In comparison to advanced fault diagnosis methods (such as MSVM, ANN, GRU, CNN, and DCNN), the suggested EDCNN method can identify all health states correctly and reliably.

(3) The robustness analysis results indicate that the suggested EDCNN approach can perform fault diagnosis of rotating equipment in complex situations with stronger feature learning and feature extraction capabilities.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant no. 51775114); the Fujian Provincial Science and Technology Major Special Project (Grant no. 2021HZ024006); and the Fujian Provincial High-End Equipment Manufacturing Collaborative Innovation (Center no. 2021-C-275).

References

[1] R. Liu, B. Yang, E. Zio, and X. Chen, "Artificial intelligence for fault diagnosis of rotating machinery: a review," *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.

[2] T. Yan, Y. Lei, N. Li, X. Si, L. Pintelon, and R. Dewil, "Online joint replacement-order optimization driven by a nonlinear ensemble remaining useful life prediction method," *Mechanical Systems and Signal Processing*, vol. 173, Article ID 109053, 2022.

[3] Z. Wang, L. Yao, and Y. Cai, "Rolling bearing fault diagnosis using generalized refined composite multiscale sample entropy and optimized support vector machine," *Measurement*, vol. 156, Article ID 107574, 2020.

[4] Z. Wang, L. Yao, Y. Cai, and J. Zhang, "Mahalanobis semi-supervised mapping and beetle antennae search based support vector machine for wind turbine rolling bearings fault diagnosis," *Renewable Energy*, vol. 155, pp. 1312–1327, 2020.

[5] L. Duan, Y. Wang, J. Wang, L. Zhang, and J. Chen, "Undecimated lifting wavelet packet transform with boundary treatment for machinery incipient fault diagnosis," *Shock and Vibration*, vol. 20169 pages, Article ID 9792807, 2016.

[6] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Universal domain adaptation in fault diagnostics with hybrid weighted deep adversarial Learning," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 7957–7967, 2021.

[7] J. Li, H. Wang, X. Wang, and Y. Zhang, "Rolling bearing fault diagnosis based on improved adaptive parameterless empirical wavelet transform and sparse denoising," *Measurement*, vol. 152, Article ID 107392, 2020.

[8] J. Wang, C. Zhan, S. Li, Q. Zhao, J. Liu, and Z. Xie, "Adaptive variational mode decomposition based on Archimedes

optimization algorithm and its application to bearing fault diagnosis," *Measurement*, vol. 191, Article ID 110798, 2022.

[9] X. Jiang, J. Wang, J. Shi, C. Shen, W. Huang, and Z. Zhu, "A coarse-to-fine decomposing strategy of VMD for extraction of weak repetitive transients in fault diagnosis of rotating machines," *Mechanical Systems and Signal Processing*, vol. 116, pp. 668–692, 2019.

[10] J. Gu and Y. Peng, "An improved complementary ensemble empirical mode decomposition method and its application in rolling bearing fault diagnosis," *Digital Signal Processing*, vol. 113, Article ID 103050, 2021.

[11] A. Mejia-Barron, M. Valtierra-Rodriguez, D. Granados-Lieberman, J. C. Olivares-Galvan, and R. Escarela-Perez, "The application of EMD-based methods for diagnosis of winding faults in a transformer using transient and steady state currents," *Measurement*, vol. 117, pp. 371–379, 2018.

[12] Z. Shen, X. Chen, X. Zhang, and Z. He, "A novel intelligent gear fault diagnosis model based on EMD and multi-class TSVM," *Measurement*, vol. 45, no. 1, pp. 30–40, 2012.

[13] K. Shao, W. Fu, J. Tan, and K. Wang, "Coordinated approach fusing time-shift multiscale dispersion entropy and vibrational Harris hawks optimization-based SVM for fault diagnosis of rolling bearing," *Measurement*, vol. 173, Article ID 108580, 2021.

[14] Z. Wang, L. Yao, G. Chen, and J. Ding, "Modified multiscale weighted permutation entropy and optimized support vector machine method for rolling bearing fault diagnosis with complex signals," *ISA Transactions*, vol. 114, pp. 470–484, 2021.

[15] M. Y. Cho and T. T. Hoang, "Feature selection and parameters optimization of SVM using particle swarm optimization for fault classification in power distribution systems," *Computational Intelligence and Neuroscience*, vol. 2017, Article ID 4135465, 9 pages, 2017.

[16] J. Liu, C. Pan, F. Lei, D. Hu, and H. Zuo, "Fault prediction of bearings based on LSTM and statistical process analysis," *Reliability Engineering & System Safety*, vol. 214, Article ID 107646, 2021.

[17] S. Gao, Z. Jiang, and S. Liu, "An approach to intelligent fault diagnosis of cryocooler using time-frequency image and CNN," *Computational Intelligence and Neuroscience*, vol. 2022, 8 pages, Article ID 1754726, 2022.

[18] J. Liu, H. Yang, J. He, Z. Sheng, and S. Chen, "Unbalanced fault diagnosis based on an invariant temporal-spatial attention fusion network," *Computational Intelligence and Neuroscience*, vol. 202215 pages, Article ID 1875011, 2022.

[19] R. Saeed, A. Galybin, and V. Popov, "3D fluid–structure modelling and vibration analysis for fault diagnosis of Francis turbine using multiple ANN and multiple ANFIS," *Mechanical Systems and Signal Processing*, vol. 34, no. 1-2, pp. 259–276, 2013.

[20] Q. Ye and C. Liu, "An unsupervised deep feature learning model based on parallel convolutional autoencoder for intelligent fault diagnosis of main reducer," *Computational Intelligence and Neuroscience*, vol. 2022, 12 pages, Article ID 8922656, 2021.

[21] S. Moosavi, A. Djerdir, Y. Ait-Amirat, and D. Khaburi, "ANN based fault diagnosis of permanent magnet synchronous motor under stator winding shorted turn," *Electric Power Systems Research*, vol. 125, pp. 67–82, 2015.

[22] X. Mao, F. Zhang, G. Wang, Y. Chu, and K. Yuan, "Semi-random subspace with Bi-GRU: fusing statistical and deep representation features for bearing fault diagnosis," *Measurement*, vol. 173, Article ID 108603, 2021.

[23] Y. Wu and X. Ma, "A hybrid LSTM-KLD approach to condition monitoring of operational wind turbines," *Renewable Energy*, vol. 181, pp. 554–566, 2022.

[24] J. Liu, M. Zhang, H. Wang, W. Zhao, and Y. Liu, "Sensor fault detection and diagnosis method for AHU using 1-D CNN and clustering analysis," *Computational Intelligence and Neuroscience*, vol. 2019, Article ID 5367217, 20 pages, 2019.

[25] Z. Chen, A. Mauricio, W. Li, and K. Gryllias, "A deep learning method for bearing fault diagnosis based on cyclic spectral coherence and convolutional neural networks," *Mechanical Systems and Signal Processing*, vol. 140, Article ID 106683, 2020.

[26] S. Plakias and Y. S. Boutalis, "Fault detection and identification of rolling element bearings with attentive dense CNN," *Neurocomputing*, vol. 405, pp. 208–217, 2020.

[27] Y. Guo, Y. Zhou, and Z. Zhang, "Fault diagnosis of multi-channel data by the CNN with the multilinear principal component analysis," *Measurement*, vol. 171, Article ID 108513, 2021.

[28] T. Han, L. Zhang, Z. Yin, and A. C. Tan, "Rolling bearing fault diagnosis with combined convolutional neural networks and support vector machine," *Measurement*, vol. 177, Article ID 109022, 2021.

[29] S. Liu, P. Minne, M. Lulić, J. Li, and E. Gruyaert, "Implementation and validation of Dewar's particle packing model for recycled concrete aggregates," *Construction and Building Materials*, vol. 294, Article ID 123429, 2021.

[30] W. Zhang, X. Li, H. Ma, Z. Luo, and X. Li, "Federated learning for machinery fault diagnosis with dynamic validation and self-supervision," *Knowledge-Based Systems*, vol. 213, Article ID 106679, 2021.

[31] V. Kumar, R. S. Singh, and Y. Dua, "Morphologically dilated convolutional neural network for hyperspectral image classification," *Signal Processing: Image Communication*, vol. 101, Article ID 116549, 2022.

[32] M. Alam, F. Abid, C. Guangpei, and L. Yunrong, "Social media sentiment analysis through parallel dilated convolutional neural network for smart city applications," *Computer Communications*, vol. 154, pp. 129–137, 2020.

[33] Y. Liu, X. Zhang, Y. Gao, T. Qu, and Y. Shi, "Improved CNN method for crop pest identification based on transfer learning," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 9709648, 8 pages, 2022.

[34] D. Yang, N. Hou, J. Lu, and D. Ji, "Novel leakage detection by ensemble 1DCNN-VAPSO-SVM in oil and gas pipeline systems," *Applied Soft Computing*, vol. 115, Article ID 108212, 2021.

[35] Z. Wang, H. Huang, and Y. Wang, "Fault diagnosis of planetary gearbox using multi-criteria feature selection and heterogeneous ensemble learning classification," *Measurement*, vol. 173, Article ID 108654, 2021.

[36] S. Padhy, S. Dash, S. Routray, S. Ahmad, J. Nazeer, and A. Alam, "IoT-based hybrid ensemble machine learning model for efficient diabetes mellitus prediction," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 2389636, 11 pages, 2022.

[37] J. Xu, Y. Liu, J. Liu, and Z. Qu, "Effectiveness of English online learning based on deep learning," *Computational Intelligence and Neuroscience*, vol. 202210 pages, Article ID 1310194, 2022.

[38] T. S. Qaid, H. Mazaar, M. Y. H. Al-Shamri, M. S. Alqahtani, A. A. Raweh, and W. Alakwaa, "Hybrid deep-learning and machine-learning models for predicting COVID-19," *Computational Intelligence and Neuroscience*, vol. 2021, 11 pages, Article ID 9996737, 2021.

[39] A. Dibaj, M. M. Ettefagh, R. Hassannejad, and M. B. Ehghaghi, "A hybrid fine-tuned VMD and CNN scheme for untrained compound fault diagnosis of rotating machinery with unequal-severity faults," *Expert Systems with Applications*, vol. 167, Article ID 114094, 2021.

[40] Z. He, H. Shao, X. Zhong, and X. Zhao, "Ensemble transfer CNNs driven by multi-channel signals for fault diagnosis of rotating machinery cross working conditions," *Knowledge-Based Systems*, vol. 207, Article ID 106396, 2020.

[41] *Case Western Reserve University Bearing Data Center Website*, http://csegroups.case.edu/%20bearingdatacenter/pages/download-data-file.

[42] Y. Ye, Y. Zhang, Q. Wang, Z. Wang, Z. Teng, and H. Zhang, "Fault diagnosis of high-speed train suspension systems using multiscale permutation entropy and linear local tangent space alignment," *Mechanical Systems and Signal Processing*, vol. 138, Article ID 106565, 2019.

[43] L. Gao, D. Li, L. Yao, and Y. Gao, "Sensor drift fault diagnosis for chiller system using deep recurrent canonical correlation analysis and k-nearest neighbor classifier," *ISA Transactions*, vol. 122, pp. 232–246, 2022.

[44] H. Lv, J. Chen, T. Zhang, R. Hou, T. Pan, and Z. Zhou, "SDA: regularization with Cut-Flip and Mix-Normal for machinery fault diagnosis under small dataset," *ISA Transactions*, vol. 111, pp. 337–349, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 9173249, 8 pages
https://doi.org/10.1155/2022/9173249

Research Article

Adaptive Fuzzy Controller Design for Uncertain Robotic Manipulators Subject to Nonlinear Dead Zone Inputs

Hua Zhang

Zhengzhou Preschool Education College, Zhengzhou 450099, China

Correspondence should be addressed to Hua Zhang; zhanghua27@163.com

Received 29 July 2022; Revised 24 August 2022; Accepted 29 August 2022; Published 20 September 2022

Academic Editor: Heng Liu

This paper comes up with the adaptive fuzzy scheme for multi-degree robotic manipulators with dead zone. Fuzzy logic system is employed to approximate unknown nonlinear functions and input nonlinear functions with dead zones that are introduced. The adaptive fuzzy technique is utilized to deal with the problems of multi-joint robotic manipulators with unknown nonlinear functions in dead zones. Based on Lyapunov criterion, all the states and signals involved in the system are maintained in a certain bounded region and the tracking error converges to a small domain of origin. Finally, a simulation example of robotic manipulators with two joints is structured to confirm the practicability of the raised scheme.

1. Introduction

In recent decades, the robotic manipulators are not only the most widely used automated mechanical device in the field of robotics technology but also have an extensive range of applications in daily life. For example, in the medical field, robotic manipulators can easily adapt to the existing operating environments to offer better control for surgeons and ultimately improve surgical outcomes. Besides, the emergence of robotic manipulators can take the place of heavy labor to achieve mechanization and automation of production and operate in harmful environments to ensure personal safety. With the exception of the abovementioned, it can also be applied in industrial manufacturing, medical care, entertainment services, military, semiconductor manufacturing, and so on. Therefore, due to its extensive application and practical value, the research of robotic manipulators is more meaningful.

Since the robotic manipulator is a relatively sophisticated multiple-input-multiple-output (MIMO) nonlinear kinematics system with time-varying, highly coupling, and nonlinear dynamic characteristics, the control of robotic manipulators has been studied extensively and a number of methods have been proposed. In [1], for a redundant manipulator, an adaptive PID fast terminal sliding mode was proposed. A sliding mode and a state observer-based controller were designed to make flexible robotic manipulators perform remote rehabilitation missions in [2]. There are also some other approaches dealing with the robotic manipulators such as sliding mode control [1–6], robust control [3, 7], adaptive fuzzy [8–12], PID controller [13–15]. However, there is a prerequisite for the literature discussed above whose models are all known. To the best of our knowledge, the model equation of the robotic manipulators is uncertain and difficult to establish accurately due to some unavoidable uncertainties such as parameter perturbation, external disturbances, and unmodeled dynamics. Compared with other existing processing methods, adaptive fuzzy control has been widely studied due to its good approximation effect for dealing with a class of unknown and uncertain systems. In [16], the adaptive fuzzy controller was designed to settle multilateral cooperative teleoperation of multiple robotic manipulators under random network-induced delays. A decentralized adaptive fuzzy scheme was proposed for robotic manipulators combining the genetic algorithm and the gradient method in [17]. The adaptive fuzzy control was raised for a robotic manipulator with asymptotic tracking performance in [18]. However, it is noteworthy that the discussion of the aforementioned literature [17, 18] is not involved in the dead zone.

In fact, dead zone is ubiquitous in many industrial systems which affect system operation, especially in the robotic manipulator's system. In [19], a robot manipulator with the dead zone was solved by using the adaptive neural network tracking control method. Backstepping funnel control was used to handle the prescribed performance of robotic manipulators with unknown dead zones in [20]. For flexible manipulators with input dead zones, adaptive neural command filter tracking control was proposed in [21]. However, the models studied in [19–21] were unknown and uncertain, which makes their research methods more consistent with the actual system.

Motivated by the above discussion, the proposed adaptive fuzzy controller for multi-degree robotic manipulators with dead zones will be an interesting and challenging topic for us. In comparison with the available literature, the main contributions of this work are listed as follows:

(1) The proposed adaptive fuzzy controller design can be utilized to approximate a large class of uncertain robotic manipulator systems and unknown nonlinear functions under actual conditions. In addition, even though the system model is fully unknown, our controller still works effectively.

(2) The uncertain multi-joint robotic manipulator system with dead-zone considered is addressed by the adaptive fuzzy control approach, and the proposed method is valid for other high-order nonlinear systems with dead-zone inputs.

In the end, this paper is settled as below. Section 2 brings in fuzzy logic system. The models of multi-joint robotic manipulators and input nonlinearity are raised in Section 3. Section 4 provides a fuzzy controller in the light of the universal approximation criterion for the sake of dealing with robotic manipulators. In Section 5, a simulation case of robotic manipulators with two joints is offered to confirm the feasibility of the proposed method. Besides, conclusions are given to the end.

2. Description of the Fuzzy Logic System

Here, we make a short presentation of the fuzzy logic system (FLS). As is known, the FLS is a formidable nonlinear function approximation containing fuzzy IF-THEN rules which are listed as [8].

$R^{(K)}$: if x_1 is P_1^K and ... and x_n is P_n^K, then $\widehat{f}$ is f^K, $K = 1, 2, \ldots, N$,where N is fuzzy rules, P_1^K, P_2^K, ... and P_n^K are a collection of fuzzy, f^K is the output of the fuzzy singleton in the Kth rule. In accordance with applying the fuzzifier, product inference, and defuzzifier, the output of the FLS is constructed as [22, 23]

$$\widehat{f}(\underline{x}) = \frac{\sum_{i=1}^{N} f^i\left(\prod_{j=1}^{n} \mu_{P_j^i}(x_j)\right)}{\sum_{i=1}^{N}\left(\prod_{j=1}^{n} \mu_{P_j^i}(x_j)\right)} = \Theta^T R(\underline{x}), \tag{1}$$

where $\mu_{P_j^i}(x_j)$ represents the membership grade which x_j is part of P_j^i, $\Theta^T = [f^1, f^2, \ldots, f^N]$ lies on the following parameters, and $R^T = [R^1 R^2 \ldots R^N]$, where

$$R^i(\underline{x}) = \frac{\prod_{j=1}^{n} \mu_{P_j^i}(x_j)}{\sum_{i=1}^{N}\left(\prod_{j=1}^{n} \mu_{P_j^i}(x_j)\right)}. \tag{2}$$

Once that fuzzy basis function has been determined, there always exists more than one rule i.e. $\sum_{i=1}^{n}(\prod_{j=1}^{n} \mu_{P_j^i}(x_j)) > 0$ [24].

3. Dynamic Model of Robotic Manipulators

In this paper, we will consider a more general multi-degree robotic manipulators whose control inputs contain dead zone. The model can be written as

$$M(q)\ddot{q} + C(q, \dot{q})\dot{q} + D\dot{q} + g(q) = \Psi(u), \tag{3}$$

where $q = [q_1, \ldots, q_n]^T \in R^n$ represents joints position, $u \in R^n$ is a control signal with dead zones, $M(q)$ is a inertia matrix, $C(q, \dot{q})\dot{q}$ is centrifugal pull, and $D\dot{q}$ and $g(q)$ are resistance and power of gravity, respectively. $M(q)$ with respect to t total derivative is $\dot{M}$, D is a semi-definite symmetric matrix [8, 25].

Let $y = q$, $x = \dot{q}$, (2) can be converted into

$$\ddot{y} = M^{-1}(y)[\Psi(u) - C(y, x) - D - g(y)]. \tag{4}$$

In the light of (4), the n joints robotic manipulators with dead zones can be written as the following compact formula:

$$\ddot{y} = F(x) + G(x)\Psi(u), \tag{5}$$

where $F(x) = -M^{-1}(y)[D + g(y) + C(y, x)]$, and $G(x) = M^{-1}(y)$.

There exists some necessary assumptions that need to be given.

Assumption 1. Presuming bounded referable signal is $y_d = [y_{d1}, \ldots, y_{dn}]^T$ and $x = [x_1, \ldots, x_n]^T$ is known and measurable. Besides, y_d is a known and continuous vector function in a compact set Ω_{yd}.

Assumption 2. The function G is a positive definite matrix, namely, $G \geqslant \sigma_0 I_{n\times n}$, with $\sigma_0 \geq 0$.

In accordance with the dynamic model of robotic manipulators system (5), the fuzzy adaptive control approach is designed as follows.

To begin with, the tracking errors are defined by

$$e_i = y_i - y_{di}, \quad i = 1, 2, \ldots, n, \tag{6}$$

and the filter tracking error is given as

$$S = [S_1, \ldots, S_n]^T, \tag{7}$$

with

$$S_i = (\lambda_i, 1)\begin{pmatrix} e_i \\ \dot{e}_i \end{pmatrix}. \tag{8}$$

In order to express more clearly, we rewrite (8) into the form of a matrix. Hence, the vector S can be expressed as

$$S = C^T E, \tag{9}$$

where

$$C^T = \begin{pmatrix} \lambda_1 & 1 & 0 & 0 \cdots 0 & 0 \\ 0 & 0 & \lambda_2 & 1 \cdots 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \vdots & \vdots \\ 0 & 0 & 0 & 0 \ldots \lambda_n & 1 \end{pmatrix}_{(n\times 2n)}, \tag{10}$$

$$E = \left(e_1, \dot{e}_1, \ldots, e_n, \dot{e}_n \right)^T_{(2n\times 1)}. \tag{11}$$

To proceed given the dynamic of S_i as the form of a matrix, we can also gain

$$\dot{S}_i = \left(0, \lambda_i \right) \begin{pmatrix} e_i \\ \dot{e}_i \end{pmatrix} + \ddot{e}_i . \tag{12}$$

Hence, the dynamic of S can be given as

$$\dot{S} = C_r^T E + \ddot{e}, \tag{13}$$

where

$$C_r^T = \begin{pmatrix} 0 & \lambda_1 & 0 & 0 \cdots 0 & 0 \\ 0 & 0 & 0 & \lambda_2 \cdots 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \vdots & \vdots \\ 0 & 0 & 0 & 0 \ldots 0 & \lambda_n \end{pmatrix}_{(n\times 2n)}, \tag{14}$$

with

$$\ddot{e} = \ddot{y} - \ddot{y}_d. \tag{15}$$

Then, substituting (15) into (13) yields

$$\dot{S} = C_r^T E + \ddot{y} - \ddot{y}_d . \tag{16}$$

3.1. Input Nonlinearity. Noting that in the robotic manipulators (3), the dead zone input exists. In this work, this input nonlinearity $\Psi_i(u_i)$ can be expressed as [25]

$$\Psi_i(u_i) = \begin{cases} Q_{i+}(u_i)(u_i - u_{iT}), & u_i > u_{iT}, \\ 0, & -u_{it} \leqslant u_i \leqslant u_{iT}, \\ Q_{i-}(u_i)(u_i + u_{it}), & u_i < -u_{it}. \end{cases} \tag{17}$$

To facilitate the transformation of the dead zone input, here, we suppose that it has some particular properties, i.e., there are two positive constants M_{i+}^* and M_{i-}^* such that the below conditions can be constructed as [16]

$$\begin{cases} (u_i - u_{iT})\Psi_i(u_i) & \geqslant M_{i+}^* (u_i - u_{iT})^2, \quad u_i > u_{iT}, \\ (u_i + u_{it})\Psi_i(u_i) & \geqslant M_{i-}^* (u_i + u_{it})^2, \quad u_i < -u_{it}, \end{cases} \tag{18}$$

with $\beta_i = \min\{M_{i+}^*, M_{i-}^*\}$.

4. The Construction of AFC

In what follows, how to construct a reasonable controller to control the robotic manipulator system (5) is given as follows.

First of all, by substituting (5) into (16), we can obtain

$$\dot{S} = C_r^T E + F(x) + G\Psi(u) - \cdot\, y_d. \tag{19}$$

Now posing $\widehat{G}_1 = G^{-1}$, we have

$$\widehat{G}_1 \dot{S} = \widehat{G_1}\left(C_r^T E - \cdot\, y_d + F(x) \right) + \Psi(u). \tag{20}$$

In order to design the optimal controller, (20) is conducted as

$$\begin{aligned} \widehat{G}_1 \dot{S} &= \widehat{G}_1\left(C_r^T E - \cdot\, y_d + F(x) \right) + \Psi(u) \\ &= \kappa(x, m) + \Psi(u), \end{aligned} \tag{21}$$

where $\kappa(x, m) = \widehat{G}_1[m + F(x)]$, and $m = C_r^T E - \dot{y}_d$.

Assumption 3. It is supposed that $\kappa_i(x, m)$ is a bound continuous function satisfying $|\kappa_i(x, m)| \leqslant \beta\overline{\kappa}_i(x)$, with $\beta = \min\{\beta_i\}$.

Remark 1. The establishment of Assumption 3 is not limited mainly due to the upper limit $\beta\overline{\kappa}_i(x)$ is unknown, in addition, $\kappa_i(x, m)$ is a continuous function in the interval Ω_x such that $\beta\overline{\kappa}_i(x)$ always exists [25].

Here, in light of description of the FLS, $\overline{\kappa}_i(x)$ can be depicted as

$$\widehat{\overline{\kappa}}_i(x, \Theta) = \Theta_i^T R_i(x), \tag{22}$$

where $R_i(x)$ is the FBF which is given by the decision maker and Θ_i can be tuned based on requirements of the FLS.

We define

$$\Theta_i^* = \arg\min_{\Theta_i} \left[\sup_{x\in\Omega_x} |\overline{\kappa}_i(x) - \widehat{\overline{\kappa}}_i(x, \Theta_i)| \right], \tag{23}$$

as the optimal estimate values of Θ_i. It is worth mentioning that artificial constant Θ_i^* is introduced only for analysis purposes in the whole process.

Let the estimate error be

$$\widetilde{\Theta}_i = \Theta_i - \Theta_i^*, \tag{24}$$

and the fuzzy approximation error is given as

$$\varepsilon_i(x) = \overline{\kappa}_i(x) - \widehat{\overline{\kappa}}_i(x, \Theta_i^*), \tag{25}$$

where $\widehat{\overline{\kappa}}_i(x, \Theta_i^*) = \Theta_i^{*T} R_i(x)$.

It must be emphasized that the input function of FLS must be contained in the scope of Ω_x, otherwise the system cannot operate normally. Assume that $\varepsilon_i(x) \in \Omega_x$ is bounded, that is to say, $|\varepsilon_i(x)| \leqslant \overline{\varepsilon}_i$, $\overline{\varepsilon}_i \in \mathbb{R}$, one can obtain

$$\begin{aligned} &\widehat{\overline{\kappa}}_i(x, \Theta_i) - \overline{\kappa}_i(x) = \widehat{\overline{\kappa}}_i(x, \Theta_i) - \widehat{\overline{\kappa}}_i(x, \Theta_i^*) + \widehat{\overline{\kappa}}_i(x, \Theta_i^*) - \overline{\kappa}_i(x) \\ &= \widehat{\overline{\kappa}}_i(x, \Theta_i) - \widehat{\overline{\kappa}}_i(x, \Theta_i^*) - \varepsilon_i(x) = \widetilde{\Theta}_i^T R_i(x) - \varepsilon_i(x). \end{aligned} \tag{26}$$

In what follows, let us consider a suitable adaptive fuzzy controller

$$u_i = \begin{cases} -Q_i(t)\text{sign}(S_i) - u_{it}, & S_i > 0, \\ 0, & S_i = 0, \\ -Q_i(t)\text{sign}(S_i) + u_{iT}, & S_i < 0, \end{cases} \tag{27}$$

with $Q_i(t) = z_{0i} + z_{1i}|S_i| + \Theta_i^T R_i(x)$, and

$$\begin{aligned} \dot{z}_{0i} &= -\zeta_{0i}\sigma_{0i}z_{0i} + \zeta_{0i}|S_i|, z_{0i}(0) \geqslant 0, \\ \dot{\Theta}_i &= -\zeta_{1i}\sigma_{1i}\Theta_i + \zeta_{1i}|S_i|R_i(x), \Theta_{ij} \geqslant 0, \end{aligned} \tag{28}$$

where ζ_{0i}, ζ_{1i}, σ_{0i}, σ_{1i}, $z_{1i} > 0$ are marked up to decision makers, z_{0i} and Θ_i are estimated values of $z_{0i}^* = \bar{\varepsilon}_i$ and Θ_i^*, respectively.

Remark 2. With $z_{0i}(0) \geqslant 0$ and $\Theta_i(0) \geqslant 0$, it is known from that solutions of adaptive laws (28) hold $z_{0i}(t) \geqslant 0$ and $\Theta_i(t) \geqslant 0$, for $t > 0$.

Based on Assumption 3 and multiplying both sides of equation (21) by $(1/\beta)S^T$, one can obtain

$$\begin{aligned} \frac{1}{\beta}S^T\widehat{G}_1(x)\dot{S} &= \frac{1}{\beta}S^T\kappa(x,m) + \frac{1}{\beta}S^T u, \\ &\leqslant \sum_{i=1}^{n}|S_i|\bar{\kappa}_i(x) + \frac{1}{\beta}S^T u. \end{aligned} \tag{29}$$

From (26) and (29), one gets

$$\begin{aligned} \frac{1}{\beta}S^T\widehat{G_1}(x)\dot{S} &\leqslant \sum_{i=1}^{n}|S_i|\bar{\kappa}_i(x) + \frac{1}{\beta}S^T u \\ &\leqslant -\sum_{i=1}^{n}|S_i|\tilde{z}_{0i} - \sum_{i=1}^{n}|S_i|\tilde{\Theta}_i^T R_i(x) + \sum_{i=1}^{n}|S_i|z_{0i} \\ &\quad + \sum_{i=1}^{n}|S_i|\Theta_i^T R_i(x) + \frac{1}{\beta}S^T u, \end{aligned} \tag{30}$$

with $\tilde{\Theta}_i = \Theta_i - \Theta_i^*$ and $\tilde{z}_{0i} = z_{0i} - z_{0i}^* = z_{0i} - \bar{\varepsilon}_i$.

Theorem 1. *When Assumptions 1–3 are satisfied, for the robotic manipulator system (5) with control law (27) and adaptive laws (28), all the states are remained as a bounded domain and the error function of the closed-loop system tends to be within a small variable range of origin via using adaptive fuzzy approach.*

Proof. We define

$$V = \frac{1}{2\beta}S^T\widehat{G}_1 S + \frac{1}{2}\sum_{i=1}^{n}\frac{1}{\zeta_{0i}}\tilde{z}_{0i}^2 + \frac{1}{2}\sum_{i=1}^{n}\frac{1}{\gamma_{1i}}\tilde{\Theta}_i^T\tilde{\Theta}_i, \tag{31}$$

$$\dot{V} = \frac{1}{\beta}S^T\widehat{G}_1\dot{S} + \sum_{i=1}^{n}\frac{1}{\zeta_{0i}}\tilde{z}_{0i}\dot{z}_{0i} + \sum_{i=1}^{n}\frac{1}{\gamma_{1i}}\tilde{\Theta}_i^T\dot{\Theta}_i, \tag{32}$$

with $\widehat{G_1} = 0$. Thus, from (28) and (41), we can obtain the following important inequality. When $S_i > 0$,

$$(u_i + u_{it})\Psi_i(u_i) = -Q_i(t)\text{sign}(S_i)\Psi_i(u_i) \geqslant M_{i-}^* Q_i^2(t) \geqslant \beta Q_i^2(t), \tag{33}$$

and when $S_i < 0$,

$$(u_i - u_{iT})\Psi_i(u_i) = -Q_i(t)\text{sign}(S_i)\Psi_i(u_i) \geqslant M_{i+}^* Q_i^2(t) \geqslant \beta Q_i^2(t). \tag{34}$$

Hence, we conclude that for all S_i holding

$$-Q_i(t)\text{sign}(S_i)\Psi_i(u_i) \geqslant \beta Q^2(t). \tag{35}$$

Owing to $S_i^2 > 0$ and $S_i\text{sign}(S_i) = |S_i|$, (35) becomes

$$-Q_i(t)S_i^2\text{sign}(S_i)\Psi_i(u_i) \geqslant \beta Q_i^2(t)S_i^2 = \beta Q_i^2(t)|S_i|^2. \tag{36}$$

For all S_i, we have

$$S_i\Psi_i(u_i) \leq -\beta Q_i(t)|S_i|. \tag{37}$$

Then, substituting (28), (30), and (37) into (32), we obtain

$$\begin{aligned} \dot{V} &\leqslant \sum_{i=1}^{n}|S_i|z0i + \sum_{i=1}^{n}|S_i|\Theta_i^T R_i(x) + \frac{1}{\beta}\sum_{i=1}^{n}S_i\Psi_i(u_i) - \sum_{i=1}^{n}\sigma_{0i}\tilde{z}_{0i}z_{0i} \\ &\quad - \sum_{i=1}^{n}\sigma_{1i}\tilde{\Theta}_i^T\Theta_i \leqslant \sum_{i=1}^{n}|S_i|z0i + \sum_{i=1}^{n}|S_i|\Theta_i^T\psi_i(x) + \sum_{i=1}^{n}-z_i(t)|S_i| \\ &\quad - \sum_{i=1}^{n}\sigma_{0i}\tilde{z}_{0i}z_{0i} - \sum_{i=1}^{n}\sigma_{1i}\tilde{\Theta}_i^T\Theta_i = -\sum_{i=1}^{n}z_{1i}S_i^2 - \sum_{i=1}^{n}\sigma_{0i}\tilde{z}_{0i}z_{0i} \\ &\quad - \sum_{i=1}^{n}\sigma_{1i}\tilde{\Theta}_i^T\Theta_i. \end{aligned} \tag{38}$$

Obviously, we have

$$\begin{aligned} -\sigma_{0i}\tilde{z}_{0i}z_{0i} &\leqslant -\frac{\sigma_{0i}}{2}\tilde{z}_{0i}^2 + \frac{\sigma_{0i}}{2}z_{0i}^* 2, \\ -\sigma_{1i}\tilde{\Theta}_i^T\Theta_i &\leqslant -\frac{\sigma_{1i}}{2}\left\|\tilde{\Theta}_i\right\|^2 + \frac{\sigma_{1i}}{2}\left\|\Theta_i^*\right\|^2. \end{aligned} \tag{39}$$

Then, (38) becomes

$$\begin{aligned} \dot{V} &\leqslant -\sum_{i=1}^{n}z_{1i}S_i^2 - \sum_{i=1}^{n}\frac{\sigma_{0i}}{2}\tilde{z}_{0i}^2 - \sum_{i=1}^{n} - \sum_{i=1}^{n}\frac{\sigma_{1i}}{2}\left\|\tilde{\Theta}_i\right\|^2 + \sum_{i=1}^{n}\frac{\sigma_{0i}}{2}z_{0i}^* 2 \\ &\quad + \sum_{i=1}^{n}\frac{\sigma_{1i}}{2}\left\|\Theta_i^*\right\|^2 \end{aligned} \tag{40}$$

Since $\widehat{G}_1 \geqslant \sigma_{g0}I_n$, then we have

$$S^T G^{-1} S = S^T\widehat{G}_1 S \leqslant \frac{1}{\sigma_{g0}}\|S\|^2. \tag{41}$$

From (40) and (41), one has

$$\dot{V} \leqslant -\mu V + \tau, \tag{42}$$

where

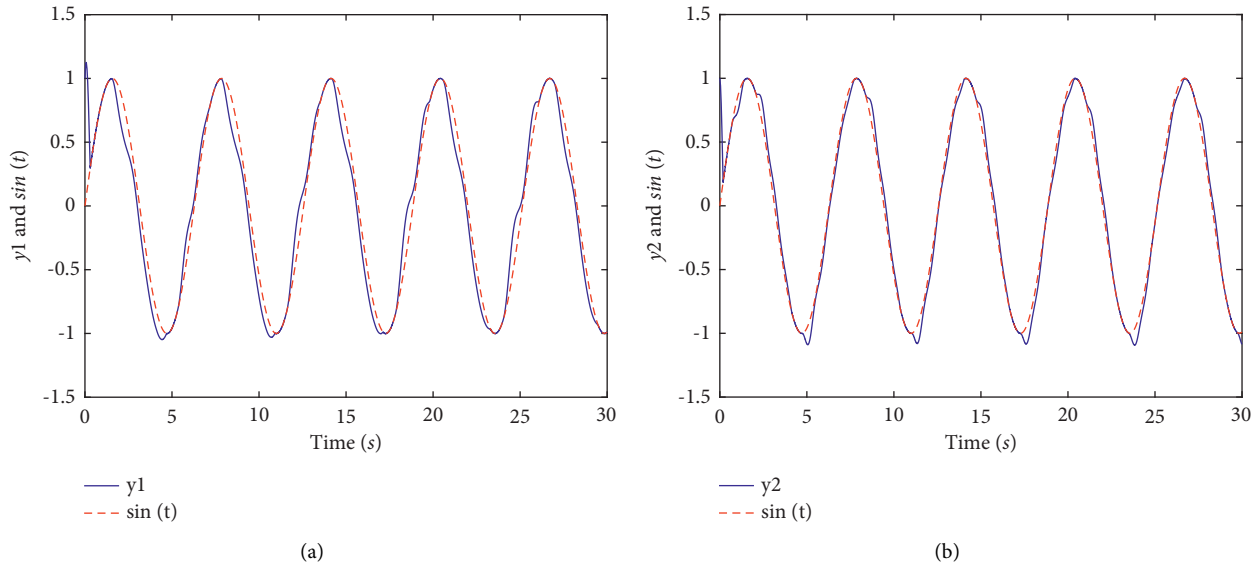


FIGURE 1: The tracking performance of two-link robotic manipulators systems without dead zones (with $\alpha = 1$). (a) Tracking of y_1 and $\sin t$. (b) Tracking of y_2 and $\sin t$.

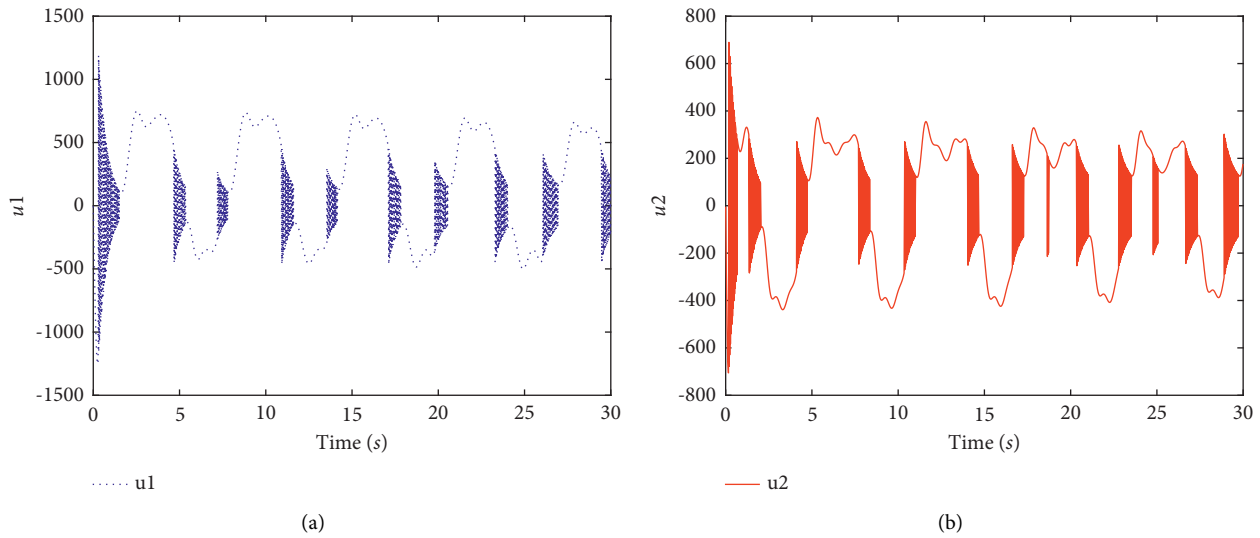


FIGURE 2: Simulation results (with $\alpha = 1$). (a) The controller u_1. (b) The controller u_2.

$$\tau = \sum_{i=1}^{n} \frac{\sigma_{0i}}{2} z_{0i}^{*} 2 + \sum_{i=1}^{n} \frac{\sigma_{1i}}{2} \left\| \Theta_i^{*} \right\|^2,$$

$$\omega = \min \left\{ \min_i \left\{ 2\eta \sigma_{g0} z_{1i} \right\}, \ \min_i \left\{ \zeta_{0i} \sigma_{0i} \right\}, \min_i \left\{ \zeta_{1i} \sigma_{1i} \right\} \right\}. \tag{43}$$

Multiplying (42) by $e^{\omega t}$ yields

$$\frac{d}{dt} \left(V e^{\omega t} \right) \leqslant \tau e^{\omega t}. \tag{44}$$

Then, integrating over the interval [0, t] of the (44), we can get

$$0 \leqslant V(t) \leqslant \frac{\tau}{\omega} + \left(V(0) - \frac{\tau}{\omega} \right) e^{-\omega t}. \tag{45}$$

By utilizing (31), $V(0)$ is built as

$$V(0) = \frac{1}{2\beta} S(0)^T \widehat{G_1} S0 + \frac{1}{2} \sum_{i=1}^{n} \frac{1}{\zeta_{0i}} z_{0i} 0 - z_{0i}^{*\,2} + \frac{1}{2} \sum_{i=1}^{n} \frac{1}{\zeta_{1i}} \Theta_i 0 - \Theta_i^{*\,T} \Theta_i 0 - \Theta_i^{*}, \tag{46}$$

where $\widehat{G_1} \geqslant \sigma_{g1} I_n$, $\sigma_{g1} > 0$.

In the light of (31) and (45), one can achieve

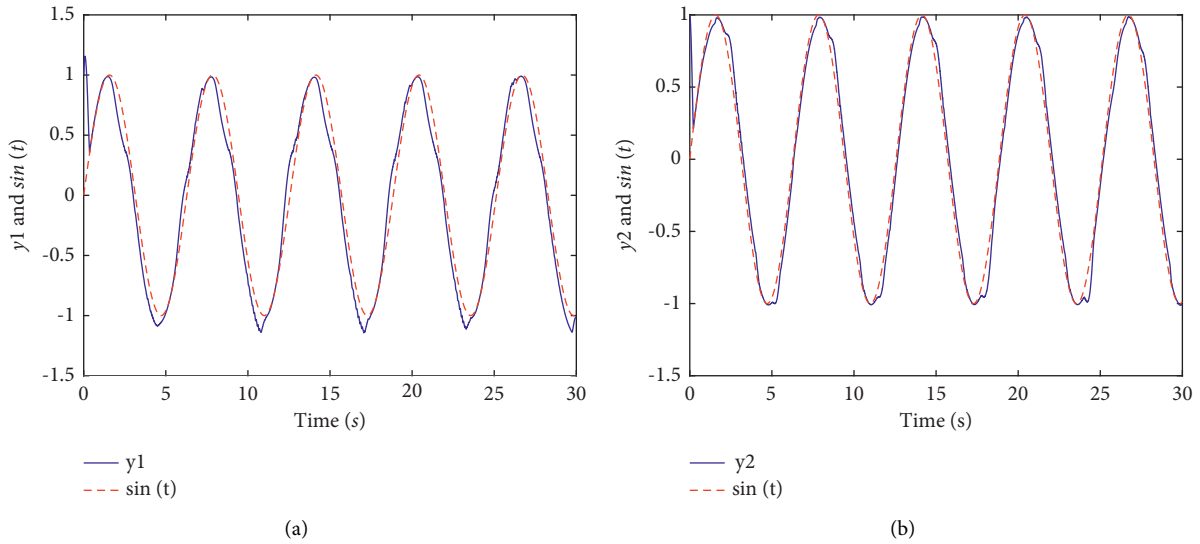


FIGURE 3: The tracking performance of two-link robotic manipulators systems with dead zones (with $\alpha = 1$). (a) Tracking of y_1 and $\sin t$. (b) Tracking of y_2 and $\sin t$.

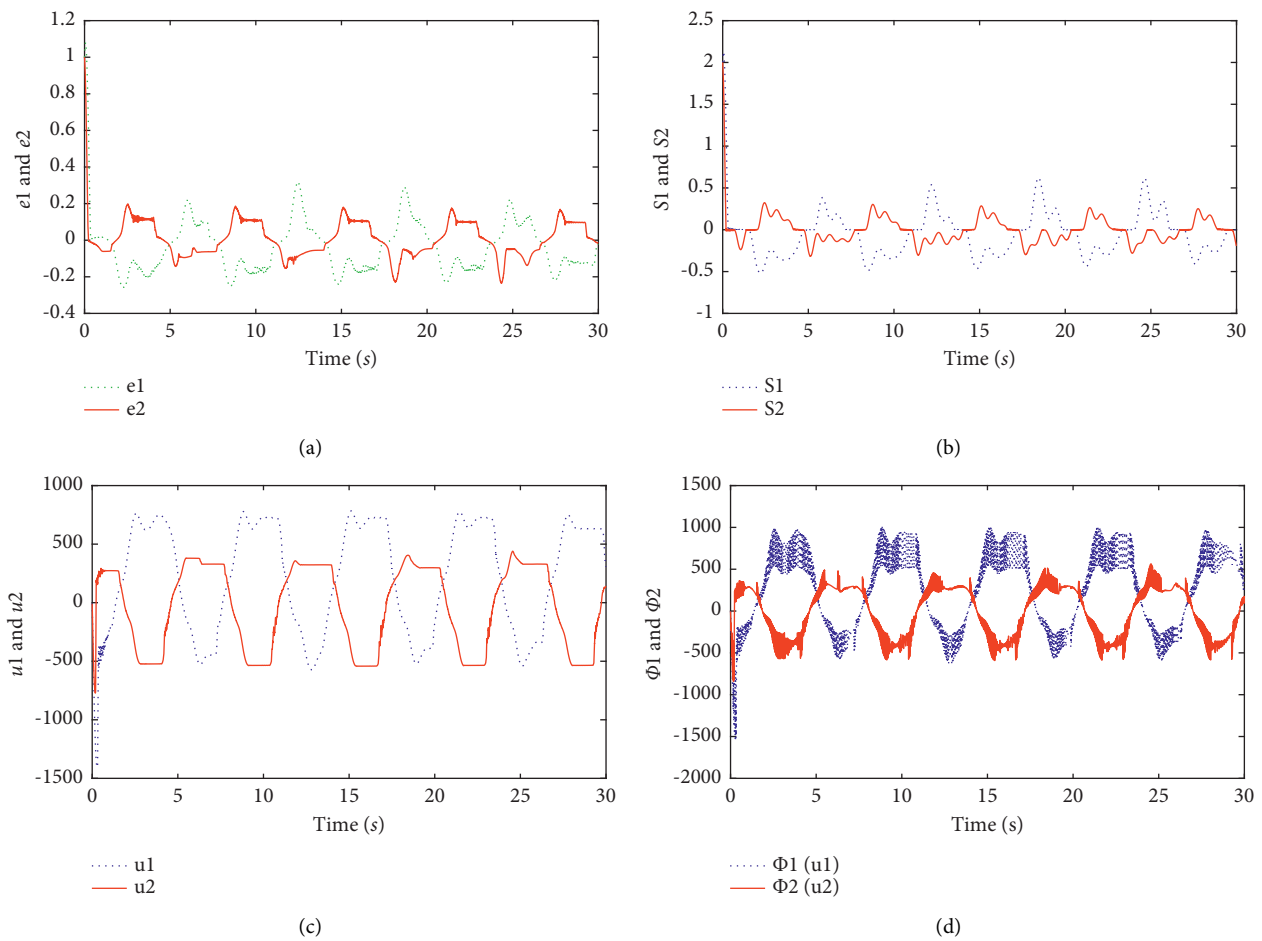


FIGURE 4: Simulation results (with $\alpha = 1$). (a) The tracking error e_1 and e_2. (b) The sliding mode S_1 and S_2. (c) The controller u_1 and u_2. (d) Nonlinear functions $\Psi_1(u_1)$ and $\Psi_2(u_2)$.

$$|S_i| \leqslant \left(\frac{2\beta}{\sigma_{g1}} \left(\frac{\tau}{\omega} + \left(V(0) - \frac{\tau}{\omega} \right) e^{-\omega t} \right) \right)^{1/2}. \quad (47)$$

From the above inequality, it can be concluded that the exponential solution of S_i gradually tends to be bounded, where $\Omega_{S_i} = \left\{ S_i \,|\, |S_i| \leqslant (2\eta/\sigma_{g1}\tau/\omega)^{1/2} \right\}$, and the proof of the theorem is finished. □

Remark 3. It is worth mentioning that there is a special example $\Psi_i(u_i) = u_i$, that is, when the system model does not contain nonlinear input functions with dead zones, the controller we structured is still applicable to system (5), and the tracking performance are shown in Figures 1 and 2.

Remark 4. (1) If $u_{iT} = u_{it} = u_{ii}$ are established, (27) can be transformed into

$$u_i = -(u_{ii} + Q_i(t)) \operatorname{sign}(S_i), \quad (48)$$

where $Q_i(t) = z_{0i} + z_{1i}|S_i| + \Theta^T R_i(x)$.

(2) It is worth highlighting that the approximation effect of the robotic manipulator is significantly better when the function $\operatorname{sign}(.)$ is substituted for a smooth function $\arcsin(.)$, $\arctan(.)$, and so on.

5. Simulation Results

With the aim of checking the control efficiency of the proposed scheme, let us consider robotic manipulators with two single joints without disturbance.

$$(\ddot{q}_1\ \ddot{q}_2) = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix}^{-1} \left\{ \begin{pmatrix} \Psi_1(u_1) \\ \Psi_2(u_2) \end{pmatrix} - \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \begin{pmatrix} \dot{q}_1 \\ \dot{q}_2 \end{pmatrix} - \begin{pmatrix} g_{11} \\ g_{21} \end{pmatrix} \right\}, \quad (49)$$

where $M_{11} = b_1 + 4b_4 \cos q_2$, $M_{12} = b_2 + b_4 \cos q_2$, $M_{21} = b_1 + b_4 \cos q_2$, $M_{22} = b_3$, $C_{11} = -b_4 \dot{q}_2 \sin q_2$, $C_{12} = -b_4 \sin q_2 (\dot{q}_1 + \dot{q}_2)$, $C_{21} = b_4 \dot{q}_1 \sin q_2$, $C_{22} = 0$, $g_{11} = c_1 \cos q_1 + c_2 \cos(q_1 + q_2)$, $g_{21} = c_2 \cos(q_1 + q_2)$ and the parameters are designed below without disturbance: $b_1 = 200.03$, $b_2 = 23.7$, $b_3 = 122.8$, $b_4 = 25.3$, $c_1 = 784.9$, $c_2 = 245.35$.

Let us denote $y = [y_1, y_2]^T = [q_1, q_2]^T$, $x = [x_1, x_2, x_3, x_4]^T = [q_1, q_2, \dot{q}_1, \dot{q}_2]^T$, and $\Psi(u) = [\Psi_1(u_1), \Psi_2(u_2)]$, thus, (49) is depicted as

$$\ddot{y} = F(x) + G(x)\Psi(u). \quad (50)$$

The input nonlinearities $\Psi_1(u_1)$ and $\Psi_2(u_2)$ are arranged as

$$\Psi_1(u_1) = \begin{cases} (u_1 - 2)(1 - 0.3 \sin u_1), & u_1 > 2, \\ 0, & -2 \leqslant u_1 \leqslant 2, \\ (u_1 + 2)(0.8 - 0.3 \cos u_1), & u_1 < -2, \end{cases} \quad (51)$$

and $\Psi_2(u_2)$ is supposed to be

$$\Psi_2(u_2) = \begin{cases} (u_2 - 5)(1 - 0.3 \sin u_2), u_2 > 5, \\ 0, -5 \leq u_2 \leq 5, \\ (u_2 + 5)(0.8 - 0.3 \cos u_2), u_2 < -5. \end{cases} \quad (52)$$

In the simulation process, a fuzzy logic system is used, which contains four identical Gaussian functions. Each Gaussian function is distributed in the interval $[-10, 10]$ and their standard deviation is selected as 0.9. The input vector of the fuzzy system $\Theta^T R(\underline{x})$ is $x = [x_1, x_2, x_3, x_4]^T$. Here, we set the step size as 5 so that there exists $5 \times 5 \times 5 \times 5 = 625$ rules in the FLS. The parameters in the simulation process are designed as $\zeta_{01} = 31$, $\zeta_{02} = 79$, $\zeta_{11} = \zeta_{12} = 4000$, $\sigma_{01} = \sigma_{02} = 0.001$, $\sigma_{11} = \sigma_{12} = 0.0005$, $\lambda_1 = \lambda_2 = 2$, and $z_{11} = z_{12} = 2$. Simulation initial parameters of adaptive laws and control law are chosen as $z_{01} = z_{02} = 0$, $\Theta_{1j} = \Theta_{2j}$, $j = 1, \ldots, 625$, and $u_{1T} = u_2 T = 2$, $u_1 t = u_2 t = 5$. Original values of the robotic manipulator system are designed as $x(0) = [1, 2, 3, 4]^T$ and $y(0) = [1, 1]^T$. Here, one can see that $\operatorname{sign}(S_i)$ is a piecewise discontinuous function that can take the place of $\tanh(z_{si} S_i)$ with $z_{si} = 20$, $i = 1, 2$. It is noteworthy that our target is to make the robotic manipulator system joints q_1 and q_2 to track the desired given trajectories $y_{d1} = y_{d2} = \sin t$.

Finally, it is apparent that presents the good tracking performances of the two-link robotic manipulators with dead zones in Figure 3. Figure 4 shows transient behaviors of the tracking error, sliding mode, controllers, and nonlinear functions with dead zones. In terms of Remark 4, it is noticeable that the tracking effect in Figure 1 is the same as Figure 3 and the range of variation of controllers with different parameter values are shown in Figure 2.

6. Conclusion

This paper adopts the adaptive fuzzy control approach for uncertain multi-degree robotic manipulators with a dead zone. The fuzzy logic system has been presented to dispose of unknown and uncertain nonlinear functions of the robotic manipulator. It has been proven in accordance with Lyapunov criterion that the proposed adaptive fuzzy control scheme can guarantee that the states and signals of the whole closed-loop system tend toward bounded and the tracking errors are close to small regions of origin gradually. The practicability of the proposed approach is affirmed by an example of robotic manipulators with two joint. The deficiency of this paper is that the tracking error observed from the image is relatively large, how to reduce the tracking error will be our future research direction.

Data Availability

All datasets generated for this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] G. Zhong, C. Wang, and W. Dou, "Fuzzy adaptive Pid fast terminal sliding mode controller for a redundant manipulator," *Mechanical Systems and Signal Processing*, vol. 159, Article ID 107577, 2021.

[2] Y. Zhong, Y. Pu, and T. Wang, "A A sliding mode and nonlinear disturbance observer based bilateral control for telerehabilitation systems with flexible manipulators sliding Mode and Non-linear Disturbance Observer Based Bilateral Control for Telerehabilitation Systems with Flexible Manipulators," *Cognitive Robotics*, vol. 2, pp. 39–49, 2022.

[3] X. Yin, L. Pan, and S. Cai, "Robust adaptive fuzzy sliding mode trajectory tracking control for serial robotic manipulators," *Robotics and Computer-Integrated Manufacturing*, vol. 72, Article ID 101884, 2021.

[4] J. Rezapour, M. Sharifi, and N. Nariman-zadeh, "Application of fuzzy sliding mode control to robotic manipulator using multi-objective genetic algorithm," in *Proceedings of the 2011 International Symposium on Innovations in Intelligent Systems and Applications*, pp. 455–459, IEEE, Istanbul, Turkey, June 2011.

[5] M. Ertugrul and O. Kaynak, "Neuro sliding mode control of robotic manipulators," *Mechatronics*, vol. 10, no. 1-2, pp. 239–263, 2000.

[6] W. Hongrui, L. Congna, and Z. Qiguang, "Fuzzy sliding mode control for robotic manipulators based on passivity theory,"vol. 1, pp. 535–538, in *Proceedings of the 2010 International Conference on Logistics Systems and Intelligent Management (ICLSIM)*, vol. 1, pp. 535–538, IEEE, Harbin, China, January 2010.

[7] H. Ho, Y.-K. Wong, and A. B. Rad, "Robust fuzzy tracking control for robotic manipulators," *Simulation Modelling Practice and Theory*, vol. 15, no. 7, pp. 801–816, 2007.

[8] Q. Zhou, H. Li, and P. Shi, "Decentralized adaptive fuzzy tracking control for robot finger dynamics," *IEEE Transactions on Fuzzy Systems*, vol. 23, no. 3, pp. 501–510, 2015.

[9] A. Boulkroune, M. Tadjine, M. MSaad, and M. Farza, "Fuzzy adaptive controller for mimo nonlinear systems with known and unknown control direction," *Fuzzy Sets and Systems*, vol. 161, no. 6, pp. 797–820, 2010.

[10] O. Tutsoy and D. E. Barkana, "Model free adaptive control of the under-actuated robot manipulator with the chaotic dynamics," *ISA Transactions*, vol. 118, pp. 106–115, 2021.

[11] Y. Gao, M. Er, and S. Yang, "Adaptive fuzzy neural control of robot manipulators," *IEEE Transactions on Industrial Electronics*, vol. 48, no. 6, pp. 1274–1278, 2001.

[12] N. Goléa, A. Goléa, K. Barra, and T. Bouktir, "Observer-based adaptive control of robot manipulators: Observer-based adaptive control of robot manipulators: Fuzzy systems approach fuzzy systems approach," *Applied Soft Computing*, vol. 8, no. 1, pp. 778–787, 2008.

[13] R. K. Gupta and S. Chauhan, "Comparision of Pid controller & adaptive neuro fuzzy controller for robot manipulator," in *Proceedings of the 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pp. 1–4, IEEE, Madurai, India, December 2015.

[14] V. Kumar, K. Rana, and D. Kler, "Efficient control of a 3-link planar rigid manipulator using self-regulated fractional-order fuzzy Pid controller," *Applied Soft Computing*, vol. 82, Article ID 105531, 2019.

[15] A. L. Luna, I. C. Vega, and J. M. Carranza, "Gain-scheduling and Pid control for an autonomous aerial vehicle with a robotic arm," in *Proceedings of the 2018 IEEE 2nd Colombian Conference on Robotics and Automation (CCRA)*, pp. 1–6, IEEE, Barranquilla, Colombia, November 2018.

[16] A. Boulkroune and M. MSaad, "A fuzzy adaptive variable-structure control scheme for uncertain chaotic mimo systems with sector nonlinearities and dead-zones," *Expert Systems with Applications*, vol. 38, no. 12, pp. 14744–14750, 2011.

[17] Y. Jin, "Decentralized adaptive fuzzy control of robot manipulators," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 28, no. 1, pp. 47–57, 1998.

[18] S. Fateh and M. M. Fateh, "Adaptive fuzzy control of robot manipulators with asymptotic tracking performance," *Journal of Control, Automation and Electrical Systems*, vol. 31, no. 1, pp. 52–61, 2020.

[19] Q. Zhou, S. Zhao, H. Li, R. Lu, and C. Wu, "Adaptive neural network tracking control for robotic manipulators with dead zone," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 12, pp. 3611–3620, 2019.

[20] K.-C. Hsu, W.-Y. Wang, and P.-Z. Lin, "Sliding mode control for uncertain nonlinear systems with multiple inputs containing sector nonlinearities and dead zones," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 374–380, 2004.

[21] H. Wang and S. Kang, "Adaptive neural command filtered tracking control for flexible robotic manipulator with input dead-zone," *IEEE Access*, vol. 7, pp. 22675–22683, 2019.

[22] H. Li, X. Liao, and X. Lei, "Two fuzzy control schemes for Lorenz-Stenflo chaotic system," *Journal of Vibration and Control*, vol. 18, no. 11, pp. 1675–1682, 2012.

[23] A. Boulkroune, A. Bouzeriba, S. Hamel, and T. Bouden, "A projective synchronization scheme based on fuzzy adaptive control for unknown multivariable chaotic systems," *Nonlinear Dynamics*, vol. 78, no. 1, pp. 433–447, 2014.

[24] A. Boulkroune, "Fuzzy adaptive controller for uncertain multivariable nonlinear systems with both sector nonlinearities and dead-zones," in *Handbook of Research on Advanced Intelligent Control Engineering and Automation*, pp. 334–363, IGI Global, Pennsylvania, USA, 2015.

[25] Y. Huangfu and K. Xue, "Synchronization of chaotic systems with dead zones via fuzzy adaptive variable-structure control," *Complexity*, vol. 2021, pp. 1–8, Article ID 3224416, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 4144569, 15 pages
https://doi.org/10.1155/2022/4144569

Research Article

A New Type of Air Conditioning System Based on Finned Ceiling Radiant Coupled with Independent Fresh Air and Its Thermal Comfort Experimental Study

Wenqi Qin,[1] **Yingning Hu**,[1,2] **Jinwen Su**,[2] **and Yubang Hu**[2]

[1]*College of Civil Engineering and Architecture, Guangxi University, Nanning 530004, China*
[2]*School of Mechanical Engineering, Guangxi University, Nanning 530004, China*

Correspondence should be addressed to Yingning Hu; hyngxu@st.gxu.edu.cn

Received 24 June 2022; Revised 4 August 2022; Accepted 11 August 2022; Published 17 September 2022

Academic Editor: Wei Xiang

The traditional radiation air conditioning system has some problems, such as easy condensation, insufficient refrigeration capacity, complex structure, and control system. Therefore, this study proposes a new type of finned metal radiant plate with large heat flow per unit area, sufficient cooling capacity, and simplified heat exchange system, in order to realize large temperature difference between cooling and heating. The temperature field uniformity and thermal comfort test of a novel type of finned ceiling radiant panel and independent fresh air linked air conditioning system under summer cooling and winter heating circumstances are accomplished through artificially generated climate environments. The study's findings demonstrate that in the radiation and fresh air modes, the maximum interior temperature differential under cooling conditions does not rise over 2.1°C. The maximum temperature differential in the space at any one moment in the radiation and fresh air modes cannot be greater than 3°C when heating conditions are present. The fresh air's cooling and dehumidifying effects are clear. The dehumidification efficiency may reach 50%, and the moisture content ranges from 5.48 to 9.63 g/kg. With PMV ranging from −0.34 to 0.54, the enhanced air conditioning system in this research provides exceptionally good thermal comfort. Additionally, the finned radiant panel's installation area occupies just 14% of the ceiling, which is sufficient to fulfill the room's cooling and heating load needs as well as provide high thermal comfort and consistent indoor temperature. The theoretical investigation and practical implementation of the direct expansion radiant air conditioning system are both strongly supported by this research.

1. Introduction

Radiant air conditioners have attracted more and more attention because of their good thermal comfort. In recent years, many experts and scholars have conducted a lot of research work on the performance of radiant air conditioners. In terms of the theoretical model and simulation research of radiant air conditioning, Wu et al. [1] proposed and established a new simplified model to calculate the surface temperature and heat flux of the floor radiant heating and cooling system and verified the model with the measured data. Using a radiative heat transfer model, Bahadori et al. [2] developed a prediction tool to estimate the characteristic temperature ratio of a plate radiator and a system of heat exchange plates with a high temperature. The computation approach is demonstrated to be in good agreement with the observed data by comparison, and the average absolute variation is just approximately 0.2%. Tang et al. [3] derived an analytical expression for the dynamic thermal response of a radiant floor slab under variable solar radiation using a new approximate harmonic method. The results show that the dynamic heat transfer process calculated by the new approximate harmonic method is in good agreement with the 3D model. Based on three radiation models, Lv et al. [4] established a synchronous thermal model to calculate the indoor air temperature and the indoor wall temperature. Merabtine et al. [5] proposed a new analytical correlation-based simplified model to evaluate the surface temperature of a heat exchange plate and study its thermal performance under dynamic conditions. Finally,

numerical models based on the finite difference method and finite volume method are established. Yu et al. [6] established a simplified model of top-insulated metal ceiling radiant cooling panels (CRCP-s) using a serpentine arrangement to predict the average temperature of the panels, outlet water temperature, and cooling capacity. Experiments were carried out on two CRCP-s, and the results showed that the outlet water temperature and cooling capacity predicted by the model were in good agreement with the experimental measurements. By introducing an inner air layer into the improved ceiling radiant cooling panels (CRCP) system, Su et al. [7] established a computational model for the heat transfer analysis of the improved CRCP system and verified it experimentally. The results show that thicker aluminum plates, thicker heat exchange plates, and thinner air layers increase cooling capacity. Zhao et al. [8] used the CFD model of the exposed capillary ceiling radiant panel (E-CCRP) system office to analyze the heat transfer characteristics and time-delay characteristics of the E-CCRP winter start-up process under unsteady conditions, and experiments are used to confirm that the simulation findings are accurate. Khatri et al. [9] performed computational fluid dynamics (CFD) simulations using ANSYS to determine the thermal performance of the radiative cooling system. Moreover, the air space distribution of the radiation system and the conventional system is simulated by the computational fluid dynamics (CFD) method.

In addition, many researchers have carried out a lot of research work on the performance of radiant air conditioners. Based on the indoor comfort and energy consumption of a demonstration office building, Zhang et al. [10] analyzed the operating conditions and key energy-saving technologies of the compound air conditioning system and conducted a large number of experimental tests and accurate simulations. Research has confirmed that the ground source heat pump combined with the floor radiant air conditioning system is more energy efficient. Du et al. [11] conducted a numerical study on the influence of the changes in the design and operating parameters of the heat exchange plate on indoor thermal comfort and energy-saving performance in the human sleep environment. The results show that the four operating parameters have significant effects on bedroom thermal comfort and energy-saving performance when the R-TAC system is used. Ding et al. [12] made a simplified calculation of the heating capacity, heat loss, and surface temperature of a radiant heating floor with a high-performance thermal insulation layer and verified it with measured data and numerical simulation data. Sun et al. [13] proposed a new type of flat heat pipe radiant heating and cooling integrated terminal and studied the thermal performance (including thermal response speed, heating and cooling capacity, heating and cooling temperature, and heat transfer coefficient) of the flat heat pipe terminal through experiments. Lv et al. [14] developed a new type of heat sink, which consists of an upper flat plate, a lower groove plate, and a heat transfer liquid filling the gap between the two plates. Tang et al. [15] carried out an experimental study on the droplet size from superhydrophobic surfaces and conventional aluminum alloy surfaces. Jin et al. [16] studied the condensation characteristics of the heat exchange plate through experiments and simulations. The temperature of the heat exchange plate and the relative humidity of the adhesion layer during the start-up phase were tested, and their effects on the setting time were analyzed. Tang et al. [15] performed a theoretical study of condensation rates using a simplified N–S approximation equation and derived a heat and mass transfer analogy. In the experiments, the condensation rates of different lengths of radiant cooling panels were measured at different locations, including floors, walls, and ceilings.

In recent years, some scholars have carried out some researches on the indoor environment of radiant air conditioning systems. Su et al. [17] conducted experiments on radiator heating and floor heating in a controlled climate room. The experiment recruited 66 college students as volunteers to investigate their physical and psychological responses. Wang et al. [18] recruited 20 volunteers to participate in thermal comfort experiments in 2 nonuniform environments and used questionnaires for subjective evaluations. Chao et al. [19] recruited 30 experiencers to participate in a questionnaire survey under 8 different experimental conditions under the fan coil and floor radiant cooling system and collected the indoor thermal environment parameters of the 2 systems at different heights through instrument monitoring. Tian et al. [20] conducted a field study of occupant thermal comfort and the thermal environment with a heat exchange slab cooling system, collecting 116 sets of data from 82 participants during summer and winter. For the segmental thermal interaction between the human body and the surrounding environment, Atmaca et al. [21] modified the Gagge-2 node model appropriately to simulate the multisegment situation and verified the local differences.

With the increasing maturity of artificial intelligence technology, artificial intelligence has been widely used in various fields, and many scholars have put forward many models and control algorithms in the field of artificial intelligence. Zhou et al. [22] proposed a fuzzy adaptive synchronization control algorithm with a composite learning technique for a class of incommensurate fractional-order chaotic systems with mismatched parametric uncertainties based on T-S fuzzy models presented. Zhou et al. [23] studied a generalized function projective synchronization of incommensurate fractional-order chaotic systems with inputs saturation. In the study of Zhou et al. [24], passive, active, and combined passive active solutions in PCMs systems have been comprehensively reviewed when being applied in heating, cooling, and electrical systems, together with a dialectical analysis of advantages and disadvantages. In another study by Zhou et al. [24], machine learning methods are effective to assist the energy-efficient renewable systems during multicriteria design and multilevel uncertainty-based operation periods. Zhou et al. [24] pointed out in the study that machine learning methods are promising for thermal and energy performance improvement, through surrogate model development, model predictive control, and optimisation function. Using appropriate intelligent control methods in the HVAC system not only can improve the

efficiency of the system and reduce the energy consumption of the system but also can create a good thermal comfort environment for people. In the heating system, model predictive control (MPC) is an advanced control technology. Siroky et al. [25] found that its energy-saving potential in the weather forecast of building heating systems is between 15% and 28%. Shafiei et al. [26] proposed a predictive control model to directly control the power consumption of large refrigeration systems. Beghi et al. [27] designed a reinforcement-learning algorithm that learns to operate a cooling system based on its interaction with the environment. The above methods improve the efficiency of the refrigeration system by 20%~30% and reduce the wear of various components.

Aiming at the problems existing in the application of traditional radiant air conditioners, this study proposes and manufactures a new type of direct-expansion ceiling radiant and independent fresh air coupled air conditioning system. The system uses R410A as the refrigerant and uses the self-designed finned metal radiant plate structure, which can effectively prevent condensation and solve the problem of insufficient cooling capacity. In the artificial simulated climate environment, the system operating characteristics of the radiant fresh air mode under the conditions of low-temperature heating in winter, normal heating in winter, normal cooling in summer, and high-temperature cooling in summer are studied. Finally, the indoor temperature distribution uniformity, cooling effect, and dehumidification effect of the system under various working conditions are studied through experiments, and then the indoor thermal comfort index (predicted mean vote-predicted percentage dissatisfied (PMV-PPD)) value is calculated. In conclusion, the work and significance of this study are as follows.

Aiming at the shortcomings of the traditional radiant air conditioning system with water as the medium, this paper designs a new radiant air conditioning system based on variable air volume (VAV), which is called a frequency conversion control air conditioning system with a roof radiation module and independent fresh air function. In the system developed in this study, the roof radiation module and the independent fresh air system directly use the refrigerant R410 as the cold medium and are driven by the frequency conversion operation of the compressor, so as to simplify the radiation air conditioning system and make the radiation air conditioning system obtain accurate control. The radiant cooling and radiant heating share a radiant roof end system and add an independent fresh air system to improve indoor comfort proposed in this study, which provides some design theory support and practical value for the application of direct expansion radiation air conditioning system.

In this paper, aiming at the problems of small cooling capacity and easy condensation at the radiation end of the traditional radiation air conditioning, we designed a finned metal radiation plate. The finned metal radiation plate has inlined fins. The aluminum fins increase the contact area with the air, strengthen the heat transfer ability of the copper tube, and increase the convective heat transfer ability of the radiation plate. In addition, the working medium in the copper tube is R410 A, and the groove-type baffle plate discharges the condensed water generated under the refrigeration condition. Our design of finned metal radiant plate provides a reference for the promotion, development, and application of inlined finned structure radiant plate.

This paper takes the direct-expansion roof radiation and independent fresh air coupling air conditioning system as the research objective. Moreover, we analyze the dehumidification effect of the system under refrigeration and dehumidification conditions by testing the distribution uniformity of the indoor temperature field and calculate the indoor thermal comfort index (predicted mean vote-predicted percentage dissatisfied (PMV-PPD)) value, which provides theoretical and application reference for the research and development of this kind of air conditioning system.

2. Design of Air Conditioning System Coupled with Direct Expansion Ceiling Radiation and Independent Fresh Air

2.1. Finned Metal Radiant Panel. Figure 1 depicts the genuine finned metal radiant panel. As can be seen from the figure, it is different from the traditional suspended radiant panel. The finned metal radiant panel has fins arranged inline, which can increase the contact area with the air, strengthen the heat transfer capacity of the copper tube, and increase the convective heat transfer capacity of the radiant panels. The medium in the copper tube is R410A, and the groove-type shielding plate drains the condensed water generated under refrigeration conditions. The groove-type shielding plate is made of stainless steel and has a built-in thermal insulation material to reduce the thermal conductivity of the shielding plate, which can reduce the risk of condensation on the lower surface. The schematic diagram of heat exchange of finned metal radiant plate is shown in Figure 2. It can be seen from Figure 2 that the lower surface of the entire finned radiant plate transfers the cooling down. Due to the temperature difference between the upper and lower surfaces of the air interlayer, natural convection and radiation heat transfer will occur in the air interlayer. The cooling energy is transferred to the upper surface of the groove-type shielding plate; then the upper surface of the rear groove-type shielding plate is transferred to the lower surface through heat conduction; and finally, the lower surface of the groove-type shielding plate participates in the heat exchange of the indoor environment. Table 1 lists the structural parameters of the finned metal radiant panels.

2.2. Air Conditioning System Coupled with Direct Expansion Ceiling Radiation and Independent Fresh Air. In this study, the self-designed direct-expansion ceiling radiation and independent fresh air coupling air-conditioning system adopts the form of one machine and multiple connections. It is made up of a radiant terminal, a fresh air ventilator, and a heat pump air conditioner host, as can be seen in Figure 3. The radiation terminal is composed of four finned heat exchange plates connected in series, which have the

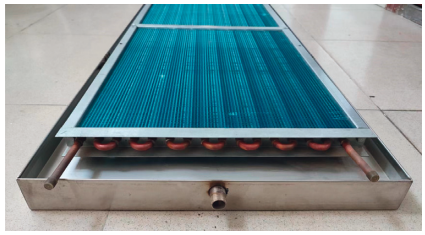


Figure 1: Finned metal radiant panel.

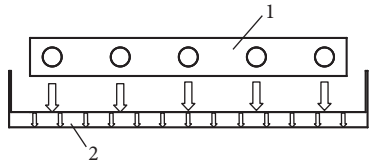


Figure 2: Schematic diagram of heat transfer of finned metal radiant plate. 1 – the finned radiant plate and 2 – grooved shielding plate.

characteristics of a compact structure and a large heat exchange area. The compound air conditioning system directly uses refrigerant R410 A as a heat transfer medium to heat or cool the air. In addition, the fresh air ventilator adopts the direct expansion freezing and dehumidification method, which can obtain a better dehumidification effect. Compared with the traditional radiant air-conditioning system using water as the medium, the air-conditioning system proposed in this study no longer needs a complex water mixing center and a dew point monitoring system because it does not require a special waterway system, so it has some advantages, including equipment small footprint, simple structure, easy control, convenient inspection, and maintenance. In addition, it has the advantages of energy saving and environmental protection, shortened response time, and avoidance of condensation.

The coupled air conditioning system designed in this study can avoid cold condensation dew, mainly for two reasons. First, the fresh air system adopts direct expansion and freezing dehumidification, and the fresh air is sent indoors after dehumidification so that the fresh air can be treated to avoid condensation. Second, each finned metal radiation plate is equipped with a groove-type shielding plate below, which is to prevent condensation and discharge condensate when refrigeration; even if the radiation plate condensation, the groove-type shielding plate can receive the drops of water falling on the surface of the finned heat exchange plate and discharge through the drain pipe.

3. Experimental Platform

3.1. Introduction of the Experimental Platform. Experimental platform structure diagram is shown in Figure 4. The experimental platform is mainly composed of two parts, one is outdoor, which simulates climate change using an artificial climate chamber. Its spatial structure size (length, width, and height) is 7.7 m × 5.7 m × 3.5 m. The other part is the room, which is used to test the effect of indoor heating, refrigeration, and fresh air exchange. Its spatial structure size (length, width, and height) is 4.16 m × 4.06 m × 3 m. The outdoor environment can be controlled by temperature and humidity (the temperature adjustment range is −20°C to 55°C, and the relative humidity of the outer room can be adjusted to more than 90%). In order to ensure that the experimental room in this study has good thermal insulation, the envelope structure of the inner room adopts polyurethane thermal insulation material.

3.2. Arrangement of Finned Metal Radiant Panels. The distribution of the four finned radiant panels on the ceiling of the room in the experiment is shown in Figure 5. The radiant panel is 100 mm away from the top of the ceiling. It can be seen from its sectional view that 1 is the thermal insulation layer, which can prevent the loss of cooling capacity. A layer of aluminum foil is attached under the heat insulation layer as a reflective layer (marked as 2 in Figure 5) to reflect the radiant cooling from the finned radiant plate. Number 3 is a finned metal radiant plate, which exchanges heat to the indoor environment in the form of natural convection and radiation heat transfer. Number 4 in Figure 5 is a groove-type shielding plate, which can collect the dew condensation water of the radiant panel. There is a heat insulation layer inside the groove-type shielding plate, which plays a buffering role in energy conduction, so that the energy conduction in the room is uniformly distributed. In the artificial simulated climate chamber designed in this study, the total area of the four finned metal radiant panels is 2.38 m^2; the total area of the inner room ceiling is 16.89 m^2; and the installation area of the finned metal radiant panels only accounts for 14% of the ceiling.

3.3. Experimental Method. In the experiment of this study, by changing the outdoor ambient temperature, the system is set to four operating conditions, including low-temperature heating, normal heating, normal cooling, and high-temperature cooling, each of which have different temperature ranges. The indoor space test point of this experiment is to measure the temperature, humidity, and wind speed of three planes, which are $Z = 0.1$ m (the position of the human ankle), $Z = 1.1$ m (head position when sitting), and $Z = 1.7$ m (head position when standing). Finally, according to the measurement data of temperature, humidity, and wind speed, it is quantitatively determined what the PMV-PPD value is for interior cooling and heating situations. The time interval for collecting experimental data in this study was 1 min. In order to ensure the accuracy of the experimental data, each plane adopts the method of multipoint measurement and averaging, thereby minimizing the error.

3.4. Test Parameters and Experimental Apparatus. The test instruments used in this study include a multichannel paperless recorder, a temperature, and humidity transmitter, a

Table 1: Fin radiation plate parameters.

Copper pipe		Radiator fan	
Material	Cold drawn tubes	Material	Hydrophilic aluminum
Length	1,650 mm	Area	350 mm × 34 mm
Thermal conductivity	379.14 W/(m.K)	Thermal conductivity	237 W/(m.K)
Diameter	3/8" × δ0.7	Thickness	0.105 mm
Bay number	Single row	Spacing	1.8 mm
Quantity	16	Quantity	841

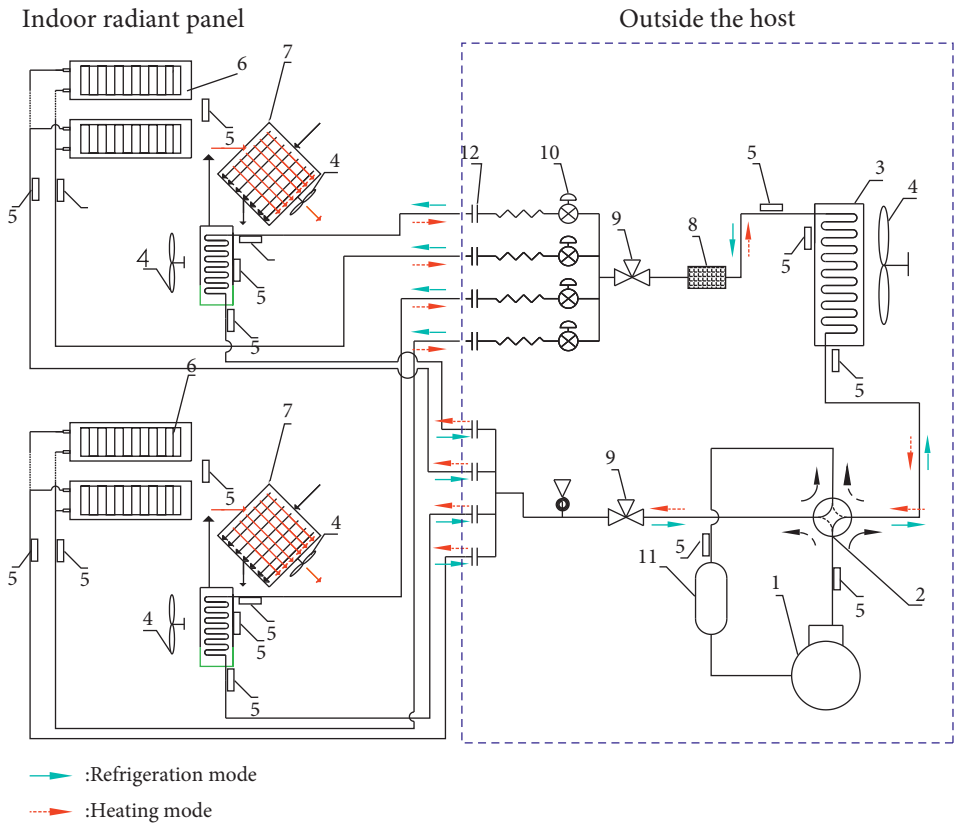


Figure 3: Schematic diagram of the direct-expansion ceiling radiation and independent fresh air coupled air conditioning system. 1 – compressor, 2 – four-way reversing valves, 3 – heat exchanger, 4 – ventilation fan, 5 – temperature sensor, 6 – radiation terminal, 7 – heat exchange core, 8 – filter, 9 – main service valve, 10 – electronic expansion valve, and 11 – liquid storage tank.

PT100 temperature sensor, an automatic temperature and humidity recorder, and an anemograph. The parameters of each instrument are shown in Table 2.

3.5. Location Distribution of Experimental Test Points. In this paper, the indoor thermal comfort of the composite air conditioning system under artificial simulated climate conditions is studied. It mainly analyzes the indoor temperature field distribution, so it is necessary to arrange temperature and humidity sensors in different parts of the room. In the experimental design, the sensors are arranged in different places, including the surrounding walls of the room, the center of the ceiling and the floor, as well as the air inlet and outlet. In addition to the enclosure structure, it is also necessary to install some sensors for the three planes of $z = 0.1$ m, $z = 1.1$ m, and $z = 1.7$ m in the indoor space and ensure that five sensors are installed on each surface. A schematic diagram of the arrangement of the temperature and humidity sensors in three planes is shown in Figure 6.

The specific position coordinates of the test points in the indoor space are shown in Table 3.

3.6. Experimental Working Condition. All the experimental instruments must be calibrated before the experiment begins. At the beginning of each experiment, firstly turn on the air conditioner of the outer room to make the temperature of the outer room reach the setting range of the corresponding working condition and then open the door of the inner room to keep the temperature of the inner room and the outer room as the same as possible. Finally, the door to the inner room needs to be closed. Table 4 is the specific experimental scheme of this study.

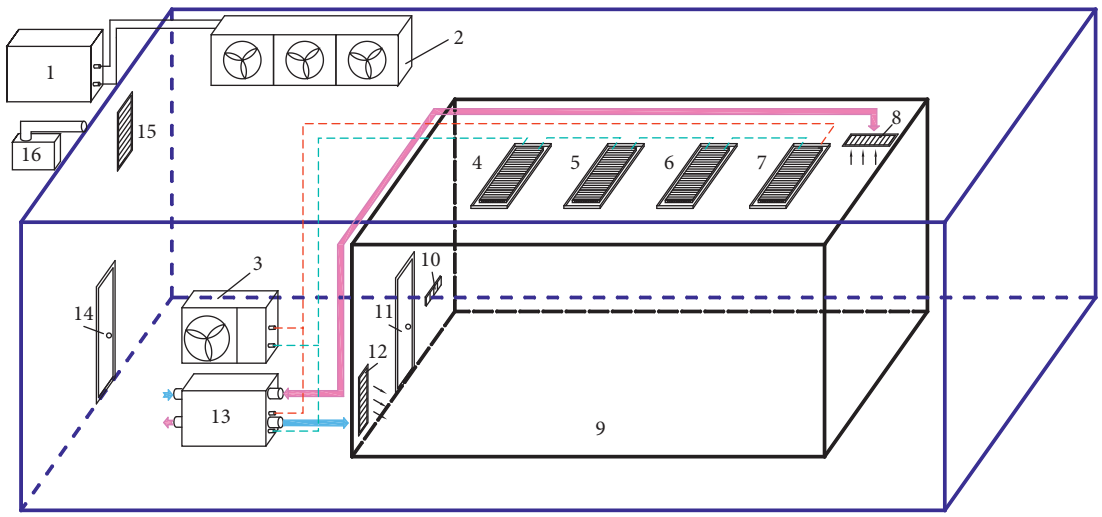


Figure 4: Schematic diagram of the experimental platform of this study. 1 – outdoor environment host, 2 – indoor machine for the outdoor environment, 3 – heat pump air conditioning unit, 4–7 – finned metal radiant panels, 8 – return air outlet, 9 – indoor environment test room, 10 – control panel, 11 – inner door, 12 – fresh air outlet, 13 – fresh air ventilator, 14 – outer door, and 15 – humidifier outlet.

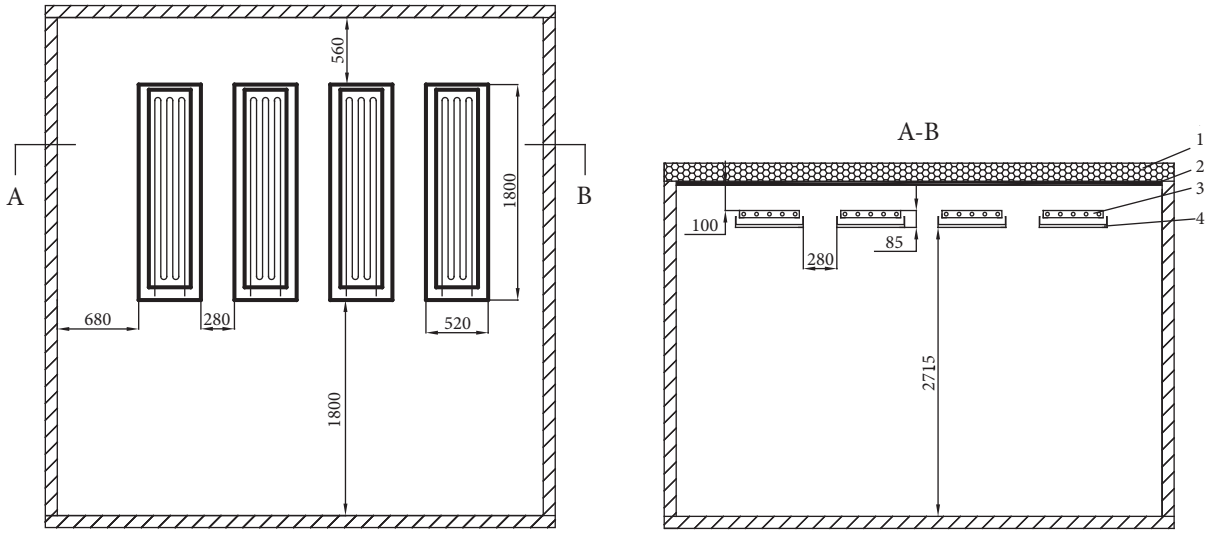


Figure 5: Plan and sectional view of radiant panel arrangement. 1 – thermal insulation layer, 2 – reflective layer, 3 – finned radiant plate, and 4 – groove-type shielding plate.

Table 2: Experimental instrument parameters used in this study.

Apparatus	Measuring range	Accuracy of measurement
Multichannel paperless recorder	Temperature: −50°C~400°C; Humidity: 0~100%RH	0.1/0.1%
Temperature and humidity transmitter	Temperature: −20°C~80°C; Humidity: 0~100%RH	±0.4/±3%
PT100 temperature sensor	−200°C~200°C	0.5
Temperature and humidity recorder	Temperature: −20°C~70°C; Humidity: 0~100%RH	0.1
Anemometer	0~30 m/s	±0.015

4. Experimental Results and Analysis

Each working condition runs for 24 hours, and each test point's temperature and humidity are automatically recorded every minute. Finally, an experimental curve is drawn based on data with a sampling interval of half an hour for analysis. Before each working condition test, turn on the air-conditioning equipment in the artificial climate environment laboratory to make the outdoor simulated temperature reach the set temperature range. Open the door of the inner room before the test so that the initial indoor temperature is basically the same as the temperature of the artificial

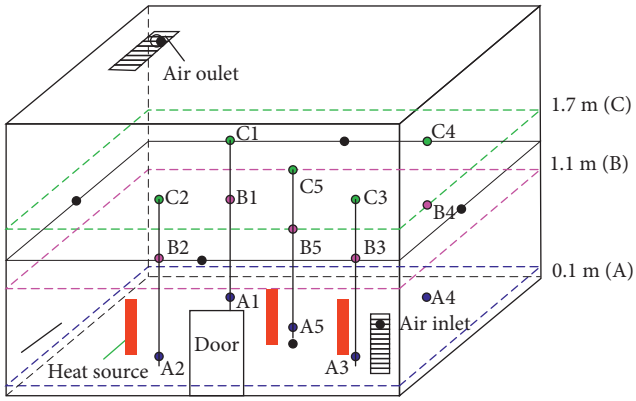


Figure 6: Sensor layout diagram in the room.

simulated outdoor environment. After closing the door of the inner room, the system operation mode is turned on, and the temperature and humidity changes in the room are recorded. This paper mainly studies the temperature distribution of five test points on the $z = 0.1$ m, $z = 1.1$ m, and $z = 1.7$ m horizontal planes.

4.1. Variation of Indoor Temperature Field under Cooling Conditions

4.1.1. Common Refrigeration Conditions. The experimental conditions of ordinary refrigeration conditions are set as the outdoor environment simulation temperature is 28°C~30°C, the relative humidity is 75%, the indoor heat source is set at 300 W, and the indoor air conditioning temperature is set at 25°C. The heat source arrangement is shown in Figure 6.

When the radiation and fresh air modes are turned on at the same time in the normal cooling mode in summer (the outdoor simulated temperature is 28°C~30°C), the temperature of the three indoor horizontal surfaces and the average temperature distribution of each horizontal surface are shown in Figure 7. Figures 7(a)–7(c) show that the temperature changes in the three horizontal planes follow the same pattern. This pattern is demonstrated by the fact that the indoor temperature drops to the set temperature (25°C), the system is turned off for an hour, and then the indoor temperature rises. When the indoor temperature exceeds 25°C, the air conditioning unit will start again. After about half an hour of operation, the indoor temperature will drop below 25°C. Then the air conditioning unit was shut down again for 1 hour. The air conditioning units of the air conditioning system designed in this study basically start and stop according to this rule. After the system is turned on for about 2 hours, it reaches a steady state. During the test time after the system is stabilized, the temperature in the room is kept between 23.4°C and 26.3°C. When the whole system is in a steady state, the temperature difference of the horizontal surface of $z = 0.1$ m is within 0.9°C; the temperature difference of the horizontal surface of $z = 1.1$ m is within 1.1°C; and the temperature difference of the horizontal surface of $z = 1.7$ m is within 0.9°C at the same time. In addition, the maximum difference in temperature of the vertical plane is within 1.8°C. The above temperature distribution characteristics indicate that in the human activity area, the temperature distribution of the indoor space is relatively uniform.

4.1.2. High-Temperature Refrigeration Conditions. In this study, under the condition of high-temperature refrigeration, the experimental settings are that the simulated outdoor temperature is 37°C~39°C, the relative humidity is 80%, the indoor heat source is set at 300 W, and the indoor air conditioning temperature is 25°C.

When the radiation and fresh air systems are opened at the same time, the distribution of the temperature in the three indoor horizontal planes and the average temperature in each horizontal plane under the condition of high-temperature refrigeration in summer (outdoor simulated temperature ranges from 37°C to 39°C) is shown in Figure 8. It can be seen from Figures 8(a)–8(c) that the temperature variation trend of the three horizontal planes is consistent. It is shown that the indoor temperature drops below the preset temperature (25°C) about 1 hour after the system is turned on, and then the system is shut down for 1 hour, followed by the indoor temperature rise. When the indoor temperature exceeds 26°C, the air conditioning unit will start again. After the equipment works for about half an hour, the indoor temperature drops to below 24.5°C again, and then the air conditioning unit shuts down again for 1 hour. The air conditioning unit of the whole air conditioning system basically starts and stops in accordance with this rule. After about 3 hours, the whole system reaches a steady state. After the system is stable, the indoor temperature is kept between 23.6°C and 26.6°C. According to the test results, at the same time, the horizontal plane temperature difference of $Z = 0.1$ m is kept within 1.5°C; the horizontal plane temperature difference of $Z = 1.1$ m is kept within 1.5°C; and the horizontal plane temperature difference of $Z = 1.7$ m is kept within 1.3°C. It can be seen from Figure 8(d) that the maximum temperature difference in the vertical plane is within 2.1°C. The above experimental analysis results show that the indoor temperature distribution of the system is relatively uniform under the high-temperature refrigeration condition.

4.2. Variation of Indoor Temperature Field under Heating Conditions

4.2.1. Normal Heating Condition. In the test experiment of this study, the experimental settings under the normal heating condition are that the outdoor environment simulation temperature is 1°C~5°C, the relative humidity is 30%, and the indoor air conditioner setting temperature is 26°C.

In this experiment, when the radiation and fresh air modes are turned on at the same time, the distribution of indoor temperature in three horizontal planes and the average temperature in each horizontal plane under the conditions of ordinary heating in winter (outdoor simulated temperature ranges from 1°C to 5°C) is shown in Figure 9. From Figures 9(a)–9(c), it can be seen that the experimental

Table 3: Test point location parameters.

Plane $Z = 0.1$	Coordinates	Plane $Z = 1.1$	Coordinates	Plane $Z = 1.7$	Coordinates
A1	(1, 2.8, 0.1)	B1	(1, 2.8, 1.1)	C1	(1, 2.8, 1.7)
A2	(1, 1.1, 0.1)	B2	(1, 1.1, 1.1)	C2	(1, 1.1, 1.7)
A3	(3, 1.1, 0.1)	B3	(3, 1.1, 1.1)	C3	(3, 1.1, 1.7)
A4	(3, 2.8, 0.1)	B4	(3, 2.8, 1.1)	C4	(3, 2.8, 1.7)
A5	(2, 1.9, 0.1)	B5	(2, 1.9, 1.1)	C5	(2, 1.9, 1.7)

Table 4: Experimental design of radiation accompanying fresh air system.

	Outdoor temperature (°C)	Outdoor relative humidity (RH; %)	Indoor setting temperature (°C)
Low-temperature heating	−10~−6	30	26
Ordinary heating	1~5	30	
Ordinary refrigeration	28~30	75	25
High-temperature cooling	37~39	80	

system of three horizontal surface temperature changing trends is consistent, if the indoor temperature exceeds the preset temperature (26°C) after two hours of continuous operation, the system automatically stops working for two hours, then the system automatically stop 2 hours, then the indoor temperature drops. When the indoor temperature is lower than the preset temperature, the air conditioning unit will start the heating work again. After about 1 hour of work, the indoor temperature exceeds the preset temperature again, and then the air conditioning unit shuts down again for 1.5 hours. In the subsequent experimental test, the air conditioning unit basically starts and stops working in accordance with this rule. In the above working cycle, the air conditioning unit is turned on for 1 hour and shut down for 1.5 hours. It takes two hours for the system to reach a steady state and start recording data. Recorded experimental data show that the temperature in most areas of the room is always maintained between 25°C and 28°C. In the steady state of the system, the horizontal plane temperature difference of $Z = 0.1$ m at the same time is within 0.9°C; the horizontal plane temperature difference of $Z = 1.1$ m is within 1.5°C; and the horizontal plane temperature difference of $Z = 1.7$ m is within 1°C. In addition, the maximum temperature difference in the vertical plane of the interior space is within 2.4°C. The above experimental analysis results show that the indoor temperature distribution of the system is relatively uniform under the normal heating condition.

4.2.2. Low-Temperature Heating Condition. In the low-temperature heating condition of this study, the experiment is set as the simulated outdoor environment temperature of [−10°C, −6°C], relative humidity of 30%, and indoor air conditioning temperature of 26°C.

When the radiation and fresh air modes are opened at the same time, the distribution of indoor temperature in three horizontal planes and the average temperature in each horizontal plane under the conditions of low-temperature heating in winter (outdoor simulated temperature ranges −10°C to −6°C) is shown in Figure 10. It can be seen from Figures 10(a)–10(c) that the temperature variation trend of the three horizontal planes is consistent. The working process of the air conditioning system is that the indoor temperature rises above the preset temperature (26°C) about three hours after the system is turned on. Then the system shuts down for an hour, and the temperature drops. When the indoor temperature is lower than the preset temperature, the air conditioning unit will start again. After the system equipment works for about 1 hour, the indoor temperature rises to above 26°C again, and then the air conditioning unit shuts down again for 1 hour. In the subsequent experimental test, the air conditioning unit basically starts and stops working in accordance with this rule. The air conditioning system works for 1 hour and stops for 1 hour in a cycle. The air conditioning system proposed in this study reaches a steady state after about 5 hours, and the temperature in most areas of the room always stays between 27°C and 30°C after the system is stabilized. In the laboratory, the temperature difference of the inner room in the horizontal plane $z = 0.1$ m is within 2.9°C, in the horizontal plane $z = 1.1$ m is within 2.4°C, and in the horizontal plane $z = 1.7$ m is within 1.9°C. In addition, the maximum temperature difference in the vertical plane of the inner room is within 3°C. In conclusion, the experimental test results of the air conditioning system designed in this study show that the indoor temperature distribution is relatively uniform under working conditions of low-temperature heating.

5. Dehumidification Efficiency Analysis of Fresh Air System

Since dehumidification process exists in both the refrigeration condition and the fresh air system, this study analyzed the dehumidification efficiency of the fresh air system in three working conditions: ordinary fresh air refrigeration (the temperature range is 31°C to 33°C), high-temperature refrigeration (the temperature range is 37°C to 39°C), and dehumidification in a transition season (the temperature range is 20°C to 22°C). The moisture content of outdoor air and indoor air under the above three working conditions is shown in Table 5.

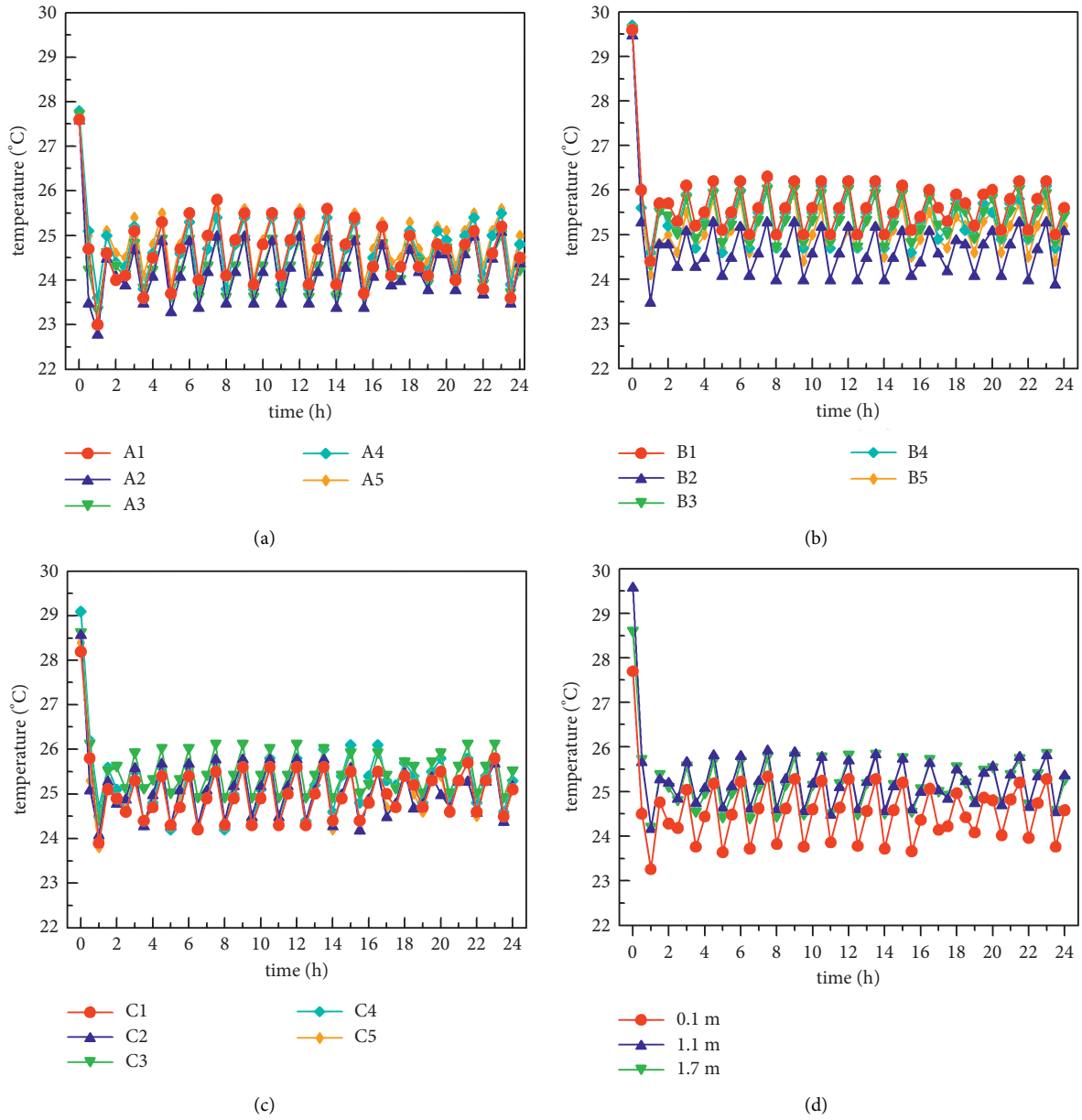


Figure 7: Indoor temperature distribution under normal refrigerating condition (28°C~30°C): (a) $Z = 0.1$ m horizontal temperature distribution, (b) $Z = 1.1$ m horizontal temperature distribution, (c) $Z = 1.7$ m horizontal temperature distribution, and (d) average temperature distribution of each horizontal plane.

It can be seen from Table 5 that the average moisture content of outdoor air and indoor air is 11.33 g/kg and 5.48 g/kg under dehumidification conditions in the transition season, indicating that the average dehumidification efficiency of the fresh air system is 51.6% under such conditions. The average moisture content of outdoor air and indoor air is 22.91 g/kg and 8.9 g/kg under ordinary fresh air refrigeration, indicating that the average dehumidification efficiency of the fresh air system is 61.2% under such conditions. The average moisture content of outdoor air and indoor air is 32.6 g/kg and 9.63 g/kg under high-temperature refrigeration, indicating that the average dehumidification efficiency of the fresh air system is 70.4% under such conditions. The experimental results show that the average dehumidification efficiency is more than 50% under the above three working conditions, which is sufficient to prove that the fresh air system designed in this study has a good dehumidification effect.

6. Thermal Comfort Analysis

6.1. PMV-PPD. The factors affecting thermal comfort can be divided into two categories: one is environmental factors, including air temperature, airflow velocity, partial pressure of water vapor in the air, and average radiant temperature. Another category is human factors, including human

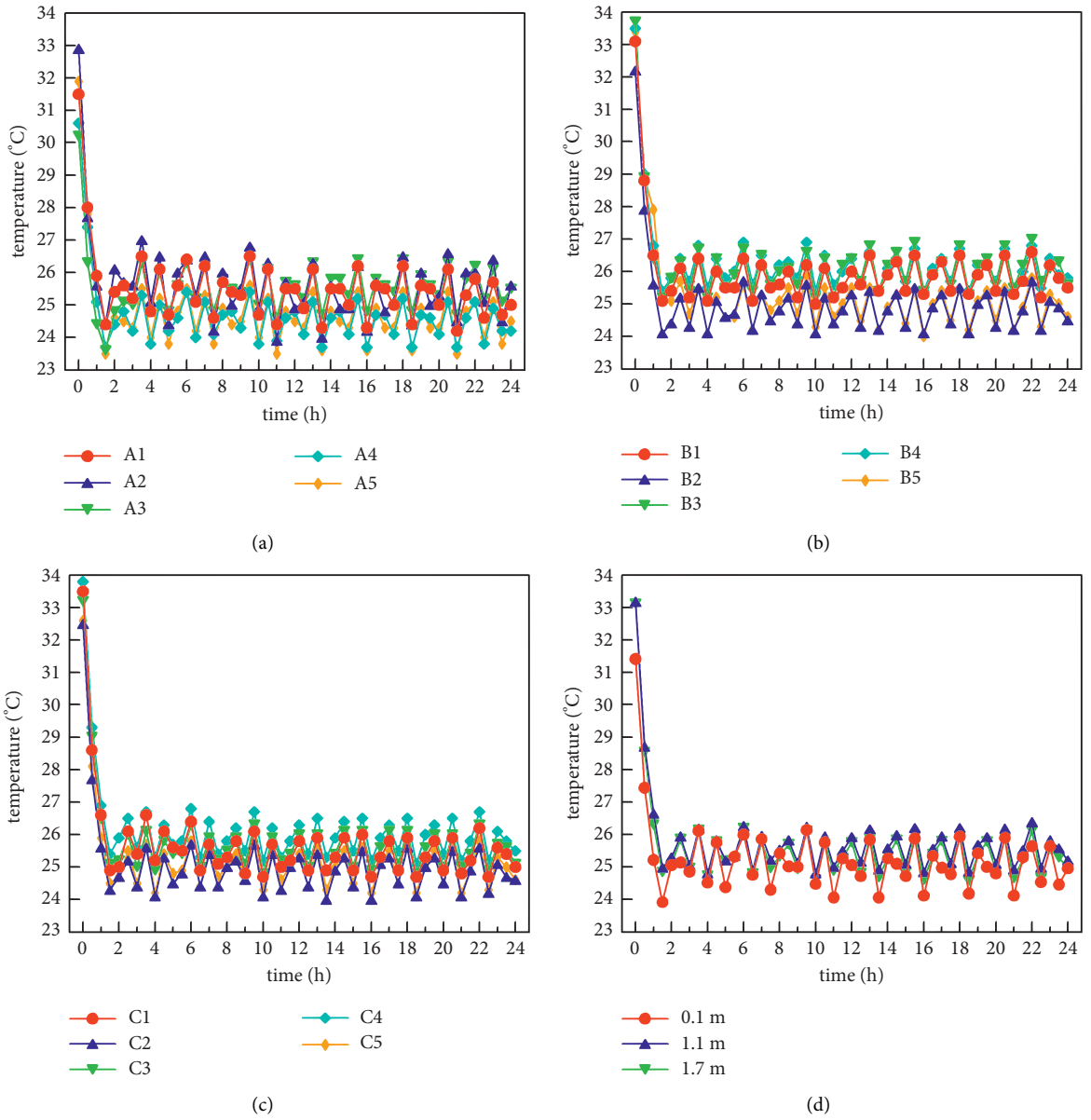


Figure 8: Indoor temperature distribution under high-temperature refrigeration condition (37°C~39°C): (a) $Z = 0.1$ m horizontal temperature distribution, (b) $Z = 1.1$ m horizontal temperature distribution, (c) $Z = 1.7$ m horizontal temperature distribution, and (d) average temperature distribution of each horizontal plane.

metabolic rate and thermal resistance of clothing. Professor Fanger expanded the thermal comfort equation and the calculation formula of PMV through the data of many subjects. The equation for calculating PMV is

$$\begin{aligned}\mathrm{PMV} = {} & \left(0.303e^{-0.036M} + 0.028\right) \\ & \times \left\{(M - W) - 3.05 \times 10^{-3} \times [5733 - 6.99(M - W) - P_a] - 0.42 \times [(M - W) - 58.15] - 1.7 \times 10^{-5} M (5867 - P_a)\right. \\ & \left. - 1.4 \times 10^{-3} M (34 - t_a) - 3.96 \times 10^{-8} f_{cl} \times \left[(t_{cl} + 273)^4 - (t_r + 273)^4\right] - f_{cl} \cdot h_c \cdot (t_{cl} - t_a)\right\},\end{aligned} \tag{1}$$

where M is metabolic rate, Met; W stands for human body power, W/s; P_a is the partial pressure of water vapor in the air, Pa; t is the air temperature, °C; f_{cl} is the ratio of the surface area of the clothed body to the naked body; t is the mean radiation temperature, °C; h_c is the convective heat exchange coefficient, W/(s·m2°C); and t_{cl} is the

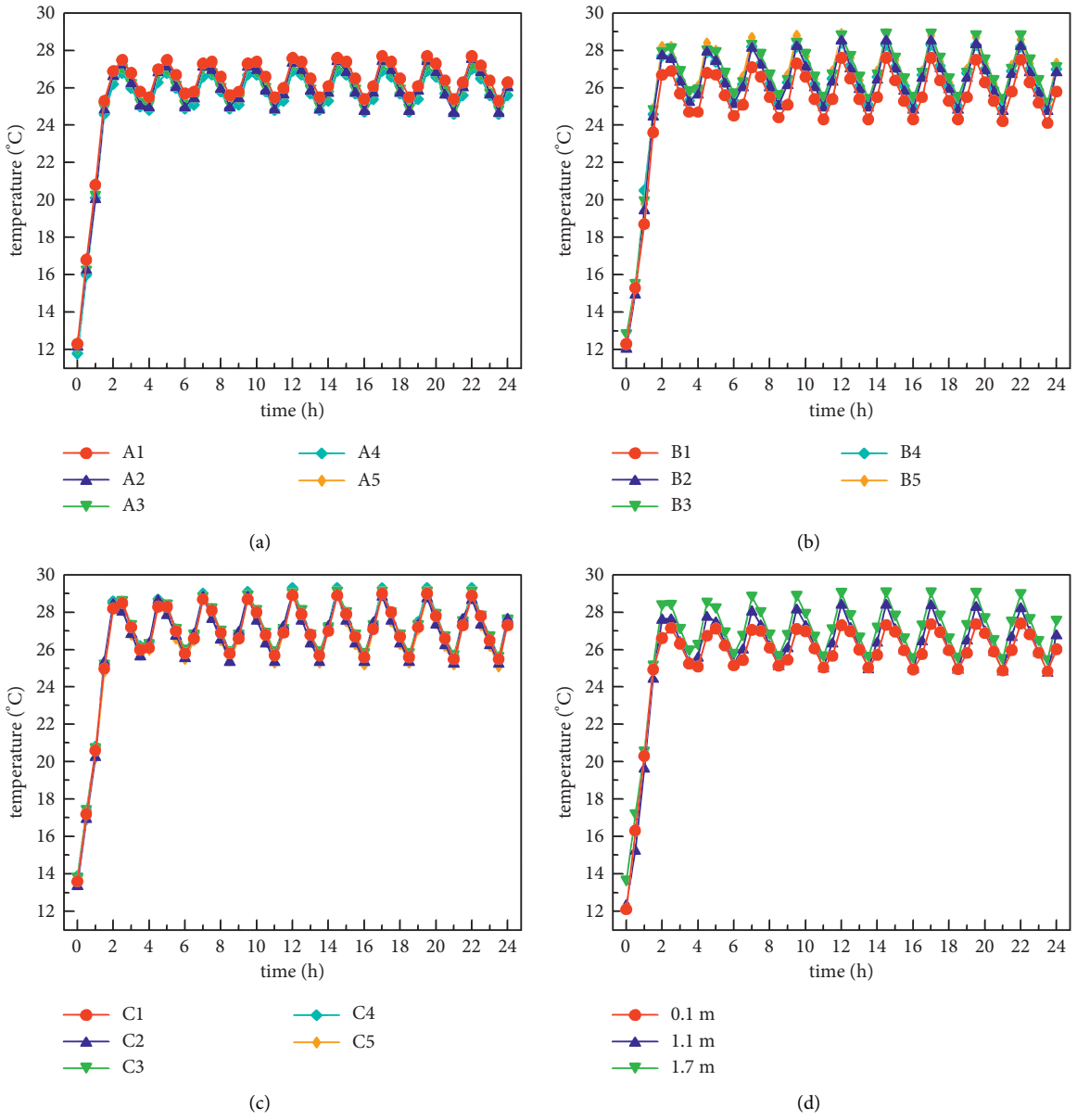


Figure 9: Indoor temperature distribution under normal heating conditions (1°C~5°C): (a) $Z = 0.1$ m horizontal temperature distribution, (b) $Z = 1.1$ m horizontal temperature distribution, (c) $Z = 1.7$ m horizontal temperature distribution, and (d) average temperature distribution of each horizontal plane.

average temperature of the outer surface of the wearer's body, °C.

When the PMV value is determined, the PPD value can be calculated by the following formula:

$$\mathrm{PPD} = 100 - 95\ \exp\left[-\left(0.03353 \times PMV^4 + 0.2179PMV^2\right)\right]. \quad (2)$$

6.2. Indoor Thermal Comfort Analysis. When the indoor temperature and humidity change is basically stable after the air conditioning unit runs stably, the thermal comfort index PMV-PPD value of the indoor $z = 1.1$ m plane is calculated. The calculation results of thermal comfort are shown in Table 6.

It can be seen from Table 6 that thermal comfort indexes of all working conditions are within the standard range ($-1 \le \mathrm{PMV} \le 1$, $\mathrm{PPD} \le 27\%$), indicating that indoor thermal comfort is better under the three air conditioning operating modes proposed in this study. However, by comparing different modes, it is found that indoor comfort with both radiation and fresh air opened at the same time is better than that with radiation or fresh air opened alone. The PMV value and the PPD value are 0.13 and 5.2, respectively, when radiation and fresh air are opened at the same time in summer under normal refrigeration conditions. However, the PMV value and the PPD value are 0.22 and 5.41,

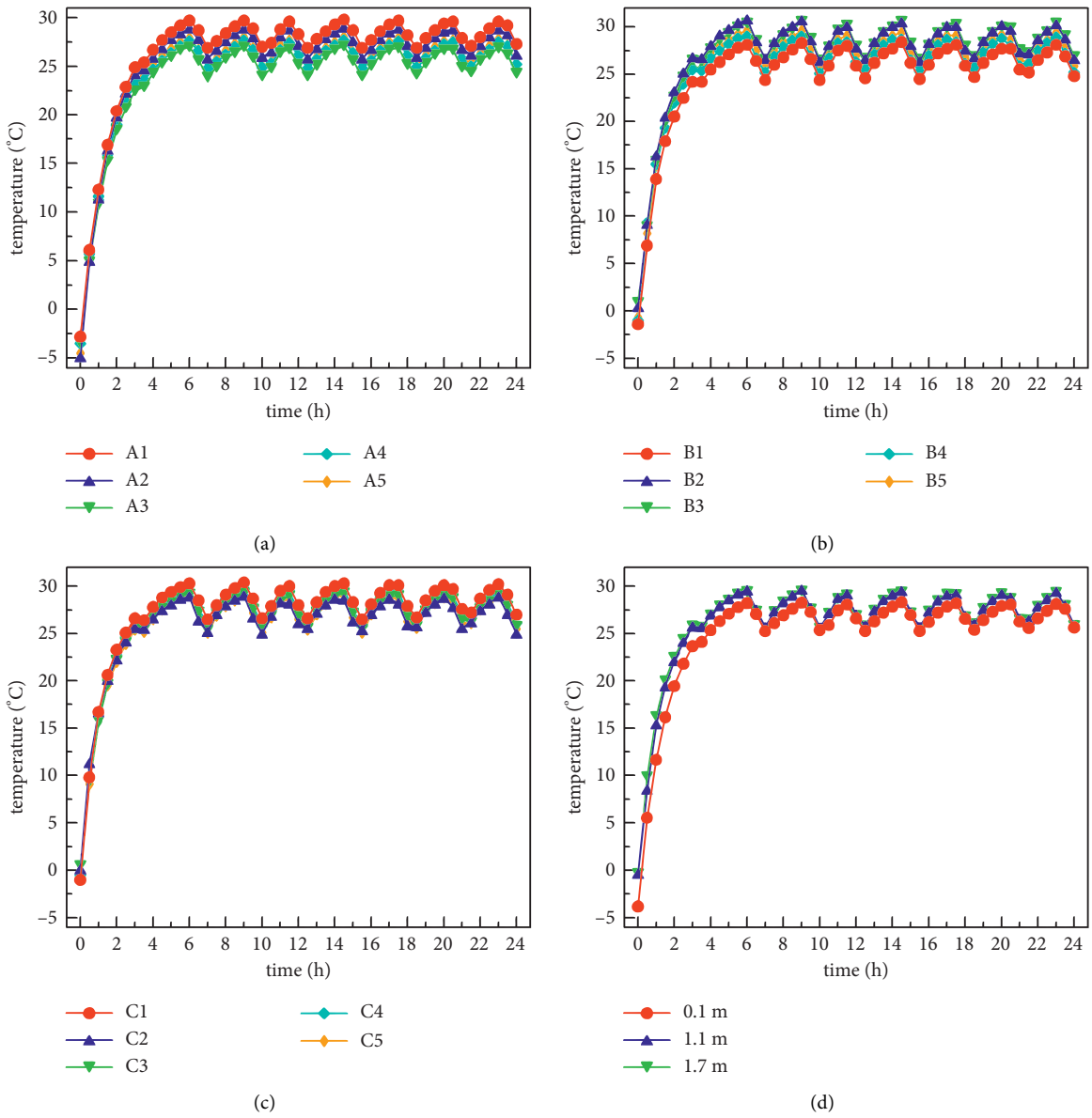


Figure 10: Indoor temperature distribution under low-temperature heating condition (–10°C~–6°C): (a) $Z = 0.1$ m horizontal temperature distribution, (b) $Z = 1.1$ m horizontal temperature distribution, (c) $Z = 1.7$ m horizontal temperature distribution, and (d) average temperature distribution of each horizontal plane.

Table 5: Moisture content of outdoor air and indoor fresh air.

Three kinds of working condition	Air samples	Average temperature (°C)	Average relative humidity (%)	Moisture content (g/kg)	Dehumidification efficiency (%)
Dehumidification in transition season (20°C~22°C)	Air from the indoor air outlet	13.6	56.80	5.48	51.6
	Outdoor air	21.5	70.69	11.33	
Ordinary refrigeration (31°C~33°C)	Air from the indoor air outlet	18.4	67.5	8.9	61.2
	Outdoor air	32.2	74.8	22.91	
High-temperature cooling (37°C~39°C)	Air from the indoor air outlet	18.6	72.1	9.63	70.4
	Outdoor air	37.5	78.2	32.6	

Table 6: Indoor thermal comfort index test results.

Air conditioner operating mode	PMV-PPD value	General refrigeration (28°C~30°C)	High-temperature refrigeration (37°C~39°C)	Common heating (11°C~15°C)	Low-temperature heating (−5°C~−1°C)
Radiation	PMV	0.25	0.54	−0.34	−0.45
	PPD	5.59	7.17	5.87	6.21
Fresh air	PMV	0.22	0.49	−0.27	-0.4
	PPD	5.41	6.91	5.61	5.94
Radiation and fresh air at the same time	PMV	0.13	0.47	−0.23	−0.38
	PPD	5.2	6.82	5.39	5.75

respectively, when only fresh air is turned on, and the PMV value and the PPD value are 0.25 and 5.59, respectively, when only radiation is turned on. Under normal heating conditions in winter, when radiation and fresh air are opened at the same time, the PMV value and the PPD value are −0.23 and 5.39, respectively. However, the PMV value and the PPD value are −0.27 and 5.61, respectively, when only fresh air is turned on, and the PMV value the PPD value are −0.34 and 5.87, respectively, when only radiation is turned on. The above comparison results are sufficient to show that indoor thermal comfort is the best when radiation and fresh air are opened at the same time, and it is the worst when only radiation is opened. However, the three operating modes can meet the requirements of indoor thermal comfort. To put it simply, in the composite air conditioning system of ceiling radiation combined with an independent fresh air system, the radiation system bears part of the sensible heat load of indoors, while the independent fresh air system bears part of the sensible heat load and all the latent heat load of indoors. The combination of the two systems can provide a good thermal comfortable environment for the indoors.

7. Conclusion

This paper focuses on the technology of direct-expansion ceiling radiation and independent fresh air coupled with air conditioning systems to carry out scientific research. The temperature distribution uniformity of the indoor vertical plane and the horizontal plane is mainly studied when the radiation and fresh air operation modes are turned on at the same time. In addition, a detailed comparative analysis of the dehumidification effect of the system under refrigeration and dehumidification conditions is carried out. Finally, according to the measured data, the indoor thermal comfort index PMV-PPD value under different modes is calculated. The conclusions reached are as follows:

(1) This research system adopts the form of coupling of direct expansion ceiling radiation and independent fresh air. The air conditioning system is composed of a heat pump air conditioning host, a new fan, and a radiation terminal. It directly uses refrigerant as a heat transfer medium to heat or refrigerate the air, and the new fan adopts direct expansion refrigeration dehumidification.

(2) The system adopts a fin heat exchanger with the direct expansion of refrigerant. The large temperature difference between cooling and heating can be realized by the inline arrangement of finned metal radiation plates. Therefore, the system we designed not only has the characteristics of large heat flow per unit area, sufficient cooling capacity, and a simplified heat exchange system but also has the installation area of the finned metal radiation plate that only accounts for 14% of the ceiling.

(3) The indoor temperature field test results show that the maximum temperature difference at the same time in the room is less than 2.1°C when radiation and fresh air mode are coupled under refrigeration conditions, and the maximum temperature difference at the same time in the room is less than 3°C when radiation and fresh air mode are coupled under heating condition. In general, the indoor temperature field is uniformly distributed when radiation and fresh air mode are coupled.

(4) The dehumidification ability test of the system shows that the fresh air moisture content after dehumidification is between 5.48 g/kg and 9.63 g/kg, and the average dehumidification efficiency is more than 50%, which proves that the fresh air mode has obvious dehumidification effect.

(5) The indoor thermal comfort index PMV and PPD calculated based on the experimental test data are $-0.34 \le \text{PMV} \le 0.54$ and $5.2 \le \text{PPD} \le 5.87$, indicating that the indoor thermal comfort of the system designed in this study is relatively good.

The above research results show that the direct-expansion radiation and fresh air coupling air conditioning system proposed in this paper can realize independent control of temperature and humidity, which is very suitable for office and home use and has a good application prospect. In addition, the large temperature difference between refrigeration and heating radiation technology applied in the system, especially the direct-expansion radiative heat transfer technology, has certain novelty and creativity. However, there are still many technical and theoretical problems that need to be further improved to fully transform these technologies into practical engineering applications. In the future, further research can be done in the following areas to solve more problems that are scientific.

(i) Relevant researchers can carry out research on innovative air-conditioning technology and theoretical system based on large temperature differences and strong cold and heat radiation, especially on the mechanism of a large temperature difference and strong cold and heat radiation to achieve a uniform temperature of indoor radiation surface.

(ii) With the support of the theory of heat transfer and computational fluid dynamics, scholars can carry out the simulation of heat and mass transfer in a complex thermal radiation environment.

(iii) The subsequent related research work can focus on establishing the numerical analysis model of temperature field and airflow field of indoor thermal environment, which can provide a strong theoretical basis and application reference for the design standards, equipment manufacturing standards, and installation specifications related to strong cold and heat uniform radiation and direct expansion radiation air conditioning in the future.

Data Availability

All data included in this study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest in this work.

Acknowledgments

This work was supported by the 12th FYPe National Key Technology R&D Program-Renewable Energy Utilization to Achieve Higher Level Building Energy Efficiency Target (Grant no. 2014BAJ01B03).

References

[1] X. Wu, J. Zhao, B. W. Olesen, L. Fang, and F. Wang, "A new simplified model to calculate surface temperature and heat transfer of radiant floor heating and cooling systems," *Energy and Buildings*, vol. 105, pp. 285–293, 2015.

[2] A. Bahadori, S. Zendehboudi, and K. Hooman, "Estimation of characteristic temperature ratios for panel radiator and high temperature radiant strip systems to calculate heat loss within a room space," *Energy and Buildings*, vol. 55, pp. 508–514, 2012.

[3] H. Tang, T. Zhang, X. H. Liu, and C. Li, "A novel approximate harmonic method for the dynamic cooling capacity prediction of radiant slab floors with time variable solar radiation," *Energy and Buildings*, vol. 223, pp. 110117–110295, 2020.

[4] L. Lv, Y. Luo, and C. Huang, "A synchronizing thermal model based on three kinds of radiant models to compute interior air temperature and interior wall temperature," *Journal of Building Engineering*, vol. 18, pp. 343–351, 2018.

[5] A. Merabtine, A. Kheiri, S. Mokraoui, and A. Belmerabet, "Semi-analytical model for thermal response of anhydrite radiant slab," *Building and Environment*, vol. 153, pp. 253–266, 2019.

[6] G. Yu, L. Xiong, C. Du, and H. Chen, "Simplified model and performance analysis for top insulated metal ceiling radiant cooling panels with serpentine tube arrangement," *Case Studies in Thermal Engineering*, vol. 11, pp. 35–42, 2018.

[7] X. Su, L. Zhang, Z. Liu, Y. Luo, J. Lian, and Y. Luo, "A computational model of an improved cooling radiant ceiling panel system for optimization and design," *Building and Environment*, vol. 163, Article ID 106312, 2019.

[8] Z. Jing and L. Jiayu, "Study on heat transfer delay of exposed capillary ceiling radiant panels (E-CCRP) system based on CFD method," *Building and Environment*, vol. 180, Article ID 106982, 2020.

[9] R. Khatri, V. R. Khare, and H. Kumar, "Spatial distribution of air temperature and air flow analysis in radiant cooling system using CFD technique," *Energy Reports*, vol. 6, pp. 268–275, 2020.

[10] L. Zhang, H. Fang, W. Wang, and J. Liu, "Energy-saving analysis of ground source heat pump combined with floor radiant air conditioning system," *Procedia Engineering*, vol. 205, pp. 4067–4073, 2017.

[11] J. Du, M. Chan, D. Pan, and S. Deng, "A numerical study on the effects of design/operating parameters of the radiant panel in a radiation-based task air conditioning system on indoor thermal comfort and energy saving for a sleeping environment," *Energy and Buildings*, vol. 151, pp. 250–262, 2017.

[12] P. Ding, Y. Li, E. Long, Y. Zhang, and Q. Liu, "Study on heating capacity and heat loss of capillary radiant floor heating systems," *Applied Thermal Engineering*, vol. 165, Article ID 114618, 2020.

[13] H. Sun, Y. Wu, B. Lin, M. Duan, Z. Lin, and H. Li, "Experimental investigation on the thermal performance of a novel radiant heating and cooling terminal integrated with a flat heat pipe," *Energy and Buildings*, vol. 208, Article ID 109646, 2020.

[14] G. Lv, C. Shen, Z. Han, W. Liao, and D. Chen, "Experimental investigation on the cooling performance of a novel grooved radiant panel filled with heat transfer liquid," *Sustainable Cities and Society*, vol. 50, Article ID 101638, 2019.

[15] H. Tang, X.-H. Liu, H. Li, Y. Zhou, and Yi Jiang, "Study on the reduction of condensation risks on the radiant cooling ceiling with superhydrophobic treatment," *Building and Environment*, vol. 100, pp. 135–144, 2016.

[16] W. Jin, L. Jia, Q. Wang, and Z. Yu, "Study on condensation features of radiant cooling ceiling," *Procedia Engineering*, vol. 121, pp. 1682–1688, 2015.

[17] X. Su, Z. Wang, Y. Xu, and N. Liu, "Thermal comfort under asymmetric cold radiant environment at different exposure distances," *Building and Environment*, vol. 178, Article ID 106961, 2020.

[18] D. Wang, G. Chen, C. Song et al., "Experimental study on coupling effect of indoor air temperature and radiant temperature on human thermal comfort in non-uniform thermal environment," *Building and Environment*, vol. 165, Article ID 106387, 2019.

[19] C. Cen, Y. Jia, K. Liu, and R. Geng, "Experimental comparison of thermal comfort during cooling with a fan coil system and radiant floor system at varying space heights," *Building and Environment*, vol. 141, pp. 71–79, 2018.

[20] Z. Tian and J. A. Love, "A field study of occupant thermal comfort and thermal environments with radiant slab cooling," *Building and Environment*, vol. 43, no. 10, pp. 1658–1670, 2008.

[21] I. Atmaca, O. Kaynakli, and A. Yigit, "Effects of radiant temperature on thermal comfort," *Building and Environment*, vol. 42, no. 9, pp. 3210–3220, 2007.

[22] Y. Zhou, S. Zheng, and G. Zhang, "A review on cooling performance enhancement for phase change materials integrated systems—flexible design and smart control with machine learning applications," *Building and Environment*, vol. 174, Article ID 106786, 2020.

[23] Y. Zhou, S. Zheng, and G. Zhang, "A state-of-the-art-review on phase change materials integrated cooling systems for deterministic parametrical analysis, stochastic uncertainty-based design, single and multi-objective optimisations with machine learning applications," *Energy and Buildings*, vol. 220, Article ID 110013, 2020.

[24] Y. Zhou, S. Zheng, Z. Liu et al., "Passive and active phase change materials integrated building energy systems with advanced machine-learning based climate-adaptive designs, intelligent operations, uncertainty-based analysis and optimisations: a state-of-the-art review," *Renewable and Sustainable Energy Reviews*, vol. 130, Article ID 109889, 2020.

[25] J. Široký, F. Oldewurtel, J. Cigler, and S. Privara, "Experimental analysis of model predictive control for an energy efficient building heating system," *Applied Energy*, vol. 88, no. 9, pp. 3079–3087, 2011.

[26] S. E. Shafiei, J. Stoustrup, and H. Rasmussen, "Model predictive control for flexible power consumption of large-scale refrigeration systems," *American Control Conference*, vol. 06, pp. 412–417, 2014.

[27] A. Beghi, M. Rampazzo, and S. Zorzi, "Reinforcement learning control of transcritical carbon dioxide supermarket refrigeration systems," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 13754–13759, 2017.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 7984852, 13 pages
https://doi.org/10.1155/2022/7984852

Research Article

Empirical Analysis of Customer Risk and Corporate Financing Constraints Based on Supply Chain Networks

Qun Bao,[1] **Ju-Ying Wang**,[1] **Rui Xie**,[1] **and Zheng-Qun Cai**[2]

[1]*College of Accounting, Anhui University of Finance and Economics, Bengbu 233030, China*
[2]*School of Foreign Studies, Anhui Jianzhu University, Hefei 230601, China*

Correspondence should be addressed to Zheng-Qun Cai; caizhengqun1983@163.com

Received 9 August 2022; Accepted 27 August 2022; Published 16 September 2022

Academic Editor: Wei Xiang

A supply chain's risk spillover effect will affect the customer's risk on the financing constraints of suppliers. This paper builds on the evaluation of customer risk by fuzzy mathematics, combines with the A-share listed companies in Shanghai and Shenzhen from 2007 to 2019 as a study sample, and empirically inspects the influence of customer risk on the level of corporate financing constraints. According to the study, it shows that the customer risk is currently at a moderate level, which will notably impair the supplier's external financing ability. This phenomenon is more remarkable when the monetary policy is tightened with fierce competition in the industry. This paper unveils the economic consequences of customer risk spillovers from a supply chain, enriches the study of the generation mechanism of corporate financing constraints, and provides investors and regulators with empirical evidence to appreciate corporate financing constraints.

1. Introduction

The issue of "difficult and exorbitant financing" has always been the main factor restricting China's economic growth. In 2016, President Xi Jinping, at an Economic Situation Expert Symposium, noted the following: "how to reduce corporate financing costs and mitigate financing constraints is pivotal in the supply-side reform." Pursuant to the Report on the Work of the Government in 2019, the alleviation of financing constraints in the real economy was listed as a priority for the government. In a multilevel capital market, information asymmetry and principal-agent problems among investors, creditors, and enterprises can crucially influence financing constraints. The enterprise financing capacity notably varies with the financial conditions, governance level, and social status among enterprises. As per the study results, it proves that the enterprise's external economic environment [1], credit environment [2], internal financial characteristic information [3], governance characteristic information [4], accounting information quality [5], and other factors affect the level of corporate financing constraints. In the real scenario, corporate financing constraints are restricted by the enterprise's internal and external environment and are affected by the risk spillover effects of upstream and downstream companies in the supply chain.

As an important stakeholder of enterprises' implicit contracts, customers have such a kind of interest relationship with suppliers that they will both prosper or lose. Customers will not only exert positive effect on upstream enterprises to integrate supply chains and conduct external supervision [6] but also bring the negative influence of risk transmission through supply chains [7]. If the enterprise of the customers of supply chain is faced with difficulties, the risks will result in spillover effects along the supply chain and bring uncertainty to the operation of enterprise, giving rise to the business crisis and financial crisis [8]. Therefore, the information about spillover effects, based on supply chain, will attract more investors and creditors. The existing study demonstrates that the risk information of customers will produce spillover effects to supply chains [9] and lead to the corresponding financial consequences of the enterprise [10]. Moreover, it will also affect the enterprise's trade credit [11], cash holdings [12], cost factors [13, 14],

investment efficiency [15], profitability [16], and so on. Hence, will customer risk affect the level of corporate financing constraints? Clarifying this problem will assist us in appreciating the enterprise financing constraints and preventing and resolving the risks brought by supply chain cooperation.

Accordingly, this paper utilizes fuzzy mathematics to evaluate customer risk and adopts the data of listed companies from 2007 to 2019 to investigate the influence of customer risk on corporate financing constraints. Firstly, this paper evaluates customer risks on the basis of fuzzy mathematics. Secondly, it establishes an investment-cash-flow sensitivity model to examine whether the company has external financing constraints. Then, it uses 1089 groups of suppliers-customers to steadily match the company's annual data. The influence of customer risk on the financing constraints of suppliers is empirically inspected, and the heterogeneity of the relationship between the two under different monetary policy tightness and different industry competition levels is conducted.

Compared with the existing study, the incremental contributions of this paper lie with the following: firstly, it enriches the study of supply chain risk spillovers. Although the existing literature gives priority to the "predatory effect" and "support effect" among enterprises in the industry, it fails to elaborate the risk transmission mechanism between supply chain enterprises. In this paper, it deepens the influence of financial information and nonfinancial information transmitted vertically on the third-party investors and creditors based on the supply chain. Secondly, it enriches the economic consequences of customer information. Most of the existing literature explores the economic consequences of customer information as per customer concentration, customer surplus, and supplier-customer relationships, and it seldom considers the extent of the economic consequences of the customer's business crisis and financial crisis. Thirdly, it enriches the study of the influencing factors of financing constraints and further investigates the level of corporate financing constraints as per supply chain risk spillovers. The credit decision-making behavior of investors and creditors is more easily affected by the financial status of corporate customers. In this paper, it dynamically examines the influence of customer risk on the level of corporate financing constraints. Hence, the conclusion is more convincing.

2. Theoretical Basis and Research Hypotheses

2.1. Customer Risk Evaluation Based on Fuzzy Mathematics. The customer risk is a significant factor that influences corporate financial decisions. The influencing factors of customer risk fall into customer external macro environment, customer financial status, customer operational risk, and customer relationship risk. There are various secondary evaluation indicators in each primary evaluation index. The construction of the customer risk evaluation index system is indicated by Table 1.

In this paper, it adopts the analytic hierarchy process (AHP) to calculate the weight of the primary indexes in the customer risk evaluation system and the secondary index weight under each primary index. In the AHP, the first step is to build a hierarchical structure model, which shall contain the objective layer, the criterion layer, and the scheme layer. In the second step, it establishes a pairwise comparison matrix, indicating the comparison of the relative importance of all factors in this layer to a certain factor in the previous layer. As for the element a_{ij} of the pairwise comparison matrix, it stands for the comparison result of the i^{th} factor relative to the j^{th} factor. In the third step, it needs to solve the feature vectors of the judgment matrix. In this paper, it adopts the square root method to calculate the approximate value of the matrix feature vectors. Firstly, the n^{th} root of the product of the elements of each row of the judgment matrix A is calculated. The formula can be seen as follows:

$$M_i = \sqrt[n]{\prod_{j=1}^{n} a_{ij.}} \tag{1}$$

Then, it is normalized.

$$W_i = \frac{M_i}{\sum_{i=1}^{n} M_i}. \tag{2}$$

It is concluded by calculating the largest eigenvalue of the judgment matrix.

$$\lambda = \sum_{i=1}^{n} \frac{(Aw)_i}{nw_i}. \tag{3}$$

In the fourth step, the consistency of the judgment matrix is checked. CI is an indicator that measures the judgment matrix's deviation consistency, and the calculation formula is as follows:

$$CI = \frac{\lambda - n}{n - 1}. \tag{4}$$

The larger the CI, the worse the consistency of the judgment matrix. If it is, then the judgment matrix has complete conformity. CI = 0, while CR represents consistency ratio, and its formula is as follows:

$$CR = \frac{CI}{RI}, \tag{5}$$

where RI is the average random consistency index. If $CR < 0.1$, it means that it passes the consistency test.

Then, calculate the comprehensive weight of each secondary indicator and sort each indicator. The results are as shown in Table 2.

Generally, the evaluation level of customer risk falls into five levels, namely, low-risk, relatively low-risk, medium-risk, relatively high-risk, and high-risk, and their risks show an increasing trend. For low-risk customers, their business environment is more stable. They have larger assets with better profitability and debt repayment capacity. They also have less operational risk. For medium-risk customers, there

Table 1: Customer risk evaluation index system.

Objective layer	Primary evaluation index	Secondary evaluation index	Secondary evaluation index description
Customer risk	Customer external macro environment	Economic policy uncertainty	The degree of economic policy uncertainty in the province where the customer is located
		Industry competition environment	The competition level in the customer's industry
		Regional legal environment	The legal environment of the province where the customer is located
	Customer's financial status	Profitability	Customer business profitability, measured by earnings per share
		Solvency	The solvency of the customer enterprise, measured by the asset-liability ratio
		Operational capability	The operating capacity of the customer enterprise, measured by the total asset turnover ratio
		Development capacity	The development capacity of the customer's enterprise, measured by the revenue growth rate
	Customer's operational risk	Senior executives	Working age of senior executives
		The situation that the company is regulated	The situation that the company is punished or inquired
		Company's internal control	The effectiveness of the company's internal control measured by the internal control index
		Operational stability	Earnings volatility, measured by ROA volatility
	Customer relationship risk	Relationship investment	Whether there are relationship investments between customers
		Relational risk	Dependence on customers, measured by customer concentration

Table 2: Customer risk evaluation index and its weights.

Objective layer	Primary evaluation index	Weight	Secondary evaluation index	Weight	Comprehensive weight	Sort
Customer risk	Customer external macro environment	0.1858	Economic policy uncertainty	0.2809	0.0522	11
			Industry competition environment	0.4140	0.0769	8
			Regional legal environment	0.3051	0.0567	10
	Customer's financial status	0.3759	Profitability	0.2929	0.1101	1
			Solvency	0.2609	0.0981	3
			Operational capability	0.2583	0.0971	4
			Development capacity	0.1879	0.0706	9
	Customer's operational risk	0.2652	Senior executives	0.1208	0.0320	13
			The situation that the company is regulated	0.1726	0.0458	12
			Company's internal control	0.3241	0.0860	6
			Operational stability	0.3825	0.1014	2
	Customer relationship risk	0.1731	Relationship investment	0.4568	0.0791	7
			Relational risk	0.5432	0.0940	5

are certain risks in their business environment, financial status, and operation. However, they also have assets available for collateral, which can be timely settled after collection in previous transactions. For high-risk customers, their business environment is uncertain. They have a poor financial situation and credit level, and they are recognized as bad enterprises in the industry. In this paper, it utilizes the method of fuzzy comprehensive evaluation to evaluate customer risk.

In the fuzzy comprehensive evaluation, in the first step, it needs to determine the customer risk evaluation index set U.

$$U = \{u_1, u_2, \ldots, u_m\}. \tag{6}$$

In the second step, it requires to confirm customer risk evaluation level set V.

$$V = \{v_1, v_2, \ldots, v_n\}. \tag{7}$$

In the third step, it needs to establish a fuzzy relationship matrix R.

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}, \tag{8}$$

Table 3: Customer risk evaluation sets.

Objective layer	Criterion layer Primary evaluation index	Index layer Secondary evaluation index	Evaluation sets 1	3	5	7	9
Customer risk	Customer external macro environment	Economic policy uncertainty	1	2	5	2	0
		Industry competition environment	1	3	4	0	2
		Regional legal environment	1	3	4	1	1
	Customer's financial status	Profitability	0	2	3	4	1
		Solvency	0	1	5	1	3
		Operational capability	1	1	4	3	1
		Development capacity	3	2	4	1	0
		Senior executives	3	2	3	1	1
	Customer's operational risk	The situation that the company is regulated	0	1	7	2	0
		Company's internal control	0	2	6	2	0
		Operational stability	1	1	3	3	2
	Customer relationship risk	Relationship investment	3	1	4	2	0
		Relational risk	1	2	5	1	1

where r_{ij} indicates the membership degree of customer risk to the fuzzy subset of v_i level in terms of index u_i.

In the fourth step, it requires determining the fuzzy weight vector $(a_1, a_2, \ldots, a_m)$for customer risk. Here, the weights obtained above using the AHP method are imported.

In the fifth step, it is to obtain the final fuzzy comprehensive evaluation model.

$$B = A \cdot R = (a_1, a_2, \ldots, a_m) \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} = (b_1, b_2, \ldots, b_n). \quad (9)$$

Among them, b_i indicates the membership of the customer risk to the fuzzy subset of the v_j level as a whole.

To make the customer risk evaluation results more intuitive, in this paper, it establishes a set C to represent the evaluation results of the judges on the risk degree of each index factor that influences the customer risk.

$$C = \{\text{Low risk, lower risk, me di um risk, higher risk, high risk}\} = \{1, 3, 5, 7, 9\}. \quad (10)$$

In this paper, it adopts the expert scoring method to determine the evaluation set for customer risk. The sorting results are indicated by Table 3.

As indicated by Table 3, it shows that the fuzzy comprehensive evaluation is conducted for each index and different risk degrees. By constructing a 13×5 weight judgment matrix R, the membership degrees of five risk level sets are finally calculated. $B = (0.138, 0.134, 0.3, 0.256, 0.171)$, as indicated by Figure 1.

As illustrated in Figure 1, it intuitively proves that "medium risk" has the highest membership degree. Meanwhile, we can get $V = B \cdot C^T = 5.377$ by calculating the evaluation score. Since 5.377 is between 5 and 7 and close to 5, we can finally judge that the evaluation result of customer risk is "medium risk." It means, considering the factors of external macro environment, the enterprise's financial situation, operating conditions, external relations, and others, the customer company generally has certain operational risks. Its financial status and reputation level cannot be fully guaranteed, however, the probability of risk is not high. Therefore, there is some uncertainty in the expected income.

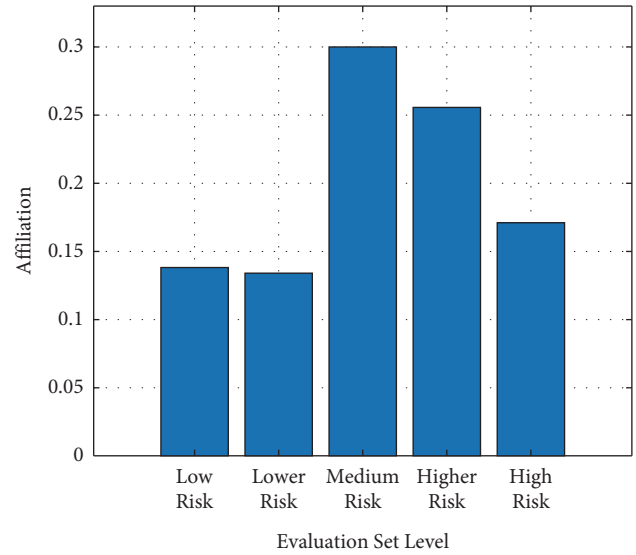


Figure 1: Judgment of the result of risk level membership degree.

2.2. The Influence of Customer Risk on Corporate Financing Constraints. Pursuant to the theory of information asymmetry and principal-agent theory, in the capital market and listed companies have always seen the phenomenon of the low quality of information disclosure, and financial institutions like banks are in a disadvantageous position in the credit process. To reduce credit risk, various financial institutions, during the process of making credit decisions, will raise loan interest rates and restrict the scope of use of loan funds, thereby substantially increasing corporate financing costs and financing constraints [17, 18].

Whether the information publicly disclosed by downstream customers can affect the credit decisions of creditors and investors of suppliers depends first on whether the

public information is valuable to investors to evaluate corporate credit risks. Meanwhile, investors and creditors have strong information collection capabilities. Besides collecting public information, they can also obtain private customer information associated with credit risk [19]. The negative spillover influence of customer risk on supplier companies affects the credit risk of supplier companies through two aspects.

Firstly, the customer risk degree will affect their ability to perform contracts and increase the commercial credit period and bad-debt losses on accounts receivable. When the customer's operating performance declines, its ability and willingness to observe the contract decrease [20, 21]. It, in turn, affects the capital turnover efficiency of suppliers, leading to a cash flow crisis and raising credit risk. Meanwhile, customer risk will make its demand for products from upstream suppliers substantially fall off, and the company cannot transform customer resources in the short term, indicating that the company's bargaining power will decline. At this time, customers require suppliers to provide more business credit concessions terms, which further impairs the cash flow of upstream companies, so that it is detrimental to creditors and investors in evaluating the company's capacity to repay loans [22]. Furthermore, enterprises cannot open up new marketing channels. To maintain a stable supply chain relationship and sustain the supply chain's coordinated operation, the supplier enterprise decreases the risk of future sales revenue and guards against weakening the enterprise's capacity to create cash flow in the future. It provides more business credit based on risk sharing motivation to help customers overcome difficulties, which will also cause short-term cash flow crisis of enterprises and affect financing capacity [23].

Secondly, customer risk will reduce the collateral value of proprietary assets of upstream supplier relationships, so that the investors' evaluation of their credit risk can be affected. In case that a customer encounters a financial crisis, its financial risk is essentially shared by all stakeholders, and it is hard for the supplier to withdraw from the supply chain relationship. This additional risk it takes will evidently reduce the value of relationship assets [24]. Particularly with the cooperation with key customers, supplier enterprises have established a close competitive and cooperative game relationship with them and invested more proprietary assets. Once their contractual relationship is terminated, caused by customer risk in the supply chain, the proprietary assets of the relationship will be considerably depreciated, meanwhile facing high conversion costs. This makes it impossible for companies to improve their credit status and weakens credit risks by mortgaging their proprietary assets. In view of the above analysis, in this paper, it proposes the following assumptions:

H1. pursuant to the supply chain risk spillover effect, the higher the customer risk, the greater the level of corporate financing constraints.

2.3. The Regulatory Role of Monetary Policy. Monetary policy transmits market signals and affects banks' market expectations and credit decisions. It influences the financing constraints of enterprises through interest rates and credit [25]. According to the monetary policy transmission theory, from the formulation to the influence on the investment and financing of enterprises and other economic activities, the monetary policy is mainly realized through the two channels of money and credit [26]. In case it has a loose monetary policy, enterprises can increase the effective demand of the product market by stimulating the aggregate demand, thus enhancing the profitability of the enterprise [27]. The weakening of profitability decreases the business operation risk of enterprises because of the negative amplification effect caused by customer risk spillover. Meanwhile, under a loose monetary policy, it can increase the total supply of bank credit and enhance the enthusiasm of bank loans. Even though the enterprise's business conditions are affected by customer risk contagion, the conditions of bank loans will be relaxed, which, in turn, can improve the external financing environment of enterprises and increase the financing ability [28]. In view of the above analysis, in this paper, it proposes the following assumptions:

H2. loose monetary policy can alleviate the deterioration of customer risk on corporate financing constraints

2.4. The Regulating Effect of Industry Competitiveness. In the capital market, the industry competitions will have a differentiated effect on the main business through the "predatory effect" opportunistic behavior. More precisely, when the enterprise industry is highly competitive, the "predatory behavior" of other competitors in the industry will create operational uncertainty risks [29]. The enterprise customer risk is relatively high. Based on the risk contagion of the supply chain, the competitive position of supplier companies is substantially decreased. At this time, provided that the company industry is highly competitive, the situation it faces is even more severe, and the "predatory behavior" of other companies in the industry will have greater impact [30, 31]. Meanwhile, highly competitive industries will face the disadvantages of small market share, slow growth, and fewer investment opportunities. Financial institutions like investors and creditors have higher requirements when making credit decisions, which notably increase the level of financing constraints for suppliers [32]. In view of the above analysis, in this paper, it proposes the following assumptions:

H3. the higher the level of competition in the industry, the worse the impact of customer risk on the financing constraints of enterprises.

3. Research Design

3.1. Data and Sample. In this paper, it takes the panel data of Shanghai and Shenzhen A-share listed companies from 2007 to 2019 as the initial sample and process the obtained samples as follows: (1) it eliminates the listed companies in the ST and financial industries. (2) Delete the companies

whose top five customers fail to disclose their sales proportion. (3) The samples with missing study data are removed. (4) To reduce the influence of extreme values on the regression results, in this paper, the authors perform a winsorization of 1% up and down for all continuous variables. In this paper, the data come from companies whose suppliers and customers are both listed companies, which have completed the annual matching data of suppliers and customers. Customer-related data were collected manually, and other data were derived from the CSMAR database. After the above screening, it ended up acquiring1089 groups of supplier-customer matching data.

3.2. Definition of Variables and Model Setting

3.2.1. Explained Variables. Financing constraints (invest): currently, there is no unified conclusion about the measurement of corporate financing constraints. Existing research measures financing constraints from three categories. Use enterprise characteristics to establish KZ, SA, WWS, and other indices: the cash holdings-cash flow sensitivity model; investment-cash flow sensitivity model. Fazzari and Peters [33] argued that the transaction cost caused by the problem of information asymmetry led to an increase in the external financing cost of enterprises, making the enterprises with poor financing ability more dependent on internal capital investment. Accordingly, they believed that the fluctuation of investment on cash flow notably influenced the cash flow. The higher the "sensitivity," the greater the sensitivity coefficient of cash flow (CashFlow), indicating that the company is more dependent on internal capital investment, which, in turn, indicates that the company faces a higher level of external financing constraints, and its value is equal to net cash flows from operating activities divided by total assets. Based on the above analysis, in this paper, it selects the investment-cash flow sensitivity model to measure the level of financing constraints, and the level of corporate capital expenditure (invest) as the explained variable, whose value is equal to the cash paid for the purchase and construction of fixed assets, intangible assets, and other long-term assets divided by total assets.

3.2.2. Explanatory Variables. Customer risk (risk): there are many models for measuring the level of enterprise risk. The existing study on its measurement indicators contains the following: *Z* index, earnings volatility, stock return volatility, asset-liability ratio, and leverage coefficient, among which earnings volatility is the most commonly used indicator to measure corporate risk. In this paper, it learns from the research of Liu Xing [34] and adopts the volatility of return on assets (ROA) to measure the level of corporate risk. The greater the volatility of this indicator, the higher the corporate risk. Learning from the study of Coles [35], it utilized the rolling year method to calculate the standard deviation of the stock returns (r_i) of the sample companies in each period with every 5 years as an observation period. The calculation process is as follows:

$$\text{risk} = \sqrt{\frac{1}{N-1}\sum_{n=1}^{N}\left(\text{ADJ\_ROA}_{in} - \frac{1}{N}\sum_{n=1}^{N}\text{ADJ\_ROA}_{in}\right)^2}, N = 5, \tag{11}$$

where $\text{ADJ\_ROA}_{in} = \text{ROA}_{in} - 1/X\sum_{K=1}^{X}\text{ROA}_{kn}$.

3.2.3. Control Variable. In this paper, the control variables fall into the supplier corporate governance level, operation level, and customer level. The variables contain corporate growth (growth), corporate age (age), return on assets (ROA), shareholding ratio of the largest shareholder (hold), change in short-term current liabilities (Std), management shareholding (Msh), customer age (KHAge), and return on customer assets (KHROA). They also contain monetary policy (MP) and industry competitiveness (HHI)-regulated variables, controlling for the year and the industry. The calculation method of the variables is indicated by Table 4.

3.2.4. Model Specification. It adopts the basic model (12) of the investment-cash flow sensitivity model to measure the level of financing constraints.

$$\left(\frac{I}{K}\right)_{it} = f\left(\frac{X}{K}\right)_{it} + g\left(\frac{CF}{K}\right)_{it} + \varepsilon_{it}. \tag{12}$$

Among them, K indicates capital stock, I indicates investment expenditure, CF refers to cash flow from operating activities, *X* stands for a variable that theoretically determines the investment demand of an enterprise, and g refers to the sensitivity of enterprise investment to fluctuations in internal cash flow. Drawing on the study of Wan Liangyong [36], it adopts the investment-cash flow sensitivity model to verify the existence of external financing constraints of enterprises, in which sensitivity coefficient is used as the factor of financing constraints. Measure the indicators and build the model (13) as follows:

$$\begin{aligned}\text{Invest}_{it} = {} & \alpha_0 + \alpha_1\text{CashFlow}_{it} + \alpha_2\text{Growth}_{it} + \alpha_3\text{Age}_{it} \\ & + \alpha_4\text{ROA}_{it} + \alpha_5\text{Hold}_{it} + \alpha_6\text{Std}_{it} \\ & + \alpha_7\text{Msh}_{it} + \sum\text{Industry} + \sum\text{year} + \varepsilon_{it}.\end{aligned} \tag{13}$$

To study the relationship between customer risk and financing constraints, namely, hypothesis H1, we add customer risk factor to build model (13) and add customer characteristic factors like customer age (KHAge) and return on customer assets (KHROA) to build a model (14).

Table 4: Summary table of variable definition 1.

Variable name	Symbols	Calculation method
Total assets level	Invest	Total cashes/assets paid for construction of fixed assets, intangible assets, and other long-term assets
Customer risk	Risk	Volatility of return on assets
Operating cash flows	CashFlow	Operating activities net cash flow/total assets
Growth	Growth	(Operating income at the end of the period–operating income at the beginning of the period)/operating income at the end of the period
Age of establishment	Age	The natural logarithm of the age of establishment + 1
Return on assets	ROA	Net income/total assets
Shareholding ratio of the largest shareholder	Hold	Number of shares held by the largest shareholder/total number of shares
Changes in short-term current liabilities	Std	(Short-term current liabilities in year t–short-term current liabilities in year t-1)/total assets in year t
Management shareholding	Msh	Number of shares held by management/total shares
Customer's age	KHAge	The natural logarithm of the customer's establishment year + 1
Return on customer assets	KHROA	Customers net income/total assets
Monetary policy	MP	M2 growth rate–GDP growth rate–CPI growth rate
Industrial competitiveness	HHI	Herfindahl–Hirschman index
Industry dummy	Industry	Sector
Year virtual variable	year	Year

$$\begin{aligned}\text{Invest}_{it} = {} & \alpha_0 + \alpha_1 \text{CashFlow}_{it} + \alpha_2 \text{CashFlow}_{it} * \text{risk}_{it} + \alpha_3 \text{risk}_{it} + \alpha_4 \text{Growth}_{it} + \alpha_5 \text{Age}_{it} \\ & + \alpha_6 \text{ROA}_{it} + \alpha_7 \text{Hold}_{it} + \alpha_8 \text{Std}_{it} + \alpha_9 \text{Msh}_{it} + \alpha_{10} \text{KHAge}_{it} + \alpha_{11} \text{KHROA}_{it} \\ & + \sum \text{Industry} + \sum \text{year} + \varepsilon_{it}.\end{aligned} \tag{14}$$

Considering the relationship among customer risk, monetary policy (industry competitiveness), and financing constraint, we verify the intermediary role of monetary policy (industry competitiveness) between customer risk and financing constraint, namely, hypothesis H2 and H3. On the basis of model (14), we construct model (15) using X variables, where X denotes monetary policy (MP) and Herfindal Hirschman index (HHI).

$$\begin{aligned}\text{Invest}_{it} = {} & \alpha_0 + \alpha_1 \text{CashFlow}_{it} + \alpha_2 \text{CashFlow}_{it} * \text{risk}_{it} + \alpha_3 \text{risk}_{it} + \alpha_4 \text{CashFlow}_{it} * \text{risk}_{it} * X \\ & + \alpha_5 X + \alpha_6 \text{Growth}_{it} + \alpha_7 \text{Age}_{it} + \alpha_8 \text{ROA}_{it} + \alpha_9 \text{Hold}_{it} + \alpha_{10} \text{Std}_{it} + \alpha_{11} \text{Msh}_{it} \\ & + \alpha_{12} \text{KHAge}_{it} + \alpha_{13} \text{KHROA}_{it} + \sum \text{Industry} + \sum \text{year} + \varepsilon_{it}.\end{aligned} \tag{15}$$

In formulas (12)–(15), α refers to the coefficient value, ε means the residual item, i represents different enterprises, t indicates the year, and the coefficient α_1 of CashFlow is the investment-cash flow sensitivity.

4. Empirical Analysis and Results

4.1. Descriptive Statistics. The full-sample descriptive statistics of each variable are indicated in Table 5. The maximum value of Invest is 0.2378, the mean value is 0.0564, and the minimum value is 0.0003, which indicates that there exist differences in the capital expenditures of listed enterprises. The mean value of operating cash flow (CashFlow) reaches 0.0399, which represents that annual operating cash flow accounts for 3.99% of total assets. The average value of customer risk reaches 0.0428, which indicates that 4.28% of business customers face high risk. It can be seen that there are, as a whole, fewer business customers with high risk.

4.2. Regression Results and Interpretation

4.2.1. Investment-Cash-Flow Sensitivity Model Regression Results. According to the result of column (1) in Table 6, it proves that the coefficient of capital expenditure and cash flow from operating activities is 0.0832, which is significantly positive at the 1% level, indicating that Shanghai and Shenzhen A-share listed companies generally have external financing constraints. The reason is that when corporate financing is blocked, to maintain corporate scale and enhance operating performance, companies generally choose to extract some funds from internal cash flow for investment. Considering the above analysis, the investment-cash flow sensitivity coefficient is notably positive.

Table 5: Descriptive statistics.

Variable	N	Min	Max	Mean	P50	Sd
Invest	1089	0.0003	0.2378	0.0564	0.0421	0.0500
Cash flow	1089	−0.1541	0.2137	0.0399	0.0385	0.0661
Risk	1089	0.0016	0.2502	0.0428	0.0183	0.0655
Growth	1089	−0.9327	2.1035	0.1066	0.0962	0.4388
Age	1089	1.0986	3.4340	2.6837	2.7726	0.4281
ROA	1089	−0.2848	0.1717	0.0384	0.0398	0.0598
Hold	1089	0.1047	0.7713	0.3629	0.3393	0.1532
Std	1089	−3.6152	0.3330	−0.0837	0.0260	0.5541
Msh	1089	≤0.001	0.6923	0.1415	0.0019	0.2110
KHAge	1089	1.6094	3.4340	2.7566	2.8332	0.3586
KHROA	1089	−0.0841	0.1783	0.0438	0.0382	0.0435

Table 6: Regression results of customer risk and corporate financing constraints.

	(1) Invest	(2) Invest
Cash flow	0.0832***	0.0690**
	(20.16)	(2.48)
Risk		0.0353
		(1.50)
Risk ∗ CashFlow		0.8925**
		(2.27)
Hold	0.0027	0.0203**
	(1.41)	(2.06)
Std	0.0036***	0.0096***
	(6.35)	(3.30)
Msh	0.0253***	0.0347***
	(16.59)	(4.75)
ROA	0.0417***	0.0520**
	(8.43)	(2.03)
Growth	0.0007*	−0.0020
	(1.85)	(−0.55)
Age	−0.0143***	−0.0077*
	(−16.64)	(−1.89)
KHROA		0.0945***
		(2.91)
KHAge		0.0056
		(1.24)
Constant	0.1005***	0.0982***
	(28.68)	(2.97)
Year	Yes	Yes
Industry	Yes	Yes
N	28391	1089
R^2	0.1508	0.2101

Note. ∗ ∗∗ , ∗ ∗ , and ∗ indicate that the test coefficients are significant at the 1%, 5%, and 10% levels, respectively, and in parentheses, it displays the value t.

4.2.2. The Impact of Customer Risk on Corporate Financing Constraints. In Column (2) of Table 6, it displays the regression result of the full-sample customer risk and corporate financing constraints. Among them, the coefficient before the interaction term between customer risk and operating cash flow represents the degree to which the investment-cash-flow sensitivity is affected by the enterprise customer risk. Provided that the coefficient of the interaction term is obviously positive, the higher the customer risk will aggravate the level of external financing constraints, the larger the coefficient, and the greater the impact. According to the result of column (2) in Table 6, it shows that the coefficient of the interaction term between customer risk and cash flow from operating activities is notably positive at the 5% level. The parameter estimate of cash flow from operating activities to capital expenditure reaches 0.0690, and at 5% is notably positive at the level. It shows that the higher the customer risk, the higher the level of enterprise financing constraints. Hence, hypothesis H1 has been verified.

4.2.3. The Influence of Monetary Policy and Industry Competitiveness on the Relationship between Customer Risk and Corporate Financing Constraints. Pursuant to the regression results in Table 7, it proves that, after the introduction of monetary policy and industry competitiveness, the coefficient of interaction among customer risk, monetary policy, and operating cash flow is −0.2433, which is significantly negative at the 10% level. It shows that compared with tight monetary policy, the level of financing constraints when monetary policy is loose is less affected by corporate customer risk. The coefficient of interaction item between customer risk, industry competitiveness, and capital expenditure reaches −6.6113, which is notably negative at the 10% level. It indicates that customer risk positively influences financing constraints compared to less competitive companies in the industry. The effect of the directional influence on the enterprises with a high degree of competition in the industry is more significant. Hence, the hypothesis H2 and H3 are supported.

4.3. Robustness Test

4.3.1. Replace the Explained Variable. Drawing on the study of Hong Jinming and Lin Runyu et al., the observation period is defined as from year t to year t+2 to redefine the level of customer risk taking and substitute it into the model (14). Upon inspection, the replaced customer risk makes the regression result consistent with the previous text. Hence, it is concluded to be robust. The regression results in column (1) of Table 8 are as follows:

4.3.2. Other Sensitivity Test. To prevent the omitted variable from causing a positive correlation between customer risk and the level of suppliers' external financing constraints, the model contains the size of the company (Size), the concurrent position of chairman and General Manager (Dual), the size of the Board of Directors (Board), and the ratio of tangible assets (PPE). Equal variables are used as control variables, and the regression results are indicated by column (2) in Table 8. Furthermore, it only considers that the manufacturing industry is with the industries other than the manufacturing industry being excluded, and the regression results are listed in Column (1) of Table 9. To prevent the influence of the privatization information of state-controlled enterprises, in this paper, it selects a sample of nonstate-owned enterprises to perform basic regression again. The empirical results are indicated by column (2) of Table 9. The

Table 7: Regression results of the adjustment effect of monetary policy and industry competitiveness.

	(1) Invest	(2) Invest
CashFlow	0.0899*** (2.74)	0.0736*** (2.61)
Risk	0.0286 (1.16)	0.0371 (1.57)
Risk ∗ CashFlow	1.4678* (1.70)	1.4242*** (2.75)
Risk ∗ CashFlow ∗ MP	−0.2433* (−1.69)	
MP	0.0008 (1.15)	
Risk ∗ CashFlow ∗ HHI		−6.6113* (−1.67)
HHI		0.0248 (1.25)
Hold	0.0235*** (2.60)	0.0204** (2.07)
Std	0.0394*** (4.38)	0.0094*** (3.10)
Msh	0.0314*** (4.74)	0.0351*** (4.79)
ROA	0.0536 (1.45)	0.0513* (1.94)
Growth	−0.0082* (−1.68)	−0.0021 (−0.57)
Age	−0.0094** (−2.34)	−0.0077* (−1.90)
KHROA	0.1127*** (3.29)	0.0989*** (3.00)
KHAge	0.0031 (0.70)	0.0058 (1.29)
Constant	0.0840*** (2.86)	0.0941*** (2.84)
Year	Yes	Yes
Industry	Yes	Yes
N	1089	1088
R^2	0.2203	0.2101

Note. ∗∗∗, ∗∗, and ∗ indicate that the test coefficients are significant at the 1%, 5%, and 10% levels, respectively, and in parentheses, it displays the value t.

above regression results are consistent with the previous text. Hence, it is concluded to be robust.

4.3.3. Alternate Models. The measurement of financing constraints in existing research is not uniform, and the use of SA index is more common. In this paper, it draws on the methods of HadLock and Pierce [37] to construct the SA index, SA = −0.737 ∗ Size + 0.043Size^2−0.04 ∗ Age, in which,

Table 8: Summary table of robustness test results 1.

	(1) Invest	(2) Invest
CashFlow	0.0734*** (2.77)	0.0649** (2.27)
Risk	0.0338 (1.58)	0.0296 (1.23)
risk ∗ CashFlow	0.9821** (2.49)	0.8667** (2.18)
Hold	0.0198** (2.01)	0.0172* (1.69)
Std	0.0097*** (3.32)	0.0107*** (3.59)
Msh	0.0351*** (4.80)	0.0301*** (3.71)
ROA	0.0481* (1.87)	0.0617** (2.25)
Growth	−0.0020 (−0.55)	−0.0035 (−0.94)
Age	−0.0078* (−1.92)	−0.0073* (−1.77)
Size		0.0003 (0.22)
Dual		0.0100*** (2.75)
Board		−0.0017 (−0.51)
PPE		−0.0020 (−0.21)
KHROA	0.0877*** (2.66)	0.0852** (2.56)
KHAge	0.0050 (1.11)	0.0073 (1.60)
Constant	0.0997*** (3.02)	0.0889** (2.05)
Year	Yes	Yes
Industry	Yes	Yes
N	1088	1062
R^2	0.2089	0.2146

Note. ∗∗∗, ∗∗, and ∗ indicate that the test coefficients are significant at the 1%, 5%, and 10% levels, respectively, and in parentheses, it displays the value t.

Size refers to the scale of the enterprise, Age means the year of establishment, and the absolute value of SA is used to measure the financing constraint (FC). The larger the absolute value of SA, the higher the level of corporate financing constraints. Moreover, given the high volatility of China's stock market, the stock return volatility index (CRT) can be used to measure customer risk. Meanwhile, control variables that may affect financing constraints are added to the model, and the model (16) is constructed as follows:

$$\begin{aligned} FC_{it} = {} & \alpha_0 + \alpha_1 CRT_{it} + \alpha_2 CashFlow_{it} + \alpha_3 Age_{it} + \alpha_4 Size_{it} + + \alpha_5 PPE_{it} + + \alpha_6 Lev_{it} \\ & + \alpha_7 Hold_{it} + \alpha_8 Msh_{it} + \alpha_9 ROA_{it} + \alpha_{10} Growth_{it} + \alpha_{11} Dual_{it} + \alpha_{12} KHROA_{it} \\ & + \alpha_{12} KHAge_{it} + \sum Industry + \sum year + \varepsilon_{it}. \end{aligned} \tag{16}$$

Table 9: Summary table of robustness test 2.

	(1) Invest	(2) Invest
CashFlow	0.0260	0.0620
	(0.69)	(1.40)
Risk	0.0119	0.0805**
	(0.42)	(2.49)
Risk ∗ CashFlow	1.5114***	0.9565*
	(3.06)	(1.90)
Hold	0.0142	0.0042
	(1.15)	(0.25)
Std	0.0099***	0.0092***
	(3.19)	(2.63)
Msh	0.0401***	0.0273**
	(4.80)	(2.49)
ROA	0.0252	−0.0070
	(0.66)	(−0.11)
Growth	−0.0081	0.0010
	(−1.51)	(0.19)
Age	−0.0081	−0.0070
	(−1.60)	(−1.18)
KHROA	0.0550	0.0435
	(1.38)	(0.84)
KHAge	0.0112**	−0.0023
	(2.04)	(−0.33)
Constant	0.0169	0.1110*
	(0.56)	(1.93)
Year	Yes	Yes
Industry	Yes	Yes
N	694	439
R^2	0.1644	0.2015

Note. ∗∗∗, ∗∗, and ∗ indicate that the test coefficients are significant at the 1%, 5%, and 10% levels, respectively, and in parentheses, it displays the value *t*.

The regression results are given in Table 10. The coefficient of CRT is notably positive, which means that the higher the customer risk, the greater the level of corporate financing constraints. It is in conformity with the previous text. Hence, it is concluded to be robust.

4.3.4. Endogeneity

(1) Two-stage least squares method: based on the previous test, it is proved that customer risk affects the level of enterprise's external financing constraints, and corporate financing capacity will also restrict the conduct of customer risk activities. Therefore, it can adopt the instrumental variable method to alleviate the endogeneity problem caused by reverse causality. In this paper, it utilizes the two-stage least squares method, draws on the study of Miao Miao and Liao Shiyu (2020) [38], and introduces a one-phase lag (L. risk) and managerial overconfidence (OC) of corporate customers as the instrumental variables of customer risk to further test model (14). Among them, the overconfidence of managers affects the level of risk taking, and the indicator of overconfidence of customer managers fails to directly affect the level of suppliers' external financing constraints to meet the exogenous requirements. It adopts the personal

Table 10: Summary table of robustness test results 3.

	(1) FC
CRT	0.0103**
	(2.21)
CashFlow	0.0303
	(1.30)
Age	0.0811***
	(19.69)
Size	0.0023
	(1.57)
PPE	0.0004
	(0.04)
Lev	0.0044
	(0.48)
Hold	−0.0371***
	(−3.70)
Msh	−0.0183**
	(−2.32)
ROA	0.0495*
	(1.66)
Growth	0.0039
	(1.22)
Dual	−0.0203***
	(−5.64)
KHROA	0.0551
	(1.64)
KHAge	0.0042
	(0.95)
Constant	1.0518***
	(18.96)
Year	Yes
Industry	Yes
N	1032
R^2	0.5525

Note. ∗∗∗, ∗∗, and ∗ indicate that the test coefficients are significant at the 1%, 5%, and 10% levels, respectively, and in parentheses, it displays the value *t*.

Table 11: Two-stage least squares regression results 1.

	(1) Risk	(2) Risk ∗ CashFlow	(3) Invest
L.risk	0.8147***	−0.0002	
	(13.90)	(-0.05)	
L.risk ∗ CashFlow	−1.1793	0.6611***	
	(−1.55)	(3.35)	
Risk			0.0847
			(1.42)
Risk ∗ CashFlow			1.6162**
			(2.52)
Control	Yes	Yes	Yes
Constant	0.0404**	0.0016	0.1400***
	(2.21)	(1.17)	(4.35)
Year	Yes	Yes	Yes
Industry	Yes	Yes	Yes
N	485	485	485
R^2	0.6173	0.6185	0.2351

Note. ∗∗∗, ∗∗, and ∗ indicate that the test coefficients are significant at the 1%, 5%, and 10% levels, respectively, and in parentheses, it displays the value *t*.

Table 12: Two-stage least squares regression results 2.

	(1) Risk	(2) Risk ∗ CashFlow	(3) Invest
OC	0.0502**	−0.0036*	
	(2.43)	(−1.82)	
OC ∗ CashFlow	−0.0781	0.0591***	
	(−1.63)	(6.40)	
Risk			0.1053
			(0.33)
Risk ∗ CashFlow			2.4237***
			(3.39)
Control	Yes	Yes	Yes
Constant	0.0973**	0.0001	0.0344
	(2.12)	(0.02)	(0.74)
Year	Yes	Yes	Yes
Industry	Yes	Yes	Yes
N	855	855	855
R^2	0.0732	0.2716	0.1756

Note. ∗∗∗, ∗∗, and ∗ indicate that the test coefficients are significant at the 1%, 5%, and 10% levels, respectively, and in parentheses, it displays the value t.

characteristics of general managers to construct a measure of managerial overconfidence. See Tables 11 and 12 for regression result. As shown in column (1) and column (2), in the regression results of the first stage, the significance levels of L. risk and OC of the first-stage regression and L. risk ∗ CashFlow and OC ∗ CashFlow of the second-stage exceed 5%, and the F statistic values of the two regressions in the first stage exceed 10. It can indicate that there is no weak instrumental variable problem. As indicated by column (3), the variable risk ∗ CashFlow has a positive and significant sign, which means that the conclusion is still robust after using instrumental variables to deal with possible endogeneity problems. It is still concluded to be robust.

(2) Heckman two-stage. In this study, the sample corresponding to customer risk may not be randomly selected; however, it is rationally selected. There may be endogeneity problems caused by sample bias and self-selection between customer risk and corporate financing constraints. To this end, in this paper, it adopts Heckman two-stage regression to test again.

In the first stage, the customer risk is transformed into a dummy variable with the average value as the demarcation point. The value above the average value of risk taking is 1, and the value below the average value is 0. Based on the study of Zhang Xin et al. [39], the size of the enterprise, the asset-liability ratio Lev, the return on assets ROA, the age of establishment Age, and the growth of the enterprise are selected as the main influencing factors of the risk taking level, and the first-stage Probit model (17) is constructed as follows:

$$\begin{aligned} \text{risk}_{it} = {} & \alpha_0 + \alpha_1 \text{Size}_{it} + \alpha_2 \text{Lev}_{it} + \alpha_3 \text{ROA}_{it} + \alpha_4 \text{Age}_{it} \\ & + \alpha_5 \text{Growth}_{it} + \sum \text{In du stry} + \sum \text{year} + \varepsilon_{it}. \end{aligned} \tag{17}$$

It obtains the inverse Mills ratio (IMR) by performing the Probit regression on model (17), substituting it into model (14) as a control variable, and reregressing to test the previous hypothesis. Pursuant to the regression result, it proves that the IMR coefficient of the second-stage regression is not notable and the coefficient of risk ∗ CashFlow is still significantly positive, indicating that the problem of sample self-selection in this paper is not serious, and the conclusion is robust.

5. Conclusions and Suggestions

On the basis of the evaluation of customer risk by fuzzy mathematics, in this paper, it adopts the Shanghai and Shenzhen A-share listed companies from 2007 to 2019 as a study sample to empirically test the impact of customer risk on the level of corporate financing constraints. It also investigates the heterogeneous effects of monetary policy and industry competitiveness on the relationship between them. Pursuant to the research result, it shows that customer risk is currently at a moderate level. Based on the spillover effect of supply chain risk, customer risk will increase the level of external financing constraints for suppliers. It is because customer risks will reduce their ability to perform contracts and the mortgage value of proprietary assets of the relationship. Buyers and sellers in the same supply chain are often a community of interests that "prosper, lose, and are mutually dependent," and it is hard for suppliers to break away from the supply chain. The information publicly disclosed by downstream customers can affect the credit decisions of creditors and investors of suppliers. Investors and creditors have strong information collection capabilities. Besides collecting public information, they can also obtain private information about customers associated with credit risk. A loose monetary policy will alleviate the problem of the external financing constraints of suppliers caused by customer risks. It is because macro monetary policy restricts the price of bank credit assets, thereby causing systemic changes in the credit market. There are relatively few financing channels for listed companies in China. Most companies conduct financing activities through the banks. When the monetary policy is loose, companies are less affected by customer risks. Hence, it is easier to obtain bank credit resources. In the event that the industry is highly competitive, it will enlarge the negative spillover effect of customer risks on supplier companies and reduce corporate financing capabilities. It is because supply chain risk spillovers affect the business performance of suppliers. Provided that the industry competition is fierce at this time, companies will be affected by the same industry. The predatory behavior of other companies is more serious, and creditors and investors will increase financing conditions as per this information.

To further activate the information transmission and financing market in the supply chain, this paper, pursuant to the above conclusions, puts forward the following suggestions: firstly, when making financing decisions, enterprises should focus on the macroeconomic policies, internal corporate governance structure, and the impact of public information and private information between affiliated enterprises, especially upstream and downstream enterprises, because suppliers and customers restrict an

enterprise's production and business activities. Secondly, in case it witnesses the tightening of monetary policy and fierce industry competitiveness, enterprises are less dependent on customer resources. Provided that the customers encounter risks at this time, the negative spillover effect of customer risks will be magnified, which will affect the credit decisions of investors and creditors. Finally, enterprises should fully exert their supervisory role and governance mechanism to curb the adverse effects caused by the financial and nonfinancial information of downstream customers. Meanwhile, the government shall unceasingly improve the capital market credit system, enhance the transparency of market information, and ease corporate financing constraints, thereby creating a good business environment.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

Qun Bao was the creator and director of this research project. She wrote the first draft of the paper and completed theoretical analysis. Ju-Yin Wang participated in data collation and analysis of empirical results. Rui Xie was involved in the empirical design and data analysis, and Zheng-Qun Cai participated in the writing and revision of the thesis. All authors read and agree to the final text.

Acknowledgments

This work was funded by the National Social Science Foundation of China (NSFC) project "Research on Enterprise Supply Chain Risk Immunity Mechanism under Major Public Emergencies" (20BGL095).

References

[1] J. Xie and Z. Z. Huang, "The impact of macro monetary policy and regional financial development on enterprise investment and financing constraints," *Financial Research*, vol. 11, pp. 64–78, 2014.

[2] Z. Ma, L. F. Wang, and R. M. Cui, "Subsidiary business, debt commitment and financing constraints of listed companies," *Economic management*, vol. 42, pp. 174–193, 2020.

[3] X. H. Shen, "Bank risk identification, government financial subsidies and enterprise debt financing costs -- an Empirical Test Based on the company data of Shanghai and Shenzhen stock markets from 2007 to 2012," *Finance and Trade Economy*, vol. 9, pp. 62–71, 2014.

[4] L. J. Yao, M. Luo, and L. Xia, "Corporate governance and bank loan financing," *Accounting Research*, vol. 8, pp. 55–61, 2010.

[5] Y. D. Xu, T. W. Li, and J. M. Hong, "Institutional environment, information disclosure quality and bank debt financing constraints--Empirical Evidence from Shenzhen A-share listed companies," *Finance and Trade Economics*, vol. 5, pp. 51–57, 2011.

[6] J. Itzkowitz, "Customers and cash: how relationships affect suppliers' cash holdings," *Journal of Corporate Finance*, vol. 19, pp. 159–180, 2013.

[7] X. Y. Wang and K. J. Gao, "Customer relationship and enterprise cost stickiness: rip off or cooperation," *Nankai Business Review*, vol. 20, pp. 132–142, 2017.

[8] P. N. Patatoukas, "Customer-base concentration: implications for firm performance and capital markets," *The Accounting Review*, vol. 87, no. 2, pp. 363–392, 2012.

[9] Q. Bao and Y. N. Mao, "Research on supply chain spillover effect of customer risk," *Journal of Anhui Agricultural University*, vol. 29, pp. 64–71, 2020.

[10] X. Peng and X. Y. Wang, "Does the risk of customer share price collapse have an infectious effect on suppliers?" *Financial Research*, vol. 44, pp. 141–153, 2018.

[11] Z. F. Lu and D. M. Yang, "Commercial credit: alternative financing or buyer's market?" *Management World*, vol. 4, pp. 6–14, 2011.

[12] X. Y. Zhao and Q. Bao, "Whether the relationship between suppliers and customers affects the cash holding level of Enterprises -- an Empirical Analysis Based on the panel data of listed manufacturing companies," *Journal of Jiangxi University of Finance and Economics*, vol. 5, pp. 41–48, 2014.

[13] Y. Wang, "Will the level of customer debt affect the supplier's credit financing cost—— empirical research based on Manufacturing Listed Companies," *Systems Engineering*, vol. 37, pp. 106–128, 2019.

[14] J. B. Kim, B. Y. Song, and Y. Zhang, "Earnings performance of major customers and bank loan contracting with suppliers," *Journal of Banking & Finance*, vol. 59, pp. 384–398, 2015.

[15] Y. S. Chen, "Social networks and enterprise efficiency: evidence based on the location of structural holes," *Accounting Research*, vol. 1, pp. 48–55, 2015.

[16] Z. W. Pei and D. F. Chen, "Mechanism and policy suggestions on the influence of commercial credit supply motivation on enterprise profitability," *Modern Management Science*, vol. 3, pp. 21–23, 2016.

[17] S. C. Myers and N. S. Majluf, "Corporate financing and investment decisions when firms have information that investors do not have," *Journal of Financial Economics*, vol. 13, no. 2, pp. 187–221, 1984.

[18] X. Liu and X. C. Chen, "CSRC punishment, analyst tracking and corporate bank debt financing -- Empirical Evidence from violations of information disclosure," *Accounting Research*, vol. 1, pp. 60–67, 2018.

[19] L. Cen, E. L. Maydew, L. Zhang, and L. Zuo, "Customer-supplier relationships and corporate tax avoidance," *Journal of Financial Economics*, vol. 123, no. 2, pp. 377–394, 2017.

[20] R. A. Jarrow and F. Yu, "Counterparty risk and the pricing of defaultable securities," *The Journal of Finance*, vol. 56, no. 5, pp. 1765–1799, 2001.

[21] R. Files and U. G. Gurun, "Lenders' response to peer and customer restatements," *Contemporary Accounting Research*, vol. 35, no. 1, pp. 464–493, 2018.

[22] Q. Bao, B. Yu, and M. Q. Sheng, "Financial flexibility, supply chain relationship and enterprise value——an Empirical Test Based on the background of the new normal," *Modern finance and Economics (Journal of Tianjin University of Finance and Economics)*, vol. 37, pp. 90–102, 2017.

[23] D. P. Han and P. Zang, "Monetary policy, financing constraints and investment efficiency-Empirical Evidence from

Chinese Private Listed Companies," *Nankai Business Review*, vol. 18, pp. 121–129+150, 2015.

[24] S. Banerjee, S. Dasgupta, and Y. Kim, "Buyer-supplier relationships and the stakeholder theory of capital structure," *The Journal of Finance*, vol. 63, no. 5, pp. 2507–2552, 2008.

[25] C. F. Baum, M. Caglayan, and N. Ozkan, "The second moments matter: the impact of macroeconomic uncertainty on the allocation of loanable funds," *Economics Letters*, vol. 102, no. 2, pp. 87–89, 2009.

[26] G. L. Zhang, Y. Xu, Y. Q. Han, and N. Y. Sun, "Economic policy uncertainty and enterprise financing constraints," *Investment Research*, vol. 37, pp. 144–159, 2018.

[27] K. Zhong, X. K. Cheng, and W. H. Zhang, "Monetary policy, information transparency and enterprise credit term structure," *Finance and Trade Economy*, vol. 3, pp. 60–77, 2016.

[28] Y. F. Gong and L. Guo, "Analyst tracking, ownership nature and financing constraints -- a study based on different property rights subjects," *Economics and Management*, vol. 34, pp. 129–137, 2012.

[29] Z. Yao, Y. Zheng, S. Q. Lu, and Y. F. Zhang, "Economic policy uncertainty and enterprise financing constraints——a study based on enterprise heterogeneity and financing channels," *Industrial Technology Economy*, vol. 39, pp. 116–125, 2020.

[30] Z. L. Chen, "Customer concentration, industry competition and commercial credit," *Accounting Research*, vol. 11, pp. 79–85, 2017.

[31] L. L. Sun and Z. Q. Wang, "Dynamic adjustment of bargaining power, monetary policy and commercial credit policy," *Journal of Jiangxi University of Finance and Economics*, vol. 5, pp. 43–53, 2016.

[32] F. Teng, X. Xia, and Y. Xin, "Customer relationship and operational performance of private placement," *Nankai Business Review*, vol. 23, pp. 212–224, 2020.

[33] S. M. Fazzari, R. G. Hubbard, B. C. Petersen, A. S. Blinder, and J. M. Poterba, "Financing constraints and corporate investment," *Brookings Papers on Economic Activity*, vol. 1988, no. 1, pp. 141–206, Article ID 2534426, 1988.

[34] L. Hang, J. Lei, and L. Juan, "House prices, collateral value and corporate risk-taking," *Journal of Financial Research*, vol. 3, pp. 107–123, 2016.

[35] J. Coles, N. Daniel, and L. Naveen, "Managerial incentives and risk-taking," *Journal of Financial Economics*, vol. 79, no. 2, pp. 431–468, 2006.

[36] L. Y. Wan, M. Q. Liao, and J. Hu, "Combination of industry and finance and corporate financing constraints——an Empirical Study Based on listed companies' participation in banks," *Nankai Business Review*, vol. 18, pp. 64–72, 2015.

[37] C. J. Hadlock and J. R. Pierce, "New evidence on measuring financial constraints: moving beyond the KZ index," *Review of Financial Studies*, vol. 23, no. 5, pp. 1909–1940, 2010.

[38] M. Miao and S. Y. Liao, "Enterprise risk bearing, ownership and litigation settlement," *China's Economic Problems*, vol. 2, pp. 91–105, 2020.

[39] X. Zhang, G. T. Qiao, and Y. R. Wang, "Does corporate risk-taking improve the level of audit fees?" *Journal of Nanjing Audit University*, vol. 17, pp. 34–44, 2020.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 2988639, 10 pages
https://doi.org/10.1155/2022/2988639

Research Article

Research and Implementation of Text Generation Based on Text Augmentation and Knowledge Understanding

Lei Liu,[1,2] **Yeguo Sun**,[3] **Yihong Liu**,[1] **Rachel Edita O. Roxas**,[2] **and Rodolfo C. Raga**[2]

[1]*School of Computer Science, Huainan Normal University, Huainan, China*
[2]*School of Computing and Information Technologies, National University, Manila, Philippines*
[3]*School of Finance and Mathematics, Huainan Normal University, Huainan, China*

Correspondence should be addressed to Yeguo Sun; yeguosun@126.com

Received 30 May 2022; Revised 17 July 2022; Accepted 11 August 2022; Published 10 September 2022

Academic Editor: Jun Zhang

Text generation has always been limited by the lack of corpus data required for language model (LM) training and the low quality of the generated text. Researchers have proposed some solutions, but these solutions are often complex and will greatly increase the consumption of computing resources. Referring to the current main solutions, this paper proposes a lightweight language model (EDA-BoB) based on text augmentation technology and knowledge understanding mechanism. Experiments show that the EDA-BoB model cannot only expand the scale of the training data set but also ensure the data quality at the cost of consuming little computing resources. Moreover, our model is shown to combine the contextual semantics of sentences to generate rich and accurate texts.

1. Introduction

Natural language generation (NLG) is an essential and challenging task in the current natural language processing (NLP) field [1]. And text generation task usually takes text as input, uses the word vector to represent it semantically, and finally, generates understandable natural language text. In Figure 1, all traditional NLP tasks can be transformed into text generation tasks. Typical text generation tasks include machine translation, document summarization, and dialogue systems. Since the Seq2Seq [2] framework was proposed in 2014, text generation has rapidly become a hot research topic. The Seq2Seq is based on an encoder-decoder mechanism, in which the encoder maps input text sequences to fixed-size vectors, and the decoder maps the vectors to target sequences. This new scheme solves the problem of many-to-many forms in NLP tasks and the varying length of input and output sequences by adding a semantic encoding layer. With the recent resurgence of deep learning technologies, the deep neural NLG model has remarkably enabled machines to understand and generate natural language [3]. And researchers have proposed a series of classic and practical models, such as recurrent neural network (RNN), convolutional neural network (CNN), and transformer [4]. Based on these models, the attention mechanism and the copy/pointer-generator [5] have also extensively promoted the research of text generation.

But the current development of text generation also faces some problems [6]. For instance, the quality of generated text is often limited due to the lack of large-scale and high-quality annotated corpus data for deep neural NLG model training. In addition, due to the limited knowledge contained in the input text, LM generally suffers from inability to understand language well, retain and recall knowledge using memory, and reason about complex concepts and relational paths. These problems often limit the quality of the output text by the NLG model.

Consequently, this paper intends to use text augmentation to generate expanded training corpus data. And, we will use a solution based on a knowledge understanding mechanism to solve the second problem. The framework of our scheme is shown in Figure 2.

The primary advantages of the scheme are summarized as follows:

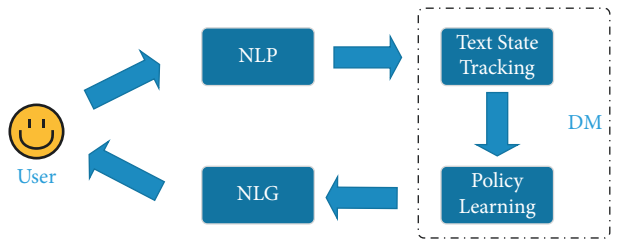


Figure 1: NLG framework.

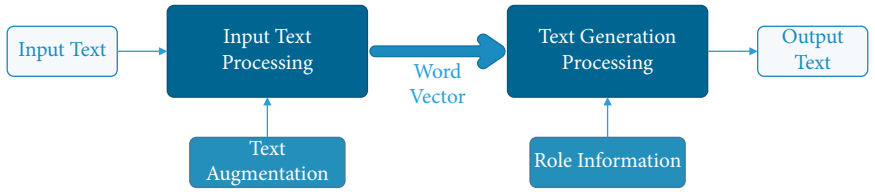


Figure 2: Our scheme framework (1).

(1) We solve the problem of limiting language model training due to the lack of corpus data by introducing lightweight text enhancement technology and ensuring training data quality.

(2) Our scheme adopts a multi-LM combination scheme, which meets the model's needs for both knowledge understanding and text generation tasks.

(3) We successfully combined the abovementioned two parts of the work. Based on this, this paper improves the performance of text-generating language models on a limited-scale corpus.

(4) This paper designs a simplified training data processing method and proposes a scheme to reasonably evaluate the generated text with a few indicators. These methods are beneficial to the realization and evaluation of text generation schemes.

In general, our model is a lightweight language model. And the EDA data augmentation method can greatly expand the training corpus data with a simple method and less resource consumption, which will help improve the accuracy of the language model. In addition, the knowledge understanding mechanism introduced by the model can combine the context information of sentences to generate richer and more accurate texts. This lightweight model makes it possible to obtain high-quality text generation in limited resource scenarios.

The remainder of our paper is organized as follows: Section 2 briefly explains the related work. Our scheme is detailed in Section 3. We discuss the performance evaluation results in Section 4. Finally, Section 5 is the conclusion.

2. Related Work

In this section, we introduce the current trend of text augmentation and text generation in detail. And, we compare the traditional scheme and the deep learning-based scheme of text augmentation. Meanwhile, we enumerate several general methods for text generation, and the method based on the knowledge understanding mechanism will be introduced emphatically.

2.1. Text Augmentation. With the development of AI, the requirements for the scale and quality of data for neural network models are gradually increasing. If the amount of data in different categories is very different in classification tasks, the model will overfit, which will seriously affect prediction accuracy. In this regard, data augmentation is a powerful solution. By using limited labeled data, more training data can be obtained. The phenomenon of overfitting in the network can be reduced, thereby training a model with more vital generalization ability.

Data augmentation techniques were initially applied in computer vision, mainly using various techniques to generate new training samples [7], which can create new data by translating, rotating, compressing, and adjusting the colors of images. Although the "new" sample data changes its appearance to a certain extent, the labels of the sample data always remain unchanged, which is beneficial to ensure the accuracy of the sample data. Consequently, data augmentation is a simple and cost-effective solution for augmenting training sample data. But the sample data in NLP is discrete, making it impossible to achieve data augmentation by simply transforming the input data because replacing one word may change the meaning of the entire sentence [8]. Accordingly, this paper presents traditional and deep learning methods for text enhancement schemes, respectively.

2.2. Traditional Text Augmentation Method. The traditional data augmentation method in the NLP field can be mainly divided into two types. The first one is the back-translation method, and the other one is the noise-addition method, both of which are supervised methods [9].

The back-translation method translates the original data into other languages and then translates it back to the original language [10]. This method is a data augmentation

technique frequently used by NLP in machine translation and has been successfully used in several malicious comment classification competitions [11]. The language logic order of back-translation methods is different, which can increase the diversity of text data. Therefore, this method not only change the syntactic structure but also preserve the semantic information, which can generate new data, that is, very different from the original data. However, the data produced by back-translation methods are often too dependent on the quality of the translation system and may not be accurate in most cases. If the interface of some translation software is used, the user may also encounter situations such as account restrictions.

Correspondingly, the noise-addition method is to create new data similar to the original data by replacing words, deleting words, etc., on the basis of the original data [12]. Some papers use Gaussian noise, dropout noise, etc., to augment data. Nevertheless, the methods represented by easy data augmentation for text classification tasks (EDA) [13] use simpler algorithms to implement data augmentation in a combined form. And the EDA is the first method to explore text editing techniques for data augmentation comprehensively. By systematically evaluating the performance of EDA on multiple benchmark classification tasks, it is shown that EDA can provide substantial improvements to the training data of the model, especially on smaller datasets, as shown in Figure 3.

EDA mainly includes four methods: synonyms replace (SR), randomly insert (RI), randomly swap (RS), and randomly delete (RD). Reasonable use of these four methods can quickly and easily expand the scale of training samples, but there are also disadvantages. For instance, RI may cause the original training data to lose semantic structure and order. The addition of synonyms does not focus on the keywords in the sentence, and the diversity of data expansion will be more limited. On the other hand, RS does not change the morphemes of the original sentence, so the generalization ability of new sentence patterns, sentence patterns, and similar words is substantially improved. In practical research, the combination of SR and RS is often used.

2.3. Text Augmentation Method Based on Deep Learning. With the widespread application of machine learning, researchers have also done a lot on deep learning-based text augmentation methods. These text augmentation schemes can mainly be divided into semi-supervised learning and unsupervised learning methods. The semi-supervised learning method is proposed to use unlabeled data better and reduce the model's dependence on large-scale labeled datasets [14, 15]. Experiments have shown that this is a powerful learning paradigm. MixMatch [16] proposed by Google represents the semi-supervised learning method, which incorporates consistency regulation, pseudo label idea, entropy regularization, and MixUp technology. It works by guessing the low-entropy labels of unlabeled samples produced by the MixUp data augmentation method and mixing the unlabeled data with the labeled data. These

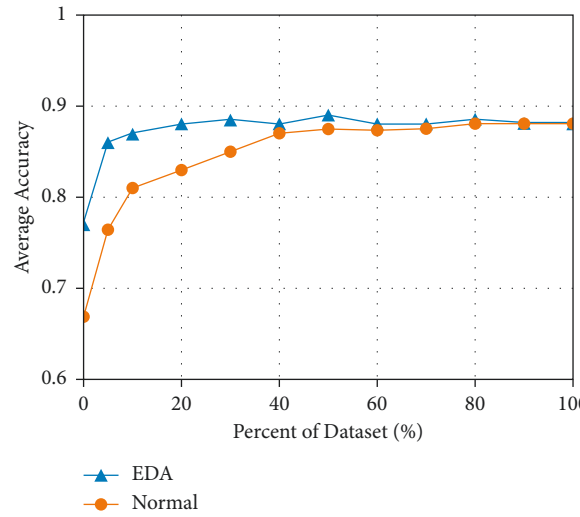


Figure 3: EDA performance comparison.

methods enable MixMatch to achieve significantly better results than previous semi-supervised learning techniques.

Unsupervised data augmentation (UDA) is also a successful data augmentation method [17]. Its success benefits from the adoption of mechanisms that use specific goals for specific tasks, that is, the targeted data augmentation (TDA). Compared to regular noise, this task-specific mechanism can generate more diverse and realistic noise and learn how to find missing or most wanted training signals in the original labeled dataset.

For unlabeled data, UDA uses KL divergence for augmented unlabeled data prediction results, which is different from MixMatch. The specific target data enhancement of TDA includes three methods: back translation, AutoAugment, and TFIDF. The Back Translation method can help enrich the sentence patterns and patterns of the data, and the TFIDF method optimizes UDA's random word processing strategy. In addition, another critical breakthrough of UDA is the use of the training signal annealing (TSA) method to gradually release the training signal during training, which avoids the situation that the labeled data and unlabeled data may be very different.

In general, traditional methods perform better on small batches of data. While deep learning-based methods can undoubtedly meet the needs of large-scale data, they are expensive to implement relative to the performance gain, so they are not often used in practice.

2.4. Text Generation. The text generation method is the foundation of many NLP tasks [18–20]. And the general text generation model is mainly based on the encoder-decoder framework [21]. The encoder learns to encode the input text into a vector representation, and the decoder is responsible for decoding this vector representation into a text sequence. In recent years, the blessing of technologies such as neural networks and deep learning has extensively promoted the development of text generation [22]. Since the Seq2Seq [23] model was proposed in 2014, text generation has rapidly become an important research point, and researchers have

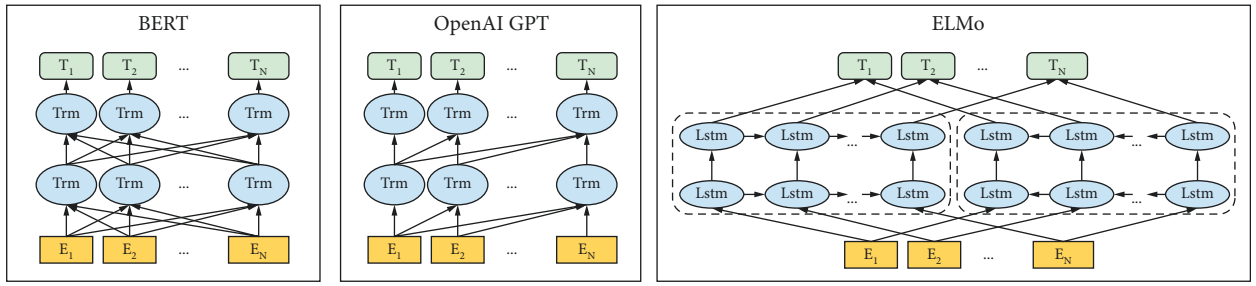


Figure 4: Structural comparison of Bert, GPT, and ELMo.

launched a series of classic and practical language models. The pretraining model has become a research trend in the current NLP field, and of course, they are also widely used in text generation. Its general and flexible framework design features can effectively solve the problem that the quality of the model is affected by the lack of labeled training data and provide good support for downstream tasks through fine-tuning. Based on these models, the attention mechanism and copy/pointer-generator mechanism has also greatly promoted text generation research [24]. Typical pretraining models include ELMo [25], GPT [26], and Bert [27] as representatives. Because Bert has added transformer [28] and bidirection technologies based on ELMo and GPT, it has become the most popular NLP language model. Figure 4 is a comparison frame diagram of Bert, GPT, and ELMo.

As a general model of NLP, Bert has been widely used in various scenarios, and its multi-stacked transformer network structure based on bidirectional encoding can significantly improve the expressiveness of the model. Meanwhile, Bert added MLM and NSP tasks in the pretraining process, making the trained model have the representational solid ability. Figure 5 shows the chart of Bert's network framework.

Meanwhile, Bert can provide significant support for various downstream tasks. Especially, when dealing with classification tasks, sequence labeling tasks, and knowledge reasoning tasks, you only need to add the relevant structure of the corresponding task after Bert's output without adjusting Bert. However, Bert is not naturally suitable for text generation tasks due to the inconsistency between Bert's MLM mechanism and the goal of generating tasks. It can be solved by adjusting Bert's mask matrix. The specific way is to make Bert predict the current word only based on the previous content of the text and ignore the latter content to continuously generate the word at the current position until the predicted word is (CLS) label, which is similar to the principle of the UNILM [27]. As shown in Figure 6, Bert can almost be competent for all NLP downstream tasks after field adaptation and task adaptation.

However, researchers also found that traditional text generation methods only rely on input text to generate, lacking richer "knowledge" information, so the generated text is often boring and lacks interesting content [29]. For text generation tasks, knowledge can transcend the semantic limitations in the input text and help the text generation system to generate richer and more interesting text.

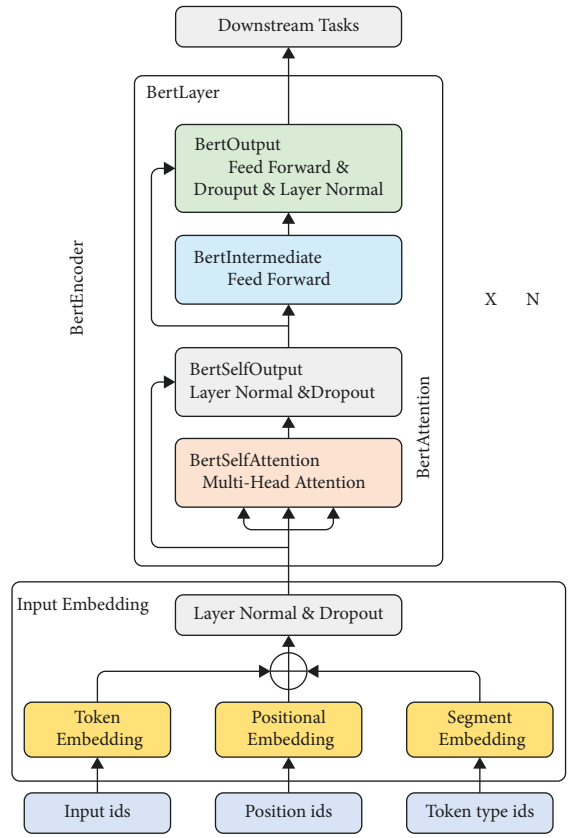


Figure 5: Bert network.

Therefore, knowledge-enhanced text generation has become the focus of researchers.

In the text generation task, knowledge can be obtained through different methods and information sources, including keywords, topics, language features, knowledge bases, knowledge maps, and the acquisition technology is very mature. Based on this, researchers have widely incorporated knowledge into text generation models to improve the quality of text generation. Reference [30] introduces an attention mechanism, which is mainly used to describe the importance of the input text to the generation process by adding a context vector to the decoder. Reference [31] proposed the CopyNet framework, which designed knowledge patterns and knowledge-related dictionaries for knowledge and used copy/ pointer-generator mechanism for the generation of text output sequences. Reference [32]

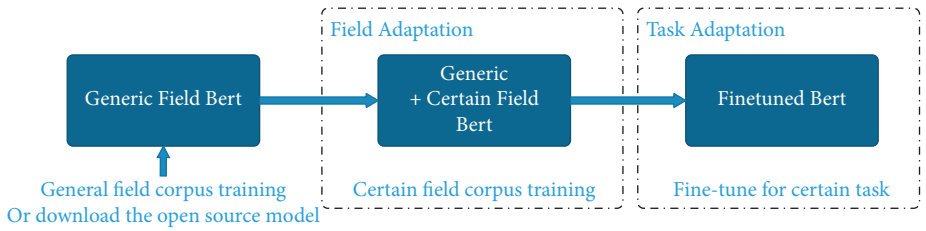


Figure 6: Bert's task processing flow.

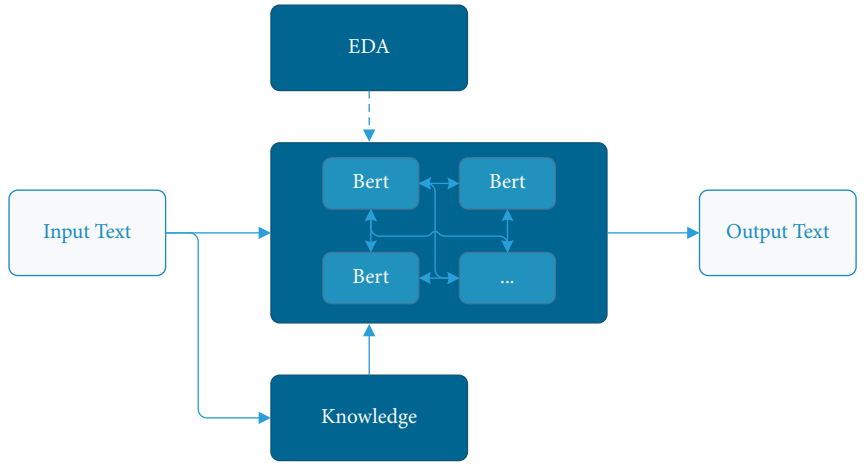


Figure 7: Our scheme framework (2).

proposes an approach to memory networks, a recurrent attention model that acts on external storage by cyclically exploiting the input sequence to read the memory representation on the storage, and then write the updated memory representation back to the storage superior. The pretraining model proposed in [33] is currently a widely used solution. By using large-scale unlabeled datasets for pretraining models, these models can provide better model initialization for the text generation task model to solve the problem of insufficient generalization ability. The proposed methods not only increase the sentiment of the generated text but also effectively combine the context and contextual information to improve the accuracy of text generation.

3. Our Scheme

As mentioned above, EDA can achieve better text augmentation effects in small-scale data application scenarios. The improved pretrained language model (Bert) can meet the needs of text generation after introducing the knowledge understanding mechanism. Therefore, our scheme will be based on EDA and Bert, as shown in Figure 7. First, the high-quality annotated training sample data of the text generation model is boosted by EDA; second, the multi-Bert combination scheme is adopted to solve the knowledge understanding task and the text generation task simultaneously.

3.1. Our EDA Method. The main body of EDA is four data augmentation algorithms, SR, IR, RS, and RD. And the SR and RS are suitable for short texts, while the RI and RD are more suitable for text augmentation of long texts. In addition, it also includes important configuration parameters such as n and α. Among them, we use n to control the number of target sentences generated corresponding to each source sentence, and α is a parameter that indicates the percent of the words in a sentence is changed by augmentation. Since the change of the words in the sentence may change the semantics of the sentence, many factors need to be comprehensively considered when selecting the four algorithms of EDA and setting n and α.

Figure 8 shows that the average performance gain of EDA operations over five text classification tasks for different training set sizes, while varying the augmentation parameter $\alpha = \{0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$. From the results, the average performance of the four algorithms of EDA is different. In contrast, the gain of RI on performance is relatively low, RD is greatly affected by the change of α value, and the realization effect of SR and RS is relatively good. Since the language model in this paper is mainly oriented to short texts, to reduce the impact of RI and RD on the semantics of the original text as much as possible, this paper only uses the SR and RS algorithms based on setting the α value at a low value. In Figure 9, our new EDA method uses random operator K to control our method to choose whether to use SR, RS, or SR + RS. We mainly use SR or RS for short texts. For longer texts, we use the SR + RS algorithm more. In the SR + RS algorithm, α is divided into $\alpha 1$ and $\alpha 2$, respectively, used to control the number of words in the text for synonym replacement and random swap.

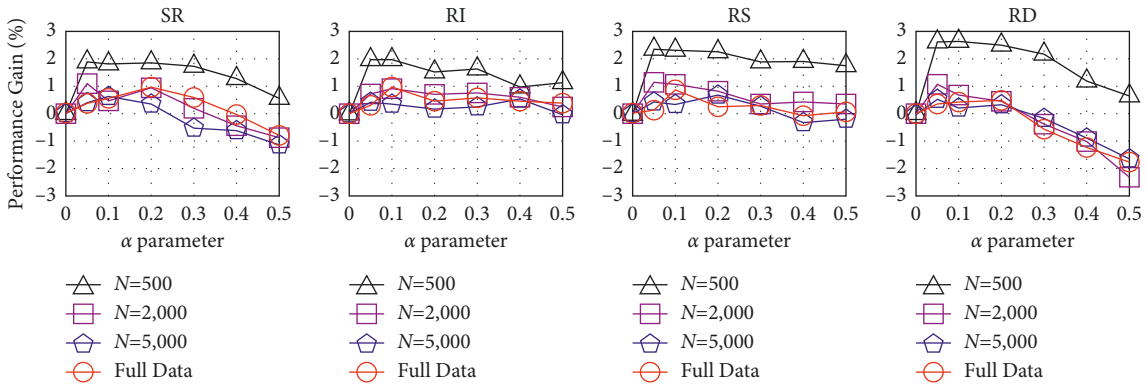


Figure 8: Average performance gain of EDA.

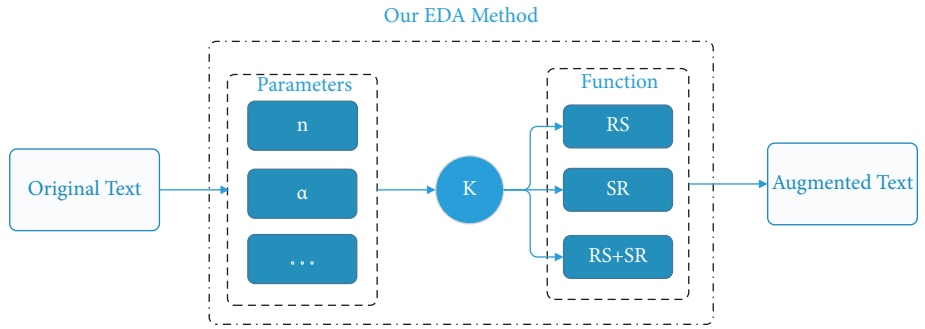


Figure 9: Our EDA method is based on SR and RS.

Figure 10 shows the augmentation result of one sample text when $n = 10$ and $\alpha = 0.1$.

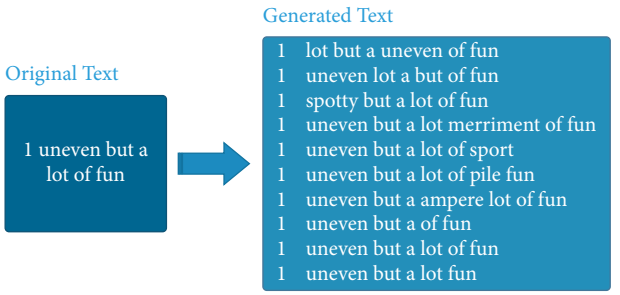


Figure 10: Text augmentation results.

3.2. Text Generation Model (EDA-BoB). On the one hand, the problem of inconsistency in dialogue responses is one of the biggest challenges in the NLP field, and it is difficult to eradicate this phenomenon with large-scale training data alone. The language model based on knowledge understanding can better understand the contextual information to generate richer and more accurate text. Language models can understand many kinds of knowledge, and the user role information is widely used because of its simple structure and easy training. On the other hand, since the independent Bert can only handle a specific NLP task, to enable the language model to handle both the knowledge understanding task and the text generation task, we adopt the Bert-over-Bert (BoB) model proposed in [34]. BoB is a new model based on the combination of multi-Bert, including a Bert encoder and two Bert decoders, in order to separate the acquisition of understanding ability and generation ability. Once, we separate the two, whether character information understanding or text generation, we can find sufficient data resources for training.

The model consists of an encoder E, an autoregressive decoder $D1$ for responding to dialogue replies, and a bidirectional decoder $D2$ for consensus understanding. Given role information P and dialogue input Q, E, and $D1$, the model operates in a classical encoder-decoder mode to learn a typical input-to-reply mapping $F_G(S|Q, P)$ and generate a preliminary dialogue-reply representation $R1$. Then $R1$ and persona P are fed to a bidirectional decoder $D2$ to map $R1$ to the final response representation $R2$: $F_U(R2|S, P)$. Since the part $F_U(R|S, P)$ that learns consistent understanding is independent of the dialogue input Q, the model can learn this part on nondialogue inference datasets. Here, we refer to the previous work [35] and introduce the unlikelihood objective function to reduce the possibility of contradictory data in the inference data so that $D2$ can obtain the ability of consistent understanding. We use the pretrained Bert model to initialize all modules. The overall structure of the BoB model and the corresponding training method are shown in Figure 11.

The working principle of the encoder E is similar to the standard Bert model, which converts the input text into a word vector through the embedding layer shared by all

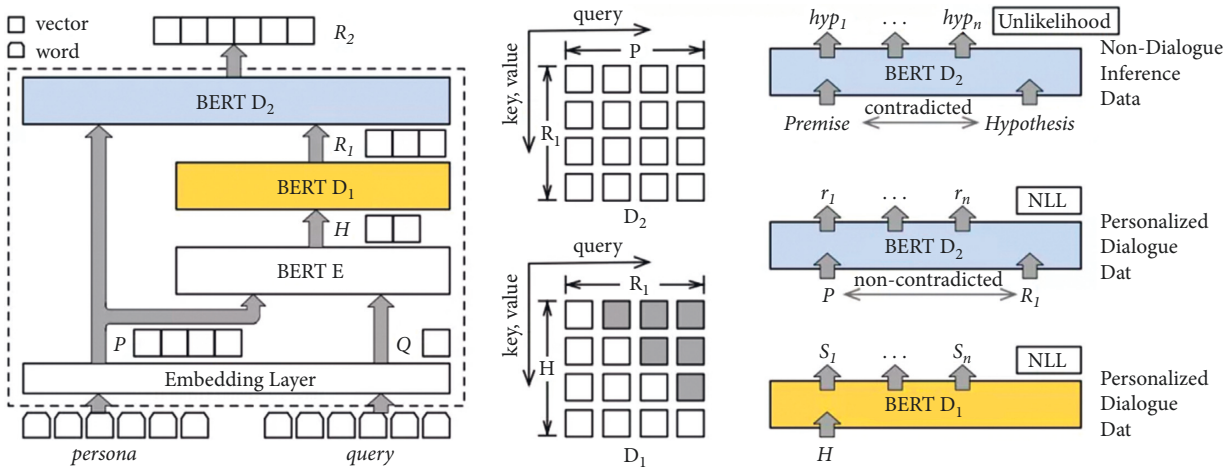


Figure 11: BoB framework and training objectives.

modules. It then is encoded into a vector representation H by the encoder E through a multilayer bidirectional self-attention mechanism.

Benefiting from initialization, *D1* inherits Bert's robustness. The difference is that *D1* works as an autoregressive decoder. When Bert predicts a mask word, it uses the left and correct bidirectional information of the word. Still, in the autoregressive generation task, the words are indicated from left to right one by one. In Figure 11, to eliminate this inconsistency, this paper is inspired by UniLM to upper-triangular mask matrices during training and prediction to ensure that the generated reply words can only rely on existing information. In addition, this scheme is similar to the classic sequence-to-sequence model, adding an attention mechanism between the decoder and encoder can effectively improve the performance.

The *D2* is the key for the BoB model to separate understanding and generation tasks. Like *E* and *D1*, *D2* is also initialized from Bert and thus inherits an excellent semantic representation for text understanding tasks. The task of *D2* is to learn how to understand coherence relations and apply this ability to the task of reply generation, so it needs to utilize bidirectional information for better learning. To achieve this, *D2* is trained with both the standard cross-entropy loss for text generation and the unlikelihood loss for consensus understanding.

The BoB model can learn from limited role-based dialogue data by separating reply generation and consistent understanding and introduces the unlikelihood training based on nondialogue reasoning data, which improves the model's consistent understanding ability. The *E*, *D1*, and *D2* are all based on the generic Bert model, and the Bert-base-uncased version can be downloaded directly from the Hugging face [36]. Since the Bert-base-uncased is an already trained model, we only need to train it for task adaptation. Experiments show that the BoB model can achieve better results than strong baseline methods trained with full data, even with less characterized data [34].

4. Experiments

This section will experiment with the performance of EDA and the BoB model augmented by EDA. The experiments mainly verify the scheme's performance by comparing the semantic similarity between the original text and the generated text. And this paper especially calculates the similarity of short texts. Generally, it is necessary to embed two short texts and then calculate the cosine similarity. Word embedding methods such as word2vec and GloVe have become standard methods for finding semantic similarities between words. Specifically, it includes the benchmark method based on the average value of word embeddings, the word shift distance method based on the shortest distance of words, the smooth inverse frequency method, and the method based on pretrained encoders [38, 39]. On this basis, many mature short text similarity schemes have been proposed. The pretrained language model (ERNIE) based on Baidu's self-developed performance performs well [40], and it is very convenient to use. The ERNIE model has a more vital semantic understanding ability and can deeply understand the semantic relationship between texts.

Baidu's solution is an online method, where users can send requests to the server in POST mode through the API provided by Baidu. The scheme can return the similarity value of two texts each time, and the maximum length of each text is 512 kB. The *score* represents the similar result, and the value range is (0, 1). The higher the *score*, the higher the similarity.

4.1. EDA Performance Analysis. In Section 3.1, n and α are two main parameters of the EDA, where n controls the number of target sentences generated corresponding to each source sentence, and α is a parameter that indicates the percent of the words in a sentence are changed by augmentation. In our experiment, we modified the value of n to {1, 2, 4, 8, 16, 32}, and α was {0.1, 0.2, 0.3, 0.4, 0.5},

respectively, and finally analyzed the text semantic similarity between the generated text after EDA augmentation and the original text. For the n texts generated from one original text, we take the average semantic similarity of the n texts as the comparison value. The experimental results are shown in Figure 12 and Table 1.

In Figure 12, it is first shown that as the value of n increases, the text semantic similarity decreases. The reason is that EDA will prioritize adopting a better modification scheme for the original text. When the number of texts generated is larger, the similarity between the generated text and the original text will be lower. Secondly, the increase of the α will also increase the degree of modification of the original text by EDA. The reason is that when the α increases, more words are changed in the original sentence. However, since EDA only changes the structure and content of the statement by simply replacing, moving, etc., when the α is too large, the semantics of the original statement will be changed. This will lead to a large difference in text semantic similarity.

After many experiments, we show that the quality of the generated text is higher when the text semantic similarity is above 75%. For the short text dataset, text augmentation can achieve better results when the combination of n and α is in the green area of Table 1.

A good evaluation metric can effectively guide the model to fit the data distribution and objectively evaluate the quality of the text generation model. However, due to the inherent complexity of natural language and the limitations of current technology, there is no perfect evaluation indicator yet. Reference [40] lists several types of common [41–44] evaluation indicators: word overlap evaluation metric (WOEM) and word vector evaluation metric (WVEM). The WOEM includes BLEU, ROUGE, and METEOR, and the WOEM is represented by greedy matching, embedding average, and vector extrema. Perplexity (PPL) is a widely used language model-based method because it can also directly compare the pros and cons of two language models in predicting samples. In addition to the above-mentioned automatic measurement methods, this paper also introduces several manual measurement methods.

$$\begin{aligned} PP(S) &= P\left(w_1 w_2 \ldots w_N\right)^{-1/N} \\ &= \sqrt[N]{\frac{1}{P\left(w_1 w_2 \ldots w_N\right)}} \\ &= \sqrt[N]{\prod_{i=1}^{N} \frac{1}{P\left(w_i | w_1 w_2 \ldots w_{i-1}\right)}}. \end{aligned} \tag{1}$$

For the BoB model, our experimental scheme consists of two parts. The first part is one horizontal comparison scheme. We compare the PPLs of transformer, GPT2, BoB, and EDA-BoB to verify that the EDA-BoB model has high text generation quality. The specific method is to calculate the PPL of each model on the respective test set. And TensorFlow can provide a simple method for calculating

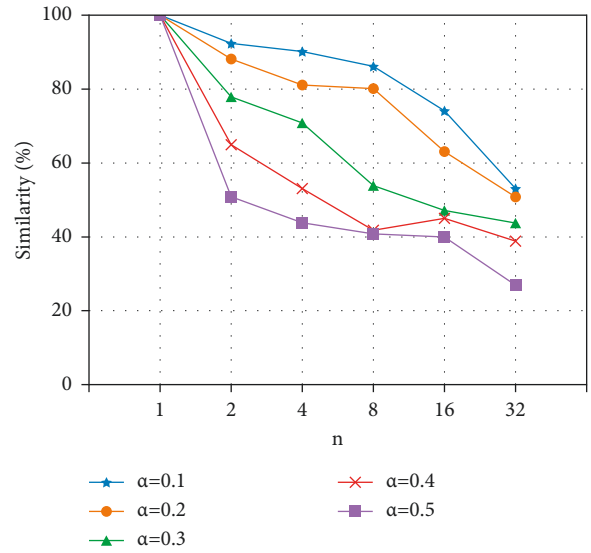


Figure 12: Text semantic similarity with EDA.

Table 1: Text semantic similarity with EDA (1).

α	n					
	$n=1$	$n=2$	$n=4$	$n=8$	$n=16$	$n=32$
$\alpha=0.1$	100	92	90	86	75	56
$\alpha=0.2$	100	88	81	80	63	51
$\alpha=0.3$	100	78	71	54	47	44
$\alpha=0.4$	100	65	53	43	45	39
$\alpha=0.5$	100	50	44	42	40	27

BoB performance analysis.

PPL. Specifically, when using the *tf.contrib.seq2seq.sequence_loss()* function to calculate the *loss* of the model, take the exponential operation on the *loss* directly. The results are shown in Table 2. Compared with the first two models, BoB has apparent advantages, and EDA can improve BoB.

On the other hand, we adopted a similar scheme to the EDA experiment. We set the α to 0.2 and analyze the similarity between the text generated by our model and the text in the test set when the $n = \{1, 2, 4, 8, 16, 32\}$. Meanwhile, we also conducted a human survey to determine the augmentation effect of EDA on BoB. The scheme is based on human subjective satisfaction with the generated text, which can better reflect the richness and consistency of the text. Thirty volunteers from the computer science department of the author's institution each obtained 20 sets of data, each consisting of a question and an answer. And volunteers were asked to judge the quality of the text on a scale of 1 to 100. The higher the ranking, the higher the satisfaction. The experiment also analyzed the change in volunteer satisfaction when the $n = \{1, 2, 4, 8, 16, 32\}$, and the results are shown in Figure 13. The volunteers are generally satisfied with the quality of the text generated by EDA-BoB, and the increase of the n value has a slight improvement to the model.

But we can also see from the results that when the n is increasing, the similarity between the generated text pair and

Table 2: Text semantic similarity with EDA (2).

Model	PPL
Transformer	28.8
GPT2	14.4
BoB	7.73
EDA-BoB	7.35

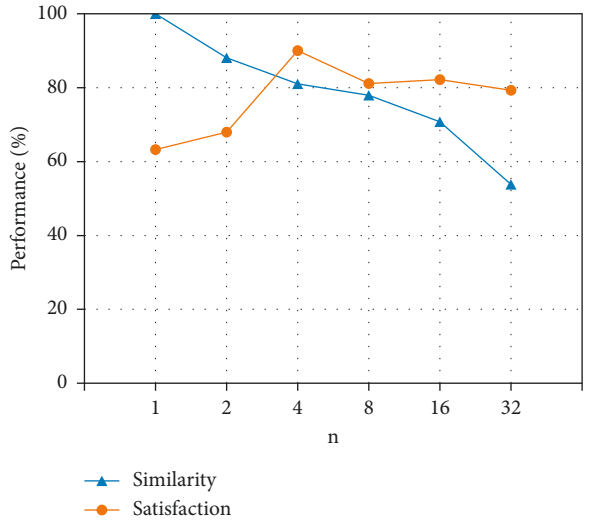


Figure 13: EDA-BoB performance.

the original text pair is always decreasing. The main reason is that the more text pairs are generated, the more different the model is from the generated results, and the lower the match between the question and the answer. Meanwhile, the satisfaction reaches the highest 92% when n is 4, and then declines and remains at 80%. This shows that the increase of the n has an obvious effect on the quality improvement of the generated text in the early stage, but too much EDA augmented data will not continue to improve the model.

5. Conclusion

The limited size of the training corpus and the lack of emotion in the generated sentences are the two major problems for text generation tasks. Based on text augmentation methods and knowledge understanding mechanisms, this paper proposes a lightweight language model (EDA-BoB). This model uses a lightweight text augmentation approach (EDA) to expand the training dataset size, which can improve model performance and reduce overfitting. Furthermore, this paper adopts a multi-Bert model (BoB) based on the understanding mechanisms for text generation. And this BoB model can learn user role information and use it to generate text with richness and consistency. Experiments show that EDA can further improve the performance of BoB with a small computational cost. Our future work will focus on researching lightweight text augmentation methods for large datasets and designing general solutions for multiple language models.

Data Availability

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study received support from the following sources: the Natural Science Foundation of Anhui Province (Grant no. 2008085MF200), the Huainan Normal University Scientific Research Project (Grant no. 2022XJZD030), the University Natural Science Foundation of Anhui Province (Grant no. KJ2021A0970), and the Key Research and Development Plan Project Foundation of Huainan (Grant no. 2021A248).

References

[1] O. Dušek, J. Novikova, and V. Rieser, "Evaluating the state-of-the-art of end-to-end natural language generation: the e2e nlg challenge," *Computer Speech & Language*, vol. 59, pp. 123–156, 2020.

[2] E. Egonmwan and Y. Chali, Eds., *Proceedings of the 3rd Workshop on Neural Generation and Translation*pp. 249–255, 2019.

[3] A. M. A. L. K. Sankaranarayanan and P. J. S. Krishnan, "Storytelling from structured data and knowledge graphs an NLG perspective," *ACLPPinforma*, p. 43, 2019.

[4] T. Iqbal and S. Qureshi, *The Survey: Text Generation Models in Deep Learning*, pp. 1–14, Journal of King Saud University-Computer Information Sciences, China, 2020.

[5] A. Galassi, M. Lippi, and P. Torroni, "Attention in natural language processing," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 10, pp. 4291–4308, 2021.

[6] A. Celikyilmaz, E. Clark, and J. Gao, "Evaluation of text generation: a survey," 2020, http://arxiv.org/abs/200614799.

[7] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of big data*, vol. 6, no. 1, pp. 60–48, 2019.

[8] M. Abulaish and A. K. Sah, Eds., in *Proceedings of the 2019 11th International Conference on Communication Systems & Networks (COMSNETS)*, pp. 625–630, IEEE, Bengaluru, India, 07-11 January 2019.

[9] M. Bayer, M.-A. Kaufhold, and C. Reuter, "A survey on data augmentation for text classification," 2021, http://arxiv.org/abs/03158.

[10] A. W. Yu, D. Dohan, M.-T. Luong, R. Zhao, K. Chen, and M. Norouzi, "Qanet: combining local convolution with global self-attention for reading comprehension," 2018, http://arxiv.org/abs/200614799.

[11] V. C. D. Hoang, P. Koehn, G. Haffari, and T. Cohn, Eds., *Proceedings of the 2nd Workshop on Neural Machine Translation and Generation*pp. 18–24, 2018.

[12] F. J. Moreno-Barea, F. Strazzera, J. M. Jerez, D. Urda, and L. Franco, Eds., in *Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 728–734, IEEE, Bangalore, India, 18-21 November 2018.

[13] J. Wei and K. Zou, "Easy data augmentation techniques for boosting performance on text classification tasks," 2019, http://arxiv.org/abs/11196.

[14] E. Cho, H. Xie, and W. M. Campbell, "Paraphrase generation for semi-supervised learning in NLU," in *Proceedings of the Workshop on Methods for Optimizing and Evaluating Neural Language Generation*, pp. 45–54, June 2019.

[15] J. Bose and S. Mukherjee, "Semi-supervised method using Gaussian random fields for boilerplate removal in web browsers," in *Proceedings of the 2019 IEEE 16th India Council International Conference (INDICON)*, pp. 1–4, IEEE, Rajkot, India, December 2019.

[16] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "Mixmatch: a holistic approach to semi-supervised learning," *Advances in Neural Information Processing Systems*, vol. 32, p. 45, 2019.

[17] Q. Xie, Z. Dai, E. Hovy, T. Luong, and Q. Le, "Unsupervised data augmentation for consistency training," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6256–6268, 2020.

[18] S. Wiseman, S. M. Shieber, and A. M. Rush, "Learning neural templates for text generation," 2018, http://arxiv.org/abs/10122.

[19] C. Song and V. Shmatikov, "Auditing data provenance in text-generation models," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 196–206, Anchorage, AK, USA, August 2019.

[20] Z. Du, Y. Qian, X. Liu, M. Ding, J. Qiu, and Z. Yang, "All nlp tasks are generation tasks: a general pretraining framework," 2021, http://arxiv.org/abs/10360.

[21] M. Lusetti, T. Ruzsics, A. Göhring, T. Samardzic, and E. Stark, "Encoder-decoder methods for text normalization," in *Proceedings of the Fifth Workshop on NLP for Similar Languages, Varieties and Dialects (VarDial)*, pp. 18–28, Santa Fe, New Mexico, USA, August 2018.

[22] S. Narayan and C. Gardent, "Deep learning approaches to text production," *Synthesis Lectures on Human Language Technologies*, vol. 13, no. 1, pp. 1–199, 2020.

[23] Z. Li, J. Cai, S. He, and H. Zhao, "Seq2seq dependency parsing," in *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 3203–3214, Santa Fe, NM, USA, 2018.

[24] J. Xie, Y. Hou, Y. Wang et al., "Chinese text classification based on attention mechanism and feature-enhanced fusion neural network," *Computing*, vol. 102, no. 3, pp. 683–700, 2020.

[25] H. Gupta and M. Patel, "Study of extractive text summarizer using the elmo embedding," in *Proceedings of the 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, pp. 829–834, IEEE, Palladam, India, October 2020.

[26] Y. Qu, P. Liu, W. Song, L. Liu, and M. Cheng, "A text generation and prediction system: pre-training on new corpora using BERT and GPT-2," in *Proceedings of the 2020 IEEE 10th International Conference on Electronics Information and Emergency Communication (ICEIEC)*, pp. 323–326, IEEE, Beijing, China, July 2020.

[27] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: pre-training of deep bidirectional transformers for language understanding," 2018, http://arxiv.org/abs/04805.

[28] A. Gillioz, J. Casas, E. Mugellini, and O. Abou Khaled, "Overview of the transformer-based models for NLP tasks," in *Proceedings of the 2020 15th Conference on Computer Science and Information Systems (FedCSIS)*, pp. 179–183, IEEE, Sofia, Bulgaria, September 2020.

[29] W. Yu, C. Zhu, Z. Li, Z. Hu, Q. Wang, and H. Ji, "A survey of knowledge-enhanced text generation," 2020, http://arxiv.org/abs/04389.

[30] B. Bi, C. Wu, M. Yan, W. Wang, J. Xia, and C. Li, "Incorporating external knowledge into machine reading for generative question answering," 2019, http://arxiv.org/abs/02745.

[31] J. Gu, Z. Lu, H. Li, and V. O. Li, "Incorporating copying mechanism in sequence-to-sequence learning," 2016, http://arxiv.org/abs/06393.

[32] S. Sukhbaatar, J. Weston, and R. Fergus, "End-to-end memory networks," *Advances in Neural Information Processing Systems*, vol. 28, p. 1312, 2015.

[33] J. Guan, F. Huang, Z. Zhao, X. Zhu, and M. Huang, "A knowledge-enhanced pretraining model for commonsense story generation," *Transactions of the Association for Computational Linguistics*, no. 8, pp. 93–108, 2020.

[34] H. Song, Y. Wang, K. Zhang, W.-N. Zhang, and T. Liu, "BoB: bert over bert for training persona-based dialogue models from limited personalized data," 2021, http://arxiv.org/abs/06169.

[35] S. Welleck, I. Kulikov, S. Roller, E. Dinan, K. Cho, and J. Weston, "Neural text generation with unlikelihood training," 2019, http://arxiv.org/abs/04319.

[36] Bert. base-uncased, "Hugging face," 2022, https://huggingface.co/bert-base-uncased.

[37] Z. Chen, J. Sadoc, L. F. D'Haro, R. Banchs, and A. Rudnicky, "Automatic evaluation and moderation of open-domain dialogue systems," 2021, http://arxiv.org/abs/02110.

[38] P. Sitikhu, K. Pahi, P. Thapa, and S. Shakya, "A comparison of semantic similarity methods for maximum human interpretability," in *Proceedings of the 2019 artificial intelligence for transforming business and society (AITB)*, pp. 1–4, IEEE, China, 2019.

[39] L. Chen, "Short text similarity API: BaiDu," 2022, https://ai.baidu.com/ai-doc/NLP/ek6z52frp.

[40] C. W. Liu, R. Lowe, I. V. Serban, M. Noseworthy, and J. Pineau, "How NOT to evaluate your dialogue system: an empirical study of unsupervised evaluation metrics for dialogue response generation," in *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pp. 2122–2132, Austin, Texas, 2016 November.

[41] H. Liu, Y. Pan, S. Li, and Y. Chen, "Synchronization for fractional-order neural networks with full/under-actuation using fractional-order sliding mode control," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 7, pp. 1219–1232, 2018.

[42] S. Ha, L. Chen, and H. Liu, "Command filtered adaptive neural network synchronization control of fractional-order chaotic systems subject to unknown dead zones," *Journal of the Franklin Institute*, vol. 358, no. 7, pp. 3376–3402, 2021.

[43] T. Huang, Q. She, and J. Zhang, "Boostingbert: integrating multi-class boosting into BERT for NLP tasks," 2020, http://arxiv.org/abs/2009.05959.

[44] H. Choi, J. Kim, S. Joe, and Y. Gwon, "Evaluation of bert and albert sentence embedding performance on downstream nlp tasks," in *Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 5482–5487, IEEE, Newon, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 9249530, 13 pages
https://doi.org/10.1155/2022/9249530

Research Article

An Efficient Deep Learning Mechanism for the Recognition of Olive Trees in Jouf Region

Hamoud H. Alshammari[1] and Osama R. Shahin[2]

[1]*Department of Information Systems, Computer and Information Sciences College, Jouf University, Sakaka, Saudi Arabia*
[2]*Department of Computer Science, College of Science and Arts in Qurayyat, Jouf University, Qurayyat, Saudi Arabia*

Correspondence should be addressed to Osama R. Shahin; orshahin@ju.edu.sa

Received 14 June 2022; Accepted 18 July 2022; Published 31 August 2022

Academic Editor: Heng Liu

Olive trees grow all over the world in reasonably moderate and dry climates, making them fortunate and medicinal. Pesticides are required to improve crop quality and productivity. Olive trees have had important cultural and economic significance since the early pre-Roman era. In 2019, Al-Jouf region in a Kingdom of Saudi Arabia's north achieved global prominence by breaking a Guinness World Record for having more number of olive trees in a world. Unmanned aerial systems (UAS) were increasingly being used in aerial sensing activities. However, sensing data must be processed further before it can be used. This processing necessitates a huge amount of computational power as well as the time until transmission. Accurately measuring the biovolume of trees is an initial step in monitoring their effectiveness in olive output and health. To overcome these issues, we initially formed a large scale of olive database for deep learning technology and applications. The collection comprises 250 RGB photos captured throughout Al-Jouf, KSA. This paper employs among the greatest efficient deep learning occurrence segmentation techniques (Mask Regional-CNN) with photos from unmanned aerial vehicles (UAVs) to calculate the biovolume of single olive trees. Then, using satellite imagery, we present an actual deep learning method (SwinTU-net) for identifying and counting of olive trees. SwinTU-net is a U-net-like network that includes encoding, decoding, and skipping links. SwinTU-net's essential unit for learning locally and globally semantic features is the Swin Transformer blocks. Then, we tested the method on photos with several wavelength channels (red, greenish, blues, and infrared region) and vegetation indexes (NDVI and GNDVI). The effectiveness of RGB images is evaluated at the two spatial rulings: 3 cm/pixel and 13 cm/pixel, whereas NDVI and GNDV images have only been evaluated at 13 cm/pixel. As a result of integrating all datasets of GNDVI and NDVI, all generated mask regional-CNN-based systems performed well in segmenting tree crowns ($F1$-measure from 95.0 to 98.0 percent). Based on ground truth readings in a group of trees, a calculated biovolume was 82 percent accurate. These findings support all usage of NDVI and GNDVI spectrum indices in UAV pictures to accurately estimate the biovolume of distributed trees including olive trees.

1. Introduction

Unmanned aerial systems (UAS) were currently used in a number of agricultural improvement projects due to their adaptability and low-cost [1]. Typically, the unmanned aerial vehicles (UAV) will fly over the area of interest, gathering aerial photos. Flight time could variety from few minutes to many hours, regardless of the model of the UAV and also the region to be explored, amongst many other things [2]. Numerous take-offs and recoveries could well be required when employing a short-endurance UAV. The UAV operator can manually assess the sensing accuracy by transferring several low-resolution photos to a bottom either during a technological stoppage. The maximum resolution photographs are saved on the camera's storage device. When the aerial job is completed, it is transported to a desktop for postprocessing [3]. The Pix4Dmapper, Agi-soft Photo Scan correlator3D, have been used to create a mosaic in which most of the images were patched together with a substantial overlapping (60–80 percent). Despite being present, autopilot information is rarely used through this postprocessing.

The sophistication of image processing methods [4], as well as the vast number of photographs captured in solitary aerial work (thousands), necessitates the use of sophisticated

computational resources. As a result, this postprocessing should be performed on the facilities of UAV operators, employing computers having substantial quantities of storage, strong coprocessors, or computing clusters. The entire procedure could easily require 2-3 days to complete, also with annoyance that this lost time implies for the end-user. In a survey report on thermal remotely sensed related to accuracy agriculture [5], the travel time is estimated to be much longer (between 1 and 3 weeks). A geographic mosaic built using two giga-bytes of the data, for example, required 25 hours of CPU period if there is no parallel used. Furthermore, the scientists reported a lengthy (about 20 h) processing time for a series of multispectral photos (around 100) obtained using a UAV [6]. The authors created orthomosaics with ultrahigh quality using a software package running on the powerful system (a 12−core Intel *i*7 computer and 64 GB of RAM) (14.8 cm per pixel) [7]. Figure 1 summarizes the technology process of the unique tree picture segmentation and recovery approach used. First, we used a UAV image (develop at different image), a DSM, and a gradient model to separate each tree crown. Next, we created a visual representation of the ground's truth map. Finally, we labeled every tree image with such a regression coefficients label [8].

The utilization of unmanned aerial vehicle (UAV) images using near-infrared (NIR), greenish, red, and blue multispectral images is effectively implemented in personalized agriculture for evaluating plant growth and condition [9]. Spectral indicators like the normalization differential vegetative indices (NDVI) or greenish normalization differential vegetative indices (GNDVI) could be used to evaluate crop kind, productivity, and maturing phase [10]. The GNDVI indicator is added responsive to the fluctuations in a crop chlorophyll concentration which that of the NDVI indicators, but it also has a comparatively higher criterion, so it could be used in crops too closely packed canopies even in a more sophisticated growth phase, as well as to assess moisture levels and nitrogen ability to concentrate levels in plant leaf. On either hand, the NDVI value is especially useful for evaluating crop vigor during the early phases of growth [11].

Object counting has been a good computer vision issue that seeks to determine the number of things in a stationary image or a video. Object counting is indeed an established research field with numerous applications in a variety of fields, including ecological studies, population counts, microcell tallying, and vehicle counts [12]. Handcrafted characteristics (such as SIFT and HOG) were retrieved from such a stationary image to identify and measure olive trees using traditional approaches. Nevertheless, several variables including scale fluctuations, climate changes, perception abnormalities, and orientation modifications have an impact on the effectiveness of these old methods [13, 14]. Deep learning identification techniques including single shot multibox detectors (SSD) and regional-convolutional neural networks (R–CNN) have newly reached great effectiveness and proposed a possible answer to these difficulties [15]. In spite of the popularity of deep learning technologies, a standardized collection of olive trees is not accessible for deep learning purposes. As a result, we began by building a large-scale of olive database with both deep learning

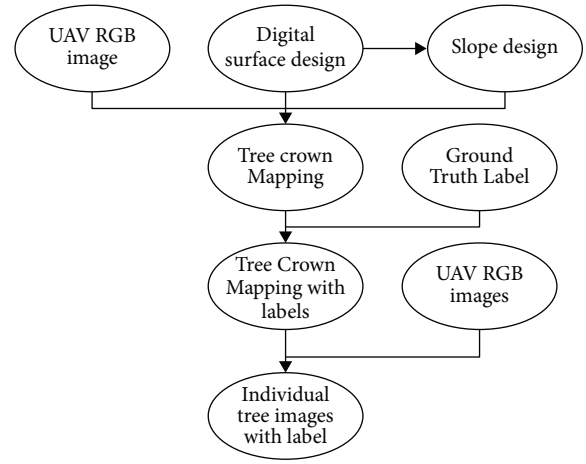


Figure 1: Image segmentation.

technology and implementations. The database is made up of 250 RGB satellite images acquired in Al-Jouf, Saudi Arabia. Satellites Pro, which offers satellite photos and mapping with most nations and towns across the world, provided the images. The olive trees were tagged with a center fact to decrease the workload and speed up the annotating process. Several olive tree deep learning applications, including identification, count, and classification, can benefit from the suggested dataset [16].

Deep learning (DL) approaches in overall, and CNNs specifically, have outperformed conventional techniques in identifying spatial characteristics from realistic RGB images. In addition, CNNs are at the cutting edge of all core computer vision applications, including classification tasks, object identification, and sample segmentation. Utilizing semantic segmentation techniques including mask regional-CNN, among the most effective CNN-based segmentation techniques, is a useful way to reliably predict olive tree crowns [17]. DL-CNNs' fundamental weakness is the systems demand a huge training of database to provide decent results. Many optimizations, such as transfer training, precise tuning, data enhancement, and possibly data fusion, were employed to solve this constraint in real systems.

The goal of this paper is to show how deep CNNs may be used to estimate the activities connected to olive-tree fields based on treetops and shadows found in ultrahigh precision photos (less than 30 cm) [18]. First, we educated CNNs to recognize olive tree crowns and shadow sections. The tree biovolumes were then estimated using the tree's crown areas, and the tree heights were calculated using the shadow distances. Earlier research on the shrubs and trees concentrated on detecting species of plants or damaging phases in photos captured by unmanned aerial vehicles (UAVs) [19]. As far as understood, it was the first study of the semantic segmentation job for plant species identification to determine tree biovolume. The following are the primary achievements of this paper.

Then, motivated by the Swin Transformer's performance, we present an actual deep learning network (SwinTU-net) for identifying and counting of olive trees. SwinTU-net is a system similar to U-net that features

encoding, decoding, and skip connectors. Swin Transformer is employed instead of the convolution technique to develop good locally and globally semantic features. According to the findings of an experimental investigation on the offered database, the offered SwinTU-net approach beats the comparable research in terms of total identification, using a 0.94 percent prediction error. OTCS-dataset is a novel documented multispectral of orthoimages database for the olive tree crown identification. At two-dimensional frequencies (3 cm/pixel and 13 cm/pixel), the OTCS-dataset is divided by the semantic segmentation of mask regional-CNN model was assessed for the objectives of olive tree crown identification and shadow identification in UAV pictures. This paper presents a model that enhances identification over algorithms involving picture fusion by fusing RGB images using vegetation indices. The biovolume for olive trees was calculated using the region of the crowns and the heights deduced after their length of shadow's. The outcome reveal using NDVI and GNDVI spectrum index data with such a resolution of 13 cm/pixel is sufficient for effectively measuring the biovolume of olive plants into four subgroups of distinct spectral channels and vegetative index (RGB, NDVI, and GNDVI).

Section 2 explains previous research related to this research, and Section 3 describes the methodology of our work. Section 4 reports on computational studies of the proposed approach, and Section 5 concludes with findings and work to be done in the future.

2. Related Work

An important field of research is the identifying and modeling of trees using remote sensing data for use in forest ecosystems. Earlier proposed approaches that use airborne and multispectral sensors to detect tree species with very great precision are expensive and thus inappropriate for small-scale of forest management. In this paper, we built a machine visualization scheme for the tree recognition and planning utilizing RGB images captured by the unmanned aerial vehicle (UAV) and then convolutional neural network (CNN). In this scheme, we initially determined the curve from UAV's three-dimensional framework, then instantly sectioned the UAV-RGB images such forest into a many of tree crown particles utilizing color and the three-dimensional details and the slope design, and finally applied the object-based CNN categorization to each of crown image. This scheme classified seven tree classifications, containing multiple tree species, with significantly greater than 90% of accuracy. The directed gradient-weighted category authentication modeling (guided the Grad-CAM) demonstrated that a CNN categorized trees based on their forms and leaf differences, enhancing the system's possibilities for categorizing specific trees to comparable shades in the cost-real way and significant point for forest ecosystems [20].

Coconut is one of India's most profitable crops. We establish an instinctive approach for detecting a number of the coconut trees using an UAV photo in this study. The accessibility of high spatial resolution satellite pictures enables users to create vast volumes of accurate digital imagery of vegetation regions. Today, the projected number of coconut trees may be determined in a short period using high-definition drone pictures at a minimal cost and manpower. The purpose is to develop new ways for determining coconut trees by remotely sensed. To identify coconut trees, deep learning approaches utilizing convolutional neural network (CNN) techniques are applied [21].

Cotton plant populations assessment is critical for making replanted selections within reduced plant densities locations prone to production consequences. Because measuring population of plant in a ground is a labor expensive and prone to the inaccuracy, this research proposes a fresh strategy of image-dependence on vegetation counts based on information from unmanned aircraft systems. Initially proposed image-based algorithms needed a priori knowledge of geometrical or spectral features underlying plant canopy structures, restricting the methods' adaptability in changeable field settings. In this context, a deep learning-based vegetation counting method is accessible to minimize the amount of data collected and eliminate the need for geometrical or statistics data integration. To distinguish, find, and measure cotton plants there at the seedling phase, the object recognition system You Only Look Once on version 3 (YOLOv3) and remote sensing data were used. The suggested method was evaluated using 4 distinct UAS datasets that differed in plant size, overall light, and the background luminance. The optimum crop count was found to have RMSE and R2 values of 0.50 to 0.60 plants per linear meter (numbers of plants across 1m away along with the planted row orientation) and 0.96 to 0.97, respectively. Unless there was a visible difference between cotton growing seasons, the object recognition system trained using varying plant density, ground moisture, and illumination conditions led to a decreased identification error. While cotton plants are usually separated at the germination stage, the preliminary design counting method performed strong with 0–14 individuals per linear meters of the row. This project is expected to give a computerized system for assessing plants emerging in situ utilizing UAS information [22].

Deep learning-dependence on superresolution (SR) is used in this study using Sentinel-2 photos of the Greek island of Zakynthos to determine stress levels in the supercentenarian on olive trees severe water shortages. The goal of this research to track stress levels in the supercentennial of olive trees across period and seasons. The Carotenoid Reflectance Indicator 2 (CRI 2) is derived especially using Sentinel−2 frequencies $B2$ and $B5$. CRI2 mapping with different magnifications of 10 m and 2.5 m was produced. Indeed, pictures of band $B2$ having an original pixel size of 10 m have been superdetermined to 2.5 m. In terms of spectrum $B5$ pictures, they are SR reduced between 20 m and 10 m and then to 2.5 m. Deep-learning-based SR approaches, especially DSen2 and RakSRGAN, have been used to increase pixel density to 10 m and 2.5 m, respectively. Autumn 2019, spring 2019, spring 2020, summer 2019, and summer 2020 are the five seasons evaluated. In the approach, correlations using measurement data may be used to better examine the suggested methodology's efficiency in detecting anxiety levels in extremely old olive trees [23].

The extension and development of olive farming have been related to the enormous development of Verticillium-wilt, a greatest serious fungal issue afflicting olive trees. Current research demonstrates that measures such as using new natural materials (Zeoshell ZF1) and useful microbes can reestablish vitality to afflicted trees. Nevertheless, to be effective, the above procedures necessitate the labeling of trees there in the initial stages of infestation—a work which is not only unreasonable using traditional methods (physical work), and also extremely difficult, because initial phases were impossible to detect with a human eye. The outcomes of My-Olive Grove Coach (MyOGC) scheme are discussed in this work, utilizing multispectral imagery from unmanned aerial vehicles to advance an olive meadows detection method based here on independent and automatic sorting of spectral information utilizing computer vision and the machine learning methods. A program aims to manage and evaluate the condition of olive trees, aid in forecast of Verticillium wilting development, and create decision support mechanism to assist farmers and agronomists [24].

This presents a novel deep learning system for automatic enumeration and localization of the palm trees using aerial photographs utilizing CNN throughout this research. This used two DJI UAVs to capture aerial photographs from two distinct places in Saudi Arabia for such a purpose and created a dataset of approximately 11,000 palm tree examples. Then, we used various contemporary convolutional neural network architectures to recognize palms or other forests (YOLOv3, Faster Regional-CNN, EfficientDet, and YOLOv4), and performed a comprehensive direct comparison in terms of overall correctness and inference performance [25]. YOLOv4 and EfficientDet-D5 provided the best balance of speed and accuracy (upto 99 percent mean regular precision and a 7.4 FPS). Additionally, photogrammetry principles and range adjustments were employed to dynamically infer the geographical position of observed palm palms utilizing geotagged metadata from aerial pictures. This localization approach was evaluated on two distinct kinds of the drones (DJI of Mavic-Pro and Phantom of 4-Pro) and found to have a median positioning precision of 1.6 m. This GPS tracking enables us to authenticate palm trees and estimate the number from a sequence of the drone photographs while handling image overlaps accurately. Furthermore, this novel combination of deep learning object recognition and geolocation could be used for some another entities in UAV images [26].

3. Methodology

3.1. UAV-RGB and Multispectral Images. The research area is positioned in Jouf, Saudi Arabia. The climate is the Mediterranean, with harsh summer temperatures and mild-wet winters. The median annual temperature is 400 mm, and the median yearly temperature is 15°C. Rainfed cereal agriculture and olive groves dominate the flatlands, including small areas of natural plants in the hills. To reduce interference with water supply, olive trees were spaced roughly 6 meters apart. The testing site is located within a 50-hectare olive orchard with 11,000 trees established in 2006. In this research study, we utilized a $560\,\text{m} \times 280\,\text{m}$ flat rectangle comprising around 4000 trees. For the first time, remotely sensed was utilized to automatically detect trees in forestry operations. In recent times, the scientific community has prioritized tree identification and counting in crop fields. There are several approaches available for effectively recognizing and measuring olive trees. Olive tree recognition and identification from UAS photos can be divided based on image analysis approaches.

3.2. Proposed Model. The suggested SwinTU-net design, seen in Figure 2, contains encoding, decoding, and skip connectors. SwinTU-net's fundamental basis is just the Swin Transformer blocks. The encoding creates a sequence of embeddings from the inputs. The olive aerial photos have been separated into four nonoverlapping sections. Each patch already has a feature size of $4 \times 4 \times 3 = 48$ as a result of this partitioning method. The predicted feature size is also turned into an unstructured length utilizing a linear-embedding layers (shown as C). Tokens (modified patched) are passed via numerous layers of Swin Converter and levels of patch combining to produce hierarchical visual features.

The patch combining layer handles downsampling and expanding size, while the Swin Transformer blocks handle feature sentence representations. This designs a synchronous transformer-based decoder influenced by the U-net. The decoding is built from Swin Transformer partitions and opposing patch widening layers [27]. The resulting perspective characteristics are combined using multiresolution characteristics output from an encoder via fully connected layers to compensate for general image loss due to downsampling. A patched increasing element, as opposed to a patched merger layer, is used particularly to promote the growth of features. The patch increasing layer resizes adjacent-dimension relevant features into large feature matrices by upsampling the frequency by two. Lastly, the last patches expansion layer is placed to the extracted features to execute four frequencies up samplings (W and H). The densities map is then created by superimposing a linear projection overlay on the front of these upsampled characteristics.

3.3. Swin Transformer Blocks. The Swin Transformer blocks, unlike typical multihead self-attention (MSA) modules, are focused on the utilization of moved panels. Every Swin Transformer block is depicted in Figure 3, which has a layer norm (LN) level, an MSA module, a residual connector, and a two of MLP levels [28]. The window-dependence on MSA (W-MSA) and shifting window-dependence on MSA (SW-MSA) components were utilized in the 2 consecutive transformers sections. Sequential Swin Transformer units could be formed utilizing a window segmentation approach as follows:

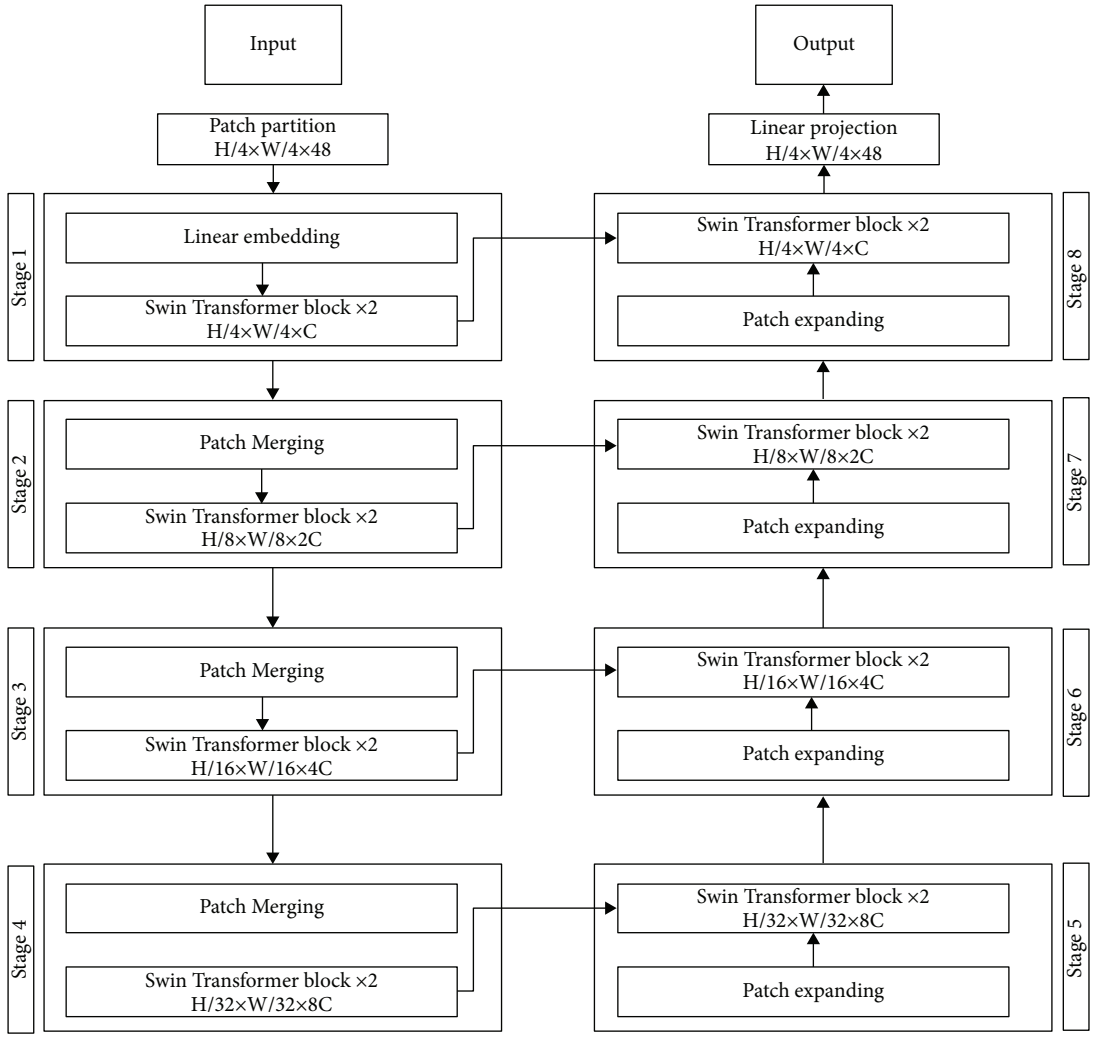


Figure 2: SwinTU-net model structure includes encoding, decoding, and skip connections. The Swin converter serves as the model's foundation.

$$\begin{aligned}
\widehat{a}^{k} &= W - MSA\left(LN\left(a^{k-1}\right)\right),\\
a^{k} &= MLP\left(LN\left(\widehat{a}^{k}\right)\right) + \widehat{a}^{k}, +x^{k-1},\\
\widehat{a}^{k+1} &= SW - MSA\left(LN\left(a^{l}\right)\right) + a^{l},\\
a^{k+1} &= MLP\left(LN\left(\widehat{a}^{k+1}\right)\right) + \widehat{a}^{k+1},
\end{aligned} \tag{1}$$

where $\widehat{a}^{k}$ and a^{l} are its lth block's outputs characteristics, (WMSA and the MLP components, correspondingly) The MSA remains determined into the similar manner as in prior studies.

$$\mathrm{MSA}\,(E,S,V) = \mathrm{softmax}\left(\frac{ES^{T}}{\sqrt{d}} + \mathrm{Bias}\right)V, \tag{2}$$

where $E,S,V \in \mathbb{R}^{N^2\times d}$ are the inquiry, secret, and value matrix, respectively. In a window, d and N^2 represent the inquiry or core size and the patching amount, correspondingly. The bias parameters are extracted from biases matrices $\widehat{B} \in \mathbb{R}^{(2N-1)\times(2N+1)}$.

3.4. Encoding. During an encoding, two successive Swin Transformer units with such a frequency of $(H/4) \times (W/4)$, and 48 variables are performed on the incoming symbols to provide a learning algorithm. The display frequency and features dimensions remained constant. By patch combining levels, the amount of the tokens is lowered as the network grows to generate a hierarchical depiction [29]. The initial patch merging layer combines the characteristics of each set of two neighboring patches. Following that, a linear layering in 4C dimensions is added to the merging features. The outputs dimensions are set to 2C, and the token count is decreased by $2 \times 2 = 4$. Swin Transformer elements were being used to modify the characteristics while retaining the quality at H/8 W/8.

The first phase of patching combining and features modification is referred to as period 2. Because the generator is too profound to be contegrated, the operation is continued twice extra, to various output decisions of $(H/16) \times (W/16)$ and $(H/32) \times (W/32)$, in both, as "period 3" and "period 4." The four phases would be enough to discover a deep functionality visualization because a transistor is too heavy to be cointegrated.

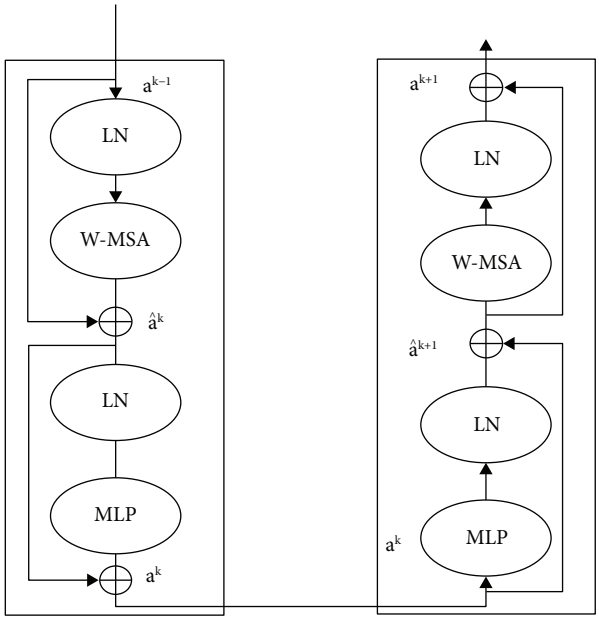


Figure 3: Two sequential blocks of swin transformer.

3.5. Decoding. The Swin Transformer component serves as the foundation both for encoding and decoding. In decoding, unlike encoding, the patching expanding level is employed rather than the patched combining layers to upsample the built characteristics. The patched expanding layer raises the precision of the characteristic map by rearranging nearby dimensions feature mappings and reduces the size of a characteristic by two of its input parameters. Considering the initial patches expansion layering, before upsampling, a quadratic surface is used to double the dimensions of the characteristic from $((H/32) \times (W/32) \times 8C)$ to $((H/32) \times (W/32) \times 16C)$. Use the reorganize procedure to double frequency and reduce the concept's size to a fraction of its initial dimension's $((H/32) \times (W/32) \times 16C \longrightarrow (H/32) \times (H/32) \times 4C)$.

The skip links, such as the U-net, were utilized to feed its encoder's upsampled characteristics via the encoder's multiscale characteristics [30]. This combines the deep and shallow characteristics just to mitigate the spatial and spectral damages produced by downsampling. After such a generative model, the size of the concatenation characteristics is kept the same as the amount of the extract feature characteristics.

3.6. Multispectral Images and UAV RGB. This conducted two of UAV missions at 120 m altitude to collect an RGB of image with ultrahigh longitudinal precision and the multispectral imaging with an actual-high of resolution to examine the impact of deep learning methods on various spectral and spatial recommendations:

(i) A sequoia hyperspectral sensor mounted on a Parrot DiscoPro AG UAV (Al-Jouf, Saudi Arabia) that acquired four spectral features (near-infrared (NIR), red, green, and red-edge). A multispectral image had a pixel density of 13 cm/pixel [31]. The vegetative indicators described in the introduction were then calculated: the normalization differential vegetative indices (NDVI) equation (3) and the greenish normalization differences vegetative indices (GNDVI):

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{red}}{\mathrm{NIR} + \mathrm{red}}, \tag{3}$$

$$\mathrm{GNDVI} = \frac{\mathrm{NIR} - \mathrm{green}}{\mathrm{NIR} + \mathrm{green}}. \tag{4}$$

(ii) This used the DJI-Phantom 4 UAV's native RGB Hasselblad 20-megapixel sensor to get higher spatial precision. The RGB image had a spatial precision of 3 cm/pixels. These RGB pictures are then transformed to the 13 cm/pixel resolution by geographical pooling to be evaluated. Granada Drone S.L. supplied the images used in both missions.

Weather parameters (sunshine and bright day) and the shooting period before nightfall are the specific criteria for current data gathering.

3.7. Construction of the Dataset. To create four subgroups of data to create a dataset to segment olive tree tops and tree darkness, which would allow us to examine the impact of reducing spatial resolution and acquiring spectral information. (i) RGB−3, (ii) RGB−13, (iii) NDVI−13, and (d) GNDVI−13, where 3 and 13 represent the picture pixel density in cm/pixel [32]. This produced 150 picture patches including 2400 trees for every batch of data, with 120 images (80% of a database) utilized for testing its system and 30 photos (20% of the dataset) utilized for validating the system here on olive tree crown category (Table 1) as well as olive tree shadows category (Table 2). Every image of patch comprised one to the eight of olive trees, complete using treetops and tree shadows.

Figure 4 depicts the overall process of creating a large dataset. Utilizing Pix4D 4.0, the initial UAV photos were integrated into an orthophoto. QGIS 2.14.21 is utilized to reduce a pixel density of an RGB−3 cm/pixel to an RGB−13 cm/pixel, as well as to calculate the NDVI and GNDI indexes. The patching was created in ENVI Classic and converted from tiff to .jpg formats (greatest appropriate arrangement for the training of deep learning techniques). The number of pixels is artificially raised to 13 cm/pixel throughout the. Tiff to .jpg transformation is done by the QGIS 2.14.21 application. VGG Image Annotator 1.0.6, standalone program for the manual feature extraction, was used to create and annotate tree crown and the tree shadow sections within every image patching. This semantic segmentation task's annotating procedure was entirely manual. In other words, the observer drew a polygon around every olive tree crown and a second-round tree shadow example. The known values developed using the VGG autocomplete feature was again saved in JSON style.

3.8. Mask Regional-CNN. Instance segmentation is the process of finding and separating all the pixels which make up a unique olive tree crown in UAV photos. This is

Table 1: Patches of images and sections in four subgroups of a crown olive tree.

A subset of tree crown	Train images	Train segments	Test images	Test images	Overall images	Overall segments
RGB–3	150	500	40	130	145	650
RGB–13	150	500	40	130	145	650
NDVI–13	150	500	40	130	145	650
GNDVI–13	150	500	40	130	145	650
Overall	600	2000	160	520	580	2600

Table 2: Patches of images and sections in the four subgroups of an olive tree shadows.

A subset of tree crown	Train images	Train segments	Test images	Test images	Overall images	Overall segments
RGB–3	150	500	40	130	145	650
RGB–13	150	500	40	130	145	650
NDVI–13	150	500	40	130	145	650
GNDVI–13	150	500	40	130	145	650
Overall	600	2000	160	520	580	2600

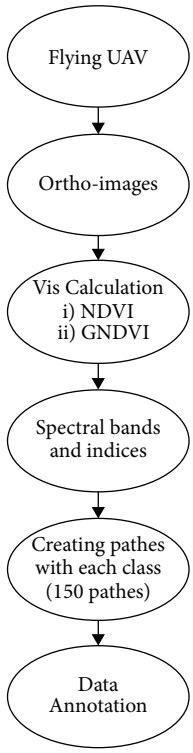


Figure 4: Preparing process of images.

among the most difficult challenges in computer vision. This employed the contemporary mask regional-CNN system (regions using convolutional neural networks) in this work that is an extension of the quicker Regional-CNN classification models [33]. Mask regional-CNN examines the contribution image and generates a three-outputs for every item class: (i) a category label indicating the object-class description, (ii) a boundary box delimiting every object class, and (iii) a mask delimiting the pixels which comprise every object class. Mask regional-CNN creates a binary mask (using values 0 and 1) for every olive tree in this study, whereas a value of 1 denotes a pixel of olive tree and value 0 denotes a pixel of nonolive tree.

To extract features, mask regional-CNN is dependent on the classification algorithm. ResNet50 CNN was used in this research to recover gradually higher-level characteristics from the weakest to deeper layer levels [34]. To enhance the categorization model's generalization performance, we evaluated the influence of data preprocessing, which contains growing the dimension of the data by implementing simple alterations including cropping (reducing columns/rows of pixel value just at edges of images), scalability, inversion, transcription, horizontal, and vertical compressive. This utilized transfer learning rather than building mask regional-CNN (inspired on ResNet50) at the start of using the database. Transfer learning involves first implementing the model's parameters using pretrained weights on the COCO-database and reskilling its system on their data source. Fine relates to retraining the last decade using a limited dataset [35].

3.9. Performance Evaluation of CNN. The $F1$-score measure was used to assess the effectiveness of its trained of mask regional-CNN upon this OCTS-database within the objective of an olive tree crown and shadow example of segmentation. It is calculated as the harmonics average of recall and precision. Mask regional-CNN generates three output results: a boundary area, masks, and probability in the projected class. The interaction overlap unions (IoU) or Jaccard ratio has been used to judge whether a forecast was right [36]. It is determined by the intersection of the expected and actual boundary boxes reduced by the unity. A projection is a true positive (T_{positive}) if the IoU is greater than 50%, and a false-positive (F_{positive}) if the IoU is minimum than 50%. The following is how IoU is measured:

$$\text{IoU} = \frac{\text{overlap area}}{\text{union area}}. \tag{5}$$

A threshold quantity of 0.5 is commonly employed since it produces high indications of the score. Accuracy (equation (6)) and recall (equation (7)) are determined as follows:

$$\text{accuracy} = \frac{T_{\text{positive}}}{T_{\text{positive}} + F_{\text{positive}}} = \frac{T_{\text{positive}}}{\#\text{ground\_truths}'}, \tag{6}$$

$$\text{recall} = \frac{\text{T}_{\text{positive}}}{T_{\text{positive}} + F_{\text{negative}}} = \frac{T_{\text{positive}}}{\#\text{Predictions}'}. \tag{7}$$

Accuracy is the proportion of properly identified labels, and recall is a component of effective label retrieval. The F_1 – measure is designed as the weighting factor of recall and precision (equation (8)). It considers includes F_{negative} and F_{negative} to determine a model's overall accuracy:

$$F_1 = \frac{2 \times \text{recall} \times \text{precision}}{\text{recall} + \text{precision}}. \tag{8}$$

3.10. Execution Details. In the development, a well-known PyTorch library was employed [37]. The suggested method then is developed and validated using an NVIDIAGeForceRTX2060, GPU. Arbitrary data modifications including rotation, scale, and turning are employed before training to increase data variation. The input image is enlarged to 224 × 224, which allowed the GPU to run out of memory during training. The scores pretrained model on the ImageNet-1K is utilized to set the set of parameters [38]. During the training phase, the model is tuned for training algorithm utilizing a well SGD algorithm, using the velocity of 0.9 and weight loss of $1e^{-4}$.

4. Results and Discussion

This section outlines the planned olive tree datasets that will be used to test the algorithm. It also discusses the measures that were utilized to evaluate the effectiveness of the suggested model. Image segmentation findings were based on RGB and vegetative indexes. The findings of the tree biovolume computations are shown.

4.1. Data Collection. According to 2019 figures from the Department of Atmosphere and Waters Department in Al-Jouf, the Al-Jouf area is home to 30 million trees, the majority of that is olive trees (18 million of trees) that provide 10 thousand tons of the oil each year. As a result, collection contains 250 photos collected around Al-Jouf, KSA region, utilizing the Satellites Pro. Satellite imagery and maps are available as with most nations and towns across the world with Satellites Pro. The particular section was photographed in RGB pictures with a resolution of 512 × 512 and a bit depth of 32. To save strain and speed up annotations, the olive trees were marked using centroids. The initial step is to mark the olive photos with bounding boxes that enclose the olive trees [39]. $\{(a_i, b_i), i = 1, 2, 3, 4\}$ were the four vertices of bounding boxes. In the next stage, use the given equation (9) to get the center of every box also as the center location:

$$(a, b) = \left(\frac{1}{4}\sum_{i=1}^{4} a_i, \frac{1}{4}\sum_{i=1}^{4} b_i\right). \tag{9}$$

4.2. Performance Metrics. The effectiveness of their model is determined by utilizing some performance criteria to compare different strategies on various datasets:

(i) Total accuracy (TA)

TA is the proportion of successfully predicted olive trees out of an overall amount of the olive trees. Between the designated trees, it represents the number of trees that were accurately recognized within the ground-truth dataset [40]. The respective equation is used to calculate total accuracy:

$$\text{TA} = \frac{\text{number of estimated olive trees}}{\text{number of actual olive trees}} \times 100. \tag{10}$$

(ii) Omission failure rate (OFR)

OFR is the proportion of a positive assessment individuals that were mistaken for negative test individuals. In other terms, OFR represents the proportion of times suggested algorithm flops to an identify olive trees as being. OFR is computed quantitatively utilizing the formula

$$\text{OFR} = \frac{\text{number of omitted olive trees}}{\text{number of actual olive trees}} \times 100. \tag{11}$$

(iii) Commission failure rate (CFR)

CFR is described by an occurrence of the negative specimens which were mislabeled as positive. It occurs when the output contains nonolive trees. CFR is determined mathematically utilizing the formula

$$\text{CFR} = \frac{\text{number of false olive trees identified}}{\text{number of actual olive trees}} \times 100. \tag{12}$$

(iv) Estimation failure (EF)

It corresponds to a discrepancy among the number of things detected and also a number of objects yet to be recognized. It is the ratio between the actual and predicted amount of olive trees within samples split by an actual amount of olive trees within a suggested model. The following formula is used to compute the EE mathematical model:

$$\text{EF} = \frac{\text{number of estimated olive trees} - \text{number of actual olive trees}}{\text{number of actual olive trees}} \times 100. \tag{13}$$

Table 3: Mask regional-CNN segmentation effectiveness model for olive tree crown.

Testing subgroups	$T_{positive}$	$F_{positive}$	$F_{negative}$	Accuracy	F_1 measure	Recall
(a) Models were trained on every subgroup without the use of additional data						
Red green blue−3	130	1	0	1	1	1
Red green blue−13	120	2	3	1	0.9915	0.9959
Normalization differential vegetative indices−13	115	0	5	0.9835	0.9501	0.9661
Greenish normalization differential vegetative indices−13	120	1	12	1.00	0.9167	0.9656
(b) Models were trained on every subgroup with the use of additional data						
Red green blue−3	130	0	0	1	1	1
Red green blue−13	114	0	5	1	0.9835	0.9917
Normalization differential vegetative indices−13	113	15	5	0.9009	0.9835	0.9403
Greenish normalization differential vegetative indices−13	114	13	5	0.9077	0.9835	0.9538
(c) Models were trained upon that merging of all 13 cm/pixel picture subgroups and with training data						
Red green blue−3	120	0	0	1	0.9915	0.9957
Normalization differential vegetative indices−13	115	0	5	1	0.9667	0.9831
Greenish normalization differential vegetative indices−13	110	0	10	1	0.9084	0.9521

Table 4: Mask regional-CNN segmentation effectiveness model for olive tree shadow.

Testing subgroups	$T_{positive}$	$F_{positive}$	$F_{negative}$	Accuracy	F_1 measure	Recall
(d) Models were trained on every subgroup with the use of additional data						
Red green blue−3	130	0	0	1.0000	1.0000	1.00
*(e) Models were trained upon that merging of overall*13 cm/pixel picture subgroups and with training data						
Red green blue−3	120	0	0	1	0.9915	0.9957
Normalization differential vegetative indices−13	115	0	8	1	0.9261	0.9811
Greenish normalization differential vegetative indices−13	110	0	4	1	0.9752	0.9572

4.3. Tree Shadow and Tree Crown Segmentation Using RGB and Vegetative Index Images. Table 3 shows the effectiveness among all mask regional-CNN systems here on relevant test subgroups of its information in terms of accuracy, recall, and *F*1-measure for the tree crowns, and Table 4 shows the effectiveness of tree shadows.

As demonstrated in Table 3, several training and testing mask regional-CNN networks with tree crown segmentation had high F1 scores, exceeding 94 percent throughout all dataset groups. The F1 score was unaffected by data augmentation. The RGB subset produced the best results (F1 = 100 percent) at a spatial and temporal precision of 3 cm/pixel. In a RGB data group, increasing the display resolution from 3 to 13 cm/pixel reduced F1 by 0.42 percent without a data augmentation (plan A) and 0.86 percent with a data preprocessing (plan B). At the 13 – cm/pixel quality, the 3-band RGB images consistently outperformed the single-band NDVI or GNDVI views in terms of F1.

Nevertheless, the model was trained using data synthesis (type C, which is trained using RGB, NDVI, and GNDVI visuals simultaneously) and demonstrated similar or higher F1 than that of the trained models without data fusion (plans A and B, either with or without data preprocessing). Data aggregation boosted the F1 score by 1.76 percent for the NDVI-13 database, whereas data augmentation reduced by it 2.68 percent, as associated with training only using a NDVI-13 input data and without data preprocessing. The F_1 measure obtained a GNDVI database which is comparable to or higher than that obtained on the NDVI database.

As demonstrated in Table 4, all designed and evaluated mask regional-CNN models for the tree shadow identification have a high F_1 measure—more than 96 percent. The classifier (type D) with the greatest F_1 measure was designed and evaluated on the RGB visuals at 3 cm/pixel. Nevertheless, the information fusion algorithm (type *E*, which was generated on RGB, NDVI, and GNDVI images simultaneously) also demonstrated a really strong F_1 measure on RGB-13 cm/pixel imagery (99.58 percent). For tree shadow identification, the information fusion approach (type E) showed improvement here on RGB−13 (99.58 percent) and GNDVI−13 (98.73 percent) datasets than with a NDVI−13 (96.10 percent) database.

Table 5 illustrates six olive trees that could be surveyed in the environment for open canopy area estimation combining tree circumference and tree heights fragmentation using the mask regional-CNN training set. Total precision is 94.51 percent for RGB-3, 75.61 percent for RGB-13, 82.58 percent for NDVI-13, and 77.38 percent for GNDVI-13. The system gave training and testing on the RGB images around 3 cm/pixel and had a highest total reliability for the estimating biovolume. The information fusion algorithm functioned well and achieved greater accuracy here on NDVI subgroups than with the GNDVI or RGB subgroups at 13 cm/pixel level.

4.4. Total Evaluation. The total prediction error for testing is 0.94 percent after implementing the proposed model. According to Table 6, for a 100% mixture of olive and nonolive trees among other items, around 0.97 percent of nonolive data is misinterpreted as an olive and 1.2 percent of the olive data were misdiagnosed as nonolive. The suggested

Table 5: Characteristics average for olive trees, where Pt represents perimeter, Ht represents height, vol represents volume, and Lt represents length.

	Ground table			Types of tree crowns and tree shadows Testing using RGB–3				Types of tree crowns and tree shadows Testing using RGB–13				Types of tree crowns and tree shadows Tested using NVDI–13				Types of tree crowns and tree shadows Testing using GNDVI–13			
Sl. no	Pt	Ht	Vol	Pt	Lt	Ht	Vol	Pt	Lt	Ht	Vol	Pt	Lt	Ht	Vol	Pt	Lt	Ht	Vol
1	6.2	2.3	6.32	6.5	4.4	2.5	6.72	7.2	4.2	2.4	7.35	7.8	3.5	1.9	6	9.5	3.7	1.9	8.98
2	6.4	2.7	7.09	6.4	4.7	2.8	7.42	8	4.4	2.5	9.87	8.3	4.6	2.4	9.16	8.5	4.6	2.3	9.17
3	8.4	2	13.72	8.9	8.7	2.7	13.01	10	5.9	3.4	22.23	10	5.3	2.5	16.5	10.7	5.3	2.7	18.43
4	8.2	2	15.38	8.4	8.6	2.8	14.12	8.8	5.2	2.8	14.35	9.2	4.9	2.7	12.25	10.7	4.9	2.5	16.67
5	8.3	2.8	12.54	8.2	8.2	3.2	13.42	8.2	5.8	3.5	14.85	8.5	4.6	2.3	9.64	9.3	4.5	2.1	11.57
6	8.9	3	16.05	8.5	8.5	3.4	17.04	8.4	5.2	2.8	`4.75	9.3	5	2.6	13.22	10.3	5	2.5	15.94

Table 6: Total evaluation.

Number of images	250
Number of trees	73285
Identifies trees	72596
EF (percent)	0.95
CFG (percent)	708
OFG (percent)	913

Table 7: Proposed techniques comparison.

	Reticular matching	Multilevel thresholding	Detection utilizing red-bands	Enhanced *K*-means techniques	Proposed techniques
Database	QuickBird	SIGPAC viewer	SIGPAC viewer	SIGPAC viewer	UAV and satellites images pro
Spectrum	Grey-scale	Grey-scale	Red-band	RGB	RGB
Number of images	Not available	96	60	110	250
			Evaluation metrics		
TA	98.0%	96.0%	Not available	98.5%	98.4%
CFR (percent)	5	5	1.3	5	0.98
OFR (percent)	7	3	5	1	1.3
EF (percent)	1.25	1.2	1.28	0.98	0.95

dataset was tested, and the results indicated an overall recognition with a 0.94 percent prediction error.

An olive image, as well as ground truth and recognition findings, illustrates a combination of olive tree patterns with significant spacing among them and that those were densely established. The suggested technique appropriately defined nearly all of an olive tree, although it underestimated the number of immature and densely implanted trees.

4.5. Comparative Evaluation. A suggested model's findings were evaluated with those of established olive identification and enumeration algorithms. The specifications of the database, the quantity of images generated, a spectra reflecting the dimension of manufactured information, and the assessed effectiveness were all utilized in the comparisons. The outcomes of an evaluation of a proposed system to existing methodologies are shown in Table 7.

The suggested approach was validated on a large database and yielded good accuracy, as seen in Figure 4, suggesting that the methodology is efficient and robust. By reliably detecting and estimating olive trees, the suggested approach resolved weaknesses in existing approaches. This innovative algorithm for identifying and identifying olive trees was verified using RGB photos with an effective level of 98.4 percent, outperforming previous work.

The suggested framework exhibited the lowest total prediction error of 0.95 percent among the available strategies when tested on a huge database of olive trees as well as other surface items. It is worth mentioning that the suggested dataset contains 250 photos of olive trees as well as other things as shown in Figure 5.

There seems to be currently a plethora of low-cost RGB and the multispectral sensors which could be installed on the multirotor and secure-wing UAVs, but these images could

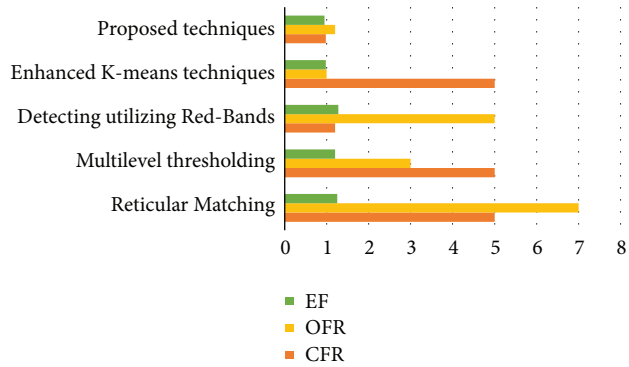


Figure 5: Performance evaluation.

be mechanically managed using CNN techniques for such a reason. On only side, the RGB security sensors on such a multirotor UAV may seizure considerably higher determination satellite data, which improves CNN prediction performance, but covers fewer regions (because of power limits) resulting in less costly visuals for each acre. Multispectral sensor arrays on fixed-wing drone attacks, on either side, could indeed encapsulate relatively coarse resolution satellite visuals over large regions, lowering the cost per acre while also integrating plant radiance in the near-infrared and red-edge, that also either someone to photosynthesis rate than only RGB. Merging both data sets could combine the benefits of both systems, such as increasing CNN precision, lowering cost per hectare, and including photosynthesis activities data.

The findings reveal whether CNN systems trained and tested at rather greater resolution (namely, RGB with 3 cm/pixel) achieved significantly higher accuracy (approximately 0.42 percent higher) than CNN algorithms were trained and

tested at coarse grain quality (namely, RGB at 13cm/pixel). Extraspecifically, findings show which training CNN concepts on a merging among all a RGB, NDVI, and GNDVI subsamples of the images at coarse-grained pixel density (namely, 13 cm/pixel precision) outcomes in the common model with very rising accuracies (also larger than 95.0 percent and 96.0 percent for the tree crown and the tree shadow, combined) regardless of the type of image selected in the diagnostics (RGB, NDVI, or GNDVI). This extension opens the door to utilizing secure-wing multispectral or RGB images finished large regions for tree volume management at a cheaper cost per hectare, with broad consequences in accuracy agriculture, accuracy forestry, and accuracy restoration.

It is worth noting that feature extraction when applied to the mask regional-CNN model had no effect on the outcomes and even appeared to slightly lower the $F1$ score. Simulations training on the RGB image database produced a better effect across databases with such a frequency of 13 cm/pixel, indicating that the approach performs better on three-band images rather than single-band images like using both NDVI and GNDVI vegetative indexes. It can be attributed to the reality that now the supplementation data provided us with some items that resembled the plants that grow beneath and amid an olive tree, resulting in a false-positive and a fall in the total $F1$. Considering this, this concept design demonstrates how a technique of a pixel segmentation utilizing deep-CNNs may be employed efficiently in forest and agricultural situations on UAV images.

5. Conclusion

Finally, an effective deep learning approach (SwinTU-net) for identifying and counting the olive trees using the satellite data and UAV was developed. SwinTU-net is a system similar to U-net that features encoding, decoding, and skip links. The SwinTU-net extended a Swin Transformer block to acquire locally and globally semantic features rather than utilizing the convolution function. Furthermore, we began by generating a large-scale of olive database for a deep learning experiments and development. The fact that the CNN segmentation findings of the tree crown and the tree shadow may be utilized to estimated biovolume in numerous trees encourages additional research in this area to enhance the organization. The collection is made up of 250 RGB photos gathered in Al-Jouf, Saudi Arabia. Experimental research on the program found that the SwinTU-net model beats similar studies in terms of total identification, having a 0.95 percent prediction failure. The estimated values match the ground measures of the sample trees with a margin of error of 5.4 percent. However, there are certain disadvantages, including the difficulties in distinguishing olive trees that are near other trees. More field observations, estimates, and tests are required to have a better grasp of the possibilities of this method, which will be the subject of future research. As a result, plans include expanding the suggested dataset with more photos from diverse sources and improving the developed framework. It is proposed to evaluate trained CNN utilizing medium quality satellite data, which is of particular importance for utilizing possible outcomes over wide positions and also estimating yields and earnings for olive trees.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors extend their appreciation to the Deputyship for Research Innovation, Ministry of Education in Saudi Arabia, for funding this research through project number 400682337.

References

[1] A. Capolupo, L. Kooistra, C. Berendonk, L. Boccia, and J. Suomalainen, "Estimating plant traits of grasslands from UAV-acquired hyperspectral images: a comparison of statistical approaches," *ISPRS International Journal of Geo-Information*, vol. 4, pp. 2792–2820, Dec. 2015.

[2] E. Salamí, A. Gallardo, G. Skorobogatov, and C. Barrado, "On-the-Fly olive trees counting using a UAS and cloud services," *Remote Sensing*, vol. 11, p. 316, Feb. 2019.

[3] Q. Vu, M. Raković, V. Delic, and A. Ronzhin, "Trends in Development of UAV-UGV Cooperation Approaches in Precision Agriculture," *Interactive Collaborative Robotics*, vol. 11097, pp. 213–221, 2018.

[4] H. Song, C. Yang, J. Zhang, W. C. Hoffmann, D. He, and J. A. Thomasson, "Comparison of mosaicking techniques for airborne images from consumer-grade cameras," *Journal of Applied Remote Sensing*, vol. 10, p. 016030, Mar. 2016.

[5] S. Khanal, J. Fulton, and S. Shearer, "An overview of current and potential applications of thermal remote sensing in precision agriculture," *Computers and Electronics in Agriculture*, vol. 139, pp. 22–32, Jun. 2017.

[6] J. Fernández-Guisuraga, E. Sanz-Ablanedo, S. Suárez-Seoane, and L. Calvo, ""Using unmanned aerial vehicles in postfire vegetation survey campaigns through large and heterogeneous areas: opportunities and challenges," *sensors*, vol. 2, p. 586, Feb. 2018.

[7] I. Ahmed, "Taloba "an artificial neural network mechanism for optimizing the water treatment process and desalination process"," *Alexandria Engineering Journal 61*, vol. 12, pp. 9287–9295, 2022.

[8] M Elloumi, M. A. Ahmad, A. H. Samak, A. M. Al-Sharafi, D. Kihara, and A. I. Taloba, Error correction algorithms in non-null aspheric testing next generation sequencing data," *Alexandria Engineering Journal*, vol. 61, pp. 9819–9829, 2022, Mourad.

[9] V. Lukas, J. Novak, L. Neudert et al., "THE combination of UAV survey and landsat imagery for monitoring of crop vigor in precision agriculture," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 953–957, Jun. 2016.

[10] D. A. García Cárdenas, J. A. Ramón Valencia, D. F. Alzate Velásquez, and J. R. Palacios Gonzalez, *"Dynamics of the Indices NDVI and GNDVI in a Rice Growing in Its Reproduction Phase from Multi-spectral Aerial Images Taken by*

Drones in *Proceedings of the Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change II*, vol. 893, pp. 106–119, Cali, Colombiam, November2019.

[11] M. Basso, D. Stocchero, R. Ventura Bayan Henriques et al., "Proposal for an embedded system Architecture using a GNDVI algorithm to support UAV-based agrochemical spraying," *Sensors*, vol. 19, p. 5397, 2019.

[12] S. S. I. Ismail, R. F. Mansour, R. M. Abd El-Aziz, A. I. Taloba, and A. I. Taloba, "Efficient E-mail spam detection strategy using genetic decision tree processing with NLP features," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–16, 2022.

[13] W. Liu, D. Anguelov, D. Erhan et al., *"SSD: Single Shot MultiBox Detector,* in *Proceedings of the Computer Vision – ECCV 2016*, pp. 21–37, 2016.

[14] S. Ren, K. He, R. Girshick, and J. Sun, *"Faster r-cnn: Towards real-time object detection with region proposal networks, " Advances in neural information processing systems*, vol. 28, 2015.

[15] G. Gao, Q. Liu, Z. Hu, L. Li, Q. Wen, and Y. Wang, "PSGCNet: a pyramidal scale and global context guided network for dense object counting in remote sensing images," 2022, http://arxiv.org/abs/2012.03597.

[16] A. I. Taloba, R. M. Abd El-Aziz, H. M. Alshanbari, A. A. H. El-Bagoury, and A.-A. H. El-Bagoury, "Estimation and prediction of hospitalization and medical care costs using regression in machine learning," *Journal of Healthcare Engineering*, vol. 2022, pp. 1–10, 2022.

[17] W. Zhang, P. Tang, and L. Zhao, "Remote sensing image scene classification using CNN-CapsNet," *Remote Sensing*, vol. 11, p. 494, Feb. 2019.

[18] Q. Zhang, R. Qin, X. Huang, Y. Fang, and L. Liu, "Classification of ultra-high resolution orthophotos combined with DSM using a dual morphological top hat profile," *Remote Sensing*, vol. 7, pp. 16422–16440, Dec. 2015.

[19] A. Safonova, S. Tabik, D. Alcaraz-Segura, A. Rubtsov, Y. Maglinets, and F. Herrera, ""Detection of fir trees (abies sibirica) damaged by the bark beetle in unmanned aerial vehicle images with deep learning, " remote sensing, vol," *11, no*, vol. 6, p. 643, Mar. 2019.

[20] M. Onishi and T. Ise, "Explainable identification and mapping of trees using UAV RGB image and deep learning," *Scientific Reports*, vol. 11, p. 903, Dec. 2021.

[21] A. P, R. Hebbar, H. Sounder, N. K, and P. V. Vinod, ""CNN based technique for automatic tree counting using very high resolution data," in *Proceedings of the 2018 international conference on design innovations for 3Cs compute communicate control (ICDI3C)*, pp. 127–129, Bangalore, Apr. 2018.

[22] S. Oh, A. Chang, A. Ashapure et al., "Plant counting of cotton from UAS imagery using deep learning-based object detection framework," *Remote Sensing*, vol. 12, p. 2981, Sep. 2020.

[23] A. Panagiotopoulou, E. Charou, K. Poirazidis et al., "Deep-Learning based super-resolution of sentinel-2 images for monitoring supercentenarian olive trees," *Volos Greece*, pp. 143–148, 2021.

[24] K. Blekos, A. Tsakas, C. Xouris et al., "Analysis, modeling and multi-spectral sensing for the predictive management of Verticillium wilt in olive groves," *Journal of Sensor and Actuator Networks*, vol. 10, p. 15, 2021.

[25] A. I. Taloba, M. R. Riad, and T. H. A. Soliman, "Developing an Efficient Spectral Clustering Algorithm on Large Scale Graphs in spark," in *Proceedings of the 2017 eighth international conference on intelligent computing and information systems (ICICIS)*, pp. 292–298, IEEE, Cairo, Egypt, December2017.

[26] A. Ammar, A. Koubaa, and B. Benjdira, "Deep-learning-based automated palm tree counting and geolocation in large farms from aerial geotagged images," *Agronomy*, vol. 11, p. 1458, Jul. 2021.

[27] Z. Xie, Y. Lin, Z. Yao et al., ""Self-Supervised learning with Swin transformers," 2021, http://arxiv.org/abs/2105.04553.

[28] Z. Liu, Y. Lin, Y. Cao et al., ""Swin transformer: hierarchical vision transformer using shifted windows," in proceedings of the IEEE/CVF international conference on computer vision," pp. 10012–10022, 2021.

[29] H. Touvron, M. Cord, A. Sablayrolles, G. Synnaeve, and H. Jégou, ""Going deeper with image transformers," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 32–42.

[30] O. Ronneberger, P. Fischer, and T. Brox, *"U-Net: Convolutional Networks for Biomedical Image Segmentation,* in *Proceedings of the Medical Image Computing and Computer-Assisted Intervention – MICCAI*, pp. 234–241, Munich, Germany, October2015.

[31] A. A. Gitelson, Y. J. Kaufman, and M. N. Merzlyak, "Use of a green channel in remote sensing of global vegetation from EOS-MODIS," *Remote Sensing of Environment*, vol. 58, pp. 289–298, Dec. 1996.

[32] R. Alanazi, A. Elhadad, A. I. Taloba, and R. M. Abd El-Aziz, "A large-scale dataset and deep learning model for detecting and counting olive trees in satellite imagery," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–8, 2022.

[33] E. Guirado, J. Blanco-Sacristan, E. Rodriguez-Caballero et al., "Mask R-CNN and OBIA fusion improves the segmentation of scattered vegetation in very high-resolution optical sensors," *Sensors*, vol. 21, p. 320, Jan. 2021.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[35] S. Tabik, D. Peralta, A. Herrera-Poyatos, and F. Herrera, "A snapshot of image pre-processing for convolutional neural networks: case study of MNIST," *IJCIS*, vol. 1, p. 555, 2017.

[36] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, ""Generalized intersection over union: a metric and a loss for bounding box regression," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition," pp. 658–666, 2019.

[37] A. Paszke, S. Gross, F. Massa et al., ""Pytorch: An imperative style," *high-performance deep learning library Advances in neural information processing systems*, vol. 32, 2019.

[38] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and Li Fei-Fei, "ImageNet: a large-scale hierarchical image database," in *Proceedings of the 2009 ieee conference on computer vision and pattern recognition*, pp. 248–255, Miami, FL U S A, JunE 2009.

[39] G. Xue, F. Lin, S. Li, and H. Liu, "Adaptive dynamic surface control for finite-time tracking of uncertain nonlinear systems with dead-zone inputs and actuator faults," *Automation and Systems*, 2021.

[40] S. Ha, L. Chen, H. Liu, and S. Zhang, "Command filtered adaptive fuzzy control of fractional-order nonlinear systems," *European Journal of Control*, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 1200611, 10 pages
https://doi.org/10.1155/2022/1200611

Research Article

Ridge Regression Method and Bayesian Estimators under Composite LINEX Loss Function to Estimate the Shape Parameter in Lomax Distribution

Mansour F. Yassen,[1] **Fuad S. Al-Duais**,[1] **and Mohammed M. A. Almazah**[2]

[1]*Mathematics Department, College of Humanities and Science in Al Aflaj, Prince Sattam Bin Abdulaziz University, Al-Kharj, Al Aflaj, Saudi Arabia*
[2]*Department of Mathematics, College of Sciences and Arts (Muhyil), King Khalid University, Muhyil 61421, Saudi Arabia*

Correspondence should be addressed to Fuad S. Al-Duais; f.alduais@psau.edu.sa

Received 18 June 2022; Accepted 3 August 2022; Published 29 August 2022

Academic Editor: Heng Liu

In this paper, the Ridge Regression method is employed to estimate the shape parameter of the Lomax distribution (LD). In addition to that, the approaches of both classical and Bayesian are considered with several loss functions as a squared error (SELF), Linear Exponential (LLF), and Composite Linear Exponential (CLLF). As far as Bayesian estimators are concerned, informative and noninformative priors are used to estimate the shape parameter. To examine the performance of the Ridge Regression method, we compared it with classical estimators which included Maximum Likelihood, Ordinary Least Squares, Uniformly Minimum Variance Unbiased Estimator, and Median Method as well as Bayesian estimators. Monte Carlo simulation compares these estimators with respect to the Mean Square Error criteria (MSE's). The result of the simulation mentioned that the Ridge Regression method is promising and can be used in a real environment. where it revealed better performance the than Ordinary Least Squares method for estimating shape parameter.

1. Introduction

Ridge Regression is a popular parameter estimation method for analyzing multiple regression data which has multicollinearity. When Multicollinearity happens, Least squares estimates are unbiased, but their variances are large. As a result, the estimator of the Ordinary Least Squares Method (OLS) becomes far from the true value. Ridge Regression decreases the standard errors when a degree of bias is added to the Regression Estimates. Several authors have addressed Ridge Regression as in [1–10].

The Lomax distribution (LD) is a proposed distribution of the Pareto distribution of type II; it was used to obtain a good model for biomedical problems. Also, it is an important model for modeling failure times. The Lomax distribution was used as a stochastic model with a decreasing failure rate for the operating times of the electronic vehicles under study. It was also used in studies related to income and studies related to the size of cities. As well as being a useful model in studying queuing theory and in analyzing data related to biostatistics.

Many theoretical and statistician's studies have given great interest to estimating the parameter and survival analysis of LD.

Al-Noor and Alwan [11] compared the Nonbayesian, Bayesian, and Empirical Bayes estimate for the parameters of the LD by considering the symmetric and asymmetric loss functions. Al-Duais and Hmood [12] compared the Bayesian estimators and Classical estimators to estimate the parameter and survival analysis depending on record values and by considering the SELF, LLF, and WLLF. Ellah [13] estimated the parameter, reliability, and hazard function of the LD by applying the Bayesian estimators and Classical estimators based on record values by considering the SELF and LLF. Asl

et al. [14] applied the Bayesian estimators and Classical estimators of prediction on unidentified parameters of an LD based on a progressively type-I hybrid censoring scheme. Mohie El-Din et al. [15] studied the Classical estimations and Bayesian estimations of the LD based on progressively type-II censored samples and by considering symmetric (SELF) and asymmetric (LLF and GELF), Okasha [16] estimated the parameters, and survival analysis of the LD by applying the E-Bayesian estimators and Bayesian estimators under type-II censored data and by regarding the balanced squared error loss function. Liu and Zhang [17] studied the Bayesian and E-Bayesian estimations of the LD Based on the Generalized Type-I Hybrid Censoring Scheme. to estimate the unknown parameter of LD and by considering SELF and LLF to estimate the parameter and reliability function. Al-Bossly [18] developed a compound LINEX loss function (CLLF) to estimate the shape parameter of the Lomax distribution utilizing the E-Bayes and Bayes estimation methods for the distributional parameters of the LD.

In the current study, Ridge regression was employed to estimate the shape parameter of LD and compare it with the classical estimators which included Maximum Likelihood, Ordinary Least Squares, Uniformly Minimum Variance Unbiased Estimator, and Median Method as well as Bayesian estimators. The uniqueness of this work comes from the fact that, to date, no attempt has been made to estimate the shape parameter of the LD using the method of Ridge regression.

The pdf of LD is given as follows [19]:

$$f(x;\vartheta,\delta)=\begin{cases}\dfrac{\vartheta}{\delta}\left(1+\dfrac{x}{\delta}\right)^{-(\vartheta+1)}; & x\geq 0;\ \vartheta,\delta>0,\\ 0; & o.w,\end{cases} \tag{1}$$

where x is a random variable, and $\delta>0$, $\vartheta>0$ are the scale and shape parameters, respectively.

The CDF and reliability function $R(t)$ of (1) are given by the following equation:

$$F(x;\vartheta,\delta)=1-\left(1+\frac{x}{\delta}\right)^{-\vartheta}; \quad x\geq 0;\ \vartheta,\delta>0, \tag{2}$$

$$R(t)=\left(1+\frac{t}{\delta}\right)^{-\vartheta}; \quad t\geq 0;\ \vartheta,\delta>0. \tag{3}$$

2. Classical Methods of Estimation of Lomax Shape Parameter

The Classical methods selected for the comparative study are (i) Maximum Likelihood Estimator (MLE), (ii) Ordinary Least Squares Method (OLS), (iii) Ridge Regression method, (iv) Uniformly Minimum Variance Unbiased Estimator (UMVUE), and (v) Median Method (M.M).

2.1. MLE's of the Shape Parameter ϑ. Suppose that $\underline{x}=x_1,x_2,\ldots x_n$ is a random sample from the LD as in (1), then the $L(\underline{x}\,|\vartheta)$ for the sample observation will be as follows:

$$L(\underline{x}\,|\vartheta)=\prod_{i=1}^{n}\frac{\vartheta}{\delta}\left(1+\frac{x}{\delta}\right)^{-(\vartheta+1)}=\left(\frac{\vartheta}{\delta}\right)^{n}\exp\left[-T(\vartheta+1)\right], \tag{4}$$

where $T=\sum_{i=1}^{n}\ln(1+x_i/\delta)$.

Log likelihood function

$$\ln L(\vartheta,\delta)=n\ln\vartheta-n\ln\delta-(\vartheta+1)\sum_{i=1}^{n}\ln\left(1+\frac{x_i}{\delta}\right). \tag{5}$$

The MLE's of ϑ denoted by $\widehat{\vartheta}_{\mathrm{MLE}}$ is given as follows:

$$\widehat{\vartheta}_{MLE}=\frac{n}{\sum_{i=1}^{n}\ln\left(1+x_i/\delta\right)}. \tag{6}$$

2.2. Ordinary Least Squares Method (OLS). The CDF in equation (2) satisfies

$$\ln(1-F(x))=-\vartheta\ln\left(1+\frac{x}{\delta}\right)=-\vartheta\ln(\delta+x)+\vartheta\ln\delta. \tag{7}$$

Now, suppose that $X_1,X_2,\ldots\ldots X_n$ form a random sample from LD defined by (1), and that $X_{(1)}<X_{(2)}<\ldots<X_{(n)}$ are the order statistics. With observed ordered observations $x_{(1)}<x_{(2)}<\ldots<x_{(n)}$ (2) gives the following equation:

$$\ln\left(1-F\left(x_{(i)}\right)\right)=-\vartheta\ln\left(1+\frac{x_{(i)}}{\delta}\right)=-\vartheta\ln\left(\delta+x_{(i)}\right)+\vartheta\ln\delta. \tag{8}$$

(8) represents a simple linear regression function corresponding to $F(x_{(i)})$

$$Y_i=\alpha+\beta X_i+\varepsilon_i, \tag{9}$$

where $Y_i=\ln(1-\widehat{F}_i)$ and $\widehat{F}_i i$ it is a point estimator of $F(x_{(i)})$ many estimators for $\widehat{F}_i$ are used.

For example, the Median Rank estimator $\widehat{F}_i=(i-0.3)/(n+0.4)$ or $\widehat{F}_i=(i-3/8)/(n+0.25)$, the mean rank estimator $\widehat{F}_i=i/(n+1)$. where i denotes the i^{th} smallest value of $x_{(1)},x_{(2)},\ldots x_{(n)}$, $i=1,2,\ldots,n$. ε_i is the random error with expected value $E(\varepsilon_i)=0$. $X_i=\ln(\delta+x_{(i)})$, $\beta=-\vartheta$, $\alpha=\vartheta\ln\delta$.

The estimates $\widehat{\alpha}$ and $\widehat{\beta}$ of the regression parameters, α and β minimize the function,

$$Q(\alpha,\beta)=\sum_{i=1}^{n}\left(Y_i-\alpha-\beta\ln\left(\delta+x_{(i)}\right)\right)^2. \tag{10}$$

Therefore, the estimates $\widehat{\beta}_{\mathrm{OLS}}$ of the parameter, β is given by the following equation:

$$\widehat{\beta}_{\mathrm{OLS}}=\frac{n\sum_{i=1}^{n}\ln\left(\delta+x_{(i)}\right)\ln\left(1-\widehat{F}_i\right)-\sum_{i=1}^{n}\ln\left(\delta+x_{(i)}\right)\sum_{i=1}^{n}\ln\left(1-\widehat{F}_i\right)}{n\sum_{i=1}^{n}\ln^2\left(\delta+x_{(i)}\right)-\sum_{i=1}^{n}\ln\left(\delta+x_{(i)}\right)^2},$$

$$\widehat{\alpha}_{\mathrm{OLS}}=\frac{1}{n}\sum_{i=1}^{n}\ln\left(1-\widehat{F}_i\right)-\widehat{\beta}_{\mathrm{OLS}}\frac{1}{n}\sum_{i=1}^{n}\ln\left(\delta+x_{(i)}\right). \tag{11}$$

The estimate $\widehat{\vartheta}_{\mathrm{OLS}}$ of the parameter ϑ is given by the following equation:

$$\widehat{\vartheta}_{\mathrm{OLS}} = -\widehat{\beta}_{\mathrm{OLS}}. \tag{12}$$

2.3. Ridge Regression Method. Ridge Regression estimates can be obtained by minimizing the function.

$$\mathcal{Q}^{*}(\alpha,\beta) = \sum_{i=1}^{n}\left(\ln\left(1-\widehat{F}_i\right) - \alpha - \beta\ln\left(\delta + x_{(i)}\right)\right)^2. \tag{13}$$

According to the following constraint

$$\alpha^2 + \beta^2 = \phi, \tag{14}$$

where ϕ is a definite positive constant.

The Lagrange multiples method requires that we derive the following:

$$\begin{aligned} L &= \sum_{i=1}^{n}\left(\left(\ln\left(1-\widehat{F}_i\right) - \alpha - \beta\ln\left(\delta + x_{(i)}\right)\right)^2 + \lambda\left(\alpha^2+\beta^2-\phi\right)\right),\\ \frac{\partial \ln L}{\partial\alpha} &= -2\sum_{i=1}^{n}\left(\ln\left(1-\widehat{F}_i\right) - \alpha - \beta\ln\left(\delta + x_{(i)}\right)\ln\left(\delta + x_{(i)}\right)\right) + 2\alpha\lambda = 0,\\ \frac{\partial \ln L}{\partial\beta} &= -2\sum_{i=1}^{n}\left(\ln\left(1-\widehat{F}_i\right) - \alpha - \beta\ln\left(\delta + x_{(i)}\right)\right)\ln\left(\delta + x_{(i)}\right) + 2\beta\lambda. \end{aligned} \tag{15}$$

Therefore, the estimates $\widehat{\alpha}_{\mathrm{Rid}}$ and $\widehat{\beta}_{\mathrm{Rid}}$ of the parameters, α and β are given by the following equation:

$$\begin{aligned} \widehat{\beta}_{\mathrm{Rid}} &= \frac{(n+\lambda)\sum_{i=1}^{n}\ln\left(\delta + x_{(i)}\right)\ln\left(1-\widehat{F}_i\right) - \sum_{i=1}^{n}\ln\left(\delta + x_{(i)}\right)\sum_{i=1}^{n}\ln\left(1-\widehat{F}_i\right)}{(n+\lambda)\sum_{i=1}^{n}\ln^2\left(\delta + x_{(i)}\right) - \left(\sum_{i=1}^{n}\ln\left(\delta + x_{(i)}\right)\right)^2},\\ \widehat{\alpha}_{\mathrm{Rid}} &= \frac{\sum_{i=1}^{n}\ln\left(1-\widehat{F}_i\right) - \beta\sum_{i=1}^{n}\ln\left(\delta + x_{(i)}\right)}{n+\lambda}, \end{aligned} \tag{16}$$

and

$$\lambda = \frac{\rho\sigma^2}{\beta'\beta};\ 0 < \lambda < 1, \tag{17}$$

where ρ represents the number of parameter of the distribution and $\beta'\beta$ represents the covariance matrix.

NOT when $\lambda = 0$, we get the estimations of the OLS.

The Ridge Regression estimate of ϑ denoted by $\widehat{\vartheta}_{\mathrm{Rid}}$ is given as follows:

$$\widehat{\vartheta}_{\mathrm{Rid}} = -\widehat{\beta}_{\mathrm{Rid}}. \tag{18}$$

2.4. UMVUE Estimator of the Shape Parameter ϑ. The pdf of LD belongs to the exponential family. Therefore, $T = \sum_{i=1}^{n}\ln(1 + (x_i/\delta))$ is a complete sufficient statistic for ϑ. Then, depending on the theorem of Lehmann-Scheffe [20], the uniformly minimum variance unbiased estimator $\widehat{\vartheta}_{\mathrm{UMVUE}}$ of ϑ, may be given by the following equation:

$$\widehat{\vartheta}_{\mathrm{UMVUE}} = \frac{n-1}{\sum_{i=1}^{n}\ln\left(1 + x_i/\delta\right)}. \tag{19}$$

2.5. Median Method (M.M). This method is dependent on the basis that the median divides the data into two equal parts

$$F\left(x_{\mathrm{med}}\right) = 0.5. \tag{20}$$

By substituting into the cumulative distribution function defined by (2). The equation will become

$$1 - \left(1 + \frac{x_{\mathrm{med}}}{\delta}\right)^{-\vartheta} = 0.5. \tag{21}$$

Therefore, the estimates $\widehat{\vartheta}_{\mathrm{Med}}$ of ϑ, can be obtained as follows:

$$\widehat{\vartheta}_{\mathrm{Med}} = \frac{-\log(0.5)}{\log\left(1 + x_{\mathrm{med}}/\delta\right)}, \tag{22}$$

where x_{med} is the median of the data.

3. Prior and Posterior Density Functions

3.1. Prior Distribution. The Bays estimators demand an appropriate selection of priors for the parameter. If we do not have sufficient knowledge about the parameter, in this case, the noninformative priors are better chosen. Or else, it

is desirable to use informative priors. In this research, we study both types of priors: informative priors and non-informative priors.

3.1.1. Non-Informative Prior. Let us assume that ϑ has noninformative prior density defined as using extended Jeffrey's prior $h_1(\vartheta)$ which is given by the following equation:

$$h_1(\vartheta) \propto [I(\vartheta)]^c;\ c > 0, \tag{23}$$

where $I(\vartheta)$ represented Fisher information matrix which defined as follows:

$$I(\vartheta) = -nE\left[\frac{\partial^2 \log f(x;\vartheta,\delta)}{\partial\vartheta}\right],$$
$$h_1(\vartheta) = \frac{1}{\vartheta^{2c}};\ c > 0. \tag{24}$$

3.1.2. Informative Priors (The Natural Conjugate Prior). In this work, three types of prior distributions were used to study the effect of the different prior distributions on a Bayesian estimate of ϑ.

(a) Chi-squared prior

$$h_2(\vartheta) = \frac{d^{k/2}}{2^{k/2}\Gamma(k/2)}\vartheta^{k/2-1}\exp\left[-\frac{d\vartheta}{2}\right];\quad \vartheta > 0; k, d > 0. \tag{25}$$

(b) Inverted levy prior

$$h_3(\vartheta) = \sqrt{\frac{k}{2\pi}}\vartheta^{-1/2}\exp\left[-\frac{d\vartheta}{2}\right];\quad \vartheta > 0; k > 0, \tag{26}$$

(c) Gamma Prior

$$h_4(\vartheta) = \frac{d^k}{\Gamma(k)}\vartheta^{k-1}\exp[-d\vartheta];\quad \vartheta > 0; k, d > 0. \tag{27}$$

3.2. Posterior Density Functions. The posterior distribution for the shape parameter ϑ can be expressed as follows:

$$\pi(\vartheta|\underline{x}) = \frac{L(\vartheta,\delta|\underline{x})h(\vartheta)}{\int_0^\infty L(\vartheta,\delta|\underline{x})h(\vartheta)\mathrm{d}\vartheta}. \tag{28}$$

Combining the $L(|\underline{x}|\vartheta)$ in (4) and the prior distribution of extended Jeffrey's prior (16), chi-square prior (17), inverted Levy prior (18), and gamma prior (19). The posterior density of ϑ It can be found on respectively as follows:

$$\pi_1(\vartheta|\underline{x}) = \frac{T^{n-2c+1}}{\Gamma(n-2c+1)}\vartheta^{n-2c}\exp[-T\vartheta];\ \vartheta > 0; c > 0,$$
$$\pi_2(\vartheta|\underline{x}) = \frac{(T+d/2)^{n+k/2}}{\Gamma(n+k/2)}\vartheta^{n+k/2-1}\exp\left[-\left(T+\frac{d}{2}\right)\vartheta\right];\ \vartheta > 0; k, d > 0,$$
$$\pi_3(\vartheta|\underline{x}) = \frac{(T+d/2)^{n+1/2}}{\Gamma(n+1/2)}\vartheta^{n-1/2}\exp\left[-\left(T+\frac{d}{2}\right)\vartheta\right];\ \vartheta > 0; k, d > 0,$$
$$\pi_4(\vartheta|\underline{x}) = \frac{(T+d)^{n+k}}{\Gamma(n+k)}\vartheta^{n+k-1}\exp[-(T+d)\vartheta];\ \vartheta > 0; k, d > 0. \tag{29}$$

4. Loss Functions

In Bayes estimation, we will consider three types of loss functions including SELL, LLF, and CLLF.

4.1. Squared Error Loss Function. The SELF is defined as follows [21]:

$$L(\widehat{\vartheta},\vartheta) = (\widehat{\vartheta}-\vartheta)^2. \tag{30}$$

The Bayes estimator of ϑ relative to SELF, signified by $\widehat{\vartheta}_{\mathrm{BSE}}$ is

$$\widehat{\vartheta}_{\mathrm{BSE}} = E_h(\vartheta|\underline{x}). \tag{31}$$

4.2. LINEX Loss Function. The LINEX loss function for ϑ can be written as follows [22, 23]:

$$L(\widehat{\vartheta},\vartheta) \propto [\exp[a(\widehat{\vartheta}-\vartheta)] - a(\widehat{\vartheta}-\vartheta) - 1];\ a \neq 0. \tag{32}$$

The Bayes estimator of ϑ relative to LLF, denoted by $\widehat{\vartheta}_{BL}$ is

$$\widehat{\vartheta}_{BL} = -\frac{1}{a}Ln[E_\vartheta \exp[-a\vartheta]];\ a \neq 0. \tag{33}$$

Provided that $E_\vartheta = \exp[-a\vartheta]$ exists and is finite, where E_ϑ denotes the expected value.

4.3. Composite LINEX Loss Function. CLLF is given by the following formula [24].

$$L(\widehat{\vartheta},\vartheta) = L_a(\widehat{\vartheta},\vartheta) + L_{-a}(\widehat{\vartheta},\vartheta)$$
$$= \exp[-a(\widehat{\vartheta},\vartheta)] + \exp[a(\widehat{\vartheta},\vartheta)] - 2a > 0. \tag{34}$$

The Bayes estimator of ϑ relative to CLLF, denoted by $\widehat{\beta}_{\mathrm{BCL}}$, is

$$\widehat{\beta}_{BCL} = \frac{1}{2a}\ln\left(\frac{E_\vartheta(\exp[a\vartheta]|\underline{x})}{E_\vartheta(\exp[-a\vartheta]|\underline{x})}\right). \tag{35}$$

Provided that $E_\vartheta = (\exp[a\vartheta]|\underline{x})$ and $E_\vartheta(|\underline{x})(\exp[-a\vartheta])$ exist and are finite.

5. Bayes Estimator

In this part, we estimate ϑ, using three various loss functions, including SELF, LLF, and CLLF. We assume four different prior distributions for ϑ including; extended Jeffrey's prior, chi-square prior, inverted Levy prior, and gamma prior [25–28].

5.1. Bayesian Estimator of ϑ under SELF. The Bayes estimates of ϑ relative to SELF depended on $\pi_1(\vartheta|\underline{x})$ which is signified as $\widehat{\vartheta}_{\mathrm{BSE1}}$ and can be acquired by using equations (21) and (26) to be

$$\begin{aligned}\widehat{\vartheta}_{BSE1} &= E(\vartheta|\underline{x}) = \int_0^\infty \vartheta\pi_1(\vartheta|\underline{x})d\vartheta,\\ \widehat{\vartheta}_{BSE1} &= \int_0^\infty \frac{T^{n-2c+1}}{\Gamma(n-2c+1)}\vartheta^{n-2c+1}\exp[-T\vartheta]d\vartheta = \frac{n-2c+1}{T}.\end{aligned} \tag{36}$$

Likewise, we can obtain the Bayesian estimates of ϑ relative to SELF depending on $\pi_2(\vartheta|\underline{x})$, $\pi_3(\vartheta|\underline{x})$, and $\pi_4(\vartheta|\underline{x})$, which are signified as $\widehat{\vartheta}_{\mathrm{BSE2}}$, $\widehat{\vartheta}_{\mathrm{BSE3}}$, and $\widehat{\vartheta}_{\mathrm{BSE4}}$ by using equations ((22) and (26)), ((23) and (26)) and ((24) and (26)), respectively, to be

$$\begin{aligned}\widehat{\vartheta}_{BSE2} &= E(\vartheta|\underline{x}) = \int_0^\infty \vartheta\pi_2(\vartheta|\underline{x})\mathrm{d}\vartheta,\\ \widehat{\vartheta}_{BSE2} &= \int_0^\infty \frac{(T+d/2)^{n+k/2}}{\Gamma(n+k/2)}\vartheta^{n+k/2}\exp\left[-\left(T+\frac{d}{2}\right)\vartheta\right]\mathrm{d}\vartheta = \frac{n+k/2}{T+d/2},\\ \widehat{\vartheta}_{BSE3} &= E(\vartheta|\underline{x}) = \int_0^\infty \vartheta\pi_3(\vartheta|\underline{x})\mathrm{d}\vartheta,\\ \widehat{\vartheta}_{BSE3} &= \int_0^\infty \frac{(T+d/2)^{n+1/2}}{\Gamma(n+1/2)}\vartheta^{n+1/2}\exp\left[-\left(T+\frac{d}{2}\right)\vartheta\right]\mathrm{d}\vartheta = \frac{\Gamma(n+3/2)}{\Gamma(n+1/2)(T+d/2)},\end{aligned} \tag{37}$$

and

$$\begin{aligned}\widehat{\vartheta}_{BSE4} &= E(\vartheta|\underline{x}) = \int_0^\infty \vartheta\pi_3(\vartheta|\underline{x})\mathrm{d}\vartheta,\\ \widehat{\vartheta}_{BSE4} &= E(\vartheta|\underline{x}) = \int_0^\infty \vartheta\pi_3(\vartheta|\underline{x})\mathrm{d}\vartheta,\\ \widehat{\vartheta}_{BSE4} &= \int_0^\infty \frac{(T+d)^{n+k}}{\Gamma(n+k)}\vartheta^{n+k}\exp[-(T+d)\vartheta]\mathrm{d}\vartheta\\ &= \frac{n+k}{T+d}.\end{aligned} \tag{38}$$

5.2. Bayesian Estimator of ϑ under LLF. We can obtain the Bayes estimator of ϑ under the LLF depending on $\pi_1(\vartheta|\underline{x})$ signified as $\widehat{\vartheta}_{BL1}$ by using equations (21) and (28) as follows:

$$\begin{aligned}\widehat{\vartheta}_{BL1} &= -\frac{1}{a}Ln[E_\vartheta\exp[-a\vartheta]] = \int_0^\infty \exp[-a\vartheta]\pi_1(\vartheta|\underline{x})\mathrm{d}\vartheta,\\ \widehat{\vartheta}_{BL1} &= -\frac{1}{a}\ln\int_0^\infty \exp[-a\vartheta]\frac{T^{n-2c+1}}{\Gamma(n-2c+1)}\vartheta^{n-2c}\exp[-T\vartheta]\mathrm{d}\vartheta\\ &= \frac{n-2c+1}{a}\ln\left(1+\frac{a}{T}\right).\end{aligned} \tag{39}$$

In the same way, the Bayes estimates of ϑ relative to LLF depended on $\pi_2(\vartheta|\underline{x})$, $\pi_3(\vartheta|\underline{x})$, and $\pi_4(\vartheta|\underline{x})$, which are signified as $\widehat{\vartheta}_{BL2}$, $\widehat{\vartheta}_{BL3}$, and $\widehat{\vartheta}_{BL4}$ by using equations ((22) and (28)) ((23) and (28)), and ((24) and (28)), respectively, to be

$$\begin{aligned}\widehat{\vartheta}_{BL2} &= -\frac{1}{a}Ln[E_\vartheta\exp[-a\vartheta]] = \int_0^\infty \exp[-a\vartheta]\pi_2(\vartheta|\underline{x})d\vartheta,\\ \widehat{\vartheta}_{BL2} &= -\frac{1}{a}\ln\int_0^\infty \exp[-a\vartheta]\frac{(T+d/2)^{n+k/2}}{\Gamma(n+k/2)}\vartheta^{n+k/2-1}\exp\left[-\left(T+\frac{d}{2}\right)\vartheta\right]d\vartheta\\ &= \frac{n+0.5k}{a}\ln\left(1+\frac{a}{T+d/2}\right),\\ \widehat{\vartheta}_{BL3} &= -\frac{1}{a}Ln[E_\vartheta\exp[-a\vartheta]] = \int_0^\infty \exp[-a\vartheta]\pi_3(\vartheta|\underline{x})d\vartheta,\\ \widehat{\vartheta}_{BL3} &= -\frac{1}{a}\ln\int_0^\infty \exp[-a\vartheta]\frac{(T+d/2)^{n+1/2}}{\Gamma(n+1/2)}\vartheta^{n-1/2}\exp\left[-\left(T+\frac{d}{2}\right)\vartheta\right]d\vartheta\\ &= \frac{n+0.5}{a}\ln\left(1+\frac{a}{T+d/2}\right),\end{aligned} \tag{40}$$

and

$$\widehat{\vartheta}_{BL4} = -\frac{1}{a} Ln\left[E_{\vartheta}\exp\left[-a\vartheta\right]\right] = \int_0^{\infty} \exp\left[-a\vartheta\right]\pi_4\left(\vartheta|\,\underline{x}\right)\mathrm{d}\vartheta,$$

$$\widehat{\vartheta}_{BL4} = -\frac{1}{a}\ln\int_0^{\infty} \exp\left[-a\vartheta\right]\frac{(T+d)^{n+k}}{\Gamma(n+k)}\vartheta^{n+k-1}\exp\left[-(T+d)\vartheta\right]\mathrm{d}\vartheta \tag{41}$$

$$= \frac{n+k}{a}\ln\left(1+\frac{a}{T+d}\right).$$

5.3. Bayesian Estimation of ϑ under CLLF. The Bayes estimate of ϑ under CLLF depended on $\pi_1(\vartheta|\,\underline{x})$, which is signified as $\widehat{\vartheta}_{\mathrm{BCL1}}$ by using equations (21) and (30) to be

$$\widehat{\vartheta}_{BCL1} = \frac{1}{2a}\ln\left(\frac{E_{\vartheta}\left(\exp\left[a\vartheta\right]|\,\underline{x}\right)}{E_{\vartheta}\left(\exp\left[-a\vartheta\right]|\,\underline{x}\right)}\right) = \frac{I_1}{I_2}, \tag{42}$$

where:

$$I_1 = E_{\vartheta}\left(\exp\left[a\vartheta\right]|\,\underline{x}\right) = \int_0^{\infty}\exp\left[a\vartheta\right]\pi_1\left(\vartheta|\,\underline{x}\right)\mathrm{d}\vartheta$$

$$= \int_0^{\infty}\exp\left[a\vartheta\right]\frac{T^{n-2c+1}}{\Gamma(n-2c+1)}\vartheta^{n-2c}\exp\left[-T\vartheta\right]\mathrm{d}\vartheta \tag{43}$$

$$= \left(\frac{T}{T-a}\right)^{n-2c+1},$$

and

$$I_2 = E_{\vartheta}\left(\exp\left[-a\vartheta\right]|\,\underline{x}\right) = \int_0^{\infty}\exp\left[-a\vartheta\right]\pi_1\left(\vartheta|\,\underline{x}\right)d\vartheta$$

$$= \int_0^{\infty}\exp\left[-a\vartheta\right]\frac{T^{n-2c+1}}{\Gamma(n-2c+1)}\vartheta^{n-2c}\exp\left[-T\vartheta\right]d\vartheta \tag{44}$$

$$= \left(\frac{T}{T+a}\right)^{n-2c+1}.$$

Therefore, the Bayes estimation of parameter ϑ is

$$\widehat{\vartheta}_{BCL1} = \frac{1}{2a}\ln\left[\frac{(T/T-a)^{n-2c+1}}{(T/T+a)^{n-2c+1}}\right] = \frac{n-2c+1}{2a}\ln\left(\frac{T+a}{T-a}\right). \tag{45}$$

Similarly, the Bayesian estimates of ϑ under CLLF depended on $\pi_2(\vartheta|\,\underline{x})$, $\pi_3(\vartheta|\,\underline{x})$, and $\pi_4(\vartheta|\,\underline{x})$, which are signified as $\widehat{\vartheta}_{\mathrm{BCL2}}$, $\widehat{\vartheta}_{\mathrm{BCL3}}$, and $\widehat{\vartheta}_{\mathrm{BCL4}}$ by using equations ((22) and (30)), ((23) and (30)), and ((24) and (30)), respectively, to be

$$\widehat{\vartheta}_{BCL2} = \frac{1}{2a}\ln\left(\frac{E_{\vartheta}\left(\exp\left[a\vartheta\right]|\,\underline{x}\right)}{E_{\vartheta}\left(\exp\left[-a\vartheta\right]|\,\underline{x}\right)}\right) = \frac{I_3}{I_4}, \tag{46}$$

where

$$I_3 = E_{\vartheta}\left(\exp\left[a\vartheta\right]|\,\underline{x}\right) = \int_0^{\infty}\exp\left[a\vartheta\right]\pi_2\left(\vartheta|\,\underline{x}\right)d\vartheta$$

$$= \int_0^{\infty}\exp\left[a\vartheta\right]\frac{(T+d/2)^{n+k/2}}{\Gamma(n+k/2)}\vartheta^{n+k/2-1}\exp\left[-\left(T+\frac{d}{2}\right)\vartheta\right]d\vartheta$$

$$= \left(\frac{T+d/2}{T+d/2-a}\right)^{n+k/2}, \tag{47}$$

and

$$I_4 = E_{\vartheta}\left(\exp\left[-a\vartheta\right]|\,\underline{x}\right) = \int_0^{\infty}\exp\left[-a\vartheta\right]\pi_2\left(\vartheta|\,\underline{x}\right)d\vartheta$$

$$= \int_0^{\infty}\exp\left[-a\vartheta\right]\frac{(T+d/2)^{n+k/2}}{\Gamma(n+k/2)}\vartheta^{n+k/2-1}\exp\left[-\left(T+\frac{d}{2}\right)\vartheta\right]d\vartheta$$

$$= \left(\frac{T+d/2}{T+d/2+a}\right)^{n+k/2}, \tag{48}$$

So

$$\widehat{\vartheta}_{BCL2} = \frac{1}{2a}\ln\left[\frac{((T+d/2)/(T+d/2-a))^{n+k/2}}{((T+d/2)/(T+d/2+a))^{n+k/2}}\right]$$

$$= \frac{n+k/2}{2a}\ln\left(\frac{T+(d/2)+a}{T+(d/2)-a}\right), \tag{49}$$

and

$$\widehat{\vartheta}_{BCL3} = \frac{1}{2a}\ln\left(\frac{E_{\vartheta}\left(\exp\left[a\vartheta\right]|\,\underline{x}\right)}{E_{\vartheta}\left(\exp\left[-a\vartheta\right]|\,\underline{x}\right)}\right) = \frac{I_5}{I_6}, \tag{50}$$

where

$$I_5 = E_{\vartheta}\left(\exp\left[a\vartheta\right]|\,\underline{x}\right) = \int_0^{\infty}\exp\left[a\vartheta\right]\pi_3\left(\vartheta|\,\underline{x}\right)d\vartheta$$

$$= \int_0^{\infty}\exp\left[a\vartheta\right]\frac{(T+d/2)^{n+1/2}}{\Gamma(n+1/2)}\vartheta^{n-1/2}\exp\left[-\left(T+\frac{d}{2}\right)\vartheta\right]d\vartheta$$

$$= \left(\frac{T+d/2}{T+(d/2)-a}\right)^{n+1/2}, \tag{51}$$

and

Table 1: MSE values for non-bayes estimators of ϑ.

Cases	n	$\widehat{\vartheta}_{\mathrm{MLE}}$	$\widehat{\vartheta}_{\mathrm{OLS}}$	$\widehat{\vartheta}_{\mathrm{Rid}}$	$\widehat{\vartheta}_{\mathrm{UMVUE}}$	$\widehat{\vartheta}_{\mathrm{Med}}$
	20	0.1457	0.2021	0.1350	0.1249	0.3042
	40	0.0645	0.1005	0.0870	0.0599	0.1440
I	60	0.0408	0.0689	0.0631	0.0388	0.0873
	80	0.0302	0.0531	0.0501	0.0291	0.0665
	100	0.0237	0.0417	0.0399	0.0229	0.0507
	20	0.2551	0.3585	0.2434	0.2196	0.5406
	40	0.1126	0.1814	0.1521	0.1047	0.2412
II	60	0.0702	0.1224	0.1105	0.0667	0.1528
	80	0.0526	0.0942	0.0882	0.0507	0.1148
	100	0.0412	0.0770	0.0732	0.0399	0.0878
	20	0.3995	0.5743	0.4357	0.3462	0.7967
	40	0.1743	0.2818	0.2455	0.1621	0.3770
III	60	0.1116	0.1864	0.1735	0.1062	0.2475
	80	0.0807	0.1478	0.1416	0.0778	0.1775
	100	0.0668	0.1215	0.1175	0.0649	0.1440

Table 2: MSE values for bayes estimators of ϑ with extended Jeffrey's prior.

Cases	n	$\widehat{\vartheta}_{\mathrm{BSE1}}$	$\widehat{\vartheta}_{BL1}$			$\widehat{\vartheta}_{\mathrm{BCL1}}$		
			$a = 0.5$	$a = 1$	$a = 1.5$	$a = 0.5$	$a = 1$	$a = 1.5$
	20	0.1457	0.1298	0.1177	0.1081	0.1464	0.1476	0.1511
	40	0.0645	0.0611	0.0560	0.0532	0.0645	0.0621	0.0614
I	60	0.0408	0.0393	0.0371	0.0371	0.0408	0.0397	0.0408
	80	0.0302	0.0294	0.0292	0.0275	0.0303	0.0307	0.0296
	100	0.0237	0.0232	0.0230	0.0223	0.0237	0.0239	0.0237
	20	0.2551	0.2200	0.1961	0.1803	0.2574	0.2628	0.2776
	40	0.1126	0.1050	0.0967	0.0952	0.1128	0.1104	0.1151
II	60	0.0702	0.0670	0.0655	0.0632	0.0703	0.0716	0.0708
	80	0.0526	0.0508	0.0504	0.0480	0.0526	0.0536	0.0526
	100	0.0412	0.0400	0.0397	0.0388	0.0412	0.0419	0.0415
	20	0.3995	0.3347	0.2922	0.2727	0.4051	0.4189	0.4535
	40	0.1743	0.1601	0.1517	0.1441	0.1748	0.1780	0.1794
III	60	0.1116	0.1054	0.1012	0.0985	0.1117	0.1126	0.1121
	80	0.0807	0.0774	0.0729	0.0723	0.0808	0.0786	0.0805
	100	0.0668	0.0646	0.0600	0.0596	0.0668	0.0636	0.0646

$$
\begin{aligned}
I_6 = E_\vartheta\left(\exp\left[-a\vartheta\right]\middle|\,\underline{x}\right) &= \int_0^\infty \exp\left[-a\vartheta\right]\pi_3\left(\vartheta\middle|\,\underline{x}\right)d\vartheta \\
&= \int_0^\infty \exp\left[-a\vartheta\right]\frac{(T+d/2)^{n+1/2}}{\Gamma(n+1/2)}\vartheta^{n-1/2}\exp\left[-\left(T+\frac{d}{2}\right)\vartheta\right]d\vartheta \\
&= \left(\frac{T+d/2}{T+d/2+a}\right)^{n+1/2},
\end{aligned}
\tag{52}
$$

So

$$
\begin{aligned}
\widehat{\vartheta}_{BCL3} &= \frac{1}{2a}\ln\left[\frac{\left((T+d/2)/(T+d/2-a)\right)^{n+1/2}}{\left((T+d/2)/(T+d/2+a)\right)^{n+1/2}}\right] \\
&= \frac{n+1/2}{2a}\ln\left(\frac{T+d/2+a}{T+d/2-a}\right),
\end{aligned}
\tag{53}
$$

$$
\widehat{\vartheta}_{BCL4} = \frac{1}{2a}\ln\left(\frac{E_\vartheta\left(\exp\left[a\vartheta\right]\middle|\,\underline{x}\right)}{E_\vartheta\left(\exp\left[-a\vartheta\right]\middle|\,\underline{x}\right)}\right) = \frac{I_7}{I_8},
$$

where

$$
\begin{aligned}
I_7 = E_\vartheta\left(\exp\left[a\vartheta\right]\middle|\,\underline{x}\right) &= \int_0^\infty \exp\left[a\vartheta\right]\pi_4\left(\vartheta\middle|\,\underline{x}\right)d\vartheta \\
&= \int_0^\infty \exp\left[a\vartheta\right]\frac{(T+d)^{n+k}}{\Gamma(n+k)}\vartheta^{n+k-1}\exp\left[-(T+d)\vartheta\right]d\vartheta \\
&= \left(\frac{T+d}{T+d-a}\right)^{n+k},
\end{aligned}
\tag{54}
$$

and

$$
\begin{aligned}
I_8 = E_\vartheta\left(\exp\left[-a\vartheta\right]\middle|\,\underline{x}\right) &= \int_0^\infty \exp\left[-a\vartheta\right]\pi_4\left(\vartheta\middle|\,\underline{x}\right)d\vartheta \\
&= \int_0^\infty \exp\left[-a\vartheta\right]\frac{(T+d)^{n+k}}{\Gamma(n+k)}\vartheta^{n+k-1}\exp\left[-(T+d)\vartheta\right]d\vartheta \\
&= \left(\frac{T+d}{T+d+a}\right)^{n+k}.
\end{aligned}
\tag{55}
$$

Therefore, the Bayes estimation of parameter ϑ is

Table 3: MSE values for bayes estimators of ϑ with chi-square prior.

Cases	n	$\widehat{\vartheta}_{BSE1}$	$\widehat{\vartheta}_{BL1}$			$\widehat{\vartheta}_{BCL1}$		
			$a = 0.5$	$a = 1$	$a = 1.5$	$a = 0.5$	$a = 1$	$a = 1.5$
	20	0.1468	0.1305	0.1177	0.1077	0.1476	0.1486	0.1522
	40	0.0648	0.0613	0.0561	0.0531	0.0649	0.0625	0.0617
I	60	0.0410	0.0395	0.0371	0.0370	0.0410	0.0399	0.0410
	80	0.0304	0.0295	0.0292	0.0275	0.0304	0.0309	0.0297
	100	0.0238	0.0232	0.0231	0.0223	0.0238	0.0240	0.0238
	20	0.2529	0.2177	0.1933	0.1771	0.2552	0.2602	0.2746
	40	0.1124	0.1045	0.0961	0.0944	0.1126	0.1101	0.1149
II	60	0.0702	0.0668	0.0653	0.0629	0.0702	0.0715	0.0708
	80	0.0526	0.0507	0.0502	0.0479	0.0526	0.0536	0.0526
	100	0.0412	0.0400	0.0397	0.0386	0.0412	0.0419	0.0415
	20	0.3895	0.3265	0.2849	0.2658	0.3948	0.4077	0.4405
	40	0.1726	0.1584	0.1499	0.1423	0.1731	0.1762	0.1776
III	60	0.1109	0.1047	0.1005	0.0977	0.1110	0.1119	0.1114
	80	0.0804	0.0770	0.0725	0.0719	0.0804	0.0783	0.0801
	100	0.0666	0.0643	0.0598	0.0593	0.0666	0.0634	0.0643

Table 4: MSE values for bayes estimators of ϑ with inverted Levy prior.

Cases	n	$\widehat{\vartheta}_{BSE1}$	$\widehat{\vartheta}_{BL1}$			$\widehat{\vartheta}_{BCL1}$		
			$a = 0.5$	$a = 1$	$a = 1.5$	$a = 0.5$	$a = 1$	$a = 1.5$
	20	0.1528	0.1352	0.1210	0.1100	0.1536	0.1544	0.1583
	40	0.0662	0.0624	0.0569	0.0536	0.0663	0.0638	0.0630
I	60	0.0416	0.0400	0.0375	0.0373	0.0416	0.0405	0.0416
	80	0.0307	0.0298	0.0294	0.0277	0.0307	0.0312	0.0300
	100	0.0240	0.0234	0.0232	0.0224	0.0240	0.0242	0.0240
	20	0.2630	0.2248	0.1974	0.1793	0.2654	0.2703	0.2858
	40	0.1146	0.1062	0.0972	0.0951	0.1149	0.1124	0.1174
II	60	0.0712	0.0676	0.0658	0.0631	0.0712	0.0726	0.0717
	80	0.0531	0.0512	0.0505	0.0481	0.0532	0.0541	0.0532
	100	0.0416	0.0403	0.0399	0.0387	0.0416	0.0422	0.0418
	20	0.4043	0.3358	0.2892	0.2662	0.4099	0.4234	0.4581
	40	0.1760	0.1607	0.1511	0.1425	0.1765	0.1798	0.1813
III	60	0.1124	0.1057	0.1011	0.0977	0.1126	0.1135	0.1128
	80	0.0812	0.0776	0.0728	0.0720	0.0813	0.0791	0.0811
	100	0.0671	0.0647	0.0599	0.0594	0.0671	0.0639	0.0649

$$\widehat{\vartheta}_{BCL4} = \frac{1}{2a}\ln\left[\frac{(T + d/T + d/2 - a)^{n+k}}{(T + d/T + d + a)^{n+k}}\right] = \frac{n+k}{2a}\ln\left(\frac{T + d + a}{T + d - a}\right). \quad (56)$$

6. Simulation Study and Results

In this part, a simulation study has been conducted to assess and examine the behavior of Classical methods and Bayes estimators for the shape parameter of LD under different cases. The following steps of the simulation are as follows:

(1) Set the true values for the parameters of LD which are varied into three cases to observe their effect on the estimates when $\delta > \vartheta$, $\delta = \vartheta$, and $\delta < \vartheta$ "case I $(\delta = 2, \vartheta = 1.5)$, case II $(\delta = 2, \vartheta = 2)$ and case III $(\delta = 2, \vartheta = 2.5)$"

(2) Determine the sample size $n = 20, 40, 60, 80$ and 100

(3) Determine the value $\lambda = 0.75$, $c = 0.5$, $(K, d) = (0.6, 0.2)$ and $a = 0.5, 1$ and 1.5

Table 5: MSE values for bayes estimators of ϑ with gamma prior.

Cases	n	$\widehat{\vartheta}_{\text{BSE1}}$	$\widehat{\vartheta}_{BL1}$			$\widehat{\vartheta}_{\text{BCL1}}$		
			$a = 0.5$	$a = 1$	$a = 1.5$	$a = 0.5$	$a = 1$	$a = 1.5$
	20	0.1482	0.1314	0.1179	0.1075	0.1489	0.1498	0.1534
	40	0.0652	0.0615	0.0562	0.0530	0.0653	0.0629	0.0621
I	60	0.0412	0.0396	0.0372	0.0370	0.0412	0.0401	0.0412
	80	0.0305	0.0296	0.0293	0.0275	0.0305	0.0310	0.0298
	100	0.0239	0.0233	0.0231	0.0223	0.0239	0.0240	0.0239
	20	0.2510	0.2156	0.1907	0.1742	0.2531	0.2578	0.2720
	40	0.1121	0.1041	0.0956	0.0937	0.1123	0.1099	0.1147
II	60	0.0701	0.0667	0.0651	0.0626	0.0702	0.0715	0.0707
	80	0.0526	0.0507	0.0501	0.0477	0.0526	0.0535	0.0526
	100	0.0412	0.0400	0.0396	0.0385	0.0412	0.0419	0.0415
	20	0.3801	0.3187	0.2779	0.2591	0.3851	0.3971	0.4284
	40	0.1709	0.1567	0.1482	0.1405	0.1714	0.1745	0.1759
III	60	0.1102	0.1040	0.0997	0.0969	0.1104	0.1113	0.1107
	80	0.0800	0.0767	0.0721	0.0714	0.0801	0.0780	0.0798
	100	0.0663	0.0641	0.0595	0.0591	0.0664	0.0632	0.0641

(4) For a given sample size n, generate $x_1, x_2, \ldots x_n$ by using the following formula: $x_i = \delta[(1 - U_i)^{-1/\vartheta} - 1]$, $i = 1, \ldots, n$, where U_i is uniform $(0, 1)$

(5) Classical methods estimation, $\widehat{\vartheta}_{MLE}$, $\widehat{\vartheta}_{\text{OLS}}$, $\widehat{\vartheta}_{\text{Rid}}$, $\widehat{\vartheta}_{\text{UMVUE}}$, and $\widehat{\vartheta}_{\text{Med}}$ of ϑ are computed from equations (6), (10), (13)–(15), respectively

(6) Under SELL and based on $h_1(\vartheta), h_2(\vartheta) h_3(\vartheta)$ and $h_4(\vartheta)$ priors, Bayesian estimation, $\widehat{\vartheta}_{\text{BSE1}}$, $\widehat{\vartheta}_{\text{BSE2}}$, $\widehat{\vartheta}_{\text{BSE3}}$ and $\widehat{\vartheta}_{\text{BSE.}}$, of ϑ are computed from equations (31), (32), (33), and (34), respectively

(7) Under LLF and based on $h_1(\vartheta), h_2(\vartheta), h_3(\vartheta)$ and $h_4(\vartheta)$ priors, Bayesian estimation, $\widehat{\vartheta}_{BL1}$, $\widehat{\vartheta}_{BL2}$, $\widehat{\vartheta}_{BL3}$ and $\widehat{\vartheta}_{BL4}$, of ϑ are computed from equations (35), (36), (37), and (38), respectively

(8) Under CLLF and based on $h_1(\vartheta), h_2(\vartheta) h_3(\vartheta)$ and $h_4(\vartheta)$ priors, $\widehat{\vartheta}_{BCL1}$, $\widehat{\vartheta}_{BCL2}$, $\widehat{\vartheta}_{\text{BCL3}}$ and $\widehat{\vartheta}_{\text{BCL4}}$, of ϑ are calculated from equations (40), (42), (44), and (46), respectively

(9) Steps 4 to 8 are replicated 10,000 times. The (MSE's) for all Estimates of the parameter ϑ are obtained, where

$$\text{MSE}\ (\widehat{\vartheta}) = \frac{1}{10000} \sum_{i=1}^{10000} \left(\widehat{\vartheta}_i - \vartheta\right)^2. \tag{57}$$

The results are displayed in the following Tables 1–5.

7. Conclusions and Recommendations

In this paper, the Ridge Regression method was employed to estimate the shape parameter of LD. Besides, researchers made a Monte Carlo simulation to test the performance of the Ridge Regression method. Then, compared the Ridge Regression estimator with the other estimators, including MLE, OLS, UMVUE, M.M, and Bayesian estimators based on SELF, LLF, and CLLF. However, the major observations are identified in the following points:

(1) Among classical estimators, in Table 1, the performance of the UMVUE was shown as better than other estimators: "MLE, OLS, Ridge, and M.M estimators" in all different cases and all samples sizes. Whereas the performance of Ridge Regression was better than MLE. Estimator especially for a small sample size $(n = 10)$. In the meanwhile, the results showed that the performance of the Ridge estimator was better than OLS estimator in all different cases and sample sizes.

(2) With Bayes estimators, gamma prior records full appearance as best prior based on LLF and CLLF for all different cases and all sample sizes. As well as that is true under SELF with $\delta = \vartheta$ and $\delta < \vartheta$, while extended Jeffrey's prior record as best prior based on SELF when $\delta > \vartheta$.

(3) The MSE values associated with each of the classical and Bayes estimate "corresponding to each prior and every loss function" reduces with the increase in the sample size. Also, the results show a convergence between most of the estimators to increase the sample sizes and this conforms to the statistical theory.

(4) For all cases and all sample sizes, LLF $(a = 1.5)$ records full appearance as the best loss function associated with Bayes estimates corresponding to gamma prior.

(5) According to the results, MSE values of all classical and Bayes estimators of shape parameters are decreasing as the shape parameter value increase.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Their authors extend their appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through Larg Groups. (Project under grant number (RGP.2/4/43).

References

[1] P. Fan, R. Deng, J. Qiu, Z. Zhao, and S. Wu, "Well logging curve reconstruction based on kernel ridge regression," *Arabian Journal of Geosciences*, vol. 14, no. 16, pp. 1559–1610, 2021.

[2] A. V. Dorugade, "New ridge parameters for ridge regression," *Journal of the Association of Arab Universities for Basic and Applied Sciences*, vol. 15, no. 1, pp. 94–99, 2014.

[3] E. C. Malthouse, "Ridge regression and direct marketing scoring models," *Journal of Interactive Marketing*, vol. 13, no. 4, pp. 10–23, 1999.

[4] M. R. Özkale, S. Lemeshow, and R. Sturdivant, "Logistic regression diagnostics in ridge regression," *Computational Statistics*, vol. 33, no. 2, pp. 563–593, 2018.

[5] A. G. Assaf, M. Tsionas, and A. Tasiopoulos, "Diagnosing and correcting the effects of multicollinearity: Bayesian implications of ridge regression," *Tourism Management*, vol. 71, pp. 1–8, 2019.

[6] M. I. Ullah, M. Aslam, and S. Altaf, "Lmridge: a comprehensive R package for Ridge regression," *Rice Journal*, vol. 10, no. 2, p. 326, 2019.

[7] V. Sevinç and A. Göktaş, "Çoklu doğrusallık ve değişen varyans altında farklı ridge parametrelerinin bir karşılaştırması," *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, vol. 23, no. 2, pp. 381–389, 2019.

[8] M. P. Rajan, "An efficient Ridge regression algorithm with parameter estimation for data analysis in machine learning," *SN Computer Science*, vol. 3, no. 2, pp. 171–216, 2022.

[9] F. Sami, M. Amin, and M. M. Butt, "On the ridge estimation of the ConwayMaxwell Poisson regression model with multicollinearity: methods and applications," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 1, p. e6477, 2022.

[10] M. Amin, M. N. Akram, and Q. Ramzan, "Bayesian estimation of ridge parameter under different loss functions," *Communications in Statistics - Theory and Methods*, vol. 51, no. 12, pp. 4055–4071, 2022.

[11] N. H. Al-Noor and S. S. Alwan, "Non-Bayes, Bayes and empirical Bayes estimators for the shape parameter of Lomax distribution," *Mathematical Theory and Modeling*, vol. 5, no. 2, pp. 17–28, 2015.

[12] F. S. Al-Duais and M. Y. Hmood, "Bayesian and non-Bayesian estimation of the Lomax model based on upper record values under weighted LINEX loss function," *Periodicals of Engineering and Natural Sciences*, vol. 8, no. 3, pp. 1786–1794, 2020.

[13] H. Ellah, "Comparison of estimates using record statistics from Lomax model: bayesian and non bayesian approaches," *Journal of Statistical Research of Iran*, vol. 3, no. 2, pp. 139–158, 2007.

[14] M. N. Asl, R. A. Belaghi, and H. Bevrani, "Classical and Bayesian inferential approaches using Lomax model under progressively type-I hybrid censoring," *Journal of Computational and Applied Mathematics*, vol. 343, pp. 397–412, 2018.

[15] M. M. Mohie El-Din, H. M. Okasha, and B. Al-Zahrani, "Empirical Bayes estimators of reliability performances using progressive type-II censoring from Lomax model," *Journal of Advanced Research in Applied Mathematics*, vol. 5, no. 1, pp. 74–83, 2013.

[16] H. M. Okasha, "E-Bayesian estimation for the Lomax distribution based on type-II censored data," *Journal of the Egyptian Mathematical Society*, vol. 22, no. 3, pp. 489–495, 2014.

[17] K. Liu and Y. Zhang, "The E-bayesian estimation for Lomax distribution based on generalized type-I hybrid censoring scheme," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–19, 2021.

[18] A. Al-Bossly, "E-bayesian and bayesian estimation for the Lomax distribution under weighted composite LINEX loss function," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–10, 2021, þ, Article ID 2101972.

[19] M. A. W. Mahmoud, "MCMC technique to study the Bayesian estimation using record values from the Lomax distribution," *International Journal of Computer Application*, vol. 73, no. 5, pp. 8–14, 2013.

[20] H Howlader and A. M. Hossain, "Bayesian survival estimation of Pareto distribution of the second kind based on failure-censored data," *Computational Statistics & Data Analysis*, vol. 38, no. 3, pp. 301–314, 2002.

[21] F. S. Al-Duais, "Bayesian estimations under the weighted LINEX loss function based on upper record values," *Complexity*, vol. 2021, pp. 1–7, 2021, þ, Article ID 9982916.

[22] F. S. Al-Duais and M. Alhagyan, "Nonlinear programming to determine best weighted coefficient of balanced LINEX loss function based on lower record values," *Complexity*, vol. 2021, pp. 1–6, 2021, þ, Article ID 5273191.

[23] F. S. Al-Duais, "Bayesian analysis of record statistic from the Inverse Weibull distribution under balanced loss function," *Mathematical Problems in Engineering*, vol. 2021, pp. 1–9, 2021, þ, Article ID 6648462.

[24] S. Wei, C. Wang, and Z. Li, "Bayes estimation of Lomax distribution parameter in the composite LINEX loss of symmetry," *Journal of Interdisciplinary Mathematics*, vol. 20, no. 5, pp. 1277–1287, 2017.

[25] G. Xue, F. Lin, S. Li, and H. Liu, "Adaptive dynamic surface control for finite-time tracking of uncertain nonlinear systems with dead-zone inputs and actuator faults," *International Journal of Control, Automation and Systems*, vol. 19, no. 8, pp. 2797–2811, 2021.

[26] A. I. Taloba, M. R. Riad, and T. H. A. Soliman, "Developing an efficient spectral clustering algorithm on large scale graphs in spark," in *In Proceedings of the 2017 Eighth International Conference on Intelligent Computing and Information Systems ICICIS*, pp. 292–298, IEEE, Cairo, Egypt, December 2017.

[27] A. Rayan, A. I. Taloba, R. M. Abd El-Aziz, and A. Amr, "IoT enabled secured fog-based cloud server management using task prioritization strategies," *International Journal of Advanced Research in Engineering & Technology*, vol. 11, p. 9, 2020.

[28] S. Ha, L. Chen, H. Liu, and S. Zhang, "Command filtered adaptive fuzzy control of fractional-order nonlinear systems," *European Journal of Control*, vol. 63, pp. 48–60, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 3585506, 13 pages
https://doi.org/10.1155/2022/3585506

Research Article

Generative Adversarial Network Combined with SE-ResNet and Dilated Inception Block for Segmenting Retinal Vessels

Chen Yue,[1,2] **Mingquan Ye**,[1,2] **Peipei Wang**,[1,2] **Daobin Huang**,[1,2] **and Xiaojie Lu**[1,2]

[1]*School of Medical Information, Wannan Medical College, Wuhu 241002, China*
[2]*Research Center of Health Big Data Mining and Applications, Wannan Medical College, Wuhu 241002, China*

Correspondence should be addressed to Mingquan Ye; ymq@wnmc.edu.cn

Received 15 May 2022; Revised 12 July 2022; Accepted 29 July 2022; Published 28 August 2022

Academic Editor: Wei Xiang

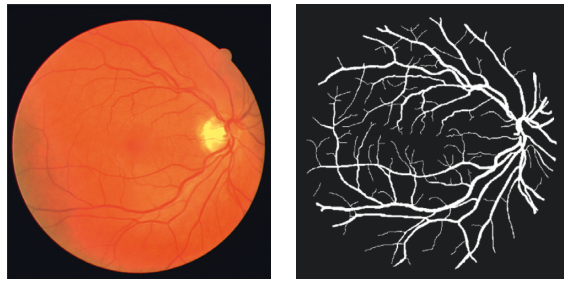
This study develops an accurate method based on the generative adversarial network (GAN) that targets the issue of the current discontinuity of micro vessel segmentation in the retinal segmentation images. The processing of images has become increasingly efficient since the advent of deep learning method. We have proposed an improved GAN combined with SE-ResNet and dilated inception block for the segmenting retinal vessels (SAD-GAN). The GAN model has been improved with respect to the following points. (1) In the generator, the original convolution block is replaced with SE-ResNet module. Furthermore, SE-Net can extract the global channel information, while concomitantly strengthening and weakening the key features and invalid features, respectively. The residual structure can alleviate the issue of gradient disappearance. (2) The inception block and dilated convolution are introduced into the discriminator, which enhance the transmission of features and expand the acceptance domain for improved extraction of the deep network features. (3) We have included the attention mechanism in the discriminator for combining the local features with the corresponding global dependencies, and for highlighting the interdependent channel mapping. SAD-GAN performs satisfactorily on public retina datasets. On DRIVE dataset, ROC_AUC and PR_AUC reach 0.9813 and 0.8928, respectively. On CHASE_DB1 dataset, ROC_AUC and PR_AUC reach 0.9839 and 0.9002, respectively. Experimental results demonstrate that the generative adversarial model, combined with deep convolutional neural network, enhances the segmentation accuracy of the retinal vessels far above that of certain state-of-the-art methods.

1. Introduction

We can infer the vascular tissue condition of the whole body by examining the retinal blood vessels of the fundus. Since there are several small capillary structures in the fundus, as shown in Figure 1, the manual segmentation of the retinal blood vessel images is complicated and cumbersome. Recently, computer vision technology has been extensively employed in image classification, target tracking [1, 2], semantic segmentation, and other fields. Therefore, scholars have proposed certain efficient automatic segmentation algorithms for the retinal blood vessels, which can assist doctors to complete the segmentation and diagnosis with the aid of computers. Traditional image segmentation methods are based on edge detection [3], threshold [4], and region [5] segmentation methods. However, since medical images tend to have complex textures and blurred boundaries, the abovementioned methods sometimes do not work well.

Recently, the application of the convolutional neural networks in various image scenes has brought breakthroughs in several fields of image processing, such as image classification [6] and semantic segmentation [7]. The capability of the automatic feature extraction of the deep learning algorithms effectively overcomes the drawbacks of the traditional medical image segmentation algorithms that exceedingly rely on the prior cognition of the medical experts. Furthermore, the deep learning algorithms are highly portable and can be quickly expanded to different task scenarios with the help of transfer learning.

The GAN model is one of the most promising methods for unsupervised and semi-supervised learning [8–10]. However, the semi-supervised methods mandate a large



(a) (b)

Figure 1: The retinal blood vessel images. (a) Fundus image. (b) Ground truth.

number of unlabeled data sets for the experiments. According to the previous studies, the experiments are complex, hence supervised learning is still considered in this paper. Luc et al. [11] have used adversarial training for semantic segmentation and proved the accuracy of the method on publicly available data sets. Xue et al. [12] have proposed the first adversarial network SegAN for medical image segmentation. SegAN has taken a fully convolutional neural network as a segment and introduced a multi-scale L1 loss function to evaluate the segmentation results, thus obtaining a higher segmentation accuracy. Isola et al. [13] have proposed a general framework of "image-to-image conversion," which solves the problems of conversion and loss function with respect to different types of images, in the traditional methods. The generator of DR-GAN, proposed by Zhou et al. [14], takes the retinal structure, pathology, and adaptive grading vectors as the conditions, and introduces a multi-scale space and channel attention mechanism to extract effective feature information. The RV-GAN proposed by Kamran et al. [15] has used multiple generators and multi-scale discriminators and introduced a new loss function that can improve the segmentation accuracy of the retinal thick and thin blood vessels. The GAN proposed by Tavakkoli et al. [16] can generate the fluorescein angiography (FA) images from the fundus photos with a high segmentation accuracy, with a quality that exceeds that of the images from the most advanced algorithms.

The model proposed in this paper, termed SAD-GAN, SAD-GAN is composed of a generator and a discriminator, as shown in Figure 2. The fundus images are input to the generator. Following the generation of the segmentation maps, they are input into the discriminator along with the label images, and the discriminator distinguishes the generated segmentation images from the labeled ones. The discriminator shares the parameters with the generator, and both play a game against each other. The segmentation effect of the generator is improved through continuously alternating iterative training.

Although most of the studies demonstrate that the GAN performs well in image segmentation, its segmentation accuracy needs improvement, since the adversarial learning of the discriminator in the GAN affects the segmentation accuracy of the final generator. Therefore, it is essential to design a discriminator that can accurately distinguish the prediction graph from the ground truth. Furthermore, the design of the loss function in the adversarial training also affects the segmentation result. Therefore, we optimize the generator and discriminator by using the convolutional neural network and add the adversarial loss function to the cross-entropy loss function to improve the segmentation result, by controlling the weight.

2. Related Works

Since Ronneberger et al. [17] have proposed U-Net, U-Net has been extensively used in image segmentation. Owing to its unique encoding and decoding structure, the performance of the image segmentation has greatly improved, hence it is extensively applied in computer vision research. However, as the network level deepens, the problem of the gradient disappearance easily occurs, and the emergence of residual learning and dense connections can alleviate this problem. For example, Ding et al. [18] have introduced the residual learning, feature fusion, and balance modules in the proposed MRU-Net, and achieved high accuracy on the retinal data set. Alom et al. [19] have proposed a cyclic residual model based on U-Net that can extract the effective segmentation features and segmentation performance of the network, and it achieved good segmentation results. Ibtehaz and Rahman [20] have proposed MultiResUNet, which employed the residual structure in the encoding stage, decoding stage, and skip connection part, and henceforth, the performance relatively improved for the multiple datasets. The MPS-Net proposed by Lin et al. [21] adopts one high-resolution and two low-resolution routes. The proposed multi-path scale module enhances the representation ability of the model, and the significantly improved multiple indicators confirm that the method has good application prospects. The DCU-Net proposed by Yang et al. [22] is a segmentation model that connects the two U-Nets. DCU-Net is primarily composed of the residual deformable convolution modules to enhance the feature extraction ability of the model. To improve the information communication between the two U-shaped networks, a residual attention module has been designed to connect the two U-shaped networks, which achieves a high segmentation accuracy on the retina datasets. Zhai et al. [23] have adopted the inception module in the encoder to obtain the multi-scale features and adopted the pyramid pooling module in the decoder to fuse the multi-scale information. Experiments show that this method improves the segmentation effect.

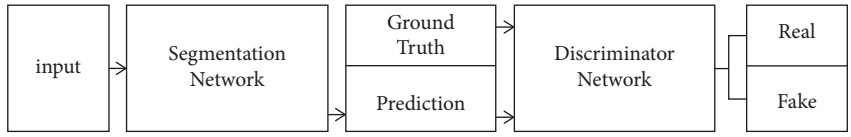


Figure 2: Overall architecture of SAD-GAN.

Once the model has excessive parameters, it will cause an information overload. However, the attention mechanism can focus on the key information and obtain the target area, besides suppressing or filtering out the irrelevant information interference and improving the efficiency of the computing processing. The SE-Net proposed by Hu et al. [24] has focused on the channel of the feature map and then multiplied it with the original feature map. The convolutional block attention module (CBAM) proposed by Woo et al. [25] is a simple and effective feedforward convolutional neural network attention module. CBAM applies attention to both the channel and spatial dimensions. CBAM has the same advantages as a SE module. They can be applied to most models, thus making the model more capable of extracting the feature information without complicating the model. The attention mechanism is employed multiple times in this article. Recently, most of the relevant works have demonstrated that the attention mechanism can enhance feature expression. According to several studies, we can see the significance of the attention mechanism and its wide range of applications. For example, Li et al. [26] have proposed a module that combines a triple attention mechanism, channel, space, and feature internal attention mechanism to improve the correlation between feature maps of various dimensions. Wang et al. [27] have proposed a residual attention network combined with a feedforward network system. The different perception functions of attention modules have been altered through the depth of stacking, and they have achieved a good target recognition on three different datasets. On the basis of FCN, Fu et al. [28] have adopted a dual attention network, channel attention, and spatial attention to obtaining the characteristics of the related dimensions. Furthermore, it has achieved a high segmentation performance on three different scene datasets. Wang et al. [29] have proposed an effective channel attention module, and through the local cross-channel interaction method, the dimensionality reduction was avoided and the parameters were reduced, besides effectively reducing the complexity of the model. Huang et al. [30] have proposed a cross-network CCNet. The proposed cyclic cross-attention module has collected the pixel information on the cross paths and captured the image dependencies. The proposed method has reduced the GPU memory usage and demonstrated a high computational efficiency. The DAR-Net proposed by Li and Wang [31] has used a dense residual module, and it added the channel and spatial attention to the module to select the feature information in the feature map that performed well on the multiple data sets. The DCACNet proposed by Lu et al. [32] has also applied the dual attention mechanism to the medical image segmentation, thus establishing the pixel associations and adding feature representations.

Since Radford et al. [33] have proposed the deep convolutional generative adversarial networks, which can combine the deep convolutional neural network with the generative adversarial network, this method is more rapid and stable in the experiment with an excellent effect. Although the generative adversarial networks are extensively used in the field of image processing, they are rarely seen in retinal image processing [34]. Therefore, this work focuses on the combination of the generative adversarial network and the deep convolutional neural network.

The main contributions from our work can be listed as follows: (1) We have proposed a module that combines ResNet and SE-Net in SAD-GAN. The module is employed as the convolutional layer of the generator. Residual connections can alleviate the gradient disappearance and improve computational efficiency. The attention mechanism can extract the global effective information, and suppress the interference of redundant information. (2) We have added a module in SAD-GAN that combines the dilated convolution and inception in the discriminator. For extracting the effective features, we have used the dilated convolution to obtain the multi-scale features through different dilation rates. Furthermore, we have obtained a larger receptive field and context information by capturing the global information. We employ the Inception-ResNet structure proposed by Szegedy et al. [35] to solve the problems of excessive calculation parameters and disappearance of gradients, which can reduce the redundant calculations through the inception structure and alleviate the phenomenon of the gradient disappearance by adding the residual connections. Furthermore, the structure improves the utilization of the computing resources. We have included the attention mechanism at the end of the discriminator to obtain spatial attention and channel attention to improve the classification accuracy of the discriminator.

3. Methods

3.1. Generator. The generator model used in this work follows the traditional U-shaped structure, as shown in Figure 3. The U-shaped network has been proposed by Ronneberger et al. whose core idea is the encoding and decoding structures. The function of the encoding structure is to obtain the feature information of the image and expand the acceptance domain through the pooling operation. The decoding structure locates and classifies the pixels, which allows the network to propagate the contextual information to higher resolution levels, and the skip connection combines the encoding layer features with the decoding layer features to reduce the loss of information and achieve an effective image segmentation.

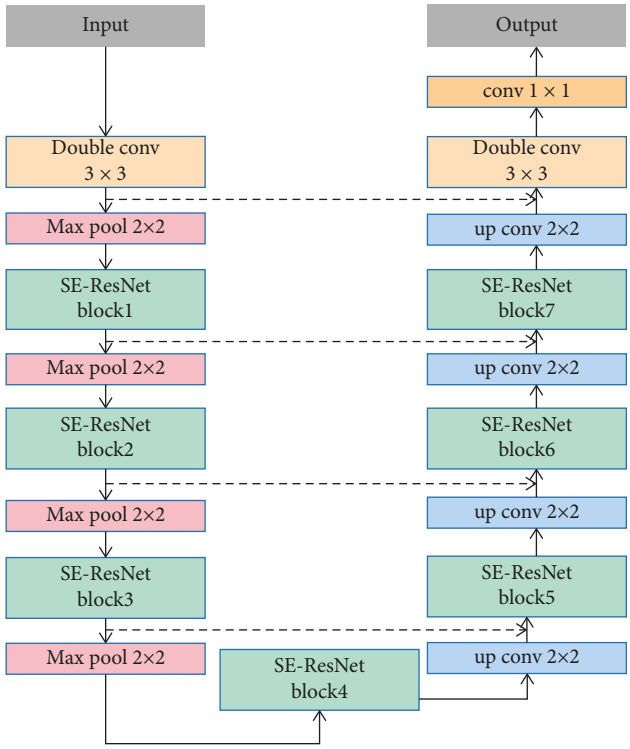


Figure 3: The structure of the generator.

To improve the segmentation accuracy, we have employed an improved convolution layer, which combines the residual structure and SE-Net, termed SE-ResNet block. The generator model employed in this work consists of two traditional convolution layers, seven SE-ResNet blocks, four max-pooling layers, four up sampling layers, and a conv 1×1 layer. Subsequent to 1×1 convolution, the pixels are divided into two categories and the segmented image is output.

In the original U-Net model, the traditional convolution layer is composed of two Conv-BN-ReLU. Frequent convolution and pooling operations result in losing a lot of feature information. Residual structure can retain the feature mapping in the deep network, thus alleviating the risk of the gradient message and easing the optimization of the model. Batch normalization (BN) method normalizes the activation values of the hidden layer neurons to prevent the gradient disappearance and improve the model expression ability. ReLU can reduce the model complexity and speed up the convergence.

SE-ResNet module combines the residual block and SE-Net. SE-Net module is an attention mechanism that extracts channel information, emphasizes global effective information features, and inhibits the irrelevant information features, which can alleviate the risk of deep network overfitting. In Figure 4, SE-Net consists of five continuous functions, viz., global average pooling, fully connected (FC) layer, ReLU, fully connected (FC) layer, and sigmoid function. Furthermore, C is the channel feature number, r is the reduction ratio that is set to 2. The global average pooling can extract the global feature information, after a fully connected layer. The channel feature number is compressed into C/r, through ReLU, to improve the nonlinear. Following another fully connected layer, the number of the channel features is restored to C, and the sigmoid function can obtain the weight of each channel. Finally, the features of each channel are multiplied by the corresponding weights as the output. SE-Net can be used along with any architecture.

3.2. Discriminator. The structure of the discriminator is displayed in Figure 5. The structure of the first convolution block is Conv 3×3-BN-ReLU, and the output of the first convolution block is used as the input of the inception block. We have used the Inception-ResNet proposed by Szegedy et al. [35]. Inception module has for long been extensively used. It has the advantages of extracting the multi-scale features, separating spatial features, and selecting the channel information. In this work, the combination of the inception module and dilated convolution can reduce the number of parameters and expand the acceptance field. In Figure 6, the values of rates 1, 2, and 3, are 1, 2, and 4, respectively. Thereafter, it passes through the down sampling layer composed of the convolution and pooling layers. The second and third convolution blocks are Conv 1×1-BN-ReLU, which reduce the dimension of the channel number, number of parameters, and the risk of network overfitting while improving the nonlinear characteristics and accelerating the training process. The pooling layer is composed of the average pooling 2×2. The reason for the adoption of the average pooling instead of max pooling is that in the image classification, the average pooling can retain the high-level semantic information in the deep network and reduce information loss. Furthermore, the size of feature maps can be changed through convolution and pooling operations. This discriminator introduces the channel and spatial attentions to combine the local and global information, thus reducing the redundancy of information in the physical signs and extracting more effective information. Through a series of operations, the sample features can be extracted. Finally, it distinguishes the ground truth and the generated retinal blood vessel segmentation images through the fully connected and SoftMax.

3.2.1. Dilated Inception Block. In a convolutional neural network, different convolution kernels represent different receptive fields, and the convolution kernel is larger for a larger feature scale. However, excessively large convolution kernels often expand the number of parameters, thus resulting in heavy computing tasks. Therefore, in the previous studies, the method of replacing the large convolution kernels with dilated convolution has appeared. The dilated convolution can obtain the acceptance domain of different scales, besides saving a lot of computation [36]. In Figure 6, we have used the dilated convolution with the dilation rates of 1, 2, and 4, thus the acceptance field can be expanded to obtain features of different scales. The inception has been proposed earlier and is widely used in various computer vision studies. Its variant plays an important role in the deep network [37]. Therefore, according to Figure 7, we combine

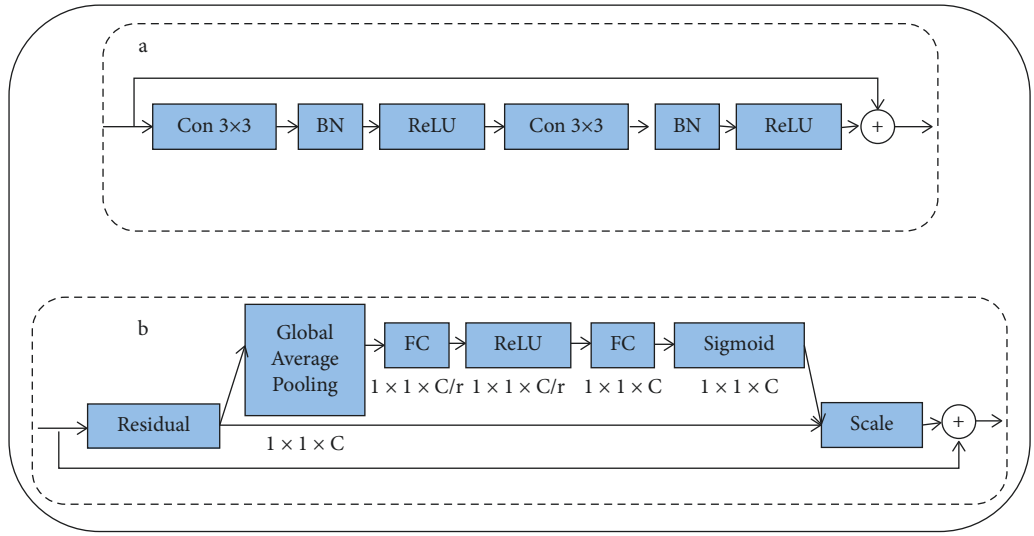


Figure 4: SE-ResNet module combines residual block and SE-Net. (a) Residual block. (b) SE-Net module.

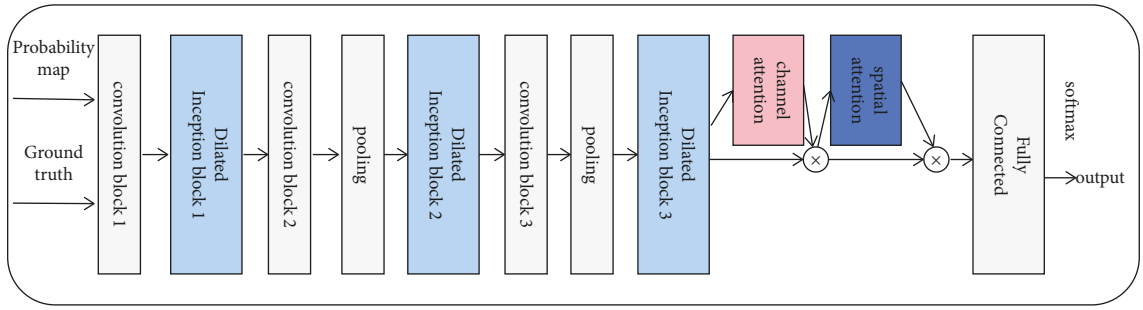


Figure 5: The structure of the discriminator.

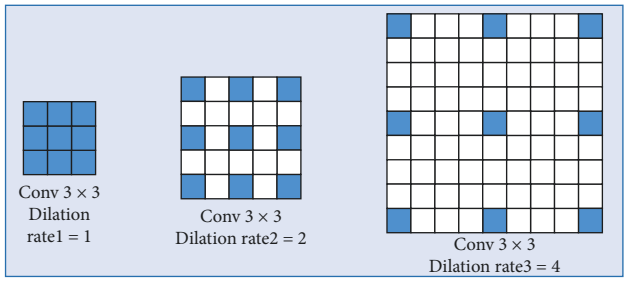


Figure 6: The dilated convolution with the different dilation rates.

inception and dilated convolution to extract the multi-scale features with fewer parameters to improve the overall performance of the network.

3.2.2. Attention Mechanism. According to Figure 8(a), the channel attention focuses on the channel dimension. Each channel of the feature map has been considered as a feature detected by a feature detector, and each channel feature is obtained by a different convolution kernel. For a feature map ($C \times H \times W$), we have calculated the importance of each channel, which is the weight ($C \times 1 \times 1$). Furthermore, we multiply the weight with the feature map to obtain a weighted feature map, which gives the channel attention.

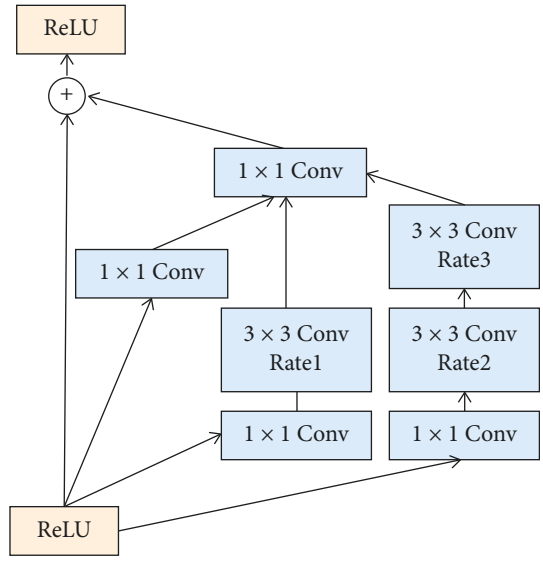


Figure 7: The dilated inception block.

The weight is calculated by MaxPool and AvgPool. To obtain the weight of the channel dimension, each channel ($1 \times H \times W$) needs to be compressed into a number ($1 \times 1 \times 1$). The two methods are given in the following descriptions. We have calculated the average value

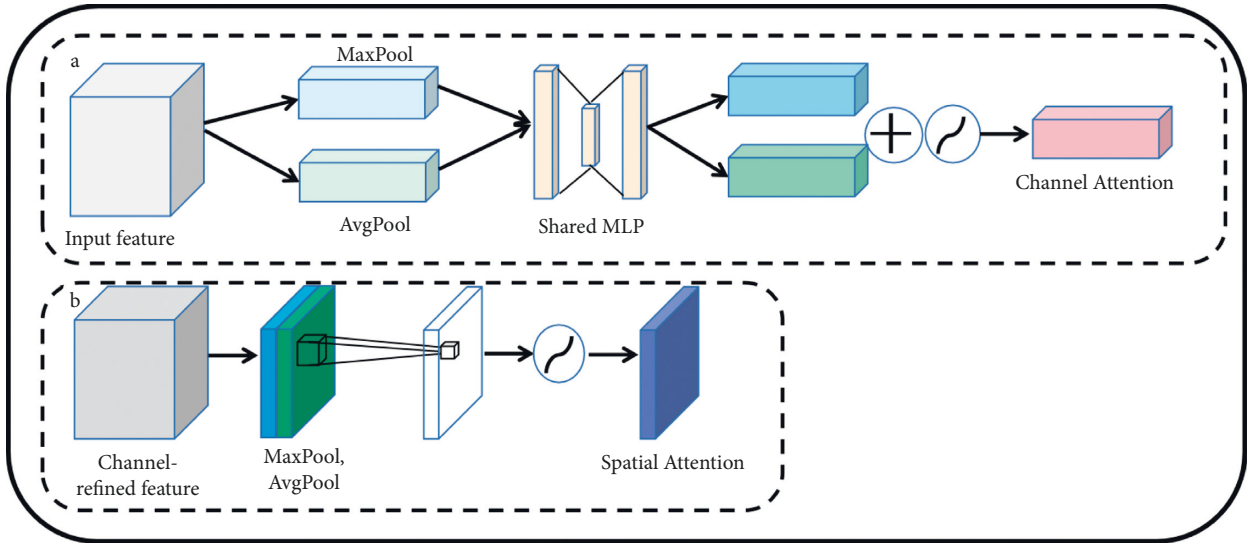


Figure 8: The attention mechanism. (a) Channel attention module. (b) Spatial attention module.

(AvgPool) and the maximum value of each channel (MaxPool). Furthermore, we pass them through a common multilayer perceptron MLP. The structure of MLP is an input or the input neuron C, and the hidden layer neuron C/r, besides the output neuron C. Thereafter, we add these two outputs and normalize them with the sigmoid function to get the final weight. Then we multiply this weight with the original feature map to get the weighted feature map in the channel dimension.

Spatial attention focuses on "where," which is an important issue in each of the channels. Its purpose is to perform AvgPool and MaxPool operations on different channel values on the same plane space point and pass a convolutional layer and a sigmoid function to obtain the weight. Furthermore, we multiply this weight with each channel in the spatial dimension, and finally, get the weighted feature map in the spatial dimension. The spatial attention module is shown in Figure 8(b). It can simulate the rich global feature context information so that the same feature in different locations can be enhanced. Furthermore, the semantic segmentation ability can be improved. In the last part, we have used the sequential model, and let the feature map pass the channel attention first, and then pass the spatial attention.

3.3. Loss Function. Formula (1) is the loss function of the discriminator, which is the minimum spatial cross-entropy loss function for two classes.

$$L_D = -\sum_{h,w}\left((1-q)\cdot \log\left(1 - D(S(x))^{(h,w)}\right)\right) + \left(q\cdot \log\left(D(y)^{(h,w)}\right)\right). \tag{1}$$

When q is zero, the sample comes from the generator, and when q is 1, it comes from the ground truth. Here, x represents the input image, y represents the ground truth, and $y(h,w)$ represents the pixel value of y at the position (h,w). Formula (2), given below, is the loss function of the generator, which is composed of the cross entropy and adversarial losses.

$$L_{\text{seg}} = L_{\text{ce}} + \lambda_{\text{adv}} L_{\text{adv}}, \tag{2}$$

λ_{adv} is the weight to minimize L_{seg}. The cross-entropy loss is shown in equation (3).

$$L_{\text{ce}} = -\sum_{h,w}\sum_{c\in C} y^{(h,w,c)} \log\left(S(x)^{(h,w,c)}\right), \tag{3}$$

where x is the input image, y is the ground truth, $S(x)$ is the prediction result, and C refers to the number of classes. Formula (4) is the adversarial loss. To train the generator and trick the discriminator, we have added adversarial loss to maximize the probability of the predicted results being generated from the ground truth distribution.

$$L_{\text{adv}} = -\sum_{h,w} \log\left(D\left(S^{(h,w)}(x)\right)\right). \tag{4}$$

4. Results and Discussion

To demonstrate the performance of the SAD-GAN model, we have tested them on two retinal image datasets, viz., DRIVE [38] and CHASE_DB1 [39]. We have employed Adam optimizer for both the networks and we trained the generator with a weight decay $= 10^{-4}$, learning rate $= 2.5 \times 10^{-4}$, and learning rate of discriminator $= 10^{-4}$. Furthermore, we have set λ_{adv} to 0.1, batch_size = 1, and the number of iterations of the two data sets to 20000. To complete the implementation, we have used the PyTorch and TensorFlow frameworks on a single GPU machine with 16G RAM and NVIDIA GEFORCE GTX-1650 SUPER.

4.1. Database and Preprocessing Methods. Database: The DRIVE dataset includes 40 retinal images. The size of each image is 584×565. The CHASE_DB1 dataset includes 28

retinal images. The size of each image is 999×960. Both the datasets have the mask and ground truth manually segmented by experts.

Preprocessing: Random rotation and mirror flip have been employed to increase the number of datasets. Since the contrast between the G channel and the blood vessel is the highest among the three RGB channels, we have extracted the G channel to obtain the grayscale image, and then normalized it. Furthermore, we have used the contrast-limited adaptive histogram equalization (CLAHE), to improve the contrast between the blood vessel and the background. Finally, the gamma transform algorithm has been employed to increase the brightness of the darker part without affecting the brighter part. The final data set is partitioned as 10% for validation and 90% for training.

4.2. Evaluation Indexes. Retinal blood vessel segmentation is intended to divide the pixels in the color retinal image into the blood vessels and background. This work has employed accuracy (ACC), sensitivity (SE), specificity (SP), ROC_AUC, and PR_AUC to judge the effectiveness of the retinal blood vessel segmentation method in this work. These indices are the five commonly used ones to measure the performance of the algorithm. Their calculation methods are listed in Table 1. ROC_AUC represents the relationship between sensitivity and specificity, whereas ROC_AUC and PR_AUC represent the prediction accuracy of the image segmentation result. The segmentation effect of the algorithm is better for the value of AUC proximate to one.

4.3. Ablation Study. Figures 9–12 are the segmentation results of SAD-GAN on DRIVE and CHASE_DB1. We have conducted comparative experiments to ensure that the generator and discriminator, which we have proposed, to improve the segmentation accuracy. SD-GAN: the generator is identical to SAD-GAN, and we replace the inception module in the discriminator with a standard convolution layer consisting of two Conv 3×3-BN-ReLU. Furthermore, we remove the attention module, and this model is called SD-GAN. RA-GAN: the generator is identical to that of SAD-GAN, and the discriminator removes the attention module from the discriminator of SAD-GAN, and this model is named as RA-GAN. SG-GAN: we replace the SE-ResNet block in the generator with two Conv 3×3-BN-ReLU. Furthermore, the discriminator is identical to SAD-GAN, and this model is termed SG-GAN. We have used SAD-GAN, SD-GAN, RA-GAN, and SG-GAN models to conduct the experiments.

To further demonstrate the effectiveness of the improvement measures, this work has employed four different GAN models to conduct experiments on the DRIVE and CHASE_DB1 datasets. The experimental results are displayed in Figure 13. The first to sixth columns represent the original image, ground truth, segmentation result of SAD-GAN, segmentation result of the RA-GAN model, segmentation result of the SG-GAN model, and segmentation result of the SD-GAN model, respectively. We have found through comparison that the SAD-GAN network structure excels other network models in the overall performance of both datasets. The segmentation results of the algorithm in this paper are consistent with the expert manual segmentation results. Compared with other algorithms, the SAD-GAN model employed in this paper can better overcome the difficulties in retinal segmentation and achieve better segmentation results for small blood vessels.

Table 1: The evaluation indexes.

Evaluation metric	Description
Accuracy (ACC)	$(P_{TP} + P_{TN})/(P_{TP} + P_{FP} + P_{FN} + P_{TN})$
Sensitivity (SE)	$P_{TP}/(P_{TP} + P_{FN})$
Specificity (SP)	$P_{TN}/(P_{TN} + P_{FP})$

4.3.1. Comparison of the Segmentation Indicators of Different GAN Algorithms in this Paper. Figures 14 and 15 show the ROC and PR curves, respectively, of several types of models under the DRIVE and CHASE_DB1 datasets. When compared with RA-GAN, SG-GAN, and SD-GAN, the PR_AUC value of RA-GAN is 0.9003 on DRIVE, which exceeds that of SAD-GAN. However, SAD-GAN has the largest area under ROC on both datasets and has the largest area under PR on the CHASE_DB1 dataset. Hence, its segmentation performance is relatively better. The segmentation results of SG-GAN, SD-GAN, RA-GAN, and SAD-GAN have been compared. The results of the experiments are listed in Tables 2 and 3. On the DRIVE dataset, the ACC value of the algorithm in this work is 0.9567, the SE value is 0.8422, the SP value is 0.9825, the ROC_AUC value is 0.9813, and the PR_AUC value is 0.8928. On the CHASE_DB1 dataset, the ACC value of the algorithm given in this work is 0.9671, the SE value is 0.8503, the SP value is 0.9850, the ROC_AUC value is 0.9839, and the PR_AUC value is 0.9002. Compared with the first three models, the recognition accuracy of SAD-GAN is higher, which demonstrates that SAD-GAN can segment the retinal vascular structure accurately. According to the tables, the indicators of SAD-GAN exceed those of the other three models, which establishes that the improved generator and discriminator proposed in this paper have marked effects, thus tremendously improving the overall performance and the network prediction ability.

4.3.2. Comparison of the Segmentation Indicators between the Algorithm in this Article and Other Algorithms. Tables 4 and 5 display the index results of SAD-GAN and recent retinal segmentation methods on DRIVE and CHASE_DB1 datasets. SUD-GAN proposed by Yang et al. [46] is also a generative adversarial network for retinal vascular segmentation. On the DRIVE dataset, the algorithm proposed in this work outperforms SUD-GAN in all indicators. Furthermore, PR_AUC has been improved by nearly 10%, whereas PR_AUC is 0.8928 and sensitivity is 0.8422. ROC_AUC is 0.9813, which is the highest value in Table 4. On the CHASE_DB1 dataset, the accuracy, specificity, and

Figure 9: From left to right are the fundus image, ground truth, and the segmentation result of SAD-GAN on DRIVE, respectively.

Figure 10: From left to right are the fundus image, ground truth, and the segmentation result of SAD-GAN on DRIVE, respectively.

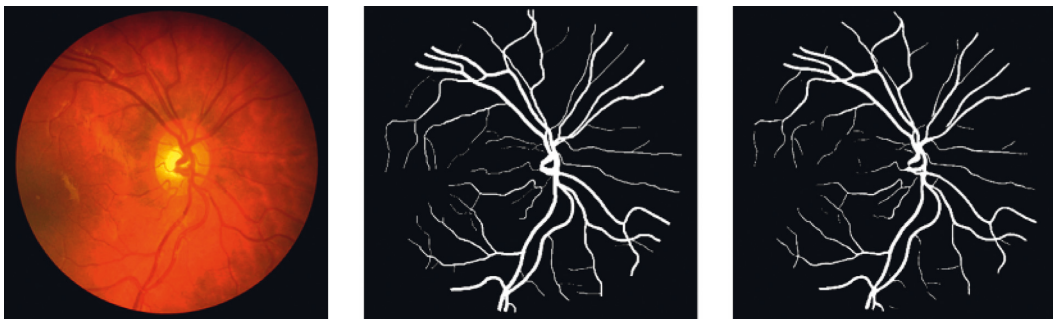


Figure 11: From left to right are the fundus image, ground truth, and the segmentation result of SAD-GAN on CHASE_DB1, respectively.

Figure 12: From left to right are the fundus image, ground truth, and the segmentation result of SAD-GAN on CHASE_DB1, respectively.

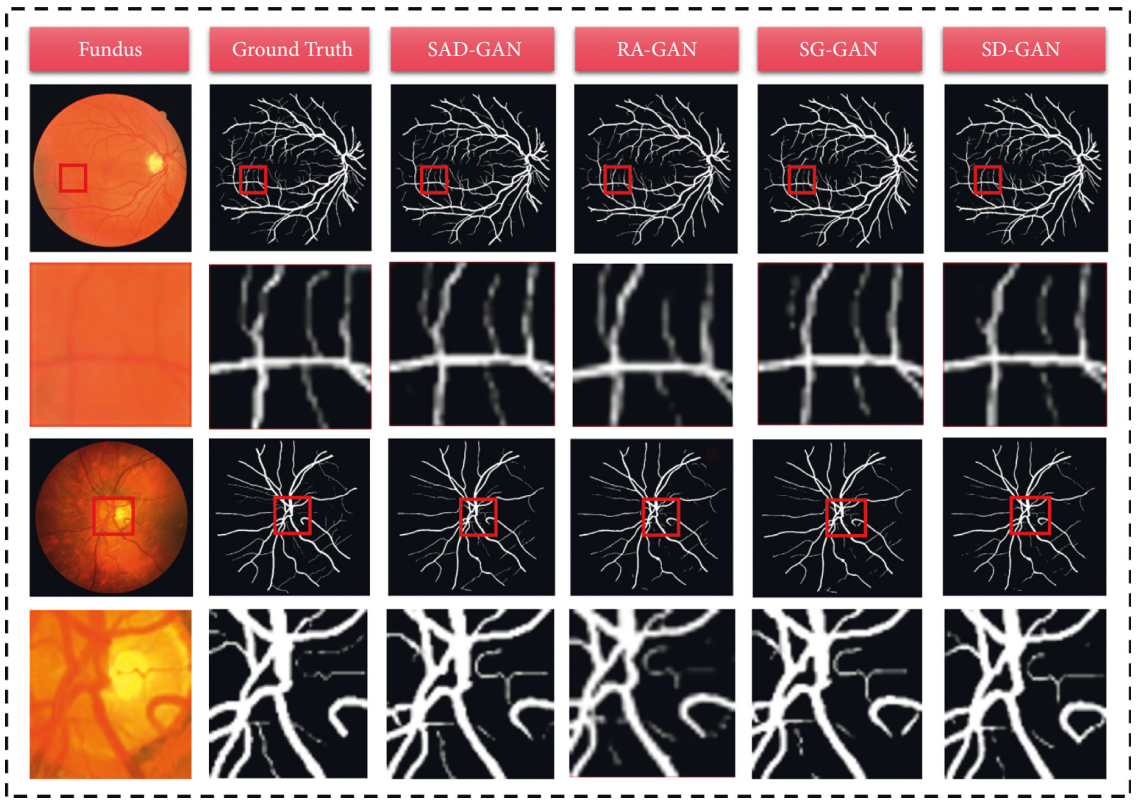


Figure 13: The first two lines and the last two lines are detailed segmentation diagrams of different GAN models on DRIVE and CHASE_DB1, respectively.

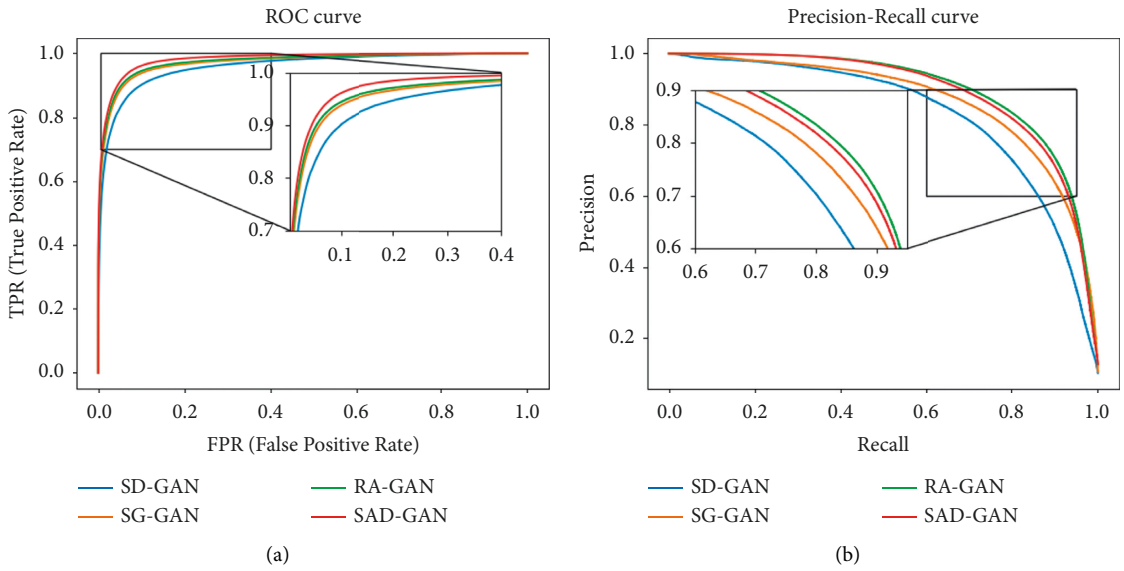


Figure 14: The curves of four GAN models. (a) ROC curves for DRIVE dataset. (b) PR curves for DRIVE dataset.

PR_AUC have been significantly improved, which are 0.9671, 0.9850, and 0.9002, respectively. According to Table 5, the index results of the algorithm proposed in this paper are not inferior to other algorithms. Particularly, the AUC indices exceed that of most of the algorithms listed in the tables. Therefore, the method proposed in this paper is accurate and robust for the segmentation of tiny retinal vessels.

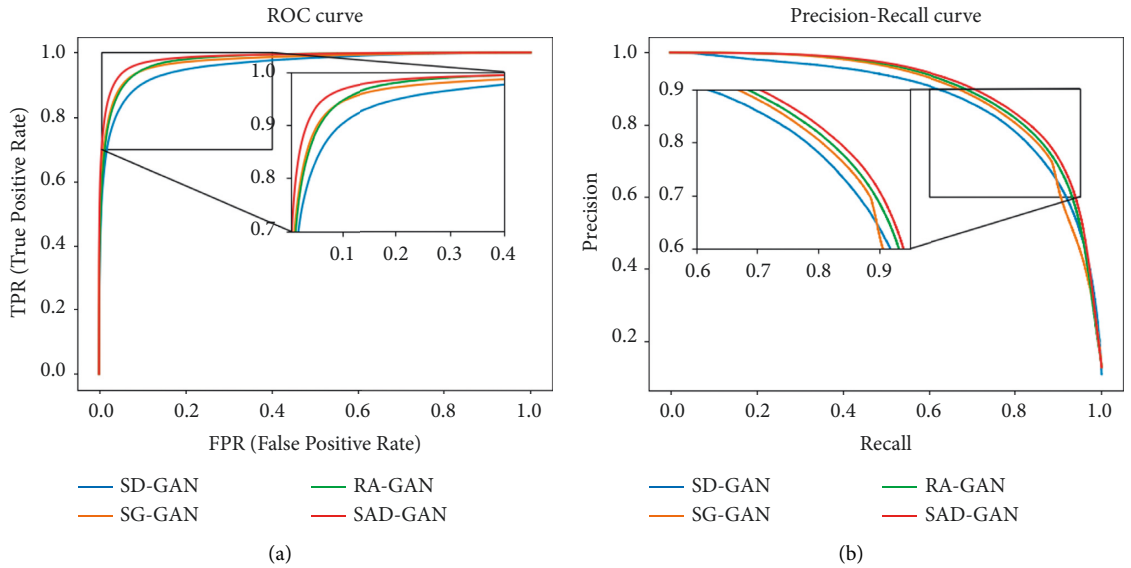


Figure 15: The curves of four GAN models. (a) ROC curves for CHASE_DB1 dataset. (b) PR curves for CHASE_DB1 dataset.

Table 2: The following table shows the segmentation results of different GAN models on DRIVE dataset in this paper.

Methods	SE	SP	ACC	ROC_AUC	PR_AUC
SD-GAN	0.7782	0.9811	0.9542	0.9585	0.8322
SG-GAN	0.7812	0.9779	0.9553	0.9711	0.8694
RA-GAN	0.8276	0.9802	0.9548	0.9737	0.9003
SAD-GAN	0.8422	0.9825	0.9567	0.9813	0.8928

Table 3: The following table shows the segmentation results of different GAN models on CHASE_DB1 dataset in this paper.

Methods	SE	SP	ACC	ROC_AUC	PR_AUC
SD-GAN	0.7893	0.9799	0.9646	0.9594	0.8698
SG-GAN	0.7939	0.9825	0.9651	0.9736	0.8829
RA-GAN	0.8126	0.9834	0.9648	0.9767	0.8927
SAD-GAN	0.8503	0.9850	0.9671	0.9839	0.9002

Table 4: The results of SAD-GAN and other algorithms on DRIVE dataset.

Methods	Year	SE	SP	ACC	ROC_AUC	PR_AUC
Wu et al. [40]	2018	0.7844	0.9819	0.9567	0.9807	—
Wang et al. [41]	2019	0.7940	0.9816	0.9567	0.9772	—
Cheng et al. [42]	2019	0.7672	0.9834	0.9559	0.9793	—
Yue et al. [43]	2019	0.8199	0.9762	0.9561	0.9796	—
Li et al. [44]	2020	0.7890	0.9799	0.9557	0.9774	0.8519
Ma et al. [45]	2020	0.7875	0.9813	0.9566	0.9794	—
Yang et al. [46]	2020	0.8340	0.9820	0.9560	0.9786	0.8821
Samuel and Veeramalai [47]	2020	0.7827	0.9821	0.9627	0.9789	
Li et al. [48]	2021	0.7931	**0.9896**	**0.9698**	0.9738	—
Xu et al. [49]	2021	0.8320	0.9885	0.9590	0.9713	
Lin et al. [21]	2021	0.8361	0.9740	0.9563	0.9805	
Yang et al. [22]	2022	0.8115	0.9780	0.9568	0.9810	—
Zhai et al. [23]	2022	0.7982	0.9818	0.9571	0.9811	—
Our proposed		**0.8422**	0.9825	0.9567	**0.9813**	**0.8928**

Table 5: The results of SAD-GAN and other algorithms on CHASE_DB1 dataset.

Methods	Year	SE	SP	ACC	ROC_AUC	PR_AUC
Wu et al. [40]	2018	0.7538	0.9847	0.9637	0.9825	—
Wang et al. [41]	2019	0.8074	0.9821	0.9661	0.9812	—
Cheng et al. [42]	2019	**0.8967**	0.9540	0.9488	0.9785	—
Li et al. [44]	2020	0.7798	0.9822	0.9620	0.9791	0.8291
Lin et al. [21]	2021	0.8488	0.9795	0.9668	0.9869	
Yang et al. [22]	2022	0.8075	0.9841	0.9664	**0.9872**	—
Our proposed		0.8503	**0.9850**	**0.9671**	0.9839	**0.9002**

5. Conclusions

Regular retinal blood vessel detection can timely detect vascular abnormalities, help patients to advance treatment, and prevent the further development and deterioration of the disease. Employing artificial intelligence to assist doctors to segment the retinal blood vessels has practical significance for rapid diagnosis. We propose an improved generative adversarial network for retinal image segmentation, which is intended to solve the problems of inaccurate and discontinuous retinal blood vessel segmentation. In the generator, we have introduced the residual structure and SE-Net to extract the global channel feature information without significantly increasing the number of parameters, which avoids overfitting and alleviation of the phenomenon of gradient disappearance, thus improving the segmentation accuracy. In the discriminator, we have replaced the standard convolution layer with the inception and dilated convolution layer, which can extract the high-level semantic information of different scale features of the deep network. This is important for accurately distinguishing the prediction maps from the ground truth. The channel attention and spatial attention that we have introduced also play an important role in the feature extraction, thus reducing the influence of the redundant information in the image classification by focusing on the useful information. By testing on the public datasets DRIVE and CHASE_DB1, our method achieves sensitivity = 0.8422, specificity = 0.9825, accuracy = 0.9567, ROC_AUC = 0.9813, PR_AUC = 0.8928 on the DRIVE dataset, whereas the sensitivity = 0.8503, specificity = 0.9850, accuracy = 0.9671, ROC_AUC = 0.9839, and PR_AUC = 0.9002 on the CHASE_DB1 dataset. The algorithm in this work has high accuracy and excels in some of the current state-of-the-art methods. It can segment the small retinal blood vessels better. The adversarial loss function that we have included plays a certain role in improving the segmentation accuracy in the overall training. We will further discuss the appropriate loss functions in future work on the experimental designs.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (61672386), the Academic Support Project for Top-notch Talents in Disciplines (Majors) of Universities in Anhui Province (gxbjZD2022042), and the Key Research and Development Plan of Anhui Province, China (2022a05020011).

References

[1] R. Xia, Y. Chen, and B. Ren, "Improved anti-occlusion object tracking algorithm using unscented rauch-tung-striebel smoother and kernel correlation filter," *Journal of King Saud University-Computer and Information Sciences*, 2022.

[2] J. Zhang, W. Feng, T. Yuan, J. Wang, and A. K. Sangaiah, "SCSTCF: spatial-channel selection and temporal regularized correlation filters for visual tracking," *Applied Soft Computing*, vol. 118, Article ID 108485, 2022.

[3] R. Muthukrishnan and M. Radha, "Edge detection techniques for image segmentation," *International Journal of Computer Science and Information Technology*, vol. 3, no. 6, pp. 259–267, 2011.

[4] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[5] H. G. Kaganami and Z. Beiji, "Region-based segmentation versus edge detection," in *Proceedings of the Fifth International Conference on Intelligent Information Hiding & Multimedia Signal Processing IEEE*, pp. 1217–1221, Kyoto, Japan, September 2009.

[6] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, no. 2, 2012.

[7] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3431–3440, Columbus, Ohio, USA, June 2015.

[8] H. Peiris, Z. Chen, and G. Egan, "Duo-SegNet: adversarial dual-views for semi-supervised medical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 428–438, Springer, Strasbourg, France, September 2021.

[9] L. Han, Y. Huang, H. Dou et al., "Semi-supervised segmentation of lesion from breast ultrasound images with attentional generative adversarial network," *Computer Methods and Programs in Biomedicine*, vol. 189, Article ID 105275, 2020.

[10] C. Decourt and L. Duong, "Semi-supervised generative adversarial networks for the segmentation of the left ventricle

in pediatric MRI," *Computers in Biology and Medicine*, vol. 123, no. 1, Article ID 103884, 2020.

[11] P. Luc, C. Couprie, and S. Chintala, "Semantic segmentation using adversarial networks," 2016, https://arxiv.org/abs/1611.08408.

[12] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "SegAN: adversarial network with multi-scale L1 loss for medical image segmentation," *Neuroinformatics*, vol. 16, no. 3-4, pp. 383–392, 2018.

[13] P. Isola, J. Y. Zhu, and T. Zhou, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, Honolulu, HI, USA, July 2017.

[14] Y. Zhou, B. Wang, X. He, S. Cui, and L Shao, "DR-GAN: conditional generative adversarial network for fine-grained lesion synthesis on diabetic retinopathy images," *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 1, pp. 56–66, 2022.

[15] S. A. Kamran, K. F. Hossain, and A. Tavakkoli, "RV-GAN: segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 34–44, Springer, Athens, Greece, October 2021.

[16] A. Tavakkoli, S. A. Kamran, K. F. Hossain, and S. L. Zuckerbrod, "A novel deep learning conditional generative adversarial network for producing angiography images from retinal fundus photographs," *Scientific Reports*, vol. 10, no. 1, pp. 21580–21615, 2020.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, Lima, Peru, October 2015.

[18] H. Ding, X. Cui, L. Chen, and K. Zhao, "MRU-Net: a U-shaped network for retinal vessel segmentation," *Applied Sciences*, vol. 10, no. 19, p. 6823, 2020.

[19] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *Journal of Medical Imaging*, vol. 6, no. 01, 2019.

[20] N. Ibtehaz and M. S. Rahman, "MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Networks*, vol. 121, pp. 74–87, 2020.

[21] Z. Lin, J. Huang, Y. Chen et al., "A high resolution representation network with multi-path scale for retinal vessel segmentation," *Computer Methods and Programs in Biomedicine*, vol. 208, Article ID 106206, 2021.

[22] X. Yang, Z. Li, Y. Guo, and D. Zhou, "DCU-Net: a deformable convolutional neural network based on cascade U-Net for retinal vessel segmentation," *Multimedia Tools and Applications*, vol. 81, no. 11, pp. 15593–15607, 2022.

[23] Z. L. Zhai, S. Feng, L. Yao, and P. Li, "Retinal vessel image segmentation algorithm based on encoder-decoder structure," *Multimedia Tools and Applications*, pp. 1–13, 2022.

[24] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, Salt Lake City, UT, USA, June 2018.

[25] S. Woo, J. Park, and J. Lee, "Cbam: convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, 2018.

[26] Y. Li, J. Yang, J. Ni, A. Elazab, and J. Wu, "TA-Net: triple attention network for medical image segmentation," *Computers in Biology and Medicine*, vol. 137, no. 2, Article ID 104836, 2021.

[27] F. Wang, M. Jiang, and C. Qian, "Residual attention network for image classification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3156–3164, Honolulu, HI, USA, July 2017.

[28] J. Fu, J. Liu, and H. Tian, "Dual attention network for scene segmentation," in *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3141–3149, Long Beach, CA, USA, June 2019.

[29] Q. Wang, B. Wu, and P. Zhu, "ECA-Net: efficient channel attention for deep convolutional neural networks," in *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11531–11539, Seattle, WA, USA, June 2020.

[30] Z. Huang, X. Wang, and L. Huang, "CCNet: criss-cross attention for semantic segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 603–612, June 2019.

[31] K. Li and B. Wang, "DAR-Net: dense attentional residual network for vehicle detection in aerial images," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 6340823, 2021.

[32] H. Lu, S. Tian, L. Yu et al., "DCACNet: dual context aggregation and attention-guided cross deconvolution network for medical image segmentation," *Computer Methods and Programs in Biomedicine*, vol. 214, Article ID 106566, 2022.

[33] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, https://arxiv.org/abs/1511.06434.

[34] A. You, JK. Kim, IH. Ryu, and T. K. Yoo, "Application of generative adversarial networks (GAN) for ophthalmology image domains: a survey," *Eye and Vision*, vol. 9, no. 1, pp. 6–19, 2022.

[35] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, pp. 4278–4284, 2017.

[36] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," 2015, https://arxiv.org/abs/1511.07122.

[37] C. Szegedy, V. Vanhoucke, and S. Ioffe, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, Honolulu, HI, USA, July 2016.

[38] J. Staal, M. D. Abramoff, M. Niemeijer, M. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 501–509, 2004.

[39] C. G. Owen, A. R. Rudnicka, R. Mullen et al., "Measuring retinal vessel tortuosity in 10-year-old children: validation of the computer-assisted image analysis of the retina (CAIAR) program," *Investigative Opthalmology & Visual Science*, vol. 50, no. 5, p. 2004, 2009.

[40] Y. Wu, X. Yong, and S. Yang, "Multiscale network followed network model for retinal vessel segmentation," in *Proceedings of the International Conference on Medical Image Computing & Computer-assisted Intervention*, pp. 119–126, Springer, Athens, Greece, October 2018.

[41] B. Wang, S. Qiu, and H. He, "Dual encoding U-Net for retinal vessel segmentation," *Proceedings of the International conference on medical image computing and computer-assisted intervention*, vol. 11764, pp. 84–92, 2019.

[42] Y. Cheng, M. Ma, L. Zhang, C. Jin, L. Ma, and Y. Zhou, "Retinal blood vessel segmentation based on densely

connected U-Net," *Mathematical Biosciences and Engineering*, vol. 17, no. 4, pp. 3088–3108, 2020.

[43] K. Yue, B. Zou, Z. Chen, and Q. Liu, "Retinal vessel segmentation using dense U-Net with multiscale inputs," *Journal of Medical Imaging*, vol. 6, no. 03, 2019.

[44] H. Li, Y. K. Wang, and C. Wan, "MAU-Net: a retinal vessels segmentation method," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 2020, p. 1961, 2020.

[45] Y. Ma, X. Li, X. Duan, Y. Peng, and Y. Zhang, "Retinal vessel segmentation by deep residual learning with wide activation," *Computational Intelligence and Neuroscience*, vol. 2020, no. 1, Article ID 8822407, 11 pages, 2020.

[46] T. Yang, T. Wu, L. Li, and C. Zhu, "SUD-GAN: deep convolution generative adversarial network combined with short connection and dense block for retinal vessel segmentation," *Journal of Digital Imaging*, vol. 33, no. 4, pp. 946–957, 2020.

[47] P. M. Samuel and T. Veeramalai, "VSSC Net: vessel specific skip chain convolutional network for blood vessel segmentation," *Computer Methods and Programs in Biomedicine*, vol. 198, Article ID 105769, 2021.

[48] Z. Li, M. Jia, X. Yang, and M. Xu, "Blood vessel segmentation of retinal image based on dense-U-Net network," *Micromachines*, vol. 12, no. 12, p. 1478, 2021.

[49] S. Xu, Z. Chen, W. Cao, F. Zhang, and B. Tao, "Retinal vessel segmentation algorithm based on residual convolution neural network," *Frontiers in Bioengineering and Biotechnology*, vol. 9, Article ID 786425, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 6319197, 10 pages
https://doi.org/10.1155/2022/6319197

Research Article

Estimation of Parameters and Pooling in Nonlinear Flooding Event Scenarios with Bayesian Model

Fuad S. Alduais[1] **and Taghreed M. Jawa**[2]

[1]*Department of Mathematics, College of Science and Humanities in Al-Kharj, Prince Sattam Bin Abdulaziz University, 11942 Al-Kharj, Saudi Arabia*
[2]*Department of Mathematics and Statistics, College of Sciences, P.O. Box 11099, Taif University, Taif 21944, Saudi Arabia*

Correspondence should be addressed to Fuad S. Alduais; f.alduais@psau.edu.sa

Received 15 June 2022; Accepted 1 August 2022; Published 28 August 2022

Academic Editor: Heng Liu

Water-related tragedies are the highest common of all proven natural calamities and pose severe attacks on people and on socioeconomic status development. Due to the obvious controversy surrounding, their volume, location and time of incidence, geological swaths, and geophysical interrelations, flood events are difficult to completely control. Hence, complete flood prevention is always considered to be a viable choice. The specialized flood occurrences are investigated by developing a structural measure. In this paper, nonlinear flood event circumstance is determined by using a statistical Bayesian parametric approach for parameter estimation. A popular tool for estimating a flood design is model of nonlinear flood event. Nonlinear flood event models are subjected to a Bayesian technique for estimating parameter. The approach is based on the minimization function of square for models with nonlinear calculated peak discharges in terms of parameters. The observed and calculated peak discharges for numerous storms in the watershed, data on the pattern of error observed, and previous information on values of parameter all influence this objective function. The subsequent matrix for covariance is a measure of the calculated parameters' accuracy. Rainfall and runoff data from a Harvey River sample are used in this study to show the process.

1. Introduction

Water-related disasters are the most prevalent of all natural disasters documented, posing major threats to people and socioeconomic growth. Between 1900 and 2006, floods were responsible for more than 30% of all natural disasters, killing more than 19% of all people and harming more than 48% of the population [1]. Floods account for 26% of all natural catastrophe expenses, with water-related calamities accounting for 72% of all costs. Change in climate, land use changes, sea level increase, deforestation, and development of population in flood-prone areas are all expected to worsen these losses in the future, pushing the global flood disaster population to two billion people [2].

For a number of reasons, optimal forecasting of flood and sustainable risk management systems for flood have been promoted as flood preparation strategies. Floods are difficult to entirely regulate because of the uncertainties surrounding their volume, time, and location of occurrence, geographical expanse, and geophysical interactions. As a result, total flood protection is not always regarded as a practical option [3]. Traditional flood control, which generally consists of structural protection measures such as dams and levees, focuses on changing the characteristics of a flood to minimize peak heights and geographical extents [4]. Although flood danger is reduced by structural solutions such as dams and embankments, it cannot be entirely eliminated. Furthermore, these interventions are impractical in certain region like remote mountain areas, ineffective for all flood procedures like freezing lake outburst floods, and have negative environmental consequences [5]. Moreover, as seen in the USA, aging flood control infrastructure has significant costs and implications in flood setup that does not provide the desired level of fortification or is vulnerable

to collapse. Furthermore, specific flood events are analyzed by designing a structural measure, such as a 1% annual exceeding [6], which is difficult to implement because it is done manually, and increased hydrologic uncertainty has resulted from channel alterations, land use changes, and climate change, making floods less expectable. As a result, structural improvements are almost always doomed to fail [7].

The pooling problem has been solved using sophisticated methods. A weighted linear combination of parameter estimations is used in the bootstrapped estimates. The validity of pooling is explicitly assumed in this procedure [8]. Parameter variation is represented in a Bayesian technique by a random parameter model, with storm event parameters chosen from a hyper-distribution that is assumed to be multivariate normal [9]. A variety of models, spanning from fixed to random parameters, can be utilized depending on the model parameterization [10].

Furthermore, there is probable cause to believe that flood event model parameters may be storm-dependent. The model's forecasting power will be harmed unless the parameters have a tight link to easily recognizable elements of the storm (e.g., storm track or rainfall temporal pattern) [11]. The pooling techniques are not useful, but they do presuppose the presence of a universal set of parameters that can be applied to all storms. Individual discrepancies in parameter estimations might be due to a single estimation factor [12].

In NFEMs, manual methods for parameter estimation include trial and error, parameter interaction diagrams, and nonlinear regression modeling [13]. These approaches, with the exception of the method, focus on fitting calculated and pragmatic overflow hydrographs for a single storm. For each storm occurrence, this generates a single set of parameter estimations [14]. When numerous storms are accessible for estimation of parameter, a pooling approach must be used to provide parameter estimation with a single set for the watershed.

The rest of the section is summarized as follows: Section 2 reveals the background of the parameter estimation model; Section 3 details the proposed methodology of Bayesian mechanism in nonlinear flood event detection; Section 5 explains the result and discussion, and finally Section 5 concludes the paper.

2. Background

2.1. Parameter Estimation and Interference. The calibration of nonlinear flood event is fitted by observed and computed hydrograph peaks with parameter a and m discrete storm event. The nonlinear flood event model is defined using the regression

$$r_t = f\left(i_t, v\right) + \varepsilon_t, \quad t = 1, 2, \ldots, m. \tag{1}$$

The input vector is represented as i_t with the observed peak discharge r_t, v is the unknown parameter with vector, and F is the response function which is continuously differential with vector and creates error randomly. The response function (2) of the nonlinear flood event is

$$f\left(i_t, v\right) = x'_{tl}, \qquad v = \sum_{i=1}^{a} x_{tl} v_l, \tag{2}$$

where x_{tl} is component of i_t and transpose matrix of the vector.

The estimation of parameter with statistical approach is based on the literature of hydrologic with least square method. The function of sum of square for estimating the least square with minimum vector is expressed as

$$S(v) = [r - f(v)]' [r - f(v)], \tag{3}$$

$$S(v) = \sum_{t=1}^{m} \left[r_t - f\left(i_t, v\right)\right]. \tag{4}$$

The column vector and the component of r and $f(v)$ are denoted by r_t and $f(i_t, v)$. The response surface of the contour is denoted by the sum of square function with the linear parameter function for flood event, which is quadratic to the surface and forms hyper-ellipsoid. The unbiased estimation parameter is obtained with the variance, and estimation is based on the interference. The three assumptions should be satisfied for unbiased estimation.

(i) Specifying the function of model response correctly

(ii) The input vector component is measured without error and is nonstochastic

(iii) The error ε_t obtained is distributed randomly with arbitrary variable normally with mean and constant unknown variance σ_t^2

The unbiased vectors hold the three assumptions with least square, distributed normally, and estimating minimum variance. The preceding properties only hold asymptotically for models with nonlinear parameters. The response surface is nonquadratic for smaller samples, and least squares estimators may be severely skewed and non-normally disseminated, and have variances greater than the smallest feasible variance [15]. The asymmetry of a estimation of least squares for a nonlinear model is the most essential aspect of its distribution. The degree to which this behavior changes depends on the data set combination, and there are no rules on how large a sample must be to approximate one-dimensionality.

2.2. Parameter Transformation. Under parameter modifications, the response surface's shape and the sampling characteristics of estimating the least squares are not invariant. It may be feasible to discover a parameterization that produces more quadratic surface response of contours and estimators that approximate the asymptotic requirements. The problem of joint confidence is obtained by simplifying the parameterization for the parameter model.

The parameterization of transformation is obtained alternatively represented by

$$w = t(v). \tag{5}$$

The new parameter w is implied with the original parameter without input vector i_t. The predicted value is provided for the reparametrized model with the response observed with the original model.

The transformation of parameter has two classes generally. The transformation with expected value and parameter with stability is proposed in this paper. The set of input vector is chosen for the analysis, and expected value is used $w_1, w_2, \ldots, w_r$.

$$wi = \text{fit}^*, \quad v, t = 1, 2, \ldots, r. \tag{6}$$

The original parameter v is eliminated after transformation of equation. The transformation with the second class is obtained for the parameter estimation.

$$w_i = w_i^i \text{ for } \gamma \neq 0, \tag{7}$$

$$w_i = \ln(u_i) \text{ for } \gamma = 0. \tag{8}$$

The transformation parameter has the advantage with constant random variable which enhance the contour form of the surface of response and the possessions of sampling with estimating the least square for the combination of model and data set. When compared to the original model parameterization, one drawback of predictable value alterations is the numerical complication of their formulations [16]. The reparametrized model response function has many converted parameters, and the difficulty of these purposes increases as the number of parameter model increases. Another disadvantage is that it is not always possible to solve the equations to create terminologies for the inventive values just in terms of the changed constraints.

2.3. Nonlinearity Measure. The sampling properties are assessed using the nonlinearity measure for the estimation of least square. The asymptotic bias is included for the calculation of the bias and the intrinsic curvature measures and effect of parameter curvature. The model specified for the identification of the bias measure and individual parameter estimation is related to parameter transformation.

The derivate matrix element is given as

$$V' = \partial f(i_t, \widehat{w}) dw_i. \tag{9}$$

The second derivative is given in

$$V'' = \partial^2 f(i_t, \widehat{w}) dw_i dw_j. \tag{10}$$

The function of J is defined as

$$J = V'.V''. \tag{11}$$

The element of W is given as

$$W = \sum_t V'.V''. \tag{12}$$

The bias measure with the definition of Hougaard is defined by

$$E(\widehat{w}_i) - w_i = -\frac{1}{2}\sigma^2 \sum W.J, \tag{13}$$

where $E(\widehat{w}_i)$ is the expected value of $\widehat{w}_i$.

The estimation of variance is given by

$$\sigma^2 = \frac{s(\widehat{w})}{n - p}. \tag{14}$$

The nonlinear behavior is indicated using the percentage bias

$$B\% = \frac{100(E(\widehat{w}_i) - w_i)}{\widehat{w}_i}. \tag{15}$$

Each individual parameter with the nonlinear behavior measure is estimated by the asymptotic moment, which is given by

$$E[\widehat{w}_i - E(\widehat{w}_i)]^3 = -\sigma^4 \sum_{qrs} J_{iq}.J_{ir}.J_{is}.\left(W_{qrs} + W_{rqs} + W_{sqr}\right). \tag{16}$$

The skewness with direct measure is given by

$$Sk = \frac{E[\widehat{w}_i - E(\widehat{w}_i)]^3}{\left(\sigma^4 g_i\right)^{3/2}}, \tag{17}$$

where g_i is nonlinearity indicator.

The shape of the locus solution is refereed using the intrinsic curvature, which is described using

$$f(w) = [f(i_1, w), \ldots, f(i_1, w)]. \tag{18}$$

The curvature measure for the element V' and V'' is assumed by the standard radius of the element:

$$A = [Q'][L'V''L]. \tag{19}$$

The equation is called square bracket multiplication. The decomposition of V' is given by(19).

$$V' = qR. \tag{20}$$

The parameter effect of the array accelerated in the curvature is denoted as A_{PE}, and intrinsic acceleration is denoted as A_{IN}. The maximum curvature of parameter effect and intrinsic culture is defined by

$$\text{intrinsic curvature} = \max\left\|u' A_{IN} u\right\|, \tag{21}$$

$$\text{parameter effect curvature} = \max\left\|u' A_{PE} u\right\|, \tag{22}$$

where u is unit vector, and norm of vector is represented using the vertical bars.

3. Methodology

3.1. Hydraulic Model. The 1-D modeling technique for flood propagation is still frequently utilized in engineering practice because of its ease of implementation, low processing time, and good real-time operational efficiency. Because of the long channel length, limited floodplain, and low sinuosity, as well as the fact that our flood forecasting is primarily focused on discharge and stage at cross sections rather than flood inundation, a 1-D modeling approach is

used to achieve real-time flood forecasting efficiently and adequately in this study. As a result, the hydraulic model used here is the basic 1-D unstable open-channel flow model, which is described using the Saint Venant formulation as follows:

$$\frac{\partial A}{\partial T} + \frac{\partial D}{\partial L} = d_L, \tag{23}$$

$$\frac{\partial D}{\partial T} + \frac{\partial}{\partial L}\left(\frac{\alpha D^2}{A}\right) + gA\frac{\partial S}{\partial L} + g\frac{D|D|n^2}{AR^{4/3}} = 0. \tag{24}$$

T is time, L is distance of space along the channel, D is discharge, S is stage, α is correction factor of momentum, A is area of cross section, R is hydraulic radius, d_L is lateral discharge, and n is coefficient of Manning's roughness.

These are hyperbolic partial differential equations that do not have an analytical solution. With sufficient beginning and boundary conditions, the equations may be numerically solved using the four-point implicit finite-difference approach. The four-point implicit finite-difference technique has the advantages of being unconditionally stable, exceedingly robust, and quick to compute. The scheme's fundamental flaw is that it does not operate well when the river topography is complex, which is not the situation with this project. Inflow discharges occur at the upstream and lateral boundary conditions, whereas outflow phases occur at the downstream boundary circumstances.

It is important to mention that the aforementioned fundamental governing equations' form implies the assumption of uniform velocity distribution over the whole cross section. When out-of-bank flow occurs in compound channels, the difference in water depth and flow resistance between the main channel and the floodplains causes a flow velocity differential between these subsections. To mimic the various flow characteristics in subsections, greater momentum correction factors, transverse changes in Manning's roughness coefficient, and momentum transfer between main channel and floodplains should be addressed in the governing equations. The cross sections are simplified as simple channels, eliminating the impact of floodplains, because the geometry of the cross sections in the study channel and the floodplains are small or not formed in the mountain valleys. As a result, the 1-D hydraulic model's essential governing equations maintain the traditional form, with an overall Manning's roughness coefficient for each cross section. Another benefit of employing an overall Manning's roughness coefficient in hydraulic models is the simplicity with which it may be updated during the assimilation process.

The collective resistance to open-channel flow is represented by Manning's roughness coefficient. Its value is influenced by flow circumstances such as the area of submerged vegetation and flow turbulence intensity, as well as physical parameters such as bed geology and cross-sectional geometry. The overall Manning's roughness coefficient for a cross section is a composite number that includes all elements' contributions to flow resistance and may be calibrated using hydrological measurements. The Manning's roughness coefficients in the flood-forecasting model should be permitted to change both geographically and temporally, given the longitudinal fluctuation of a channel's physical properties and the unstable flooding flow. A method for segmenting a river reach is provided, with each segment defined by two adjacent hydrological stations and a series of spatially unique cross sections. Although the cross sections within a river segment have the same roughness coefficient, each river segment has its unique Manning's roughness coefficient. The Manning's roughness coefficient values are calibrated using a trial-and-error technique based on historical flow data and then updated using real-time stage measurements from hydrological stations to represent the unstable flood's instantaneous changes in flow resistance.

3.2. Bayesian Framework. A competent hydrologist is typically aware of the potential values, or range of values, of model parameters before any observed data are collected. This knowledge might be based on personal preferences, theoretical considerations, or other factors or an objective assessment of indirect data, such as model parameter regressions on catchment feature. To give reliable parameter estimates, Bayesian estimation combines pre-existing knowledge with observed data (sample information). The unknown parameter vector u is regarded as a random variable and distributed using a probability density function in this approach. The information of posterior probability density function is used to combine the Bayes theorem.

$$f(\theta|O) = cL(O|\theta)f(\theta). \tag{25}$$

O is the vector of observation, the likelihood function $L(O|\theta)$ which contains the information sample. The prior information is given as $f(\theta)$, and c is the constant of normalization which need a density function $f(\theta|O)$. By locating the typical value of posterior distribution, various estimations of θ are analyzed.

Because the exact form of $f(\theta|O)$ is too complex to calculate, a strict request of Bayesian approximation is not practicable for the types of replicas and sample sizes under consideration.

3.3. Bayes Theorem. A Bayesian technique to hydrological modeling offers a persuasive mechanism for integrating precise parameter uncertainty estimations with readily available expert knowledge. Bayesian statistics recognizes preceding data based on historic data or knowledge form expert, as well as data composed via investigation and opinion, as two types of data for knowledge about unidentified parameters. The unknown vector parameter θ is modeled as a random variable with a probability density function that represents uncertainty. This function of density is designated $P(\theta)$ and detains all known knowledge about the parameters prior to gathering data. The density is represented by $P(\theta|Q)$ prior to gathering data Q. Prior and posterior densities are the terms used to describe these densities. Prior densities are less concentrated than posterior

densities, which are analyzed using the knowledge about the θ is acquired through the process of collecting data is reflector with a decrease in uncertainty. By using Bayes theorem, the process of updating posterior distribution from prior distribution is given by

$$P(\theta|Q) = \frac{P(\theta|Q)P(\theta)}{P(Q)}. \tag{26}$$

The proportionality constant (27) is represented by $P(Q)$ which is required for

$$\int P(\theta|Q)d\theta = 1. \tag{27}$$

The likelihood function of the sample is given as $P(\theta|Q)$

$$P(\theta|Q) = \prod_{l,j} 2\pi\sigma_l^2 \exp\left(-\frac{[Q - x(r;\theta)]}{2\sigma_l^2}\right). \tag{28}$$

The information available about the vector parameter θ is contained in the posterior distribution. The posterior probability distribution is summarized by reducing the interference of the statistical Bayesian.

3.4. Markov Chain Monte Carlo (MCMC) Method. It is difficult to describe the subsequent dispersal by direct computation for even modestly complicated issues with realistic prior belief conditions. MCMC techniques provide an alternate way. This is a method for creating samples from high-dimensional distributions, such as the posterior distribution P(OIQ). The goal is to obtain a big enough sample to correctly characterize any desired aspect of the posterior.

The basic concept is to begin with a random beginning value o and produce a sequence of reliant on parameter values $\{\theta^t : t = 1, 2, \ldots\}$ from a well-prepared Markov chain: in other words, a random walk across the parameter state space. To create this arbitrary walk, need to design a transition density function that describes the move $\theta^t \longrightarrow \theta^{t+1}$ in such a way that the chain's observed values converge in distribution to the posterior. Before the chain's limiting distribution is achieved, there is usually an unstable transient phase. The chain is considered to have congregated in dispersal after this phase. The iterations collected before convergence are eliminated, leaving a reliant on sample drawn from the subsequent. If the chain is protracted after convergence to gather a large enough sample, the replicated standards of the chain may be utilized to summarize the characteristics of the posterior distribution.

The most commonly used Markov Chain Monte Carlo method is metropolis hasting algorithm. The transition form of $\theta' = \theta$ to $\theta^{t+1} = \theta'$ is described using the probability function. $P(\theta, \theta')$ is constructed using metropolis hasting algorithm, which is given as follows:

(i) The candidate value θ' is generated from the probability distribution $Q(\theta, \theta')$ for essentially arbitrary

(ii) The probability $R(\theta, \theta')$ is accepted and moved the value to $\theta^{t+1} = \theta'$

(iii) With the probability 1-$R(\theta, \theta')$, $\theta = \theta^{t+1}$ is set, if the move is rejected

The transition probability with Markov chain is given by

$$\begin{aligned} P(\theta, \theta') &= Q(\theta, \theta').R(\theta, \theta'), \\ P(\theta, \theta'') &= \int_{\theta''} Q(\theta, \theta'')[1 - R(\theta, \theta'')]d\theta'', \\ P(\theta, \theta'') &= 1 - \int_{\theta''} Q(\theta, \theta'')[1 - R(\theta, \theta'')]d\theta''. \end{aligned} \tag{29}$$

Let $R(\theta, \theta')$ be defined for the convenience of $P(\theta|Q) = \pi(\theta)$

$$R(\theta, \theta') = \begin{cases} \min\left(\dfrac{\pi(\theta')Q(\theta', \theta)}{\pi(\theta)Q(\theta, \theta')}, 1\right), & \text{if } \pi(\theta)Q(\theta, \theta') > 0, \\ 1, & \text{if } \pi(\theta)Q(\theta, \theta') = 0. \end{cases} \tag{30}$$

Then,

$$\pi(\theta)\,Q(\theta, \theta') = \pi(\theta')\,Q(\theta', \theta). \tag{31}$$

The reversibility condition is defined by the above equation which has the limited distribution for $\pi(\theta)$ with sufficient condition for the provided chain $Q(\theta, \theta')$ irreducibly chosen.

The probability with univariant density function is given by

$$R_t(\theta, \theta^*) = \min\left(\frac{\pi(\theta^*)\,Q(\theta^*, \theta)}{\pi(\theta)\,Q(\theta, \theta^*)}, 1\right). \tag{32}$$

3.5. Particle Filter. The state variables and constraints are characterized by likelihood distributions that provide for their uncertainties to account for the state-space formulation's stochastic character. Because the state-space preparation is recursive, it may be used to absorb experimental data as it becomes accessible in a sequential manner. From time t-1 through t, the conditional probability recursive equation may be used to represent the association among observational data and model outputs. The conditional probability is given

$$P(u_t, \theta_t/v_t) = \frac{P(v_t/u_t, \theta_t)}{P(v_t/v_{t-1})} P(u_t, \theta_t/v_t), \tag{33}$$

$P(u_t, \theta_t/v_t)$ is probability distribution of posterior model. $P(u_t, \theta_t/v_{t-1})$ is probability distribution of prior mode at time t, which is distributed as time $t-1$. $P(v_t/u_t, \theta_t)$ is the likelihood function. $P(v_t/v_{t-1})$ is normalized constant.

Analytic functional formulations of the likelihood distributions of variables and constraints for complex systems are difficult to come by. To approximate the distributions, the PF uses an ensemble particle. When the number of particles is high enough, particles created via Monte Carlo simulation can reasonably resemble real probability densities. The likelihood distribution of posterior state variable model and constraints at time t may be estimated using the ensemble of particles, as follows:

$$P\left(u_t, \frac{\theta_t}{v_t}\right) = \sum_{i=1}^{N} w_t \delta\left(u_t - u_t', \theta_t - \theta_t'\right). \quad (34)$$

N is number of particle, w_t is weight based on ith particle, and $\delta(.)$ is Dirac delta function.

The two phases of recursive model approaches based on the particle filter are model forecast and filter correction:

(1) *Model Forecast Phase*. For each particle, from time $t-1$ until time t, each set of state variables and parameters in the dynamic nonlinear model is performed. Each particle represents the prior probability distribution at time t following integration, as well as the subsequent probability distribution of a set of state variable quantity and parameter model at time $t-1$ prior to combination.

(2) *Filter Correction Phase*. Filter correction consists of three steps such as likelihood computation, resampling, and agitation. The goal of probability computation is to inform each particle's weight depending on available data. The Gaussian likelihood function for calculating the likelihood value of each particle is given by (33).

$$w_t = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{\left(u_t - v_t\right)^2}{2\sigma^2}\right), \quad (35)$$

where u_t is prior state variable at time t, v_t is state variable observation at time t, and σ^2 is standard deviation.

By normalizing the likelihood value, the particle weight is determined by summing all the likelihood value of the particle.

$$W_t = \frac{w_t^i}{\sum_{i=1}^{N} w_t^i}, \quad (36)$$

where W_t is normalized weight.

The random noise of the perturbed particles incorporated to the parameter model is given by

$$\varnothing_{t+1} = \varnothing_t + \epsilon, \quad (37)$$

where ϵ is Gaussian distributed with random noise.

After determining the normalized weights, resampling is done. Particle degeneracy occurs when the majority of particles have insignificant weights and just a few elements are efficient in the strainer. It is necessary to resample the particles in order to remove the small weighted elements and develop auspicious new elements based on the high weighted elements in order to decrease the meaningless calculations and more precisely predict the current time step's condition.

To keep particle variety, perturbation is used. Though resampling can reduce immorality, the variety of elements may suffer as a result of sample impoverishment, in which some particular elements with great weights are reproduced multiple times. It is a waste of time to keep running the same model with the same elements. The disturbed sub-divisions can move about in space in a stochastic way to follow their development over time. By including random noise in the model parameter, the particles can be disturbed.

3.6. Parameter Estimation. The assortment of an objective function that shows quality of fit is required for parameter estimation by comparing calculated and experimental hydrographs. This function must be appropriate for the model's intended application. Reproduction of experimental flood hydrograph peaks is regarded to be extremely important in many flood investigations. As a result, the suggested goal function focuses on the fitting of calculated and pragmatic hydrograph peaks.

3.7. Objective Function. Consider a rainfall runoff data issue with a p parameter model and n discrete storm occurrences. Assume the observed and calculated peak discharges are correlated as follows:

$$O = h(\theta) + u, \quad (38)$$

where O is the observed vector peak discharge, $h(\theta)$ is the computed vector of the peak discharge, θ is the vector parameter, and R is the covariance and zero mean with Gaussian error. The prior information is pooled with the sample information and the parameter vector with prior mean and covariance M:

(i) First, when just the prior mean and standard deviation of u are known, the supposition of a Gaussian prior probability density function contributes the least volume of unnecessary data to the valuation issue.

(ii) Second, the Gaussian assumption is not dangerous in repetition and its influence is smaller for large n.

(iii) Third, the section that contains the inverse of M works as a constraint on the components of u m. However, when additional floods are included in the sample data, the influence reduces.

(iv) Fourth, timing errors caused by the rainfall with poor synchronization and runoff time series have little effect on parameter estimations.

3.8. Algorithm of Optimization. Minimization necessitates the employment of a nonlinear, unconstrained optimization technique for representations in which $h(u)$ is a function of nonlinear u. For the sort of issue at hand, the algorithms provided based on the Gauss technique are preferable than flexible metric approaches and direct search methods. As a result, the aforementioned optimization issue was solved using the Gauss–Marquardt method. The best aspects of the steepest descent and Newton techniques are combined in this method. The basic repetition in these Gauss–Marquardt methods is

$$x_{i-1} = x_i - AG_i. \quad (39)$$

The matrix A is defined as

$$A = (M + \mu B)^{-1}. \tag{40}$$

Gauss approximation represented M, and μ is the chosen scalar with positive derivatives. The optimization of gradient vector is described as

$$G_i = -2L^{-1}(\theta_0 - \theta) - 2DR(x - h(\theta)). \tag{41}$$

3.9. Estimation of Pooling. NFEMs explain a strategy for parameter estimate across storm occurrences. This approach includes the concept that pooling is justifiable, as well as the common parameter model (CPM) with parameter vector. ts is a vector holding storm-specific parameters, whereas t is a vector containing parameters that are common to all storm occurrences in the CPM. If pooling is not allowed, a more generic data model is needed.

The competing model is selected by specifying the beliefs prior to the parameter, and common parameter model with prior odds is specified for easy processing.

MC = common parameter model is true and M_G = general parameter model is true. The odd prior is represented as $P(M_C)/P(M_G)$. The odd posterior is represented as M_C and M_G as given in

$$\frac{P(M_C/Q)}{P(M_G/Q)} = \frac{P(Q/M_C)}{P(Q/M_G)}.\frac{P(M_C)}{P(M_G)}. \tag{42}$$

The Bayes quantity factor for representing the common and general parameter is given as B_{CG}, which is used to select the parameter among the hydrological model. The marginal probability of B_{CG} is denoted as

$$P\left(\frac{Q}{M}\right) = \int_\theta P\left(\frac{Q}{M}, \theta\right) P(\theta) \mathrm{d}\theta. \tag{43}$$

The equation is related to Monte Carlo average and concentrated with posterior mode distribution. The harmonic mean estimation is given as (42).

$$P\left(\frac{Q}{M}\right) = \left[N^{-1} \sum P\left(\frac{Q}{\theta}\right)^{-1}\right]^{-1}. \tag{44}$$

The large fluctuation in the likelihood value is estimated occasionally. The sophisticated technique is applied for potentially generating the finer answer.

3.10. Uncertainty in Flood Estimation. There are two sorts of prediction failures in flood forecasting: (a) the system may fail to give a warning for a flood event, and (b) it may issue a warning for an event that does not materialize due to an error of commission. A flood in the first situation might result in the loss of lives, infrastructure, and property [17]. People may lose faith in the forecast in the second situation and fail to respond to the following warning. To reduce the chances of any form of failure, it is vital to analyze and communicate forecasting mistakes and uncertainties.

3.11. Source of Uncertainty and Error. In flood forecasting, there are several sources of uncertainty such as input data uncertainty, model uncertainty, and model parameter uncertainty. The several causes of mistakes in the flood warning process have been recognized, and general methodologies for analyzing uncertainty have been offered. Precipitation and anticipated is a crucial meteorological input in forecasting flood. Forecasted precipitation is usually produced from numerical weather prediction's quantitative precipitation predictions [18]. The numerical weather prediction grid size may be a substantial source of mistake in rainfall forecasting, which is exacerbated by the grids' positional error. Even observed precipitation is subject to considerable uncertainty. Rain gauges only cover a small region, and there might be significant gaps between them, resulting in huge precipitation inaccuracies, especially in mountainous places. Weather radars may cover wide regions but cannot directly monitor rainfall, and conversion from reflectance to rainfall might be problematic. Flood-producing storm events often occur on small scales, and gridded remotely sensed products may not catch them.

Besides from precipitation, a variety of additional mistakes and uncertainties can be significant. When the models are applied to individual storm occurrences, for example, inaccuracies connected with the beginning circumstances, e.g., soil moisture, are particularly critical. Furthermore, any model update or downscaling, as well as infrastructure activities, might introduce mistakes and uncertainty. With the usage of rating curves, errors might be introduced. Flood forecasts are usually presented as level (gauge data), but hydrological models usually compute discharge. To convert computed fluxes to water levels, a rating curve is employed. Rating curves are often created using a small number of discharge observations that may not cover severe flood occurrences, leaving enough opportunity for error. Furthermore, the gauge observations may have uncertainties. Furthermore, erroneous or missing data, human processing mistakes, or unforeseen actions can all contribute to operational uncertainty in flood forecasting [19].

Finally, model structural errors, parameter mistakes, and spatial discretization errors can all result in large inaccuracy when forecasting models misrepresent hydrologic processes. Model structure arises from the simplification of intricate catchment physics into models of physical processes. More data are not going to fix these systemic issues. There may be a lot of uncertainty in model parameters, but this usually goes away when more recorded runoff data become available and are utilized to change the model parameters. The size of the research region, the diversity of its characteristics, the number of sub-divisions of the area, and the data resolution all impact parameter uncertainty. The geographical representation of the catchment, of which there are three popular techniques such as sub-catchments, rectangular grid model, and response units such as SWAT model, is a significant inaccuracy related to forecasting models.

3.12. Uncertainty Quantification. Quantifying flood-forecasting predictive uncertainty is critical for conveying flood danger and minimizing uncertainties. Uncertainty analysis is one of the most challenging components of flood forecasting since a thorough definition of the cascade of uncertainties from input to warning is a time-consuming and computer-intensive approach. There are several methods for quantifying the uncertainty of specific forecasting components. One method is to modify the expected rainfall and model parameters to test the model's sensitivity to a believable range of inputs. Because parameter model may be dependent on previous standardization and knowledge, there is a lot of uncertainty in the early stages of forecasting. Improved model calibration raises parameter confidence and reduces uncertainty as the event progresses and more observable data become accessible. The relevance of observed data is underscored by the findings, which show that additional stage data are more useful than parameterization of flood models in lowering barrage depths and range uncertainty [20].

The predictive uncertainty paradigm depicts the likelihood of a forthcoming value of a forecast, such as water level and discharge, occurring. This likelihood is based on all available knowledge about the future value, which is often supplied via forecast modeling. The predictive uncertainty technique has the benefit of quantifying uncertainty in terms of a likelihood distribution.

The impacts of a wide variety of forecast uncertainties can be included into the abovementioned ensemble prediction systems (EPSs), yielding a probabilistic forecast with directly expressed uncertainty. To indicate uncertainty, individual models within the ensemble may have different assumptions regarding beginning circumstances, boundary conditions, model parameterization, model structure, or any combination of these. Ensembles can be assembled in a variety of ways. The most common technique involves feeding numerical weather predictions from an ensemble prediction system into hydrological models to obtain hybrid ensemble prediction system-based forecasts. This technology was used to create the European Flood Awareness System [21].

Within the ensemble, some ensembles utilize many models or the same model with various physical parameterization methodologies. Ensemble forecasting strategies include using climatology and starting circumstances to build an ensemble, leveraging error distributions derived from previous hydrological forecasts to improve current predictions, and handling spatial uncertainty in rainfall forecasting. Rather than a single deterministic forecast, a set of projections may be used to quantify and explain the uncertainty surrounding the flood event [22].

Hybrid ensemble prediction methods, while widely used, are not without their drawbacks. In many circumstances, hybrid ensemble prediction systems must analyze enormous volumes of data produced by collaborative models that are not always available or computationally practical. Understanding how to base flood warning decisions on probabilistic projections is difficult. Emergency response agencies, for example, may be perplexed by probabilistic estimates and react with fear or apathy. Furthermore, efforts must be made to describe and convey the elements that influence the accuracy of forecasts based on hybrid ensemble prediction systems. Probabilistic hybrid rainfall forecasts influence the performance of hydrological predictions. An evaluation of hydrological performance is based on estimating the bias of deterministic and collaborative hydrological estimates impacting the catchment outlet discharge threshold. To evaluate ensemble predictions, a peak-box technique is used. The extent and effectiveness of peak discharges are the emphasis of this method [23].

Furthermore, decision-making based on probabilistic predictions provided by hybrid ensemble prediction systems needs visualization tools and user-focused ensemble forecast assessment. Despite the significance of prediction visualization and communication approaches, individuals who receive forecasts see, understand, and act on them. If predictions are to be effective for flood control, developing visualization tools and estimating the products for expressing indecision is a critical challenge.

4. Results and Discussion

The study analyzes the application of nonlinear flood event in the Harvey River catchment with the description of data based on rainfall, model, and prior information available. The procedure of estimation is checked based on the several assumptions. The various stages of the flood event are analyzed based on the parameter estimation using Bayesian theorem. The information about the region of study is given in Table 1.

The error in least square is assumed with the residual plot, and the correlation in the serial is tested. The error variance with constant value is sampled using the Bayesian parameter and pooling diagnosis, which is given in Figure 1.

The relationship among the mean depth of the rainfall and error variance is revealed using river catchment. The data set of each element in the matrix is established in Figure 2. The variance of the peak discharge is observed using the regression model.

The transformed residual plot is shown in Figure 3, which indicated the variability of the residual transformed and transformed dependent on the weighted least square function.

The plot of normal probability is shown in Figure 4, which is observed for transmuted residuals for Harvey River samples. The curve denotes the Gaussian distribution and critical value of the statistical analysis. The altered residuals are carefully conformed to the Gaussian distribution function.

Table 1: Prior information about the Harvey River sample.

Area	148
Relief	Low to moderate
Length of record available	13
Measurement of maximum discharge	20
Recording the maximum number of rain gauge	4
Number of storm	35

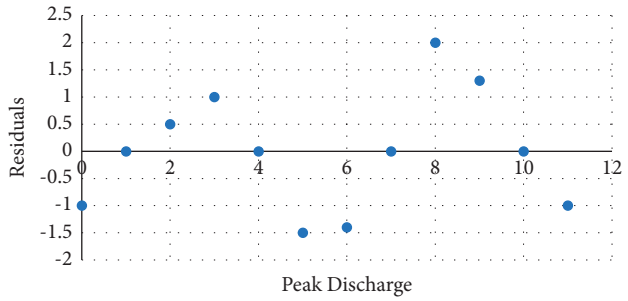


Figure 1: Least square with the standardized residual plot.

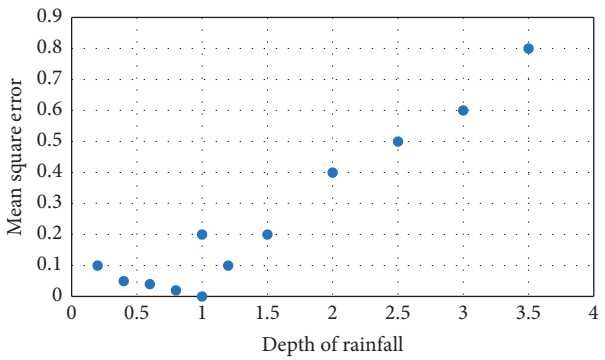


Figure 2: Quadratic relation.

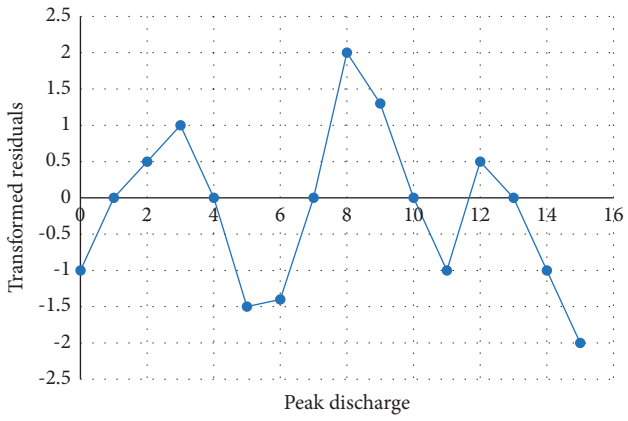


Figure 3: Plot of residual for weighted least square.

The result of maximum a posterior is obtained in Table 2 for Harvey River sample. The estimation of MAP is precise which is prior to the corresponding mean. The utilization of MAP estimation enhances the precision improvement.

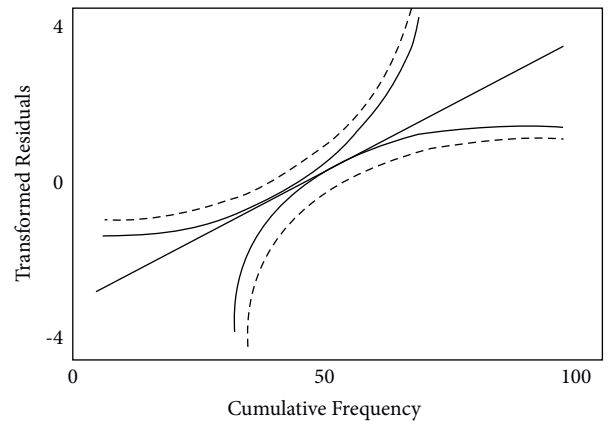


Figure 4: Plot of normal probability.

Table 2: Model parameter with mean and variance.

Parameter	Mean	Variance
Ordinary least square	0.038	0.054
Weighted least square	0.034	0.058
Prior parameter	0.13	0.62
Maximum a posterior	0.095	0.064

5. Conclusion

A Bayesian technique is developed for estimation of parameter in flood event with nonlinear models in this research. The approach is fairly broad, and it may be used to convert any perfect into a nonlinear regression framework. The use of a maximum a posteriori (MAP) approach for estimation of parameter in flood event models is discussed, with the goal of improving the dependability of design flood hydrographs. The suggested technique combines previous information on the model parameters as well as information of the mistake structures of the input data into the approximation process. The matrix of posterior covariance is measured using the calculated parameters' accuracy.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This research was supported by Taif University Researchers Supporting Project (TURSP-2020/318), Taif University, Taif, Saudi Arabia.

References

[1] I. L. Rubin and H. Falk, "Natural disasters and vulnerable populations: a commentary," *International Journal of Child Health and Human Development*, vol. 12, no. 4, pp. 303–318, 2019.

[2] X. Xu, X. Zhang, H. Fang et al., "A real-time probabilistic channel flood-forecasting model based on the Bayesian particle filter approach," *Environmental Modelling & Software*, vol. 88, pp. 151–167, 2017.

[3] E. P. Campbell, D. R. Fox, and B. C. Bates, "A Bayesian approach to parameter estimation and pooling in nonlinear flood event models," *Water Resources Research*, vol. 35, no. 1, pp. 211–220, 1999.

[4] B. C. Bates, "Use of parameter transformations in nonlinear, discrete flood event models," *Journal of Hydrology*, vol. 117, no. 1–4, pp. 55–79, 1990.

[5] S. K. Jain, P. Mani, S. K. Jain et al., "A Brief review of flood forecasting techniques and their applications," *International Journal of River Basin Management*, vol. 16, no. 3, pp. 329–344, 2018.

[6] M. Kefi, B. K. Mishra, P. Kumar, Y. Masago, and K. Fukushi, "Assessment of tangible direct flood damage using a spatial analysis approach under the effects of climate change: case study in an urban watershed in Hanoi, Vietnam," *ISPRS International Journal of Geo-Information*, vol. 7, no. 1, p. 29, 2018.

[7] B. C. Bates and E. P. Campbell, "A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling," *Water Resources Research*, vol. 37, no. 4, pp. 937–947, 2001.

[8] A. I. Taloba, M. R. Riad, and T. H. A. Soliman, "Developing an efficient spectral clustering algorithm on large scale graphs in spark," in *Proceedings of the 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 292–298, IEEE, Cairo, Egypt, December 2017.

[9] K. N. Teja and N. Umamahesh, "Application of ensemble techniques for flood forecasting in India," in *Proceedings of the Roorkee Water Conclave*, Roorkee, India, 2020.

[10] B. C. Bates and L. R. Townley, "Nonlinear, discrete flood event models, 1. Bayesian estimation of parameters," *Journal of Hydrology*, vol. 99, no. 1–2, pp. 61–76, 1988.

[11] A. Rayan, A. I. Taloba, R. M. E. Aziz, M. Rasha, and A. Amr, "IoT enabled secured fog based cloud server management using task prioritization strategies," *International Journal of Advanced Research in Engineering & Technology*, vol. 11, p. 9, 2020.

[12] S. Michaels, "Probabilistic forecasting and the reshaping of flood risk management," *Journal of Natural Resources Policy Research*, vol. 7, no. 1, pp. 41–51, 2015.

[13] A. I. Taloba, "An artificial neural network mechanism for optimizing the water treatment process and desalination process," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9287–9295, 2022.

[14] V. A. Rangari, N. V. Umamahesh, and C. M. Bhatt, "Assessment of inundation risk in urban floods using HEC RAS 2D," *Modeling Earth Systems and Environment*, vol. 5, no. 4, pp. 1839–1851, 2019.

[15] M. Damadipour, R. Tarinejad, and M. H. Aminfar, "Weighted transmissibility-based operational modal analysis for identification of structures using seismic responses," *Iranian Journal of Science and Technology, Transactions of Civil Engineering*, vol. 45, no. 1, pp. 43–59, 2021.

[16] F. Luino, C. G. Cirio, M. Biddoccu et al., "Application of a model to the evaluation of flood damage," *GeoInformatica*, vol. 13, no. 3, pp. 339–353, 2009.

[17] R. Egorova, J. M. Noortwijk, and S. R. Holterman, "Uncertainty in flood damage estimation," *International Journal of River Basin Management*, vol. 6, no. 2, pp. 139–148, 2008.

[18] S. Blazkova and K. Beven, "Uncertainty in flood estimation," *Structure and Infrastructure Engineering*, vol. 5, no. 4, pp. 325–332, 2009.

[19] P. D. Bates, F. Pappenberger, and R. J. Romanowicz, "Uncertainty in flood inundation modelling," in *Applied Uncertainty Analysis for Flood Risk Management*, pp. 232–269, 2014.

[20] A. Guo, J. Chang, Y. Wang, Q. Huang, and Y. Li, "Uncertainty quantification and propagation in bivariate design flood estimation using a Bayesian information-theoretic approach," *Journal of Hydrology*, vol. 584, Article ID 124677, 2020.

[21] K. D. Berends, M. W. Straatsma, J. J. Warmink, and S. M. J. H. Hulscher, "Uncertainty quantification of flood mitigation predictions and implications for interventions," *Natural Hazards and Earth System Sciences*, vol. 19, no. 8, pp. 1737–1753, 2019.

[22] G. Xue, F. Lin, S. Li, and H. Liu, "Adaptive dynamic surface control for finite-time tracking of uncertain nonlinear systems with dead-zone inputs and actuator faults," *International Journal of Control, Automation and Systems*, vol. 19, no. 8, pp. 2797–2811, 2021.

[23] S. Ha, L. Chen, H. Liu, and S. Zhang, "Command filtered adaptive fuzzy control of fractional-order nonlinear systems," *European Journal of Control*, vol. 63, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 5169259, 9 pages
https://doi.org/10.1155/2022/5169259

Research Article

Improved Method of Blockchain Cross-Chain Consensus Algorithm Based on Weighted PBFT

Liu Lei,[1,2] **Liangtu Song**,[1] **and Jiahua Wan**[3]

[1]*Institute of Intelligent Machines, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei 230031, China*
[2]*University of Science and Technology of China, Hefei 230026, China*
[3]*Anhui Xinhua University, Hefei 230088, China*

Correspondence should be addressed to Jiahua Wan; wanjiahua@axhu.edu.cn

Received 30 May 2022; Accepted 1 August 2022; Published 25 August 2022

Academic Editor: Heng Liu

Aiming to solve the problems of low fault tolerance, low throughput, and high delay in traditional methods, an improved method of the blockchain cross-chain consensus algorithm based on weighted PBFT is proposed. This article constructs a blockchain cross-chain exchange model based on cluster centers and divides the nodes in the blockchain system into consensus service nodes, cross-chain exchange nodes, and application nodes to improve the performance of consensus computing services. On this basis, according to the weighted PBFT consensus mechanism, the blockchain consensus environment is set up, and the distribution of nodes in the consensus domain and the blockchain signature scheme are obtained. Therefore, the blockchain cross-chain consensus optimization algorithm is designed to reduce throughput and delay and optimize the consensus effect. The experimental results show that the proposed method can effectively improve the shortcomings of traditional methods, with high throughput and low latency, and strong security. It shows that it is a low resource consumption and secure consensus method.

1. Introduction

Blockchain is a new technology integrated with distributed storage, peer-to-peer (P2P) networking, consistency verification, consensus algorithm, cryptography, and other computer technologies [1]. It uses blockchain data structure to verify and store data, uses a consensus algorithm to generate and update data, uses cryptography to ensure the security of data transmission and access, uses an intelligent contract composed of automatic script code to program and operate data, and realizes trusted data management in the incomplete trusted environment [2]. Amongst them, the consensus algorithm is the core part of the blockchain, which directly affects the efficiency, security, and stability of the whole system. At this stage, blockchain is accelerating the development of the digital economy and is deeply integrated with real industries. However, selecting or designing an appropriate blockchain cross-chain consensus algorithm according to business needs is difficult for researchers and developers [3, 4].

Under the above background, reference [5] proposed a master-slave multichain blockchain consensus mechanism based on reputation, designed a two-layer blockchain structure to build a master-slave multichain mechanism, and connected multiple chains through the main blockchain. From the blockchain, the global consistency of digital assets is guaranteed, and the performance of the blockchain is improved. A reputation evaluation is introduced into the consensus mechanism based on proof of rights and interests, and a joint consensus mechanism integrating multiple consensus mechanisms is designed to ensure data consistency and tamper-proof modification. By generating dynamic verification nodes, the decentralization of nodes is ensured, and malicious attacks are prevented. The simulation experiment results show that this method has the advantage of high security in dealing with right smashing attacks and bribery attacks, but it has the problem of weak fault tolerance. Reference [6] proposed an efficient blockchain consensus algorithm based on directed acyclic graphs. The algorithm uses the directed acyclic graph data

structure based on ID classification, which can reach a consensus more simply, and is suitable for multiple users to confirm transactions simultaneously. The experimental results show that the consensus algorithm can save a large amount of hardware resources and improve the transaction processing effect of the blockchain, but it has the problem of low throughput. At the system operation level, an adaptive controller together with fractional-order parameter adaptation laws is designed based on combining the parallel distributed compensation technique and the fractional Lyapunov stability theory to guarantee the Mittag-Leffler stability in the closed-loop system [7]. Reference [8] presents a composite learning fuzzy control to synchronize two different uncertain incommensurate fractional-order time-varying delayed chaotic systems with unknown external disturbances and mismatched parametric uncertainties via the Takagi-Sugeno fuzzy method. An adaptive controller together with fractional-order composite learning laws is designed based on both a parallel distributed compensation technology and a fractional Lyapunov criterion. The boundedness of all variables in the closed-loop system and the Mittag-Leffler stability of tracking error can be guaranteed. Reference [9] proposed a high-reliability blockchain consensus mechanism based on contribution value and difficulty value. According to the node's contribution value ranking, the nodes are assigned the corresponding proof of work (PoW) difficulty value, and the nodes then compete for the accounting rights through the PoW consensus mechanism. The consensus mechanism after the introduction of PoW respects the proof of capacity (PoC) contribution value ranking to the greatest extent. Thus, the node's accounting block rate is highly proportional to its contribution value. At the system operation level, the accounting block rate is guaranteed to reach or infinitely approach 100%, which effectively solves the problem of system suspension caused by PoC. The PoW difficulty value distribution algorithm is designed from the perspectives of the node contribution value ranking, the value difference between adjacent contribution value nodes, and the grouping method. The rationality and the effectiveness of the difficulty value distribution algorithm are verified through experiments. The superiority and the feasibility of this scheme are verified through experiments, but when the nodes are many, this method has a certain time delay, and the real-time performance is poor.

The above discussion shows that despite many blockchain formulas proposed by related scholars, certain problems remain, such as excessive delay and low throughput. Each consensus algorithm has problems, which is also one of the main issues restricting the development of blockchain. Consensus algorithms are an important part of the blockchain. Improving the consensus algorithm in the blockchain is the most important link in improving the performance of the blockchain. Therefore, starting from improving throughput, reducing delay, and improving fault tolerance performance, this article proposes an improved method of the blockchain cross-chain consensus algorithm based on weighted PBFT and improves the blockchain consensus algorithm according to the blockchain use scenario.

2. Blockchain Cross-Chain Exchange Model Based on Cluster Centre

In view of the existing problems of blockchain systems, this article proposes a blockchain cross-chain exchange model based on cluster centers. This model divides the nodes in the consortium blockchain system into three different types of nodes: consensus service nodes, cross-chain exchange nodes, and application nodes [10]. The consensus service nodes with efficient computing capability through a high-speed network are connected to form a blockchain P2P network serving a business field [11], which specifically provides consensus computing services for the application nodes in the blockchain network. The cross-chain switching node is connected to different blockchain networks at the same time, the block data of different blockchains are synchronized, the status database of different blockchain networks on this node is formed, and a blockchain switching network between switching nodes based on the P2P protocol is formed to provide cross-chain access services for application nodes of different blockchain networks. The application node can synchronize data from its blockchain network consensus service node, access cross-chain exchange nodes, and send intrachain or cross-chain transactions. The schematic diagram of the blockchain cross-chain exchange model is shown in Figure 1.

In Figure 1, the blockchain network is divided into different specialized blockchain networks according to service or application fields and businesses; different blockchain networks have different requirements for service capacity and transaction efficiency, and they have corresponding calculations. Computing nodes that are capable and can connect to the main blockchain network at a high speed are used as consensus service nodes that provide consensus computing services to the blockchain network application nodes through computing power competition.

In the blockchain cross-chain exchange model, the collection of transaction nodes is denoted as D, the single node is denoted by d_i, and the transaction data are denoted by s_d. The consensus node collection is denoted as A, the nodes in the set are numbered as $\{0, 1, \ldots, n-1\}$, and a single node is denoted as a_i. The collection of data storage nodes is denoted by E, and a single node is denoted by e_i.

Each node d_i in D sends a transaction α, which is propagated to the consensus node set A through the P2P network. In the absence of intermediate nodes and because the consensus nodes do not trust one another, the nodes must exchange information and verify one another to reach a consensus [12]. If the transaction data sent by a certain consensus node A_{ij} to other nodes are s_{ij}, the transaction received by this node from other nodes is represented by a vector $G_t(t_0, t_1, \ldots, t_n)$, $(t_0, t_1, .., t_n)$ is a set of vectors for

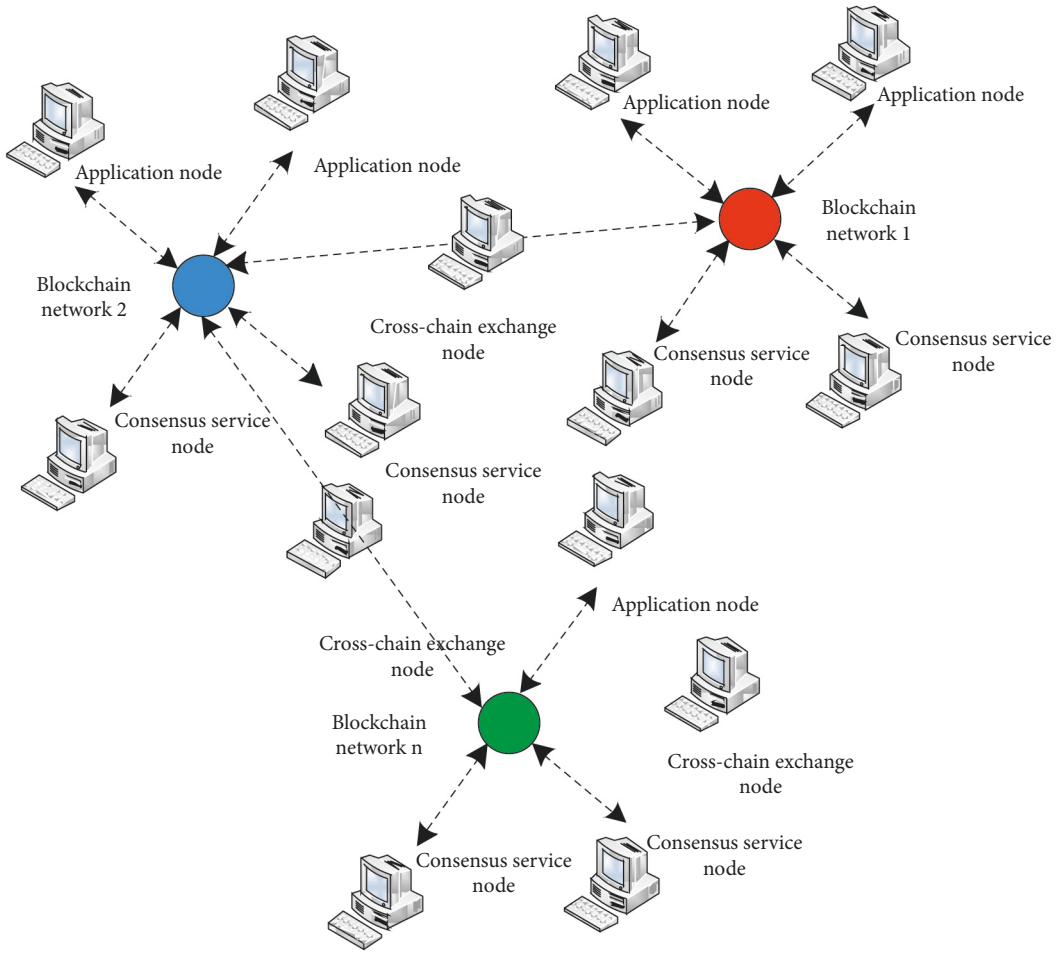


Figure 1: Schematic diagram of blockchain cross-chain exchange model.

transactions of different nodes, and the modulus of the vector is $|G_t| \le |D| - 1$.

A function $f(x)$ and an algorithm (protocol/process) Y are designed such that

$$G_t = f(t_0, t_1, \ldots, t_n),$$
$$\begin{cases} \text{Store}, & Y(t) = \text{ture}, \\ \text{Discard}, & Y(t) = \text{false}. \end{cases} \tag{1}$$

The final consensus result obtained through the transaction vector received by function $f(x)$ and consensus node A_{ij} is equal to the original transaction information. After a transaction of the algorithm Y, if it is approved by most nodes, the algorithm result is TRUE, and the "Store" instruction is sent to the data storage node. If it is FALSE, the "Discard" instruction is executed and current transaction data are discarded. The model is essentially a data pipeline, and transaction data flow between different types of nodes.

Malicious nodes are in the consensus node, deliberately tampering with transaction data, causing the problem of "double consumption." Therefore, the consensus nodes are divided into three categories, namely, A_1, A_2, and A_3. The first two are honest nodes, and the latter is a malicious node. If A_1 receives A_2's transaction vector as ∂_A, and the transaction vector received by A_3 is $\partial_{A'}$, then function $f(x)$ and algorithm Y must meet the following conditions:

$$\begin{cases} \partial_{A'} = \dfrac{f(\partial_{A'}, \partial_A)}{a_i}, \\ \text{Store}(\partial_A) \ne \text{True}. \end{cases} \tag{2}$$

Node A_{ij} can identify which transactions are illegal and which transactions are legal without intermediate nodes, and store the legal transactions as the execution result in the blockchain database. Illegal transactions cannot be executed at any time.

3. Blockchain Cross-Chain Consensus Algorithm Optimization

3.1. PBFT Consensus Mechanism. The consensus mechanism layer is responsible for the data consistency amongst the nodes in the whole blockchain system network. The data of all nodes in the blockchain are stored independently. Therefore, a mechanism is needed to ensure that the ledger data stored by each node of the blockchain are consistent, and the role of the

consensus layer is to ensure that the information on the chain is transparent and the data can be shared [13].

PBFT is an algorithm used to solve the Byzantine Generals problem, which can ensure consistency between nodes in the presence of malicious nodes in the network. The PBFT algorithm has three roles: client, master node, and slave node. The master node and the slave node perform data backup. During PBFT, the consensus process led by a master node is in one view [14]. The view is the definition of the node relationship in the PBFT consensus mechanism, and the number of the view is marked as Z. In a view, different nodes have different numbers, and each view has only one main node. In the event of a failure of the master node and timeout of the consensus process, the master node needs to be switched according to the view switching protocol in the PBFT consensus mechanism to generate a new view Z' and continue the consensus process in this view. Figure 2 shows the PBFT consensus flow chart.

The specific PBFT consensus steps are as follows:

(1) Request stage: The client sends a request $\langle \text{Request} : r, c, v \rangle$.

(2) Prepreparation phase: The master node assigns a number M to the received request and broadcasts a prepreparation message in the format $\text{pre}\langle R: r, c, v, p \rangle$, where p is the digest value of the request, which is generated using a secure hash function.

(3) Preparation stage. The node that receives the prepreparation message verifies p in the prepreparation message and enters the preparation phase if the verification passes. The preparation phase broadcasts the preparation message. The format is $\langle \text{R} : \text{r}, \text{c}, \text{l} \rangle$, and l is the number of the node itself. The $\text{pre}\langle \text{R} : \text{r}, \text{c}, \text{v}, \text{p} \rangle$ and $\langle \text{R} : \text{r}, \text{c}, \text{v} \rangle$ messages are written to the log whilst broadcasting. When the master node receives the $\text{pre}\langle \text{R} : \text{r}, \text{c}, \text{v}, \text{p} \rangle$ and $\langle \text{R} : \text{r}, \text{c}, \text{v} \rangle$ messages from $N + 1$ different nodes and passes the verification, the preparation phase is completed, and the main verifications are r, c, and v.

(4) Confirmation stage. After the node completes the preparation phase, it enters the confirmation phase and broadcasts the confirmation message in the format $\langle \text{Comment} : \text{r}, \text{c}, k_f, \text{l} \rangle$, where k_f is the signature set of the slave node. When the $\langle \text{Comment} : \text{r}, \text{c}, k_f, \text{l} \rangle$ message is confirmed, k_f and l are mainly verified. When $N + 1$ confirmation messages including themselves are received and verified, the confirmation phase is completed.

(5) Recovery stage. The response format is $\langle \text{Reply} : \text{r}, \text{c}, p, \text{l}, \text{h} \rangle$, where h is the execution result of the request. When the client receives the same request result from $N + 1$ different nodes, it believes that the network has reached this consensus.

3.2. Blockchain Cross-Chain Consensus Algorithm Optimization. According to the weighted PBFT consensus mechanism, the blockchain consensus environment is set up, and the distribution of nodes in the consensus domain is obtained, thereby designing the blockchain cross-chain consensus optimization algorithm to reduce throughput and delay and optimize the consensus effect.

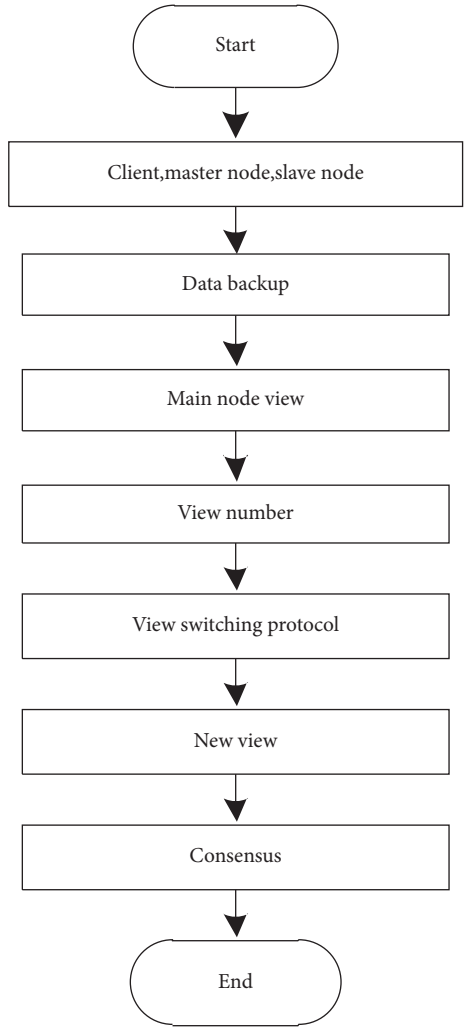


Figure 2: PBFT consensus flow chart.

3.2.1. Blockchain Consensus Environment. Transactions can be created and propagated on any node of the alliance chain. Blocks are created by consensus nodes and propagated to other nodes. The verification and processing rules of transactions and block data are governed by the Ethereum protocol; consensus data are propagated on the consensus network by consensus mechanism constraints. The data flow between nodes is shown in Figure 3.

In the alliance chain network, the set of alliance chain nodes is B, its size is U_0, the set of consensus nodes is B^{η}, its size is U_1, the set of nonconsensus nodes is B^{μ}, and its size is U_2, satisfying

$$\begin{cases} U_0 = U_1 \cup U_2, \\ U_1 \cap U_2 = \varnothing. \end{cases} \tag{3}$$

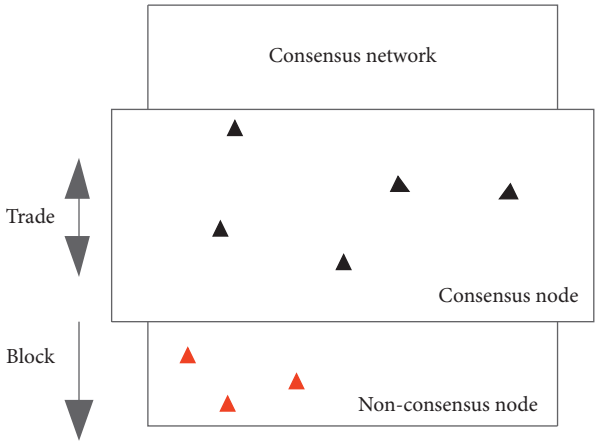


Figure 3: Blockchain consensus environment.

The nodes in set B are numbered as $(b_1, b_2, \ldots, b_t)$, and the single node is numbered as ω is denoted as B_ω [15], which satisfies

$$\forall B_\omega \in B\left(B^\eta + B^\mu\right), \tag{4}$$

and at the same time meets

$$\begin{aligned} B &\in B^\eta; \\ B &\in B^\mu. \end{aligned} \tag{5}$$

The consensus network may have multiple error nodes. The set of consensus nodes is B^η, and the set of error nodes is B^τ, i is the weight coefficient in the node-set m, and j is the weight coefficient in the node set n, which satisfies

$$B \geq \sum_{i=1}^{m} \sum_{j=1}^{n} \left[B_{ij}^\eta - B_{ij}^\tau\right]^2. \tag{6}$$

In the blockchain consensus environment based on the Byzantine fault-tolerant mechanism, if the number of error nodes that exist can meet formula (6), the Byzantine General problem can be solved, the fault tolerance is good, and the correctness and activity are strictly proven. In practice, the number of consensus nodes is at least 4.

3.2.2. Distribution of Consensus Domain Nodes. In the blockchain consensus environment, all consensus nodes in the consensus domain are divided into four groups, namely, β_1, β_2, β_3, and β_4, and each group member has a different weight in the consensus process, as shown in Figure 4.

Given that the voting weight of group β_1 members is $W_1 = 1$, the weight of group β_2 is $W_2 = 2$, the weight of group β_3 is $W_3 = 3$, and the weight of group β_4 is $W_4 = 4$. The members of group β_4 have the highest voting weight and believe that they are the most secure and least prone to evil nodes. Through this division, the security of the consensus network depends more on those groups with high voting weight [16].

The Byzantine decision criterion undergoes several changes because nodes have different voting weights. In

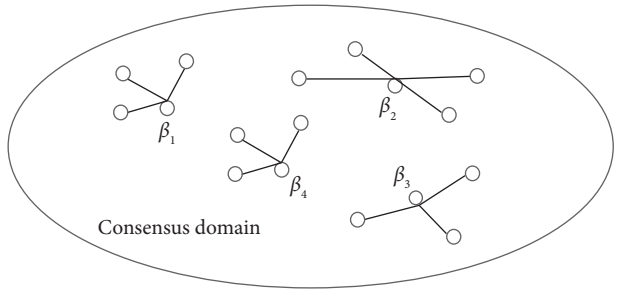


Figure 4: Schematic diagram of node distribution.

traditional algorithms, the voting weights of nodes are the same, which can be considered 1. Assuming that the total number of nodes in the entire network is N', the maximum number of malicious nodes allowed in the consensus network is C, and the Byzantine judgment criterion requires $N' = 3C + 1$. In the new consensus environment, the weighted sum of votes of the entire network is defined as $\rho_{N'}$, and the weighted sum of the maximum allowable votes of malicious nodes as W_f. Then, $\rho_{N'}$ is calculated as:

$$\rho_{N'} = \frac{\rho_l \varphi_l \left(T_i - T_j\right)}{T_W}. \tag{7}$$

Amongst them, ρ_l represents the weighted similarity value, φ_l represents the trustworthiness, T_i represents the density of secure nodes, T_j represents the density of malicious nodes, and T_W represents the total weight of the vote.

W_f is calculated as:

$$W_f = \int_{T_f}^{T_W} c_{ij}\left(T_i - T_j\right) d\left(T_i - T_j\right). \tag{8}$$

Amongst them, c_{ij} represents the average trust degree of the node.

When all nodes belong to group β_4, the algorithm becomes the traditional PBFT consensus algorithm.

The weighted PBFT optimization algorithm is explained based on the PBFT algorithm. All consensus nodes are divided into main domain nodes and subdomain nodes. The nodes of group β_4 constitute the main domain, and the nodes of group β_1, β_2, and β_3 constitute the subdomain [17]. The number of nodes in the main domain is defined as $\beta_4(N)$, and the upper limit of the malicious nodes that can be accommodated in the main domain is $W_f(\max)$.

The selection rule of the master node is to select from the main domain randomly but this is different from the traditional algorithm, which is randomly selected from the entire network. This has the advantage of reducing the frequency of master node changes. The switch of the master node causes a view switch. To ensure consistency, a large amount of additional communication is required. Therefore, the switch of the master node reduces consensus efficiency. Given that the security level of the nodes in the main domain is the highest, the probability of the master node's error is low because the view switching caused by the master node's error occurs less frequently, which reduces consensus overhead and improves consensus efficiency.

3.2.3. Blockchain Signature Scheme. The consensus node is required to sign the block because the traditional PBFT algorithm does not verify whether the received block comes from a legal consensus node. The specific rules are as follows:

(1) When creating a proposal block, the consensus node list is used as the block header data to participate in the calculation of the block hash value to ensure that the set of nodes participating in each round of consensus negotiation cannot be tampered with.

(2) When the proposed block is submitted to the chain, the cache list consisting of the consensus node index and the signature of the block is synchronously submitted. Although the cache list of different nodes may be different, the number of list elements strictly meets the vote requirement.

(3) When synchronizing the history and adding a new block, the sender node broadcasts the block whilst sending the list cache list, and the receiver node verifies it. If the signature verification passes, the block is uploaded to the chain [18].

3.2.4. Realisation of Blockchain Cross-Chain Consensus. When traditional algorithms execute consensus protocols, a large amount of communication occurs between nodes. As the number of nodes and transactions increases, network communication increases rapidly, which increases bandwidth pressure and affects the consensus efficiency of the algorithm. Therefore, this article aims to solve this problem, adopts the weighted PBFT algorithm, combined with the characteristics of the alliance chain, and in the absence of Byzantine nodes, optimizes the consensus protocol to reduce the amount of communication between nodes. Furthermore, the integration mechanism and the elevator mechanism are introduced, such that the algorithm can quickly restore to the optimal state when Byzantine nodes appear in the network and execute the optimized consensus protocol most of the time [19].

The weighted PBFT algorithm is an improvement on the PBFT algorithm. It also completes the consensus operation through mutual communication between nodes in the network, and its communication is executed according to the consensus protocol. The algorithm in this article improves the consensus protocol of the PBFT algorithm and designs an optimized consensus protocol to reduce the amount of communication between nodes in the consensus process.

The specific execution of the weighted PBFT algorithm is as follows:

(1) Initialize the node. First, the nodes in the network are numbered, and the points of the nodes are initialized to 100 points[20]. Second, the consensus node set B^{η} and the candidate node set B^{λ} are initialized.

$$B^{\lambda} = \{1, 2, 3, \ldots N - 1\}. \tag{9}$$

Then, the number of consensus nodes and candidate nodes is

$$\begin{cases} \left|B^{\eta}\right| = N_{f'}^{2}, \\ \left|B^{\lambda}\right| = N_{d'}^{2}. \end{cases} \tag{10}$$

In (10), f' represents the consensus weight of consensus nodes. d' represents the expected weight of the candidate node candidate set.

(2) The client sends a transaction request to the master node. After the master node receives the request, the request message is numbered, and then the master node executes the optimized consistency protocol.

(3) All consensus nodes execute the optimized consistency protocol. In the confirmation phase of the optimized consistency protocol, the state of the consensus node is judged. At this stage, the master node receives the feedback messages from all consensus nodes, judges the correctness of the feedback messages, compares them with the locally saved prepreparation messages, and judges whether the corresponding values of the fields are the same. Once the transaction information is tampered with, its hash value changes. Therefore, two comparison results appear in the system, and the algorithm performs different operations according to different comparison results [21].

(4) The master node upgrades the node according to the integration of the node, and updates the consensus node set and candidate node set to ensure that the high probability of consensus nodes are honest nodes, and then continues to implement the optimized consistency protocol in the next consensus process [22]. In summary, the use of the weighted PBFT algorithm to improve blockchain cross-chain consensus is completed.

4. Simulation Experiment

In order to verify the effectiveness of the blockchain cross-chain consistency algorithm based on weighted PBFT, simulation and comparison experiments are carried out. It is compared with the master-slave multichain blockchain consensus mechanism based on reputation and the efficient blockchain consensus algorithm based on a directed acyclic graph. Next, they are compared and evaluated in terms of fault tolerance, throughput, and delay. This test is conducted in a CloudSim cloud computing environment, and the network bandwidth is 80 mi S^{-1}, 8 GB memory, 10 servers, and SPSS simulation software is used to process the experimental data.

4.1. Experimental Indicators

(1) Throughput refers to the number of network nodes that the algorithm can carry after being used in the system.

(2) Fault tolerance refers to the tolerance value of the consensus algorithm to nodes that have non-Byzantine faults and the tolerance value of nodes that

have Byzantine faults in the system. Fault tolerance is also one of the references for algorithm security.

(3) Latency mainly reflects the efficiency of the consensus algorithm in an application.

(4) Safety: Consensus algorithm safety mainly refers to malicious behavior incidence. In the current complex network environment, ensuring the security and stable operation of the blockchain is the focus of the current research.

4.2. Analysis of Experimental Results

Experiment 1. To obtain transaction data better, the transaction simulation module is introduced into the blockchain system. Through the newly built 100 simulated transaction accounts, to make the transaction sustainable, the initial quota of each transaction account is 10000, and the volume of each transaction is 1 to ensure that the transaction can occur at the fastest speed and test the throughput of the consensus algorithm. The throughput comparison results of the three methods are shown in Figure 5.

According to the analysis of Figure 5, with the increase in the number of nodes in the network, the throughputs of the three methods show an upward trend, but overall, the throughput of the method in this article is much higher than that of the master-slave multichain blockchain consensus mechanism based on reputation and the efficient blockchain consensus algorithm based on the directed acyclic graph because the traditional method suspends the consensus process for a period of time, during which the algorithm only completes the consensus work of one transaction, which affects the throughput of the algorithm. This method does not have this problem. Thus, the throughput is substantially higher than that of the two traditional methods.

Experiment 2. Comparing the fault tolerance of different methods, the results are shown in Table 1, where n represents the total number of nodes in the system.

Table 1 shows that this method can tolerate that the proportion of non-Byzantine error nodes is $n/2$, with high consistency, high availability, and strong antifraud ability. The fault tolerance of the master-slave multichain blockchain consensus mechanism based on reputation is $n/3$, and any attempt to destroy the system entails a large amount of costs that outweighs the losses. The fault tolerance of the efficient blockchain consensus algorithm based on the directed acyclic graph is $n/5$, and the security is poor. The comparison shows that the fault tolerance of this method is good, which shows that it has high security.

Experiment 3. To verify the effectiveness of the method in this article further, the time delay is taken as the experimental index to compare the application effects of different methods. The results are shown in Table 2.

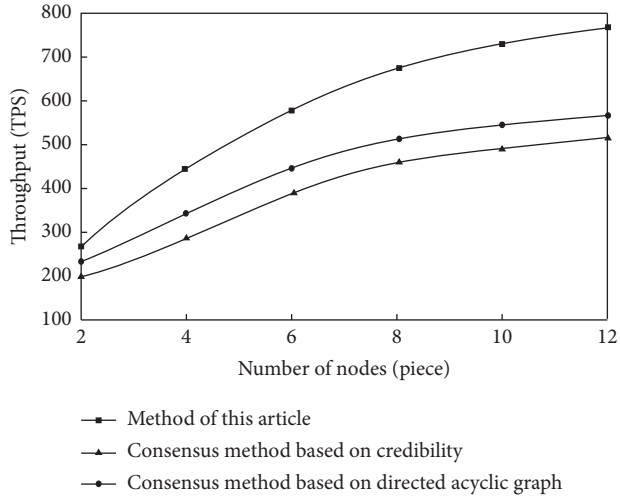


Figure 5: Throughput comparison results.

Table 1: Comparison results of fault tolerance.

Method	Fault tolerance
Method of this article	$n/2$
Master-slave multichain blockchain consensus mechanism based on reputation	$n/3$
Efficient blockchain consensus algorithm based on directed acyclic graph	$n/5$

Table 2 shows that in the case of relatively few transaction nodes, the three methods spend almost the same time, and the method in this article has a slight advantage. However, with the increase in the number of transaction nodes, the advantage of the method in this article is gradually evident, which is far from the time spent by the traditional method. This method has great advantages in transaction efficiency and is suitable for the application scenario of large-scale blockchain deployment.

Experiment 4. In order to further verify the security of the consensus algorithm designed in this article, the security of the three methods is compared with the proportion of malicious behaviors as the experimental indicator. Given the same system parameters, 20 transaction information consensuses were performed for the three consensus algorithms. The occurrence of malicious behaviors in the system is recorded separately, and then the proportion of malicious behaviors in the consensus algorithm is statistically compared. The experimental results are shown in Figure 6.

As can be seen from Figure 6, the average malicious behaviors are 28.47%, 35%, and 39.1%, respectively, of this method. As can be seen from the experimental results, this article has significant advantages in reducing the probability of malicious behavior in the system, effectively improving system security.

Table 2: Delay comparison results.

Number of transaction nodes/piece	Method of this article	Consensus mechanism of master-slave multichain blockchain based on reputation	Efficient blockchain consensus algorithm based on the directed acyclic graph
4	2.5	2.6	2.6
8	2.8	2.8	3.0
12	3.2	3.2	3.4
16	3.5	4.1	4.7
20	3.6	4.5	5.0
24	3.7	4.9	5.5
28	3.8	5.2	6.0
32	3.9	5.7	6.6

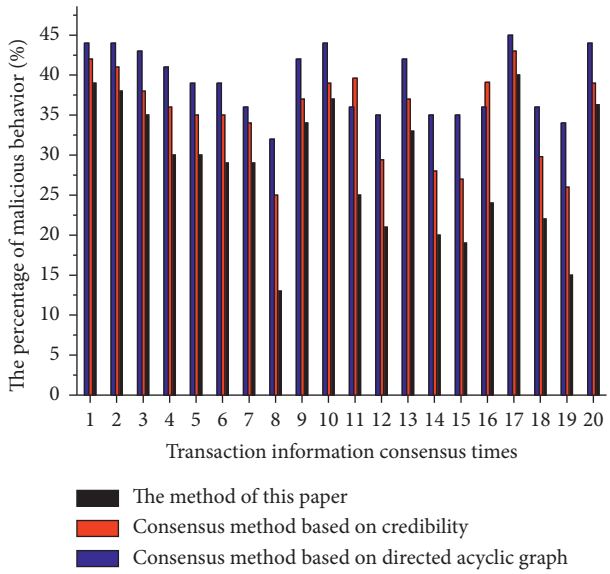


Figure 6: Safety contrast.

5. Conclusion

Summarizing the advantages and disadvantages of traditional methods, to improve the effect of blockchain consensus further, a method for improving the blockchain cross-chain consensus algorithm based on weighted PBFT is proposed. A blockchain cross-chain exchange model based on a cluster center is established, the blockchain consensus environment is set, the distribution of consensus domain nodes and the blockchain signature scheme is obtained, and the blockchain cross-chain consensus optimization algorithm is designed. The experimental results show that the proposed method has higher throughput, lower latency, and stronger security, which fully verifies the effectiveness of the method. It is expected to provide a reference for researchers and developers when selecting or innovatively designing consensus algorithms and to promote the evolution of blockchain consensus algorithms.

Data Availability

The data used to support the findings of the study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was financially supported by the Universities Blockchain Technology Innovation Action Plan 2020jyxm0789 and 2020qk121 fund.

References

[1] G. Strawn, "Block chain," *IT Professional*, vol. 21, no. 1, pp. 91-92, 2019.

[2] S. M. Alrubei, E. A. Ball, J. M. Rigelsford, and C. A. Willis, "Latency and performance analyses of real-world wireless IoT-blockchain application," *IEEE Sensors Journal*, vol. 20, no. 13, pp. 7372–7383, 2020.

[3] L. R. Abbade, F. M. Ribeiro, M. H. Silva et al., "Blockchain applied to vehicular odometers," *IEEE Network*, vol. 34, no. 1, pp. 62–68, 2020.

[4] M. Ahmed and A. S. K. Pathan, "The Blockchain: can it be trusted?" *Computer*, vol. 53, no. 4, pp. 31–35, 2020.

[5] H. Z. Liu, S. S. Li, W. L. Lv, and S. J. Wei, "Master-slave multiple-blockchain consensus based on credibility," *Journal of Nanjing University of Science and Technology*, vol. 44, no. 3, pp. 325–331, 2020.

[6] Y. M. Wang, J. F. Chu, Y. J. Wang, and Y. D. Chen, "Efficient blockchain consensus algorithm based on directed acyclic graph," *Journal of Jilin University (Science Edition)*, vol. 58, no. 5, pp. 1167–1172, 2020.

[7] Y. Zhou, H. Liu, J. Cao, and S. Li, "Composite learning fuzzy synchronization for incommensurate fractional-order chaotic systems with time-varying delays," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 12, pp. 1739–1758, 2019.

[8] Y. Zhou, H. Wang, and H. Liu, "Generalized function projective synchronization of incommensurate fractional-order chaotic systems with inputs saturation," *International Journal of Fuzzy Systems*, vol. 21, no. 3, pp. 823–836, 2019.

[9] J. S. He, K. Zhang, R. X. Xue et al., "A highly reliable consensus mechanism for blockchain based on contribution and difficulty values," *Chinese Journal of Computers*, vol. 44, no. 1, pp. 162–176, 2021.

[10] R. Riesco, X. Larriva-Novo, and V. A. Villagra, "Cybersecurity threat intelligence knowledge exchange based on blockchain," *Telecommunication Systems*, vol. 73, no. 2, pp. 259–288, 2020.

[11] Y. Zhan, B. Wang, R. Lu, and Y. Yu, "DRBFT: delegated randomization Byzantine fault tolerance consensus protocol

for blockchains," *Information Sciences*, vol. 559, pp. 8–21, 2021.

[12] W. Li, C. Feng, L. Zhang, H. Xu, B. Cao, and M. A. Imran, "A scalable multi-layer PBFT consensus for blockchain," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 5, pp. 1146–1160, 2021.

[13] R. Hou, H. Liu, Y. Hu, and Y. H. Zhao, "Research on secure transmission and storage of energy IoT information based on Blockchain," *Peer-to-Peer Networking and Applications*, vol. 13, no. 4, pp. 1225–1235, 2020.

[14] X. Zheng, W. Feng, M. Huang, and S. Feng, "Optimization of PBFT algorithm based on improved C4.5," *Mathematical Problems in Engineering*, vol. 2021, no. 2, pp. 1–7, 2021.

[15] P. Bagga, A. K. Sutrala, A. K. Das, and P. Vijayakumar, "Blockchain-based batch authentication protocol for Internet of Vehicles," *Journal of Systems Architecture*, vol. 113, no. 8, Article ID 101877, 2021.

[16] T. Maksymyuk, J. Gazda, M. Volosin et al., "Blockchain-empowered framework for decentralized network management in 6G," *IEEE Communications Magazine*, vol. 58, no. 9, pp. 86–92, 2020.

[17] G. Nagasubramanian, R. K. Sakthivel, R. Patan, A. H. Gandomi, M. Sankayya, and B. Balusamy, "Securing e-health records using keyless signature infrastructure blockchain technology in the cloud," *Neural Computing & Applications*, vol. 32, no. 3, pp. 639–647, 2020.

[18] G. Mwitende, Y. Ye, I. Ali, and F. Li, "Certificateless authenticated key agreement for blockchain-based WBANs," *Journal of Systems Architecture*, vol. 110, no. 4, Article ID 101777, 2020.

[19] H. Kim, J. Park, M. Bennis, and S. L. Kim, "Blockchained on-device federated learning," *IEEE Communications Letters*, vol. 24, no. 6, pp. 1279–1283, 2020.

[20] B. Wu and J. F. Yan, "Distributed trusted network access authentication based on blockchain technology," *Computer Simulation*, vol. 38, no. 1, pp. 277–281, 2021.

[21] F. Leal, A. E. Chis, and H. González–Vélez, "Multi-service model for blockchain networks," *Information Processing & Management*, vol. 58, no. 3, Article ID 102525, 2021.

[22] P. S. R. Garcia and J. H. Kleinschmidt, "Sharing health and wellness data with blockchain and smart contracts," *IEEE Latin America Transactions*, vol. 18, no. 06, pp. 1026–1033, 2020.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 8005249, 19 pages
https://doi.org/10.1155/2022/8005249

Research Article

Dynamic Evaluation of Transformation Ability for Emergency Scientific Research Achievements Based on an Improved Minimum Distance-Maximum Entropy Combination Weighting Method: A Case Study of COVID-19 Epidemic Data

Qingmei Tan and Juan Hui

College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu 211106, China

Correspondence should be addressed to Juan Hui; hwshj22@nuaa.edu.cn

Received 16 May 2022; Revised 23 June 2022; Accepted 1 July 2022; Published 12 August 2022

Academic Editor: Heng Liu

In the process of responding to major public health emergencies, the transformation of emergency scientific research results often faces many unfavourable factors such as limited resources, tight time, changes in needs, and lack of results. It is necessary to evaluate and analyze the ability to transform emergency scientific research results under public health emergencies, so as to rationally allocate emergency scientific research resources between subjects and regions, improve the efficiency of emergency results transformation, enhance emergency scientific research capabilities, and efficiently support incident prevention, control, and treatment. Starting from the patent level, this paper constructs an indicator system to evaluate the transformation ability of emergency scientific research results under major public health emergencies. It improves the minimum distance-maximum entropy combination weighting method to realize the static evaluation of transformation ability for emergency scientific research results from the perspective of patents, then constructs the dynamic evaluation model of transformation ability for emergency scientific research results in public health emergencies from the perspective of patents, and carries out the dynamic evaluation of the emergency scientific research achievements transformation ability of different subjects and different regions. We also improve the ER index, measure the static polarization effect of the transformation ability for regional emergency scientific research results, and consider the time factor to construct a dynamic polarization effect measurement model for the transformation ability of emergency scientific research achievement. Furthermore, this paper improves the measurement model of contribution degree to the polarization effect, and analyzes the contribution degree to polarization of the transformation ability for regional emergency scientific research results.

1. Introduction

Since the 21st century, public emergencies have occurred from time to time around the world, and the urgency and uncertainty of major public emergencies have posed a great challenge to global economic development. The sudden outbreak of the "COVID-19" epidemic in 2020 has had a great impact on the economic development of various countries and even the world, which is considered to be one of the greatest challenges and tragedies of the century after World War II [1]. It has resulted in the death of hundreds of thousands of people around the world, posed an urgent challenge to medical professionals, and exposed the weaknesses of national health systems [2, 3]. COVID-19 rapidly caused major economic and social disruptions on an unprecedented scale [4, 5].

Under the background of the above events, it is crucial to efficiently play the supporting role of scientific research results in the process of emergency management. However, the result transformation of emergency scientific research during the process of public health emergency response is often faced with many unfavourable factors such as limited

resources, time constraints, and changing needs [6]. There is a great lack of scientific research results that can meet the emergency needs of health emergencies, and the update rate is fast. Therefore, it is of great significance to form a strong capacity to transform emergency scientific research results. It is beneficial to enhance the efficiency of emergency scientific research resources allocation among subjects and regions, strengthen the strength of emergency scientific research under emergencies, improve the efficiency of emergency results transformation, and provide scientific, safe, and effective emergency results guarantee for responding to public emergencies, so as to efficiently support the prevention, control, and treatment of events.

The transformation of scientific and technological achievements is a complex project, which is the key to the deep integration of scientific & technological innovation and economic development [7, 8]. In-depth studies have been conducted in the existing literature on the emergency scientific research system and its achievement transformation mechanism, the rapid response mechanism of emergency R&D, and the collaborative integration process and corresponding methods of emergency scientific research, which have laid a solid theoretical foundation for subsequent studies. However, there is a lack of empirical studies related to the transformation of emergency scientific research results in the existing literature, and no empirical analysis of the transformation capacity of emergency scientific research results and its spatial distribution characteristics under emergencies has been conducted. In addition, the evaluation of achievement transformation capability belongs to the multi-criteria decision-making problem. Due to many index attributes, insufficient index data and various evaluation methods, it is often difficult to accurately measure the science and technology achievements transformation ability, and a systematic, standard, and comprehensive evaluation model of achievement transformation ability has not been formed, and the influence of the time factor is mostly not considered. In the above research context, it is necessary to evaluate and analyze the transformation capacity of emergency scientific research results under public health emergencies. On the one hand, it is beneficial to enrich the theoretical achievements related to emergency scientific research management, provide the evaluation index system and transformation ability evaluation model of emergency scientific research results, and provide theoretical support for improving the transformation ability of emergency scientific research results under public health emergencies. On the other hand, it is conducive to clarifying the objective status of the transformation capacity of emergency scientific research results under public health emergencies, specifying the characteristics and differences of the transformation capacity of emergency results of different subjects and regions, clarifying the spatial distribution characteristics of the transformation of regional emergency results, and providing a realistic basis for improving the transformation capacity of emergency scientific research results under public health emergencies.

As one of the core measures of innovation output, patents are effective carriers of science and technology innovation information, which are more accurate, informative, and forward-looking than journal literature. Patents often contain huge technical and commercial values. According to the statistics of the World Intellectual Property Organization, 80% of global inventions and technical knowledge can be traced back to patent literature, and more than 90% of the world's innovation information originates from patents [9, 10]. In recent years, more and more experts and scholars have analyzed the capacity, efficiency, and status of scientific and technological innovation and transformation based on patent indicators and data.

To sum up, from the perspective of the patent, the indicator system to evaluate the transformation ability of emergency scientific research results under major public health emergencies is constructed. Using the G1 method, deviation maximization method, index difficulty weighting method, and CRITIC weighting method, the minimum distance-maximum entropy combined weighting method is improved to achieve the static evaluation of transformation ability of emergency scientific research achievements for each subject and each region. Then, a dynamic evaluation model of the transformation ability of emergency scientific research achievements under public health emergencies is constructed, and the transformation ability of emergency scientific research achievements for different subjects and regions is dynamically evaluated.

Combined with the characteristics of data and index, this paper improves the ER index to measure the static polarization effect of regional emergency scientific research achievement transformation ability. On this basis, considering the influence of event development time, the dynamic polarization effect measurement model of the transformation ability of emergency scientific research achievements is established. Furthermore, the existing measurement model of polarization contribution degree (it is abbreviated to PCD in this work) is improved to analyze the polarization effect contribution degree of transformation ability of emergency scientific research achievements in different regions.

In order to achieve the above research objectives, this paper is structured as follows(the frame structure diagram of this work is shown in Figure 1): Section 2 adopts the G1 method, deviation maximization method, index difficulty weight method, and CRITIC weight method to improve the minimum distance-maximum entropy combination weight method; based on this, static evaluation and dynamic evaluation models of the transformation ability of emergency scientific research achievements in the context of public health emergencies are constructed; the static polarization effect and dynamic polarization effect of transformation ability of emergency scientific research achievement in each region are calculated by improving ER index. Furthermore, it improves the measurement model of contribution degree to polarization effect. Section 3 selects the number of valid invention patents ($X1$), patent maintenance period ($X2$), and the number of patent claims ($X3$) of emergency scientific research results under public health emergencies to construct an evaluation indicator system for the transformation ability of emergency scientific research results from the patent perspective. Section 4 carries out

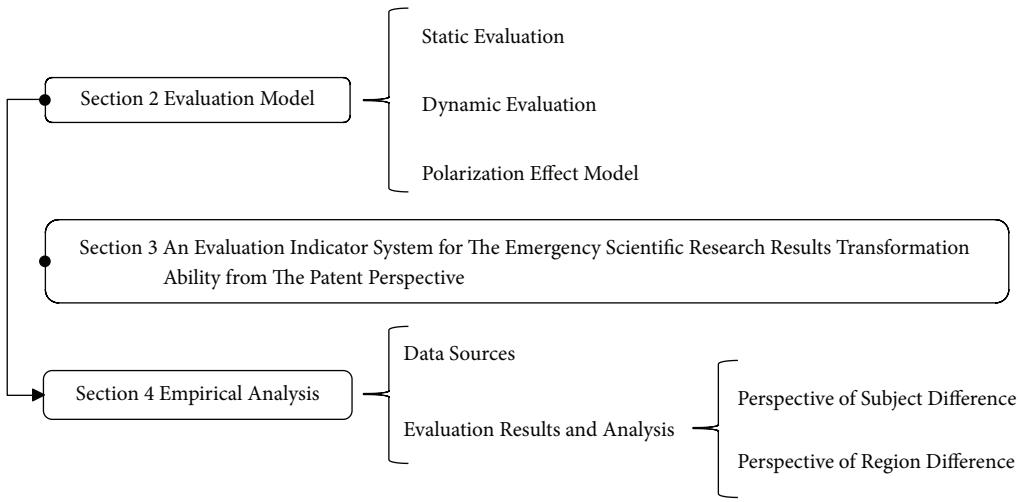


Figure 1: The logical frame structure diagram.

static evaluation and dynamic evaluation for the transformation ability of emergency scientific research achievements under public health emergencies in China from the perspective of subject difference and regional difference. Based on this, the improved polarization effect index measurement model is used to analyze the polarization effect of the transformation ability of emergency achievements in the national and eight comprehensive economic zones; furthermore, the improved measurement model of PCD is applied to analyze the polarization effect contribution of the transformation ability of emergency scientific research achievements in different regions. Finally, Section 5 presents the important findings, limitations, and possible future research directions.

2. Evaluation Model

The existing literature mostly uses the combined subjective-objective weighting method to avoid the defects of subjective and objective weighting methods, to take into account the advantages of both methods, so as to improve the rationality, accuracy, and credibility of evaluation results. The minimum distance-maximum entropy combination weighting method integrates the two factors of "distance" and "entropy" to calculate the combined weights of subjective and objective multiple weighting methods [11].

The reason for applying the combined subjective-objective weighting method in this work mainly includes the following two aspects: (1) The combined subjective-objective weighting method can take into account the advantages of both subjective and objective weighting methods, and avoid the disadvantages of both methods [12, 13]. According to their experience and knowledge, the experts rank the importance of the evaluation indexes and assign the subjective weight. This method reflects the subjective will of the experts on the importance of the evaluation indexes, and the weight obtained is highly interpretable. However, the subjective weight cannot reflect the data information of the evaluation index, and the decision or evaluation results have strong subjective arbitrariness with poor objectivity, which also increases the burden on the decision analysts. The objective weight reflects the data information of the evaluation index and determines the weight mainly according to the relationship between the original data. It does not depend on subjective judgment of people and does not increase the burden of decision analysts. The decision-making or evaluation results have a solid mathematical theoretical basis. However, it will change with the change of the evaluation object set, that is, the stability is weaker than the subjective weight. It cannot reflect the importance of decision-makers to different indicators. At the same time, the objective weight cannot reflect the importance of the evaluation index itself, and is less explanatory than the subjective weight. The combined subjective-objective weighting method embodies the thought of system analysis. The combined subjective-objective weighting method can take into account the preference of decision-makers for evaluation indicators, and at the same time, reduce the subjective arbitrariness of weight assignment effectively, so as to achieve the unity of subjective and objective weight assignment of indicators, and make the evaluation results more accurate and reliable. (2) There is a lack of data related to the transformation of emergency scientific research achievements, and the evaluation indicators are limited. In this context, if only the subjective weighting method is used for evaluation, the index weight is greatly affected by the subjective preference of the evaluators, which will seriously reduce the rationality, accuracy, and credibility of evaluation results. To sum up, this work uses the combined subjective-objective weighting method to avoid the defects of subjective and objective weighting methods, to take into account the advantages of both methods, so as to improve the rationality, accuracy, and credibility of evaluation results.

Above all, on the basis of the existing literature, the minimum distance-maximum entropy combination weighting method is improved by applying the G1 method, deviation maximization method, index difficulty weighting method, and CRITIC weighting method, so as to achieve the

static evaluation of the transformation ability of emergency scientific research achievements. Furthermore, a dynamic evaluation model of the transformation ability of emergency scientific research achievements under the background of public health emergencies is constructed.

2.1. Static Evaluation. The target standardization method is selected to normalize the original data of each index without dimensions. The calculation formula is as follows:

$$x_{ij}^{*} = \begin{cases} \dfrac{x_{ij}}{x_{0j}}, & x_{ij} \text{ is the positive index,} \\ \dfrac{x_{0j}}{x_{ij}}, & x_{ij} \text{ is the negative index,} \\ \dfrac{1}{1+\left|x_{ij}-x_{0j}\right|}, & x_{ij} \text{ is the moderate index,} \end{cases} \tag{1}$$

where x_{ij} is the original data of the index, x_{ij}^{*} is the index data after dimensionless processing, x_{0j} is the ideal value of the jth indicator, i is the ordinal number of the evaluation objects $(i = 1, 2, \ldots, n)$, and j is the ordinal number of the evaluation indicators $(j = 1, 2, \ldots, m)$.

2.1.1. G1 Method. As an efficient method for subjective assignment of indicators, the G1 method determines the relationship between the importance of indicators based on the sequential relationship between indicators, and then determines the G1 subjective weight of each indicator. This method facilitates the integration of the rich theoretical and practical experience of assessment experts [14]. The main steps are as follows:

(a) The experts rank the indicators x_j according to their importance, satisfying $x_1 \geq x_2 \geq \cdots \geq x_m$ $(j = 1, 2, \ldots, m)$.

(b) Calculate the importance ratio $r_k (k = m, m-1, \ldots, 3, 2)$ between adjacent indicators, i.e., the ratio of the relative importance degree of indicators $x_{(k-1)}$ and x_k. The reference for the assignment of r_k is listed in Table 1.

(c) Obtain the weight of the last indicator x_m.

$$w_m = \left(1 + \sum_{i=2}^{m}\prod_{k=i}^{m} r_k\right)^{-1}. \tag{2}$$

(d) Calculate the weights of other indicators in reverse order,

$$w_{(k-1)} = w_k \cdot r_k (k = m, m-1, \ldots, 3, 2). \tag{3}$$

Finally, the subjective weighting of indicators is realized and the subjective weight $[w_1, w_2, \ldots, w_m]$ of each indicator by G1 method is obtained..

Table 1: The assignment reference of r_k.

r_k	Interpretation
1.000	Indexes $x_{(k-1)}$ and x_k are equally important
1.200	Index $x_{(k-1)}$ is slightly more important(than x_k)
1.400	Index $x_{(k-1)}$ is more important(than x_k)
1.600	Index $x_{(k-1)}$ is strongly more important(than x_k)
1.800	Index $x_{(k-1)}$ is seriously more important(than x_k)

2.1.2. Deviation Maximization Method. The deviation maximization method is based on the statistical distribution characteristics of the specific data of the evaluation indexes, and the objective weighting of the indexes is achieved based on the difference between the data of each evaluation object under different indexes [15]. It is considered that the greater the difference between different data of an indicator, the more significant the impact of the indicator on the final evaluation results and the greater its weight. The calculation process is as follows:

(a) The deviation between the evaluation object M_i and other evaluation object M_j is denoted by $Q_{ij}(w)$,

$$\begin{aligned} Q_{ij}(w) &= \sum_{r=1}^{n}\left|x_{ij}^{*}\cdot w_j - x_{rj}^{*}\cdot w_j\right| \\ &= w_j \cdot \sum_{r=1}^{n}\left|x_{ij}^{*} - x_{rj}^{*}\right|, \quad r = 1, 2, \cdots, n, \end{aligned} \tag{4}$$

where, x_{ij}^{*} is the index data value after dimensionless processing.

(b) Let $Q_j(w)$ be the total deviation of all evaluation objects from other objects under the jth index,

$$\begin{aligned} Q_j(w) &= \sum_{i=1}^{n} Q_{ij}(w) \\ &= w_j \cdot \sum_{i=1}^{n}\sum_{r=1}^{n}\left|x_{ij}^{*} - x_{rj}^{*}\right|. \end{aligned} \tag{5}$$

Constructing and solving the following optimization model on the basis of the deviation maximization principle,

$$\begin{aligned} &\max Q(w) = w_j \cdot \sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{r=1}^{n}\left|x_{ij}^{*} - x_{rj}^{*}\right| \\ &s.t. \begin{cases} \sum_{j=1}^{m} w_j^2 = 1. \\ w_j \geq 0. \end{cases} \end{aligned} \tag{6}$$

(c) The optimal solution of the optimization model (6) is normalized to obtain the weights of each index,

$$w_j = \frac{\sum_{i=1}^{n}\sum_{r=1}^{n}\left|x_{ij}^{*} - x_{rj}^{*}\right|}{\sum_{j=1}^{m}\sum_{i=1}^{n}\sum_{r=1}^{n}\left|x_{ij}^{*} - x_{rj}^{*}\right|}. \tag{7}$$

2.1.3. Indicator Difficulty Weighting Method. The indicator difficulty weighting method completes objective evaluation based on the growth difficulty of evaluation indicators, and analyzes the growth difficulty of indicators according to the deviation degree between the maximum value and the mean value of indicator data, so as to ensure that the evaluation results do not lose balance due to the skewing of evaluation objects toward the large weighted indicators [16]. Evaluation indicators with greater growth difficulty usually have greater differentiation, and their weights can enhance the final evaluation result differentiation when they are larger. The weighting steps are as follows:

(a) Calculate the deviation degree z_j between the maximum value and the mean value of the j th evaluation index data,

$$z_j = \frac{\max_i\left(x_{ij}^*\right) - \overline{x_j^*}}{\sigma_j}. \tag{8}$$

(b) Obtain the objective weight w_j of the jth evaluation indicator under the indicator difficulty weighting method,

$$w_j = \frac{z_j}{\sum_{j=1}^m z_j}. \tag{9}$$

2.1.4. CRITIC Method. The CRITIC indicator weighting method calculates the objective weight of each indicator based on "contrast strength" and "conflict" [14, 17]. Among them, the "contrast strength" between indicators can be calculated by the standard deviation of indicators, and the "conflict" is mostly reflected by the correlation between indicators. This method is widely used in decision-making problems in many fields such as economics, management, information, and so on. The weighting procedure is as follows:

(a) Calculate the correlation coefficient $r_{jj'}$ of the $j(j = 1, 2, \ldots, m)$th indicator with other indicators based on dimensionless processing of data values,

$$r_{jj'} = \frac{1/n - 1\sum_{i=1}^n\left(y_{ij} - \overline{y}_{\cdot j}\right)\left(y_{ij'} - \overline{y}_{j'}\right)}{\sigma_j\sigma_{j'}}, \tag{10}$$

where, $\overline{y}_j$ is the mean value of the jth index and σ_j is the standard deviation of the jth index.

(b) Obtain the total conflict f_j of the jth index,

$$f_j = \sum_{j'=1}^m\left(1 - r_{jj'}\right). \tag{11}$$

(c) Calculate the amount of information c_j for the jth indicator.

$$c_j = \sigma_j f_j. \tag{12}$$

(d) Obtain the weight w_j of the jth index under the CRITIC weighting method,

$$w_j = \frac{c_j}{\sum_{j=1}^m c_j}. \tag{13}$$

2.1.5. Improved Combination Weighting Method with Minimum Distance-Maximum Entropy. The weights $w_j^{(q)}$ $(q = 1, 2, 3, 4)$ are obtained by combining the G1 method, the deviation maximization method, the indicator difficulty weighting method, and the CRITIC weighting method, respectively. Based on the existing literature [11], the improved combined weights are obtained as

$$w_j = \sum_{q=1}^4 \lambda_q w_j^{(q)}, \tag{14}$$

where λ_q denotes the weight combination coefficient $(\sum_{q=1}^4 \lambda_q = 1)$, and $w_j^{(q)}$ denotes the weight of the jth indicator obtained by the qth weighting method. The combination coefficient λ_q is determined by considering the factors of "distance" and "entropy value" [18–20].

(a) Take the minimum value of the generalized distance between weighted values of different evaluative objects and the ideal point,

$$\min\sum_{i=1}^n d_i = \sum_{i=1}^n\sum_{j=1}^m\sum_{q=1}^4 \lambda_q w_j^{(q)}\left(1 - x_{ij}^*\right). \tag{15}$$

(b) The "maximum entropy" principle is applied to prevent individual weighting methods from being removed due to their little impact on the combined assignment results. The following objective function is constructed based on the principle of "consistency maximization" of the weight assignment results,

$$\max H = -\sum_{q=1}^4 \lambda_q \ln\lambda_q. \tag{16}$$

Considering the above objectives together, the model is constructed:

$$\min\left[\mu\sum_{i=1}^n\sum_{j=1}^m\sum_{q=1}^4 \lambda_q w_j^{(q)}\left(1 - x_{ij}^*\right) + (1-\mu)\sum_{q=1}^4 \lambda_q \ln\lambda_q\right]$$
$$s.t.\begin{cases}\sum_{q=1}^4 \lambda_q = 1,\\ \lambda_q \geq 0,\end{cases} \tag{17}$$

where the parameter $0\langle\mu\langle 1$ represents the balance coefficient between the two objectives (distance minimization, entropy maximization), usually taken as $\mu = 0.5$, and x_{ij}^* is the data value after dimensionless process of the index.

(c) By solving model (17), the optimal solution λ_q of the weight combination coefficients is obtained. The solved λ_q is substituted into equation (14) to obtain

Table 2: The importance of time.

θ	Interpretation
0.100	Great attention to recent data
0.300	Attention to recent data
0.500	Equal attention to data from all periods
0.700	Attention to long-term data
0.900	Great importance to long-term data
0.200, 0.400, 0.600, 0.800	Corresponding to the intermediate situation of the above two adjacent judgments

the minimum distance-maximum entropy combination weight w_j for each evaluation index.

Furthermore, the static evaluation value I_i of the transformation ability of scientific research achievements under public health emergencies is calculated,

$$I_i = \sum_{j=1}^{m} w_j \cdot x_{ij}, \quad i = 1, 2, \cdots, n. \tag{18}$$

2.2. Dynamic Evaluation. The realization of dynamic comprehensive evaluation must first determine the time series weights [11, 21]. On the basis of static evaluation, the influence of time factors is emphasized, and the time-series weighted average operator is applied for secondary weighting, so as to achieve a comprehensive dynamic evaluation of each evaluation object in the time interval $[1, h]$. The main evaluation steps are as follows:.

(a) Obtain the static evaluation value $I_i(t)$ of the ith evaluated object at moment t based on the static evaluation model,

(b) Calculate the time series weights. Set the time series weight vector be w_t: $w_t = (w_1, w_2, \ldots, w_h)^T$, solve the model of the nonlinear programming problem below, and obtain w_t,

$$\max\left(-\sum_{t=1}^{h} w_t \ln w_t\right)$$
$$s.t.\begin{cases} \sum_{t=1}^{h} (h-t)/(h-1) w_t = \theta, \\ \sum_{t=1}^{h} w_t = 1, \end{cases} \tag{19}$$

where the parameter θ is the importance degree of time, and its assignment values are listed in Table 2.

(c) Using the time-series weighted average operator, the static evaluation value $I_i(t)$ obtained from (18) is secondarily weighted to realize the dynamic comprehensive evaluation of the ith evaluation object,

$$L_i = A\left(\langle 1, I_i(1)\rangle, \langle 2, I_i(2)\rangle, \cdots, \langle h, I_i(h)\rangle\right) = \sum_{t=1}^{h} I_i(t) \cdot w_t. \tag{20}$$

2.3. Polarization Effect Model. The theory of polarization effect was first proposed by Perroux [22]. The most commonly used polarization effect indices in the existing literature are ER index, TW index, KZ index, and Wolfson index. Scholars use single polarization index or multi-polarization index to measure the level of regional polarization effect, and mostly use the above polarization indexes to analyze the spatial polarization effect in the fields of regional economic development, scientific and technological innovation, and industrial innovation, etc. At the same time, many research results show that under the influence of institutions and policies, science and technology innovation mostly presents spatially unbalanced development [23]. Theoretical models such as "growth poles," "gradient shift," and "core-edge" all emphasize the unbalanced allocation of scientific and technological innovation resources under the condition of limited resources. In the abovementioned research context, scholars have used the polarization effect model to analyze the spatial evolution characteristics of the regional innovation polarization effect.

2.3.1. Improved ER Index. Based on the basic principle of ER index and referring to existing literature [24], the static polarization effect measurement model of transformation capability of regional emergency scientific research results is constructed, and the static polarization degree (it is abbreviated to SPD in this work) index of emergency scientific research results transformation capability f_{ER_S} is calculated,

$$f_{ER\_S} = A^* \sum_{i=1}^{n} \sum_{j=1}^{n} p_i^{1+\partial} p_j \left| I_i - I_j \right|, \tag{21}$$

where n is the total number of evaluation objects in a region, i.e., the number of evaluated areas in a region; p_i and p_j are the proportion of a certain variable value (in the ith and jth area) in the whole country or the region, respectively. A^* is the standardized coefficient of polarization effect index, $A^* = 1 + 2\mu^{1+\alpha}$ (where, $\mu = \sum_{i=1}^{n} p_i I_i$). Meanwhile, $\alpha \in [0, 1.6]$, the value should be as large as possible, which is conducive for reflecting the polarization characteristics of the region, and usually takes $\alpha = 1.5$. Different from the original model, I_i and I_j are the static evaluation values of emergency scientific research results transformation ability in ith and jth areas within a certain region under a public health emergency, respectively.

2.3.2. Dynamic Polarization Degree. Based on the static polarization index f_{ER_S} of the transformation capability of emergency scientific research achievements, considering the influence of time factors, the time-series weighted average operator is applied to realize the secondary weighting of f_{ER_S}. The dynamic polarization effect measurement model of the transformation capability of emergency scientific research results during the time interval of $[1, h]$ is constructed, and the dynamic polarization degree (which is abbreviated to DPD in this work) index f_{ER_D} of the transformation capability of regional emergency scientific research results is calculated,

$$f_{ER\_D} = A\left(\langle 1, f_{ER\_S}(1)\rangle, \langle 2, f_{ER\_S}(2)\rangle, \cdots, \langle h, f_{ER\_S}(h)\rangle\right) = \sum_{t=1}^{h} f_{ER\_S}(t) \cdot w_t, \tag{22}$$

where $f_{ER\_S}(t)$ is the SPD index of emergency scientific research results transformation capability at moment t, w_t is the time series weight, and $t \in [1, h]$.

2.3.3. Polarization Contribution Degree. In the context of emergencies, the emergency scientific research resources are limited, the time is urgent, the demand is changeable, and the relevant evaluation index data of achievement transformation ability are limited. In order to further analyze the contribution of a certain area to the polarization effect of the transformation ability of emergency scientific research achievements in the region, on the basis of existing literature and combined with the characteristics of the evaluation index data of the transformation ability of emergency science research achievements, the measurement model of PCD is improved, so as to construct the measurement model of polarization contribution degree for the transformation ability of emergency science research results under public health emergencies, and calculate the polarization contribution C_{ERi} of the transformation ability of emergency scientific research achievements in i th area,

$$C_{ERi} = 1 - \frac{f_{ER\_D}^{(i)}}{f_{ER\_D}}, \tag{23}$$

where $f_{ER\_D}^{(i)}$ denotes the dynamic polarization degree index of a region without ith area, and f_{ER_D} denotes the dynamic polarization degree index of a region containing ith area. The PCD C_{ERi} is greater than 0, which indicates that there is a promoting effect of ith area on the polarization effect of emergency scientific research results transformation ability in the region; conversely, it produces an inhibiting effect.

3. An Evaluation Indicator System for the Emergency Scientific Research Results Transformation Ability from the Patent Perspective

The patent is an important carrier of achievement transformation and scientific research innovation, and an important indicator for judging the level of achievement transformation and the strength of scientific research innovation, which has an important impact on technology breakthroughs and innovation development. The number of patents and their changing trends often reflect the scientific research strength and innovation competitiveness of a country or region [25]. In recent years, patent data and information have become an important evaluation index for domestic and foreign experts and scholars to analyze the achievement transformation ability and innovation potential of scientific research. The U.S. "Bayh-Dole Act" promotes the achievement transformation in the form of patent law amendments, which also shows the particularity and importance of patent for the achievement transformation from the side [26]. Therefore, it is reasonable and feasible to analyze the ability, efficiency, and current situation of scientific & technological innovation and achievement transformation based on patent data.

In the context of major public health emergencies, the statistical caliber and cycle of monthly progress data in various regions are different. Some districts only publish quarterly data, not monthly progress data. Therefore, according to the principle of index system construction, "systematic, typical, dynamic, scientific, operable, and comprehensive," applying with the "grounded theory" and Delphi method, the primary index set and the optimized index set are gradually obtained through the rigorous index system construction process. Then, combining the availability of progress data (monthly), the index set is revised and improved after discussion with experts and team members. Finally, the evaluation indicator system for the emergency scientific research results transformation ability from the patent perspective is constructed (as shown in Table 3). Due to the limited space, more specific construction process, selection basis, and connotation explanation of the index system can be obtained by contacting the author.

After a rigorous screening process of indicators, the number of valid invention patents ($X1$), patent maintenance period ($X2$), and number of patent claims ($X3$) of emergency scientific research results under public health emergencies are selected to construct an index system to evaluate the transformation ability of emergency science research results from the patent perspective, as shown in Table 3. Among them, the number of valid invention patents and the number of patent claims are both positive indicators, and the patent maintenance period is an adaptive indicator. Specifically, the number of valid invention patents is a key indicator reflecting the quality of patents and an important indicator measuring the contribution of industrial economy to employment; the patent maintenance period is one of the important indicators reflecting the operation effectiveness of the patent system and the advantages and disadvantages of the patent maintenance system, and is regarded as a key factor measuring the technical and economic value of patents; the number of patent claims is the core of patents and determines the protection scope of patent rights. Based on the above evaluative indicator system, the transformation ability of emergency science research results in public health emergencies is evaluated from the patent perspective.

Table 3: Evaluation index.

Index name	Index symbol	Index attribute	Index interpretation
Number of valid invention patents	$X1$	+	The number of patents approved and granted by the SIPO. It is not only a key indicator of patent quality but also an important indicator of industrial economic employment contribution. A valid invention patent is more conducive to obtaining the competitive advantage in technology and product innovation.
Patent maintenance period	$X2$	+/−	It is one of the critical indicators to reflect the effectiveness of patent system and the merits of patent maintenance system, which is seen as a key factor in measuring the value of technology and economics for the patent. The longer the patent maintenance period, the higher the innovation quality, competitive strength, and economic value of the patented technology, that is, the more likely the patented technology to create economic benefits for the patent holder. At the same time, the cost of patent maintenance should be considered and the maintenance period should be set rationally.
Number of patent claims	$X3$	+	It is the core of the patent, and determines the scope of patent protection. The more claims a patent has, the more key features of the patented technology, and the more cutting-edge, important, and economically valuable the patent is.

4. Empirical Analysis

The static evaluation and dynamic evaluation of the transformation ability of emergency science research results under public health emergencies in China are conducted from two perspectives: subject differences and regional differences. From the perspective of subject differences, the transformation ability of emergency scientific research results of five major subjects–enterprise, institution (mainly hospital), individual, universities, and research institute–are evaluated. From the perspective of regional differences, the capacity of 31 provinces, municipalities, and autonomous regions in China are evaluated (Hong Kong, Macao, and Taiwan are not evaluated for the time being due to insufficient data); based on this, the improved polarization effect index measurement model is used to analyze the polarization effect of the transformation ability of emergency achievements in the national and eight comprehensive economic zones; furthermore, the improved measurement model of PCD is applied to analyze the polarization effect contribution of the transformation ability of emergency scientific research achievements in different regions.

4.1. Data Sources. The empirical data originate from "Patyee" patent search database. The database is an authoritative search and analysis platform with independent intellectual property rights, which provide data services for government departments, universities, research institutes, and enterprises all year round; the data are accurate, comprehensive, and timely updated. It covers more than 100 countries and regions, containing more than 140 million patent data. The search keywords are set to "COVID-19" or "Novel coronavirus," the patent status is set to "valid," and the time limit for patent disclosure is set from "March 2020" to "May 2021." After filtering and collating, we obtained a total of 458 (for 15 months) valid patents related to "COVID-19" or "Novel coronavirus" from 31 provinces, municipalities, and autonomous regions in China.

4.2. Evaluation Results and Analysis

4.2.1. Perspective of Subject Difference

(1) Static Evaluation.

Based on the improved minimum distance-maximum entropy combination weighting method, the combination weights $w_j\,(j = 1, 2, 3)$ of the evaluation indexes $X1$–$X3$ for the transformation ability of emergency scientific research results in public health emergencies under the perspective of differences in scientific research subjects are calculated and shown in Table 4.

Based on equation (18), the static evaluation results of the transformation ability of emergency research results are obtained for each subject (from March 2020 to May 2021) under the background of "COVID-19" epidemic (shown in Table 5). In order to analyze the static evaluation results more visually and present the rules and differences, Figure 2 is drawn according to Table 5.

As can be analyzed from Table 5 and Figure 2, in the static evaluation of the transformation ability of emergency scientific research results of different scientific research subjects under the "COVID-19" epidemic, the average value of the transformation ability (from March 2020 to May 2021) ranks as follows: enterprise (0.920), institution (0.256), individual (0.158), university (0.152), and research institute (0.126). Among them, the transformation ability of enterprise emergency scientific research achievements is generally strong, far exceeding the average value of the same period, and the transformation power of achievements is lasting; transformation ability of emergency scientific research results of institution is lower than enterprise, close to the average value of the same period, with an overall "W"-type change trend; emergency scientific research results transformation ability of the individual, university, and research institute is close, lower than the average value of the same period. At the same time, enterprise, institution, and individual all demonstrated a strong transformation ability to emergency scientific research results at the early stage of the epidemic, providing strong scientific research results to

Table 4: Improved combination evaluation weight (subject difference perspective).

Index symbol	202003	202004	202005	202006	.....	202102	202103	202104	202105
$X1$	0.355	0.369	0.342	0.310	.....	0.321	0.323	0.325	0.337
$X2$	0.319	0.354	0.367	0.352	.....	0.388	0.381	0.386	0.372
$X3$	0.325	0.277	0.290	0.338	.....	0.291	0.295	0.289	0.290

Table 5: Static evaluation results of the transformation ability of emergency scientific research achievements (subject difference perspective).

Subject	202003	202004	202005	.....	202102	202103	202104	202105	Mean	Rank
Enterprise	0.584	0.999	0.999	.....	0.950	0.920	0.980	0.999	0.920	1
Institution	0.674	0.235	0.254	.....	0.197	0.160	0.337	0.326	0.256	2
Individual	0.692	0.122	0.097	.....	0.089	0.051	0.099	0.158	0.158	3
University	0.007	0.003	0.251	.....	0.184	0.089	0.127	0.161	0.152	4
Research institute	0.007	0.003	0.083	.....	0.196	0.113	0.189	0.004	0.126	5

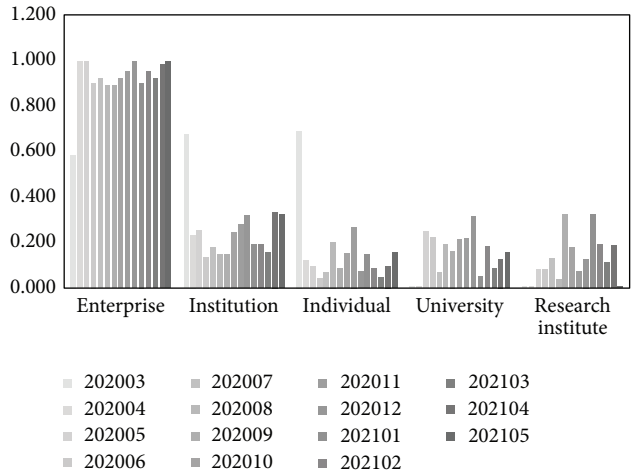


Figure 2: Static evaluation results of the transformation ability of emergency scientific research achievements (subject difference perspective).

support the epidemic response. In addition, enterprise is an important innovation subject in science and technology R&D, and hospital is an important scientific research subject in the clinical front line, both of which have a strong transformation ability of emergency scientific research results, The evaluation results are in line with objective facts, which proves the reasonableness, reliability, and scientificity of the evaluation model in this paper from the side.

(2) Dynamic Evaluation.

Based on the dynamic evaluation step (b), the time series weights $w_t\,(t = 1, 2, \ldots, 15)$ are determined (as shown in Table 6). To highlight the importance of recent data, set $\theta = 0.4$.

Based on equation (20), the dynamic evaluation results of the transformation ability of emergency research results of each subject under the "COVID-19" epidemic are obtained (as shown in Table 7). The dynamic evaluation values of each indicator are also calculated (seen in Table 8). For visual analysis, Figure 3 is drawn.

As can be seen from Table 7, in the dynamic evaluation of the transformation ability of emergency scientific research achievements for each subject, enterprise has the strongest ability to transform emergency scientific research results and exceeds the average value; institution (mainly hospital) has weaker ability to transform emergency scientific research results than enterprises and is close to the average value; the transformation ability of emergency science research results of individual, university, and research institute is weaker and lower than the average value. As can be seen from Table 8 and Figure 3, enterprise ranks high in all aspects of the transformation of emergency achievements, and the advantage of patent maintenance period is more prominent; other subjects have smaller differences in the number of valid invention patents and claims, and the gap between their patent maintenance period and enterprise is larger.

4.2.2. Perspective of Region Difference

(1) Static Evaluation.

Based on the improved minimum distance-maximum entropy combination weighting method, the combination weights $w_j\,(j = 1, 2, 3)$ of the evaluation indexes $X1$–$X3$ for the transformation ability of emergency scientific research results in public health emergencies under the perspective of regional differences are calculated and shown in Table 9.

Based on equation (18), the static evaluation results of the transformation capacity of emergency research results by region and month are obtained under the "COVID-19" epidemic (as shown in Table 10). In order to analyze the static evaluation results more visually and present the rules and differences, Figure 4 is drawn according to Table 10.

As can be seen from Table 10 and Figure 4, the top 5 regions in the static evaluation of the transform ability of emergency scientific research achievements in different regions (from March 2020 to May 2021) under the "COVID-19" epidemic are Guangdong (0.673), Beijing (0.580), Jiangsu (0.541), Shanghai (0.272),and Tianjin (0.180). Among them, the transformation ability of emergency scientific research achievements of Guangdong is generally strong, exceeding the average value of the same period as a whole; Beijing is weaker than Guangdong, but still at a higher level and exceeds the mean value of the same period;

Table 6: The time series weights.

Month	2020.03	2020.04	2020.05	2020.06	.....	2021.02	2021.03	2021.04	2021.05
w_t	0.037	0.040	0.043	0.046	.....	0.086	0.093	0.100	0.108

Table 7: Dynamic evaluation results of the transformation ability of emergency scientific research achievements (subject difference perspective).

Subject	Enterprise	Institution	Individual	University	Research institute	Mean
L_j	0.933	0.251	0.141	0.156	0.134	0.333
Rank	1	2	4	3	5	

Table 8: Dynamic evaluation results of first-level indicators (subject difference perspective).

Index symbol	Enterprise	Institution	Individual	University	Research institute	Mean
$X1$	0.311 (1)	0.136 (2)	0.071 (4)	0.074 (3)	0.066 (5)	0.135
$X2$	0.342 (1)	0.017 (2)	0.003 (4)	0.003 (3)	0.003 (5)	0.077
$X3$	0.280 (1)	0.098 (2)	0.067 (4)	0.079 (3)	0.065 (5)	0.121

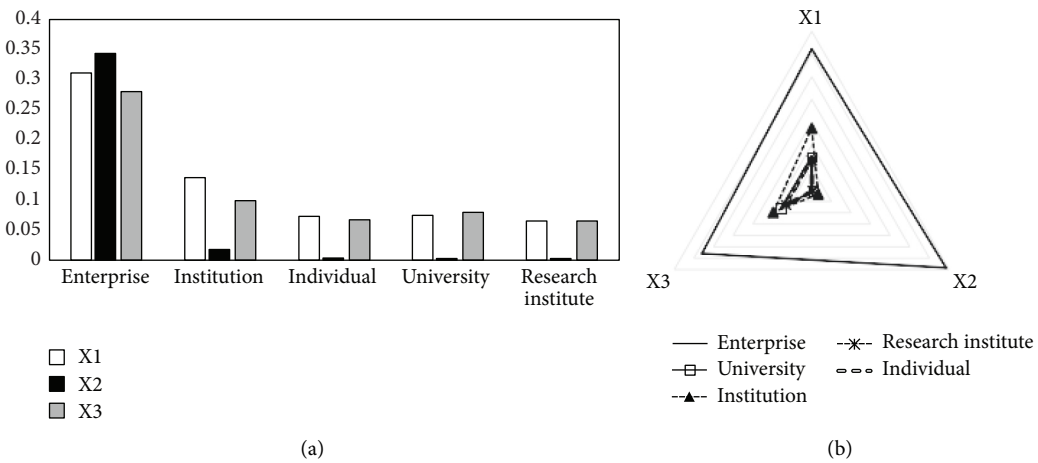


Figure 3: Dynamic evaluation results of first-level indicators (subject difference perspective).

Table 9: Improved combination evaluation weight (region difference perspective).

Index symbol	202003	202004	202005	202006	.....	202102	202103	202104	202105
$X1$	0.509	0.414	0.434	0.456	.....	0.395	0.474	0.407	0.486
$X2$	0.193	0.236	0.189	0.187	.....	0.179	0.155	0.131	0.154
$X3$	0.298	0.350	0.378	0.357	.....	0.426	0.371	0.462	0.360

Jiangsu exceeds the average value in the same period but fluctuates more obviously and shows a downward trend in the near future; Shanghai shows a significant upward trend in the near future and has outstanding development potential; Tianjin is lower than the average value of the same period, and the declining trend is more significant. Further analyze the transformation characteristics of emergency achievements for Guangdong, Beijing, and Jiangsu. Specifically enterprises and institutions in Guangdong have a large number of patents related to "COVID-19" with good quality; the transformation work of emergency scientific research results in Guangdong started quickly (after the outbreak of the epidemic, a strong transformation ability of emergency scientific research results is rapidly formed), and the advantages of emergency R&D is outstanding, the achievements of which are mostly concentrated in reagents (boxes), pharmaceutical compositions, epidemic prevention equipment, etc. Enterprises and institutions in Beijing also show strong advantages in emergency R&D, with achievements mostly concentrated in reagents (boxes), vaccines, and their preparation methods, peptides or their combinations for detecting (or inhibiting) viruses, etc. Enterprises and research institutes in Jiangsu are more prominent in emergency R&D, with results mostly focused on epidemic

Table 10: Static evaluation results of the transformation ability of emergency scientific research achievements (region difference perspective).

Region	202003	202004	202005	202006	……	202102	202103	202104	202105	Mean	Rank
Beijing	0.004	0.244	0.298	0.465		1.000	0.558	0.872	0.874	0.580	2
Tianjin	0.004	0.003	0.091	0.167		0.132	0.001	0.001	0.004	0.180	5
……	……	……	……	……	……	……	……	……	……	……	
Liaoning	0.702	0.003	0.001	0.001		0.017	0.001	0.001	0.198	0.067	10
……	……	……	……	……	……	……	……	……	……	……	
Shanghai	0.004	0.003	0.253	0.274		0.052	0.182	0.191	0.750	0.272	4
Jiangsu	0.714	0.117	0.237	0.111		0.197	1.000	0.541	0.500	0.541	3
Zhejiang	0.004	0.003	0.001	0.001		0.174	0.076	0.257	0.004	0.096	9
Anhui	0.814	0.003	0.001	0.001		0.001	0.001	0.038	0.126	0.098	8
Fujian	0.004	0.003	0.001	0.001		0.001	0.076	0.001	0.004	0.051	15
……	……	……	……	……	……	……	……	……	……	……	
Shandong	0.004	0.003	0.087	0.001		0.042	0.068	0.206	0.004	0.062	12
……	……	……	……	……	……	……	……	……	……	……	
Hubei	0.004	0.072	0.075	0.001		0.049	0.119	0.067	0.181	0.065	11
Hunan	0.004	0.269	0.091	0.119		0.056	0.001	0.001	0.004	0.101	7
……	……	……	……	……	……	……	……	……	……	……	
Guangdong	0.004	1.000	1.000	1.000		0.439	0.382	0.832	0.592	0.673	1
Guangxi	0.004	0.003	0.079	0.001		0.001	0.001	0.169	0.004	0.051	14
……	……	……	……	……	……	……	……	……	……	……	
Sichuan	0.004	0.003	0.001	0.062		0.235	0.283	0.001	0.346	0.136	6
……	……	……	……	……	……	……	……	……	……	……	
Ningxia	0.004	0.003	0.001	0.001		0.001	0.001	0.366	0.172	0.054	13
……	……	……	……	……	……	……	……	……	……	……	
Mean	0.076	0.067	0.078	0.075		0.083	0.100	0.128	0.129		

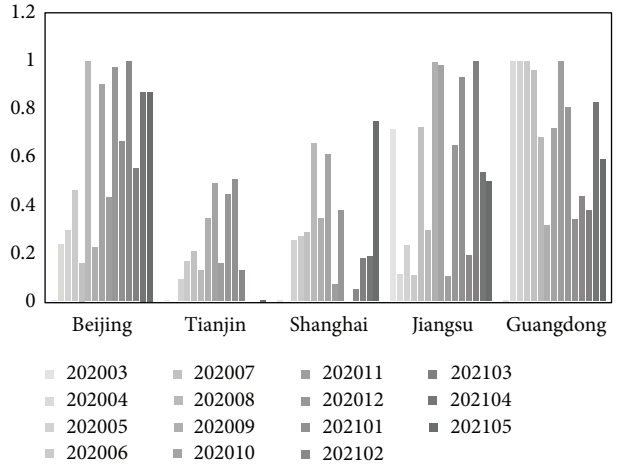


Figure 4: Static evaluation results of the transformation ability of emergency scientific research achievements (region difference perspective).

prevention equipment, reagents (boxes), interfering nucleic acids and their combinations, etc.

(2) Dynamic Evaluation.

According to the dynamic evaluation step (b), the time series weights w_t ($t = 1, 2, \cdots, 15$) are determined by making $\theta = 0.4$, and the results are consistent with Table 6. Based on equation (20), the dynamic evaluation results of the transformation capacity of emergency scientific research results in each region under the "COVID-19" epidemic are obtained (as shown in Table 11). The dynamic evaluation values of each indicator are also calculated (As shown in Table 12.). For visual analysis, Figure 5 is drawn.

As can be seen from Table 11, the top 5 regions in the dynamic evaluation of the transformation ability of the emergency scientific research results are Guangdong (0.654), Beijing (0.651), Jiangsu (0.567), Shanghai (0.288), and Tianjin (0.178). Among them, Guangdong has the strongest ability to transform the emergency scientific research results, which is much larger than the average value; Beijing is second only to Guangdong in the transformation ability of the emergency scientific research achievements, which is also much larger than the average value. From Table 12 and Figure 5, it can be analyzed that Guangdong ranks first in the dynamic evaluation of all indicators of transformation of emergency achievements, the dynamic evaluation of each indicator exceeds its average value, and the advantages of the patent maintenance period and the number of claims for Guangdong are more prominent; Beijing ranks first in the dynamic evaluation of the number of valid invention patents, and its dynamic evaluation of other indicators also exceeds the average value; Jiangsu, Shanghai, and Tianjin are relatively consistent in the dynamic evaluation of each indicator, and Jiangsu has certain advantages in patent maintenance period. The dynamic evaluation of all indicators of Shanghai and Tianjin are lower than the average value.

(3) Transformation Ability of Emergency Scientific Research Achievements in the Eight Comprehensive Economic Zones.

For the analysis of spatial distribution characteristics of the transformation capability of emergency scientific

Table 11: Dynamic evaluation results of the transformation ability of emergency scientific research achievements (region difference perspective).

Region	L_j	Rank
Beijing	0.651	2
Tianjin	0.178	5
Hebei	0.050	17
Shanxi	0.022	21
Inner Mongolia	0.002	27
Liaoning	0.055	13
Jilin	0.037	19
Heilongjiang	0.013	22
Shanghai	0.288	4
Jiangsu	0.567	3
Zhejiang	0.113	7
Anhui	0.082	9
Fujian	0.050	16
Jiangxi	0.002	27
Shandong	0.067	12
Henan	0.054	14
Hubei	0.071	11
Hunan	0.087	8
Guangdong	0.654	1
Guangxi	0.051	15
Hainan	0.002	27
Chongqing	0.006	25
Sichuan	0.157	6
Guizhou	0.031	20
Yunnan	0.042	18
Tibet	0.002	27
Shaanxi	0.008	23
Gansu	0.007	24
Qinghai	0.002	27
Ningxia	0.075	10
Xinjiang	0.005	26

research results more accurately, the average value of the static (dynamic) evaluation value of the transformation capability of emergency scientific research results in each province and city within each economic zone is taken as the static (dynamic) evaluation value of the transformation capability of emergency scientific research achievements in that economic zone (according to the principle of dividing eight comprehensive economic zones in China), as shown in Tables 13 and 14. From the table, it could be analyzed that, under the background of the "COVID-19" epidemic, the top 3 comprehensive economic zones in terms of dynamic evaluation of transformation ability of emergency scientific research results are the Eastern Coastal Economic Zone (0.322), Northern Coastal Economic Zone (0.237), and Southern Coastal Economic Zone (0.235), which have stronger transformation ability of emergency scientific research results; Middle Yellow River Economic Zone and Northwest Economic Zone have lower dynamic evaluation value of transformation ability of emergency scientific research results and weaker transformation ability of emergency scientific research results. In order to more intuitively analyze the static evaluation results of transformation ability of emergency scientific research results for each comprehensive economic zone, Figure 6 is drawn. It can be seen from this figure that the Eastern Coastal Economic Zone, Northern Coastal Economic Zone, and Southern Coastal Economic Zone maintain a high level of transformation ability of emergency scientific research results in each month from March 2020 to May 2021, which is consistent with the dynamic evaluation results. In conclusion, during the response to the "COVID-19" epidemic, the eight comprehensive economic zones have significant differences in the transformation capacity of emergency scientific research results and unbalanced spatial distribution.

(4) Polarization Effect Analysis.

Based on the static polarization effect measurement model (equation (21)), the SPD of transformation capability of emergency scientific research results in China is calculated; furthermore, according to equation (22), the dynamic polarization degree (which is abbreviated to DPD in this work) of emergency scientific research results in China is calculated, and the results are illustrated in Figure 7. The analysis shows that the SPD of transformation capability of emergency scientific research results in China fluctuates under the "COVID-19" epidemic, and the polarization level is low, indicating that the resource aggregation status of national emergency research is constantly adjusting with the epidemic development stage, and the resource aggregation process in the global scope has not yet shown the characteristics of order and large scale. The "polarization" and

Table 12: Dynamic evaluation results of first-level indicators (region difference perspective).

Index symbol	Beijing	Tianjin	Shanghai	Jiangsu	Guangdong	Mean
*X*1	0.346 (1)	0.099 (5)	0.149 (4)	0.283 (3)	0.308 (2)	0.237
*X*2	0.055 (2)	0.002 (5)	0.005 (4)	0.051 (3)	0.064 (1)	0.035
*X*3	0.250 (2)	0.077 (5)	0.134 (4)	0.233 (3)	0.283 (1)	0.195

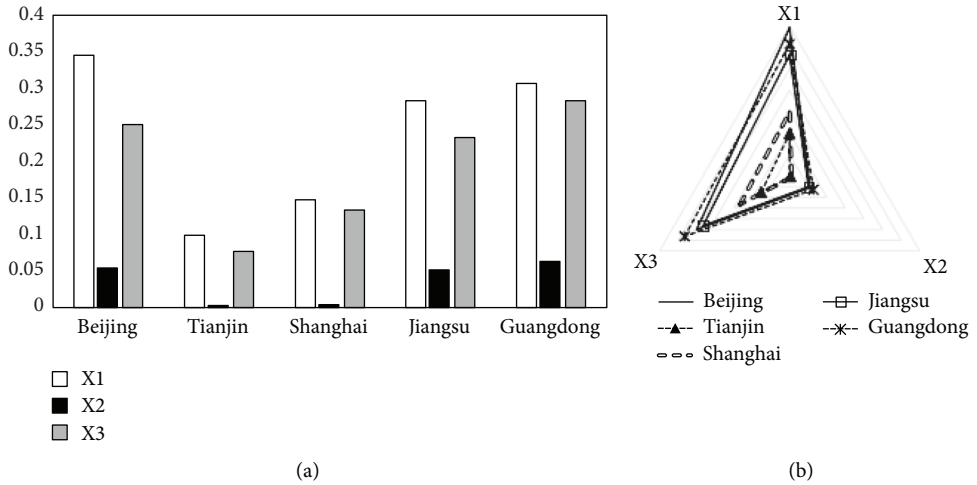


Figure 5: Dynamic evaluation results of first-level indicators (region difference perspective).

"diffusion" of transformation capability of emergency scientific research results in China are alternating, and the polarization effect is weak. In addition, the dynamic polarization level of transformation ability of emergency scientific research results in China is low, indicating that the transformation ability of emergency scientific research results at the national level does not form a significant polarization effect considering the influence of time factor.

Furthermore, the top 3 comprehensive economic zones in terms of the dynamic evaluation value of the transformation capacity of emergency scientific research results under the "COVID-19" epidemic are taken as the research objects to analyze their polarization effect of the transformation capacity of emergency scientific research results, as shown in Tables 13 and 14. It can be seen from the table that the static polarization level of the transformation capacity of emergency scientific research achievements in the Northern Coastal Economic Zone, Eastern Coastal Economic Zone, and Southern Coastal Economic Zone is significantly different, and the DPD is low; the static (dynamic) polarization effect of transformation ability of emergency scientific research results in the Southern Coastal Economic Zone is weaker. The results show that the above three comprehensive economic zones have not formed a more obvious and stable polarization effect of emergency scientific research results transformation ability; under the "COVID-19" epidemic, their resources gathering of emergency scientific research results transformation has not formed an orderly and large-scale development trend.

In order to more intuitively analyze the polarization effect of the top 3 comprehensive economic zones in terms of the transformation ability of emergency scientific research results, Figure 8 is drawn according to Table 14. It can be seen from the figure that under the "COVID-19" epidemic, the polarization effect of the transformation ability of emergency scientific research achievements in the above three comprehensive economic zones fluctuates obviously, and the polarization level is high in some months; the polarization level of the transformation capacity of emergency scientific research results in the Southern Coastal Economic Zone is low and the most unstable. The results show that the trend of resource aggregation of emergency research results transformation in the above three economic zones is unstable, and the "polarization" and "diffusion" effects of transformation ability of emergency research results occur alternately. This is consistent with the characteristics of emergency science research results transformation. Within the process of responding to major public health emergencies, due to multiple factors such as time urgency, limited resources, changing needs, and differences in the foundation, the trend of resource aggregation and capacity polarization of emergency scientific research results in each region is unstable, which are constantly adjusted and optimized with the development stage of the event. Through coordination and cooperation with surrounding areas, an efficient and stable transformation capacity of emergency scientific research achievements are gradually formed. This is a long-term process of continuous iteration and evolution.

Based on the conclusion of the above polarization degree analysis and on equation (23), the PCD of emergency scientific research results transformation capacity for each province and city within the three major economic zones of the Northern Coastal Economic Zone, Eastern Coastal

Table 13: Transformation ability of emergency scientific research achievements in the eight comprehensive economic zones.

Comprehensive economic zones	202003	202004	202005	202006	202007	202008	202009	202010	202011	202012	202101	202102	202103	202104	202105	Dynamic evaluation	Rank
Northeast economic region	0.237	0.003	0.001	0.017	0.001	0.052	0.002	0.062	0.054	0.086	0.002	0.006	0.001	0.001	0.068	0.035	6
Northern coastal economic zone	0.004	0.063	0.139	0.158	0.095	0.314	0.145	0.414	0.173	0.413	0.317	0.305	0.190	0.290	0.221	0.237	2
Eastern coastal economic zone	0.241	0.041	0.164	0.129	0.372	0.320	0.450	0.534	0.150	0.471	0.364	0.141	0.419	0.329	0.418	0.322	1
Southern coastal economic zone	0.004	0.335	0.334	0.334	0.367	0.229	0.108	0.395	0.334	0.271	0.137	0.147	0.153	0.278	0.200	0.235	3
Middle yellow river economic zone	0.004	0.029	0.001	0.015	0.001	0.002	0.002	0.002	0.001	0.054	0.025	0.001	0.048	0.034	0.048	0.022	7
Middle yangtze river economic zone	0.207	0.087	0.042	0.030	0.082	0.108	0.022	0.065	0.020	0.103	0.072	0.027	0.031	0.027	0.078	0.061	4
Southwest economic zone	0.004	0.044	0.039	0.013	0.043	0.114	0.050	0.103	0.027	0.071	0.032	0.066	0.058	0.071	0.072	0.058	5
Northwest economic zone	0.004	0.003	0.001	0.001	0.022	0.002	0.002	0.002	0.001	0.051	0.002	0.008	0.001	0.074	0.037	0.018	8

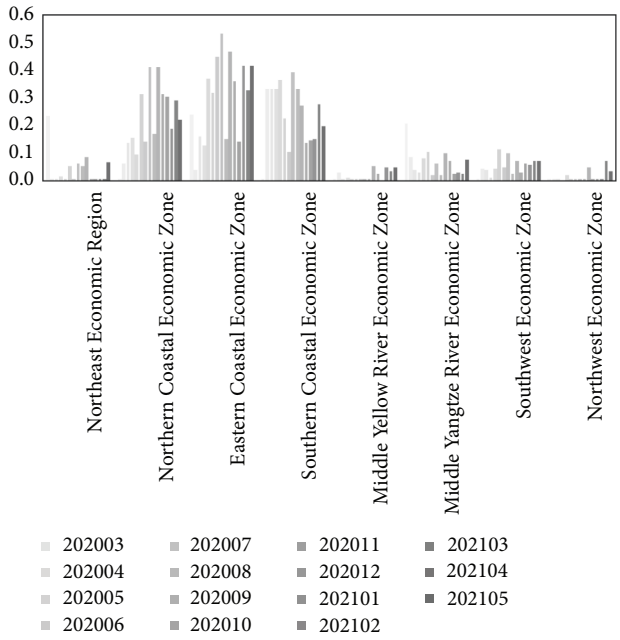


Figure 6: The static evaluation of transformation ability of emergency scientific research achievements in the 8 comprehensive economic zones.

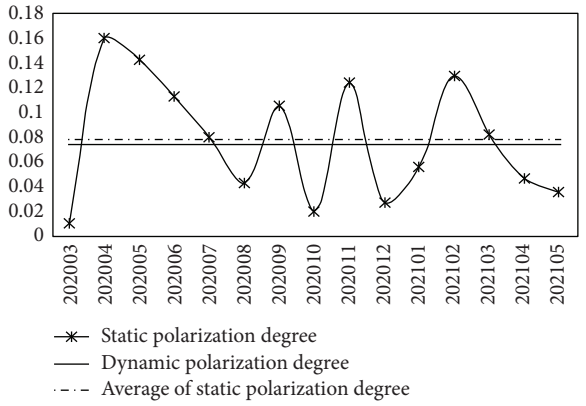


Figure 7: Polarization degree of transformation ability of emergency scientific research achievements (nationwide).

Economic Zone, and Southern Coastal Economic Zone are calculated, as shown in Table 15. Specifically: (1) Within the Northern Coastal Economic Zone, the PCD of transformation capacity of emergency scientific research results in Tianjin is the largest; the PCD of Tianjin, Hebei, and Shandong are all greater than 0, which promote the polarization effect of transformation capacity of emergency scientific research results in the Northern Coastal Economic Zone. The PCD of Beijing is less than 0. Further analysis suggests that Beijing is the core city for dispatching emergency research resources in China, and its transformation ability of emergency scientific research results has a frequent "polarization" and "diffusion" effect on other regions, so it has an inhibiting effect on the polarization effect of emergency research results transformation capacity in the Northern Coastal Economic Zone, and its PCD is less than 0, which is reasonable. (2) Within the Eastern Coastal Economic Zone, the PCD of emergency scientific research results transformation capacity in Shanghai is the largest; the PCD of Shanghai, Jiangsu, and Zhejiang are all greater than 0, which all have the positive contribution to the polarization effect of transformation capacity of emergency scientific research results in the Eastern Coastal Economic Zone. (3) Within the Southern Coastal Economic Zone, the PCD of emergency scientific research results transformation ability in Guangdong is the largest; the PCD of Fujian, Guangdong, and Hainan are all greater than 0, which all promote the polarization effect of transformation ability of emergency

Table 14: Polarization degree of transformation ability of emergency scientific research achievements (part of comprehensive economic zones).

Comprehensive economic zones	Static polarization degree																	Dynamic polarization degree
	202003	202004	202005	202006	202007	202008	202009	202010	202011	202012	202101	202103	202104	202105	Mean	Min	Max	
Northern coastal Economic zone	≤0.001	≤0.001	0.392	0.147	0.141	0.123	0.164	0.112	0.261	0.145	0.079	0.180	0.135	≤0.001	0.132	≤0.001	0.392	0.128
Eastern coastal economic zone	≤0.001	≤0.001	0.031	0.283	0.155	0.120	0.125	0.071	0.404	0.101	0.121	0.130	0.203	0.083	0.133	≤0.001	0.404	0.141
Southern coastal economic zone	≤0.001	≤0.001	≤0.001	≤0.001	0.050	≤0.001	≤0.001	0.083	≤0.001	≤0.001	0.240	0.177	≤0.001	≤0.001	0.037	≤0.001	0.240	0.043

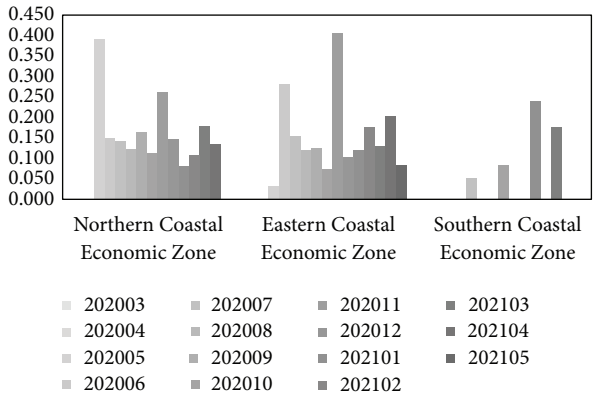


Figure 8: Static polarization degree of transformation ability of emergency scientific research achievements (part of comprehensive economic zones).

Table 15: Polarization contribution degree of transformation ability of emergency scientific research achievements (part of comprehensive economic zones).

Region	Polarization contribution degree	Rank
Northern coastal economic zone		
Beijing	−0.597	4
Tianjin	0.344	1
Hebei	0.179	2
Shandong	0.176	3
Eastern coastal economic zone		
Shanghai	0.483	1
Jiangsu	0.223	2
Zhejiang	0.065	3
Southern coastal economic zone		
Fujian	0.206	2
Guangdong	0.612	1
Hainan	0.030	3

scientific research results in the Southern Coastal Economic Zone.

5. Conclusion

The analysis of the transformation capacity of emergency scientific research results under public health emergencies is conducive for the reasonable allocation of inter-subject and inter-regional emergency scientific research resources, the improvement of the transformation efficiency of emergency results, and the enhancement of emergency scientific research strength, so as to efficiently support the prevention, control, and treatment of the event. Based on the improved minimum distance-maximum entropy combination weight, from the patent level, it carries out the static and dynamic evaluation for the transformation ability of emergency science research results of different subjects and different regions under major public health emergencies. Furthermore, the polarization effect model is improved, and the polarization effect and PCD of transformation ability of emergency scientific research achievements in different regions under major public health emergencies are analyzed. The study finds that:

(1) In the static evaluation of transformation ability of the emergency scientific research results of different scientific research subjects under the "COVID-19" epidemic, the average value of the transformation ability of the emergency results (from March 2020 to May 2021) is ranked as enterprise, institution, individual, university, and research institute. In the dynamic evaluation of each subject's transformation ability of the emergency scientific research results, enterprise has the strongest ability to transform emergency scientific research results, which exceeds the average value. In addition, enterprise ranks high in all aspects of the transformation of emergency achievements, and the advantage of patent maintenance period is more prominent; other subjects have smaller differences in the number of valid invention patents and claims, and the gap between their patent maintenance period and enterprises is larger.

(2) In the static evaluation of the transformation ability of emergency science research achievements in different regions, the top 5 regions in terms of the average value of transformation ability are Guangdong, Beijing, Jiangsu, Shanghai, and Tianjin. In the dynamic evaluation of the transformation ability of emergency scientific research results in each region, Guangdong has the strongest transformation ability of emergency scientific research results, which exceeds the average value. In addition, Guangdong ranks first in the dynamic evaluation of all indicators of emergency achievements transformation ability, with more prominent advantages in patent maintenance period and number of claims; Beijing ranks first in the dynamic evaluation of the number of valid invention patents; Jiangsu, Shanghai, and Tianjin are more consistent in the dynamic evaluation of all indicators, and Jiangsu has certain advantages in the patent maintenance period.

(3) During the response to the "COVID-19" epidemic, the top 3 comprehensive economic zones in terms of dynamic evaluation of the transformation capacity of emergency scientific research results are: Eastern Coastal Economic Zone, Northern Coastal Economic Zone, and Southern Coastal Economic Zone, with the strong transformation capacity of emergency scientific research results; the dynamic evaluation values of transformation ability of emergency scientific research achievement in Middle Yellow River Economic Zone and Northwest Economic Zone are low, and the transformation ability is weak. In the process of epidemic response, the eight comprehensive economic zones have significant differences in the transformation ability of emergency scientific research results and unbalanced spatial distribution.

(4) The polarization degree of transformation capability of emergency science research results is measured and analyzed for the whole country and eight comprehensive economic zones. Firstly, the SPDs of

transformation capability of emergency scientific research results in China fluctuates during the response to the "COVID-19" epidemic, and the polarization level is low, indicating that the resource gathering status of national emergency research is constantly adjusting with the epidemic development stage, and the resource aggregation process in the global scope has not yet shown the characteristics of order and large scale. The "polarization" and "diffusion" of transformation capability of emergency scientific research results in China are alternating, and the polarization effect is weak; the dynamic polarization level of transformation ability of emergency scientific research results in China is low, indicating that the transformation ability of emergency scientific research results at the nationwide level does not form a significant polarization effect under the consideration of the influence of time factors. Secondly, the static polarization level of transformation ability of emergency scientific research results in the Northern Coastal Economic Zone, Eastern Coastal Economic Zone, and Southern Coastal Economic Zone in each month differ significantly, and the dynamic polarization degree is low; the transformation ability of emergency scientific research results has not formed a more obvious and stable polarization effect, and under the "COVID-19" epidemic, their resources gathering of emergency scientific research results transformation has not formed an orderly and large-scale development trend. Third, within the Northern Coastal Economic Zone, Tianjin has the largest PCD of transformation capability of emergency scientific research results; the PCDs of Tianjin, Hebei, and Shandong are all greater than 0, which promote the polarization effect of transformation capacity of emergency scientific research results in the Northern Coastal Economic Zone. The PCD of Beijing is less than 0, and it has an inhibiting effect on the polarization effect of emergency research results transformation capacity in the Northern Coastal Economic Zone. Within the Eastern Coastal Economic Zone, the PCD of emergency scientific research results transformation capacity in Shanghai is the largest; the PCDs of Shanghai, Jiangsu, and Zhejiang are all greater than 0, which all have a positive contribution to the polarization effect of transformation capacity of emergency scientific research results in the Eastern Coastal Economic Zone. Within the Southern Coastal Economic Zone, Guangdong has the largest PCD of emergency scientific research results transformation ability; the PCDs of Fujian, Guangdong, and Hainan are all greater than 0, which all promote the polarization effect of transformation ability of emergency scientific research results in the Southern Coastal Economic Zone.

It should be noted that there are also limitations in this paper. In the context of COVID-19, the statistical caliber and focus of progress data related to emergency scientific research result transformation within nationwide are inconsistent, which limits the availability of evaluation index data and affects the diversification of evaluation indicators. In the follow-up study, relevant data can be further collected and organized, and relevant achievements information of emergency scientific research projects can be supplemented, so as to improve the evaluation indexes of the transformation capability of emergency scientific research achievements under public health emergencies. In addition, considering the strong innovation strength and R&D potential possessed by enterprises, we can focus on the transformation of enterprise's emergency scientific research results under major public health emergencies and further analyze the spatial-temporal evolution characteristics of emergency scientific research results transformation ability, the transformation efficiency of emergency science research results, and emergency scientific research resources allocation.

6. Innovations

The innovations of this work are reflected in the following aspects. (1) The innovation of the evaluation method. Considering the lack of data directly related to major public emergencies, the statistical cycles and calibers of different regions are different, and the data vary greatly in different regions and different periods; neither the singly subjective weighting method nor objective weighting method is suitable for the evaluation of the transformation ability of emergency scientific research achievements. Therefore, an improved minimum distance-maximum entropy combination weight method is proposed to evaluate the transformation capacity of emergency scientific research results under public health emergencies, so as to improve the rationality, accuracy, and credibility of evaluation results. (2) The innovation of the evaluation index system. The major public emergencies have a huge impact on the transformation of emergency scientific research results and bring complex consequences. It is necessary to combine the characteristics of scientific research results transformation in the context of public emergencies, and comprehensively select evaluation indicators to construct a comprehensive evaluation index system of transformation ability for emergency scientific research achievements under the background of major public emergencies. (3) This paper uses progress data (monthly) to realize the timely evaluation of transformation capability for emergency scientific research achievements. Different from most existing literature that uses annual data, the use of monthly progress data for analysis is more conducive to obtaining the immediate characteristics of achievements transformation capability under the impact of emergencies, and provides a reliable basis for efficient, accurate, and timely decision-making. (4) The ER index and the PCD are further improved. This will enrich the research methods of the spatial effect of the transformation capacity of emergency scientific research achievements and provide a reliable analysis

tool for the subsequent study of the transformation capacity of emergency scientific research achievements.

Data Availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors declare that they have no conflict of interest in this manuscript.

Acknowledgments

This study was funded by The National Social Science Fund of China (No. 19AGL003) and The National Social Science Fund of China (No. 20&ZD127).

References

[1] S. Gautam, "The influence of COVID-19 on air quality in India: a boon or inutile," *Bulletin of Environmental Contamination and Toxicology*, vol. 104, no. 6, pp. 724–726, 2020.

[2] Y. Liu, J. M. Lee, and C. Lee, "The challenges and opportunities of a global health crisis: the management and business implications of COVID-19 from an Asian perspective," *Asian Business & Management*, vol. 19, no. 3, pp. 277–297, 2020.

[3] H. Y. Sun, W. Xu, Y. Y. Cai, and G. Cai, "An intelligent mechanism for covid-19 emergency resource coordination and follow-up response," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 2005188, 1 page, 2022.

[4] C. Gibson, "Guidepost from 'social distancing'to 'care in connecting': an emerging organizational research agenda for turbulent times," *Academy of Management Discoveries*, vol. 6, no. 2, pp. 200–212, 2020.

[5] E. León-Sandoval, M. Zareei, L. I. Barbosa-Santillán, L. E. Falcón Morales, A. Pareja Lora, and G. Ochoa Ruiz, "Monitoring the emotional response to the covid-19 pandemic using sentiment analysis: a case study in Mexico," *Computational Intelligence and Neuroscience*, vol. 2022, Article ID 4914665, 11 pages, 2022.

[6] J. Hui and Q. M. Tan, "The operation mechanism of sci-tech R&D emergency system in major public health emergencies," *Science & Technology Progress and Policy*, vol. 37, no. 9, pp. 11–20, 2020.

[7] Y. N. Zhu and Y. L. Yu, "Research on the influencing factors and countermeasures of the transformation of scientific and technological achievements : a case study of local universities in Yangtze River Delta region," pp. 92–96, Chinese University Science & Technology, China, 2021.

[8] W. J. Zhang and P. Zhang, "Developing the transformation of scientific and technological achievements in colleges and universities to boost the development of low-carbon economy," *International Journal of Low Carbon Technologies*, vol. 16, no. 2, pp. 305–316, 2021.

[9] X. Zhang and R. M. Ma, "Research on the discovery of core patents based on improved PageRank algorithm," *Library and Information Service*, vol. 62, no. 10, pp. 106–115, 2018.

[10] H. Y. Song and J. H. Zhang, Y. Zhang, Patent protection: does it promote or inhibit the patented technological knowledge diffusion?" *Scientometrics*, vol. 127, no. 5, pp. 2351–2379, 2022.

[11] D. C. Fan, L. Fang, and Z. L. Song, "Dynamic comprehensive evaluation of China's industrial structure transformation capability based on the combination weighting method of index variation degree and index correlation degree," *Operations Research and Management Science*, vol. 29, no. 6, pp. 166–178, 2020.

[12] C. Xiang, W. Xing, Z. Hubiao, X. Yuheng, C. You, and W. Xiaotian, "Interval TOPSIS with a novel interval number comprehensive weight for threat evaluation on uncertain information," *Journal of Intelligent and Fuzzy Systems*, vol. 42, no. 4, pp. 4241–4257, 2022.

[13] C. Zhang, Y. Y. Zu, W. He, H. Zu, Y. Zhang, and W. He, "An ensemble learning method based on an evidential reasoning rule considering combination weighting," *Computational Intelligence and Neuroscience*, vol. 202217 pages, Article ID 1156748, 2022.

[14] W. M. Wang, H. Y. Xu, and F. M. Zhang, "Research on comprehensive evaluation of academic journal based on CRITIC-G1 and Bonferroni operator," *Information Studies: Theory & Application*, vol. 43, no. 5, pp. 104–109, 2020.

[15] Z. R. Li, X. J. Ma, and Z. L. Peng, "Research on the combination evaluation method based on maximizing deviations," *Chinese Journal of Management Science*, vol. 21, no. 1, pp. 174–179, 2013.

[16] F. P. Yu, L. F. Liu, and B. Meng, "Research on corporate social responsibility evaluation of transportation industry based on set pair analysis," *Management Review*, vol. 32, no. 1, pp. 219–234, 2020.

[17] Y. Zhou, H. Wang, and H. Liu, "Generalized function projective synchronization of incommensurate fractional-order chaotic systems with inputs saturation," *International Journal of Fuzzy Systems*, vol. 21, no. 3, pp. 823–836, 2019.

[18] G. T. Chi, F. Qi, and N. Zhang, "The city's ecosystem evaluation model based on optimal combination weights and its application," *Operations Research and Management Science*, vol. 21, no. 2, pp. 183–191, 2012.

[19] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 4, pp. 623–656, 1948.

[20] C. E. Shannon, "Communication theory of secrecy systems," *Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, 1949.

[21] Y. Zhou, H. Liu, J. Li, and S. Li, "Composite learning fuzzy synchronization for incommensurate fractional-order chaotic systems with time-varying delays," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 12, pp. 1739–1758, 2019.

[22] F. Perroux, "Economic space: theory and applications," *Quarterly Journal of Economics*, vol. 64, no. 1, pp. 89–104, 1950.

[23] M. Akinci, "Inequality and economic growth: trickle-down effect revisited," *Development Policy Review*, vol. 36, no. 3, pp. O1–O24, 2017.

[24] M. Zhou and Y. X. Sheng, "The effect of economic polarization on technological innovation in the heterogeneous space: base on the empirical research of two-level hierarchical line model," *Nankai Economic Studies*, vol. 3, pp. 65–78, 2012.

[25] J. Chen, Y. J. Zhang, and J. Wang, "A Comparative study on the development of AI industry between China and U.S.A. based on patent analysis," *Journal of Intelligence*, vol. 38, no. 1, pp. 41–47, 2019.

[26] W. Long and C. X. Long, "Does the Chinese version of Bayh-Dole Act promote university innovation?" *China Economic Quarterly International*, vol. 1, no. 3, pp. 244–257, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 6398039, 12 pages
https://doi.org/10.1155/2022/6398039

Research Article

Anticollision Decision and Control of UAV Swarm Based on Intelligent Cognitive Game

Huan Zhou, Yintong Li, and Tong Han

Aviation Engineering School, Air Force Engineering University, Xi'an, China

Correspondence should be addressed to Yintong Li; yintongli0007@163.com

Received 17 May 2022; Accepted 13 July 2022; Published 11 August 2022

Academic Editor: Wei Xiang

UAV swarm anticollision system is very important to improve the flight safety of the whole swarm formation, while the existing system design methods are still insufficient in realizing autonomous and cooperative anticollision. Based on the cognitive game theory, an intelligent decision-making and control method for UAV swarm anticollision is designed. Firstly, by using the idea of swarm intelligence, basic flight behaviors of UAV swarm are defined as five basic flight rules, such as cohesion, following, self-guidance, dispersion, and alliance. Further, the cognitive security domain of UAV swarm is constructed by setting the overall anticollision rules of the swarm and the anticollision rules of individual members. On this basis, the anticollision problem of UAV swarm is transformed into a game problem involving two parties, and the solution method of decision and control strategy set is proposed. Finally, the stability of anticollision decision and control method is proved through eigenvalue theory. The simulation results show that the method proposed in this paper can effectively realize the autonomous cooperative anticollision of UAV swarm and also has good algorithm real-time solution ability while ensuring flight safety.

1. Introduction

The operational mission requirements such as cooperative reconnaissance, cooperative tracking, and cooperative strike faced by UAV autonomous cooperation determine that its operational use mode is multi-aircraft swarm system [1, 2]. The application of swarm enables UAV to cover a wide area in less time in reconnaissance, search, and rescue tasks, which greatly improves the use efficiency of UAV. However, the increase in the number and density of space UAVs has also brought great challenges to flight safety at the same time [3, 4]. Autonomous flight and evasion control of swarm system has become an urgent problem to be solved.

For a swarm system, there are currently two main control methods: leader follower [5–7] and behavior control [8–10]. The idea of the leader follower method is to introduce a leader and use the distributed control method to control other followers in the swarm, so that the state (such as speed) of all followers gradually follows the leader and finally achieves consistency. In reference [7], considering system noise and time delay, the coordination problem of second-order definite topology multi-agent system with leader is studied by using control theory. Behavior control mainly draws lessons from the idea of swarm intelligence, designs the motion behavior of swarm system based on evolutionary mechanism, and has adaptive ability through certain evolution and development. Among them, literature [8] established a two-dimensional in-plane swarming algorithm based on the UAV model of constant speed and inclined turning. The algorithm includes two basic rules: alliance and cohesion. A great contribution of this paper is to prove that the weight of the rules will determine the flight behavior of the swarm.

At present, the commonly used swarm system avoidance control methods mainly include geometric vectors [11, 12], artificial potentials [13–15], and model predictive control (MPC) [16–18]. Geometric vector method often has no speed limit on the internal members of the swarm system, which is not allowed by UAV, so it is mainly used in the control of ground robot. The artificial potential field method regards the motion of UAV as the result of the interaction of attraction and repulsion, but this method is easy to fall into

local minimum. One of the main advantages of model predictive control is that it can effectively solve the problems of constraints or high dimension of the system, but the amount of system calculation increases significantly with the increase of the number of swarm members, which is difficult to meet the real-time requirements [19].

In order to overcome the shortcomings of the above methods, inspired by the idea of swarm intelligence in reference [8], this paper studies the UAV swarm flight and evasion control method based on rules. Firstly, the basic flight rules of UAV members in the swarm are defined to control the normal level flight of the swarm. On this basis, the evasion action is regarded as another flight rule of the swarm. By setting the corresponding rule weight coefficient, two mechanisms of the whole and members are established for swarm evasion guidance and control. Simulation results show that the proposed method can effectively realize swarm flight and evasion, and has good real-time performance.

2. Basic Flight Behavior Rules of Swarm

In order to control the normal flight and formation reconstruction of the swarm system, using the idea of swarm intelligence for reference, the basic flight rules of UAV members in the swarm system are abstracted as cohesion, following, self-guidance, dispersion, and alliance.

2.1. Swarm Intelligence. Swarm intelligence is a computing technology based on the behavior law of biological groups. There are two main algorithms in this theoretical research field: ant colony algorithm [20] (ACA) and particle swarm optimization [21] (PSO). In the field of computer science and application research, ant colony algorithm is a probability theory method to solve mathematical problems. It uses graph theory to find optimization and optimal path. Dr. Kennedy and Dr. Eberhart first proposed the particle swarm optimization algorithm in 1995, inspired by the social behaviors such as predation of birds and fish. PSO algorithm is similar to genetic algorithm in evolutionary computation, but different from genetic algorithm, it has no evolutionary mechanism such as crossover and mutation.

At present, because ant colony algorithm and particle swarm optimization algorithm have many advantages, they are widely used in research and application fields. At the same time, swarm intelligence method [22] is also gradually rising. Based on the traditional swarm intelligence algorithm, this method aims to realize the distributed intelligent cooperative control of multirobots, multiagents, and multiaircraft platforms. Swarm intelligence algorithm regards each agent/member in the swarm system as an individual in the biological system. Individuals with the same or similar characteristics form a new swarm. The interaction between swarm individuals or swarms adopts the influence mechanism of traditional swarm intelligence algorithm. In this way, the swarm intelligent algorithm will have the characteristics of simple algorithm, easy to meet real time, high degree of intelligence, and good robustness.

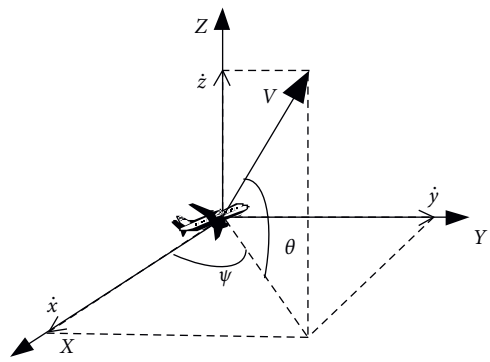


Figure 1: Geometric parameters of a single UAV member.

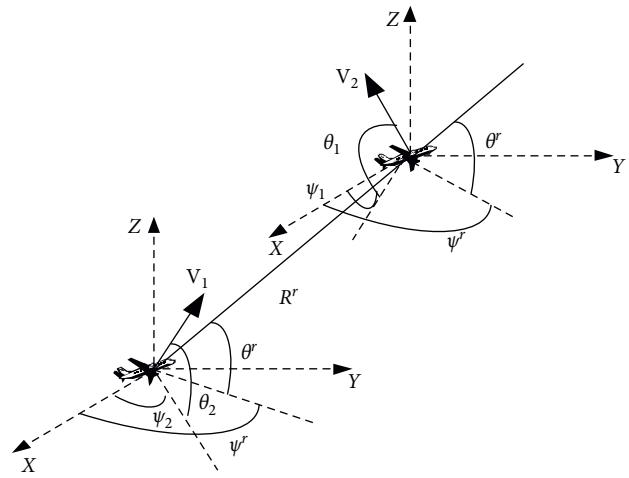


Figure 2: Relative geometric parameters of two UAV members.

2.2. Basic Flight Rules for Internal Members of the Swarm System. The geometric parameters of single and two UAV members in the swarm system are shown in Figures 1 and 2 respectively.

According to the six degree-of-freedom motion equations of UAV [23], the partial velocities of each member in the swarm system in the three axes in the geographic coordinate system are $\dot{x}, \dot{y}, \dot{z}$, respectively:

$$\dot{x} = V \cos\theta \cos\psi, \tag{1}$$

$$\dot{y} = V \cos\theta \sin\psi, \tag{2}$$

$$\dot{z} = V \sin\theta, \tag{3}$$

where x, y, z are the three-axis coordinates of the UAV, ψ, θ are the track deflection angle and track inclination angle of the UAV, V is the flight speed scalar of the UAV, and the two angular speeds $\dot{\psi}, \dot{\theta}$ of the UAV are expressed by the following formula:

$$\dot{\psi} = \frac{\kappa_\psi}{V \cos\theta}, \tag{4}$$

$$\dot{\theta} = \frac{\kappa_\theta}{V}, \tag{5}$$

where κ_ψ and κ_θ are the acceleration term coefficients of angular velocity.

In Figure 2, ψ^r, θ^r, R^r are the relative track deflection angle, relative track inclination angle, and relative distance on the sight distance of two UAV members, respectively. Make the angular velocity $\dot{\psi}, \dot{\theta}$ of UAVs in the swarm proportional to the deviation between the expected attitude angle and the current attitude angle:

$$\kappa_\psi = k_1 \Delta\psi_{\mathrm{d}}, \tag{6}$$

$$\kappa_\theta = k_2 \Delta\theta_{\mathrm{d}}, \tag{7}$$

$$\Delta\psi_{\mathrm{d}} = \psi_{\mathrm{d}} - \psi, \tag{8}$$

$$\Delta\theta_{\mathrm{d}} = \theta_{\mathrm{d}} - \theta, \tag{9}$$

where k_1, k_2 are proportional constants, $\psi_{\mathrm{d}}, \theta_{\mathrm{d}}$ are the expected track deflection angle and the expected track inclination angle, respectively, ψ, θ are the current track deflection angle and the current track inclination angle, respectively, and $\Delta\psi_{\mathrm{d}}, \Delta\theta_{\mathrm{d}}$ are the deviation between the expected track deflection angle and the current track deflection angle and the deviation between the expected track inclination angle and the current track inclination angle, respectively. Among them, the desired attitude angle of UAV is calculated by the flight rules of UAV swarm. When the number of effective flight rules of UAV is greater than 1, the expected attitude angle deviation can be obtained from the following formula:

$$\Delta\psi_{\mathrm{d}} = \tau_1\left(\psi_{\mathrm{d}_1} - \psi\right) + \tau_2\left(\psi_{\mathrm{d}_2} - \psi\right) + \cdots + \tau_n\left(\psi_{\mathrm{d}_n} - \psi\right), \tag{10}$$

$$\Delta\theta_{\mathrm{d}} = \tau_1\left(\theta_{\mathrm{d}_1} - \theta\right) + \tau_2\left(\theta_{\mathrm{d}_2} - \theta\right) + \cdots + \tau_n\left(\theta_{\mathrm{d}_n} - \theta\right), \tag{11}$$

where $\psi_{\mathrm{d}_1}, \psi_{\mathrm{d}_2}, \ldots, \psi_{\mathrm{d}_n}$ and $\theta_{\mathrm{d}_1}, \theta_{\mathrm{d}_2}, \ldots, \theta_{\mathrm{d}_n}$ are the expected track deflection angle and expected track inclination under the action of the first to n-th flight rules, respectively, and $\tau_1, \tau_2, \ldots, \tau_n$ is the weight corresponding to the first to n-th flight rules.

For the i-th UAV member in the swarm system, the basic flight rules are abstracted as the following five.

Cohesion. The cohesion feature can make the internal members of the swarm system close to each other to maintain the swarm formation. Each UAV member in the swarm approaches the UAV within its detection distance, that is, the expected velocity vector of the UAV points to the centroids of several UAV members within its detection distance ρ. Suppose that the i-th UAV member in the swarm can detect n_i UAV members; $(X_{i_M}, Y_{i_M}, Z_{i_M})$ is the average centroid coordinate of n_i UAV members, and (X_i, Y_i, Z_i) is the centroid coordinate of the i-th UAV member; then under this flight rule, the expected track deflection angle $\psi_{\mathrm{d}i_M}$ and expected track inclination angle $\theta_{\mathrm{d}i_M}$ of the i-th UAV member are, respectively,

$$\psi_{\mathrm{d}i_M} = \arctan\left(\frac{Y_{i_M} - Y_i}{X_{i_M} - X_i}\right), \tag{12}$$

$$\theta_{\mathrm{d}i_M} = \arctan\left(\frac{Z_{i_M} - Z_i}{\sqrt{\left(X_{i_M} - X_i\right)^2 + \left(Y_{i_M} - Y_i\right)^2}}\right). \tag{13}$$

Follow. The follow feature enables each member in the swarm system to follow the other two members. One member is closest to the UAV member, and the other member is randomly selected within the system. Let $(X_{i_L}, Y_{i_L}, Z_{i_L})$ be the centroid coordinate of the UAV nearest to the UAV member and $(X_{i_R}, Y_{i_R}, Z_{i_R})$ be the centroid coordinate of the randomly selected UAV; then under this flight rule, the expected track deflection angle $\psi_{\mathrm{d}i_F}$ and expected track inclination angle $\theta_{\mathrm{d}i_F}$ of the i-th UAV member can be obtained by the following formula:

$$\psi_{\mathrm{d}i_L} = \arctan\left(\frac{Y_{i_L} - Y_i}{X_{i_L} - X_i}\right), \tag{14}$$

$$\theta_{\mathrm{d}i_L} = \arctan\left(\frac{Z_{i_L} - Z_i}{\sqrt{\left(X_{i_L} - X_i\right)^2 + \left(Y_{i_L} - Y_i\right)^2}}\right), \tag{15}$$

$$\psi_{di_R} = \arctan\left(\frac{Y_{i_R} - Y_i}{X_{i_R} - X_i}\right), \tag{16}$$

$$\theta_{di_R} = \arctan\left(\frac{Z_{i_R} - Z_i}{\sqrt{\left(X_{i_R} - X_i\right)^2 + \left(Y_{i_R} - Y_i\right)^2}}\right), \tag{17}$$

$$\psi_{\mathrm{d}i_F} = \frac{\psi_{\mathrm{d}i_L} + \psi_{\mathrm{d}i_R}}{2}, \tag{18}$$

$$\theta_{\mathrm{d}i_F} = \frac{\theta_{\mathrm{d}i_L} + \theta_{\mathrm{d}i_R}}{2}. \tag{19}$$

Homing. The homing feature enables swarm system members to track specific signals to fly to a specified area. Let $(X_{i_I}, Y_{i_I}, Z_{i_I})$ be the coordinate where the tracking signal is located; then under this flight rule, the expected track deflection angle $\psi_{\mathrm{d}i_I}$ and expected track inclination angle $\theta_{\mathrm{d}i_I}$ of the i-th UAV member are, respectively,

$$\psi_{\mathrm{d}i_I} = \arctan\left(\frac{Y_{i_I} - Y_i}{X_{i_I} - X_i}\right), \tag{20}$$

$$\theta_{\mathrm{d}i_I} = \arctan\left(\frac{Z_{i_I} - Z_i}{\sqrt{\left(X_{i_I} - X_i\right)^2 + \left(Y_{i_I} - Y_i\right)^2}}\right), \tag{21}$$

Dispersion. The dispersion feature can keep enough safe distance between members in the swarm, that is, ensure that the flight interval between any two members does not exceed

$s_{\min}$. Therefore, each UAV member flies in the opposite direction of other UAVs in the swarm to prevent UAVs from getting too close. Let $(X_{i_S}, Y_{i_S}, Z_{i_S})$ be the average centroid coordinate of the i-th UAV whose distance from the UAV member is less than $s_{\min}$; then under this flight rule, the expected track deflection angle ψ_{di_S} and expected track inclination angle θ_{di_S} of the i-th UAV member are, respectively,

$$\psi_{di_S} = -\arctan\left(\frac{Y_{i_S} - Y_i}{X_{i_S} - X_i}\right), \tag{22}$$

$$\theta_{di_S} = \arctan\left(\frac{Z_{i_S} - Z_i}{\sqrt{\left(X_{i_S} - X_i\right)^2 + \left(Y_{i_S} - Y_i\right)^2}}\right). \tag{23}$$

Alliance. The alliance feature can keep the internal members of the swarm system in a certain order and ensure the flight of the swarm system as a whole. Each UAV member flies in the direction of the average velocity vector of the UAV within its detection range. Under this flight rule, the expected track deflection angle ψ_{di_A} and expected track inclination angle θ_{di_A} of the i-th UAV member are, respectively,

$$\psi_{di_A} = \left(\frac{1}{n_i}\right)\sum_{j=1}^{n_i}\left(\psi_{dj_A}\right), \tag{24}$$

$$\theta_{di_A} = \left(\frac{1}{n_i}\right)\sum_{j=1}^{n_i}\left(\theta_{dj_A}\right). \tag{25}$$

3. Cognitive Security Domain for Swarm Anticollision

The swarm collaboration adopts a distributed mechanism. Each member of the system uses airborne sensors to detect the unknown environment. The members communicate based on the global fully connected topology to realize information sharing. When sensing the threat of obstacles, they can evade in real time and independently. It is assumed that accurate environmental state information has been obtained after fusion filtering of sensor data. Based on the basic flight rules of the internal members of the swarm system, the swarm evasion action is also regarded as the flight rules of the internal members of the swarm. Similarly, under the evasion flight rules, the expected track deflection angle and expected track inclination angle of the i-th UAV member are ψ_{di_E} and θ_{di_E}, respectively. Therefore, the swarm system avoiding collaborative control can be described as follows.

By calculating such a desired track deflection angle ψ_{di} and desired track inclination angle θ_{di} (for the i-th UAV member in a swarm), $\Delta\psi_{di}$ and $\Delta\theta_{di}$ are obtained, and the guidance command is transmitted to the autopilot of the corresponding UAV. Under the action of the flight control system, it can ensure that no member in the swarm system will collide with other members in the swarm or other threats outside the swarm, so as to realize safe flight.

The above calculation formulas of $\Delta\psi_{di}$ and $\Delta\theta_{di}$ are as follows:

$$\begin{aligned}\Delta\psi_{di} = {} & \tau_{i_M}\left(\psi_{di_M} - \psi\right) + \tau_{i_F}\left(\psi_{di_F} - \psi\right) + \tau_{i_I}\left(\psi_{di_I} - \psi\right) \\ & + \tau_{i_S}\left(\psi_{di_S} - \psi\right) + \tau_{i_A}\left(\psi_{di_A} - \psi\right) + \tau_{i_E}\left(\psi_{di_E} - \psi\right),\end{aligned} \tag{26}$$

$$\begin{aligned}\Delta\theta_{di} = {} & \tau_{i_M}\left(\theta_{di_M} - \theta\right) + \tau_{i_F}\left(\theta_{di_F} - \theta\right) + \tau_{i_I}\left(\theta_{di_I} - \theta\right) \\ & + \tau_{i_S}\left(\theta_{di_S} - \theta\right) + \tau_{i_A}\left(\theta_{di_A} - \theta\right) + \tau_{i_E}\left(\theta_{di_E} - \theta\right).\end{aligned} \tag{27}$$

By setting different weights of flight rules, swarm evasion cooperative control is divided into two decision-making mechanisms: overall evasion mechanism and member evasion mechanism. In the first mechanism, UAV swarm as a whole can avoid collision threat. In the second mechanism, the UAV swarm can avoid the collision threat through the flight behavior of each member.

3.1. Overall Avoidance Mechanism. In the overall evasion mechanism, the UAV swarm as a whole, in order to ensure that the internal members of the swarm do not separate during the evasion process, the weight of the cohesive flight rules of the UAV swarm is large. Under the action of avoiding flight rules, the expected track deflection angle of the i-th UAV member in the swarm is

$$\psi_{di_E} = f_1\left(\frac{\pi}{2}\right) + \psi_i, \tag{28}$$

where f_1 is the transformation function of track deflection direction in the horizontal plane set to complete evasion. Let $(X_{i_T}, Y_{i_T}, Z_{i_T})$ be the average centroid coordinates of all collision threats outside the swarm at the detection distance of the i-th UAV member, then $f_1(x)$ is expressed as

$$f_1(x) = \begin{cases} -x, \dfrac{\pi}{4} < \psi_i \le \dfrac{3\pi}{4} \text{ and } X_i \ge X_{i_T} \\ x, \dfrac{\pi}{4} \le \psi_i < \dfrac{3\pi}{4} \text{ and } X_i < X_{i_T} \\ x, \dfrac{5\pi}{4} < \psi_i \le \dfrac{7\pi}{4} \text{ and } X_i \ge X_{i_T} \\ -x, \dfrac{5\pi}{4} \le \psi_i < \dfrac{7\pi}{4} \text{ and } X_i < X_{i_T} \\ -x, \dfrac{3\pi}{4} < \psi_i \le \dfrac{5\pi}{4} \text{ and } Y_i \ge Y_{i_T} \\ x, \dfrac{3\pi}{4} \le \psi_i < \dfrac{5\pi}{4} \text{ and } Y_i < Y_{i_T} \\ x, \left(\dfrac{7\pi}{4} < \psi_i \le 2\pi \text{ or } 0 < \psi_i \le \dfrac{\pi}{4}\right) \text{and } Y_i \ge Y_{i_T} \\ -x, \left(\dfrac{7\pi}{4} \le \psi_i < 2\pi \text{ or } 0 \le \psi_i < \dfrac{\pi}{4}\right) \text{and } Y_i < Y_{i_T} \end{cases}. \tag{29}$$

According to equation (29), the overall avoidance mechanism can be explained as follows: when the distance between the swarm and the collision threat gradually increases, the swarm flies along the pre-planned track; when the swarm and the collision threat gradually approach and the distance between them is less than a certain value, according to equation (28), the swarm system performs avoidance cooperative control as a whole. The transformation of track deflection angle of each UAV member depends on the position coordinate relationship between the UAV and the collision threat.

Similarly, in the overall avoidance mechanism, the expected track inclination of the i-th UAV member in the swarm is

$$\theta_{AVi} = f_2\left(\frac{\pi}{2}\right) + \theta_i, \tag{30}$$

where f_2 is the transformation function of track inclination direction in the vertical plane set to complete evasion and $f_2(x)$ is expressed as

$$f_2(x) = \begin{cases} x, & Z_i \geq Z_{i_T}, \\ -x, & Z_i < Z_{i_T}. \end{cases} \tag{31}$$

For the i-th UAV member, the weight of circumvention rule is defined as follows:

$$\tau_{i_E} = \begin{cases} 1, & \left\|\mathbf{P}_i - \mathbf{P}_{\rho i}\right\| \leq \rho_c, \\ 0, & \left\|\mathbf{P}_i - \mathbf{P}_{\rho i}\right\| > \rho_c, \end{cases} \tag{32}$$

where $\mathbf{P}_i = (X_i, Y_i, Z_i)$, $\mathbf{P}_{\rho i} = (X_{i_T}, Y_{i_T}, Z_{i_T})$, ρ_c is defined as the evasion distance of UAV members; that is, the evasion control is carried out only when the distance between two centroids is less than ρ_c.

3.2. Member Avoidance Mechanism. In the member evasion mechanism, the control method is similar to the overall evasion mechanism, except for the weight of flight rules. At the same time, the extended distance ρ_e is defined based on the available detection distance ρ and evasion distance ρ_c of the sensor. The three distances defined in this paper have the following relationship: $\rho_c < \rho_e < \rho$.

Set the minimum interval $s_{\min}$ between any two UAV members in the same swarm to

$$s'_{\min} = \begin{cases} s_1, & \left\|\mathbf{P}_i - \mathbf{P}_{\rho i}\right\| \leq \rho_e, \\ s_2, & \left\|\mathbf{P}_i - \mathbf{P}_{\rho i}\right\| > \rho_e, \end{cases} \tag{33}$$

where $s_1 > s_2$.

Then, the member avoidance mechanism can be explained as follows: when the collision threat is within the extended distance ρ_e, of any UAV member in the swarm system, the value of $s_{\min}$ increases from s_2 to s_1, so that the swarm system can allow the collision threat to pass through the swarm during the avoidance process. When the collision threat enters the swarm and is within the avoidance distance ρ_c of its internal members, the avoidance decision will be made, and the rule weight, expected track deflection angle, and expected track inclination angle can be obtained from equations (28)–(32). Under this mechanism, the cohesion weight τ_{i_M} is reduced, so that the distance expansion between members in the swarm system can be better realized.

3.3. Cognitive Security Domain. The autonomous anticollision decision and control of UAV swarm include anticollision and separation guarantee, that is, to keep any member in the swarm outside the area surrounding each member adjacent to it. For the sake of safety, assuming that only UAV maneuvers to avoid collision, the above area is defined as the safety area of UAV, and the safety goal is to ensure that no other UAV penetrates the area. The exact mathematical expression of cognitive security domain is given below.

Let $\mathbf{z}(t) \in \mathbb{R}^{\mathrm{d}}$ represent the state vector of two conflicting UAVs at any time $t > 0$. It is assumed that its variation law can be described by a differential equation:

$$\begin{cases} \dot{z}(s) = f(\mathbf{z}(s), s, \mathbf{u}(s), \mathbf{r}(s)), \forall s \in [0, t], \\ \mathbf{z}(0) = \mathbf{z}_0, \end{cases} \tag{34}$$

where $\mathbf{u}(\cdot)$ represents the decision and control quantity of UAV, and $\mathbf{r}(\cdot)$ represents some uncertainties considered in the model.

It can be seen that the change law depends on two different kinds of decision and control vectors $\mathbf{u}(\cdot)$ and $\mathbf{r}(\cdot)$. Let $\mathbf{M}$ and $\mathbf{N}$ be two nonempty compact subsets of $\mathbb{R}^m$ and $\mathbb{R}^P$, respectively. Suppose that $\mathbf{A}$: = $\{\mathbf{u}$: $(0, t) \longrightarrow \mathbf{M}\}$, $\mathbf{B}$: = $\{\mathbf{u}$: $(0, t) \longrightarrow \mathbf{N}\}$, for each $\mathbf{z}(0) \in \mathbb{R}^{\mathrm{d}}$ and $(\mathbf{u}, \mathbf{r}) \in \mathbf{A} \times \mathbf{B}$, $\mathbf{z} = \mathbf{z}_0^{\mathbf{u},\mathbf{r}}$ represents the relevant trajectory, which is defined as the system composed of equations (1)–(9).

Let $\mathbf{O} \subset \mathbb{R}^{\mathrm{d}}$ be an open set, which is called "collision region." All relative positions in the set are equivalent to collisions. The precise definition of the set will depend on the selection of dynamic state space. Here, the collision region $\mathbf{O}$ is regarded as an obstacle, and its complement $\mathbf{K}$= $\mathbb{R}^{\mathrm{d}}/\mathbf{O}$ represents the set of state constraints.

Different safety zones are defined as follows:

(1) Set $\mathbf{W}_1$. It is defined as a subset of the initial position, so that for any control strategy, UAV cannot guarantee to avoid collision.

(2) Set $\mathbf{W}_2(t_f)$. It is defined as the set of all initial states in $\mathbb{R}^{\mathrm{d}}/\mathbf{W}_1$, so that if there is no maneuver, there will be a risk that the system reaches region $\mathbf{W}_1$ before time t_f.

4. Swarm Anticollision Game Decision and Control

4.1. Game Theory Model of Swarm Anticollision Problem. In order to analyze the security of systems with both control and disturbance, the worst-case method in antinatural game can be used. Disturbances are regarded as opponents of control and undermine security. For autonomous anticollision control, a game problem involving two parties is considered. One party is any member in the swarm, and the

other party is each member adjacent to it. Here, the dynamic model of the relative position between each member adjacent to it and any member is

$$\dot{\mathbf{X}} = f'(\mathbf{X}(t), \mathbf{u}(t), \mathbf{r}(t)), \tag{35}$$

where $\mathbf{X}$ represents the state variable, $\mathbf{u}$ represents the control quantity of UAV, and $\mathbf{r}$ represents the uncertainty of the system, which plays the role of controlling the second player.

Meanwhile, the above formula meets the following conditions:

(1) For each $(\mathbf{X}, \dot{\mathbf{X}}, \mathbf{u}, \mathbf{r}) \in \mathbb{R}^{\mathrm{d}} \times \mathbb{R}^{\mathrm{d}} \times \mathrm{M} \times \mathbf{N}$, there is $L_f > 0$ such that $|f'(\mathbf{X}, \mathbf{u}, \mathbf{r}) - f'(\dot{\mathbf{X}}, \mathbf{u}, \mathbf{r})| \le L_f |\mathbf{X} - \dot{\mathbf{X}}|$, where L_f is the Lipschitz constant

(2) For each $\mathbf{X} \in \mathbb{R}^{\mathrm{d}}$ and $\mathbf{r} \in \mathbf{M}$, $f'(\mathbf{X}, \mathbf{M}, \mathbf{r})$ is a convex set of $\mathbb{R}^{\mathrm{d}}$

4.2. Solution of Decision and Control Strategy Set. As can be seen from Sections 3 and 3.1, the anticollision problem is described as a problem staying in a given closed set $\mathbf{K} = \mathbb{R}^{\mathrm{d}}/\mathbf{O}$, and the autonomous anticollision problem is transformed into a game framework involving two parties. At this time, the set $\mathbf{W}_1$ can be described by the worst case of the game, which is that the first party player (UAV) wants to avoid the collision area $\mathbf{O}$, or equivalent to keeping the system in the safe area $\mathbf{K}$ (through its own input $\mathbf{u}$). The unexpected strategy set of player 1 (UAV) is defined as

$$\begin{aligned}\Gamma: \{\mathbf{u}: \mathbf{B} \longrightarrow \mathbf{A}, \forall s \in [0, \infty], (\mathbf{r}(\xi) = \tilde{\mathbf{r}}(\xi), \forall \xi \in \\ [0, s]) \Rightarrow (\mathbf{u}[\mathbf{r}](\xi) = \mathbf{u}[\tilde{\mathbf{r}}](\xi), \forall \xi \in [0, s])\}.\end{aligned} \tag{36}$$

Then, the selection of UAV control law is limited to the set Γ, and the victory domain of player 1 (UAV) is defined.

The victory domain of player 1: the set $\mathbf{V}_1(\mathbf{K})$ of all initial positions z_0. There is an unexpected strategy $F \in \Gamma$, which can ensure that all corresponding trajectories $z_0^{\mathbf{u}[\mathbf{r}],\mathbf{r}}(t)$ avoid collision set $\mathbf{O}$ for all $\in$ and the allowable control $\mathbf{r} \in \mathbf{B}$. And $\mathbf{V}_1(\mathbf{K})$ is

$$\begin{aligned}\mathbf{V}_1(\mathbf{K}) = \{z_0 \in \mathbf{K} | \exists F \in \Gamma, \forall \mathbf{r}(\cdot) \in \mathbf{B}, \\ \forall t \ge 0, z_0^{\mathbf{u}[\mathbf{r}],\mathbf{r}}(t) \in \mathbf{K}\}.\end{aligned} \tag{37}$$

According to the definition, the two decision and control policy sets are

$$\mathbf{W}_1 = \frac{\mathbb{R}^3}{\mathbf{V}_1(\mathbf{K})}, \tag{38}$$

$$\begin{aligned}\mathbf{W}_2(t_f) = \{z_0 \in \mathbb{R}^{\mathrm{d}}, \exists \mathbf{r}(\cdot) \in \mathbf{B}, \exists s \in [0, t_f], \\ \text{s.t.} \ge 0, z_0^{\mathbf{u}_0,\mathbf{r}}(s) \in \mathbf{W}_1\}.\end{aligned} \tag{39}$$

4.3. Swarm System Stability Analysis. This section analyzes the stability of swarm system based on linear feedback control theory. Taking the cohesion rule as an example, it is proved that each UAV member in the swarm system flies with the centroid of other UAVs in the swarm, and the swarm system will eventually reach a stable equilibrium state.

4.3.1. Stability Analysis of Two UAV Members. The relative position relationship of the two UAV members is shown in Figure 2. The velocity vectors $\mathbf{V}_1, \mathbf{V}_2$ of the UAV are orthogonally decomposed along the sight distance line and the direction perpendicular to the sight distance line of the two UAV members, and the following can be obtained:

$$\begin{aligned}\dot{R}^r = V_2 \cos\theta_2 \cos\theta^r \cos(\psi_2 - \psi^r) + V_2 \sin\theta_2 \sin\theta^r \\ - V_1 \cos\theta_1 \cos\theta^r \cos(\psi_1 - \psi^r) - V_1 \sin\theta_1 \sin\theta^r,\end{aligned} \tag{40}$$

$$\dot{\psi}^r = \frac{V_2 \cos\theta_2 \sin(\psi_2 - \psi^r) - V_1 \cos\theta_1 \sin(\psi_2 - \psi^r)}{R^r \cos\theta^r}, \tag{41}$$

$$\begin{aligned}\dot{\theta}^r = \frac{-V_2 \cos\theta_2 \sin\theta^r \cos(\psi_2 - \psi^r) + V_2 \sin\theta_2 \cos\theta^r}{R^r} \\ + \frac{V_1 \cos\theta_1 \sin\theta^r \cos(\psi_1 - \psi^r) - V_1 \sin\theta_1 \cos\theta^r}{R^r},\end{aligned} \tag{42}$$

$$\dot{\psi}_1 = \frac{-k_1(\psi_1 - \psi^r)}{V_1 \cos\theta_1}, \tag{43}$$

$$\dot{\theta}_1 = \frac{-k_2(\theta_1 - \theta^r)}{V_1}, \tag{44}$$

$$\dot{\psi}_2 = \frac{k_1[\pi - (\psi_2 - \psi^r)]}{V_2 \cos\theta_2}, \tag{45}$$

$$\dot{\theta}_2 = \frac{k_2(-\theta_2 - \theta^r)}{V_1}. \tag{46}$$

Let $\delta\psi_1 = (\psi_1 - \psi^r), \delta\psi_2 = (\psi_2 - \psi^r)$, select six state variables as $R^r, \delta\psi_1, \delta\psi_2, \theta^r, \theta_1, \theta_2$, respectively, and linearize the above equations to obtain the following state equations:

$$\begin{bmatrix} \dot{R}^r \\ \delta\dot{\psi}_1 \\ \delta\dot{\psi}_2 \\ \dot{\theta}^r \\ \dot{\theta}_1 \\ \dot{\theta}_2 \end{bmatrix} = \begin{bmatrix} 0 & -V_1 & -V_2 & 0 & 0 & 0 \\ \frac{V_1 + V_2}{R^{r2}} & \frac{-k_1}{V_1} & 0 & 0 & 0 & 0 \\ \frac{V_1 + V_2}{R^{r2}} & 0 & \frac{-k_1}{V_2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{-V_1}{R^r} & \frac{V_2}{R^r} \\ 0 & 0 & 0 & \frac{-k_2}{V_1} & \frac{k_2}{V_1} & 0 \\ 0 & 0 & 0 & \frac{-k_2}{V_2} & 0 & \frac{-k_2}{V_2} \end{bmatrix} \begin{bmatrix} R^r \\ \delta\psi_1 \\ \delta\psi_2 \\ \theta^r \\ \theta_1 \\ \theta_2 \end{bmatrix}. \tag{47}$$

In order to simplify the calculation, let $V_1 = V_2 = V$ and $k_1 = k_2 = k$; according to the Laplace transform and the maximum value theory, the state of the equilibrium point is

$$\begin{aligned} R^r &= \frac{2V^2}{(k\pi/2)}, \\ \delta\psi_1 &= \frac{-\pi}{2}, \\ \delta\psi_2 &= \frac{\pi}{2}, \\ \theta^r &= 0, \\ \theta_1 &= 0, \\ \theta_2 &= 0. \end{aligned} \tag{48}$$

Let $V = x_1, 2V/R^{r2} = x_2, k/V = x_3, V/R^r = x_4$, then the characteristic equation becomes

$$\det(\lambda I - A) = 0. \tag{49}$$

The obtained eigenvalues are

$$\lambda = -x_3, -x_3, -\frac{1}{2}x_3 \pm \sqrt{{x_3}^2 - 8x_1x_2}, -\frac{1}{2}x_3 \pm \sqrt{{x_3}^2 - 8x_3x_4}. \tag{50}$$

Because $x_1, x_2, x_3, x_4 > 0$, ${x_3}^2 = (k/V)^2$ and $8x_1x_2 = \pi^2(k/V)^2, 8x_3x_4 = 2\pi(k/V)^2$, we have $8x_1x_2 > {x_3}^2, 8x_3x_4 > {x_3}^2$, and the eigenvalues of the characteristic equation are negative real numbers. Therefore, $\forall k, V$, the dual computer system composed of two UAV members is stable.

4.3.2. Stability Analysis of n UAV Members. Under the cohesion rule, the i-th UAV member in the swarm system follows the centroid motion of the other $(n-1)$ UAV members. Here, the average centroid of these $(n-1)$ UAV members is regarded as a virtual UAV member, and its track deflection angle and track inclination angle are expressed as

$$\psi_{i_{n-1}} = \left(\frac{1}{n-1}\right) \sum_{j=1, j\neq i}^{n} \left(\psi_j\right), \tag{51}$$

$$\theta_{i_{n-1}} = \left(\frac{1}{n-1}\right) \sum_{j=1, j\neq i}^{n} \left(\theta_j\right). \tag{52}$$

At this time, the line of sight vector $R^r_{ii_{n-1}}$, expected track deflection angle $\psi^r_{ii_{n-1}}$, and track inclination angle $\theta^r_{ii_{n-1}}$ are, respectively,

$$R^r_{ii_{n-1}} = \left(\frac{1}{n-1}\right) \sum_{j=1, j\neq i}^{n} \left(R^r_{ij}\right), \tag{53}$$

$$\psi^r_{ii_{n-1}} = \left(\frac{1}{n-1}\right) \sum_{j=1, j\neq i}^{n} \left(\psi^r_{ij}\right), \tag{54}$$

$$\theta^r_{ii_{n-1}} = \left(\frac{1}{n-1}\right) \sum_{j=1, j\neq i}^{n} \left(\theta^r_{ij}\right). \tag{55}$$

Then, the speed acceleration term is

$$\kappa_{\psi i} = -k\left(\psi_i - \psi^r_{ii_{n-1}}\right), \tag{56}$$

$$\kappa_{\theta i} = -k\left(\theta_i - \theta^r_{ii_{n-1}}\right). \tag{57}$$

Combining Equations (56) and (57) with Equations (4) and (5), respectively, we have

$$\dot{\psi}_i = \frac{-k\left(\psi_i - \psi^r_{ii_{n-1}}\right)}{V\cos\theta_i}, \tag{58}$$

$$\dot{\theta}_i = \frac{-k\left(\theta_i - \theta^r_{ii_{n-1}}\right)}{V}. \tag{59}$$

Similarly, it can be considered that the virtual point of $(n-1)$ UAV member follows the flight of the i-th UAV member, and the track deflection angular velocity acceleration term is expressed as

$$\begin{aligned} \kappa_{\psi i_{n-1}} &= \left(\frac{1}{n-1}\right) \sum_{j=1, j\neq i}^{n} \kappa_{\psi j} = \left(\frac{1}{n-1}\right) \sum_{j=1, j\neq i}^{n} -k\left(\psi_j - \psi^r_{jj_{n-1}}\right) \\ &= \left(\frac{k}{n-1}\right)\left(-\sum_{j=1, j\neq i}^{n} \psi_j + \sum_{j=1, j\neq i}^{n} \psi^r_{jj_{n-1}}\right). \end{aligned} \tag{60}$$

Here

$$\begin{aligned} \sum_{j=1, j\neq i}^{n} \psi^r_{jj_{n-1}} &= \left(\psi^r_{11_{n-1}} + \psi^r_{22_{n-1}} + \cdots + \psi^r_{(i-1)(i-1)_{n-1}}\right. \\ &\quad \left. + \psi^r_{(i+1)(i+1)_{n-1}} + \cdots + \psi^r_{nn_{n-1}}\right), \\ &= \left(\frac{1}{n-1}\right)\left(\left(\psi^r_{12} + \psi^r_{13} + \cdots + \psi^r_{1n}\right)\right. \\ &\quad + \left(\psi^r_{21} + \psi^r_{23} \cdots + \psi^r_{2n}\right) + \cdots \\ &\quad + \left(\psi^r_{(i+1)1} + \psi^r_{(i+1)2} + \cdots + \psi^r_{(i+1)(i-2)}\right. \\ &\quad \left. + \psi^r_{(i-1)i} + \cdots + \psi^r_{(i-1)n}\right) \\ &\quad + \cdots + \left(\psi^r_{(i+1)1} + \psi^r_{(i+1)2} + \cdots + \psi^r_{(i+1)i}\right. \\ &\quad + \psi^r_{(i+1)(i-2)} + \cdots \\ &\quad \left.\left. + \psi^r_{(i+1)n}\right) + \cdots + \left(\psi^r_{n1} + \psi^r_{n2} + \cdots + \psi^r_{n(n-1)}\right)\right). \end{aligned} \tag{61}$$

According to the relative position relationship of UAV, it can be seen that $\psi^r_{ij} = \pi - \psi^r_{ji}$; then,

$$\sum_{j=1, j\neq i}^{n} \psi^r_{jj_{n-1}} = \pi(n-1) + \sum_{j=1, j\neq i}^{n} \psi^r_{ij}. \tag{62}$$

Substituting equation (62) into equation (60), the following can be obtained:

$$\kappa_{\psi i_{n-1}} = k\left\{\pi - \left(\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\psi_j - \frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\psi_{ij}^{r}\right)\right\}. \tag{63}$$

It can be obtained from equations (50) and (54) that

$$\kappa_{\psi i_{n-1}} = k\left(\pi - \left(\psi_{i_{n-1}} - \psi_{ii_{n-1}}^{r}\right)\right). \tag{64}$$

Simultaneously equation (62) and equation (4) can obtain

$$\dot{\psi}_{i_{n-1}} = \frac{k\left(\pi - \left(\psi_{i_{n-1}} - \psi_{ii_{n-1}}^{r}\right)\right)}{V\cos\theta_{i_{n-1}}}. \tag{65}$$

Similarly, the track inclination velocity acceleration term is expressed as

$$\begin{aligned}\kappa_{\theta i_{n-1}} &= \left(\frac{1}{n-1}\right)\sum_{j=1,j\neq i}^{n}\kappa_{\theta j},\\ &= \left(\frac{1}{n-1}\right)\sum_{j=1,j\neq i}^{n} -k\left(\theta_j - \theta_{jj_{n-1}}^{r}\right),\\ &= \left(\frac{k}{n-1}\right)\left(-\sum_{j=1,j\neq i}^{n}\theta_j + \sum_{j=1,j\neq i}^{n}\theta_{jj_{n-1}}^{r}\right),\end{aligned} \tag{66}$$

where

$$\begin{aligned}\sum_{j=1,j\neq i}^{n}\theta_{jj_{n-1}}^{r} &= \left(\theta_{11_{n-1}}^{r} + \theta_{22_{n-1}}^{r} + \cdots + \theta_{(i-1)(i-1)_{n-1}}^{r}\right.\\ &\left. + \theta_{(i+1)(i+1)_{n-1}}^{r} + \cdots + \theta_{nn_{n-1}}^{r}\right),\\ &= \left(\frac{1}{n-1}\right)\left(\left(\theta_{12}^{r} + \theta_{13}^{r} + \cdots + \psi_{1n}^{r}\right) + \left(\theta_{21}^{r} + \theta_{23}^{r}\cdots + \theta_{2n}^{r}\right) + \cdots\right.\\ &+ \left(\theta_{(i+1)1}^{r} + \theta_{(i+1)2}^{r} + \cdots + \theta_{(i+1)(i-2)}^{r} + \theta_{(i-1)i}^{r} + \cdots + \theta_{(i-1)n}^{r}\right)\\ &+ \cdots + \left(\theta_{(i+1)1}^{r} + \theta_{(i+1)2}^{r} + \cdots + \theta_{(i+1)i}^{r} + \theta_{(i+1)(i-2)}^{r} + \cdots\right.\\ &\left.\left. + \theta_{(i+1)n}^{r}\right) + \cdots + \left(\theta_{n1}^{r} + \theta_{n2}^{r} + \cdots + \theta_{n(n-1)}^{r}\right)\right).\end{aligned} \tag{67}$$

According to the relative position relationship of UAV, it can be seen that $\theta_{ij}^{r} = -\theta_{ji}^{r}$; then,

$$\sum_{j=1,j\neq i}^{n}\theta_{jj_{n-1}}^{r} = -\sum_{j=1,j\neq i}^{n}\theta_{ij}^{r}. \tag{68}$$

Substituting equation (68) into equation (66), the following can be obtained:

$$\kappa_{\theta i_{n-1}} = \frac{k}{n-1}\left(-\sum_{j=1,j\neq i}^{n}\theta_j - \sum_{j=1,j\neq i}^{n}\theta_{ij}^{r}\right). \tag{69}$$

It can be obtained from equations (52) and (55) that

$$\kappa_{\theta i_{n-1}} = -k\left(\theta_{i_{n-1}} + \theta_{ii_{n-1}}^{r}\right). \tag{70}$$

Simultaneously equation (70) and equation (5) can obtain

$$\dot{\theta}_{i_{n-1}} = \frac{-k\left(\theta_{i_{n-1}} + \theta_{ii_{n-1}}^{r}\right)}{V}. \tag{71}$$

According to equations (40)–(46), the virtual point of $(n-1)$ UAV members is regarded as the second member of the two UAV members; that is, ψ_2, θ_2 is replaced by $\psi_{i_{n-1}}$, $\theta_{i_{n-1}}$, and the sight distance vector R^r, expected track deflection angle, and track inclination angle ψ^r, θ^r are replaced by $R_{ii_{n-1}}^{r}$ and $\psi_{ii_{n-1}}^{r}$, $\theta_{ii_{n-1}}^{r}$, respectively. The forms of equations (58), (59), (65) and (70), (47) are similar to equations (48)–(53) established by two UAV members in the sight distance direction. Therefore, the stability of two UAV members in the swarm system can be analyzed and proved by the stability of two UAV members. According to Section 4.1, the swarm system composed of n UAV members has a stable equilibrium state; that is, the swarm system is stable.

5. Simulation Experiment and Verification

In order to facilitate the result analysis and discussion, the simulation verification of UAV swarm flight and cooperative avoidance in two-dimensional plane is mainly carried out. Without losing generality, the simulation scenario is set as avoidance between two swarm systems. For one swarm, the other swarm is collision threat. Both swarm systems adopt the autonomous cooperative control method of flight and avoidance proposed in this paper. In this scenario, the following three simulation experiments are carried out.

5.1. Simulation 1: Swarm System Avoidance Process. The overall evasion mechanism and member evasion mechanism are used to simulate the evasion process of two UAV swarm systems in two-dimensional space. The swarms are in the two-dimensional horizontal plane, and the avoidance decision mainly depends on the change of track deflection angle of UAV members. The speed of UAV members is constant as $20\,m/s$, and $s_{\min} = 60\,m$, $s_1 = 150m$, and $s_2 = 60\,m$. The available detection distance of airborne sensors is set to a fixed value $\rho = 600\,m$, and the other two are $\rho_c = 400\,m$ and $\rho_e = 600\,m$. The initial track deflection angles of the two swarm systems are $\psi_1 = 0^\circ$, $\psi_2 = 180^\circ$. The dynamic characteristics of UAV and the constraints are roll angle $-20^\circ \le \phi_c \le 20^\circ$, overload in pitch plane $n_y \le 2\,g$, overload in yaw plane $n_z \le 4\,g$, and turning radius $r_d \ge 200\,m$. When the swarm system flies normally, the initial value of the weight corresponding to each flight rule is

$$\begin{aligned}\tau_{i_M} &= 0.55, \tau_{i_F}.\\ &= 0.22, \tau_{i_I},\\ &= 0.78, \tau_{i_S},\\ &= 0.55, \tau_{i_A},\\ &= 0.42, \tau_{i_E},\\ &= 0.\end{aligned} \tag{72}$$

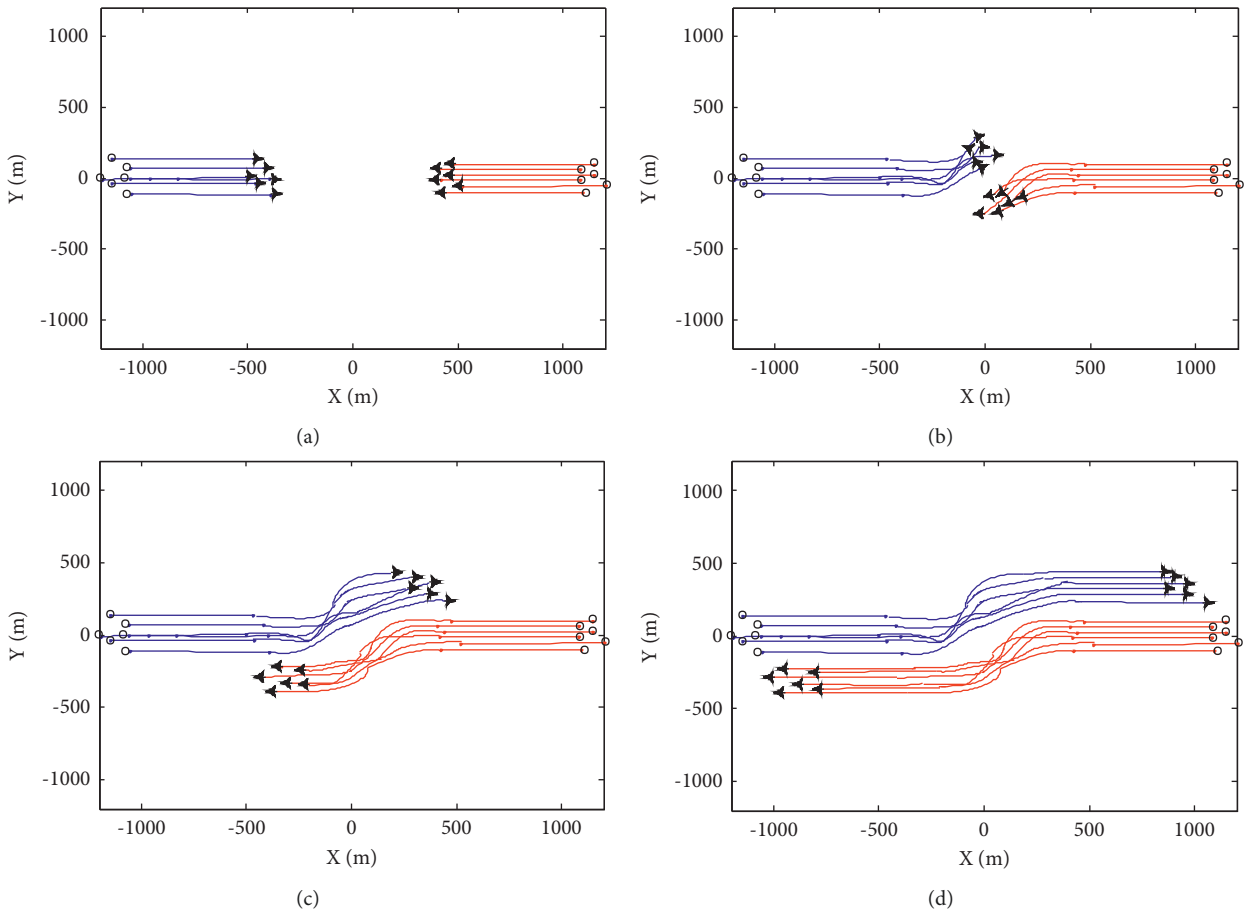


Figure 3: Swarm avoidance process under the overall mechanism. (a) $t = 25s$. (b) $t = 50s$. (c) $t = 75s$. (d) $t = 100s$.

If the swarm system detects external UAV members, under the overall and member avoidance mechanism, the weights change to, respectively,

$$\begin{aligned}
\tau_{i_M}' &= 0.55, \tau_{i_F}', \\
&= 0.22, \tau_{i_I}', \\
&= 0.38, \tau_{i_S}', \\
&= 0.55, \tau_{i_A}', \\
&= 0.42, \tau_{i_E}', \\
&= 0.55, \\
\tau_{i_M}'' &= 0.30, \tau_{i_F}'', \\
&= 0.30, \tau_{i_I}'', \\
&= 0.38, \tau_{i_S}'', \\
&= 0.55, \tau_{i_A}'', \\
&= 0.42, \tau_{i_E}'', \\
&= 0.55.
\end{aligned} \tag{73}$$

The avoidance process of UAV swarm is shown in Figures 3 and 4. The flight track of swarm 1 is represented by blue line and that of swarm 2 is represented by red line. During the avoidance process of the swarm system, the variation curves of the minimum distance $R_{\min}$ between UAV members and the maximum distance $d_{\max}$ from all UAV members to the swarm centroid are shown in Figure 5(a) and 5(b), respectively.

From the simulation 1 experimental results, it can be seen that under the action of basic flight rules, UAV swarm can maintain stable flight in a certain formation and can realize formation reconstruction after evasion. Under the action of overall evasion mechanism and member evasion mechanism, the swarm system can complete evasion and realize safe flight. The difference between the two mechanisms is that in the overall mechanism, the variation of UAV track deflection angle relative to the initial flight direction is relatively large; that is, the overall track cost is high. In the member mechanism, the overall change of UAV track deflection angle is small, but the change rate is large, and the avoidance process is more complex.

5.2. Simulation 2: The Influence of Flight Rule Weight on the Stable Scale of the Swarm System. The influence of several flight rule weight coefficients τ_M, τ_I, τ_S on the stable scale of swarm system is studied. First, make the weight of the self-guidance rule constant $\tau_I = 0.6$, and the weight of the dispersion rule takes values $\tau_S = 0.2, \tau_S = 0.4, \tau_S = 0.6$, respectively. The weight of the cohesion rule τ_M changes from

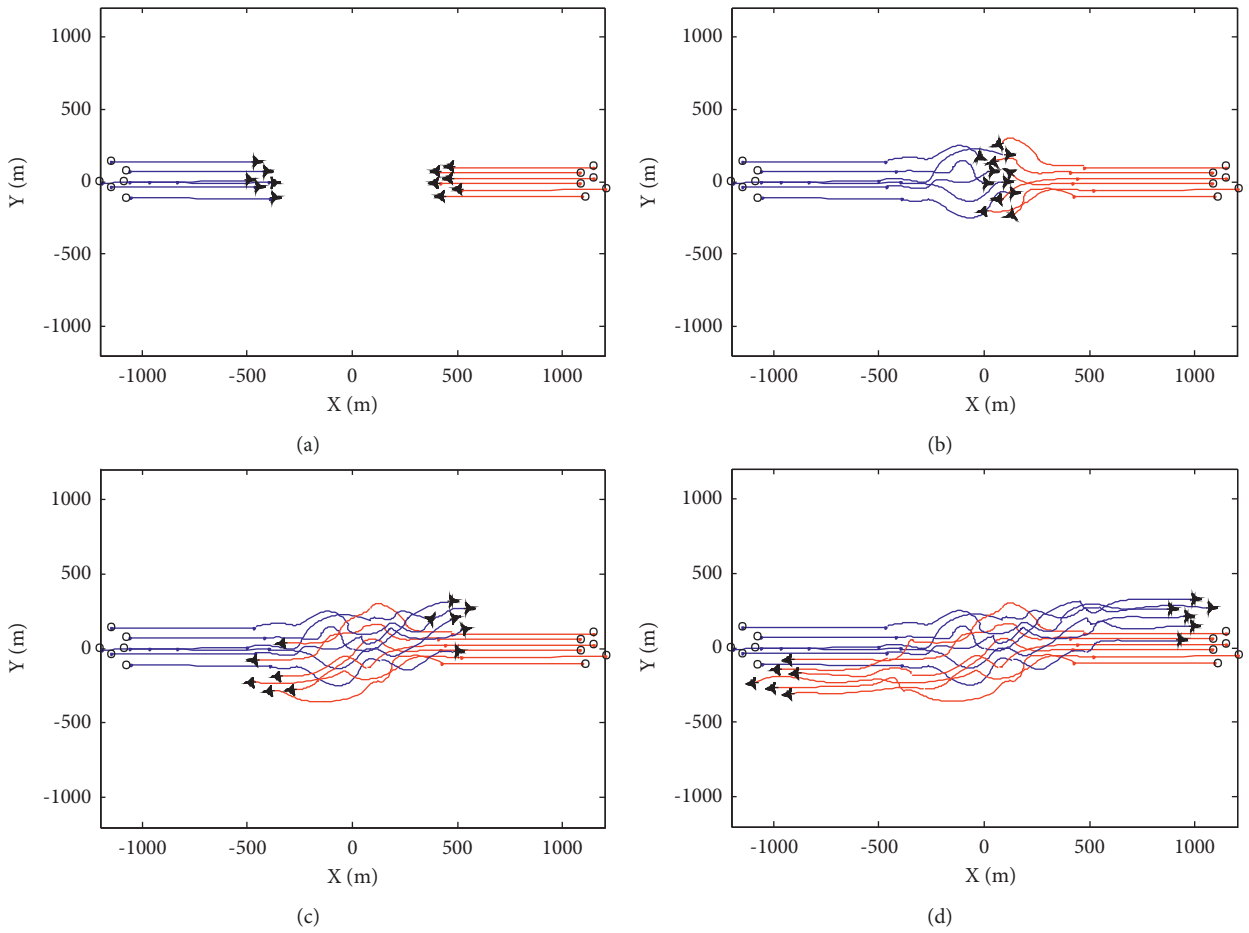


Figure 4: Swarm avoidance process under membership mechanism. (a) $t = 25s$. (b) $t = 50s$. (c) $t = 75s$. (d) $t = 100s$.

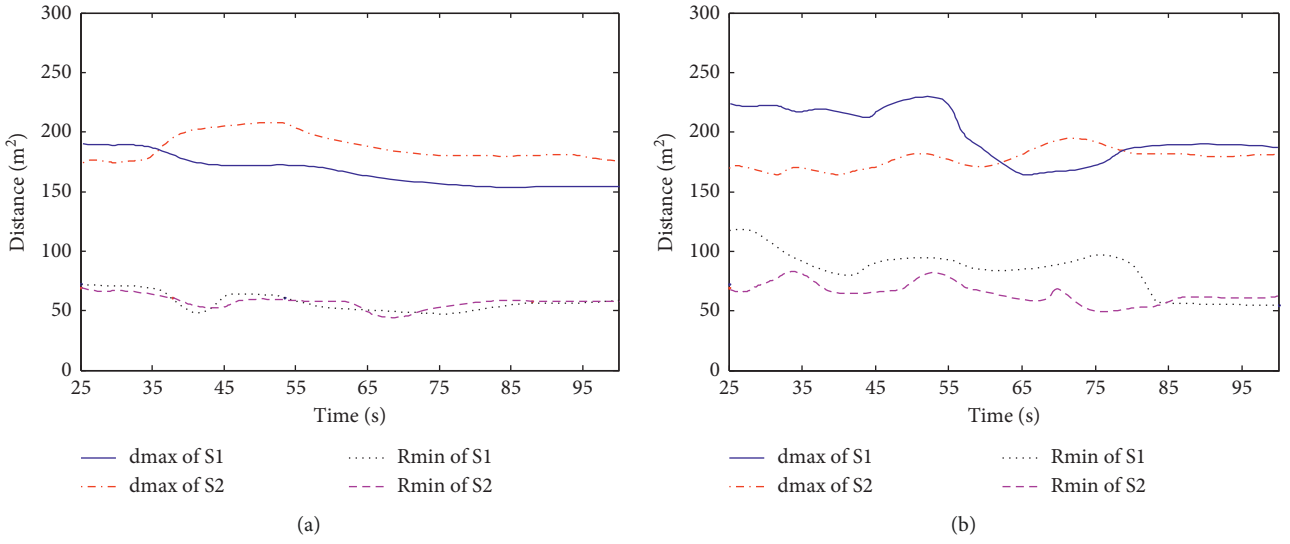


Figure 5: Variation curves of R_{min} and d_{max}. (a) Overall mechanism. (b) Membership mechanism.

0.1 to 1.0. When the swarm system is stable, the scale changes with the weight of the cohesion rule, as shown in Figure 6(a). Then, let the weight of the dispersion rule be a constant value $\tau_S = 0.4$, and the weight of the self-guidance rule takes values $\tau_I = 0.3$, $\tau_I = 0.5$, $\tau_I = 0.7$, respectively. The weight of the cohesion rule τ_M changes from 0.1 to 1.0. When the swarm system is stable, the scale changes with the weight of the cohesion rule, as shown in Figure 6(b).

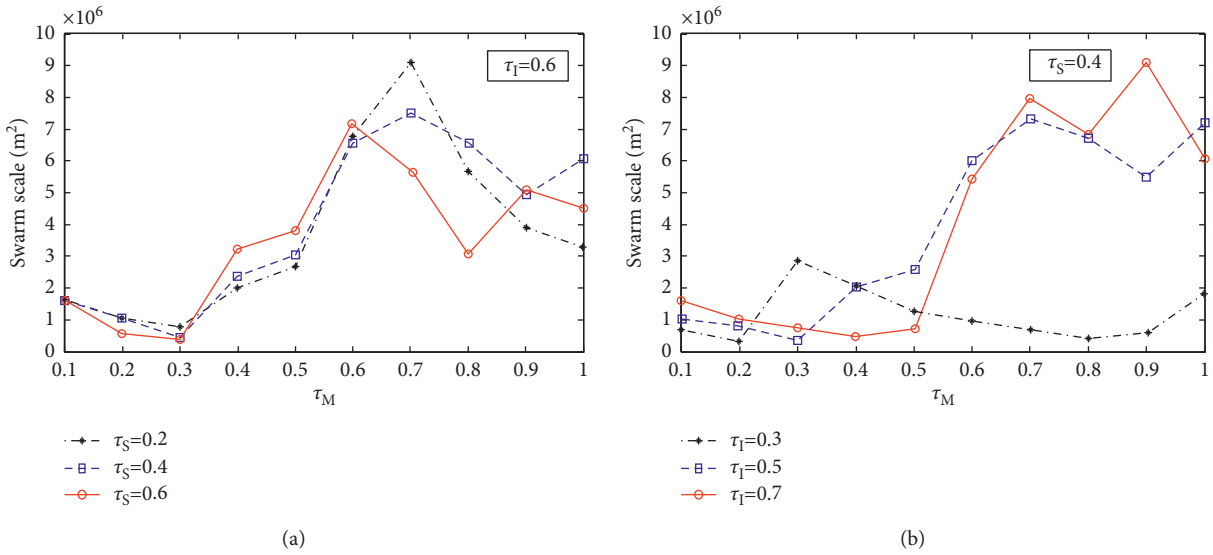


FIGURE 6: Influence of weight change of flight rules on system scale. (a) $\tau_I = 0.6$, $\tau_S = 0.2, 0.4, 0.6$. (b) $\tau_S = 0.4$, $\tau_I = 0.3, 0.5, 0.7$.

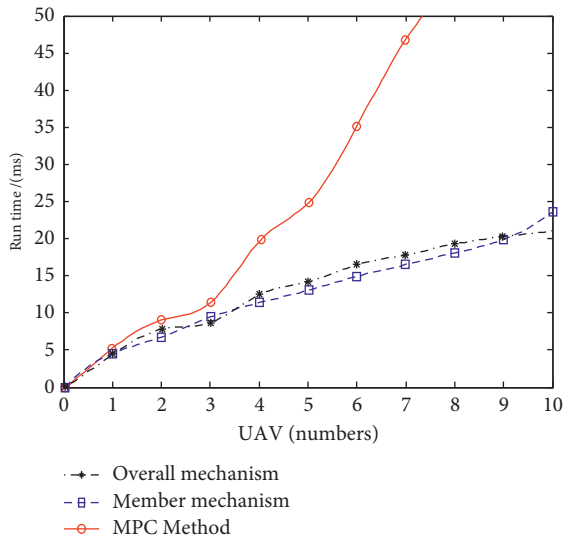


FIGURE 7: Comparison of average running time of algorithm.

From the simulation 2 experimental results in Figure 6(a), it can be seen that the change trend of the scale of the swarm system is as follows: first, it decreases with the increase of the weight of the cohesion rule τ_M and then increases with the increase of τ_M. This is because the cohesion within the swarm system is obvious at the beginning, and the stability of the swarm system gradually decreases with the continuous increase of τ_M. As can be seen from Figure 6(b), the scale of the swarm system has a similar change trend and increases with the increase of the weight of the self-guidance rule τ_I, because the cohesion decreases with the increase of τ_I. It can be seen that the value of τ_M between $0.2 \sim 0.4$ is more appropriate for the small scale and stability of the system.

5.3. Simulation 3: Real-Time Performance of Evasion Control. This paper discusses the real-time performance of swarm system evasion control method and compares it with MPC algorithm, in which MPC algorithm is proposed in reference [17]. In the two algorithms, the simulation scene settings are the same, and they are implemented in Visual *C*++ 6.0 and MATLAB R2021a environment. The computer is configured as Intel Core i5 processor, main frequency 3.10 GHz, memory 8 G, and 32-bit operating system. After multiple simulations, the average running time of the algorithm is shown in Figure 7.

It can be seen from the simulation 3 experimental results that the running time of MPC method increases exponentially with the increase of the number of formation UAVs. When the number of UAVs is large, the calculation efficiency is greatly reduced and the running time of the algorithm is long. The running time of the overall mechanism and member mechanism in the swarm algorithm increases slowly with the increase of the number of UAVs in the swarm. When the number of UAVs in the swarm is no more than 10, the execution time of the algorithm is within 25 ms, which can meet the needs of online real-time control.

6. Conclusions

In this paper, a cognitive game method is used to study the anticollision cooperative decision-making and control of UAV swarms, and the method is used to simulate the flight and avoidance process between UAV swarms. The basic flight behavior rules of UAV swarm are designed. Under the action of these rules, UAV swarm can realize normal flight and formation reconstruction. Aiming at the problem of anticollision, two anticollision mechanisms, named whole and member mechanism, are put forward. At the same time, a cognitive security domain facing swarm anticollision is constructed. On this basis, the game theory model of swarm anticollision problem is established, and the solution method of swarm anticollision game decision and control strategy is designed. Finally, the stability of UAV swarm system using the above method is analyzed. Simulation

results show that compared with the traditional MPC method, the swarm control proposed in this paper is simpler and more efficient.

Data Availability

The data used to support the findings of this study are included within the article.

Disclosure

Huan Zhou and Yintong Li are the co-first authors.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors acknowledge funding received from the following science foundations: National Natural Science Foundation of China (Nos. 62101590 and 62176214) and the Natural Science Foundation of the Shaanxi Province, China (2020JQ-481 and 2019JQ-014).

References

[1] J. Kusyk, M. U. Uyar, K. Ma et al., "Artificial intelligence and game theory controlled autonomous UAV swarms," *Evolutionary Intelligence*, vol. 14, no. 4, pp. 1775–1792, 2021.

[2] A. Puente Castro, D. Rivero, A. Pazos, and E Fernandez Blanco, "A review of artificial intelligence applied to path planning in UAV swarms," *Neural Computing & Applications*, vol. 34, no. 1, pp. 153–170, 2022.

[3] L. Wang, J. Hu, Z. Xu, and C. Zhao, "Autonomous maneuver strategy of swarm air combat based on DDPG," *Autonomous Intelligent Systems*, vol. 1, no. 1, p. 15, 2021.

[4] D. H. Stolfi, M. R. Brust, G. Danoy, and P. Bouvry, "CONSOLE: intruder detection using a UAV swarm and security rings," *Swarm Intelligence*, vol. 15, no. 3, pp. 205–235, 2021.

[5] R. Rahimi, F. Abdollahi, and K. Naqshi, "Time-varying formation control of a collaborative heterogeneous multi agent system," *Robotics and Autonomous Systems*, vol. 62, no. 12, pp. 1799–1805, 2014.

[6] C. Liu, M. Wang, Q. Zeng, and W. Huangfu, "Leader-following flocking for unmanned aerial vehicle swarm with distributed topology control," *Science China Information Sciences*, vol. 63, no. 4, pp. 140312–141153, 2020.

[7] M. A. Trujillo, R. Aldana-López, D. Gómez-Gutiérrez, M. Defoort, J. Ruiz-León, and H. M. Becerra, "Autonomous and non-autonomous fixed-time leader-follower consensus for second-order multi-agent systems," *Nonlinear Dynamics*, vol. 102, no. 4, pp. 2669–2686, 2020.

[8] W. Crowther, "Rule-based guidance for flight vehicle flocking," in *Proceedings of the Institution of Mechanical Engineers - Part G: Journal of Aerospace Engineering*, vol. 218, no. 2, pp. 111–124, Professional Engineering Publishing LTD, 1 Birdcage Walk, Westminister, England, 2004.

[9] X. Wang, L. Shen, Z. Liu et al., "Coordinated flight control of miniature fixed-wing UAV swarms: methods and experiments," *Science China Information Sciences*, vol. 62, no. 11, Article ID 212204, 2019.

[10] M. Çolak, İ. Kaya, A. Karaşan, and M. Erdoğan, "Two-phase multi-expert knowledge approach by using fuzzy clustering and rule-based system for technology evaluation of unmanned aerial vehicles," *Neural Computing & Applications*, vol. 34, no. 7, pp. 5479–5495, 2022.

[11] H. S. Su, X. F. Wang, and Z. L. Lin, "Flocking of multi-Agents with a virtual leader," *IEEE Transactions on Automatic Control*, vol. 54, no. 2, pp. 293–307, 2009.

[12] H. Oh, S. Kim, H. S. Shin, B. A. White, A. Tsourdos, and C. A. Rabbath, "Rendezvous and standoff target tracking guidance using differential geometry," *Journal of Intelligent and Robotic Systems*, vol. 69, no. 1-4, pp. 389–405, 2013.

[13] W. h. Liu, X. Zheng, and Z. h. Deng, "Dynamic collision avoidance for cooperative fixed-wing UAV swarm based on normalized artificial potential field optimization," *Journal of Central South University*, vol. 28, no. 10, pp. 3159–3172, 2021.

[14] A. A. Massoud, "Managing the dynamics of a harmonic potential field-guided robotic in a cluttered environment," *IEEE Transactions on Industrial Electronics*, vol. 56, pp. 488–496, 2009.

[15] R. Shao, R. Tao, Y. Liu, Y. Yang, D. Li, and J. Chen, "UAV cooperative search in dynamic environment based on hybrid-layered APF," *EURASIP Journal on Applied Signal Processing*, vol. 2021, no. 1, p. 101, 2021.

[16] J. Dentler, M. Rosalie, G. Danoy et al., "Collision avoidance effects on the mobility of a UAV swarm using chaotic ant colony with model predictive control," *Journal of Intelligent and Robotic Systems*, vol. 93, no. 1-2, pp. 227–243, 2019.

[17] J. Erskine, R. Balderas-Hill, and I. Fantoni, "Model predictive control for dynamic quadrotor bearing formations," in *Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 124–130, IEEE, Xi'an, China, 2021.

[18] E. Nejabat and A. Nikoofard, "Switched robust model predictive based controller for UAV swarm system," in *Proceedings of the 2021 29th Iranian Conference on Electrical Engineering (ICEE)*, pp. 721–725, IEEE, Xi'an, China, 2021.

[19] V. Izhboldina, I. Lebedev, and A. Shabanova, "Approach to UAV swarm control and collision-free reconfiguration," in *Proceedings of the 15th International Conference on Electromechanics and Robotics Zavalishin's Readings*, pp. 81–92, Springer, Singapore, 2020.

[20] Z. Cao, Y. Zhang, J. Guan, S. Zhous, and G. Wen, "A chaotic ant colony optimized link prediction algorithm," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 51, no. 9, pp. 5274–5288, 2021.

[21] P. Li, X. Liu, H. Chen, B. Li, T. Ma, and W. Jiang, "Optimization of three-dimensional magnetic field in vacuum interrupter using particle swarm optimization algorithm," *IEEE Transactions on Applied Superconductivity*, vol. 31, no. 8, pp. 1–4, 2021.

[22] V. Kaltsa, A. Briassouli, I. Kompatsiaris, L. J. Hadjileontiadis, and M. G. Strintzis, "Swarm intelligence for detecting interesting events in crowded environments," *IEEE Transactions on Image Processing*, vol. 24, no. 7, pp. 2153–2166, 2015.

[23] J. Yang, X. Xu, D. Yin, Z. Ma, and L. Shen, "A space mapping based 0-1 linear model for onboard conflict resolution of heterogeneous unmanned aerial vehicles," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 7455–7465, 2019.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 2500772, 12 pages
https://doi.org/10.1155/2022/2500772

Research Article

Analysis of e-Mail Spam Detection Using a Novel Machine Learning-Based Hybrid Bagging Technique

Alanazi Rayan

Department of Computer Science, Jouf University, Sakaka, Saudi Arabia

Correspondence should be addressed to Alanazi Rayan; rmalanazi@ju.edu.sa

Received 16 June 2022; Accepted 13 July 2022; Published 9 August 2022

Academic Editor: Heng Liu

e-mail service providers and consumers find it challenging to distinguish between spam and nonspam e-mails. The purpose of spammers is to spread false information by sending annoying messages that catch the attention of the public. Various spam identification techniques have been suggested and evaluated in the past, but the results show that the more research in this regard is required to enhance accuracy and to reduce training time and error rate. Thus, this research proposes a novel machine learning-based hybrid bagging method for e-mail spam identification by combining two machine learning methods: random forest and J48 (decision tree). The proposed framework categorizes the e-mail into ham and spam. The database is split into multiple sets and provided as input to each method in this procedure. Moreover, tokenization, stemming, and stop word removal are performed in the preprocessing stage. Further, correlation feature selection (CFS) is employed in this research to select the required features from the preprocessed data. The effectiveness of the presented method is evaluated in terms of true-negative rates, accuracy, recall, precision, false-positive rate, *f*-measure, and false-negative rate; the outcomes of three studies are compared. According to the results, the presented hybrid bagged model-based SMD technology achieved 98 percent accuracy.

1. Introduction

Over the years, e-mail has proven to be an immensely important method of communication, offering near-instant access to anyone around the world with an Internet connection. According to Tschabitscher, there was about 5 billion monthly active e-mail account in 2017, with that number likely to rise to over 5.5 billion per month in 2019. The investigator also points out that, even though over 270 billion e-mails are being sent every day, around 57 percent of them are trash. To detect spam or phishing e-mails, there seem to be several emerging machine learning algorithms and also strategies that closely match immune defense mechanisms, but overall effectiveness has been a big worry. The majority of the strategies are effective in preventing spam, but they also prevent certain nonspam communications, known as ham. This is an issue since it may result in the users losing pertinent data. Diverse kinds of e-mail assaults, including phishing, e-mail spoofing, and variations of spam e-mails, such as a covert redirect, clone phishing, spear phishing, and whaling, are continually bombarding customers worldwide. To make the warning message get sent from a valid user, e-mail spoofing frequently entails faking the e-mail header (the from portion). Spammers use mail impersonating since individuals are more likely to read an e-mail if it appears to be from someone they know [1].

Spamming, which would be described as the practice of delivering unwanted communications to a large group of people, is on the rise right now. Since reputational risk and economical disruptions are one of the spammers' highest significant things, the spam influx has prompted academia scholars to investigate this issue as a promising proposed study. As a result, both organizations and people require effective and trustworthy junk e-mail filtration solutions. To respond to the growing volume of unwanted e-mails, these computers must apply sophisticated tactics. For a long period, spam e-mail identification has indeed been intensively researched [2]. There are two types of spam e-mail filtering techniques now available: unimodal and multimodal. Message and image-based are the two categories of

the first group. The earlier was established to identify spam e-mails that simply contained information, whereas the latter was designed to handle embedding textual inside graphics, understanding that fraudsters used it to bypass message solutions. Multimodal technologies, the subordinate group, are required to identify spam e-mails that contain either text or images, sometimes known as hybrid spam e-mails. Considering the number of data contained, hybrid spam e-mail seems to be more dangerous and sophisticated than image-based and message-spam e-mail. This makes it an essential process. As a result, effective methods for processing both text and visual material are required for hybrid spam e-mail. The goal of the multifunctional system is to produce relevant features using visual and textual modes before integrating these at the characteristic or choice levels [3].

Many e-mail providers have included an automated trash detection performance based on the previous e-mail transmitted records as the premise of classification. Nevertheless, there are still e-mail servers, especially those run by individuals, that do not have trash detection performance. This is because not all private e-mail products include phishing identification as one of their application software packages by the standard. As a result, a spam detection mechanism should be developed and implemented as a software program or an add-on component. However, owing to the inherent flexibility of unsolicited messages by use of mail systems, the efficacy of trash identifying systems has frequently been restricted, or in some cases made useless or corrupted, necessitating the development of stronger spam identification techniques [4]. Many spam identifying techniques have indeed been suggested and evaluated in the research; however, the claimed accuracy nevertheless calls for more research in this area to improve accuracy. Researchers used an artificial neural network-based model to identify spamming but only managed to achieve an efficiency of 86 percent, which is still far from optimal. The researchers utilized a naïve Bayes strategy for extracting features while combining expense multi-objective genetic programming for phishing detection, with an efficiency of 79.4 percent properly detecting e-mail kinds. Furthermore, the paper developed a spam identification structure based on interval derivative fuzzy sets, investigating the potential of type-2 fuzzy logic, and was only able to achieve a spam prediction performance of 86.8% for the test dataset [5].

Individuals and businesses gain greatly from the advancement of technological advancement. Simultaneously, technological advancement acts as a conduit for illicit activity. While hacker or virus utilization is a specialized talent that necessitates behaviors that are not performed by the average individual, social engineering is not as technically demanding. As a result, social engineering is becoming increasingly prevalent. In the third period of 2019, social engineering attacks have been the most common hazard to individuals and the second most common threat to businesses [6]. In most situations, media manipulation is the first stage in the progression of cybercriminals. "Cyberattacks of a network of an organization began in 81 percent of the cases with a phishing message." A malicious program is a form of harassment that aims to steal private and personal data out of its potential victims. Phishing attacks involve a range of different methods, the much more popular of which include e-mail postings, phone conversations, social networking site messaging, and many others [7]. To combat phishing and minimize the primary cybercrime, it is critical to recognize phishing e-mails. While technological anti-phishing measures are insufficient, personal and organizational security is based on education and knowledge of the misleading assault environment. Researchers will concentrate on e-mail-based phishing assaults in this article. e-mail communications are utilized as an interactive setting to gather customized content in e-mail-based phishing scams [8].

Every day, people are inundated with hundreds of commercials. Users know how to avoid and decrease their access to information and messages that are uninteresting or unreliable throughout the term. Personal factors are one technique for advertisers to pique the interest of the customer base [9]. Marketers employ Internet shopping data to increase the relevancy of advertising and distribute them to particular customers to make companies stick out now to customers in an electronic medium bombarded with the ever-growing congestion of adverts. Personalized advertising is a paid communication "matched to a user's attributes, hobbies, or preferences." Because tailored information is recognized being a persuasive communication tactic that favorably affects attentiveness, mental reactions, and perceptions toward advertising, the capacity to gather customer information Internet is critical [10]. Customization is intended to produce favorable responses because consumers demand to advertise that are pertinent to them, but nonrelated SNS commercials are likely ignored because they are viewed as intrusive or aggressive. As a result, prior research demonstrates this need for customization in the digital world because it enhances marketing effectiveness by making adverts more unforgettable and personable, ramping up inspiration to handle advertising messages, instituting behavior changes, and ramping up the marketing rate of response [11].

Another benefit of marketing comes from the ability to identify a particular targeted audience due to self-characteristics such as age and sexuality available on profile pages. Nevertheless, there is little study on personalized social media ads, and future studies should go beyond statistics to include information specific to social media. Conventional systems for gathering customer data may provide marketers with limited or erroneous data about customers. Because social media profiles are based on voluntary self-disclosure of private details, there is the potential for purposeful distortion, allowing for the creation of fake accounts, online trolls, and chatbots. As a result, aggregating SNS information may result in mistakes in customer profiles and focused marketing [12]. The following are the particular research materials for message spam filtering: research employs the classification tree approach to filter trash, which selects text characteristics using the RLM perceptions rather than the data gain technique. The reliability and memory rates of this technique are both over 89 percent, according to the trial data. To acquire multiple classification likelihood functions

of an e-mail as junk, the study employed numerous strong classifiers, each of which uses a logarithmic regress approach to achieve the categorization confidence interval. Finally, a promoting approach is used to aggregate the several categorization likelihood functions to obtain actual data for the e-mails as junk, which will then be matched to a benchmark to evaluate whether or not the e-mail is malware [13].

Section 2 explains previous research related to this research, and Section 3 describes the methodology of our work. Section 4 introduces to the basics of decision tree algorithms random forest and J4. Section 5 explains the system design. Section 6 reports on computational studies of the proposed approach, and Section 7 concludes with findings and work to be done in the future.

2. Related Works

Their daily lives are becoming increasingly reliant on social media. Our communication through these platforms has only grown in tandem with their fast expansion. Twitter is among the most widely used social media platforms in the Middle East. Tweet, like other social networking sites, is susceptible to spam profiles that distribute part of the strategy. Leading to a shortage of appropriate technology that serves the Arabic language, Arab countries have been among the most attacked. Furthermore, because Arabic is a sophisticated language with multiple varieties and significant grammatical rules, retrieving text data might be difficult. Several recent researches have looked into inventive ways to reduce spamming in tweets. The research collected Arabic datasets appropriate for junk identification to tackle the problem of identifying spamming identities in Arab on tweet. Utilizing Twitter's subscription application programming interface, the database included information from premium content. Abandoned identities were flagged, and information was labeled. A hybrid approach that is based on deep learning algorithms was developed, which has numerous benefits, along with more efficient and timely outcomes while using fewer system resources. Text-based statistical analyses were performed using a convolution neural network (CNN) approach, while information was analyzed using a basic neural network-based framework. When the two algorithms' outputs were pooled, they categorized identities as spam or not spam. The suggested approach surpassed the best designs evaluated thus far in the research, achieving an efficiency of 94.28 percent with the hybrid algorithm employing superior extracted features. In interaction via social media platforms like Twitter, several Arabic accents and informal idioms intersect. This makes it more difficult to identify spam identities utilizing solely text-based characteristics and therefore necessitates several earlier processes to acquire correct categorization. Additional research into a preprocessing stage that might accommodate Arabic accents with minimum impact on intentions and interpretation would have been beneficial. This is considered the major limitation of spam accounts combined with the text and meta-based deep learning framework proposed by [14].

In numerous businesses, especially marketing, the categorization, and suggestion method for recognizing social networking site (SNS) members' preferences play an important part. Customized adverts help firms stick in a sea of digital advertising by increasing relevancy to customers and eliciting positive reactions [15]. The comprehensive evaluation of photos and messages on client postings could more accurately forecast a user's preferences, even though almost all user preference categorization research had concentrated on text information. As a result, this study uses both language and visuals to classify SNS participants' preferences. The Curlie directory was used to describe the interests of consumers, and researchers compared alternative convolutional neural network (CNN) and recurrent neural network (RNN)-based models for the user preference categorization systems [16]. Convolutional neural classification methods have been used to categorize photos via individuals' SNS posts, whereas RNN-based classification methods were utilized to classify text information in their hybrid neural net system. The categorization of users' preferences fared best while utilizing text and graphics combined, at 96.5 percent, vs texts alone, 41.39 percent, or photos only, 93.2 percent, according to the findings of the thorough trials. Our suggested approach helps marketers make (1) interest-based suggestions, (2) ranked-order suggestions, and (3) real-time suggestions by giving insight into tailored SNS marketing communications. To their understanding, this is one of the earliest articles to leverage combined image and message statistics using user-generated material to enhance the effectiveness of reliably identifying the political inclinations of SNS users for such aim of improving targeted advertising experiences [17].

Single-modal spam filtering algorithms have obtained a maximum classification performance for picture and text junk in the latest days. To remain undetected by single-modal spamming filtering techniques, fraudsters introduce garbage data into the multi-modality component of an e--mail and blend it to lower the single-modal spam detection processor architectures' classification performance, so achieving the goal of dodging identification. In light of this, the latest product known as multimodal design obtained from the numerical fusing (MMA-MF) has been presented, which employs a multimodal fusing strategy to ensure that trash can be efficiently filtered whether it has been concealed in word or phrase. To filter trash, the approach integrates a convolutional neural network (CNN) with a long short-term memory (LSTM) framework. The visual and textual components of an e-mail were processed individually to use the long short-term and convolutional neural network models to generate two categorizations posterior distribution, which are then merged into a hybrid framework to estimate if the message is trash or otherwise. Researchers are using a grid search optimization algorithm to determine the most appropriate hyper-parameters for the MMA-MF designer's hyper-parameters and a k-fold cross-validation technique to assess the effectiveness of the algorithm. The findings of the experiments reveal that this approach outperforms typical spam detection algorithms, with accuracies ranging from 92.65 to 98.49 percent. Researchers believe using the novel methodology, as well as the one-class classification

algorithm and the few-shot pedagogical practices, to address the problem of the discrepancy between the number of spam and nonspam e-mails, and we will keep collecting extra realistic combined electronic mail data sources to enhance the system framework of the prototype so that it could function better at phishing detection [18].

Spam e-mail makes it difficult for users of e-mail accounts to receive critical data. e-mail spam prevention has indeed been implemented in public mail servers using a variety of ways. However, not all e-mail systems offer to spam e-mail significant findings for the usage of a limited amount of enterprise e-mail addresses. To safeguard e-mail users from junk, the system administrator should implement a distinct or module junk mail detection rate. The goal of this research is to find the most effective strategy for identifying spam e-mails. To determine the most effective technique for identifying spam e-mail, researchers used machine learning techniques such as decision tree, logistic regression, and random forest, and evaluated the findings. The speed of train and test procedures, and the reliability of spam message detection are used to determine effectiveness. According to the findings of this investigation, the random forest approach has an outstanding outcome, with a testing data velocity of 0.19 seconds and a 98 percent reliability. This finding could be utilized as a basis for the formulation of various algorithms for spam filtering. The limitation of the research is the researcher assumed that more specific algorithms, like the approximate solution and the database process, would be used to enhance efficiency [19].

Smart objects supply the preponderance of computational services near to the end customers for the next-generation Internet of things. These gadgets with built-in knowledge may make autonomous choices in the context where they are placed by applying diverse AI methodologies. In response to these challenges, researchers propose a cognitive incursion prevention mechanism that prevents brand loyalty is important from infiltrating the web address bar image data, hence maintaining the legitimacy of search engine result pages. By addressing three separate levels, namely, data collecting services, edge computing services, and cloud services, the proposed model provides ambient knowledge for web data filtering and identifies web spam. The goal is to identify harmful images. The average, image gradient, and volatility of a picture are first retrieved, and afterward, the received information is analyzed in the proposed framework. For the performance evaluation of the proposed method, deep learning techniques are being used. It achieved a 98.77 percent accuracy when tested on a real-time available dataset [1].

3. Methodology

The e-mail is classified by the classification model based on its content and other criteria. The procedure of feature extraction and selection is critical for the majority of classification issues. In the categorization process, characteristics are quite important. The correlation-oriented feature extraction (CFS) approach is utilized for features extracted in this research [20]. For effective classification performance, the CFS technique is used to extract the best characteristics from a collection of characteristics. In the suggested spam mail detection (SMD) method, a unique hybrid bagged method is provided to address the shortcomings of the current approach. The basic mechanism of e-mail classification is depicted in Figure 1.

The suggested spam mail identification structure is predicated on the machine learning approach's efficiency. Firstly, electronic mail information is gathered through a spam mail identifying scheme. The e-mail information gathered is unorganized and unfiltered. e-mail information must be preprocessed in terms of reducing operations and providing an exact outcome. To obtainnecessary information, the information is automatically by eliminating text content, stemmed, and term text categorization. Each categorization technique uses the database, which is arbitrarily partitioned into multiple sets. To assess the ultimate classifier performance, the bagging method integrates the categorization performance of the different machine learning techniques [21].

4. Preliminaries

The basics of the random forest and J48 decision tree algorithms are explained in this part.

4.1. Random Forest. The random forest technique implements bagging by building each tree in an ensemble of decision trees from the bootstrap sampling of data from the training dataset. The length of each random subset of texts is determined by repeating randomized systematic sampling till its bootstrapping sample equals the size of the image training dataset. Just a random selection number of features is examined while creating each decision component for each decision tree [22]. The yes/no criterion that effectively lowers the unpredictability of the information is chosen for the next node in the tree from the "k" randomly picked attributes to examine for creating each decision point and mentioned in

$$\text{Entropy} \coloneqq = \text{pro}(\text{Spam})\log_2\text{pro}(\text{Spam}) - \text{pro}(\text{Not Spam})\log_2\text{pro}(\text{Not Spam}). \quad (1)$$

Whenever the classification is undecided as to whether a text is a junk, the unpredictability is highest. Here is an instance of a simple spam detection random model mentioned in Algorithm 1.

Unlike many other decision tree optimization models, the random forest algorithm dataset is divided set iteratively when it is no longer effective to minimize the randomness of each leaf node (whether if all the model learns texts for a

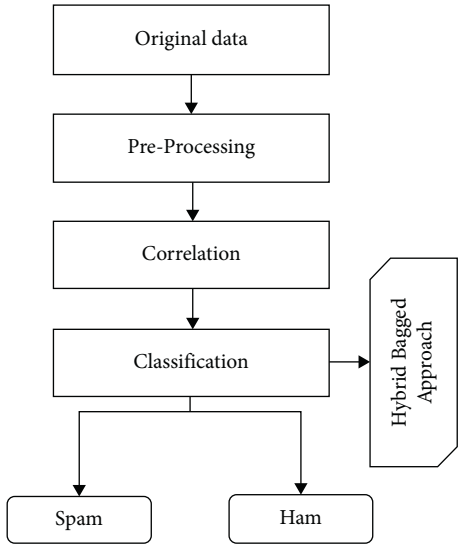


Figure 1: Common e-mail filter procedure.

```
if req("eliminate") >0.01 then
        reoccurrence "spam"
else
        if frq("free") >0.10 then
        else
        return "not spam"
```

Algorithm 1: Spam detection random model algorithm.

binary tree have the same categorization but because it is not able to successfully detach the spam e-mails from the nonspam texts). To use a voting approach, the full ensembles of decision nodes are then utilized to classify fresh communications [23].

Random forest is a higher-level variation of CART that uses the bootstrap bagging approach and random feature selection. In this approach, a forest is created from a large number of trees, which are subsequently analyzed to produce categorization input data [24]. The outcome consensus for each classification stage determines the outcome of categorization evaluation inputs, as shown in Figure 2.

4.2. Decision Tree. The J48 classification is predicated on the notion of probability and is a decision tree classification. It is a multiclass classifier that uses the training knowledge to generate decision trees. For the categorization of the new data element, the decision tree built utilizing J48 is based on the training examples feature values. J48 is based on the idea that dividing data into numerous sets allows any feature characteristic to be utilized to generate a decision [25]. The method performs in a nonlinear manner until every information characteristic is analyzed and classified; i.e., the characteristics retrieved using this method are the greatest feasible characteristics for the data category in question. The following are among some of the factors that the system takes into account:

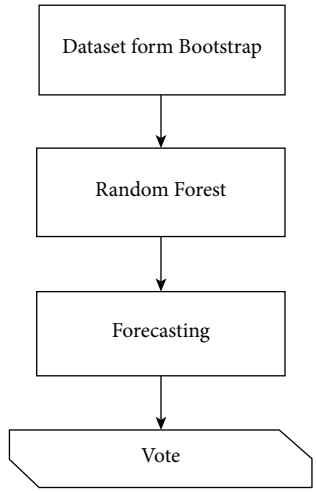


Figure 2: Mechanism process of RF.

(i) The technique creates judgment nodes higher in the tree whenever examples of originally thought unsupervised classification are detected

(ii) When the data collected correspond to a single class, the algorithm is used to generate a prediction model with a leaf node and requests that category be considered

(iii) If the characteristics or obtained features do not give any mutual information, a judgment node is generated just above the current tree using predicted values

The root of the tree, internal nodes, and leaf nodes make up a tree structure. Leaf nodes indicate the class, while nodes in the network indicate the constraints attached to characteristics and traits [26]. A sample decision tree is shown in Figure 3. For the J48 decision tree algorithm, Algorithm 2 is provided.

DT is the collection of training examples in Algorithm 2, and F is the decision tree. Splitting requirement is an attribute selection strategy in this method that divides the data objects into certain individual courses.

5. System Design

Figure 4 depicts a spam e-mail test method that employs machine learning. The first phase is data preparation, which includes data analysis and splitting. This approach sought to segregate half of the data into three categories: data training spam, data training nonspam (ham), and data testing. The data filtering process is the following step, which removes any extraneous words and expressions [27]. The investigator then constructed a model to reflect each technique that was discovered. The procedure of training and modeling testing, which has resulted in the acquisition of accuracy levels from spam e-mail classification, is the final phase. Aside from the accuracy number, another comparative criterion is the speed of the training phase when utilizing an existent database [28].

5.1. Framework of Spam Identification. The process of the spam mail detection (SMD) program for classifying e-mail into ham and spam messages is depicted in this part. The

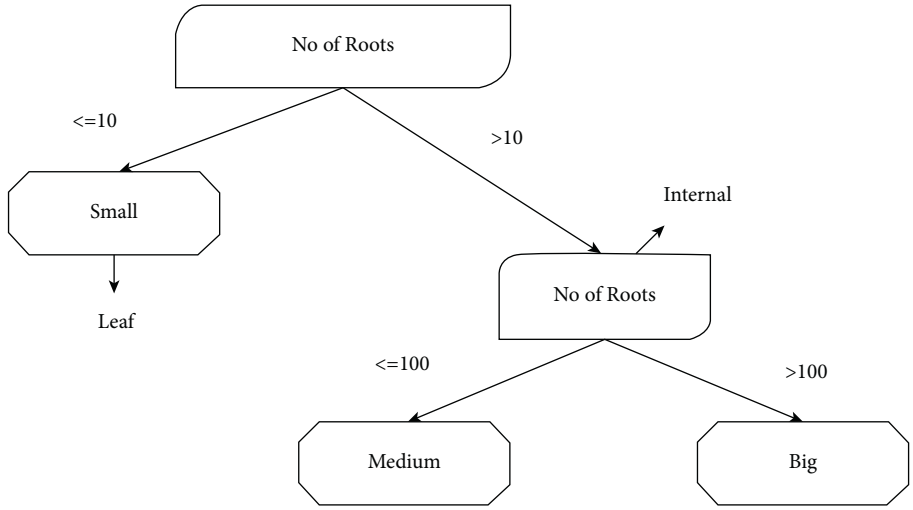


Figure 3: Decision tree structure.

```
build_decision_tree ( * DT)
{
assign root node F and using splitting trait label it
add arc (root node for F)
for each arc do
      DT =dataset developed by splitting trait to DT
      if stop measures met then
            Develop leaf node F′
            Label leaf node with suitable class
else
            F′ = build_decision_tree (DT)
F = add F′ to arc
}
```

Algorithm 2: Decision tree algorithm.

SMD network is comprised of strong division capabilities that were first established with the hybrid bagged method concept. The feature identification is executed using a correlation-based feature selection technique, and the analysis is performed using a unique hybrid bagging methodology [29].

The bagging strategy is a hybrid method in which the categorization aim is served by a decision tree-based J48 algorithm and random forest. Figure 5 shows the flowchart of the SMD system for text categorization. The SMD model divides the e-mails into two categories: spam and ham e--mails [30]. The text-based e-mail database is preprocessed to allow for effective extracting features. A hybrid bagged categorization strategy is being explored. e-mail datasheet preparation, preprocessing of information, selection of features, and hybrid bagged technique are the four modules of the SMD method. In the sub-section, a working prototype was also presented [31].

5.2. e-Mail Dataset. The database e-mail spamming code project machine learning and AI assert were primarily seen as the information training and testing data in this study. This database is provided to competition competitors as a reference for effectively detecting spam e-mails. Information is then classified as trash or nonspam intended to assist in the detection and verification of results using machine learning methods. Again, for the spam mail detection technique, an e-mail database is created [11]. From the Ling Spam database, various messages are chosen at random. For supervised classification, the database comprises a collection of 1000 e-mails, including both ham and spam e-mail messages. The database is separated into sets for each classification method because the organization procedure is a bagged technique. A total of two pairs of 500 e-mails have been generated. Each one of the random forest and decision tree algorithms uses 300 e-mails for training and 200 e-mails for testing. Table 1 displays the dataset's statistics.

5.3. Preprocessing. The message database under consideration is unprocessed. As a result, it must be preprocessed until being considered anymore. There are 3 phases in the preprocessing stage. The tokenization of the text information is measured first. Tokens are words that are separated from the rest of the phrase. Stop words are eliminated from the

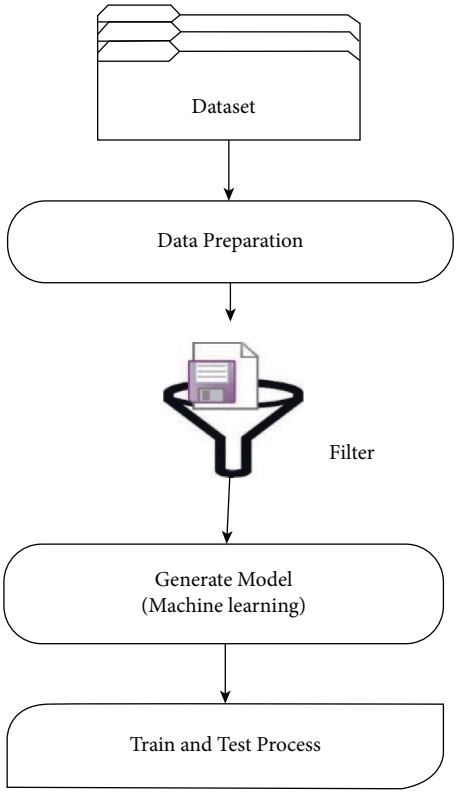


Figure 4: System design.

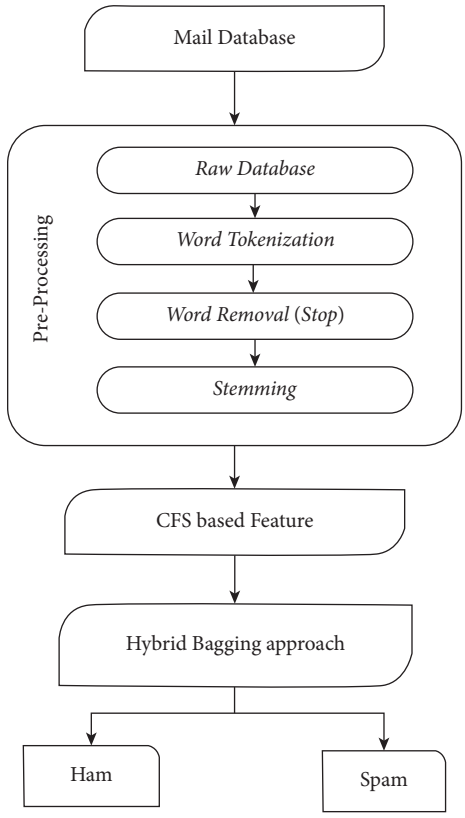


Figure 5: e-mail classification based on spam mail identification.

Table 1: Mail database.

Database	Random forest	Decision tree
Training ham mails	180	120
Training spam mails	130	170
Testing ham mails	150	50
Testing spam mails	110	90
Average mails	500	500
Overall	1000	

tokenized phrases. Inappropriate phrases with no linguistic meaning are known as stop words. During preprocessing, a document with around 670 stop words is routinely generated, and words are eliminated from the content. Stemming is the third stage in the preprocessing component. The stemming method removes a word from its root word. Stop word removal and stemming are key preprocessing processes since they significantly decrease the search area for effective extracting features [26].

5.4. Feature Selection. Feature selection is an important concern that has sparked a bit more articles. That has three objectives: (i) improving classification predictive performance, (ii) developing a better and more cost classifier, and (iii) gaining a greater understanding of fundamental processes related to information production. Two recommended strategies for reducing the feature set size are dimensionality reduction and relevant feature selection. Although relevant feature selection entails extracting a subset characteristic, image compression entails combining the original features and functionality linearly [32]. In every categorization system, features are quite significant. The SMD approach is based on the notion that spam mail contains different information than ham e-mail. The feature collection includes numerical keywords, languages, grammar or typographical problems, improper terms (words connected to product/service advertisements, dating, adult phrases, and so on), recurrence number, and document size, among other things. Correlation feature selection (CFS) is employed in the SMD technology. CFS simply selects the best characteristics from a range of options for increasing the overall system efficiency. "Good feature selections comprise characteristics correlated significantly with the categorization, but uncorrelated to one another," according to the correlation-based feature selection technique.

Text data with extracted features are primarily thought of as a bag of visuals. The term frequency technique is used to display the total number of terms in a material. All phrases are counted for recurrence, and those with a recurrence under a certain level are removed. The plan proves the words' utility while concurrently dipping the search area [33]. Utilizing a correlation-based feature selection strategy, the acquired set of features is even further decreased. The correlation-based feature selection approach chooses only the feature set that is most closely connected to the given class. Equation (2) offers the system of linear equations of the correlations-driven feature selection technique if c is the feature set with n number of features and a is the set of training.

$$\text{Correlation feature selection} = \max_{s_n}\left[\frac{r_{ac_1}, r_{ac_2}, r_{ac_3}\ldots r_{ac_n}}{\sqrt{n+2\left(r_{c_1c_2}+\ldots r_{c_ic_j}+\ldots r_{c_nc_1}\right)}}\right]. \tag{2}$$

Here, the average correlation feature class is denoted as r_{ac} and the average correlation feature-feature is denoted as r_{cc}.

5.5. Hybrid Bagging Technique. The classification design is the fourth and final system. For categorization, a hybrid bagged technique combining the decision tree-based J48 algorithm with the random forest is being examined. The bagged strategy, also known as the bootstrapping aggregating method, reduces variability by combining numerous repeating subsets of the same database. Multiple models have been created in this method by arbitrarily partitioning the e-mail database into two independent sample mail data sources: SD1 and SD2. Separate classifiers are trained using each instance of e-mail collection. The outcome of the entire process is the average of two categorization systems' results. For multiclass recognition and classification, the J48 method and random forest are utilized [34]. The mean of the anticipated values is used to determine the classification accuracy and the idea of bagging as depicted in Figure 6.

5.6. Working Process. Only with the assistance of the accompanying instance, a complete explanation of the components of the spam mail detection (SMD) process is achieved [35]. An instance of a randomized mail is used to demonstrate the spam mail identification program's step-by-step operation. As illustrated in Table 2, the SMD form includes an e-mail as input and provides spam or ham as an outcome.

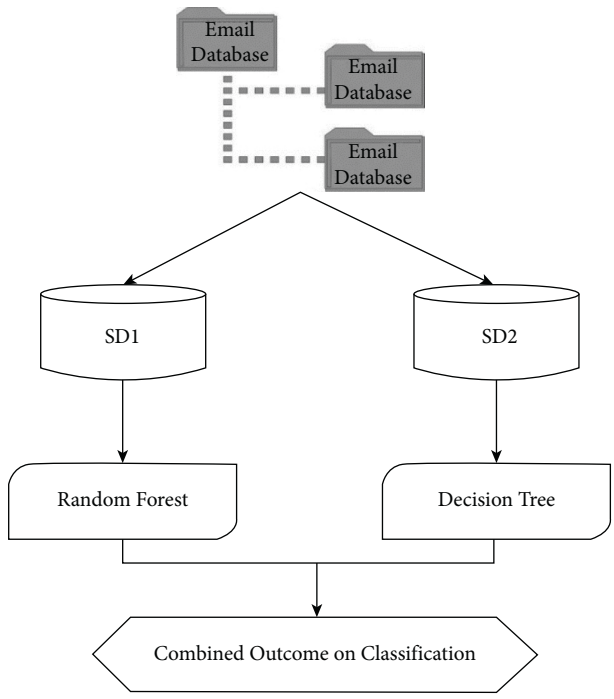


Figure 6: Bagging approach.

Table 2: Working process.

Parameter	Example
I/P	Subject: A new way to shop! Get newpass free for a year & enjoy benefits across brands! Continue to earn a minimum 5% Newcoins! Terms and conditions applied. Click here for more detail
Tokenization	"Subject" ":" "A" "new" "way" "to" "shop" "!" "Get" "newpass" "free" "for" "a" "year" "&" "enjoy" "benefits" "across" "brands" "!" "Continue" "to" "earn" "minimum" "5%" "Newcoins" "!" "Terms" "and" "condition" "applied" "." "Click" "here" "for" "more" "detail"
Stop word elimination	"new" "way" "shop" "Get" "newpass" "free" "year" "enjoy" "benefits" "across" "brands" "Continue" "earn" "minimum" "5%" "Newcoins" "Terms" "condition" "applied" "." "Click" "here" "more" "detail"
Stemming	"new" "way" "shop" "Get" "newpass" "free" "year" "enjoy" "benefits" "across" "brands" "Continue" "earn" "minimum" "5%" "Newcoins" "Terms" "condition" "applied" "." "Click" "here" "more" "detail"
Outcome	Spam mail

6. Results and Discussion

The obtained measurements of the spam mail detection (SMD) method are mentioned in this report. For the experiments, an e-mail database of 1000 e-mails is used, including 500 e-mails in each of the two classification methods [36]. Three tests were carried out in all, and the findings are evaluated. Two studies employing separate random forest classification algorithms, J48 decision tree algorithms, and a third investigation utilizing a hybrid bagged technique is done for spam e-mail identification. The RF method is a straightforward supervised learning technique that is simple to comprehend and execute. Even with an insignificant number of training trials, the method generates good results. However, the technique is based on the premise that the database contains separate class features. On either side decision, the tree-based method can deal with feature relationships, incomplete information, as well as other issues. However, decision tree algorithms struggle with data stream sets and the over-fitting problem. The finest of both methods were combined in the hybrid bagged technique of random forest and the J48 method [32].

The overall outcome of a spam mail recognition model is the combination of both models' forecasts, resulting in a system that is efficient and dependable. The efficiency characteristics are used to assess the effectiveness of the implemented method spam mail detection technique. To analyze the efficiency of the spam mail detection technique, measures such as accuracy, false-positive rate, recall, precision, true-negative rate, F-measure, and false-negative rate are computed. The effectiveness of the spam mail detection system is assessed using the criteria listed in Table 3.

The algorithm received an average accuracy of 95 percent, which is the average of the two classification systems' efficiency. Random forest classifier has an accuracy of 84 percent, with precision and recall values of 86 percent and 82 percent, correspondingly. The J48 method, on the other hand, achieves 92 percent accuracy, with precision and recall values of 94 percent and 90 percent, correspondingly. The assessed outcomes of the three trials are presented in Table 4:

Table 3: SMD evaluation measure.

Assessment parameter	Specification	Model
Precision	The efficacy of the classifier is defined by precision	$T_P/T_P + F_P$
Accuracy	The proportion of positive forecasted value to the overall set	$T_P + T_N/T_P + F_P + T_N + F_N$
Recall	The positively labeled information provided by the classification out of the entire data	$T_P/T_P + F_N$
F-score	Overall quality is demonstrated by the classifier's ability to produce efficient beneficial results.	$2 \times P.R/P + R$
True-negative rate (TN_R)	Spam mails managed to identify as a percentage of all spam mails.	$T_N/T_N + F_P$
False-negative rate (FN_R)	It detects the number of spam e-mails that have been missed.	$F_N/F_N + T_P$
False-positive rate (FP_R)	The number of spam e-mails mistakenly detected as a proportion of overall spam mails	$F_P/F_P + T_N$
True positive (T_P)	The sum of ham electronic mails that were accurately detected.	—
False negative (F_N)	The sum of ham mails that have been mistakenly classified as spam.	—
False positive (F_P)	The sum of spam messages that were mistakenly recognized as ham.	
True negative (T_N)	The sum of spam e-mails that were appropriately detected.-	—

Table 4: Analysis outcome (1).

Parameter	Random forest	Decision tree	Hybrid bagging
Accuracy	84	92	88
Precision	85	94	90
Recall	82	90	86
F-score	90	85	88

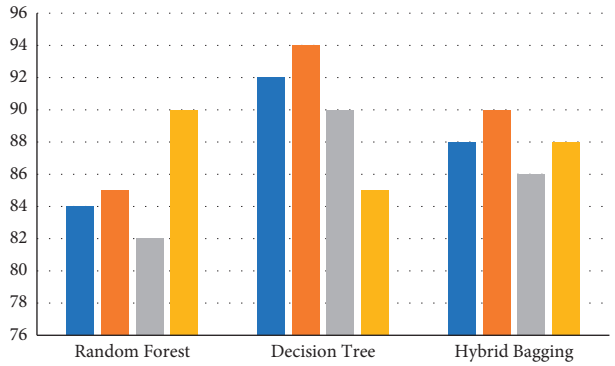


Figure 7: SMD parameter outcome on accuracy, precision, recall, and *F*-score.

random forest, J48 algorithm, and hybrid bagged technique, correspondingly. The graphical representation of Table 4 is presented in Figure 7, in which the graph is plotted for the analysis outcome of spam detection for three algorithms. The blue bar represents accuracy rate, orange bar represents the precision, gray represents the recall value, and the yellow bar represents the *F*-score.

Table 5 and Figure 8 show the comparative analysis of the true-positive (blue bar) and true-negative (gray bar) as well as the false-positive (orange bar) and false-negative (yellow bar) cases for the three algorithms. Based on the graph, it is verified that the true-positive case and the true-negative case have more performance than the others. Owing to this, Table 6 and Figure 9 show the comparative analysis of true-positive rate (blue bar), false-positive rate (orange bar), and false-negative rate (gray bar).

Table 5: Analysis outcome (2).

Parameter	Random forest	Decision tree	Hybrid bagging
True positive	82	90	86
False positive	15	5	11
True negative	87	93	84
False negative	20	10	14

By comparing the J48 decision tree method to the random forest and the hybrid bagged method, the findings in Tables 4–6 clearly show that the J48 decision tree algorithm achieves higher outcomes of precision, recall, and accuracy. Nevertheless, in the instance of random forest (90 percent), the proportion value of the F-measure is larger than that in either J48 (85 percent) or the hybrid bagged technique (90 percent) (88 percent). The graphical depiction of the contrast of the SMD state's outcome and the related

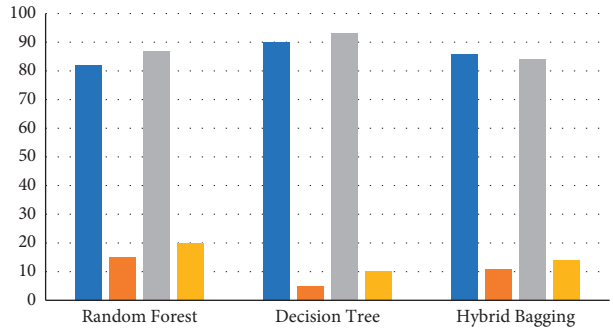


Figure 8: Analysis outcome of true and false positive and negative.

Table 6: Analysis outcome (3).

Parameter	Random forest	Decision tree	Hybrid bagging
True-positive rate	87	95	89
False-positive rate	20	10	16
False-negative rate	15	5	12

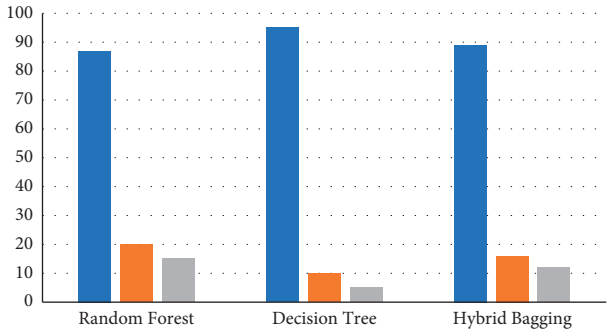


Figure 9: Analysis outcome of TN_R, FN_R, and FP_R.

classification methods independently is shown in Figures 7–9.

7. Conclusion

Today, spammers are among the most demanding and unpleasant issues related to communication and information technologies. Paid trolls abuse this communications device by sending spam e-mails, which has a negative impact on productions and numerous Internet consumers. This research presents a spam mail detection mechanism that uses a hybrid bagging technique for execution. Random forest and decision tree (J48) are the categorization techniques employed in this technique. The hybrid bagging method-based spam mail detection system attained an overall rate of 95 percent, indicating that the testing findings are superior when using simply the J48 method. The idea of enhancing technique could have been used for future studies to improve the system's effectiveness. The enhancing strategy substitutes the weak classifier's learning features with those of the classification model, enhancing overall design competence. In future consideration, the researcher assumed that even more sophisticated techniques, like the evolutionary algorithm and the dataset procedure, will be more widely used to enhance effectiveness.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

The author thanks the Deanship of Scientific Research at Jouf University for funding this work through research grant no. DSR2020-06-3680.

References

[1] A. J. Karim, Shanmugam, Azam, Kannoorpatti, Jonkman, and Boer, "An intelligent spam detection model based on artificial immune system," *Information*, vol. 10, no. 6, p. 209, Jun. 2019.

[2] E. G. Dada, J. S. Bassi, H. Chiroma, A. O. Abdulhamid, O. E. Adetunmbi, and Ajibuwa, "Machine learning for email spam filtering: review, approaches and open research problems," *Heliyon*, vol. 5, no. 6, Article ID e01802, 2019.

[3] G. Hnini, J. Riffi, M. A. Mahraz, A. Yahyaouy, and H. Tairi, "MMPC-RF: a deep multimodal feature-level fusion architecture for hybrid spam E-mail detection," *Applied Sciences*, vol. 11, no. 24, p. 11968, 11968, Dec.

[4] R. Talaei Pashiri, Y. Rostami, and M. Mahrami, "Spam detection through feature selection using artificial neural network and sine-cosine algorithm," *Mathematical Sciences*, vol. 14, no. 3, pp. 193–199, 2020.

[5] S. O. Olatunji, "Improved email spam detection model based on support vector machines," *Neural Computing & Applications*, vol. 31, no. 3, pp. 691–699, Mar. 2019.

[6] I. Ahmed, "Taloba "an artificial neural network mechanism for optimizing the water treatment process and desalination process," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9287–9295, 2022.

[7] M. Elloumi, M. A. Ahmad, A. H. Samak, A. M. Al-Sharafi, D. Kihara, and A. I. Taloba, "Error correction algorithms in non-null aspheric testing next generation sequencing data,"

Alexandria Engineering Journal, vol. 61, no. 12, pp. 9819–9829, 2022.

[8] J. Rastenis, S. Ramanauskaitė, J. Janulevičius, A. Čenys, A. Slotkienė, and K. Pakrijauskas, "E-mail-Based phishing attack taxonomy," *Applied Sciences*, vol. 10, no. 7, p. 2363, 2020.

[9] S. S. I. Ismail, R. F. Mansour, A. b. d. El-Aziz, M. Rasha, and A. I. Taloba, "Efficient E-mail spam detection strategy using genetic decision tree processing with NLP features," *Computational Intelligence and Neuroscience*, vol. 2022, 2022.

[10] A. I. Taloba, M. R. AbdEl-Aziz, and A. A. H. El-Bagoury, "Estimation and prediction of hospitalization and medical care costs using regression in machine learning," *Journal of Healthcare Engineering*, vol. 2022, 2022.

[11] D. J. Kerin and R. A. Kerin, "The effects of personalized product recommendations on advertisement response rates: the "try this. It works!" technique," *Journal of Consumer Psychology*, vol. 14, no. 3, pp. 271–279, 2004.

[12] J. A. Lim and K. Lim, "Identifying machine learning techniques for classification of target advertising," *ICT Express*, vol. 6, no. 3, pp. 175–180, 2020.

[13] G. Jain, M. Sharma, and B. Agarwal, "Optimizing semantic LSTM for spam detection," *International Journal of Information Technology*, vol. 11, no. 2, pp. 239–250, 2019.

[14] A. S. Rassam and M. A. Rassam, "A combined text-based and metadata-based deep-learning framework for the detection of spam accounts on the social media platform twitter," *Processes*, vol. 10, no. 3, p. 439, 2022.

[15] A. I. Taloba, R. Alanazi, O. R. Shahin et al., "Machine algorithm for heartbeat monitoring and arrhythmia detection based on ECG systems," *Computational Intelligence and Neuroscience*, vol. 2021, 2021.

[16] A., O. R. S. El-Komy, m Rasha, M. A. El Aziz, and A. I. Taloba, "Integration of computer vision and natural language processing in multimedia robotics application," *Information Sciences Letters*, vol. 7, p. 6, 2022.

[17] T. Hong, J. A. Choi, K. Lim, and P. Kim, "Enhancing personalized ads using interest category classification of SNS users based on deep neural networks," *Sensors*, vol. 21, no. 1, p. 199, 2020.

[18] H. Yang, Q. Liu, S. Zhou, and Y. Luo, "A spam filtering method based on multi-modal fusion," *Applied Sciences*, vol. 9, no. 6, p. 1152, 2019.

[19] B. Santoso, "An analysis of spam email detection performance assessment using machine learning," *Jurnal Online Informatika*, vol. 4, no. 1, pp. 53–56, 2019.

[20] T. S. Caminhas and W. M. Caminhas, "A review of machine learning approaches to spam filtering," *Expert Systems with Applications*, vol. 36, no. 7, pp. 10206–10222, 2009.

[21] G. Sahin and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.

[22] D. Wechsler and H. Wechsler, "Spam detection using random boost," *Pattern Recognition Letters*, vol. 33, no. 10, pp. 1237–1244, 2012.

[23] B. Abdulazeez and A. Abdulazeez, "Classification based on decision tree algorithm for machine learning," *Journal of Applied Science and Technology Trends*, vol. 2, no. 01, pp. 20–28, 2021.

[24] B. Zaman, A. Justitia, K. N. Sani, and E. Purwanti, "An Indonesian hoax news detection system using reader feedback and naïve Bayes algorithm," *Cybernetics and Information Technologies*, vol. 20, no. 1, pp. 82–94, 2020.

[25] Y. Zhang, S. Wang, and L. Wu, "Spam detection via feature selection and decision tree," *Advanced Science Letters*, vol. 5, no. 2, pp. 726–730, 2012.

[26] H. Shen, X. Liu, and X. Zhang, "Boosting social spam detection via attention mechanisms on twitter," *Electronics*, vol. 11, no. 7, p. 1129, 1129, Apr.

[27] A. I. Taloba and S. I. I. Safaa, "An intelligent hybrid technique of decision tree and genetic algorithm for e-mail spam detection," in *Proceedings of the 2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 99–104, IEEE, Cairo, Egypt, 2019.

[28] M. Z. Gashti, "Detection of spam email by combining harmony search algorithm and decision tree," *Engineering, Technology & Applied Science Research*, vol. 7, no. 3, pp. 1713–1718, 2017.

[29] A. Zainab, S. S. Refaat, and O. Bouhali, "Ensemble-based spam detection in smart home IoT devices time series data using machine learning techniques," *Information*, vol. 11, no. 7, p. 344, 2020.

[30] A. I. Taloba and M. R. Riad, "Developing an efficient spectral clustering algorithm on large scale graphs in spark," in *Proceedings of the Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 292–298, IEEE, Cairo, Egypt, December 2017.

[31] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, "Survey of review spam detection using machine learning techniques," *J. Big Data*, vol. 2, no. 1, pp. 1–24, 2015.

[32] N. Hussain, H. Turab Mirza, G. Rasool, I. Hussain, and M. Kaleem, "Spam review detection techniques: a systematic literature review," *Applied Sciences*, vol. 9, no. 5, p. 987, 2019.

[33] M. Iqbal, M. M. Abid, M. Abid, F. Ahmad, and Khurshid, "Study on the effectiveness of spam detection technologies," *International Journal of Information Technology and Computer Science*, vol. 8, no. 1, pp. 11–21, 2016.

[34] Q. Abu Al-Haija and M. Al-Dala'ien, "ELBA-IoT: an ensemble learning model for botnet attack detection in IoT networks," *Journal of Sensor and Actuator Networks*, vol. 11, no. 1, p. 18, Mar. 2022.

[35] Z. Han, S. Li, and H. Liu, "Composite learning sliding mode synchronization of chaotic fractional-order neural networks," *Journal of Advanced Research*, vol. 25, pp. 87–96, 2020.

[36] S. Ha, L. Chen, and H. Liu, "Command filtered adaptive neural network synchronization control of fractional-order chaotic systems subject to unknown dead zones," *Journal of the Franklin Institute*, vol. 358, no. 7, pp. 3376–3402, 2021.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 5086290, 13 pages
https://doi.org/10.1155/2022/5086290

Research Article

Fast and Intelligent Irrigation System Based on WSN

G. Oussama[1], **A. Rami**[1], **F. Tarek**[2], **Ahmed S. Alanazi**[1], **and M. Abid**[2]

[1]*Department of Computer Science, Jouf University, Al Qurayyat, Saudi Arabia*
[2]*CES-Lab, Sfax University, Sfax, Tunisia*

Correspondence should be addressed to G. Oussama; oaghorbel@ju.edu.sa

Received 19 May 2022; Accepted 18 June 2022; Published 14 July 2022

Academic Editor: Heng Liu

In the agricultural industry, wireless sensor networks (WSNs) can be an important tool to promote economic growth. Using network devices in agriculture has the potential to enhance the production process. One of the key challenges WSN faces is energy efficiency. A model based on the water pipeline method is proposed in order to efficiently utilize sensor nodes in agricultural production and water distribution. A water pipeline serves as an important structure for transporting potable water across a distance for consumption or irrigation. In contrast, the biggest transportation problems of water pipelines are leaks. So, water resources may be lost as a result. These pipes need real-time monitoring to prevent such problems. The wireless sensor network technique, however, is considered one of the best solutions available today for monitoring water pipelines. A detailed analysis of agriculture is provided by the model. Aspects of WSN are discussed and their agricultural use is expounded. Moreover, this paper describes the various types of applicability of existing sensor networks in the field of agriculture, along with some technical perspectives. To achieve the best power consumption and communication for the two types of range, ZigBee wireless protocols are utilized. As such, high-performance information that provides a platform for WSNs to better support agricultural production is also included in our proposed model in order to address the shortcomings of existing WSNs regarding energy efficiency. This paper presents the improvements of the proposed solution compared to other techniques in the context of energy conservation in wireless sensor networks and in the monitoring water-saving process.

1. Introduction

Most countries around the world have always seen agriculture as an important industry in order to build their economies. A phenomenon such as farmland pollution, the degradation of the environment, and biodiversity destruction has been described as a result of excessive fertilization in agricultural production efforts [1]. Producing, growing, and developing the agriculture industry are heavily impacted by these factors. A good administration of agricultural production activities is essential to the sustainable development of modern agriculture. Both need to be taken into consideration in order to develop a scientific production technology. The application of WSN for energy efficiency can have a profound impact on agriculture. Management strategies for agriculture can greatly improve energy efficiency. Agricultural production is improved by utilizing information technology in a reasonable and continuous way. It also reduces pollution levels by using a variety of resources. Energy and water are part of the used resources. The production of agricultural products becomes more advanced and more accurate.

Human survival requires access to water. Approximately 96.6% of the water on Earth comes from the ocean. Of that, approximately 2.5% is fresh and can be used for agriculture. A vast amount of fresh water is transferred worldwide via millions of kilometers of pipelines at present.

According to this work [2], a vast network of underground pipelines has been laid to connect Saudi Arabia's cities with the Arabian Gulf. A pipeline can be used for irrigation, making it easier for farmers to water their land. World Bank estimates that more than 45% of freshwater produced newly is lost in nonrevenue water (NRW) in third world countries. A recent study [3] reported the NRW indicator of real loss in terms of liters/connection/day in the world as described in Figure 1.

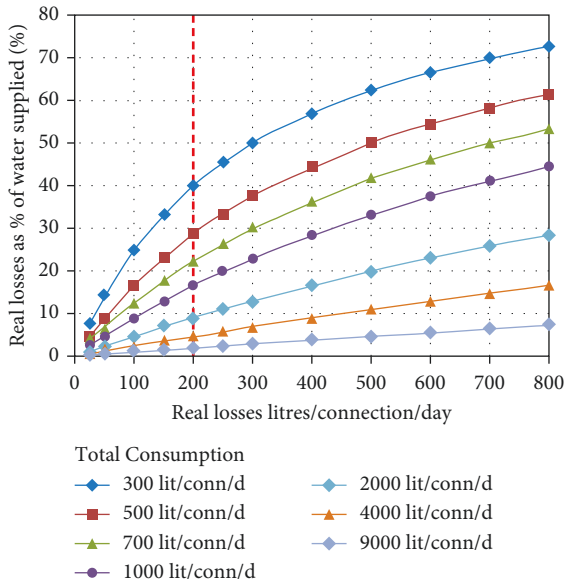


Figure 1: The NRW indicator of real water loss.

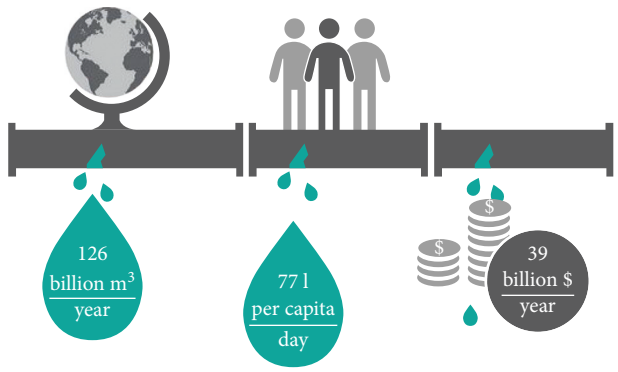


Figure 2: The wasted water around the globe.

It is expected that a pipeline with a large diameter will leak at least once per year, according to statistical analysis. Several environmental factors may cause damage to pipelines, including extreme soil conditions, aging of materials, bedding, and excessive loads. Excessive costs for hydraulic leaks can lead to infrastructure damage and environmental pollution. Thus, maintaining the pipeline infrastructure and ensuring its security is one of the main concerns. Based on Figure 2, the wasted water in the globe is calculated as 126 billion m^3/per year, 77 liters per day with an annual amount of 39 billion dollars.

Traditionally, pipeline monitoring has been researched through architecture, sensing devices, and modules. For example, the authors of this work [4] assessed the impact of network architecture on network consistency. Recent studies and advanced works on this topic are reviewed in [5] by reviewing WSN technology applied to the supervision of water pipelines as well as underwater, aboveground, and underground pipelines. Also, in their study [6], they described and compared the characteristics of microcontrollers used to monitor water pipelines. Research in this work proves useful in helping to select the best existing systems and technologies, compared to work in other described works.

In Saudi Arabia, agricultural development has yet to reach its full potential and continues to face many challenges. The use of irrigation and fertilizers should be fully regulated and monitored using modern scientific methods. A proper monitoring and control process should also be in place [7, 8]. Agricultural research is still in its early stages, and there is a number of key technical challenges to overcome. Chemical fertilizers and irrigation, for example, can be used to make food production much better. A lack of standardization research has led to the product not reaching market scale because sensors are too diverse and expensive to be used in agriculture to overcome these challenges. Agricultural sensors have many limitations at present: poor irrigation techniques, poor working conditions, and short battery life. Additionally, since the farming production zone is located in a remote area, the public telecommunications network is inadequate and sometimes costly. Also, interference issues and spectrum compatibility make communication difficult. Agricultural applications employ WSNs for a range of activities, including measuring temperature, monitoring the environment, monitoring irrigation systems, and measuring water supply. Farmers can generate high-quality crops through the use of WSNs. To power sensor nodes, a battery power supply is required. So, as a result of these networks, crops are more expertly produced and the economy is benefited.

Several recent studies have found that the use of these networks will solve a lot of obstacles in agriculture implementation. This latter utilizing wireless sensor networks is a highly efficient method for improving crop yield, lowering farmer burdens, and improving productivity. It is important to use WSN for agriculture to ensure the healthy growth of crops [9, 10]. By using this method, weeds and pests can be effectively controlled and sustainable agriculture is achieved. With the use of the WSN, we can collect and sense all types of data in real time in the agricultural production process which will provide users with timely feedback. Users are informed of the results of data analysis and processing to ensure efficiency in agriculture management.

An analysis of communication protocol development potential in WSN for agricultural production is presented in this paper. This article focuses on the characteristics of WSNs and how they can be applied to agriculture. It outlines crucial factors that influence monitoring models. Additionally, it discusses some technical prospects that can be helpful in improving agriculture's overall development level effectively and comprehensively. Besides, this study explains how existing sensor networks are applied to agricultural production technology and their limitations. The organization parts of this paper are presented as follows. The state-of-the-art studies about our work, in Section 2, are described. Then, our proposed Smart Water Irrigation System (SWIS) is discussed in Section 3. Also, we demonstrate our monitoring of diseases in Section 4. Finally, simulation results are provided in Section 5 and followed by concluding remarks in Section 6.

2. Related Work

Battery-powered sensor nodes are used as an energy supply in the evolution of technology that is deploying rapidly in WSNs [11, 12]. This latter is formed by detached and dedicated sensors that observe, record, and transmit information about the surrounding environment to a central hub. A detailed representation of the WSN architecture and structure is given in Figure 3. The battery-powered sensor network includes RF modules, processors, and sensor nodes.

Communication is possible between the sensor node and the base station or coordinator node via a wireless communication link. Each node in a network collects, computes, and communicates data and information with the associated nodes. Information like humidity level, pressure, and temperature collected from sensor nodes can range from simple to complex. Then combined with each other, they achieve real-time monitoring in WSN. Accordingly, the sensor nodes monitor, store, and process data during the entire monitoring process [13]. Information will be sent to the WSN in order to perform real-time monitoring. The use of WSNs in farming has a wide range of benefits in terms of obtaining accurate and sufficient environmental information [14]. So, the sink node, base station, and sensor node are considered the three main components of WSN [15].

Nodes on the sensor network can communicate and compute. Wireless connections are made using short-range wireless technology, which can create multihop wireless networks.

It is possible to monitor fully the crop data to obtain general agricultural information and support agriculture development activities through the use of wireless sensor networks [16]. As wireless technology itself, WSNs are simple and of low cost and low power. It adapts to the agricultural production environment and is stable over time. Integrating modern agricultural characteristics into data acquisition algorithms can significantly increase data collection rates and bring on a better agricultural production. For better quality of service (QoS) and more efficient farming production, WSNs are being widely used in agriculture right now. The sensors are being used for collecting various information types in agriculture immediately (e.g., carbon dioxide levels, humidity, etc.). In addition to supporting agriculture and improving yields, the WSN can also be utilized for other agricultural applications. Using soil nutrient data, it is possible to predict crop health and product quality [17]. Also, you can use it to predict irrigation needs as well as soil moisture levels. The network can be expanded by adding relevant sensor nodes to improve parameter agricultural monitoring. Although WSNs have some applications in agriculture, some challenges remain, like deriving the optimal deployment scheme, communication range, and others [18]. It takes a long time to collect data from decentralized sensor nodes, and signal attenuation can weaken or destroy the communication link.

Power consumption is identified as a major problem in wireless sensor networks and, also, battery life extension because these nodes are powered by batteries, which have a limited amount of power. Even though WSN has been gaining popularity for several years, its application has some limitations

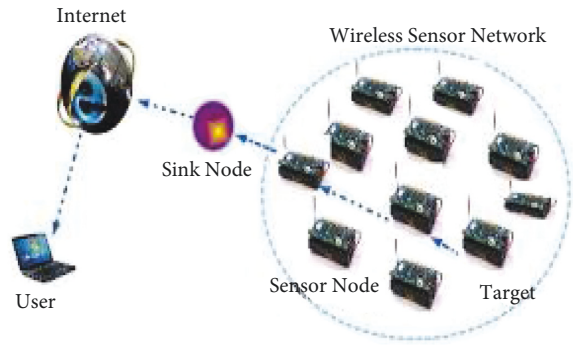


Figure 3: Wireless sensor networks architecture.

[19]. For example, to eliminate the problem of sending data from farmland to the base station, a mobile data connection service is used. A base station located in a farmland region can communicate wireless sensor data to the sensor nodes, which enables the wireless sensor battery to be buffered. The power consumption of WSNs is extremely low. A design that has low power consumption, wide coverage, multiple connections, and low cost is adopted. WSN is extensively used in agricultural production to advance the quality of crops in the field of agriculture, where it plays an important role in monitoring the agriculture field [20]. With WSN technology rapidly developing, battery power is still a major challenge.

In another part, we have also investigated the concept of a wireless sensor network to detect leaks. A multimodal system was usually considered to detect leaks in most studies. Using humidity and temperature sensors, Sadeghioon et al. [21] detected leaks. Data was collected for a few days from the humidity sensor and the temperature sensor. Analyzing the pressure profile and temperature data, they concluded that there was a leak. Studying wireless sensor network standards was the focus of this work [21]. To determine where the leak is, they used five flow sensors. To find a leak and its location [22], the authors used a variety of sensors such as humidity, temperature, pressure, and a gas detection sensor in the same way as [23]. Furthermore, they demonstrated the wireless data transfer between Arduino and XBee. They calculated the optimal placement of flow sensors as seen in [24, 25]. Our system was built using these analyses. In the next section, we will discuss the power management used in irrigation when based on the use of water pipelines as well as different materials, monitoring techniques, data used, etc.

3. Different Monitoring Models

3.1. Sensors Types. A pipeline monitoring system is typically composed of three kinds of sensors, the most commonly used of which are as follows:

(i) First model: A static sensor is usually attached to the pipeline surface and reads the actual state of the pipeline [26].

(ii) Second model Mobile sensor. The cells on this model travel from a source point to a sink point inside the pipe. Sensor data is copied from its

memory to the backend system. Sensor observations, as well as location information, are provided in this information [27].

(iii) Third model entitled Automotive Vehicles: Robots can either repair the damaged segment or gather information from it, but in most cases, static sensors capture information from pipelines, whereas robots or vehicles just collect the information.

3.2. Data Exploitation Types. Various methods of pipeline monitoring are used to carry out the monitoring process, and they include the following:

(i) The first method named centralized exploitation: the collected data is sent to the sink for processing, control, and alerting events when problems occur. Compared to other, a centralized algorithm is more accurate in providing location information. The aggregated data can also be analyzed to find the most interesting trends. However, its high bandwidth and power consumption make it expensive [28].

(ii) The second method named distributed exploitation: it was designed to distribute the monitoring area into smaller sections in order to reduce the loading on nodes. Monitoring with this model, it is possible to locate faulty pipeline areas more quickly and begin the maintenance process right away. It aims to enhance the lifetime of the WSN by transmitting just the important data to a central unit. As a result, there are a lot of messages exchanged, which makes distributed exploitation energy-intensive [29].

3.3. Sensor's Coverage. Diverse sensor topologies and implementation methods are used. Monitoring pipeline length is the main objective of sensor deployment. We will examine the sensors' range and direction of communication in this section. There are two categories of sensors coverage, which are as follows:

(i) Horizontal coverage: some vehicles move at different angles, inside or outside of the pipeline, to improve monitoring coverage. They are known as autonomous underwater vehicles (AUVs). A vehicle collects information on its own or gathers sensor data in a horizontal fashion in order to send the information vertically to the subsequent sensors. This approach is used in some systems such as [30].

(ii) Vertical coverage: it is achieved by implanting the motes along the monitoring pipeline, and they communicate linearly. Taking network operations centers into account, it can be found at the end of a pipeline or at both ends [31].

3.4. Power Management. Pipeline monitoring systems are sensitive to the power supply. Throughout the lifetime of a pipeline, which will last for many years, it needs to be monitored. In addition, the system must be durable.

Various methods have been developed for this proposal, including data filtering and aggregation, solar cells, and routing protocol management [32].

4. Monitoring Techniques for Water Pipeline

Over the years, methods of measuring total losses including actual and apparent leakages have been investigated. It is, therefore, necessary that new technologies are required to detect, locate, and estimate leak sizes.

4.1. Hardware-Based Methods. Through the use of specially designed accurate devices, these techniques can detect and locate leaks. Based on the nature of the detection equipment, hardware-based methods are categorized into five categories.

4.1.1. Visual Inspection Technique. A leak can be detected this way traditionally. Personnel who are experienced in this technique are also required when inspecting prospective leaks. Using video, drones, and others to capture images of pipelines ground, visual inspections can reliably detect bursts and other problems along the pipelines. The software provides user-friendly real-time three-dimensional (3D) images [26]. Underground pipelines are not adequately supervised with this technique.

4.1.2. Fiber Optic Sensing (FOS). In order to provide information about the circumference of a pipeline, fiber optic cables are buried directly beneath the pipe and/or helically instrumented around it. Leak detection can be done using this method by analyzing the temperature change caused by water leaks [4, 32]. Due to its resistance to interference, this technique has the significant advantage of being unaffected by electromagnetic waves. Because the method requires an additional excavation for installation, it cannot be implemented in underground pipelines.

4.1.3. Ground-Penetrating Radar (GPR). The scanning of the ground surface is done using this technique. A buried object's size and shape can be determined with these drawings. It is possible to deploy some miles of ground-penetrating radar per day with ground-penetrating radar (GPR) [33]. So, investigation and time-consuming represent the costly drawbacks of this technique.

4.1.4. Soil Properties (SP). In order to detect leaking pipeline branches, the transported water such as well as soil temperature must be abnormal in value [34]. Explicitly uncovered pipelines without soil surrounding them cannot be treated with this method.

4.1.5. Acoustic Detection (AD). Excessive use of junctions in underground pipelines can result in leakage occurrences. The entire pipeline network is therefore usually monitored

using an acoustic sensor. Its main benefit is the ability to detect damage indefinitely [24].

The disadvantages of this famous technique include its high cost, which necessitates the installation of many nodes in the monitoring area, which limits its application.

4.2. Software-Based Methods. By constantly monitoring various factors of the pipeline, like pressure, temperature, and so on, software-based techniques are primarily aimed at detecting and localizing leaks. The methods based on software can further be categorized into five groups.

(i) Negative Pressure Wave method (NPW): It is a method used for identifying leaks in a large space using stationary running strategies. Various parameters affect this technique including wave speed, noise from the device, and pressure changes.

(ii) Mass Balance method (MB): It controls the difference between the upstream and downstream flow measurements within a specified interval to detect damage.

(iii) Inverse Transient Analysis (ITA): A comparison is established between the signal reflected by monitoring devices and the signal observed in normal system operation.

(iv) Pressure Point Analysis (PPA): Various pressure measurements need to be taken along the pipeline on a continuous basis. Detection of small leaks will be easy.

(v) Real-Time Transient Modeling (RTTM): It provides a fast rate of detection by continuously analyzing the pipe configuration and the product characteristics at the same time.

4.3. Monitoring Water Pipelines: Methods Evaluation. In Tables 1 and 2, we summarize the key features of spoiled detection systems for every technique so that we can compare their performance. According to the studies referred to previously, these results are based on confident research. In Table 1, the utility of implementation and positioning of sensors are presented for each technique. Furthermore, in Table 2, the techniques' accuracy is also evaluated by calculating the percent of damaged objects identified perfectly. In addition to accuracy, there are four other key characteristics to assess. The first one is sensibility, which represents the capacity to detect small leaks. The second one is named the availability, which means the ability to conduct continuous monitoring 24 hours a day. The third type is the false alarm. This latter involves fewer false alarms in a used good technique. Finally, detection speed is defined as high when the detection time is up to a few minutes. It is then classified as medium when it is discovered within a few hours and as low for longer periods.

The implementation characteristics are also compared in Table 3, considering factors such as cost, retrofitting, complexity, and maintenance. The supported function is mentioned by (+), and the not supported function is mentioned by (/).

Table 1: Architectural damage monitoring methods evaluation.

	Methods	Detection	Localization	Sensors positions	
				Inside	Outside
Hardware	VI	+	+	—	+
	FOS	+	+	—	+
	GBR	+	+	—	+
	SP	+	+	—	+
	AD	+	+	+	—
Software	NPW	+	+	+	—
	MB	+	—	+	—
	ITA	+	+	+	—
	PPA	+	—	+	—
	RTTM	+	+	+	—

All of these tree tables mention that for all the key factors, tools that are rated "good" cannot be found. So, hardware-based models, ground-penetrating radar, and optical fibers cannot be used because of their limitations and they are expensive. Also, software-based techniques cannot be used because they cannot estimate the leak location perfectly and are unreliable in detecting damages [22]. The progress in recent years has been considerable; however, there is still much scope for improvement, especially in terms of real-time modeling and detecting pipe damages. In essence, the purpose is to provide a system of heterogeneous wireless sensor nodes that can detect leaks in water pipelines and monitor them with different nodes of different architectures. High safety levels, high localization precision, and real-time detection are all characteristics of these technologies.

5. Water Pipeline Monitoring System Based on WSNs

According to the communications technology used, monitoring water pipelines can be categorized into several categories. So, supervisory techniques include manual techniques, wired systems, SCADA systems, and wireless sensor networks.

Table 3 outlines a general assessment of each technique. As an additional comparison, Table 4 presents some of their key features. It takes into account a variety of factors, including flexibility, autonomy, security, and cost. In the table, the (/) indicates that there is not enough information to draw any conclusions.

Manual networks and WSNs both demonstrate their versatility with regard to extending or deleting models without network problems in Tables 3 and 4. Then, SCADA and wired networks show low levels of flexibility. Moreover, SCADA and wired systems are very expensive to install and maintain, while manual inspection is expensive, but WSN systems are even cheaper. The application based on WSNs is more accurate, of lower cost, and easy to set up than traditional methods.

Table 2: Damage monitoring techniques evaluation.

	Methods	Cost	Accuracy	Sensibility	Size estimation	Availability
Hardware	VI	H	H	+	—	—
	FOS	H	H	+	—	—
	GBR	L	L	+	—	—
	SP	L	L	+	—	—
	AD	M	H	+	—	+
Software	NPW	L	M	+	+	+
	MB	L	L	—	+	+
	ITA	L	L	—	—	+
	PPA	L	H	+	+	+
	RTTM	M	H	—	+	—

NB: H = high, M = medium, L = low.

Table 3: General assessment of different techniques.

	Advantages	Disadvantages
Manual technique Wu et al. [27] Jawhar et al. [23]	Need a real-time information	Not applicable to underwater applications Complex & costly
Wired technique Wong et al. [35]	Faster & secure compared to a wireless network	Rapidly technique failure Hard access to pipeline region
SCADA Cheng et al. [30] Enache et al. [36]	Rapid installation Adaptable for application with short range	Lower sense ranging Single point of system failure
WSNs Kartakis et al. [37] Cai et al. [38]	Portable & scalable Low corrosion	High conception in energy Sensitive & not reliable

Table 4: Monitoring techniques evaluation-based communication.

	Used by:	Cost	Security	Flexibility	Autonomy	Complexity
Manual technique	Wu et al. [27] Jawhar et al. [23]	M	L	H	L	—
Wired technique	Wong et al. [35]	H	M	L	M	M
SCADA	Cheng et al. [30] Enache et al. [36]	H	H	L	H	H
WSNs	Kartakis et al. [37] Cai et al. [38]	L	H	H	H	—

NB: H = high, M = medium, L = low.

5.1. Wireless Sensor Networks (WSNs). Science and industry have become more interested in wireless sensor networks. Wireless development, microelectronics, energy conservation, and microelectromechanical systems (MEMSs) have all been impacted by the development of this innovation over the last few years. There are many sensor nodes in a wireless sensor network, which are constantly monitored. A wireless connection is used to connect these wirelessly embedded nodes. Their joint effort collects large quantities of accurate information about many locations. Data is then processed and sent to sink points, also called gateway nodes. Their efficiency has been demonstrated in a variety of domains. An information structure transforms physical sensor information into general electronic measurements that can then be exploited by the wireless sensor node.

Sensor nodes consist of a variety of units, such as sensing, processing, and communication. In a measurement network, in general, the power consumption is managed via batteries, which is an important constraint. Moreover, it is advisable to implement sensors in an elegant manner to reduce communication costs and extend the lifespan of the network.

5.2. Current Work. Monitoring systems have been implemented using WSNs in several projects. Our analysis summarizes the characteristics of damage detection systems according to each of the different techniques in order to compare their performance. Data exploitation and placement pipeline depend on sensor coverage, parameters, and sensor types. Pipeline monitoring applications implemented in existing systems use a centralized monitoring system like [39] or [21] in which a powerful node is present. A large number of sensors used in this type of system create scalability problems. Some projects, such as [22], implemented a distributed monitoring system which was more appropriate

for monitoring a large pipeline network. They made use of mobile sensors to overcome the scalability issue. In a highly dynamic network, the number of messages and energy used by this solution will be expensive.

WSNs-based systems present some challenges and limitations as they require protocol management as well as tool qualification. Modern algorithms are needed for outlier detection, energy conservation, security management, and self-location of targets to fulfill these purposes. The goal of outlier detection methods is to cleanse the gathered data and build the most helpful information for end-users, so they can make the most informed decisions. The source of outlier data is divided into three categories: event, malicious attacks, and noise. The outlier data detected in water pipelines may be due to damage [40]. This latter can cause more abnormal data to be detected. The purpose of outliers is to identify sudden changes in pressure profiles caused by noise, events, or malicious attacks, as shown in Figure 4.

Additionally, the purpose of these approaches is to conserve energy while preserving low communication exchanges so that a decision can be made accurately and efficiently. The tools are also designed to detect and localize damages in a water network, ensuring continuous monitoring of the system [41].

6. LEACH Algorithm

6.1. Network Architecture. LEACH is a protocol that is based on the concept of classification (clustering) shown in Figure 5, which consists of partitioning the network into groups (clusters). The nodes send their data to the CHs which in turn send this data to the BS. The CHs perform simple processing (e.g., aggregation) on the data before transmitting it.

6.2. LEACH Protocol. The operation of the LEACH protocol is as follows: the nodes self-form to be CHs. They are based on the desired percentage of CH in the network and on the number of iterations during which a node has taken the role of CH. A node "n" takes a value between 0 and 1; if this value is less than a threshold $T(n)$, the node declares itself CH.

$$T(n) = \frac{K}{N - K \bmod (r, N/K)} \text{ if } n \in G,$$
$$T(n) = 0 \text{ if not,} \quad (1)$$

where k is the desired percentage of CHs, r is the current iteration, and G is the set of nodes that have not been selected as CHs during the last (N/K) iterations. Then, each CH informs its neighbors of its election, and each node chooses the closest CH.

7. Existing Upgrades to Extend Lifetime

The objective was to favor the nodes which have more energy to become CH. To achieve this, the authors use the formula which makes it possible to calculate the ratio of energy present in the node (E_{pres}) on the initial energy (E_{ini}) and introduce it into the classification formula, the new threshold considering the energy below:

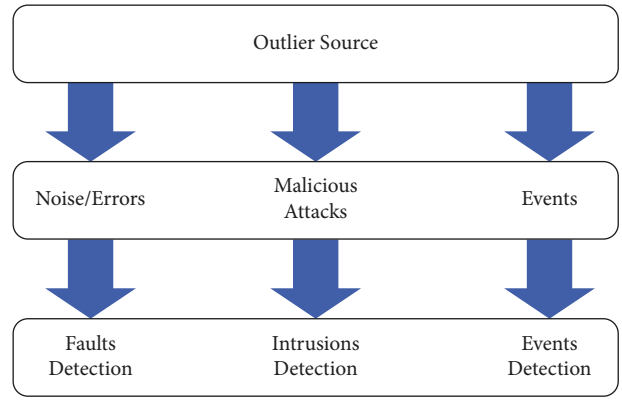


Figure 4: Different types of outlier sources in wireless sensor networks.

$$T_1(n) = T(n)\frac{E_{\text{pres}}}{E_{\text{ini}}}. \quad (2)$$

8. Proposed Improvements

The goal of our solution is to favor the nodes which have more energy and are close to the BS to become CHs. So, we introduce the metric of the distance between the node and the base station.

8.1. Using Distance between the Node and the Base Station in the Selection Formula. The objective of this method is to give more chances to the nodes which are close to the base station so that they become CHs. So, the threshold equation becomes

$$T_2(n) = T(n)\frac{d_{\max}}{d}, \quad (3)$$

where (d) is the distance between the node and the base station and $(d_{\max})$ is the maximum distance between the base station and a node.

As we have already presented previously that the nodes self-form and randomly choose a number between 0 and 1 and this value must be lower than a threshold for the nodes to be CHs, so, this threshold must have a greater or lesser value so that we will have a high probability of having CHs nodes. If we take two nodes $(n1)$ and $(n2)$ which, respectively, have distances $d1 < d2$, we will trivially have $T2\ (n1) > T2\ (n2)$.

8.2. New Selection Formula: Energy and Distance. This formula consists in considering the nodes having more energy and less distance compared to the base station as CHs; the idea is to make a combination between the two formulas $T2$ and $T1$ with such percentages (a) and (b). After some changes to this latter, we got the following result:

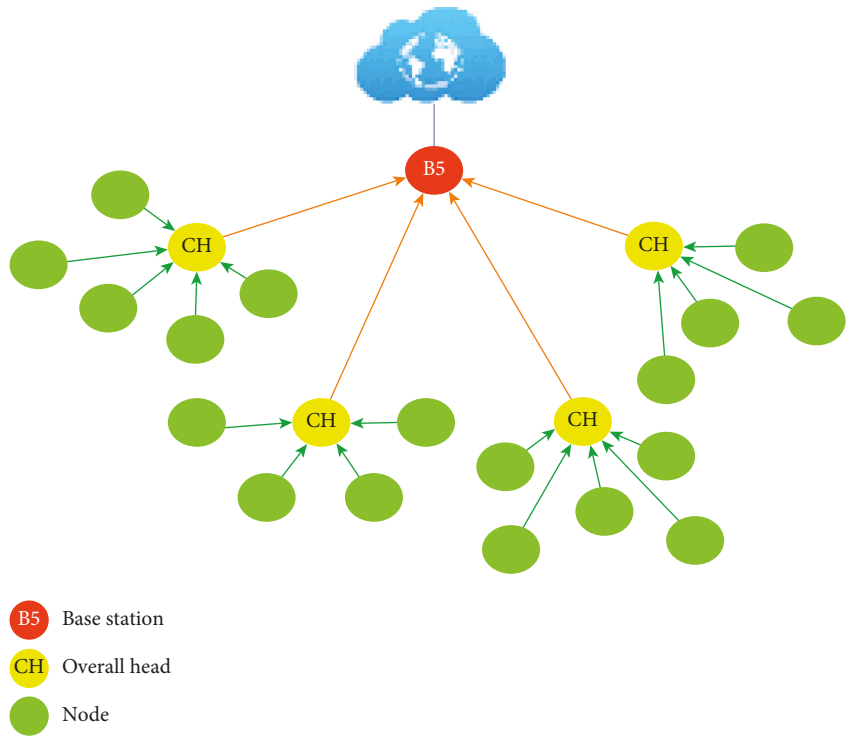


Figure 5: Classification method based on WSNs.

$$T_4(n) = aT_2(n) + bT_1(n), \quad (4)$$

avec $a = 0.15$ and $b = 0.85$ $(a + b = 1)$.

8.3. Experimentation and Results. To evaluate the performance of the LEACH protocol, we based ourselves on the existing improvements as well as on the improvement that we have proposed. We used the NS2 simulator (Network Simulator). The simulation results exhibited in Figure 6 clearly show that there is an extension of the network lifetime, applying the new formula T4.

Figure 6 shows that the T4 formula keeps all nodes together alive for 94% which reflects a further extension in network lifetime. On the other hand, by applying the original LEACH formula, the nodes could only resist 62%. However, we managed to extend the lifetime of the network by 32%. This is an interesting result since, in wireless sensor networks, all the nodes must collaborate to accomplish a well-defined objective.

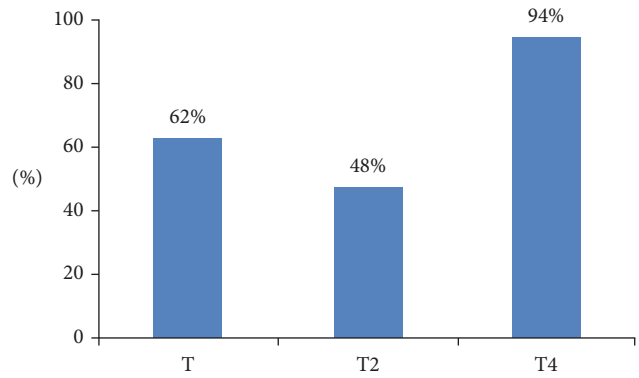


Figure 6: Network lifetime percentage of our proposed method compared to original leach.

9. Development of Smart Water Irrigation System with Moisture Monitoring

According to the proposal irrigation system for water-saving plus moisture, monitoring is described in this section. The proposed system was developed taking into account the extremely hot climate of Saudi Arabia. Hence, the vast majority of the country is covered by the Sahara Desert. In our work, we have developed a water-saving irrigation control system that is based on wireless sensor networks [37], whereby the system comprises low-power wireless sensor nodes that communicate through an ad hoc ZigBee network. We will monitor soil moisture information parameters such as soil water degree, temperature, and relative humidity that can all be used to measure moisture potential. A four-channel temperature and humidity transmitter will be used to collect this data. The information details are determined by using the various component of the sensor. In the following subsections, we will describe in detail the proposed design [42].

9.1. Smart Water Irrigation System for Farmers. As shown in Figure 7, our system of agriculture irrigation is based on famous technology entitled WSN which is made up of four components: an irrigation controller, receiving sensors, a set of sensors, and a network of irrigation pipes. To

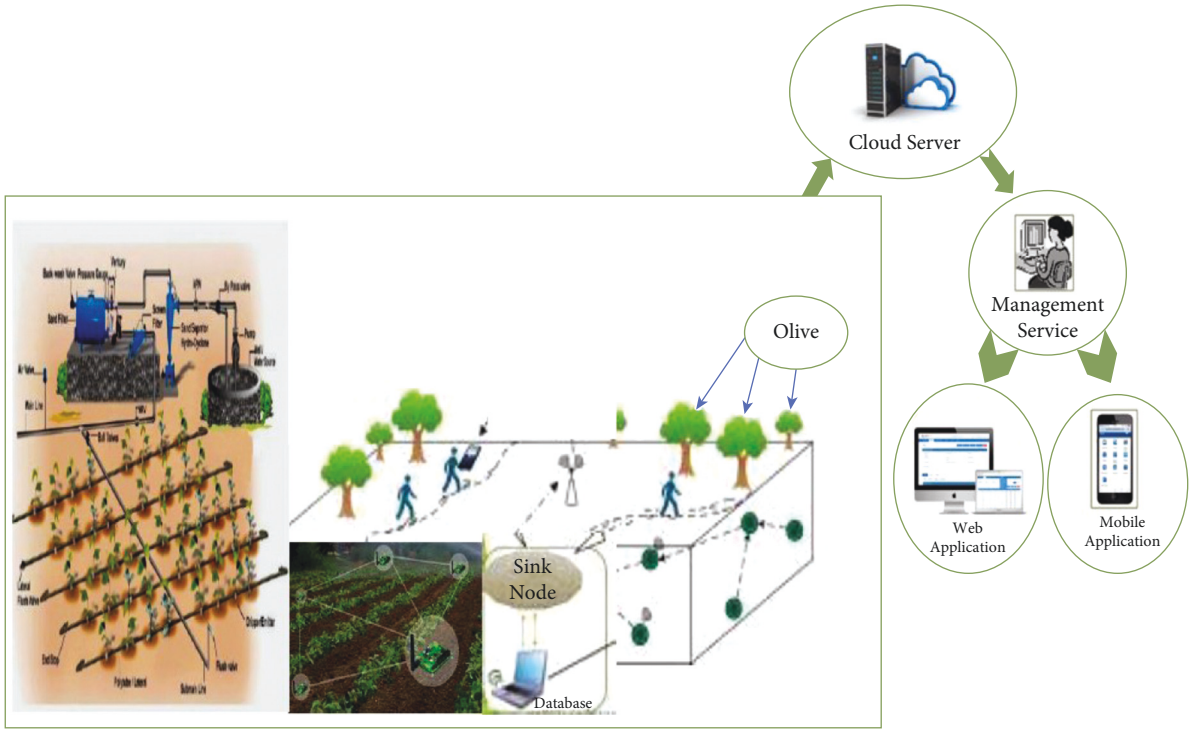


Figure 7: Smart Water Irrigation System (SWIS) based on wireless sensor networks.

construct an irrigation node group, sensor nodes that bear soil moisture are disseminated in accordance with the planting and irrigation status of farming. Each node is in charge of keeping track of the soil moisture in a certain area [43].

A standard WSN, utilizing ZigBee technology based on transmitted wireless data, consists of an irrigation region and receiving sensors. Wireless multihop is used to transfer sensor data to the receiving node. Our discipline is designed to install a network for an irrigation pipe in the farming area in the irrigation region and an electrical control valve on the pipe to create an automated water-saving irrigation system. If the control of smart water irrigation is adaptable, the total system will be more versatile [44].

A water-saving irrigation system may be modified based on the original irrigation pipe network. For greater deployment of irrigation system pipe network and to save money, an electronic control valve can be fitted. The irrigation controller in the WSN coverage region may spray irrigation in specified locations based on sensor data. This system contains a specific module taking charge of network supervision. The proposed Smart Water Irrigation System is mainly dependent on wireless sensor networks and water pipelines [45].

9.2. System Hardware Structure. In this part, the hardware structure of the sensor node implemented in the proposed architecture is addressed and illustrated in Figure 8. The controller module, sensor module, ZigBee protocol communication module, and solar self-powered module make up the majority of the hardware structure [46].

The irrigation controller is built using an embedded system development board as the mainboard. The receiving node receives information through a serial connection and processes the control data. The system is very scalable. The WSNs measure humidity which is realized 5 times in a minute (on a 12-second cycle) and send the data to the irrigation controller. When the irrigation controller detects that the humidity sensed by the WSN nodes in a specific location is lower than the prescribed value, it activates the irrigation network's electric control valve. The system will start irrigation and close the electric control valve of the pipe network in this region when the soil humidity in the area reaches a particular level. This hardware structure presents a very effective impact of improvements of the proposed solution compared to other techniques in the context of energy conservation in wireless sensor networks and in the water-saving process. It both minimizes energy consumption and wasted water [47].

10. Data Collection Methods for Prevention Diseases

An early warning system takes into account the Saudi Arabian country's environment. The data used in our work are humidity, pressure, and temperature. It can be provided by normal events, nose, or malicious attacks. Also, it is divided into data acquisition, data transmission, data processing, and data application. Diseases are a threat to plants of all kinds, but olives are most susceptible to pests because it affects the quality of the oil.

A series of information is formed through the collection of data, data management, data processing, and

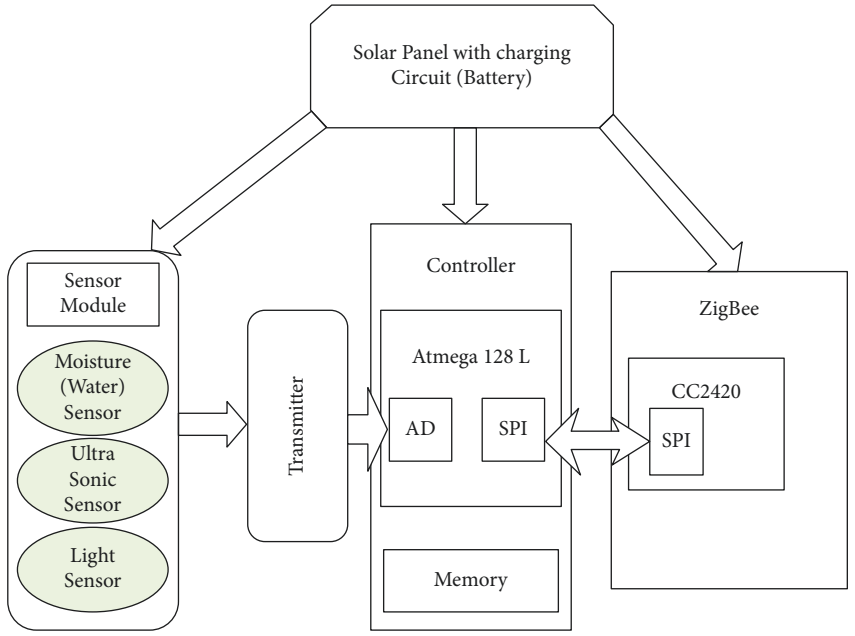


Figure 8: Hardware structure of the sensor node used in the proposed architecture.

forecasting, and finally, the forecasting information is released for monitoring and early warning. A diagram presented in Figure 9 is attached showing the information chain and its main components:

The technique of data gathering, transmission, processing, and application in information technology corresponds to each link of the information chain. In this section, we will discuss monitoring and early warning technologies. Sensor technology, database technology, expert systems technology, global positioning systems (GPS), geographic information systems (GISs), network technologies, and communication technology are all components of the disease monitoring and early warning information chain. For manual analysis of conventional diseases, the use of pad and GPS data acquisition and recording technology is most widely used. Data, as well as GPS positioning information, are simultaneously recorded. During the same time, data on field microclimates are collected, which are closely related to diseases. To obtain real-time microclimate data automatically, the field microclimate data monitoring technology utilizes sensors and communication technologies such as GPRS. Diseases are represented by the data, which is then transmitted to the database for analysis.

In another part, a data collection technology combined with data transmission technology forms the key components of the proposed system. Data management systems can benefit from real time, reliable monitoring data on diseases, and microclimates. This system takes into account environmental factors when it develops the expert system. This latter inference engine produces forecasts of diseases, which are then released through a forecast information release system in order to guide plant protection efforts [42, 43, 45].

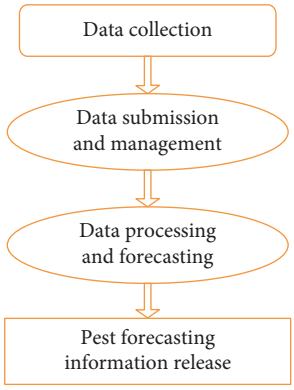


Figure 9: A diagram information chain and its main components.

11. Simulation of Agriculture Effects

For the agriculture effect, simulation analyses are performed here. A variety of physiological and ecological indicators of plant growth are necessary in order to monitor the growth status of plants and improve their scientific management level. Plant growth control activities were adversely affected by the outdated monitoring methods of the past. The use of WSNs in agriculture production is currently being actively introduced, which can study the state of crop growth automatically at a lower cost. The physiological and ecological parameters of plants are also monitored, analyzed, and recorded by an intelligent system. Accordingly, Figure 7 demonstrates how this latter can significantly improve the efficiency of monitoring.

Economic growth has been hindered by a lack of water resources, but agricultural water is being wasted in large amounts. Irrigation research must be intensified to reduce water waste. Monitoring irrigation water and soil moisture

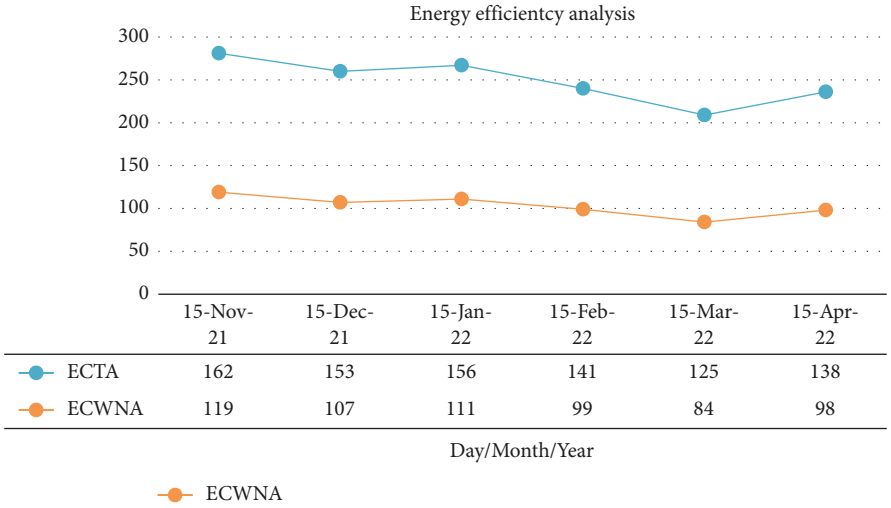


Figure 10: Energy consumption based on WSNs.

during crop growth is a good application of the wireless network agriculture technology. Essentially, this latter system is a simple wireless sensor network (WSN) that allows for agriculture applications. Additionally, an irrigation scheme is developed in conjunction with the development of these crops. There is however little evidence of WSNs being applied to agricultural production, which mostly takes place in smaller production environments such as tea gardens, orchards, and greenhouses, and more research is required. In order to increase production, large-scale farmland is needed. With the increase of time, agricultural irrigation based on wireless sensor networks has played a very important role in water saving. To understand the growth status of plants and improve the science of management, it is important to monitor their physiological and ecological indicators. Plants' growth control activities were affected by the old-fashioned monitoring methods. Currently, WSN is actively utilized in agriculture production to sense crop growth at a lower cost while automatically sensing it. In addition to tracking, analyzing, and recording the physiological and ecological characteristics of plants, intelligent systems serve as a good premise for improving the efficiency of monitoring. Based on data obtained by sensors like temperature, humidity, and pressure, we analyze the state of the land and the consumed energy. A study of data between November 2021 and April 2022 shows that WSN-based agriculture has considerably lower energy consumption than traditional agriculture (mentioned in Figure 10).

12. Conclusion

One of the most challenging aspects of wireless sensor networks is their energy efficiency. Based on the water pipeline, this paper proposes an efficient way of utilizing the energy of wireless sensor networks for agriculture production. We are particularly interested in various monitoring techniques suitable for preserving the number of water pipelines as part of our work. To provide WSN with an information platform that can support a better role in agricultural production, the proposed model aims to improve on the shortcomings of existing WSN in the context of energy efficiency. Agriculture is a key factor impacting the global economy; therefore, WSN is important.

In the end, based on our solution, we conclude that 19% of water could be saved using this irrigation practice. Also, wireless sensor networks can improve accuracy and efficiency to achieve agricultural goals as well as reduce costs associated with wireless protocol systems. Agriculture can improve its management and production efficiency as a result of using an agricultural automation system.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest to report regarding the present study.

Acknowledgments

This work was funded by the Deanship of Scientific Research at Jouf University under grant No (DSR-2021-02-0374).

References

[1] H. Alshammari, K. Gasmi, M. Krichen, L. B. Ammar, and M. O. Abdelhadi, "Optimal deep learning model for olive disease diagnosis based on an adaptive genetic algorithm," *Wireless Communications and Mobile Computing*, vol. 10, no. 8, pp. 102–110, 2022.

[2] K. H. Al-Zahrani and M. Baig, "Water in the kingdom of Saudi Arabia: sustainable management options," *J. Anim. Plant Sci*, vol. 21, pp. 601–604, 2011.

[3] R. Liemberger and A. Wyatt, "Quantifying the global non-revenue water problem," *Water Supply*, vol. 14, no. 5, pp. 831–837, 2019.

[4] A. Ayadi, O. Ghorbel, M. S. BenSaleh, and M. Abid, "A framework of monitoring water pipeline techniques based on sensors technologies," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 2, pp. 47–57, 2022.

[5] A. Ayadi, O. Ghorbel, M. S. BenSaleh, and M. Abid, "Spatio-temporal correlations for damages identification and localization in water pipeline systems based on WSNs," *Computer Networks*, vol. 17, no. 1, pp. 107–134, 2020.

[6] A. Ayadi, O. Ghorbel, A. Obeid, M. S. Ben Saleh, and M. Abid, "Leak detection in water pipeline by means of pressure measurements for WSN," in *Proceedings of the International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 1–6, IEEE, Fez, Morocco, May 2017.

[7] R. Aworka, L. S. Cedric, W. Y. H. Adoni et al., "Agricultural decision system based on advanced machine learning models for yield prediction: case of east african countries," *Smart Agricultural Technology*, vol. 10, no. 48, pp. 185–198, 2022.

[8] A. Ayadi, O. Ghorbel, M. S. Abid, and M. Abid, "Kernelized technique for outliers detection to monitoring water pipeline based on WSNs," *Computer Networks*, vol. 150, no. 13, pp. 179–189, 2019.

[9] J. Zhao, S. Sun, and J. Guo, "Design of intelligent monitoring system for facility agricultural environment based on wireless sensor network," *China Agricultural Machinery Chemical Report*, vol. 41, no. 4, pp. 152–157, 2020.

[10] L. S. Cedric, W. Y. H. Adoni, R. Aworka et al., "Crops yield prediction based on machine learning models: case of west african countries," *Smart Agricultural Technology*, vol. 10, no. 49, pp. 172–189, 2022.

[11] Y. Xie, Z. Li, and Z. Wang, "Research on key technologies of wireless sensor network for fine agriculture," *Computer Science and Application*, vol. 10, no. 10, 2020.

[12] S. M. Antony, S. Pandey, and R. Pandey, "An efficient solar energy harvesting system for wireless sensor network nodes," *Journal of Information and Optimization Sciences*, vol. 41, no. 1, pp. 39–50, 2020.

[13] F. K. Zeadally and S. Zeadally, "Energy harvesting in wireless sensor networks: a comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 55, pp. 1041–1054, 2016.

[14] O. H. Yahya, H. Alrikabi, and I. A. Aljazaery, "Reducing the data rate in internet of things applications by using wireless sensor network," *International Journal of Online and Biomedical Engineering*, vol. 16, no. 3, pp. 107–116, 2020.

[15] E. Zhang, Z. Zhang, and J. Chang, "Development of intelligent agricultural crop environment multi parameter monitoring system based on wireless sensor network," *Journal of the Yellow River Academy of Science and Technology*, vol. 22, no. 2, pp. 84–88, 2020.

[16] Y. Liang, "Development of remote intelligent control system for facility agriculture based on WSN," *Chongqing Three Gorges College*, Chongqing, China, 2020.

[17] M. Chen, X. Wang, and H. Zhang, "Wireless visual intelligent agricultural management system based on WSNs," *Agricultural Mechanization Research*, vol. 10, no. 7, 2021.

[18] J. J. Hao, "Application of wireless sensor network in modern agriculture in Guanzhong region," *Information and Computer*, vol. 32, no. 19, pp. 128-129, 2020.

[19] C. Huang, Z. D. Wang, F. Wang, L. Mei, and X. S. Chen, "Research and design of agricultural environment monitoring system based on STM32 + ZigBee technology," *Agriculture and Technology*, vol. 40, no. 23, pp. 62–64, 2020.

[20] L. Yang, B. Zhang, and T. T. Chen, "Design of cluster agricultural environment information collection network based on nRF24L01," *Digital Technology and Application*, vol. 38, no. 1, pp. 134–136, 2020.

[21] A. M. Sadeghioon, N. Metje, D. N. Anthony, and C. J. Anthony, "Smartpipes: smart wireless sensor networks for leak detection in water pipelines," *Journal of Sensor and Actuator Networks*, vol. 3, no. 1, pp. 64–78, 2014.

[22] M. Abdelhafidh, L. C. Fourati, M. Fourati, and A. Abidi, "Hybrid mechanism for remote water pipeline monitoring system," in *Proceedings of the IEEE 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pp. 2140–2145, Valencia, Spain, June 2017.

[23] I. Jawhar, N. Mohamed, J. Al-Jaroodi, and S. Zhang, "An architecture for using autonomous underwater vehicles in wireless sensor networks for underwater pipeline monitoring," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1329–1340, 2018.

[24] M. Abdelhafidh, M. Fourati, L. C. Fourati, and A. Abidi, "Remote water pipeline monitoring system iot-based architecture for new industrial era 4.0," in *Proceedings of the IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, pp. 1184–1191, Hammamet, Tunisia, October 2017.

[25] M. Murad, A. A. Sheikh, M. A. Manzoor, E. Felemban, and S. Qaisar, "A survey on current underwater acoustic sensor network applications," *International Journal of Computing Theory Engineering*, vol. 7, pp. 51–59, 2015.

[26] K. Rehman and F. Nawaz, "Remote pipeline monitoring using wireless sensor networks," in *Proceedings of the Communication, Computing and Digital Systems (C-CODE)*, pp. 32–37, Islamabad, Pakistan, March 2017.

[27] D. Wu, K. Y. Toumi, S. Mekid, and R. BenMansour, "Wireless communication systems for underground pipe inspection," *US Patent*, vol. 9, pp. 421–448, 2017.

[28] A. Abdulshaheed, F. Ghavamian, and A. Ghavamian, "A Pressure-Based Method for Monitoring Leaks in a Pipe Distribution System: A Review," *Renewable Sustained Energy*, vol. 69, pp. 902–911, 2017.

[29] K. Saravanan, E. Anusuya, and R. Kumar, "Real-time water quality monitoring using internet of things in scada," *Environmental Monitoring and Assessment*, vol. 19, no. 10, pp. 556–564, 2018.

[30] W. Cheng, H. Fang, G. Xu, and M. Chen, "Using scada to detect and locate bursts in a long-distance water pipeline," *Water*, vol. 17, no. 27, 2018.

[31] J. Chen, J. Zheng, F. Xiong, Q. Ge, Q. Yan, and F. Cheng, "Experimental investigation of leak detection using mobile distributed monitoring system," *Smart Materials and Structures*, vol. 27, 2017.

[32] Y. Wu, S. Liu, K. Smith, and X. Wang, "Using correlation between data from multiple monitoring sensors to detect bursts in water distribution systems," *Journal of Water Resources*, vol. 11, no. 23, pp. 73–79, 2017.

[33] S. Grosswig, E. Hurtig, S. Luebbecke, and B. Vogel, "Pipeline leakage detection using distributed fiber optical temperature sensing," *17th International Conference on Optical Fiber Sensors*, vol. 58, no. 55, pp. 226–229, 2015.

[34] A. Cataldo, E. D. De Benedetto, G. Cannazza, D. G. Leucci, and C. Demitri, "Enhancement of leak detection in pipelines through time-domain reflectometry/ground penetrating radar measurements," *IET Science, Measurement & Technology*, vol. 11, no. 6, pp. 696–702, 2017.

[35] L. Wong, R. N. Deo, S. Rathnayaka et al., "Leak detection and quantification of leak size along water pipe using optical fibre

sensors package," *Electronic Journal of Structural Engineering*, vol. 18, no. 1, pp. 47–53, 2018.

[36] D. Enache, O. Chenaru, L. Ichim, and D. Popescu, "Residual water burst detection using wsn measurements and cloud analysis," in *Proceedings of the 2018 26th Mediterranean Conference on Control and Automation (MED)*, pp. 926–931, IEEE, Zadar, Croatia, June 2018.

[37] S. Kartakis, E. Abraham, and J. A. M. Cann, "Waterbox: a testbed for monitoring and controlling smart water networks," in *Proceedings of the 1st ACM International Workshop on Cyber-Physical Systems for Smart Water Networks*, pp. 3–8, ACM, New York, NY, USA, April 2015.

[38] L. Cai, R. Wang, J. Sun, S. Li, and Y. Jing, "Urban water supply network monitoring and management platform based on wireless sensor network," in *Proceedings of the Wireless Communications, Networking and Applications*, pp. 351–361, Springer, New Delhi, India, October 2016.

[39] Z. Kleiner and Y. Kleiner, "State of the art review of inspection technologies for condition assessment of water pipes," *Measurement*, vol. 46, no. 1, pp. 1–15, Measurement, 2013.

[40] Y. Jia, *Research on Wireless Irrigation System Based on ZigBee*, pp. 22–29, Xi'an University of Electronic Technology, Xi'an, China, 2020.

[41] A. I. Taloba, M. R. Riad, and T. H. A. Soliman, "Developing an efficient spectral clustering algorithm on large scale graphs in spark," in *Proceedings of the 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, pp. 292–298, IEEE, Cairo, Egypt, December 2017.

[42] O. Ghorbel, R. Ayadi, A. S. Alanazi, R. Alanazi, R. N. Alabdali, and M. Abid, "Adaptive KPCA for outlier detection in wireless sensor networks: water pipeline case," in *Proceedings of the ISNCC*, pp. 1–5, Dubai, United Arab Emirates, October 2021.

[43] I. Ahmed, "Taloba an artificial neural network mechanism for optimizing the water treatment process and desalination process," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9287–9295, 2022.

[44] A. Rayan, A. I. Taloba, M. Rasha Abd El-Aziz, and A. Amr, "IoT enabled secured fog based cloud server management using task prioritization strategies," *International Journal of Advanced Research in Engineering & Technology*, vol. 11, p. 9, 2020.

[45] G. Xue, F. Lin, S. Li, and H. Liu, "Adaptive dynamic surface control for finite-time tracking of uncertain nonlinear systems with dead-zone inputs and actuator faults," *International Journal of Control, Automation and Systems*, vol. 19, no. 8, pp. 2797–2811, 2021.

[46] M. Elloumi, M. A. Ahmad, A. H. Samak, A. M. Al-Sharafi, D. Kihara, and A. I. Taloba, "Error correction algorithms in non-null aspheric testing next generation sequencing data," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9819–9829, 2022.

[47] S. Ha, L. Chen, H. Liu, and S. Zhang, "Command filtered adaptive fuzzy control of fractional-order nonlinear systems," *European Journal of Control*, vol. 63, pp. 48–60, 2022.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 8077664, 8 pages
https://doi.org/10.1155/2022/8077664

Research Article

K-Mer Spectrum-Based Error Correction Algorithm for Next-Generation Sequencing Data

Hussah N. AlEisa,[1] **Safwat Hamad**,[2] **and Ahmed Elhadad**[3]

[1]*Department of Computer Sciences, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia*
[2]*Department of Scientific Computing, Faculty of Computer and Information Sciences, Ain Shams University, Cairo, Egypt*
[3]*Department of Computer Science, Faculty of Computers and Information, South Valley University, Qena, Egypt*

Correspondence should be addressed to Hussah N. AlEisa; haleisa@pnu.edu.sa

Received 18 May 2022; Accepted 13 June 2022; Published 14 July 2022

Academic Editor: Wei Xiang

In the mid-1970s, the first-generation sequencing technique (Sanger) was created. It used Advanced BioSystems sequencing devices and Beckman's GeXP genetic testing technology. The second-generation sequencing (2GS) technique arrived just several years after the first human genome was published in 2003. 2GS devices are very quicker than Sanger sequencing equipment, with considerably cheaper manufacturing costs and far higher throughput in the form of short reads. The third-generation sequencing (3GS) method, initially introduced in 2005, offers further reduced manufacturing costs and higher throughput. Even though sequencing technique has result generations, it is error-prone due to a large number of reads. The study of this massive amount of data will aid in the decoding of life secrets, the detection of infections, the development of improved crops, and the improvement of life quality, among other things. This is a challenging task, which is complicated not just by a large number of reads and by the occurrence of sequencing mistakes. As a result, error correction is a crucial duty in data processing; it entails identifying and correcting read errors. Various k-spectrum-based error correction algorithms' performance can be influenced by a variety of characteristics like coverage depth, read length, and genome size, as demonstrated in this work. As a result, time and effort must be put into selecting acceptable approaches for error correction of certain NGS data.

1. Introduction

Nature methods named next-generation high-throughput DNA sequencing techniques as the method of the year in 2007. These methods are creating interesting new potential in biology [1]. The road to garnering the approval of the revolutionary technology, on the other hand, was not simple. Until recently, the Sanger enzymatic dideoxy method, first explained in 1977, and the Maxam and Gilbert chemical degradation technique, first mentioned in the same year, were the methodologies used for sequence analysis. The Maxam and Gilbert chemical degradation technique was used in sequential cases that could not be solved easily with the Sanger method [2]. The potential to decipher genomes and conduct ground-breaking biomedical sciences has been made possible by the rapid synthesis and accessibility of enormous amounts of DNA sequencing obtained by next-generation sequencing (NGS) technology at a lower cost than traditional Sanger sequencing [3]. There has been a significant trend apart from using automated Sanger sequencing for genome analysis in the last four years. Previous to this departure, the automated Sanger sequencing had taken over the market for half a century, resulting in a slew of significant achievements, such as the production of the only completed human genome sequence.

Despite numerous technological advances during this period, the drawbacks of automated Sanger sequencing demonstrated the need for new and superior methods for sequencing huge numbers of human genomes [4]. Sanger sequencing has seen less documented advancements as

recent attempts have been focused on the development of novel technologies [5]. As a result, automated Sanger sequencing is not addressed in this paper, and curious readers should go to earlier pages. The mechanized Sanger approach is a "first-generation" approach, whereas "next-generation sequencing" (NGS) refers to newly developed technology. A range of methods combining templates' sequencing, processing, but also imaging, and also genomic alignments and assembling approaches are included in these modern technologies [6]. The introduction of next-generation sequencing (NGS) techniques to the markets has altered how we thought regarding scientific methodologies in fundamental, practical, and healthcare findings. In certain ways, the promise of NGS is similar to that of PCR in the initial periods, with the major limitations being one's imagination [7]. The capability to create a tremendous volume of information for a low cost—in certain cases exceeding of one billion shorter readings each instrument's cycle—is NGS's biggest advantage. This component broadens the scope of exploration beyond defining basic ordering [8]. Microarrays, for instance, are being phased out of gene expressing research in favour of sequence-based approaches, which may detect and estimate uncommon transcripts without previous understanding of a gene and offer data on alternate sequencing and splicing variations in discovered genes [9].

The capacity to sequence the entire genomes of several closely connected organisms has enabled large-scale comparing and developmental investigations that were previously unthinkable [10]. The sequence analysis of individual genomes to improve our knowledge of how genetic variants impact diseases and health might be the most broad application of NGS [11]. Several platforms are expected to coexist in the industry due to the variety of NGS characteristics, with several providing apparent benefits for specific implementations over another [12]. Template development, sequential and visualization, and information processing are only a few of the processes used in sequenced techniques [13–16]. The kinds of information generated by every platform are determined by the distinctive combined effect of specialized standards that differentiates one technique from others [17]. When evaluating technologies depending on information cost and quality, these disparities in information production provide a hurdle. Even though every manufacturer provides performance ratings and precision predictions, there is no agreement that a 'reliability foundation' from one platform was similar to that from others [18].

2. Related Works

Six k-spectrum-based approaches, namely Trowel, Reptile, Bloocoo, Musket, Lighter, and Bless, were evaluated utilizing six generated collections of paired-end Illumina sequencing information in this technique [19]. The genome size, coverage depth, and read length of these NGS databases changed. The Error Rectification Evaluation Toolkit (ECET) was used to provide a set of metrics (such as true positives, false positives, false negatives, recall, accuracy, gain, and F-score) for evaluating every programme's correcting performance. Musket exhibited the best total effectiveness throughout the spectrum of studied variations indicated in the six databases, according to findings from computational simulations. Musket's efficiency was lesser in a population with a moderate read length, coverage depth, and a short genome (F-score = 0.81). The remaining five techniques executed poorly (F-score <0.80) or refused to analyse one or many data sources. Individual k-spectrum-oriented error correction algorithms' effectiveness can be influenced by a variety of characteristics including such genome size, coverage depth, and read lengths, as demonstrated in this work. As a result, time and attention must be put into selecting acceptable approaches for error correction of particular NGS databases. Due to its continuously improved effectiveness through every six experimental databases, we propose Musket as the best choice depending on our comparison analysis.

Longer read error correction has become a concern for bioinformaticians, motivating the implementation of novel error correction algorithms tailored to NGS technology. We introduced a new approach for error correction of LLRs created by an NGS sequencing utilizing another LLRs produced by the similar NGS sequencing in this publication [20]. As a result, our method is self-correcting. They describe a novel de novo self-error correction technique that uses just lengthy reads in this study. Our technique is divided into two parts: to begin, they employ a quick hashing algorithm that enables us to locate connections among the largest reads as well as other reads in a collection of extensive reads. Then, utilizing a dynamic programming approach in a band of width w, they employ the largest reads as seeds to determine the overall alignments of long reads. In contrary to previous hybrid error correction algorithms, the error correction technique does not require high-quality reads. They are now doing an experimental investigation on our self-correction method. They would utilize fictional information in which they have developed and actual information produced by Oxford Nanopore and Pacific Biosciences sequencers to conduct this research. The findings would be compared with those achieved by other self-correction methods.

They provide two novel effective error correction methods tailored for viral amplicons in this study [21]: (a) k-mer-based error correction (KEC) and (b) empirical frequency threshold (ET). Both were tested against a recently reported clustering technique (SHORAH) on 25 testing databases generated by 454-sequencing of amplicons with specified sequencing. Discovering actual haplotypes is comparable for all three techniques. ET and KEC, on the other hand, were far more effective than SHORAH at eliminating erroneous haplotypes and determining the frequencies of actual haplotypes. Both the ET and KEC techniques are well suited to the quick extraction of error-free haplotypes achieved from 454-sequencing of amplicons from heterogeneous viruses. In terms of detecting actual KEC, SHORAH, haplotypes, and ET are all similarly effective. Meanwhile, newer methods, ET, and KEC are more effective than SHORAH at eliminating erroneous haplotypes and determining the frequencies of actual haplotypes. Both techniques are well suited to quickly recovering high-quality

haplotypes from reads generated by NGS of amplicons from diverse viruses like HIV and HCV.

In this study [22], two distinct error-correcting types of software, ECHO and Quake, are tested on next-generation sequencing information from heterozygous genomes to assess how well they function. Quake and ECHO performed admirably and were capable to rectify most of the information's problems. Errors that occurred in heterozygous sites, on the other hand, have a distinct pattern. Inaccuracies at these locations were occasionally wrongly rectified, bringing errors into the database and perhaps resulting in a chimeric read. The quake had a substantially lower chance of producing chimeric readings. Read cutting in Quake deleted a significant amount of the original information, leaving reads with fewer heterozygous markers. ECHO produced more chimeric reads and generated more mistakes than Quake, but heterozygous markers were maintained. The assembling statistics were enhanced by utilizing genuine *E. coli* sequenced information and their assemblies following error correction. It was also discovered that sorting reads by haplotype improves assembling performance. These results imply that when employed to heterozygous information, both ECHO and Quake have benefits and drawbacks. With the growing popularity of haplotype-specific research, newer technologies that are meant to be haplotype-aware and do not have the flaws of ECHO and Quake are required.

They introduce Karect and unique multidimensional alignment-based error-correcting approach for next-generation sequencing information [23]. Replacement, addition, and removal errors are all supported by our method. Karect is dependent on multiple orientations; therefore, it can manage nonuniform coverage and also somewhat occupied sections. It also allows substitutions, inserting, and deleting mistakes. This could manage nonuniform coverage and also portions of the sequencing genomes that are only partly covered. Tests using data sets from 454 FLX, Ion Torrent, and Illumina sequenced machinery show that Karect is more efficient than earlier approaches in both fixing individual dependent errors (up to a 10 per cent improvement in efficiency gains) and post-de novo assembling efficiency (up to 10 per cent enhancing in NGA50). They also present a new paradigm for assessing the accuracy of error corrections. Karect provides improved error-correcting comparison to current state-of-the-art approaches, according to comprehensive experimental assessment. When employed as a preprocessing phase for newer assemblers, Karect also results in much enhanced components. They presently do not handle Pacific Biosciences data due to chimeric readings; however, they are functioning to resolve this.

To reduce the impact of mistakes on the detection of minority variants, they established a probabilistic Bayesian strategy [24]. Pyrosequencing information from a 1.5-kb segment of the HIV-1 gag/pol gene in two controls and two medical samples was used to test it. The impact of PCR amplification was looked into. In the PCR-amplified and non-PCR data sets, error correction reduced the pyro sequenced based substitution rates by two and five times, correspondingly, from 0.05 per cent to 0.03 per cent and from 0.25 per cent to 0.05 per cent. With complete sequencing reconstructions, they were capable to discover viral clones as rare as 0.1 per cent. In terms of recall and precision, probabilistic haplotype inference exceeds counting-based identification. The genetic diversity found inside and among two medical data sets resulting in a variety of phenotypic drug-resistant characteristics, implying a strong epidemiological relationship. They conclude this, if technological problems are appropriately addressed, pyrosequencing could be utilized to analyse genetically heterogeneous materials with great efficiency.

3. Methodology

Error correction approaches depending on k-spectrum are derived from prior spectral alignments implementations of de Bruijn graph assemblers and following an extended architecture as illustrated in Figure 1. In a collection of readings, a k-spectrum is the probability of a collection of decomposing separate substrings of length k (in other words, k-mer). Within the spectra characteristic area, it estimates the presence of every k-length contiguous strings expressed as vectors. When contrasted to sequencing without errors, errors in sequencing should result in a large diverging at low k-mer frequency. Inconstant genomic repetitions and genome sampling can arise at high frequencies in mistake correction, resulting in a plethora of identically susceptible correction alternatives. This means that k-mers with short hamming distances are likely from the identical genomic region and must be adjusted depending on their frequency of occurrences. After extracting k-mers from sequenced reads, the k-spectrum-based correction begins by providing a weighed number to every k-mer. Depending on categorized counting frequency or basic performance ratings, the number is provided. Weak (untrusted or insolid) k-mers with lower frequency are differentiated from solid (trusted) k-mers by evaluating and setting an appropriate error threshold (with higher frequency). Error correction is applied to reading with weak k-mers by continuously transforming those to solid k-mers until there are no more weakly k-mers in the sequencing. After validation, only solid k-mers would be maintained.

3.1. Bloom Filter. Bloom filtering is used as the data framework in the bulk of the approaches explored in this research. A Bloom filtering is a space-effective probability information framework that uses binary arrays and several hash operations to determine whether an entry is a component of a subset. This could correctly identify a collection's nonmember component. A Bloom filtering has a 100 percentage recall probability because a query might generate false positives, however no false negatives. A Bloom filtering does not save the components themselves, but it does enable you to see whether an object is definitely missing from the Bloom filtering or if it has been included to the Bloom filtering. Most approaches use the counted Bloom filter variation for sequencing error corrections, wherein array locations are not individual values but an n-bit counter. The

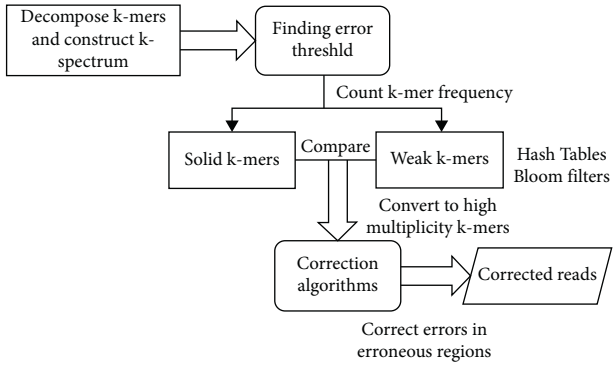


Figure 1: Framework for the proposed method.

amount of bits in the arrays, the number of hashing algorithms, and, more significantly, the reliability of the hash processes all affect the effectiveness of Bloom filtering.

3.2. KEC Algorithm. The k-mer-based error correction (KEC) system consists of four phases:

(1) Determine the number of k-mers s and their frequency $kc(s)$ (k-counts). We presume that k-mers with a high k-count ("solid") are right, but k-mers with a low k-count ("weak") include errors.

(2) Establish the k-count threshold (threshold error) that separates solid k-mers from weak k-mers.

(3) Locate the error locations. The read's error zone is the section $[i, j]$, where the k-mer beginning at position p is evaluated weak for each $[i, j]$.

(4) Errors in error sections should be corrected. Let $r = (r_1, ..., r_n)$ be the constant reads, and $r_i \in \{A, T, G, C\}$ be the variable read. $S_k(i)$ denotes the k-mer of r beginning at location i, and $KC_k(i)$ is the k-count of this k-mer. Let $\text{pref}_j(s)$ be the prefixed of length j of an unspecified sequence s.

3.2.1. Calculating K-Mers and K-Counts. The individual reads r, as well as their frequency f_r, were saved. Due to the often huge size of the knowledge collection, simple calculations of k-mers and k-counts are ineffective. We utilize a hashed map with every key being a k-mer s and the value being the arrays $v(s) = (r, i)$: $s = S_k(i)$ in the read r. Including very huge database collections, the hash mapping could be quickly created.

3.2.2. Finding the Error Threshold. To determine the error threshold, the concept suggested is employed considering the frequency distributions of k-count variables. The frequency of the k-count integer v is denoted by $f(v)$. The k-counts of erroneous k-mers and accurate k-mers are considered to have distinct probabilities. It was discovered that the prototype for the distributions does not need to be directly considered because the initial minimum of $f(v)$ satisfactorily differentiates distinct probabilities and could thus be utilized as the error threshold. Furthermore, due to the more discontinuous distributions of k-count readings in amplicon information than in shotgun trials, this technique is frequently inapplicable. The break in the distributed relates to the first minimum of $f(v)$, which is normally equivalent to 0 (in other words, to the first k-count value, there is no equivalent k-mers). The error threshold t_{er} is defined as the end of the first suitably lengthy section of consecutive 0ıs in $f(v)$. The method's component is the duration of the section.

3.2.3. Finding Error Regions. Each read's error regions are estimated as followed. We start by looking for independent sections $[i, j]$ in order to identify $KC_k(p) \le t_{er}$ for each and every $p \in [i, j]$. The read's k-mers are then categorized as per their k-counts utilizing the varied bandwidth mean-shift technique of cluster analysis. We utilize the variable bandwidth mean-shift technique's fast deployment FAMS. Furthermore, each segment is expanded in both orientations by inserting sequential places q according to the accompanying rule: q is inserted if and only if $p \epsilon [i, j]$ occurs and k-mers $S_k(p)$ and $S_k(q)$ correspond to the similar clustering. Overlapping sections are linked together, and the resultant sections are called error areas.

3.2.4. Error Correction. There are three components in this phase:

(a) Correction of errors in "short" error areas (with lengths not exceeding k)

(b) Correction of errors in "long" error areas (with lengths higher than k)

(c) Postprocessing and reconstruction of haplotypes

Phases (a) and (b) can be applied to every sequenced data and are regarded independent algorithms. The amplicon information is processed in phase (c).

Instruments for assessment to take advantage of its neutral structure recognized as targeted error formats (TEF), and the Error Correction Evaluation Toolkit (ECET) versions were employed for effectiveness assessment. ECET also generates error correction data and measurements, which could be utilized to evaluate efficiency immediately. The discrepancy between the erroneous sequence and the reference genome was determined by read alignments. The conversion of SAM (sequence alignment/mapping) structured data to FASTQ documents performed by the samanalysis.py script in ECET was validated using SamToFastq, one of the Picard command line instruments.

Sequencing database simulations, postcorrection, and precorrection alignment to reference genomes and derivation of assessment measures and statistics are all processes in the approach, as illustrated in Figure 2. In a nutshell, FASTQ format was used to construct both error-containing paired-end and error-free sequencing. During the procedure, the error-free information was used for QA/QC. Simulated sequences were associated with a reference genome utilizing BWA afterwards translating FASTQ to FASTA (preprocess owing to ECET's header demands before alignment). ECET

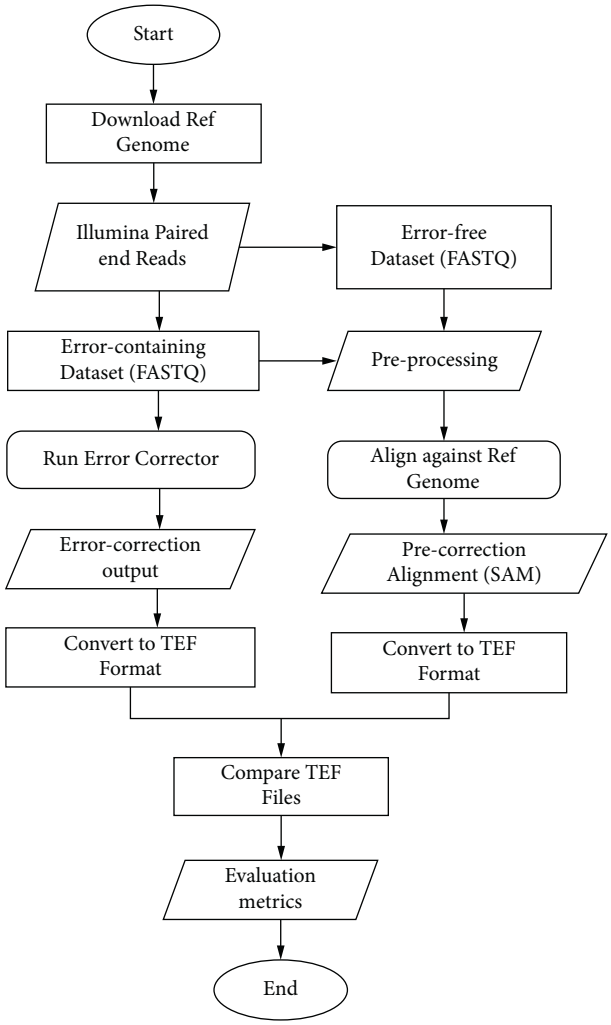


Figure 2: Flowchart for the proposed method.

was used to transfer the SAM alignments documents generated by BWA to TEF format. Error-infested statistics were cleaned up with error-correcting software. ECET was used to transform the error correction findings from these instruments to TEF documents. The Comp2PCAlign tool included in ECET was used to evaluate the TEF records that produced postcorrection and precorrection measurements and statistics for effectiveness evaluation.

4. Experimental Results

The effectiveness measures that were obtained are shown in the tables. A negative gain number indicates that additional errors are generated than rectified in the data set. Negative gains were seen in five approaches (Bless, Trowel, Bloocoo, Lighter, and Reptile), particularly in the synthesis of EC-3. The most thorough measurement of mistake correcting effectiveness is the F-score. With the exception of Musket, every approaches outperformed with at least one data set when F-score = 0.96 was used as the threshold for satisfactory efficiency. As a result, Musket had the total finest efficiency, while Trowel had the lowest, with five examples of poor effectiveness.

4.1. Analysis of the Context of the Error Correction Base Sequence. We looked at the sequenced tuples surrounding error locations to see whether there were any preferences for sequencing composing near them. We selected the appropriate section of the reference sequencing for the study because errors at location 1 do not have previous bases and errors at the place last do not have subsequent characters in the reads. This also prevents the analysis of incorrect base calls in error-prone read sequencing near to the mistake site in question. Since this section of the source sequences is not section of the sequencing fragments, the sequencing component preceding the read beginning is not recognized to be the source of an error at position 1. The bases succeeding the conclusion of the read, on the other hand, may have an impact on base calling. We chose to search for the surrounding bases in the reference sequencing for every mistake locations in the similar way. We did not regard every tuples in the referenced sequencing as referenced tuples, but rather every tuples across all individually paired reads (added 6 bases preceding and following the relevant read section from the reference sequencing). This is done to avoid read coverage bias towards GC-rich sections of the reference sequenced in the study. We produced sequencing logos for Beta and *Helicobacter* using comparative frequency for 4 to 12 base tuples encompassing the mistake in the centre location. We exhibit the 4 and 6 bases tuple findings in scatterplots with the tuple frequency for all data set collections to visualize the overall pattern. In every information collections, all 3-base tuples are beginning with a *G* are obviously dominant, with *G*-error-A and *G*-error-*G* becoming the leading choices. Errors containing tuples beginning with *A or T* are underestimated, but errors containing tuples beginning with *C* are as common as referenced tuples. *T* is the third base in the three least frequently tuples *T*-error-*T*, *C*-error-*T*, and *A*-error-*T*, and is the least recurring base following an error. In the scatterplot with 6-base tuples, the pattern of *G* becoming the most common base preceding an error is kept and emphasized even better. Before an error, *Gs* are usually the favoured bases, and *Ts* are the least commonly used bases. The incorrect location was preceded by *G* in 36 and 33 per cent of instances (*Helicobacter* and Beta, correspondingly).

4.2. Error Correction Techniques' Efficiency and Productivity. To test the efficiency and productivity of every error correction approach, we utilized a comprehensive collection of evaluating measures. True positives (TP) were described as errors that were properly resolved by the error correction instrument, false positives (FP) were characterized as appropriate bases that were wrongly modified by the instrument, false negatives (FN) were characterized as inaccurate bases that were not corrected or inaccurately resolved by the instrument, and true negatives (TN) were characterized as appropriate bases that were not impacted by the instrument. To evaluate the effectiveness of every error-correcting technology, we employed the gained measure [25]. A positive gain implies that the error correction procedure had a positive overall influence, whereas a negative gain suggests

that the device performed more incorrectly than correctly. A gain of 1.0 indicates that the error correction tool performed completed required adjustments without affecting the FP. Precision was described as the percentage of appropriate corrections made out of the overall number of corrections made by the error correction instrument. Sensitivity measures the proportion of rectified errors between every detected error in the data set; in other terms, sensitivities show that techniques repair the greatest number of generated errors [25]. Lastly, we looked to see whether the error correction approaches removed bases from updated reads at the starting or finish. Reducing the bases could be part of an attempts to rectify a deletions (TP trimming), or it could just be a case of eliminating a proper base (FP trimming).

4.3. Correction of Errors and Reconfiguration of Haplotypes. The per-base error ratio for the non-PCR-amplified sample was 0.05 per cent, while the expanded collection had a greater error ratio of 0.30 per cent. As a result, ~12 per cent of non-PCR reads have one or more sequenced errors, while 44 per cent of PCR reads have one or more sequenced errors. After the error correction technique, the error rate for the PCR and non-PCR instances lowers to 0.04 per cent and 0.06 per cent, correspondingly, resulting in ~93 per cent error-free readings in both data sets.

4.4. Estimation of Frequency. The ShoRAH method was tested for its capacity to determine the frequencies of particular clones in a population. Practically, every haplotypes were accurately reconstituted in the non-PCR-amplified data set, and their frequencies estimations were well associated with the grounded reality (Pearson's association coefficient $r = 0{:}\ 92$ for overall haplotypes and $r = 0{:}\ 98$ if outliers, in other words haplotypes with ≥1 mismatched, are eliminated). The PCR-amplified samples have a higher number of flawed haplotype matches and much more discrepant frequency estimations ($r = 0{:}\ 89$ for all and $r = 0{:}\ 97$ for every perfectly matched haplotypes), indicating that both haplotype reconstructing and frequency assessment are more complicated. For several haplogroups, accurate reconstructions and probability assessment were feasible, with frequencies as lower as 0.1 per cent for the PCR samples and 1 per cent for the non-PCR samples. The disparity in resolution could be described by the PCR and non-PCR samples having distinct mean coverage of 1050 and 3000 base pairs per sequencing location, correspondingly. The summation of every actual haplogroups we were capable to identify in every window was usually >99.5 per cent for the PCR data and >97.6 per cent for the non-PCR samples. Figures 3–6 show a comparative of the best precision, recall, F-score, and gain assessment parameters for each classification.

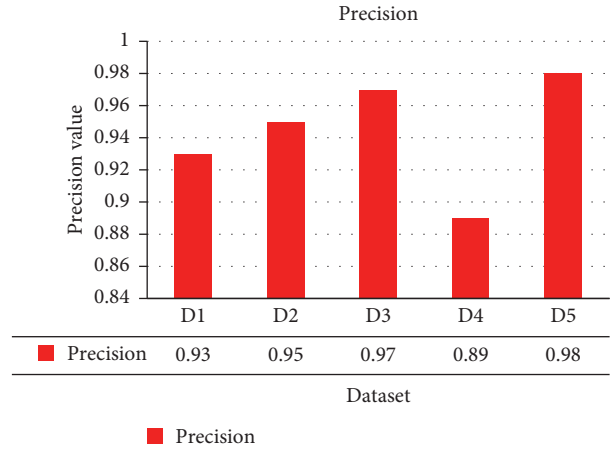


Figure 3: Precision comparison.

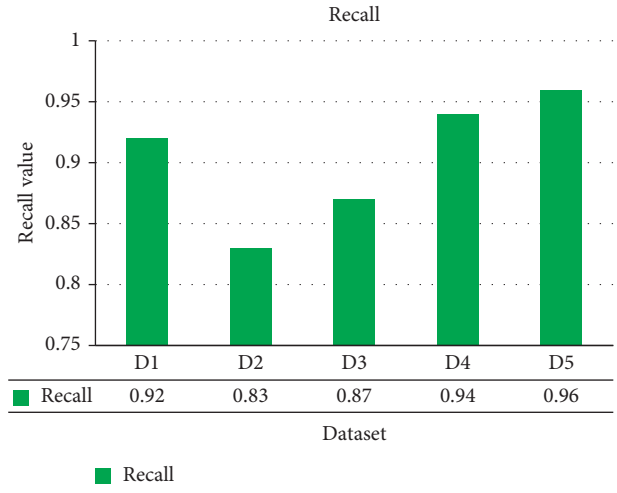


Figure 4: Recall comparison.

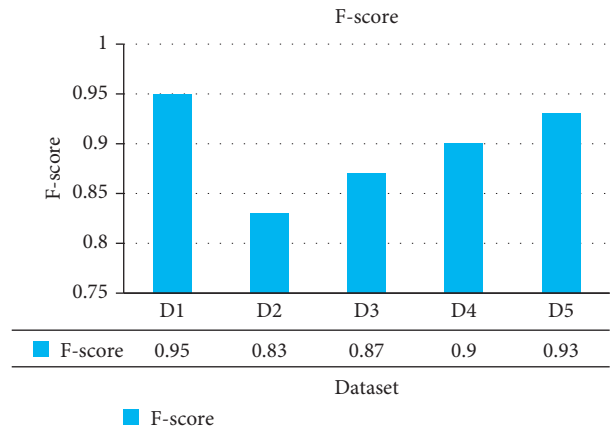


Figure 5: F-score comparison.

5. Discussion

Four of the six approaches tested used various Bloom filter variations to provide for filter compressing, storing of counted values, and representations of mapping in addition to collections [26]. Hash databases were employed in the other existing approaches, which did not produce false positives. Despite the fact that the storage effectiveness of the Bloom filter occurs at the expense of false positives, several

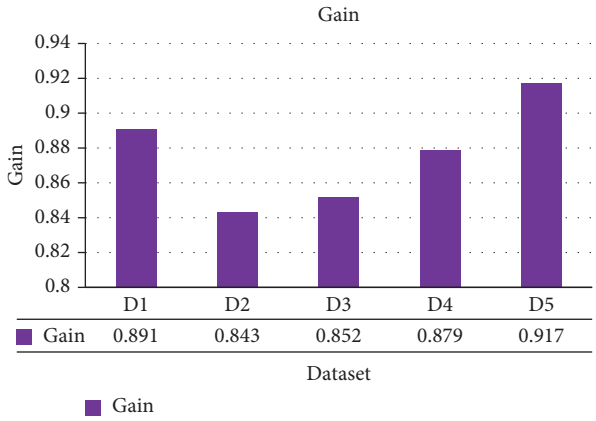


Figure 6: Gain comparison.

significant error correction programmes have decreased or limited false-positive rates by adopting diverse techniques. The developers of these six applications worked hard to increase performance and reduce storage footprints while preserving or enhancing the accuracy of their corrections. Since efficiency and storage are no longer bottlenecking concerns that restrict the implementation of these technologies, we elected to focus entirely on rectification accuracy in this research.

Since correcting efficiency might be tested precisely, generated databases were employed. Only indirect evaluating measures (e.g., genome coverage of de novo assembly and N50 contig size, and percentage of mapping reads in genome alignments) may be obtained for effectiveness analysis when genuine experimental samples are employed. We suggest that using actual databases in instrument assessment could reveal insights that simulated research cannot deliver. However, considerable testing must be carried out on generated databases before going on to actual databases. The researchers of the six instruments studied in this research evaluated them utilizing both synthesized and actual databases. The effectiveness indicators for modelled and actual data sources are highly correlated. When contrasted to various well-known programmes, including as SGA, HiTEC, DecGPU, SHREC, Quake, Coral, and Reptile, Musket was continuously one of the best functioning correctors for all modelled and actual databases. We also showed that Musket outperformed Reptile in terms of efficiency measures. Whenever the developers of Lighter, Bless [27], and Trowel [28] compared their instruments, they reported that theirs outscored Musket by an insufficient margin. Musket, on the other hand, fared just and along with the other three techniques when it came to generated databases. Musket and Bloocoo are very equivalent, particularly in terms of the multistage error correction technique [29]. On a generated database with a 1 per cent error rate at 75 per cent coverage, they apparently obtained equal corrective efficiency as evaluated by recall and precision (refer to "Appendix Materials" in [29]). These two programmes performed similarly effectively on three databases in the existing investigation. Bloocoo, on the other hand, underperformed or collapsed on the three databases with higher reads (110 bp), implying that there are potentially bottleneck concerns in Bloocoo's scripting that restrict its implementation to extended reads.

The task of selecting ideal characteristics is an inherent problem in utilizing every corrector [30]. Only a limited programme feature automatic parametric selection that is responsive to the data sets being analysed. Bless, like the other instruments assessed in this work, may identify a suitable value for M, k-mer multiplicity thresholds, but it could not determine an ideal k (excluding for reptile, that selects $k = log4|G|$, wherein G is the genomic lengths). For every data set, we utilized KmerGenie [25] to find the best k. Although it is likely that the k chosen by KmerGenie is not the best measure including all six applications assessed, we ran a few experiments altering k and other user-defined, tool-specific characteristics and found no significant differences in efficiency measures. For equivalent purposes, we established the read alignments edit distance at 3 (47/67-bp reads) or 5 (110-bp reads) depending on the 5 paper centre length suggestion. As a conclusion, we only presented the outcomes produced with the tool-specific characteristics adjusted to defaults, the selected k values, and the specified modification intervals.

6. Conclusion

Before proceeding with any further downstream in-depth assessment, it is critical to detect and repair flaws in NGS information. The goal of this comparison research was to offer an impartial and unbiased assessment of the impact of five NGS database properties on the effectiveness of k-spectrum-based error correction algorithms, with a focus on reliability. We discovered that genome size, coverage depth, and read length all influenced the effectiveness of the chosen approach.

Data Availability

The data used to support the findings of this study are included in the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest regarding the present study.

Acknowledgments

The author would like to thank the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University.

References

[1] W. J. Ansorge, "Next-generation DNA sequencing techniques," *New Biotech*, vol. 25, no. 4, pp. 195–203, Apr. 2009.

[2] F. Sanger, S. Nicklen, and A. R. Coulson, "DNA sequencing with chain-terminating inhibitors," *Proceedings of the National Academy of Sciences*, vol. 74, no. 12, pp. 5463–5467, 1977.

[3] M. L. Metzker, "Sequencing technologies - the next generation," *Nature Reviews Genetics*, vol. 11, no. 1, pp. 31–46, Jan. 2010.

[4] D. Dressman, H. Yan, G. Traverso, K. W. Kinzler, and B. Vogelstein, "Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations," *Proceedings of the National Academy of Sciences*, vol. 100, no. 15, pp. 8817–8822, 2003.

[5] T. D. Harris, P. R. Buzby, H. Babcock et al., "Single-molecule DNA sequencing of a viral genome," *Science*, vol. 320, no. 5872, pp. 106–109, 2008.

[6] J. Shendure, G. J. Porreca, N. B. Reppas et al., "Accurate multiplex polony sequencing of an evolved bacterial genome," *Science*, vol. 309, no. 5741, pp. 1728–1732, 2005.

[7] Y. Erlich, P. P. Mitra, M. delaBastide, W. R. McCombie, and G. J. Hannon, "Alta-Cyclic: a self-optimizing base caller for next-generation sequencing," *Nature Methods*, vol. 5, no. 8, pp. 679–682, 2008.

[8] J. C. Dohm, C. Lottaz, T. Borodina, and H. Himmelbauer, "Substantial biases in ultra-short read data sets from high-throughput DNA sequencing," *Nucleic Acids Research*, vol. 36, no. 16, p. e105, 2008.

[9] K. E. Holt, J. Parkhill, C. J. Mazzoni et al., "High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi," *Nature Genetics*, vol. 40, no. 8, pp. 987–993, 2008.

[10] S. Ossowski, K. Schneeberger, R. M. Clark, C. Lanz, N. Warthmann, and D. Weigel, "Sequencing of natural strains of *Arabidopsis thaliana* with short reads," *Genome Research*, vol. 18, no. 12, pp. 2024–2033, 2008.

[11] D. A. Wheeler, M. Srinivasan, M. Egholm et al., "The complete genome of an individual by massively parallel DNA sequencing," *Nature*, vol. 452, no. 7189, pp. 872–876, 2008.

[12] Y. Li and J. Wang, "Faster human genome sequencing," *Nature Biotechnology*, vol. 27, no. 9, pp. 820-821, Sep. 2009.

[13] B. Wold and R. M. Myers, "Sequence census methods for functional genomics," *Nature Methods*, vol. 5, no. 1, pp. 19–21, Jan. 2008.

[14] Z. Han, S. Li, and H. Liu, "Composite learning sliding mode synchronization of chaotic fractional-order neural networks," *Journal of Advanced Research*, vol. 25, pp. 87–96, 2020.

[15] S. Ha, L. Chen, and H. Liu, "Command filtered adaptive neural network synchronization control of fractional-order chaotic systems subject to unknown dead zones," *Journal of the Franklin Institute*, vol. 358, no. 7, pp. 3376–3402, 2021.

[16] G. Xue, F. Lin, S. Li, and H. Liu, "Adaptive dynamic surface control for finite-time tracking of uncertain nonlinear systems with dead-zone inputs and actuator faults," *International Journal of Control, Automation and Systems*, vol. 19, no. 8, pp. 2797–2811, 2021.

[17] J.-B. Fan, M. S. Chee, and K. L. Gunderson, "Highly parallel genomic assays," *Nature Reviews Genetics*, vol. 7, no. 8, pp. 632–644, 2006.

[18] M. Pop and S. L. Salzberg, "Bioinformatics challenges of new sequencing technology," *Trends in Genetics*, vol. 24, no. 3, pp. 142–149, 2008.

[19] I. Akogwu, N. Wang, C. Zhang, and P. Gong, "A comparative study of k-spectrum-based error correction methods for next-generation sequencing data analysis," *Human Genomics*, vol. 10, no. S2, p. 20, 2016.

[20] M. Kchouk, J.-F. Gibrat, and M. Elloumi, "An error correction algorithm for NGS data," in *2017 28th International Workshop on Database and Expert Systems Applications (DEXA)*, pp. 84–87, DEXA, Lyon, France, 2017.

[21] P. Skums, D. Zoya, D. S. Campo et al., "Efficient error correction for next-generation sequencing of viral amplicons," *BMC Bioinformatics*, vol. 13, no. S10, p. S6, 2012.

[22] M. S. Fujimoto, P. M. Bodily, N. Okuda, M. J. Clement, and Q. Snell, "Effects of error-correction of heterozygous next-generation sequencing data," *BMC Bioinformatics*, vol. 15, no. S7, p. S3, 2014.

[23] A. Allam, P. Kalnis, and V. Solovyev, "Karect: accurate correction of substitution, insertion and deletion errors for next-generation sequencing data," *Bioinformatics*, vol. 31, no. 21, pp. 3421–3428, 2015.

[24] O. Zagordi, R. Klein, M. Däumer, and N. Beerenwinkel, "Error correction of next-generation sequencing data and reliable estimation of HIV quasispecies," *Nucleic Acids Research*, vol. 38, no. 21, pp. 7400–7409, 2010.

[25] R. Medvedev and P. Medvedev, "Informed and automated k-mer size selection for genome assembly," *Bioinformatics*, vol. 30, no. 1, pp. 31–37, 2014.

[26] L. Song, L. Florea, and B. Langmead, "Lighter: fast and memory-efficient sequencing error correction without counting," *Genome Biology*, vol. 15, no. 11, pp. 509–513, 2014.

[27] Y. Heo, X.-L. Wu, D. Chen, J. Ma, and W.-M. Hwu, "BLESS: bloom filter-based error correction solution for high-throughput sequencing reads," *Bioinformatics*, vol. 30, no. 10, pp. 1354–1362, 2014.

[28] E.-C. Lim, J. Müller, J. Hagmann, S. R. Henz, S.-T. Kim, and D. Weigel, "Trowel: a fast and accurate error correction module for Illumina sequencing reads," *Bioinformatics*, vol. 30, no. 22, pp. 3264-3265, 2014.

[29] Y. Liu, J. Schröder, and B. Schmidt, "Musket: a multistage k-mer spectrum-based error corrector for Illumina sequence data," *Bioinformatics*, vol. 29, no. 3, pp. 308–315, 2013.

[30] X. Yang, K. S. Dorman, and S. Aluru, "Reptile: representative tiling for short read error correction," *Bioinformatics*, vol. 26, no. 20, pp. 2526–2533, 2010.

Hindawi
Computational Intelligence and Neuroscience
Volume 2022, Article ID 1332664, 12 pages
https://doi.org/10.1155/2022/1332664

Research Article

An Enhanced Ant Colony Optimization Mechanism for the Classification of Depressive Disorders

Abed Saif Alghawli [1] **and Ahmed I. Taloba** [2,3]

[1]*Computer Science Department, Prince Sattam Bin Abdulaziz University, Aflaj, Saudi Arabia*
[2]*Department of Computer Science, College of Science and Arts in Qurayyat, Jouf University, Sakaka, Saudi Arabia*
[3]*Information System Department, Faculty of Computers and Information, Assiut University, Assiut, Egypt*

Correspondence should be addressed to Abed Saif Alghawli; a.alghauly@psau.edu.sa

Received 20 May 2022; Accepted 13 June 2022; Published 28 June 2022

Academic Editor: Heng Liu

Bipolar disorder is marked by mood swings that alternate between mania and depression. The stages of bipolar disorder (BD), as one of the most common mental conditions, are often misdiagnosed as major depressive disorder (MDD), resulting in ineffective treatment and a poor prognosis. As a result, distinguishing MDD from BD at an earlier phase of the disease may aid in more efficient and targeted treatments. In this research, an enhanced ACO (IACO) technique biologically inspired by and following the required ant colony optimization (ACO) was utilized to minimize the number of features by deleting unrelated or redundant feature data. To distinguish MDD and BD individuals, the selected features were loaded into a support vector machine (SVM), a sophisticated mathematical technique for classification process, regression, functional estimates, and modeling operations. In respect of classifications efficiency and frequency of features extracted, the performance of the IACO method was linked to that of regular ACO, particle swarm optimization (PSO), and genetic algorithm (GA) techniques. The validation was performed using a nested cross-validation (CV) approach to produce nearly reliable estimates of classification error.

1. Introduction

Information science and data capture systems have made it easier to collect and store large datasets using extended time series. These datasets are becoming more and more popular in a variety of sectors, including astronomy, environment, economics, business, medicine, data analysis, and knowledge mining [1]. The assessment of large datasets is a significant activity in and of itself, and current studies highlight the usage of FS approaches and their talented findings. Seizure identification, sleep state identification, and activity recognition categorization are just a few of the applications where digital analysis of EEG signals is useful. As shown in Figure 1, digital EEG signal treatment consists of three components: a signal acquisition device, an extraction of features unit, and a decision mechanism. The EEG signal collected from a scalp, brain surfaces, or brain internal represents the service's input in Figure 1. Electrodes, either invasive or noninvasive, are used to indicate the signal acquisition component [2]. The extraction of features unit seems to be a signal processing component that extracts distinguishing features from such a stream (s). In a brain-computer interface (BCI), for instance, the decision section is a hybrid component that performs classifications, decision-making, and transmitting choices to remote systems that reflect the subject's desire [3].

Metaheuristic algorithms are often described as combining probability, randomness, and mathematical notation to mimic usual processes. Biological evolutionary processes such as the genetic algorithm (GA) and differential evolution (DE), behavioral genetics such as ACO and particle swarm optimization (PSO), and physically hardening processes such as evolutionary algorithms are examples of these processes (SA) [4]. In recent research, several metaheuristic algorithms and enhanced versions have been effectively applied to a variety of optimization situations. Those

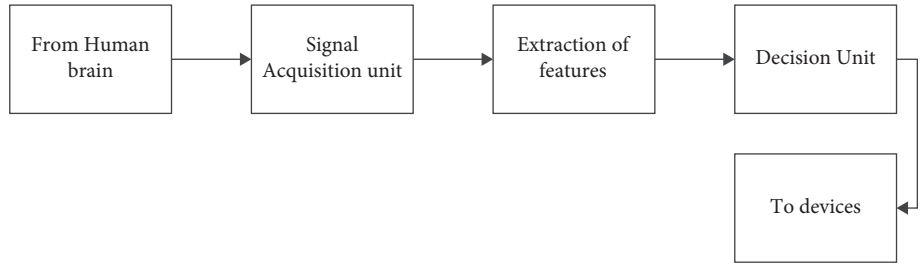


Figure 1: EEG signal processing.

techniques outperformed traditional mathematical solutions in terms of generating better results. ACO is a random optimization approach based on empirical evidence of social behaviors of real animals or insects, but it is an effective algorithm for FS issues between these metaheuristic techniques. Nevertheless, ACO still has several shortcomings in practice. The algorithm's primary criteria are the possibility of becoming caught in a local optimum, the high computing effort and system resources required to get the best solution, and the difficulty of setting the heuristic method to attain high efficiency [5]. As a result, similar enhanced ACO approaches have been developed in recent studies to minimize such faults.

As a result, categorizing MDD and BD at an earlier stage of illness may aid in more efficient and targeted treatment. Neuroimaging approaches for BD and MDD were employed in certain research to demonstrate distinct patterns of structural and functional brain alterations in brain systems involved in emotion regulation [6]. Statistical inference methods were employed in several other investigations, and because those approaches rely solely on the concept of linear combinations, they may not be suited for such assignments. Machine learning approaches have become increasingly popular in the research of anxiety symptoms and comparing these individuals to others with different psychiatric diseases over the last decade. SVM was applied in a recent study to examine the detection accuracy of MDD and BD, and the participants were categorized with an accuracy of 54.76% [7].

Information processing analysis was used in a similar investigation with compartments of the anterior cingulate cortex (ACC) blood circulation in repose. SVM diagnosed MDD and BD patients with 81% accuracy using subgenual ACC blood flow [8]. Another study used multidimensional pattern classification methods and brain morphologic indicators to differentiate the two sad groups with up to 79.3% accuracy. In that other work, the automated identifying technique is used to predict epileptic seizures in multichannel EEG recordings. For feature extraction, approximated probability and statistic estimates were used for both MDD and BD EEG recordings. The effectiveness of SVM with different functions of kernel in detecting seizures is also investigated in this work utilizing MDD and BD EEG signals. The RBF kind kernel SVM algorithm has a predictive performance of 97.17% [9]. Aside from their contributions to diagnosis through great classification accuracy, computerized feature extraction approaches are increasingly being used in the research of anxiety symptoms and assessments of psychiatric disorders. With such a lifetime frequency of up to 4-5%, depression episodes in BD are considered one of the most common psychiatric diseases. While MDD and BD are treated as separate clinical situations with different therapeutic approaches, previous studies found that 60% of BD situations were misdiagnosed as UD and treated wrongly as a result [10].

Consequently, identifying biomarkers that indicate unique pathophysiologic mechanisms in BD and MDD is vital. One of the possible neurophysiological indicators for brain waves is EEG consistency. Different spatial consistency gestures were gathered across large distances as parallelization, and coherence represents the interaction among signals in a certain frequency range [11]. EEG coherence is a unique large number of functional linkages coordinated effective among cortical region pairings, making it a biomarker for the brain's functional connectivity. The clinical importance of cohesiveness as a biomarker in the identification of psychiatric diseases has been highlighted in recent research. Using SVM and IACO, this research intends to expose the distinguishing features of MDD and BD individuals without compromising their accuracy of classification [12]. The levels of coherence, an indicator of functional connectivity within a brain, were first determined utilizing a method that had been previously identified.

Section 2 explains previous work related to this research, and Section 3 lists the topic, EEG data, and procedures for calculating coherence. Section 3.4 describes the FS approach and a hybrid architecture with SVM. Section 4 presents and discusses the results of experiments and validation of performance with current work. Finally, Section 5 concludes the study.

2. Related Work

Motor activity has been the most constant sign of bipolar disorder, according to a growing body of scientific research. Body position, cognitive responsiveness, degree of the psychomotor action, and speech-related motor functions are all examples of motor activity. Self-reported questions with medical observer-rated scales have traditionally been employed in investigations of motor activity in bipolar disorder, which are sensitive and often ineffective. Motor activity data could be used to define phase kind in bipolar patients, which is important because severe sadness and manic episodes can be fatal. This research describes a system

that uses sensing data on phones to identify the status of bipolar illness sufferers. Use real-life activities, collected voice, accelerometer, and self-assessment information from 5 patients more than 12 weeks [13]. The effectiveness of numerous classifiers, the distinct sets of characteristics, and the role of questions in categorizing bipolar disorder events were assessed in this study. This has already shown, for example, that the trajectory of mood disturbances or recurrence in bipolar patients can be classified with high confidence ($\approx 85\%$). To our knowledge, no study has concentrated on observational research of everyday phone conversations to diagnose reduced life performance in bipolar disorder patients. But, there are also plenty of other semisupervised techniques that could be used to enhance such performance in the future [14].

Major depressive disorder (MDD) is the important source of blindness and mortality, affecting around 10% of people worldwide. There are currently no clinically effective diagnostic indicators that can distinguish MDD and bipolar disorder (BD) in an early stage of a period of depression. Exploring machine learning-based integrative indicators of mood disorders is thus an urgent, albeit difficult, need with significant promise to increase the knowledge of such disorders. In this study, we look at some of the most common machine learning algorithms for brain scan prediction and classification, as well as a summary of research that used diagnosis methods imagery information to differentiate MDDs from normal as well as other mood disturbances and look into individual patient diagnostic accuracy determinants. Finally, problems, future trends, and potential limits linked to MDD biomarker development are reviewed, to provide a thorough overview that will help readers in understanding how neuroimaging data gathering might be used to treat depression. Thus, we believe that such efforts would draw attention to the critical want for a medical paradigm change that will drive tailored, optimal patient care [15].

Due to the lack of identified biomarkers, distinguishing between bipolar disorder (BD) and major depressive disorder (MDD) is a significant medical difficulty; thus, a deeper knowledge of their aetiology and brain abnormalities is urgently required. In neuroimaging applications, feature selection is extremely critical due to its intricacy; yet, feature dimension and model understanding are significant issues. A wrapper feature selection method is proposed based on the forward selection method (SVM-FoBa) and used to collect anatomical and resting-state functional magnetic character imaging information from 21 BD, 25 MDD, and 23 healthy controls in this investigation. Discriminative characteristics were extracted between both data modalities, and the categorization of MDD and BD was 92.1% accurate. In adding to the default mode networks and the cerebellum, value assessment of the features extracted suggested that the inferior frontal gyrus, along with the default mode system and the cerebellum, may play a significant role in BD-MDD distinction. When identifying the two clinical illnesses, functional information trumps biological information by a considerable margin, according to a modality-by-modality comparison. This study confirmed the benefits of multimodal collaborative efforts and the efficacy of SVM-FoBa that has the potential to be used to detect a potential biomarker for a variety of mental diseases [16].

Human-computer intelligent interaction (HCII) is a growing branch of study that strives to improve and develop human-computer interaction. As a result, several research works have been carried out to identify human emotion using various techniques including facial expression, voice, galvanic skin response (GSR), and heart rate (HR). However, such strategies have issues with reliability and validity, because people can lie about their feelings and responses. The electroencephalogram (EEG), on the other hand, has proven to be a very successful method of identifying people's emotions since it captures human brain function, which cannot be tricked by voluntary control. Despite the approval of EEG in identifying human reaction, research in this subject is difficult since the EEG signal is nonlinear, complex, and disordered. As a result of these concerns, there is a high-dimensional feature problem with poor classification performance [17]. This paper proposes a novel computational methodology to solve such issues, which consists of three basic stages: (a) extraction of features, (b) feature selection, and (c) classification. To obtain EEG signal characteristics, the discrete wavelet packet transform (DWPT) was utilized, and 204, 800 features from 32 subject-independent subjects were acquired. Meanwhile, as a feature selection method and classification, the genetic algorithm (GA) and least squares support vector machine (LS-SVM) were applied. This computer model is used to distinguish three levels of emotion and attentiveness using the typical DEAP preprocessed EEG dataset. The suggested GA-LS-SVM increased the classification performance to 49.22% and 54.83% for emotion and excitation, respectively, whereas when no feature selection strategy was applied, the classification performance was 46.33 percent and 48.30 percent for emotion and excitation, respectively [9].

With technological advancements, we now have new options for the long-term management of health issues. In psychiatry, there are numerous chances where the diagnosis is based on the patient's historical information as well as their current emotional responses, which adds to the difficulty of differentiating between bipolar illness and borderline disease after analysis. This publication was influenced by previous research that classified illnesses by treating symptoms as a time sequence phenomenon. This work presents a signature-based machine learning strategy for extracting distinctive temporal patterns that can be associated with a given illness. One of the primary criteria used to determine the disorder in this approach is the consecutive quality of the information. This study looks at examples of borderline syndrome which are either biologically passed on from parents or arise from childhood effects of extreme fear and stress. The model is validated using the Alan Turing Institute synthesized signature database within pattern psychiatric repository. The end outcome has an AUC of 0.95, which is higher than the previous result of 0.90 [18].

3. Materials and Methods

3.1. Topics. Consider the patients who registered to a Hospital Department of Psychiatric Outpatient Clinic for Neuropsychiatry and underwent a retrospective assessment. This study included 101 patients who were diagnosed with BD and MDD at the time of their admission. A total of 48 bipolar disease patients (18 men and 30 women) and 58 MDD participants (25 men and 33 women) were compared [19]. Outpatients with a period of depression of BD or MDD who were identified on the Individual Examination Meeting for Axis I Disorders as per the Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV criteria for any a initial analysis of bipolar effective long-term disorder depressive incident or major depressive disorder (SCID-I) were selected.

Individuals with a diagnosis of MDD and a value of at least 8 here on Hamilton Depression Rating Scale (HDRS) 17-item form or focuses with an analysis of the BD depressive episodes and a score with at most 13 here on Young Mania Rating Scale (YMRS) were included. Subjects who had their first period of depression, a chapter with present psychotic features, a background of rapid cycling, a history of mixture occurrences, existing psychiatric comorbidity on axis I, a serious unstable medical illness or neurocognitive disorder, or alcohol and substance violence within the previous 6 months and patients hospitalized for electroconvulsive therapy within the previous 3 months were excluded from the study [20]. Individuals with fewer than four psychiatric admissions were also excluded. This factor was established to assure the diagnosis's long-term consistency. All of the patients were assessed by four psychiatrists with at least five years of clinical experience. There was no assessment of interrater reliability. Subjects with fewer than four healthcare practitioners were also selected. This criterion was established to assure diagnosis consistency throughout time. Out of caution, none of the patients was taking antidepressants at the moment of the EEG recording. QEEG recording was once a standard practice performed before treatment decisions for individuals who sought therapy at a neuropsychiatry hospital. BD patients who were given a mood preservative or a combination of mood stabilizers, on the other hand, had a better outcome. At research screening, individuals had to pass normal laboratory tests, a urine toxicology test, and an electrocardiogram, and they had to be physically steady before enrolling in the research.

3.2. EEG Recording Data. EEG was obtained for 5 minutes in an eyes closed sleeping steady state for all the individuals. Patients were told not to take any medicine for 12 hours well before EEG recording. Qualitative EEG (QEEG) data are gathered from 101 people who were sat in a sound-attenuated, electronically insulated environment in a chair in the living room with eyes closed (waking state situation) to monitor and show the efficiency of coherence [21]. To avoid sleepiness, the technicians maintained the QEEG information during the recordings and re-altered the patients every minute as necessary. In accordance with the international 10/20 system configuration, 19 taped electrodes were implanted over the skull.

The scan LT EEG amplifiers and electrode caps were used to record 3 minutes of eyes closed EEG during rest at a sampling frequency of 250 Hz. An array of 19 sintered Ag/AgCl electrodes were positioned to give a 10/20 international organization signal. From Figure 2, intrahemispheric synchronization was evaluated on the left hemisphere using electrode pairs *F*3–C3 *F*3–P3, F3–T5, *C*3–P3, C3–T5, P3–T5 and then on the right hemisphere using electrode pairs *F*4–C4, F4–P4, F4–T6, C4–P4, *C*4–T6, *P*4–T6. Electrode pairs *F*3–F4, C3–C4, P3–P4, and *T*7–T8 were used to evaluate interhemispheric synchronization. Before artifact removal, the raw EEG data were filtered using a band-pass filter (0.15–30 Hz) and EEG sections with an evident eye and head, and muscular artifacts were carefully eliminated.

3.3. QEEG Biomarkers. The magnitude-squared synchronization function, which is dependent on the Fourier series of the frequencies, f values, was used in traditional EEG spectral analysis. The normalized frequency response per wavelength of two frequencies collected simultaneously at distinct places on the scalp is described as synchronization [22]. For each combination of electrodes, total magnitude-squared coherence $H_{xy}(f)$ is computed as squares of exponent of average cross-power spectral density (PSD) regularized to a product of the mean auto PSD. The coherent frequency for the waveform x and y of an electrode's combination is determined as

$$H_{xy}(f) = \frac{\left|P_{xy}(f)\right|^2}{\left|P_{xx}(f)\right|\left|P_{yy}(f)\right|}, \tag{1}$$

where $P_{xy}(f)$ denotes the cross PSD approximate of x and y, while $P_{xx}(f)$ and $P_{yy}(f)$ denote the PSD estimations of x and y, respectively. The cross-power spectra and the power spectrum (wide range of measures) are described as

$$\begin{aligned} P_{xy}(f) &\coloneqq \left|x''(f)\right|^2 = x''(f)\overline{x''(f)}, \\ P_{xy}(f) &\coloneqq x''(f)\overline{x''(f)}, \end{aligned} \tag{2}$$

where x'' is a complex conjugate of the x:

$$x''(f) \coloneqq \int_{-\infty}^{\infty} x(t)e^{-iwt}\mathrm{d}t. \tag{3}$$

Equation (3) is the Fourier transform which provides data on the frequency which occurs in signals as well as their dominant frequency. The raw EEG signals were split into 650 ms intervals with a 50% overlap, but the settings of every period are displayed using a Hanning window even during the calculating process. The MATLAB application was then used to perform coherent analyses of each participant using 10–14 artifact-free periods [21]. Among the targeted and non-targeted activation situations, the scan frequencies were chosen at random. For long-range intrahemispheric and interhemispheric couples in the delta Δ, theta γ, and alpha α radio frequencies, coherence standards are determined for

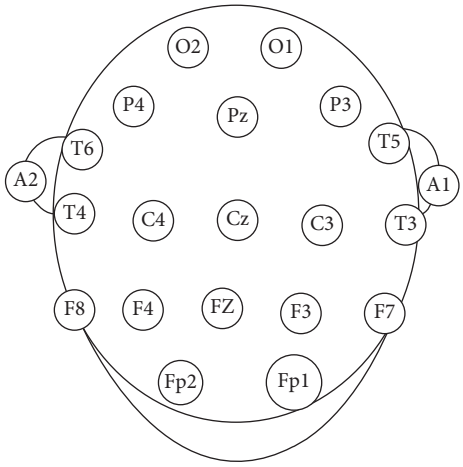


Figure 2: EEG electrodes.

such target and non-target stimulation. Then, Fisher's Z conversion was utilized to standardize the distributions of coherence levels.

3.4. Phase of Feature Selection. This paper offers an FS stage to handle the large dimensional feature storage problem, in which a sample of important characteristics is picked to improve classification performance. In this research, a genetic algorithm is used with an LS-SVM classification [23]. GA is a multidimensional problem-solving technique that, similar to gradient search approaches, avoids the problem of local optima. Furthermore, the GA algorithm has a broad and deep search space exploration capability. This method is inspired by the natural evolving criteria mechanism, which was created to emulate biological behavior to optimize a complex objective function. The optimization was simply achieved by allowing a community composed of many people to progress under specified and specific restrictions to advise that maximizes actual health. GA is used in this research under the MATLAB programming framework. The wrapper decision is built using GA and LS-SVM classifier algorithms [24]. Individuals within a population were assigned binary strings 1 and 0, indicating whether the feature was taken into account and whether it was included or not. However, the classification accuracy is defined also as a fitness function. Table 1 provides how the GA algorithm's choices variables were set up in this investigation.

The FS process is a widely used approach for reducing data dimensions and improving the effectiveness of the learning algorithms. For an FS procedure, the entire search space encompasses all potential subsections of features, and the degree of subsections is computed using the following equation:

$$\sum_{z=0}^{x}\left(\frac{x}{z}\right)=\left(\frac{x}{0}\right)+\left(\frac{x}{1}\right)+\cdots\left(\frac{x}{x}\right)=2^{x}, \tag{4}$$

where n denotes the number of characteristics in the existing subset of features and z is the size of the entire feature subset. FS approaches typically necessitate heuristics or randomized search tactics, which add to the complexity of the resulting group, lowering its degree of optimization problem [25]. Depending on their evaluation approach, FS techniques can be divided into three categories. A filter method is when a procedure conducts FS independently of a learning algorithm, and it usually involves picking feature subsets according to the interclass conditional independence principle. The filter technique is popular for working with high-dimensional data because of its processing efficiency. When a classification method is used to conduct the evaluation stage, the FS method is referred to as a wrapper method.

Table 1: Setup of parameters for GA.

Sl.no	Parameters	Description
1	Scale	0.8
2	Crossover point	2 points
3	Rate of mutation	0.2
4	Values of fitness	Classification of accuracy
5	Patients	String (0, 1)

In a wrapper method, selected structures are placed into a preprogrammed learning study to estimate the subset's effectiveness. In comparison to filters, the prediction accuracy of final specific subset in wrappers is connected with specified relevance measure; however, trade with a large dataset could grow the difficulty owing to the employment of learning techniques in feature subset assessment. To conclude, similarly to wrapper techniques, the FS and learning algorithms are interspersed within the embedded method, and the link between the FS and the classification is deeper; however, wrapper methods cover more of the search space [26]. As shown in Figure 3, wrappers are composed of three main components: a learning machine, feature assessment methods, and an FS technique. Metaheuristic search approaches appear to be significant due to their versatility in randomized searches to assist with the FS procedure since a large selection of attributes may generate high complications. Nature-inspired techniques including genetic algorithm (GA)-based attribute reductions, particle swarm optimization (PSO), and gravitational search algorithm (GSA) are highlighted in recent publications. Besides these techniques, which aim to improve solutions by applying knowledge from earlier iterations, ACO seems to be another promising method for solving combinatorial optimization issues, and it has been frequently used in FS.

3.5. Feature Selection with an ACO Algorithm. In pattern recognition problems, feature selection and dimensional reductions are critical phases. Even though the set of features was not extensive in this investigation and the results were adequate, utilizing the most significant attributes improved the classification performance [27]. The classifier was also able to develop a more efficient approach and obtain higher generalization performance by reducing the number of characteristics. The ACO method was used to obtain an optimal subset of characteristics, and the flowchart of the optimization procedure is shown in Figure 4.

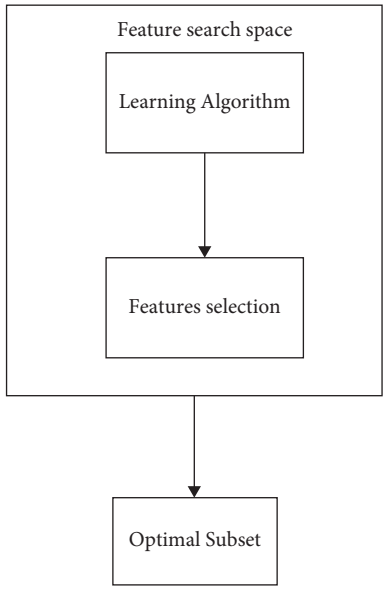


Figure 3: Wrapper-based approach process.

The optimization method and the BPNN classifiers are used to perform the feature extraction optimization processing steps. ACO's chosen features are transmitted to the classifiers, and the resulting model is evaluated with such a test set containing the allocated features. Lastly, the MSE loss is used to assess every ant's performance to modify a pheromone database. The procedure continues to meet the halting criteria, which is specified as a standard error that is as low as possible. ACO is a stochastic metaheuristic for solving complex optimization problems that are continuous. It is inspired by the foraging strategy used by real biological ants when they are trying to discover a simple path between their colony and a source of food [28]. The ants interact informally while hunting using pheromones to indicate their pathways and attract additional ants. Artificial ants produce simulated pheromones to modify their route across the decision network, which would be the route that indicates which alternative nodes an agent will choose, in the ACO algorithm. The volume and intensity of pheromones that the agent utilizes to modify its path are determined by how effective the response is when compared to the previous iteration's competitor species. While making their own decisions to determine the best route of all feasible possibilities, the ensuing representatives employ the pheromone markers of preceding good agents as a reference.

The ACO was initially used for the optimization of the traveling salesman issue because it resembled taking the shortest route to a food supply [29]. A set of cities (nodes) is supplied in this problem, and the distances between them are known. The goal is to determine the quickest route that permits you to visit each city only once. Alternative routes are developed using a probabilistic model, and these trails are stated to build by the artificial ants moving across a network that describes the problem, with each vertex representing a town and every edge representing a connection among two cities, according to the ACO metaphor. Attempts

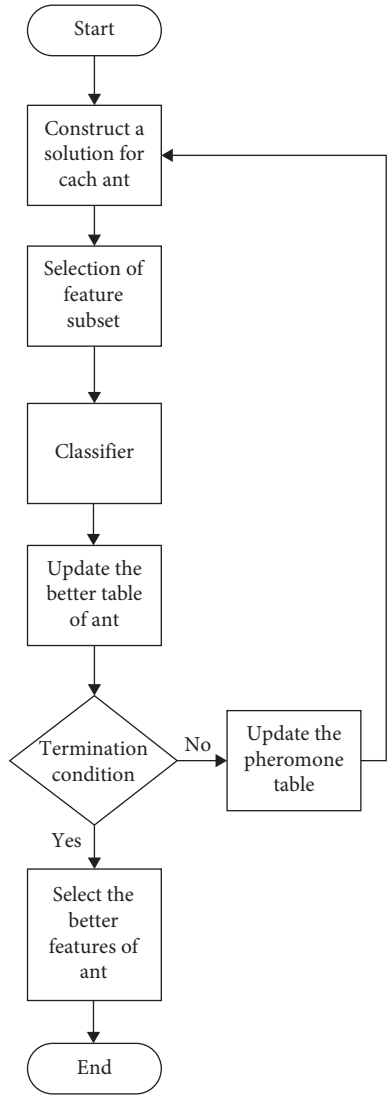


Figure 4: Proposed design of ACO-based parameter of feature selection.

to construct an ACO algorithm were unsuccessful again until the technique was combined with local optimization. One issue is early converging to a suboptimal solution due to the rapid application of too much synthetic pheromone. Pheromone absorption is used to solve this issue. To put it another way, the pheromone linked with solution vanishes later after a certain amount of period. Ants use a stochastic process to choose the next city to visit in the creation of a solution. Assuming ant r is in city p and has completed the partial equation T_s, the chance of it traveling to city q is

$$S_{pq}^{r} = \begin{cases} \dfrac{\tau_{pq}^{\sigma}.\eta_{ps}^{v}}{\sum_{C_{ps}} \in N(T^s)\tau_{pq}^{\sigma}.\eta_{ps}^{v}}, & \text{if} \quad C_{ps} \in N(T^s), \end{cases} \tag{5}$$

where $N(T^s)$ is the quantity of possible nodes, which would be provided as follows: σ and v are the parameters to regulate the relative relevance of the pheromone's vs. the heuristic input η_{pq}.

$$\eta_{pq} = \frac{1}{d_{pq}}, \tag{6}$$

where d_{pq} is the distance between p and q.

During every phase, all of the effective and efficient which have built responses in that iteration modify the pheromones quantities. The following is an update to the pheromone τ_{pq}, which is connected with the edge connecting cities p and q:

$$\tau_{pq} \leftarrow (1-\rho).\tau_{pq} + \sum_{r=1}^{m} \eta \tau_{pq}^{r}, \tag{7}$$

where $\Delta\tau_{pq}^{r}$ is the volume of pheromones placed on edge (p, q) by ant r, ρ is also the evaporation speed, and m represents the number of ants.

Here,

$$\Delta\tau_{pq}^{r} = \begin{cases} \dfrac{\delta}{Sq}, & \text{if ant used edge } (p, q), \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where S_q is the duration of a tour built by ant q, and δ is a variable.

3.6. Support Vector Machines. A training set of values is given as $Z = \left\{(p_i, q_i) | a_i \in H, q_i \in \left\{ \begin{matrix} + \\ - \end{matrix} 1 \right\}, i = 1, 2, \ldots, s \right\}$, where p_i represents the input variables and q_i represents the x_i label; the objective purpose is determined as

$$\begin{cases} \min \varphi, & (W) = \dfrac{1}{2} W.W + C \sum_{i=1}^{s} \delta_i, \\ \text{subject to,} & q_i (W.\varphi(x_i)) + b \geq 1 - \delta_i, \delta_i \geq 0 \quad i = 1, 2, \ldots, s, \end{cases} \tag{9}$$

where W is the hyperplane normal variable and C is a penalty ratio that regulates the trade-off between maximization of a valuation width and minimization of the number of misidentified instances in the training collection to 10 [30]. The width of the kernel is controlled by the hyperparameter δ_i which is set at 0.2. Lastly, the ideal hyperplane is written as the quadratic function in the following equation:

$$\begin{cases} \max S(\alpha) = \sum_{i=1}^{s} \alpha_i - \dfrac{1}{2} \sum_{ij} \alpha_i \alpha_i q_j q_j S(p_i p_j), \\ \text{subject to} \sum_{i=1}^{s} \alpha_i \quad q_i = 0, 0 \leq \alpha_i \leq C, \; i = 1, 2, \ldots, s. \end{cases} \tag{10}$$

The output data can be stated in the following equation as

$$f(p) = \text{sign}\left[\sum_{j=1}^{s} \alpha_i q_i S(p_i, p) + b \right]. \tag{11}$$

Different kernel functions might be employed in the decision function based upon the data. The most commonly used kernel functions include polynomial, radial basis function (RBF), and linear, which are all listed in the following equation:

$$S(p, p_i) = \exp\left(-\frac{p - p_i^2}{2\sigma^2} \right). \tag{12}$$

In this investigation, 3 kinds of kernels have been used, and the RBF kernel was chosen because of its superior performance, as given in Table 2.

3.7. IACO-SVM Proposed Techniques. In the proposed method, IACO processes the inputs as a specified feature subset, while the outputs are categorized as psychiatric illness types. To accomplish this, all characteristics of alpha, delta, and theta coherence values are given in the FS step, and more relevant characteristics are picked using a metaheuristic approach involving artificial ants [31]. The fitness value well over the SVM classifier can then be utilized to evaluate the performance of the various feature subsets. The pheromone updating process is started based on the fitness value, and heuristic parameters are changed as a result. The modeling procedure repeats itself until the halting requirement is met. The classification performance is evaluated after each modeling session using exterior testing data which is novel to the system.

A nested-CV is made up of two-layered cross verifications, exterior one and interior one. To begin, the dataset is divided into six pieces, each with a 6-fold outer-CV and sampling methods. Varies from one person fold is set aside for the outer-CV test, the remaining 5 folds are saved for the inner-CV cycle's training phase. A 5-fold CV method is used in the inner-CV cycle, and 5 models are generated in the end [32]. The classification performance of the models using the specified subsets of features was then ordered. The greatest of those 5 models is evaluated using restricted outer-CV assessment folds which are totally unfamiliar to the classifiers to minimize leaking from testing and train the model.

During the interior CV loop, the conserved outer-CV test folds are exchanged with a few of five training folded, allowing the inner-CV models to be tested with entirely different test data for every external CV cycle. Inner-CV training data is used to identify features utilizing IACO in internal CV. Four folds are utilized for training and one for verification in the IACO method. The aforementioned procedures are performed five extra times to obtain the local best system for each loop, as the data was separated into six pieces in outer-CV. Last, the arrays are sorted by classification results to find the globally highest of six regionally best systems with their optimized subset of features.

3.8. Fitness Value. The level of usefulness of a specific subset is calculated using a fitness function. If two subgroups with various numbers of characteristics perform similarly well in a classification issue, the subset with the fewer characteristics takes the lead [33]. As a result, the fitness value is evaluated

```
Initialize ACO
Repeat
    Set the parameters of heuristics as v and σ
        For ant r ∈ {1, 2, ..., m}
            Select a set of features utilizing rule of probabilistic
            Measure the value of fitness
            Update the limited pheromone
        End for
    Assess the solutions produced by all the ants in a tour
    choose the tour's greatest and worst performers.
    Update the pheromone globally
    The rate of evaporation reduces pheromone intensity
        If after the tour a shorter route is discovered
        Update the parameter of heuristics v and σ
        End if
Until the criteria for stopping are met
Choose the attributes that have the greatest pheromone levels
```

Algorithm 1: ACO algorithm.

Table 2: SVM classifier utilizing three distinct kernel types.

Types of kernels	Accuracy (%)
Polynomial	58.42
Linear	56.42
Radial basis function (RBF)	62.38

in terms of two factors: classification accuracy and the number of components in the subset. To address these problems, the fitness value is meant to be as accurate as possible while also having a large variety of features in the following equation:

$$f\left(a_j\right) = m^* J\left(A_j\right) + n^* \left(\frac{1}{\left|A_j\right|}\right), \tag{13}$$

where A_j is the j^{th} ant's subset, $J(A_j)$ is the classifying accuracy utilizing (A_j), $|A_j|$ is the variety of attributes in A_j, and $m \in [0,1]$ and $n \in [0,1]$ are two variables used to provide relative priority to organization performance and the number of subset strictures. Since classification performance is crucial in comparison to the number of characteristics, we selected $m = 0.92$ and $n = 0.78$ in the analysis. The fitness values are computed by replacing the coefficients within a solution, as given in Table 3.

3.9. Analysis of Complexity

3.9.1. Time Complexity. To construct a model with fewer but more useful features, the hybrid strategy integrating IACO and SVM algorithms was used in this research [34]. Using the values in (14), the computational difficulty of the suggested method in terms of time may be calculated:

$$N_{\text{iter}} \times \left(N_{\text{Ants}} \times \left(T_{\text{selection of features}} + T_{\text{SVM}}\right) + T_{\text{Updating the pheromne}}\right). \tag{14}$$

The number of iterations is N_{iter}, while the number of ants within every iteration is N_{Ants}. $T_{\text{selection of features}}$ is the time it takes an ant to produce a subset of features, T_{SVM} learning is the time it takes an ant to train an SVM classifier using selected features, which expands under a particular $(O(n^3))$ with the samples in the training set (n), and $T_{\text{Updating the pheromne}}$ is the time it takes an ant to upgrade the pheromone table after a model has been produced.

The fitness value calculation method, described in (13), is performed for all iterations, as shown in Figure 5. The effectiveness of the methods improves as the number of iterations grows. The convergence efficiency of both the ACO and IACO approaches was investigated by running the algorithms ten times [35–37]. In Figure 5, the average values of the minimal fitness functions of those 10 sessions are shown against the number of iterations. It is clear that the suggested method, IACO, converges faster and more accurately than the classic ACO algorithm. In addition, 20 trials were conducted to evaluate the average execution duration of regular ACO, PSO, GA, and IACO, the outcomes of which are given in Table 4.

3.10. Complexity of Space. A system's space complexity is measured by the amount of memory used across the entire procedure. The algorithm's memory footprint is estimated utilizing both fixed and variable storage. The variables utilized in the program take up a set amount of space, whereas the component whose size is determined by repetitions and recurrent operations takes up a variable amount of space [38–40]. Two space characteristics are taken into account while determining the overall amount of storage used against the algorithm. Factors, data structures, allotted memory, and other data components are used to describe the quantity of subspace. The space difficulty is determined as in the following equation, where the magnitude of the challenge is m and the number of ants is n:

$$S(m) = S(m) = O\left(4m^2\right) + O(2mn). \tag{15}$$

Table 3: FS algorithm performance.

FS techniques	No. of features	Sensitivity	Accuracy	Value of fitness	AUC
None	47	0.637	62.38	0.591	0.632
GA	25	0.8	75.25	0.725	0.778
PSO	26	0.783	73.27	0.706	0.740
ACO	26	0.837	78.22	0.751	0.780
IACO	23	0.855	80.18	0.774	0.794

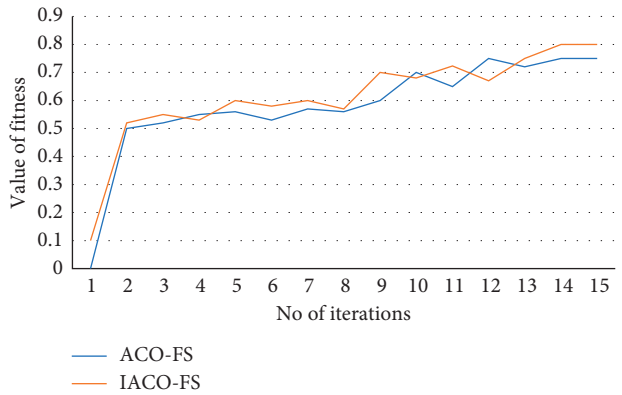


Figure 5: With different numbers of iterations, the ACOFS and IACOFS techniques have different fitness values.

Because the issue dimension should be the same for IACO and simple ACO, the case complexity will be the same.

4. Result and Discussion

A series of tests are carried out in this work to demonstrate the efficacy of the suggested FS method. SVM with a selection of features extracted was used to classify 46 BD and 55 MDD cases. The genetic algorithm (GA), particle swarm optimization (PSO), and ACO and IACO algorithms were used in the FS process. The heuristics variables for IACO were changed during the optimization phase to mitigate general ACO limitations such as earlier stagnation and slow convergence speed. Twelve interhemispheric and four intrahemispheric QEEG synchronization measures from alpha, delta, and theta bandwidths make up the initial feature set. To train and test the SVM classifier, nested-CV was used. This used ROC and area under the curve (AUC) analyses for every combined and standalone classification model because ROC analysis is a useful tool for displaying how classifications and threshold selections work. Figure 6 shows the effectiveness of each technique.

The suggested approach, IACO-SVM, improves GA-SVM, PSO-SVM, ACO-SVM, and standalone SVM classifiers, as shown in the figure. The number of characteristics, accuracy rate, sensitivity, and AUC levels of classifiers with different FS approaches are given in Table 5 to stress the importance of dynamic heuristics utilized in IACO.

Table 4: The average execution duration of the FS algorithm.

Techniques	Average execution duration
GA	21 minutes, 13 sec
PSO	19 minutes, 39 sec
Standard ACO	23 minutes, 8 sec
IACO	17 minutes, 12 sec

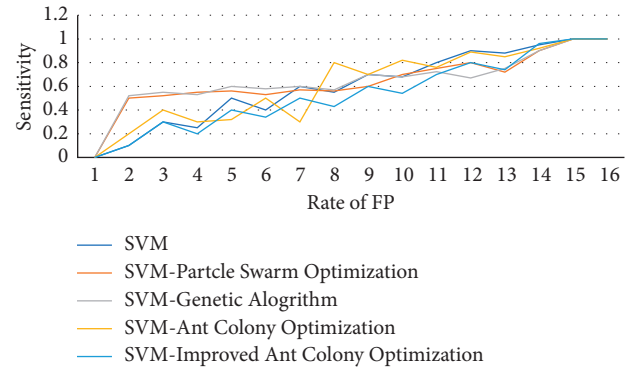


Figure 6: The effectiveness of each technique.

The efficiency, sensitivity, and AUC ratings of the standalone SVM classifier are not sufficient, even though it has more parameters than classifiers using FS techniques. With limited parameters, assigning FS techniques to the classifier improved classification results by 73.26%. According to IACO's actual performance, variable heuristics variables boost the accuracy of classification better. As a result, we improved regular ACO and saw a significant rise in both total classification results and AUC value, highlighting the necessity of giving relevant attributes to a model. Table 6 also includes the specific characteristics that lead to the classification results, as well as a visualization of the resulting extracted features using Table 6.

Only a few studies have looked into the differences in brain connectivity between MDD and BD in terms of clinical translation of a processing variable in the IACO-SVM model used in this research. Using univariate approaches, research group discovered a paucity of frontal interhemispheric alpha connection in BD patients in comparison to MDD patients in such a new analysis. The study found that BD patients had more alpha activity in the contralateral temporoparietal

Table 5: FS algorithm performance.

FS technique	Quantity of features	Sensitivity	Fitness rate	Accuracy (%)	AUC
GA	25	0.800	0.725	75.25	0.772
PSO	26	0.783	0.704	78.22	0.738
ACO	25	0.837	0.751	78.22	0.778
IACO	21	0.855	0.794	80.18	0.792
None	49	0.637	0.591	62.37	0.632

Table 6: Selection of feature subset by using IACO.

Δ – frequency band	γ – frequency band	α – frequency band
F3-F4, *P3-T5*, *P4-T6*, *F3-P3*,	*F3-C3*, *F3-T5*, *P4-T6*, *C4-P4*, *C4-T6*, *C3-C4*, *T5-T6-C3-P3*,	*C4-P4*, *F3-P3*, *F3-T5*, *C3-P3*, *F4-C4-F6-T6*, *P4-T6*, *C3-C4*, *P3-P4*

areas than UD patients. Furthermore, a study conducted argued that the absence of interhemispheric synchronization in the slow-wave frequency ranges could be one of the characteristics distinguishing BD from UD. In contrast to the previous studies, delta frequency range was included in the results.

The explanation seemed to be that the delta frequency range was linked to cognitive performance, which was found to be worse in bipolar disorder patients than in those with significant depression. Second, specific changes in the delta band were observed during the sleep of depressive individuals, suggesting that the delta band may be used as a control band because recordings were taken in an eyes closed, nonsleep state. This could explain why delta frequency contributed fewer characteristics in the current investigation. Furthermore, it was discovered that the alpha band seems to have more characteristics than the beta and delta bands. In EEG examinations of depressed people, the most common area which has been observed to be changed is the alpha frequency spectrum.

5. Conclusion

An IACO-SVM algorithm is presented in depth in this study. The suggested framework is comprised of two steps: FS and categorization. In summary, research has discovered an aberrant frontal alpha asymmetry that impairs behavioral approach avoiding inclinations, which can lead to sadness. Finally, changes in anxiety symptoms may be due to greater activity in the beta frequency spectrum, but this may vary across the patients in this study. This could explain why the beta frequency spectrum has fewer features than the alpha frequency range. Taken together, the criteria used in this study are usually complementary to studies that demonstrated group-level differences between UD and BD. Anyway, the final goal of studying the differences between psychiatric conditions is to find molecular markers that seem to be unique to every condition. Univariate statistics, in particular, avoid making conclusions on an individual basis, making them less helpful for clinical usage. In this case, SVM-based techniques are critical since they allow for individual-based predictions as well as the specificity and sensitivity levels of the selected variables.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] M. A. AbdelAal, A. A. Alsawy, and H. A. Hefny, "EEG-based emotion recognition using a wrapper-based feature selection method," in *Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2017*, A. E. Hassanien, K. Shaalan, T. Gaber, and M. F. Tolba, Eds., vol. 639, pp. 247–256, Springer International Publishing, Cham, 2018.

[2] T. T. Erguzel, S. Ozekes, S. Gultekin, and N. Tarhan, "Ant colony optimization based feature selection method for QEEG data classification," *Psychiatry Investigation*, vol. 11, no. 3, p. 243, 2014.

[3] M. A. A. Al qaness, A. A. Ewees, H. Fan, A. M. AlRassas, M. A Elaziz, and M. Abd Elaziz, "Modified aquila optimizer for forecasting oil production," *Geo-Spatial Information Science*, pp. 1–17, 2022.

[4] E. Hopper and B. C. H. Turton, "An empirical investigation of meta-heuristic and heuristic algorithms for a 2D packing problem," *European Journal of Operational Research*, vol. 128, no. 1, pp. 34–57, 2001.

[5] T. Tekin Erguzel, C. Tas, and M. Cebi, "A wrapper-based approach for feature selection and classification of major depressive disorder-bipolar disorders," *Computers in Biology and Medicine*, vol. 64, pp. 127–137, 2015.

[6] T. Alotaiby, F. E. A. El Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for EEG signal processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, no. 1, p. 66, 2015.

[7] D. Yousri, A. M. AbdelAty, M. A. A. Al-qaness, A. A. Ewees, A. G. Radwan, and M. Abd Elaziz, "Discrete fractional-order Caputo method to overcome trapping in local optima: manta Ray Foraging Optimizer as a case study," *Expert Systems with Applications*, vol. 192, Article ID 116355, 2022.

[8] N. Mahendran, P. M. Durai Raj Vincent, K. Srinivasan, V. Sharma, and D. K. Jayakody, "Realizing a stacking

generalization model to improve the prediction accuracy of major depressive disorder in adults," *IEEE Access*, vol. 8, pp. 49509–49522, 2020.

[9] F. K. Ahmad, A. Y. Awwad Al-Qammaz, and Y. Yusof, "Optimization of least squares support vector machine technique using genetic algorithm for electroencephalogram multi-dimensional signals," *Jurnal Teknologi*, vol. 78, pp. 5–10, 2016.

[10] A. I. Taloba, I. A. Mohamed, A. Ben Aissa, and L. F. Hussein, *IoT enabled modulated residential surveillance system using Fuzzy logic*, Materials Today Proceedings, Sciencedirect, Amsterdam, Netherlands, 2021.

[11] M. Zelenina and D. Prata, "Machine Learning with Electroencephalography Features for Precise Diagnosis of Depression Subtypes," 2019, https://arxiv.org/abs/1908.11217.

[12] W. L. Toh, N. Thomas, and S. L. Rossell, "Auditory verbal hallucinations in bipolar disorder (BD) and major depressive disorder (MDD): a systematic review," *Journal of Affective Disorders*, vol. 184, pp. 18–28, 2015.

[13] I. Ahmed, "Taloba "an artificial neural network mechanism for optimizing the water treatment process and desalination process," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9287–9295, 2022.

[14] A. Maxhuni, A. Muñoz-Meléndez, V. Osmani, H. Perez, O. Mayora, and E. F. Morales, "Classification of bipolar disorder episodes based on analysis of voice and motor activity of patients," *Pervasive and Mobile Computing*, vol. 31, pp. 50–66, 2016.

[15] S. Gao, V. D. Calhoun, and J. Sui, "Machine learning in major depression: from classification to treatment outcome prediction," *CNS Neuroscience & Therapeutics*, vol. 24, no. 11, pp. 1037–1052, 2018.

[16] N. F. Jie, H. Z Mao, Y. M Xiao et al., "Discriminating bipolar disorder from major depression based on SVM FoBa: efficient feature selection with multimodal brain imaging data," *IEEE Trans. Auton. Mental Dev*, vol. 7, no. 4, pp. 320–331, 2015.

[17] M. Elloumi, M. A. Ahmad, A. H. Samak, A. M. Al-Sharafi, D. Kihara, and A. I. Taloba, "Error correction algorithms in non-null aspheric testing next generation sequencing data," *Alexandria Engineering Journal*, vol. 61, no. 12, pp. 9819–9829, 2022.

[18] S. Ghosh, K. Pahuja, J. M. Jiji, A. Puthussery, S. Shukla, and A. Unal, "Classifying bipolar personality disorder (BPD) using long short-term memory (LSTM)," in *Data Science and Security*, D. S. Jat, S. Shukla, A. Unal, and D. K. Mishra, Eds., vol. 132, pp. 169–176, Springer Singapore, Singapore, 2021.

[19] J. P. de Oliveira, A. Jansen, T. d AzevedoCardoso et al., "Predictors of conversion from major depressive disorder to bipolar disorder," *Psychiatry Research*, vol. 297, Article ID 113740, 2021.

[20] S. Isomura, O. Toshiaki, R. Tsuchimoto et al., "Differentiation between major depressive disorder and bipolar disorder by auditory steady-state responses," *Journal of Affective Disorders*, vol. 190, pp. 800–806, 2016.

[21] X. Chen and Z. J. Wang, "Design and Implementation of a Wearable, Wireless EEG Recording System," in *Proceedings of the if the 5th International Conference on Bioinformatics and Biomedical Engineering*, pp. 1–4, Wuhan, China, May 2011.

[22] A. M. Hunter, B. O. Muthén, I. A. Cook, and A. F. Leuchter, "Antidepressant response trajectories and quantitative electroencephalography (QEEG) biomarkers in major depressive disorder," *Journal of Psychiatric Research*, vol. 44, no. 2, pp. 90–98, 2010.

[23] E. Yılmaz and Ç. Kılıkçıer, "Determination of fetal state from cardiotocogram using LS-SVM with particle swarm optimization and binary decision tree," *Computational and Mathematical Methods in Medicine*, vol. 2013, Article ID 487179, 2013.

[24] C. Qian, X. Yin, and Q. Ouyang, "Modeling and parameter identification of the MR damper based on LS SVM," *International Journal of Aerospace Engineering*, vol. 2021, Article ID 6648749, 2021.

[25] H. Amazal and M. Kissi, "A New Big Data Feature Selection Approach for Text Classification," *Scientific Programming*, vol. 2021, Article ID 6645345, 2021.

[26] H. W. Chang, Y. H. Chiu, H. Y. Kao, C. H. Yang, and W. H. Ho, "Comparison of classification algorithms with wrapper-based feature selection for predicting osteoporosis outcome based on genetic factors in a Taiwanese women population," *International journal of endocrinology*, vol. 2013, Article ID 850735, 2013.

[27] E. Saraç and S. A. Özel, "An ant colony optimization-based feature selection for web page classification," *The Scientific World Journal*, vol. 2014, Article ID 649260, 2014.

[28] L. G. Fahad, S. F. Tahir, W. Shahzad, M. Hassan, H. Alquhayz, and R. Hassan, "Ant colony Optimization-Based Streaming Feature Selection: An Application to the Medical Image Diagnosis," *Scientific Programming*, vol. 2020, Article ID 1064934, 2020.

[29] H. Chen, W. Jiang, C. Li, and R. Li, "A Heuristic Feature Selection Approach for Text Categorization by Using Chaos Optimization and Genetic Algorithm," *Mathematical problems in Engineering*, vol. 2013, Article ID 524017, 2013.

[30] M. S. Uzer, N. Yilmaz, and O. Inan, "Feature Selection Method Based on Artificial Bee colony Algorithm and Support Vector Machines for Medical Datasets Classification," *The Scientific World Journal*, vol. 2013, Article ID 419187, 2013.

[31] S. Rongali and R. Yalavarthi, "Parameter optimization of support vector machine by improved ant colony optimization," in *Proceedings of the Second International Conference on Computer and Communication Technologies*, pp. 671–678, Andhra Pradesh, india, September2016.

[32] S. S. Baboo and S. Sasikala, "Multicategory classification using an Extreme Learning Machine for microarray gene expression cancer diagnosis," in *Proceedings of the International Conference on Communication Control and Computing Technologies*, pp. 748–757, Nagercoil, India, December 2010.

[33] G. De Jong, "The fitness of fitness concepts and the description of natural selection," *The Quarterly Review of Biology*, vol. 69, no. 1, pp. 3–29, 1994.

[34] A. F. Bourke, "The validity and value of inclusive fitness theory," *Proceedings of the Royal Society B: Biological Sciences*, vol. 278, no. 1723, pp. 3313–3320, 2011.

[35] L. Wang, J. Cai, M. Li, and Z. Liu, "Flexible Job Shop Scheduling Problem Using an Improved Ant colony Optimization," *Scientific Programming*, vol. 2017, Article ID 9016303, 2017.

[36] A. A. Mukheimer, "Some Common Fixed-point Theorems in Complex Valued-Metric Spaces," *The Scientific World Journal*, vol. 2014, Article ID 587825, 2014.

[37] M. H. R. Khalaf, Z. M. Abdel Azim, H. Walid, A. H. Elkhateeb, O. R. Shahin, and A. I. Taloba, "Explore the E-learning management system lower usage during COVID-19 pandemic," *Information Sciences Letters*, vol. 11, no. 2, pp. 537–548, 2022.

[38] E. K Amir, O. R. Shahin, M. Rasha, E. A Abd, and A. I. Taloba, "Integration of computer vision and natural language processing in multimedia robotics application," *Information Sciences Letters*, vol. 11, no. 3, pp. 765–775, 2022.

[39] Z. Han, S. Li, and H. Liu, "Composite learning sliding mode synchronization of chaotic fractional-order neural networks," *Journal of Advanced Research*, vol. 25, pp. 87–96, 2020.

[40] S. Ha, L. Chen, and H. Liu, "Command filtered adaptive neural network synchronization control of fractional-order chaotic systems subject to unknown dead zones," *Journal of The Franklin Institution*, vol. 358, no. 7, pp. 3376–3402, 2021.