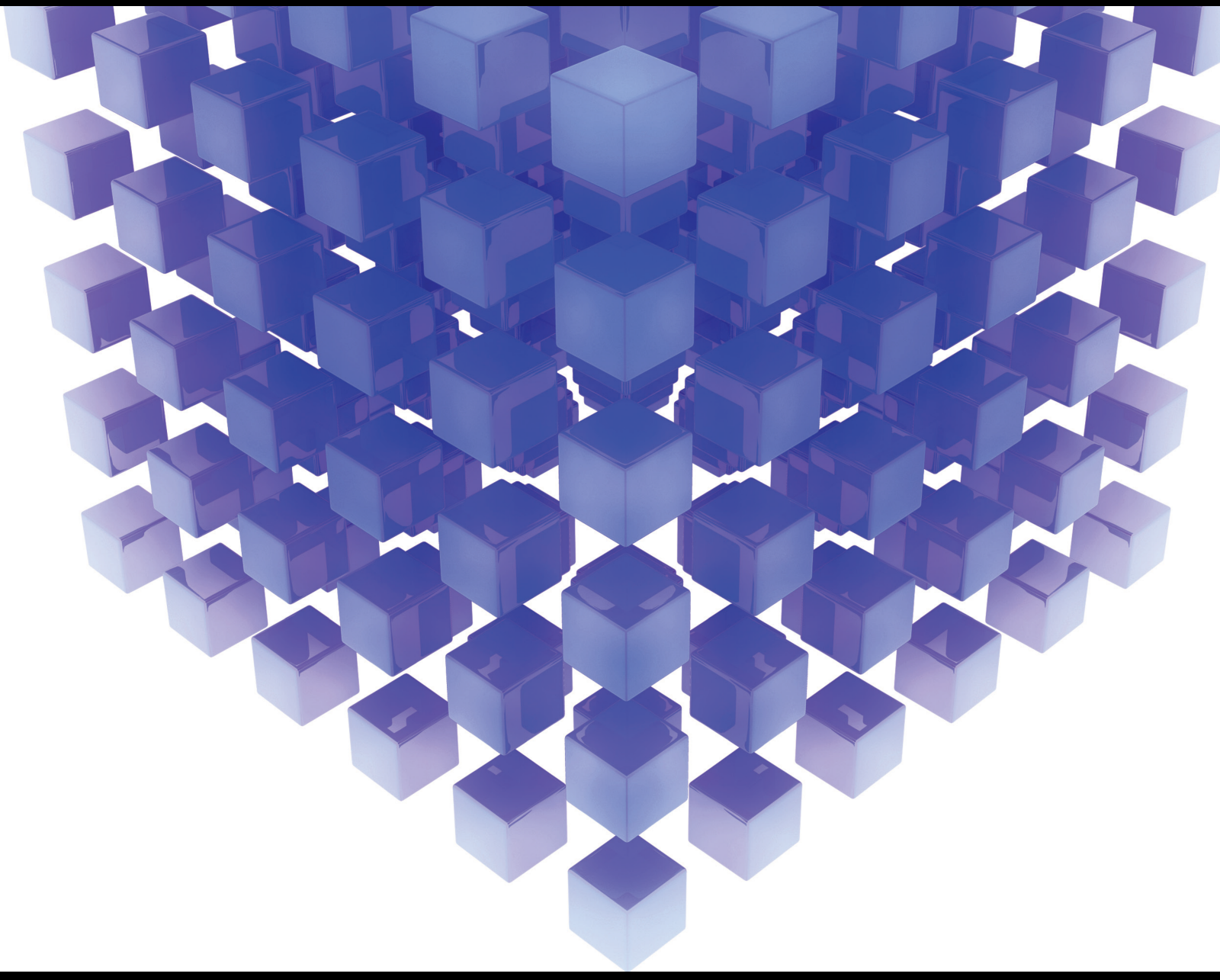
Mathematical Problems in Engineering

Fuzzy Based Many-Objective Evolutionary Algorithms for Non-Linear Optimization Problems

Lead Guest Editor: Mukesh Soni
Guest Editors: Hemant Ghayvat and Bharat Bhushan



Fuzzy Based Many-Objective Evolutionary Algorithms for Non-Linear Optimization Problems

Mathematical Problems in Engineering

Fuzzy Based Many-Objective Evolutionary Algorithms for Non-Linear Optimization Problems

Lead Guest Editor: Mukesh Soni
Guest Editors: Hemant Ghayvat and Bharat Bhushan

Chief Editor

Academic Editors

Contents

Hindawi
Mathematical Problems in Engineering
Volume 2023, Article ID 5268340, 13 pages
https://doi.org/10.1155/2023/5268340

Research Article

Multicriteria Ordered the Profile Clustering Algorithm Based on PROMETHEE and Fuzzy c-Means

Muhammad Adnan Bashir,[1] **G. Muhiuddin**,[2] **Tabasam Rashid**,[1] **and Muhammad Shoaib Sardar**[3]

[1]*University of Management and Technology, Lahore 54770, Pakistan*
[2]*Department of Mathematics, Faculty of Science, University of Tabuk, P.O. Box 741, Tabuk 71491, Saudi Arabia*
[3]*School of Mathematics, Minhaj University Lahore, Lahore, Pakistan*

Correspondence should be addressed to G. Muhiuddin; chishtygm@gmail.com

Received 23 June 2022; Accepted 22 July 2022; Published 9 May 2023

Academic Editor: Mukesh Soni

The purpose of multicriteria clustering is to locate groups of alternatives that have comparable qualities and have been examined across multiple criteria. An ordered profile clustering is a well-known problem, and the fuzzy c-means clustering (FCM) technique is one of the most broadly used in every field of life. At present, FCM is for the partitioning of information into numerous clusters which are still lacking priority relations. To address the problem of finding ranking in clusters based on multicriteria in the fuzzy environment, we propose a multicriteria ordered clustering algorithm based on the partial net outranking flow of the preference organization for enrichment evaluations method (PROMETHEE) and fuzzy c-means. Lastly, we apply the proposed algorithm to solve a real-world targeted clustering problem regarding the human development indexes. To test the efficacy of the proposed algorithm, a comparative analysis of ordered K-means clustering (OKM) and FCM is carried out with it.

1. Introduction

A classic challenge in multicriteria clustering is supervised categorization or allocating options to predefined classes. To be more specific, this subject has gotten a lot of interest from sectors like machine learning in every field [1], pattern recognition clustering [2], data retrieval [3], data mining [4], the clinical [5], sales [6], health [7], operational systems [8], and humanistic systems performance evaluation [9]. The primary concept of clustering, according to the clustering goal, is to divide a collection of data into a specified number of clusters (categories, groups, or subsets) with high similarity within each cluster.

Furthermore, several researchers have created innovative and outstanding methodologies in the framework of MCDA, including ELECTRE-SORT and TRI [10], Flowsort, a clustering algorithm [11], PROAFTN [12], UTADIS [13], PAIRCLASS [14], and so on. These approaches, in general, presuppose that the classifications are specified prior to a collection of options and their central or limit profiles.

The difficulty with supervised classification is that the groups are occasionally unknown a priori, i.e., no data structure information has been provided. Various solutions have recently been developed based on a multicriteria approach to deal with the difficulty of classifying related possibilities into undefined categories [15]. This area highlights three sorts of problems in clustering: relational, nonrelational Boujelben [16] and ordered Meyer and Olteanu [17]. Complete ordered clustering, in particular, may aid in the formation of targeting ordered clusters and the establishment of priority linkages among a subset of alternates for each cluster, making it an advantageous addition to the ranking process. Furthermore, multicriteria ordered clustering enables the display of created ranks in terms of their complexity, such as rankings of the foreign universities, economy, and the human development index (HDI), among other things.

Even though several statistical and mathematical methods, such as cluster analysis [18] and regression trees [19], may be used to find ordered clusters in a multicriteria environment, they have substantial drawbacks, such as the mischaracterization of independent/dependent variables and the removal of critical aspects [20]. To rank groupings of alternatives regarding a country's risk, De Smet et al.'s [21] initially created an enhanced and optimized PROMETHEE approach. Then, using the inconsistency matrix and the relations between pairwise preferences, they improved on their previous work and proposed an accurate technique for identifying a complete ordered partition [20]. On the other hand, De Smet et al.'s techniques partially used the structure of the data and measured the ordinal profile of the paired preference degrees [22].

Consequently, Chen et al. [22] have created an ordered K-means clustering algorithm (OKM) by combining the PROMETHEE method's partial net outranking flow with the K-means clustering algorithm. It has been observed that OKM produced a more robust and consistent outcome in HDI rankings. OKM has two primary advantages due to the partial net outranking flow: (1) it examines the weight of each criterion; and (2) it consists of the preferences in the same cluster of all the alternatives. Later, Liu et al. [23] presented a multicriterion ordered clustering algorithm based on PROMETHEE and K-Medoids clustering algorithms. The larger the cluster number, the better the results. However, this may not be appropriate for data in real-time with a large number of features and objects. As a result, a diagram method approach has been developed based on the net outranking of PROMETHEE and fuzzy c-means (OFCM) [24]. Consequently, we proposed a multi-criterion ordered profile clustering algorithm (MOPFCM) by utilizing the partial net outranking flow of PROMETHEE and fuzzy c-means, which not only provides the ordered clusters but also addresses the aforementioned problem through their centroids. In addition to that, earlier ordered clustering algorithms rewarded little attention to the fuzziness of each cluster's alternatives because they could resolve uncertainty in real-world situations.

The fuzzy c-means (FCM) algorithm is widely recognized as one of the most prominent fuzzy clustering algorithms 25]. Its ability to handle uncertainty in practical scenarios and ease of explanation has made it a popular choice. Over the years, researchers have extensively studied this innovative method, and its effectiveness has been further enhanced, as demonstrated by Wang et al. [26]. In addition, Zheng et al. [27] proposed a universal FCM and a hierarchical FCM to address image noise issues. Xu and Wu [28] present an extension of the fuzzy c-means approach to an intuitionistic fuzzy environment. Beg and Rashid [29] propose a modified dissimilarity measure for fuzzy data that is proposed and applied to real data with mixed feature variables. By employing relative entropy to optimize cluster dissimilarity, Zarinbal et al. [30] enhanced the goal. On the other hand, the current FCM algorithms are best suited to problems involving categorizing alternatives into predefined categories with no links between them. Decision-makers (DMs) in the field of MCDA may desire to receive "ordered clusters" with targeted ranking information between the clusters.

The goal of this article is to build on previous work and present a new multicriteria targeted rank clustering algorithm. Motivated by the partial net outranking flow/profile of the PROMETHEE [31] and the FCM, we propose a multicriteria-ordered profile clustering algorithm. MOPFCM first randomly assigns the membership values, then generates the cluster centroids, updates the membership values, and minimizes the partial net outranking flow of all rank clusters through the construction of an optimization model.

The rest of the paper is arranged as follows: Section 2 overviews the fuzzy c-means clustering algorithm and partial net outranking flow of the classical PROMETHEE methods. In Section 3, we propose an optimization model of MOPFCM based on the profile of PROMETHEE. In Section 4, to show the applicability and implementation procedure of the suggested method, we solved the human development index problem. This section also includes a comparison with other ordered clustering algorithms. Finally, Section 5 contains the paper's conclusion.

2. Preliminaries

This section focusing on the review of the FCM and the partial net outranking flow of the PROMETHEE methods is as follows.

2.1. Fuzzy c-Means Clustering Algorithm. To solve the clustering challenge, Bezdek [25] presented the classical fuzzy c-means clustering (FCM). The following objective function was determined by FCM to get the clustering result with fuzzy membership μ_{ij} and cluster centroid $\widehat{\mathbf{V}}_i$ cluster centroid:

$$J_{\mathrm{FCM}} = \sum_{\mathbf{i}=1}^{\mathbf{c}} \sum_{\mathbf{j}=\mathrm{sent1}}^{\mathbf{l}} \mu_{ij}^{\mathbf{m}} \left\| \widehat{\mathbf{x}}_{\mathbf{j}} - \widehat{\mathbf{V}}_{\mathbf{i}} \right\|, \tag{1}$$

where c is the cluster number of n objects, m is the weighting exponent ($m \in [1.5,\ 3.0]$ is determined to be analytically acceptable [32]), $\widehat{\mathbf{x}}_{\mathbf{j}}$ is an object of the set $X = \left\{\widehat{\mathbf{x}}_{\mathbf{j}} \mid j = 1, 2, \ldots, l\right\} \subseteq \mathbb{R}^q, q$ and l is the number of dimensions, and $\widehat{\mathbf{x}}_{\mathbf{j}}$ is the number of objects, respectively. Whereas, indicates the distance norm applied by FCM in Algorithm 1 [25]:

2.2. PROMETHEE Methods and Its Partial Net Outranking Flow. The following subsection introduces the PROMETHEE method's core ideas and the partial net outranking flow (see Brans and Mareschal [33] for more details). Let us consider the following multicriteria problem:

$$\max \left\{ g_1(a), g_2(a), \ldots, g_j(a), \ldots, g_k(a) \mid a \in A \right\}, \tag{2}$$

where A is a finite set of possible alternatives $\{a_1, a_2, \ldots, a_i, \ldots, a_n\}$ and $G = \left\{g_1(.), g_2(.), \ldots, g_j(.), \ldots, g_l(.)\right\}$ is a set of evaluation criteria, in which are may be consider maximized or to be minimized. The decision-maker expects to identify an alternative that optimizing all the criteria. Let $g_l(a_m)$ denotes the alternative a_m

Step 1: Randomly initialize memberships μ_{ij} of $\hat{\mathbf{x}}_\mathbf{j}$ belonging to cluster i, and $\sum_{i=1}^{l}\mu_{ij}=1$.
Step 2: Centroid $\hat{\mathbf{V}}_\mathbf{i}$ of FCM is calculated as $\hat{\mathbf{V}}_\mathbf{i}=\sum_{j=1}^{l}(\mu_{ij})^m\hat{\mathbf{x}}_\mathbf{j}/\sum_{J=1}^{l}(\mu_{ij})^m$.
Step 3: Update μ_{ij} based on $\mu_{ij}=(1/\hat{\mathbf{x}}_\mathbf{j}-\hat{\mathbf{V}}_\mathbf{i})^{(1/m-1)}/\sum_{i=1}^{c}(1/\hat{\mathbf{x}}_\mathbf{j}-\hat{\mathbf{V}}_\mathbf{i})^{(1/m-1)}$.
Step 4: Steps 2 & 3 are repeated till the convergence of J_{FCM} i.e., J_{FCM} has negligible variations. Finally, we get the cluster results comprising of objects $\{\hat{\mathbf{x}}_\mathbf{j}\mid j=1,2,\ldots,l\}$ i.e., $C_k(k=1,2,\ldots,c)$.

Algorithm 1: Fuzzy c-mean clustering algorithm.

evaluation with respect to the criteria g_l The preference function $p_l(a_i,a_j)$, which is apply on the deviation $d_l(a_i,a_j)$. The preference function gives the degree of preference of the alternatives a_i to a_j with respect to the criterion g_l In PROMETHEE, six types of preference functions are can be utilized [33]. We can write preference function as:

$$\begin{aligned} P_l\left(a_i,a_j\right)&=G_l\left(d_l\left(a_i,a_j\right)\right);\forall a_i,a_j\in A,\\ \text{Where } d_l\left(a_i,a_j\right)&=g_l(a_i)-g_l\left(a_j\right),\end{aligned}\tag{3}$$

and $G_l(.)$ Is a function which is monotonically non-decreasing and varying 0 to 1. $\forall a_i,a_j\in A$, we have

$$\pi\left(a_i,a_j\right)=\sum_{l=1}^{m}w_lP_l\left(a_i,a_j\right),\tag{4}$$

where $\pi(a_i,a_j)$ denotes the total preference of the alternative a_i over the a_j when all criteria are taken into consideration. Whereas, w_l is the relative weight of the criterion g_l from the set of all criteria.

Similarly,

$$\pi_{gl}\left(a_i,a_j\right)=\sum_{l=1}^{m}w_{gl}P_{gl}\left(a_i,a_j\right),\tag{5}$$

where $\pi_{gl}(a_i,a_j)$ indicates the profile (partially degree of preference) of the alternative a_i over the alternative a_j captivating into account of single criteria.

In order to acquire the ranking of all the alternatives, positive, negative and total net outranking flow were established by [33] as follows:

$$\begin{aligned}\varnothing^{+}(a_i)&=\frac{1}{m-1}\sum_{x\in A\setminus\{a_i\}}\pi(a_i,x),\\ \varnothing^{-}(a_i)&=\frac{1}{m-1}\sum_{x\in A\setminus\{a_i\}}\pi(x,a_i),\\ \varnothing(a_i)&=\varnothing^{+}(a_i)-\varnothing^{-}(a_i).\end{aligned}\tag{6}$$

Similarly, the partial positive outranking, negative outranking, and net outranking flows are given:

$$\begin{aligned}\varphi_{gl}^{+}(a_i)&=\frac{1}{m-1}\sum_{x\in A\setminus\{a_i\}}\pi_{gl}(a_i,x),\\ \varphi_{gl}^{-}(a_i)&=\frac{1}{m-1}\sum_{x\in A\setminus\{a_i\}}\pi_{gl}(x,a_i).\end{aligned}\tag{7}$$

Therefore,

$$\varnothing_{gl}(a_i)=\sum_{l=1}^{m}\varphi_{gl}^{+}(a_i)-\sum_{l=1}^{m}\varphi_{gl}^{-}(a_i).\tag{8}$$

Net outranking flow is

$$\varphi(a_i)=\sum_{l}^{m}\sum_{i}^{n}\varphi_{gl}(a_i),\tag{9}$$

where the partial positive flow $\varphi_{gl}^{+}(a_i)$ indicates how much the alternatives a_i prefers to the rest of all other alternatives in a single criterion. The larger $\varphi_{gl}^{+}(a_i)$ the better the alternatives a_i Similarly, the partial negative flow $\varphi_{gl}^{-}(a_i)$ indicates how much the alternatives are preferred by all other alternatives a_i in a single criterion. The smaller $\varphi_{gl}^{-}(a_i)$, the better the alternatives a_i. Usually, the larger the partial net outranking flow $\varnothing(a_i)$, the better the alternatives a_i.

3. Multicriteria Ordered Clustering Algorithm

We deal with a unique form of clustering problem called as ranks clustering in this study, which was first addressed by [21] for the nation risk rating problem. Identifying the ordered clusters can help the DM sort the possibilities, as previously noted. Unlike typical clustering problems, the ordered clustering problem separates the alternatives into a predetermined number of groups and also contains a complete ordering relationship between these clusters.

Let $A=\{a_1,a_2,\ldots,a_i,\ldots,a_n\}\subseteq\mathbb{R}^m$ be a sample set is assessing through a set of criteria $G=\{g_1,g_2,\ldots,g_i,\ldots,g_m\}$. We call a partition an ordered partition if it meets the following criteria:

(i) $A=\cup_{i=1,2,\ldots,K}\widehat{C}_i$;

(ii) $\forall_{i\neq j},\widehat{C}_i\cap\widehat{C}_j=\varphi$;

(iii) $\widehat{C}_1\succ\widehat{C}_2\succ\cdots\widehat{C}_K$

where $\widehat{C}_i$ signifies the i^{th} ordered cluster and $\succ$ indicating the priority relation among the clusters, such as if $\widehat{C}_i\succ\widehat{C}_j$, then the elements in $\widehat{C}_i$ are better than $\widehat{C}_j$.

By employing the Euclidean norm to assess similarity, FCM has been frequently utilized to cluster the alternatives with respect to their criteria. However, Euclidean norm cannot take into account the relative importance of the criteria being assessed. The PROMETHEE approach can determine the priority degree for each pair of alternatives based on their differences. As a result, we offer a novel

supervised clustering technique based on partial net outranking, a multicriteria ordered clustering algorithm based on partial net outranking, and fuzzy c-means (MOPFCM). This algorithm will search for the best c-ordered partition of alternatives built on the FCM in composition with the partial net outranking of PROMETHEE.

3.1. Minimum Partial Net Outranking Flow Objective Function. Similar to the FCM clustering algorithm, we propose an objective function based on partial net outranking flow of PROMETHEE to minimize:

$$\widehat{J}_{\mathrm{MOPFCM}}(U,V)=\sum_{i=1}^{K}\sum_{a_j\in C_i}\left(\mu_{ij}\right)^m\left|\varphi_{gl}^{C_i}(a_i)\right|^2, \tag{10}$$

where C_i is the group of alternatives in i^{th} cluster of the ordered clustering and

$$\varphi_{gl}^{C_k}(a_i)=\frac{\left(\mu_{ij}\right)^m}{|C_k|}\left(\sum_{a_i\in C_k}\pi_{gl}\left(a_i,a_j\right)-\sum_{a_i\in C_k}\pi_{gl}\left(a_j,a_i\right)\right). \tag{11}$$

With the fuzzy centroids θ_{gl}^{i} is

$$\theta_{gl}^{i}=\frac{\sum_{i=1}^{n}\left(\mu_{ij}\right)^m\varphi_{gl}(a_i)}{\sum_{i=1}^{n}\left(\mu_{ij}\right)^m}. \tag{12}$$

To capture the similarity of alternatives with respect to criteria and reconstruct the optimization model (10) for clustering of the alternatives, we propose the partial net outranking flow. The suggested model offers the following important benefits over the traditional FCM clustering algorithm: (1) $\widehat{J}_{\mathrm{MOFCM}}$ takes the relative importance of each criterion into account. (2) each criterion partial net outranking flow considers the preferences of all the alternatives in the same cluster; (3) centroids of partial net outranking clustering are used to classify the clusters and criteria; and (4) the net outranking will be used to rank the alternatives within and between clusters.

Theorem 1. *Clustering based on partial net outranking converges to local minima of $\widehat{J}_{MOFCM}$ in the finite iterations.*

Proof. In the PROMETHEE method, there are six kinds of preference functions [33]. These functions are monotonically increasing or decreasing and the partial net outranking flow $\varnothing_{gl}(a_i)$ for the alternatives a_i is one-to-one and converges in MOPFCM. Bezdek et al. [32] has previously mathematically validate the convergence of FCM and consequently, convergence of local optimal solution of $\widehat{J}_{\mathrm{MOFCM}}$. □

Property 1. Suppose that, $A=\{a_1,a_2,\ldots,a_i,\ldots,a_n\}$ be the set of alternatives having fuzzy centroids $\theta_{gl}^{i}\in\mathbb{R}^m$ for $i=1,2,3,\ldots,C$ of each cluster C_i, the membership μ_{ij} of the alternative a_j belonging to C_i, and $\widehat{J}_{\mathrm{MOPFCM}}(U,V)$ is the objective function of MOPFCM. If $\widehat{J}_{\mathrm{MOPFCM}}(U,V)\mathrm{is}<1$, then there will be a distinct partitions of A.

Proof. According to Algorithm 1 and (12), we have

$$\begin{aligned}\theta_{gl}^{i}&=\frac{\sum_{i=1}^{n}\left(\mu_{ij}\right)^m\varphi_{gl}^{C_i}(a_i)}{\sum_{i=1}^{n}\left(\mu_{ij}\right)^m}\geq\frac{\sum_{i=1}^{n}\left(\mu_{ij}\right)^m\varphi_{\min(gl)}^{C_i}(A)}{\sum_{i=1}^{n}\left(\mu_{ij}\right)^m}=\varphi_{\min(gl)}^{C_i}(A),\\ \theta_{gl}^{i}&=\frac{\sum_{i=1}^{n}\left(\mu_{ij}\right)^m\varphi_{gl}^{C_i}(a_i)}{\sum_{i=1}^{n}\left(\mu_{ij}\right)^m}\geq\frac{\sum_{i=1}^{n}\left(\mu_{ij}\right)^m\varphi_{\max(gl)}^{C_i}(A)}{\sum_{i=1}^{n}\left(\mu_{ij}\right)^m}=\varphi_{\max(gl)}^{C_i}(A),\end{aligned} \tag{13}$$

where, $\varphi_{\min(gl)}^{C_i}(A)$ and $\varphi_{\max(gl)}^{C_i}(A)$ represents the maximal and minimal of partial net outranking flows $\varphi_{gl}^{C_i}(a_i)$ respectively. □

3.2. Update the Cluster. The shortest Euclidean distance is being used for updating the membership of clusters in the conventional FCM techniques. Only the actual distance between alternatives is taken into account by the Euclidean distance. But it is not appropriate for MCDM's targeting ordered clustering of criterion. The partial net outranking flow can identify the relative relevance of each choice for the cluster. We attempt to figure out the relationship between the partial net outranking and the cluster center that corresponds to it, and then we get the result. From (11), we have

$$\varphi_{gl}^{C_k}(a_i) = \frac{(\mu_{ij})^m}{|C_k|}\left(\sum_{a_i \in C_k} \pi_{gl}(a_i, a_j) - \sum_{a_i \in C_k} \pi_{gl}(a_j, a_i)\right),$$

$$\varphi_{gl}^{C_k}(a_i) = \frac{(\mu_{ij})^m}{|C_k|}\left(\sum_{a_i \in C_k} \sum_{l=1}^{m} w_{gl} P_{gl}(a_i, a_j) - \sum_{a_i \in C_k} \sum_{l=1}^{m} w_{gl} P_{gl}(a_j, a_i)\right),$$

$$\varphi_{gl}^{C_k}(a_i) = \frac{(\mu_{ij})^m}{|C_k|}\left(\sum_{a_i \in C_k} \sum_{l=1}^{m} w_{gl} G_{gl}(d_{gl}(a_i) - d_{gl}(a_j)) - \sum_{a_i \in C_k} \sum_{l=1}^{m} w_{gl} G_{gl}(d_{gl}(a_j) - d_{gl}(a_i))\right), \quad (14)$$

$$\varphi_{gl}^{C_k}(a_i) = \frac{(\mu_{ij})^m}{|C_k|}\left(\sum_{a_i \in C_k} \sum_{l=1}^{m} w_{gl} G_{gl}(d_{gl}(a_i) - d_{gl}(\theta_{gl}^i) + d_{gl}(\theta_{gl}^i) - d_{gl}(a_j)) - \sum_{a_i \in C_k} \sum_{l=1}^{m} w_{gl} G_{gl}(d_{gl}(a_j) - d_{gl}(\theta_{gl}^i) + d_{gl}(\theta_{gl}^i) - d_{gl}(a_i))\right),$$

where $d_{gl}(a_i)$ denotes the calculation of the alternatives a_i with respect to the criteria gl, and $\theta_{gl}^i \varepsilon \mathbb{R}^m$ represents the center of the k^{th} cluster. If the function $G_{gl}(.)$ is a linear, then

$$\varphi_{gl}^{C_k}(a_i) = \varphi_{gl}^{\theta_{gl}^i}(a_i) + \varphi_{gl}^{C_k}(\theta_{gl}^i), \quad (15)$$

where θ_{gl}^i represents the center of the k^{th} cluster C_k and

$$\varphi_{gl}^{\theta_{gl}^i}(a_i) = \frac{(\mu_{ij})^m}{|C_k|}\left(\sum_{a_i \in C_k} \pi_{gl}(a_i, \theta_{gl}^i) - \sum_{a_i \in C_k} \pi_{gl}(\theta_{gl}^i, a_i)\right). \quad (16)$$

Brans and Vincke [34] presented six types of specific preference functions which are linear (functions except the 6^{th} (Gaussian criterion). Consequently, the conversion can be established in transformation. If the linear preference functions are applied on a big data set, we can see $\varphi_{gl}^{C_k}(a_i) \approx \varphi_{gl}^{\theta_{gl}^i}(a_i) + \varphi_{gl}^{C_k}(\theta_{gl}^i)$. Where, θ_{gl}^i is the cluster center to denote the relevant cluster and calculate the distance between cluster's center and the alternatives.

3.3. Update the Centroids. In classical fuzzy c-means clustering algorithm we used Algorithm 1 to calculate the fuzzy centroids and update the membership value. To apprehend the ordered centroids of the clustering, the cluster centers can be obtained by

$$\theta_{gl}^i = \arg\min \left|\varphi_{gl}^{C_k}(\theta_i)\right|^2, i = 1, 2, \ldots, n, \quad (17)$$

where $\varphi_{gl}^{C_k}(\theta_i)$ represents the partial net outranking flow of data and $\theta_{gl}^i \varepsilon \mathbb{R}^m$ can be computed using (10).

3.4. Classification of Ordered Clusters and Criteria. Centroids of clusters $\theta_{gl}^{C_i} \in \mathbb{R}^m$ are representing the middle value of all the alternatives lies within the clusters. These centroids are being used for classification of different levels of clusters and their corresponding criteria.

3.5. Ranking of Alternatives within and between the Clusters. Assign ranks to each clustered alternatives based on net outranking within and between the clusters i.e., (Cluster #, Rank of $\varphi_{gl}^{C_k}(a_i)$, Rank of $\varnothing(a_i)$).

3.6. The Proposed Algorithm. (Algorithm 2)

4. Case Study and Comparative Analysis

In this section, we apply the MOPFCM on real-life situation problem associated with the Human Development Index (HDI) adapted from [20, 22] to validate the efficiency of MOPFCM. The United Nations Development Program (UNDP) ranks the 179 countries in the HDI ranking based on three criteria i.e., $G = \{g_1, g_2, g_3\} = \{\text{life expectancy, education, income index}\}$. We are not concerned with the precise ranking problem of nations in this section, and instead divide the countries according to the three criteria. Our goal is to use a targeted rank-based regrouping method that takes partial net outranking of all three criteria into consideration, then compare the results to the OKM to verify the MOPFCM. Moreover, our framework not only ranks the clusters but also ranks the alternative within-cluster and between the clusters. In the end, cluster centroids of partial net outranking are being used for classification of ordered clusters and their criteria.

4.1. MOPFCM Algorithm for Regrouping on the Basis of Partial Net Outranking Flow. This subsection demonstrates how it will implement for clustering the countries (see Appendix of [20]) based on their performance in aforementioned three criteria for the year 2008. Let $A = \{A = \{a_i \mid \text{i} = 1, 2, \ldots, \text{n}\}$ represented as alternatives of all countries against the three criteria $G = \{g_1, g_2, g_3\}$. The a_i links to the g^{th} place of country in ranking of HDI. Step-by-step implementation process for ordered clustering built on Algorithm 2 is stated as follows.

The process of the MOPFCM is as follows:
Step 1: Compute the partial net outranking flow $\varphi_{gl}(a_i)$ and $\varphi(a_i)$ net outranking.
Step 2: Randomly initialize membership of $\varphi_{gl}^{C_k}(a_i)$ belonging to the cluster i.
Step 3: Compute the fuzzy centroids θ_{gl}^{i} $\theta_{gl}^{i} = \sum_{i=1}^{n}(\mu_{ij})^m \varphi_{gl}(a_i)/\sum_{i=1}^{n}(\mu_{ij})^m$.
Step 4: Targeted rank the clusters based on the fuzzy centroids θ_{gl}^{i} of clusters. For example, if $\theta_{gl}^{i} \succ \theta_{gl}^{j}$ then $\varphi_{gl}^{C_i}(a_i) \succ \varphi_{gl}^{C_j}(a_i)$, where $\succ$ is determined by net outranking flow.
Step 5: update the μ_{ij} based on $\mu_{ij} = (1/\|\varphi_{gl}^{C_k}(a_i) - \theta_{gl}^{i}\|)^{(1/m-1)}/(1/\sum_{i=1}^{c}\|\varphi_{gl}^{C_k}(a_i) - \theta_{gl}^{i}\|)^{(1/m-1)}$.
Step 6: Step 3 and 4 are repeating till the convergence of $\widehat{J}_{\text{MOPFCM}}$ in (10) i.e., $\widehat{J}_{\text{MOPFCM}}$ has negligible change.
Step 7: Centroids of clusters $\theta_{gl}^{C_i} \in \mathscr{R}^m$ are being used for classification of ordered clusters and their profiles.
Step 8: Assign ranks to each clustered alternatives based on net outranking within and between the clusters i.e., (Cluster #, Rank of $\varphi_{gl}^{C_k}(a_i)$, Rank of $\varnothing(a_i)$).

Algorithm 2: Multicriteria ordered profile fuzzy c-means clustering algorithm.

Step 1: Calculate preference degree of each criterion $\pi_{gl}(a_i, a_j)$ between each pair of alternatives and compute the partial net outranking flow $\varnothing_{gl}(a_i)$ and net outranking flow $\varnothing(a_i)$ of each country using equations (8) and (9). We select the same linear preference function for each criterion as mentioned in [20]:

$$f_k(v) = \begin{cases} 0, & v \le 0, \\ \dfrac{v}{p_l}, & 0 \le v \ge p_l, \quad l = 1, 2, 3, \\ 1, & v \ge p_l, \end{cases} \tag{18}$$

where threshold value p_l and the corresponding weights of each criterion has been determined in Table 1.

Then, for each criterion we can find the preferences between two alternatives are shown in Figures 1–3 and their partial net outranking $\varnothing_{gl}(a_i)$ for i = 1, 2, . . . , 179 is obtained as shown in Figure 4.

Step 2: Randomly initialize membership of $\varphi_{gl}^{C_k}(a_i)$ that belongs to the cluster i.

Step 3: By using equation (12) compute cluster centroids θ_{gl}^{i} of each cluster and let $m = 2$.

Step 4: Based on partial net outranking flow, compute the targeted ranked clusters based on the fuzzy centroids θ_{gl}^{i} of each cluster. Such as, if $\theta_{gl}^{i} \succ \theta_{gl}^{j}$ then $\varphi_{gl}^{C_i}(a_i) \succ \varphi_{gl}^{C_j}(a_i)$, where $\succ$ is determined by the net outranking flow.

Step 5: Update the memberships μ_{ij} of $\varphi_{gl}^{C_k}(a_i)$ using the Step 5 of Algorithm 2.

Step 6: Steps 3 and 4 are repeated till the convergence of $\widehat{J}_{\text{MOPFCM}}$ i.e., $|\widehat{J}_{\text{MOPFCM}}^{t} - \widehat{J}_{\text{MOPFCM}}^{t-1}| \le \varepsilon$, where t indicates the iteration and $\varepsilon = 0.001$.

Step 7: Classification of ordered clusters with respect to centroids of clusters $\theta_{gl}^{C_i} \in \mathscr{R}^m$.

Step 8: Assign ranks to each clustered alternatives based on net outranking within and between the clusters i.e., (Cluster # , Rank of $\varphi_{gl}^{C_k}(a_i)$, Rank of $\varnothing(a_i)$) see Table 2.

In order to assess the ordering in the HDI problem, 4 clusters are pre requests which is justified later by De Smet et al. [20, 22]. Figure 5 validate the aforementioned claim by drifting of the total sum of all country's' partial net outranking flow in respect of the cluster numbers. As a result, we select 4 clusters, which are equal to very high, high, medium, and low human-developed countries. Then, we extant the ordered clustering results got from MOPFCM in Figure 6. The x-axis indicates the HDI aggregate scores for the 179 nations, whereas the y-axis indicates the number of clusters are 4. MOPFCM results are very similar with HDI ranks and OKM and De Semet et al. [20]. First 52 countries are belonging to cluster very high human development index, 74 country belongs to high human development index, 30 countries are belonging to medium level human development index, whereas 23 are belonging to low level human development index as shown in Figure 6. All alternatives are assigned ordered ranking i.e., cluster number, ranking within cluster and ranking among the clusters. For example, the status of QATAR is (1, 34, 36) which it belongs to very high human development countries having 34^{th} position among 52 and 36^{th} position among 179 countries. In summary, Table 2 shows the overall rankings of nations as well as their grouping using various clustering approaches.

4.2. Comparison of MPOFCM with Other Clustering Algorithms with FCM and OKM. We compare the outcomes of MOPFCM with the traditional FCM and the OKM for addressing the identical HDI issue as indicated above in order to further verify our suggested clustering technique. We apply a Python "FCM" module to group the alternatives that use the conventional FCM clustering technique based on three different HDI criteria (see Table 2), showing that the partition result and HDI ranking are incompatible. The fundamental reason for this might be that the classic FCM uses the Euclidean distance to evaluate the degree of similarity between any two alternatives. In other words, the typical FCM is unable to offer preference correlations between alternatives and clusters owing to the symmetry of the Euclidean distance.

Table 1: Normalized data based, indifference, weight and preference thresholds of each criterion.

Parameters	Life expectancy	Adult literacy index	GDP
Strict preference threshold: p_l	0.704	0.719	0.828
Indifference threshold: q_l	0	0	0
Weight of criteria: w_l	0.333	0.333	0.333

Table 2: Ranking of alternatives based on net outranking within and between the clusters, i.e., (Cluster #, Rank of $\varphi_{gl}^{C_k}(a_i)$, Rank of $\varnothing(a_i)$) is given.

Countries	FCM	MOPFCM	OKM
Iceland	(1, 1, 1)	(1, 1, 1)	(1, 1, 1)
Norway	(1, 2, 2)	(1, 2, 2)	(1, 2, 2)
Canada	(1, 3, 3)	(1, 3, 3)	(1, 3, 3)
Australia	(1, 4, 4)	(1, 4, 4)	(1, 4, 4)
Ireland	(1, 5, 5)	(1, 6, 6)	(1, 6, 6)
Netherland	(1, 6, 6)	(1, 7, 7)	(1, 7, 7)
Sweden	(1, 7, 7)	(1, 5, 5)	(1, 5, 5)
Japan	(1, 8, 8)	(1, 8, 8)	(1, 8, 8)
Luxembourg	(1, 8, 8)	(1, 11, 11)	(1, 11, 11)
Switzerland	(1, 10, 10)	(1, 10, 10)	(1, 10, 10)
France	(1, 11, 11)	(1, 9, 9)	(1, 9, 9)
Finland	(1, 12, 12)	(1, 12, 12)	(1, 12, 12)
Denmark	(1, 13, 13)	(1, 13, 13)	(1, 13, 13)
Austria	(1, 14, 14)	(1, 14, 14)	(1, 14, 14)
United States	(1, 14, 14)	(1, 16, 16)	(1, 16, 16)
Spain	(1, 16, 16)	(1, 15, 15)	(1, 15, 15)
Belgium	(1, 17, 17)	(1, 17, 17)	(1, 17, 17)
Greece	(1, 18, 18)	(1, 18, 18)	(1, 18, 18)
Italy	(1, 19, 19)	(1, 20, 20)	(1, 20, 20)
New Zealand	(1, 20, 20)	(1, 19, 19)	(1, 19, 19)
United Kingdom	(1, 22, 22)	(1, 21, 21)	(1, 21, 21)
Hong Kong, China (SAR)	(1, 21, 21)	(1, 22, 22)	(1, 22, 22)
Germany	(1, 23, 23)	(1, 23, 23)	(1, 23, 23)
Israel	(1, 24, 24)	(1, 24, 24)	(1, 24, 24)
Korea (Republic of)	(1, 25, 25)	(1, 25, 25)	(1, 25, 25)
Slovenia	(1, 26, 26)	(1, 26, 26)	(1, 26, 26)
Brunei Darussalam	(1, 27, 27)	(1, 27, 27)	(1, 27, 27)
Singapore	(1, 28, 28)	(1, 28, 28)	(1, 28, 28)
Kuwait	(1, 29, 29)	(1, 30, 30)	(1, 30, 30)
Cyprus	(1, 30, 30)	(1, 29, 29)	(1, 29, 29)
United Arab Emirates	(1, 31, 31)	(1, 32, 32)	(1, 32, 32)
Bahrain	(1, 32, 32)	(1, 33, 33)	(1, 33, 33)
Portugal	(1, 33, 33)	(1, 31, 31)	(1, 31, 31)
Qatar	(1, 34, 34)	(1, 36, 36)	(1, 36, 36)
Czech Republic	(1, 35, 35)	(1, 34, 34)	(1, 34, 34)
Malta	(1, 36, 36)	(1, 35, 35)	(1, 35, 35)
Barbados	(1, 37, 37)	(1, 37, 37)	(1, 37, 37)
Hungary	(1, 39, 39)	(1, 39, 39)	(1, 39, 39)
Poland	(1, 40, 40)	(1, 40, 40)	(1, 40, 40)
Chile	(1, 41, 41)	(1, 38, 38)	(1, 38, 38)
Slovakia	(1, 38, 38)	(1, 46, 47)	(1, 48, 47)
Estonia	(1, 42, 42)	(1, 41, 41)	(1, 41, 41)
Lithuania	(1, 43, 43)	(1, 42, 42)	(1, 42, 42)
Latvia	(1, 44, 44)	(1, 43, 66)	(1, 43, 66)
Croatia	(1, 45, 45)	(1, 44, 44)	(1, 45, 44)
Argentina	(1, 46, 46)	(1, 45, 45)	(1, 46, 45)
Uruguay	(2, 1, 47)	(1, 1, 43)	(1, 44, 43)
Cuba	(2, 2, 48)	(1, 2, 46)	(1, 47, 46)
Bahamas	(1, 47, 49)	(1, 47, 48)	(1, 49, 48)
Costa Rica	(2, 3, 50)	(1, 3, 49)	(1, 50, 49)
Mexico	(2, 4, 51)	(1, 4, 50)	(1, 51, 50)
Libyan Arab Jamahiriya	(1, 48, 52)	(1, 48, 51)	(1, 52, 51)

Table 2: Continued.

Countries	FCM	MOPFCM	OKM
Oman	(1, 49, 53)	(1, 49, 52)	(2, 1, 52)
Seychelles	(1, 51, 55)	(1, 50, 55)	(2, 4, 55)
Saudi Arabia	(1, 50, 54)	(1, 51, 56)	(2, 5, 56)
Bulgaria	(2, 5, 56)	(2, 5, 53)	(2, 2, 53)
Trinidad and Tobago	(1, 52, 57)	(1, 52, 58)	(2, 7, 58)
Panama	(2, 6, 58)	(2, 6, 54)	(2, 3, 54)
Antigua and Barbuda	(1, 53, 59)	(1, 53, 71)	(2, 19, 71)
Saint Kitts and Nevis	(2, 7, 60)	(2, 7, 57)	(2, 6, 57)
Venezuela (Bolivarian Republic of)	(2, 8, 61)	(2, 8, 59)	(2, 8, 59)
Romania	(2, 9, 62)	(2, 9, 60)	(2, 9, 60)
Malaysia	(2, 10, 63)	(2, 12, 63)	(2, 12, 63)
Montenegro	(2, 11, 64)	(2, 10, 61)	(2, 10, 61)
Serbia	(2, 12, 65)	(2, 11, 62)	(2, 11, 62)
Saint Lucia	(2, 13, 66)	(2, 13, 64)	(2, 13, 64)
Belarus	(2, 14, 67)	(2, 14, 65)	(2, 14, 65)
Macedonia (TFYR)	(2, 15, 68)	(2, 16, 68)	(2, 16, 68)
Albania	(2, 16, 69)	(2, 15, 67)	(2, 15, 67)
Brazil	(2, 17, 70)	(2, 18, 70)	(2, 18, 70)
Kazakhstan	(2, 17, 70)	(2, 19, 72)	(2, 20, 72)
Ecuador	(2, 19, 72)	(2, 17, 69)	(2, 17, 69)
Russian Federation	(2, 20, 73)	(2, 21, 74)	(2, 22, 74)
Mauritius	(2, 21, 74)	(2, 22, 75)	(2, 23, 75)
Bosnia and Herzegovina	(2, 22, 75)	(2, 20, 73)	(2, 21, 73)
Turkey	(2, 23, 76)	(2, 24, 76)	(2, 25, 76)
Dominican Republic	(2, 24, 77)	(2, 23, 88)	(2, 24, 88)
Lebanon	(2, 25, 78)	(2, 25, 77)	(2, 26, 77)
Peru	(2, 26, 79)	(2, 27, 79)	(2, 28, 79)
Colombia	(2, 27, 80)	(2, 26, 78)	(2, 27, 78)
Thailand	(2, 28, 81)	(2, 29, 81)	(2, 30, 81)
Ukraine	(2, 29, 82)	(2, 28, 80)	(2, 29, 80)
Armenia	(2, 31, 84)	(2, 30, 82)	(2, 31, 82)
Iran (Islamic Republic of)	(2, 30, 83)	(2, 32, 84)	(2, 33, 84)
Tonga	(2, 32, 85)	(2, 31, 83)	(2, 32, 83)
Grenada	(2, 32, 85)	(2, 45, 98)	(2, 46, 98)
Jamaica	(2, 34, 87)	(2, 33, 85)	(2, 34, 85)
Belize	(2, 43, 96)	(2, 42, 95)	(2, 43, 95)
Suriname	(2, 35, 88)	(2, 35, 87)	(2, 36, 87)
Jordan	(2, 36, 89)	(2, 34, 86)	(2, 35, 86)
Dominican Republic	(2, 37, 90)	(2, 36, 88)	(2, 37, 88)
Saint Vincent and the Grenadines	(2, 38, 91)	(2, 38, 90)	(2, 39, 90)
Georgia	(2, 39, 92)	(2, 37, 91)	(2, 38, 91)
China	(2, 40, 93)	(2, 39, 92)	(2, 40, 92)
Tunisia	(2, 41, 94)	(2, 41, 94)	(2, 42, 94)
Samoa	(2, 42, 95)	(2, 40, 93)	(2, 41, 93)
Azerbaijan	(2, 44, 97)	(2, 43, 96)	(2, 44, 96)
Paraguay	(2, 45, 98)	(2, 44, 97)	(2, 45, 97)
Maldives	(2, 46, 99)	(2, 46, 99)	(2, 47, 99)
Algeria	(2, 47, 100)	(2, 49, 102)	(2, 50, 102)
El Salvador	(2, 48, 101)	(2, 48, 101)	(2, 49, 101)
Philippines	(2, 49, 102)	(2, 47, 100)	(2, 48, 100)
Fiji	(2, 50, 103)	(2, 51, 104)	(2, 52, 104)
Sri Lanka	(2, 51, 104)	(2, 50, 103)	(2, 51, 103)
Syrian Arab Republic	(2, 52, 105)	(2, 52, 105)	(2, 53, 105)
Occupied Palestinian Territories	(2, 53, 106)	(2, 53, 106)	(2, 54, 106)
Gabon	(2, 54, 107)	(2, 61, 114)	(2, 62, 114)
Turkmenistan	(2, 55, 108)	(2, 56, 109)	(2, 57, 109)
Indonesia	(2, 56, 109)	(2, 54, 107)	(2, 55, 107)
Guyana	(2, 57, 110)	(2, 55, 108)	(2, 56, 108)
Bolivia	(2, 58, 111)	(2, 60, 113)	(2, 61, 113)
Mongolia	(2, 59, 112)	(2, 58, 112)	(2, 59, 112)

Table 2: Continued.

Countries	FCM	MOPFCM	OKM
Moldova	(2, 60, 113)	(2, 59, 111)	(2, 60, 111)
Viet Nam	(2, 61, 114)	(2, 57, 110)	(2, 58, 110)
Equatorial Guinea	(2, 62, 115)	(2, 67, 120)	(2, 68, 120)
Egypt	(2, 63, 116)	(2, 62, 115)	(2, 63, 115)
Honduras	(2, 64, 117)	(2, 63, 116)	(2, 64, 116)
Cape Verde	(2, 65, 118)	(2, 64, 117)	(2, 65, 117)
Uzbekistan	(2, 66, 119)	(2, 65, 118)	(2, 66, 118)
Nicaragua	(2, 67, 120)	(2, 66, 119)	(2, 67, 119)
Guatemala	(2, 68, 121)	(2, 69, 122)	(2, 70, 122)
Kyrgyzstan	(2, 69, 122)	(2, 68, 121)	(2, 69, 121)
Vanuatu	(2, 70, 123)	(2, 71, 124)	(2, 72, 124)
Tajikistan	(2, 71, 124)	(2, 70, 123)	(2, 71, 123)
South Africa	(2, 72, 125)	(2, 72, 125)	(2, 73, 125)
Botswana	(2, 73, 126)	(2, 73, 126)	(2, 74, 126)
Morocco	(3, 1, 127)	(3, 2, 128)	(3, 2, 128)
Sao Tome and Principe	(3, 2, 128)	(3, 1, 127)	(3, 1, 127)
Namibia	(3, 3, 129)	(3, 3, 129)	(3, 3, 129)
Congo	(3, 4, 130)	(3, 4, 130)	(3, 4, 130)
Bhutan	(3, 5, 131)	(3, 6, 132)	(3, 6, 132)
India	(3, 9, 135)	(3, 9, 135)	(3, 9, 135)
Lao People's Democratic Republic	(3, 6, 132)	(3, 5, 131)	(3, 5, 131)
Solomon Islands	(3, 7, 133)	(3, 7, 133)	(3, 7, 133)
Myanmar	(3, 8, 134)	(3, 8, 134)	(3, 8, 134)
Cambodia	(3, 10, 136)	(3, 10, 136)	(3, 10, 136)
Comoros	(3, 11, 137)	(3, 11, 137)	(3, 11, 137)
Yemen	(3, 12, 138)	(3, 12, 138)	(3, 12, 138)
Pakistan	(3, 13, 139)	(3, 13, 139)	(3, 13, 139)
Mauritania	(3, 14, 140)	(3, 14, 140)	(3, 14, 140)
Swaziland	(3, 15, 141)	(3, 27, 153)	(3, 27, 153)
Ghana	(3, 16, 142)	(3, 16, 142)	(3, 16, 142)
Madagascar	(3, 17, 143)	(3, 15, 141)	(3, 15, 141)
Kenya	(3, 18, 144)	(3, 18, 144)	(3, 18, 144)
Nepal	(3, 19, 145)	(3, 17, 143)	(3, 17, 143)
Sudan	(3, 20, 146)	(3, 21, 147)	(3, 21, 147)
Bangladesh	(3, 21, 147)	(3, 19, 145)	(3, 19, 145)
Haiti	(3, 22, 148)	(3, 20, 146)	(3, 20, 146)
Papua New Guinea	(3, 23, 149)	(3, 22, 148)	(3, 22, 148)
Cameroon	(3, 24, 150)	(3, 23, 149)	(3, 23, 149)
Djibouti	(3, 25, 151)	(3, 24, 150)	(3, 24, 150)
Tanzania (United Republic of)	(3, 26, 152)	(3, 25, 151)	(3, 25, 151)
Senegal	(3, 27, 153)	(3, 26, 152)	(3, 26, 152)
Nigeria	(4, 1, 154)	(4, 2, 155)	(4, 2, 155)
Lesotho	(4, 2, 155)	(4, 3, 156)	(4, 3, 156)
Uganda	(4, 3, 156)	(4, 1, 154)	(4, 1, 154)
Angola	(4, 4, 157)	(4, 4, 159)	(4, 4, 159)
Timor-Leste	(3, 28, 158)	(3, 28, 157)	(3, 28, 157)
Togo	(3, 29, 159)	(3, 29, 158)	(3, 29, 158)
Gambia	(4, 5, 160)	(4, 5, 160)	(3, 30, 160)
Benin	(4, 6, 161)	(4, 7, 162)	(4, 6, 162)
Malawi	(4, 7, 162)	(4, 6, 161)	(4, 5, 161)
Zambia	(4, 8, 163)	(4, 8, 163)	(4, 7, 163)
Eritrea	(4, 9, 164)	(4, 9, 164)	(4, 8, 164)
Rwanda	(4, 10, 165)	(4, 10, 165)	(4, 9, 165)
Cote d'Ivoire	(4, 11, 166)	(4, 11, 166)	(4, 10, 166)
Guinea	(4, 12, 167)	(4, 12, 167)	(4, 11, 167)
Mali	(4, 13, 168)	(4, 14, 169)	(4, 13, 169)
Ethiopia	(4, 14, 169)	(4, 13, 168)	(4, 12, 168)
Chad	(4, 15, 170)	(4, 17, 172)	(4, 16, 172)
Guinea-Bissau	(4, 16, 171)	(4, 16, 171)	(4, 15, 171)
Burundi	(4, 17, 172)	(4, 15, 170)	(4, 14, 170)

Table 2: Continued.

Countries	FCM	MOPFCM	OKM
Burkina Faso	(4, 18, 173)	(4, 19, 174)	(4, 18, 174)
Niger	(4, 19, 174)	(4, 18, 173)	(4, 17, 173)
Mozambique	(4, 20, 175)	(4, 22, 177)	(4, 21, 177)
Liberia	(4, 21, 176)	(4, 20, 175)	(4, 19, 175)
Congo (Democratic Republic of the)	(4, 22, 177)	(4, 21, 176)	(4, 20, 176)
Central African Republic	(4, 23, 178)	(4, 23, 178)	(4, 22, 178)
Sierra Leone	(4, 24, 179)	(4, 24, 179)	(4, 23, 179)

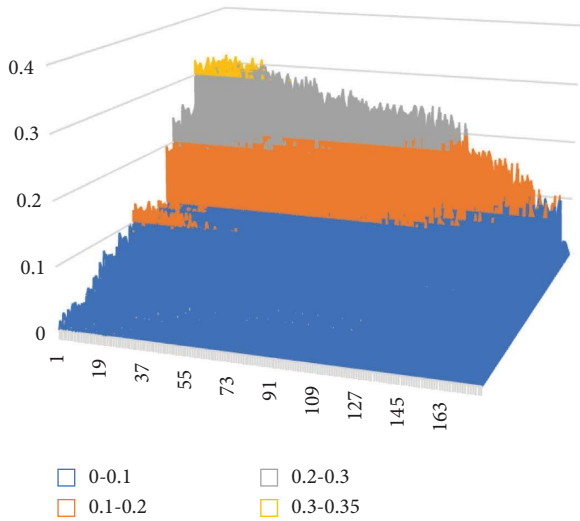


Figure 1: The pairwise preference degree of life expectancy for each pair among 179 countries. The yellow color shows highest preference degree and the blue color indicates the low preference degree in targeting number of intervals.

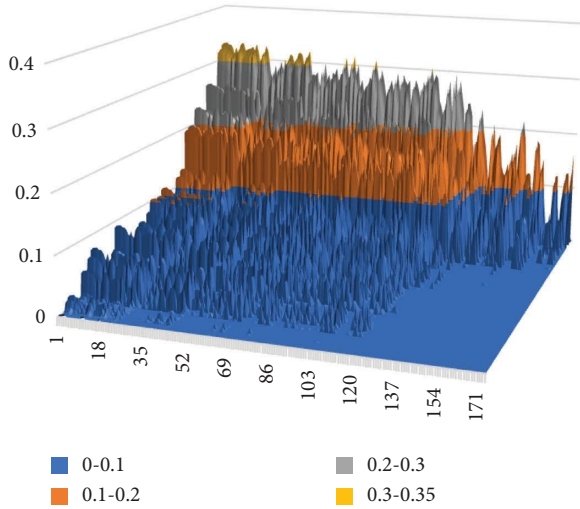


Figure 2: The pairwise preference degree of adult literacy index for each pair among 179 countries. The yellow color shows a high preference degree and the blue color indicates the low preference degree in targeting the number of intervals.

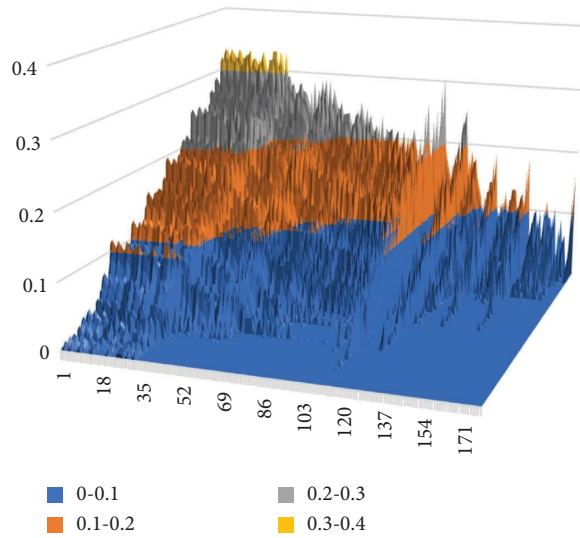


Figure 3: The pairwise preference degree of GDP for each pair among 179 countries. The Yellow color shows a high preference degree and blue color indicates a low preference degree in targeting the number of intervals.

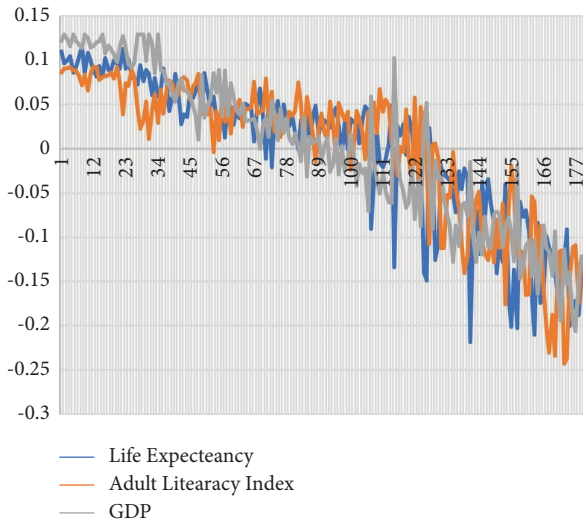


Figure 4: Partial net outranking's of life expectancy, adult literacy & GDP of 179 countries are showing very close moments with each other.

By using the Python "K-Means" package, we use the OKM clustering technique to regroup the nations-based partial net out rankings of three unique HDI criteria see Table 2, which show that the results are similar to the suggested MOPFCM and the partitioning results of HDI-ranking. The fundamental reason for this might be

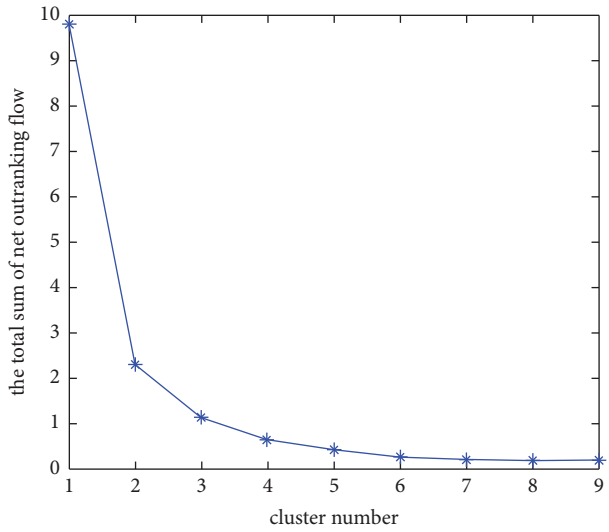


Figure 5: The sum of partial net outranking flows is proportional to the number of clusters.

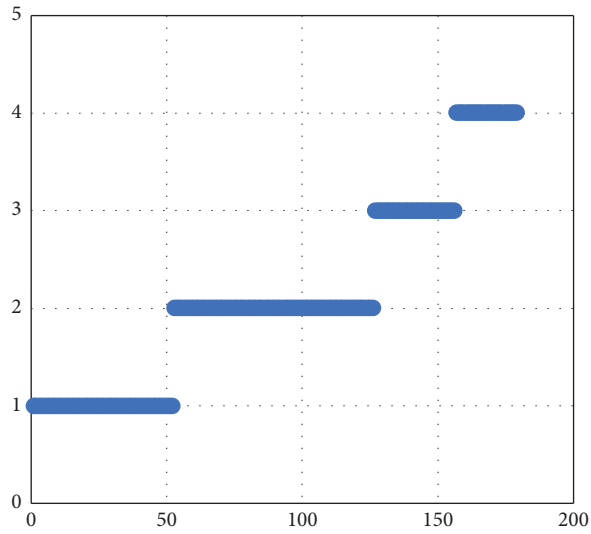


Figure 6: The x-axis represents the HDI ranking and the y-axis represents the its number of clusters.

because we estimate the similarity of any two nations based on their preference profile. In other words, preference correlations between items and clusters are provided by partial net outranking the symmetry of the Euclidean distance.

4.3. Analysis of MOPFCM Based on Centroids. The Ward minimum variance clustering technique is being used to create a double dendrogram heatmap based on four MOPFCM centroids. Rows and columns are converted into similar groups in this heat map based on their ordered centroids. Where, c_1 is the first cluster which is "Very High Human Development Index", c_2 "High Human Development Index", c_3 is "Medium Human Development Index" and c_4 is "Low Human Development Index"

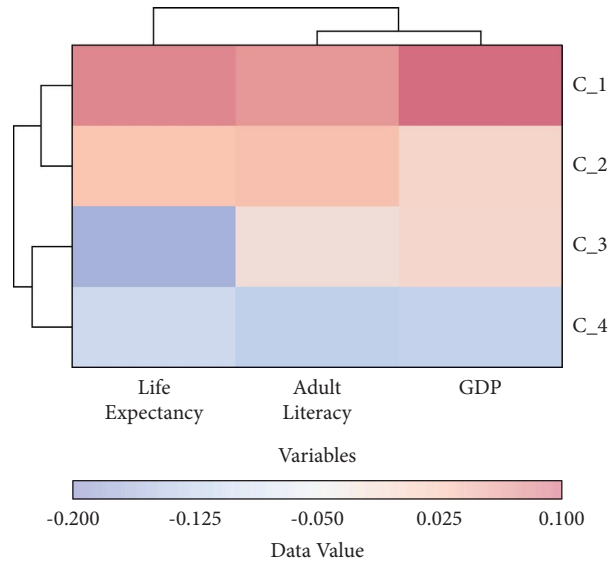


Figure 7: Double dendrogram heatmap have been applied on centroids of 4 clusters.

countries. It has been observed that, very high human development countries have low results in adult literacy. Whereas, a group of countries belongs to high human development have low GDP and group of countries belongs to medium level index have very low results in life expectancy as shown in Figure 7.

In summary, we present the multi-criteria ordered clustering algorithm for rankings of countries based on the partial and total net outranking in fuzzy environment. Rank the clustering results between the clusters and among the 179 based on total net outranking (cluster number, ranking within the cluster, overall ranking) as shown in Appendix A. The MOPFCM segmentation is compatible with OKM's ranks. The boundaries between the various clusters are adequate because they split developing, developed, and undeveloped countries in proportions that are appropriate.

5. Conclusion

In this research, we present MOPFCM, a targeted ranked clustering model based on partial net outranking and fuzzy c-means (FCM) clustering algorithm, to handle multi-criteria ordered clustering. The partial net outranking flow\ profile in the PROMETHEE approach differs from the classical FCM utilizing Euclidean norms. Several significant MOPFCM features are also theoretically supported. The human development index (HDI) issue has proved the usefulness of MOPFCM. Meanwhile, the ordered K-means (OKM) clustering technique and the traditional FCM clustering algorithm are given for comparison. The findings of the targeted rank clustering show that MOPFCM not only assists decision-makers in ranking clusters, but also in obtaining alternate rankings inside clusters and among all based on total net outranking. Based on the clustering and validation results, the advantages of MOPFCM can be summarized as follows:

(i) MOPFCM takes the weight of each criterion into account.

(ii) Each criterion profile reflects the alternatives preferences in the same cluster.

(iii) Centroids of the ordered profile clustering clarify the reasons that are being included in the cluster i.e., low, medium, high, and very high level of that particular group of countries.

(iv) Based on clustering results and total net outranking, complete ranking results are presented for clarification of the position of alternatives within the cluster and between the clusters.

As a result, we will use MOPFCM to execute big data clustering in a much quicker iterative manner in the future. More study is being investigated to see how the nonlinear preference function in PROMETHEE might be used to improve performance.

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] R. Xu and D. WunschII, "Survey of clustering algorithms," *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.

[2] M. Filippone, F. Camastra, F. Masulli, and S. Rovetta, "A survey of kernel and spectral methods for clustering," *Pattern Recognition*, vol. 41, no. 1, pp. 176–190, 2008.

[3] P. Berkhin, "A survey of clustering data mining techniques," in *Grouping Multidimensional Data*, pp. 25–71, Springer, Berlin, Heidelberg, 2006.

[4] A. Baraldi and P. Blonda, "A survey of fuzzy clustering algorithms for pattern recognition. I," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 29, no. 6, pp. 778–785, 1999.

[5] H. Zhu, E. C. Tsang, X. Z. Wang, and R. Aamir Raza Ashfaq, "Monotonic classification extreme learning machine," *Neurocomputing*, vol. 225, pp. 205–213, 2017.

[6] Y. Siskos, E. Grigoroudis, C. Zopounidis, and O. Saurais, "Measuring customer satisfaction using a collective preference disaggregation model," *Journal of Global Optimization*, vol. 12, no. 2, pp. 175–195, 1998.

[7] W. Michalowski, S. Rubin, R. Slowinski, and S. Wilk, "Triage of the child with abdominal pain: a clinical algorithm for emergency patient management," *Paediatrics and Child Health*, vol. 6, no. 1, pp. 23–28, 2001.

[8] L. Shen, F. E. Tay, L. Qu, and Y. Shen, "Fault diagnosis using rough sets theory," *Computers in Industry*, vol. 43, no. 1, pp. 61–72, 2000.

[9] I. Beg and T. Rashid, "An improved clustering algorithm using fuzzy relation for the performance evaluation of humanistic systems," *International Journal of Intelligent Systems*, vol. 29, no. 12, pp. 1181–1199, 2014.

[10] A. Ishizaka and P. Nemery, "Assigning machines to incomparable maintenance strategies with ELECTRE-SORT," *Omega*, vol. 47, pp. 45–59, 2014.

[11] P. Nemery and C. Lamboray, "Flow $\mathcal {S} $ ort: a flow-based sorting method with limiting or central profiles," *Top*, vol. 16, no. 1, pp. 90–113, 2008.

[12] N. Belacel, "Multicriteria assignment method PROAFTN: methodology and medical application," *European Journal of Operational Research*, vol. 125, no. 1, pp. 175–183, 2000.

[13] C. Zopounidis and M. Doumpos, "Business failure prediction using the UTADIS multicriteria analysis method," *Journal of the Operational Research Society*, vol. 50, no. 11, pp. 1138–1148, 1999.

[14] M. Doumpos and C. Zopounidis, "A multicriteria classification approach based on pairwise comparisons," *European Journal of Operational Research*, vol. 158, no. 2, pp. 378–389, 2004.

[15] S. Eppe, J. Roland, and Y. D. Smet, "On the use of valued action profiles for relational multi-criteria clustering," *International Journal of Multicriteria Decision Making*, vol. 4, no. 3, pp. 201–233, 2014.

[16] M. A. Boujelben, "A unicriterion analysis based on the PROMETHEE principles for multicriteria ordered clustering," *Omega*, vol. 69, pp. 126–140, 2017.

[17] P. Meyer and A. L. Olteanu, "Formalizing and solving the problem of clustering in MCDA," *European Journal of Operational Research*, vol. 227, no. 3, pp. 494–502, 2013.

[18] J. C. Cosset and J. Roy, "The determinants of country risk ratings," *Journal of International Business Studies*, vol. 22, no. 1, pp. 135–142, 1991.

[19] J. L. Mumpower, S. Livingston, and T. J. Lee, "Expert judgments of political riskiness," *Journal of Forecasting*, vol. 6, no. 1, pp. 51–65, 1987.

[20] Y. De Smet, P. Nemery, and R. Selvaraj, "An exact algorithm for the multicriteria ordered clustering problem," *Omega*, vol. 40, no. 6, pp. 861–869, 2012.

[21] Y. De Smet and F. Gilbart, *A Class Definition Method for Country Risk Problem*, 2001.

[22] L. Chen, Z. Xu, H. Wang, and S. Liu, "An ordered clustering algorithm based on K-means and the PROMETHEE method," *International Journal of Machine Learning and Cybernetics*, vol. 9, no. 6, pp. 917–926, 2018.

[23] X. Liu, H. Yu, G. Wang, and L. Guo, "A multi-criteria ordered clustering algorithm based on PROMETHEE," in *Proceedings of the Developments of Artificial Intelligence Technologies in Computation and Robotics: Proceedings of the 14th International FLINS Conference*, pp. 43–51, Cologne, Germany, August 2020.

[24] C. Bai, R. Zhang, L. Qian, L. Liu, and Y. Wu, "An ordered clustering algorithm based on fuzzy c-means and PROMETHEE," *International Journal of Machine Learning and Cybernetics*, vol. 10, no. 6, pp. 1423–1436, 2019.

[25] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Springer Science & Business Media, Heidelberg, Germany, 2013.

[26] X. Wang, Y. Wang, and L. Wang, "Improving fuzzy c-means clustering based on feature-weight learning," *Pattern Recognition Letters*, vol. 25, no. 10, pp. 1123–1132, 2004.

[27] Y. Zheng, B. Jeon, D. Xu, Q. J. Wu, and H. Zhang, "Image segmentation by generalized hierarchical fuzzy C-means algorithm," *Journal of Intelligent and Fuzzy Systems*, vol. 28, no. 2, pp. 961–973, 2015.

[28] Z. Xu and J. Wu, "Intuitionistic fuzzy C-means clustering algorithms," *Journal of Systems Engineering and Electronics*, vol. 21, no. 4, pp. 580–590, 2010.

[29] I. Beg and T. Rashid, "Fuzzy distance measure and fuzzy clustering algorithm," *Journal of Interdisciplinary Mathematics*, vol. 18, no. 5, pp. 471–492, 2015.

[30] M. Zarinbal, M. Fazel Zarandi, and I. B. Turksen, "Interval type-2 relative entropy fuzzy C-means clustering," *Information Sciences*, vol. 272, pp. 49–72, 2014.

[31] J. Clímaco and J. Craveirinha, "Multicriteria analysis in telecommunication network planning and design—problems and issues," in *Multiple Criteria Decision Analysis: State of the Art Surveys*, pp. 899–941, Springer, New York, NY, 2005.

[32] J. C. Bezdek, R. Ehrlich, and W. Full, "FCM: the fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2-3, pp. 191–203, 1984.

[33] J. P. Brans and B. Mareschal, "PROMETHEE methods," in *Multiple Criteria Decision Analysis: State of the Art Surveys*, J. Figueira, S. Greco, and M. Ehrgott, Eds., Springer, Newyork, NY, 2005.

[34] J. P. Brans and P. Vincke, "Note—a preference ranking organisation method: (the PROMETHEE method for multiple criteria decision-making)," *Management Science*, vol. 31, no. 6, pp. 647–656, 1985.

Hindawi
Mathematical Problems in Engineering
Volume 2023, Article ID 5926162, 8 pages
https://doi.org/10.1155/2023/5926162

Research Article

Financial Performance Assessment by a Type-2 Fuzzy Logic Approach

Hong Wang,[1] Samyabrata Bhattacharjee,[2] Nasreen Kausar,[3] Ardashir Mohammadzadeh,[4] Dragan Pamucar,[5] and Nasr Al Din Ide[6]

[1]*School of Tourism, Hainan University, Haikou 570228, China*
[2]*Department of Chemical Engineering, Haldia Institute of Technology, Pin-721657, Haldia, India*
[3]*Department of Mathematics, Faculty of Arts and Science, Yildiz Technical University, Esenler 34210, Istanbul, Turkey*
[4]*Multidisciplinary Center for Infrastructure Engineering, Shenyang University of Technology, Shenyang 110870, China*
[5]*Faculty of Organizational Sciences, University of Belgrade, Belgrade 11000, Serbia*
[6]*Department of Mathematics, University of Aleppo, Aleppo, Syria*

Correspondence should be addressed to Nasr Al Din Ide; ide1112002@yahoo.ca

Received 7 July 2022; Revised 4 August 2022; Accepted 26 November 2022; Published 5 May 2023

Academic Editor: Mukesh Soni

Any company must constantly innovate if they want to maintain its market share in the present cutthroat and unstable industry. Innovation has a big influence on consumer behavior, yet it goes against the principles of sustainability. The issue of sustainability has become crucial to their company's growth. In order to evaluate a business firm's sustainability performance statistically, a new and effective fuzzy logic tool is created. Evolution and assessment are performed by a novel interval type-2 fuzzy logic inference system. The judgment of the inference system is carried out on the basis of type-2 fuzzy logic (T2FL), principal component analysis (PCA), and statistical data analysis. The main input variables include corporate environmental performance (CEP) and corporate financial performance (CFP). The suggested approach can effectively examine a corporation's sustainable performance, according to experimental findings. A unique approach that makes use of language variables and if-then logic to assist quantitative business sustainability events is the link between CEP and CFP. The recommended test will provide senior administrative leaders with useful information to supervise natural concerns correctly and gauge their commitment to company success.

1. Introduction

The nature of CSP is generally enormously nonlinear, partially inconsistent and multidimensional, vague, and uncertain [1, 2]. A large number of operators, each competent to forecast the extent of business responses to environmental challenges, define the CSP. The main purpose of the CSP is to present an imprecise or uncertain model to evaluate a CEP and CFP. Many studies have recently documented and produced sustainable performance based on two dimensions, which comprise a number of significant affecting factors. There are two dimensions: CEP and CFP. They took some relevant measurements of the financial and environmental markers in order to analyze the two aspects.

The performance assessment has been widely studied. For example in [3], a fuzzy system is suggested to develop the conventional method based on complex proportional assessment approach. A fuzzy model is designed in [4] to construct a model for budget management. In the study of [5], a decision-making system is designed by fuzzy systems, and it is shown that the amusement accuracy of fuzzy approach is better than conventional step-wise weight assessment approach. A fuzzy assessing approach is suggested in [6], for the assessment and selection of fuel vehicles, and it is shown that the fuzzy approach leads to the vehicles with least decreasing carbon emissions. A risk assessment approach is studied in [7], and the risk awareness is classified. In the study of [8], an empirical approach is proposed to

analyze the financial scheme of private enterprises in China. The structural equation modeling approach is developed in [9], and the effect if COVID-19 is analyzed.

Creativity is the only way to solve any complex, multilevel problems [10] in every sector or in every domain [11]. Human development is dependent on the psychology of creativity. The employment of creative practices supports the efficacy of a continuous scientific literacy to assist society, and creative thinking influences the growth of cognitive capabilities [11]. Creativity is not synonymous with innovation since it may express in a variety of ways that are unrelated to commercial activity [12]. Although both are required to solve complicated technical, managerial, or social problems [10]. When innovative ideas are applied successfully, they are regarded as the germ of innovation inside a business [13]. A crucial component of innovation is change, as well as novelty. How do companies maintain a competitive edge in environments and markets that change quickly while still promoting long-term growth? One increasingly important way for organizations to achieve this is through sustainability-driven innovation strategies [14, 15].

The literature on the connection between creativity and sustainability has a knowledge gap. This limited understanding of the relationship is caused by the ideas' ambiguous perception [16]. Moldavska et al. [17] have introduced a special CSA technique for manufacturing industries where corporate sustainability is viewed as a situation and is connected to development goals, modeling the industrial firm utilizing systems representation estimates the evaluation tool, and corporate sustainability is seen as a procedure of constant innovation. Besne et al. [18] Utilize fuzzy logic with LCA to provide effective analytical methodologies. His work includes doing LCAs and cost analyses, normalizing the results of various financial and environmental influencing factors, implementing fuzzy logic by integrating the results, and then computing the fuzzy eco-efficiency index.

The fuzzy systems are extensively employed in different problems such as forecasting [19], controller designing [20], assessing systems [21], sentence representation technique [22], modeling problems [23], mapping problems [24], synchronization systems [25], and many others. However, it is rarely used for sustainability. In this study, a T2FS is developed for attaining corporate sustainability. An IT2FLC is an extensive version of T1-FLS, and it is more accurate and sensitive as reported in literature [26, 27]. Comparative reviews are tabulated in Table 1.

2. Principal Components Analysis

Principle components analysis (PCA) produces an orthogonal transformation from a set of observations of various variables. The order of the principle components is set up such that the first component explains the biggest proportion of variation in data, and the remaining in a decreasing order, explain the remaining proportions of variance. The principle component analysis (PCA) is a statistical method that converts a series of observations of variables that are presumably linked into a set of new values of variables that are linearly uncorrelated, or principal components. The PCA is used to create a low-dimensional representation of a set of data. The most varied linear combinations of variables that are unrelated to one another were presented. In addition to giving extracted parameters for supervised machine learning problems, the PCA may be used to visualize data. The first principal component of a set of features $y_1, y_2, \ldots, y_n$ is the normalized linear combination of the features as

$$Z_1 = \phi_{11} y_1 + = \phi_{21} y_2 + \cdots + \phi_{n1} y_n. \tag{1}$$

It takes the largest variance and it can be calculated by

$$\sum_{i}^{n} \alpha_{i1}^2 = 1. \tag{2}$$

Let us considered first principal component elements $\phi_{11}, \phi_{21}, \ldots, \phi_{n1}$ which are the loading of the first principal component and there exists some the principal component loading vectors $\phi_1 = (\phi_{21}, \phi_{21}, \ldots, \phi_{n1})^T$. The loadings are set up in such a way that the total of their squares equals one. Otherwise, arbitrarily large absolute values for these components may result in an arbitrarily bulky variance.

3. Motivation and Model Formulation

A T2-FLS is proposed to estimated the values of corporate environmental performance (CEP) and CEF. The designed T2-FLS is used to assess the sustainability of corporate. The general scheme is shown Figure 1. The developed model features a multi-variable, nonlinear structure that takes many economic factors into account. CEP and CFP are the two primary inputs of T2-FLS. The T2-FLS rules are optimized in a way that the input variables are used to model the corporate sustainability determinant (CORSUS). A novel method for quantifying CORSUS is produced by the developed rules based on linguistic characteristics. The following is a description of the key variables:

(i) WaC: it determines the influence of the company on water resource. Every industry has a limit to use minimum amount of water for product manufacture so that remaining are available for humans and other biotic source.

(ii) SS & TM: it measures the pollutant before they are exposed to ecosystem. Lower the discharge higher the sustainable value.

(iii) GHG: it comprises of release of GHG like carbon dioxide from various operation. It is essential to minimise the rate of discharge in atmosphere.

(iv) VOC: it measures the emission in atmosphere when the body work is painted i.e., body assembly plant.

(v) (SO_2 and NO_2): it measures the emission in atmosphere due to the burning of fossil fuel and from transport.

(vi) OIW: it is the solid waste obtain from industry which is landfill. It is essential for a company/industry to minimise the amount of waste landfill rather to reuse or recycle.

Table 1: Comparative review on sustainability performance assessment.

Authors, years	No. of inputs	Uncertainty/fuzzy	Statistical analysis
Wicher et al. 2019, [28]	3	Type-1	ANP
Pislaru et al. 2019, [29]	2	Type-1	PCA
Proposed model	2	Type-2	Statistical parameter

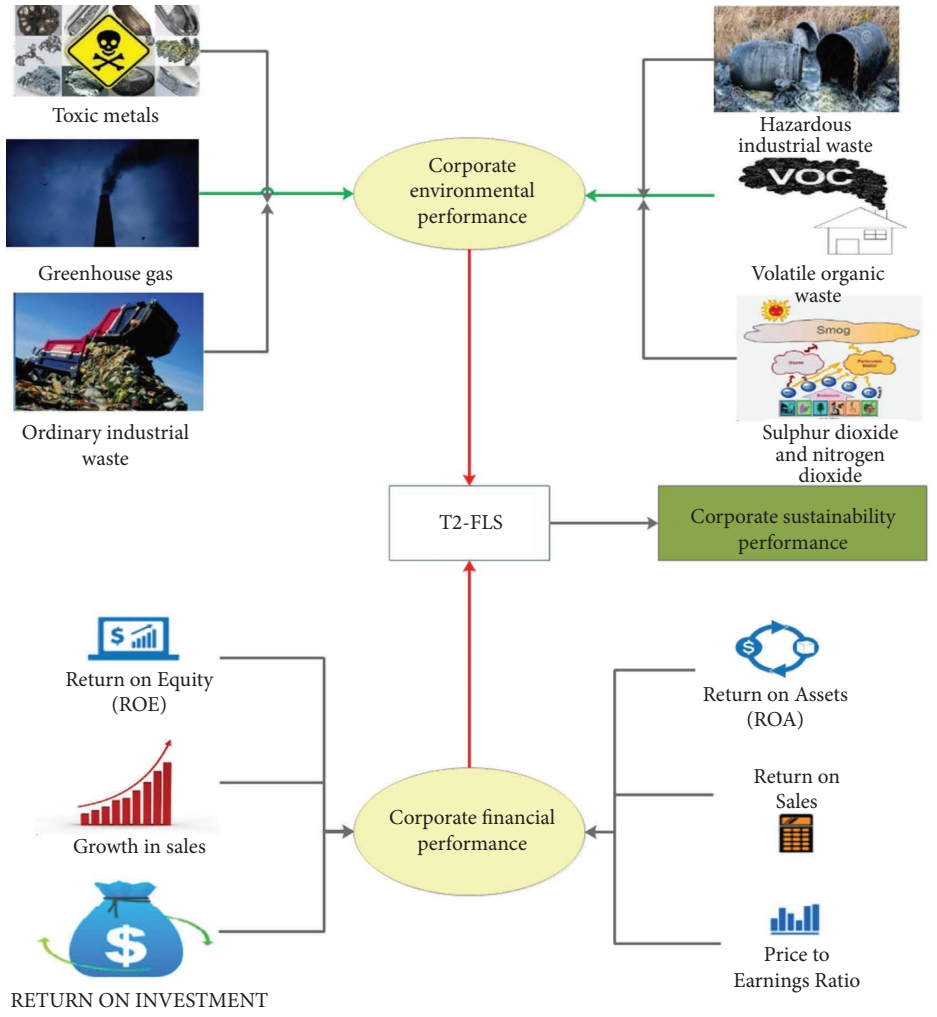


Figure 1: Systematic diagram for CSP model.

(vii) HIW: it includes all the waste which is harmful for biotic and abiotic component. It will be more sustainable for company if the amount of production of waste is less. In case of CFP consist of eight parameters which includes:

1 PE: it is a ratio that is used to compare the stock's current market price. Investors are more prepared to pay a larger share if the ratio is higher.

2 ROI: it is a ratio which is used to measure and compare the efficiency of investment.

3 ROE: it actually measures the net assets. It is use to determine how company use investment profitably.

4 ROA: it is used to measure the efficiency of the company to use asset in future profitably.

5 ROS: it measures the profit obtain by the company in selling the product.

6 OM: t measures the profit obtained by considering the cost of product sold and the operating expense.

7 GS: it signifies the difference obtain in previous sales record and current sales which is used to measure customer satisfaction.

8 DP: it measures the time taken by the company to pay the supplier. High ratio indicates company have insufficient cash.

PCA used the beginning factors to get the projection scores for the CEP and CFP magnitudes. CEP and CFP were valued for the two dimensions using the statistical method PCA by taking into account their initial points in the following equations. CEP is modeled as

$$\begin{aligned}\text{CEP} &= \beta_1 \text{WaC} + \beta_2 TM + \beta_3 SS + \beta_4 \text{GHG} + \beta_5 \text{VOC} \\ &\quad + \beta_6 SO_2 + \beta_7 NO_X + \beta_8 \text{OIW} + \beta_9 \text{HIW}, \\ \text{CFP} &= \delta_1 PE + \delta_2 \text{ROI} + \delta_3 \text{ROE} + \delta_4 \text{ROA} + \delta_5 \text{ROS} + \delta_6 OM + \delta_7 GS + \delta_8 DP.\end{aligned} \tag{3}$$

To evaluate the number of T2FS must be computing for determining the final indicator in corporate sustainable performance (CORSUS), positive values for μ and ν must be allocated which indicating the importance of the two input parameter in the development of the output parameter, CORSUS. The linguistic values are represented as follows. The value 0 is used for very bad, 1 is used for bad, 2 is used for medium, 3 is used for good, and 4 is for very good. For the weight of CEP and CFP, an index have allocated for corporate sustainable performance (CORSUS) of the overall sustainability. This index for CORSUS can be evaluated with the help of the following equation:

$$\text{CORSUS} = \mu \text{CEP} + \nu \text{CFP}. \tag{4}$$

When $\mu = \nu = 0$, the minimum value of CORSUS is zero and the maximum value is $4\mu + 4\nu$. To strike a balance between the two variables of sustainability's environmental and economical components, we have considered the hypothesis= ν. During this spontaneous process in a corporate sustainability, we have computed a new function pattern for the proposed model which is depicted in Figure 2.

4. Solution Procedure

In this section, the fuzzy approach is illustrated. The input variables are CEP and CFP, and the output is CORSUS. The structure is shown in Figure 3. As it is seen for inputs, we have 4 membership functions MFs. The MFs for CEP are denoted by $\widetilde{A}_i, i = 1, \ldots, 4$, and for CFP are denoted by $\widetilde{B}_i, i = 1, \ldots, 4$. The MFs represent the value of CEP and CFP in various range such as: very-bad ($\widetilde{A}_1$, $\widetilde{B}_1$), bad ($\widetilde{A}_2$, $\widetilde{B}_2$), good ($\widetilde{A}_3$, $\widetilde{B}_3$) and very-good ($\widetilde{A}_4$, $\widetilde{B}_4$) see Figures 4 and 5. Also, for CORSUS, 6 MFs are considered. The MFs of output are denoted by $\widetilde{O}_i, i = 1, \ldots, 6$ (see Figure 6). The MFs of output represent the value of CORSUS in various range such as: very-low$(\widetilde{O}_1)$, low $(\widetilde{O}_2)$, medium-low$(\widetilde{O}_3)$, medium $(\widetilde{O}_4)$, medium-high$(\widetilde{O}_5)$, and high $(\widetilde{O}_6)$ The details of computations are given as follows.

Step 1: get the inputs u_{CEP} and u_{CFP}, that represent CEP and CFP, respectively.

Step 2: compute the upper/lower membership of $\widetilde{A}_i, i = 1, \ldots, 4$ and $\widetilde{B}_i, i = 1, \ldots, 4$ as

$$\begin{aligned}\overline{\mu}_{\widetilde{A}_i}(\text{CEP}) &= \exp\left(-\frac{\left(\text{CEP} - m_{\widetilde{A}_i}\right)^2}{\overline{\sigma}^2_{\widetilde{A}_i}}\right), \\ \underline{\mu}_{\widetilde{A}_i}(\text{CEP}) &= \exp\left(-\frac{\left(\text{CEP} - m_{\widetilde{A}_i}\right)^2}{\underline{\sigma}^2_{\widetilde{A}_i}}\right), \\ \overline{\mu}_{\widetilde{B}_i}(\text{CFP}) &= \exp\left(-\frac{\left(\text{CFP} - m_{\widetilde{B}_i}\right)^2}{\overline{\sigma}^2_{\widetilde{B}_i}}\right), \\ \underline{\mu}_{\widetilde{B}_i}(\text{CFP}) &= \exp\left(-\frac{\left(\text{CFP} - m_{\widetilde{B}_i}\right)^2}{\underline{\sigma}^2_{\widetilde{B}_i}}\right),\end{aligned} \tag{5}$$

where $m_{\widetilde{A}_i}$, $\overline{\sigma}_{\widetilde{A}_i}$, and $\sigma_{\underset{A_i}{\sim}}$ denote the center of $\widetilde{A}_i$, upper and lower width of $\widetilde{A}_i$, respectively. Similarly, $m_{\widetilde{B}_i}$, $\overline{\sigma}_{\widetilde{B}_i}$, and $\sigma_{\underset{B_i}{\sim}}$ denote the center of $\widetilde{B}_i$, upper and lower width of $\widetilde{B}_i$, respectively.

Step 3: the rules are given as follows:

1 If u_{CEP} is $\widetilde{A}_1$ and u_{CFP} is $\widetilde{B}_1$, Then y_{CORSUS} is $\widetilde{O}_1$
2 If u_{CEP} is $\widetilde{A}_1$ and u_{CFP} is $\widetilde{B}_2$, Then y_{CORSUS} is $\widetilde{O}_1$
3 If u_{CEP} is $\widetilde{A}_1$ and u_{CFP} is $\widetilde{B}_3$, Then y_{CORSUS} is $\widetilde{O}_2$
4 If u_{CEP} is $\widetilde{A}_1$ and u_{CFP} is $\widetilde{B}_4$, Then y_{CORSUS} is $\widetilde{O}_3$
5 If u_{CEP} is $\widetilde{A}_1$ and u_{CFP} is $\widetilde{B}_5$, Then y_{CORSUS} is $\widetilde{O}_4$
6 If u_{CEP} is $\widetilde{A}_2$ and u_{CFP} is $\widetilde{B}_1$, Then y_{CORSUS} is $\widetilde{O}_1$
7 If u_{CEP} is $\widetilde{A}_2$ and u_{CFP} is $\widetilde{B}_2$, Then y_{CORSUS} is $\widetilde{O}_2$
8 If u_{CEP} is $\widetilde{A}_2$ and u_{CFP} is $\widetilde{B}_3$, Then y_{CORSUS} is $\widetilde{O}_3$
9 If u_{CEP} is $\widetilde{A}_2$ and u_{CFP} is $\widetilde{B}_4$, Then y_{CORSUS} is $\widetilde{O}_4$
10 If u_{CEP} is $\widetilde{A}_2$ and u_{CFP} is $\widetilde{B}_5$, Then y_{CORSUS} is $\widetilde{O}_5$
11 If u_{CEP} is $\widetilde{A}_3$ and u_{CFP} is $\widetilde{B}_1$, Then y_{CORSUS} is $\widetilde{O}_2$
12 If u_{CEP} is $\widetilde{A}_3$ and u_{CFP} is $\widetilde{B}_2$, Then y_{CORSUS} is $\widetilde{O}_3$
13 If u_{CEP} is $\widetilde{A}_3$ and u_{CFP} is $\widetilde{B}_3$, Then y_{CORSUS} is $\widetilde{O}_4$
14 If u_{CEP} is $\widetilde{A}_3$ and u_{CFP} is $\widetilde{B}_4$, Then y_{CORSUS} is $\widetilde{O}_5$
15 If u_{CEP} is $\widetilde{A}_3$ and u_{CFP} is $\widetilde{B}_5$, Then y_{CORSUS} is $\widetilde{O}_6$
16 If u_{CEP} is $\widetilde{A}_4$ and u_{CFP} is $\widetilde{B}_1$, Then y_{CORSUS} is $\widetilde{O}_3$
17 If u_{CEP} is $\widetilde{A}_4$ and u_{CFP} is $\widetilde{B}_2$, Then y_{CORSUS} is $\widetilde{O}_4$
18 If u_{CEP} is $\widetilde{A}_4$ and u_{CFP} is $\widetilde{B}_3$, Then y_{CORSUS} is $\widetilde{O}_5$
19 If u_{CEP} is $\widetilde{A}_4$ and u_{CFP} is $\widetilde{B}_4$, Then y_{CORSUS} is $\widetilde{O}_6$

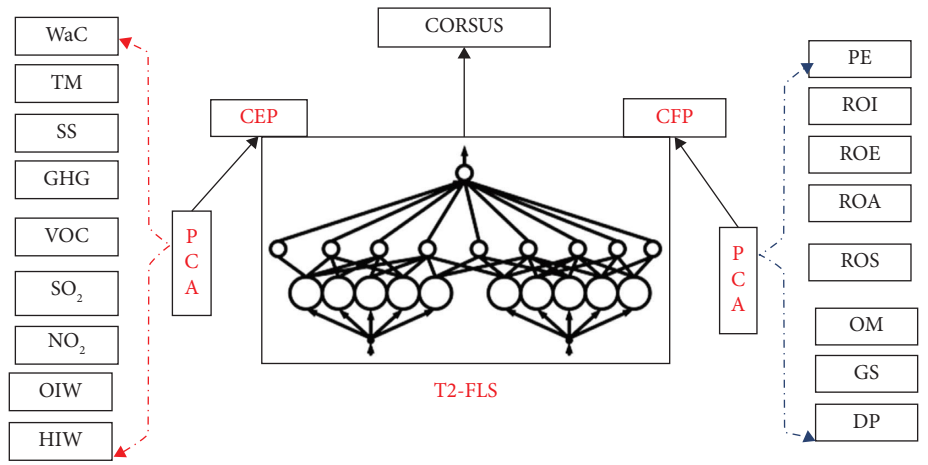


Figure 2: Block diagram for T2-FIS.

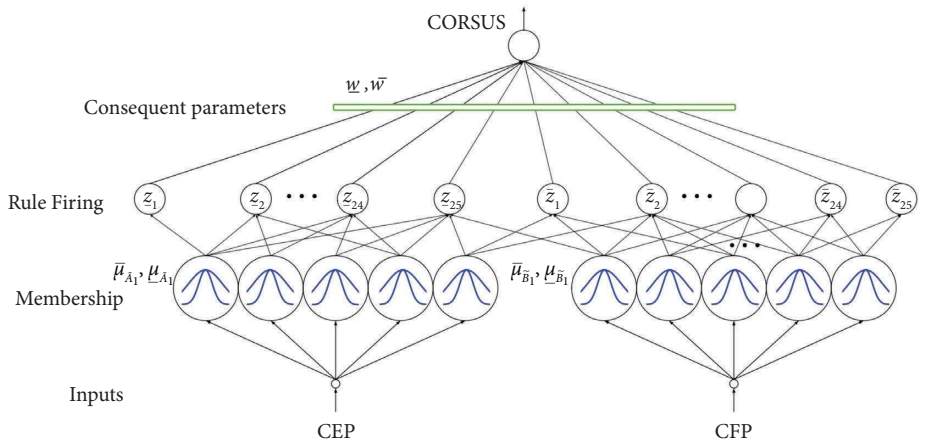


Figure 3: The designed FLS.

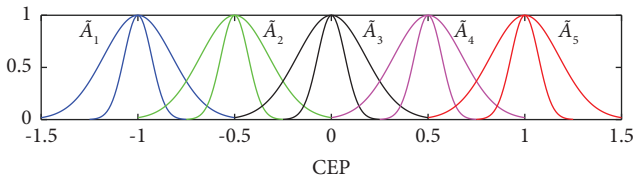


Figure 4: MFs for CEP indicators.

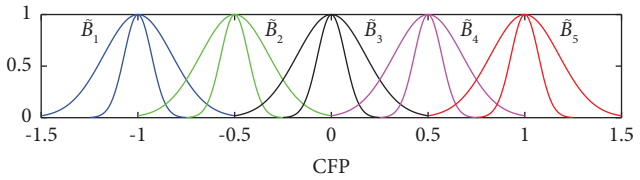


Figure 5: MFs for CFP indicators.

20 If u_{CEP} is $\tilde{A}_4$ and u_{CFP} is $\tilde{B}_5$, Then y_{CORSUS} is $\tilde{O}_1$
21 If u_{CEP} is $\tilde{A}_5$ and u_{CFP} is $\tilde{B}_1$, Then y_{CORSUS} is $\tilde{O}_4$
22 If u_{CEP} is $\tilde{A}_5$ and u_{CFP} is $\tilde{B}_2$, Then y_{CORSUS} is $\tilde{O}_5$
23 If u_{CEP} is $\tilde{A}_5$ and u_{CFP} is $\tilde{B}_3$, Then y_{CORSUS} is $\tilde{O}_6$
24 If u_{CEP} is $\tilde{A}_5$ and u_{CFP} is $\tilde{B}_4$, Then y_{CORSUS} is $\tilde{O}_1$
25 If u_{CEP} is $\tilde{A}_5$ and u_{CFP} is $\tilde{B}_5$, Then y_{CORSUS} is $\tilde{O}_1$

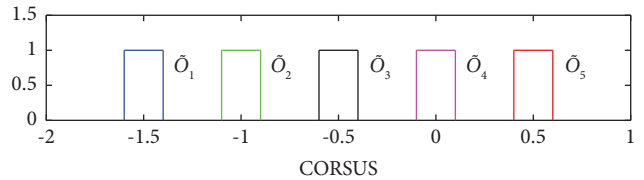


Figure 6: MFs for CORSUS.

If p_1-th MF is fired for CEP, and p_2-th MF is fired for CFP, then the upper/lower rule firing are obtained as

$$\begin{aligned}\overline{z}_l &= \overline{\mu}_{\tilde{A}_{p_1}}(\text{CEP})\overline{\mu}_{\tilde{B}_{p_2}}(\text{CFP}),\\ \underline{z}_l &= \underline{\mu}_{\tilde{A}_{p_1}}(\text{CEP})\underline{\mu}_{\tilde{B}_{p_2}}(\text{CFP}).\end{aligned} \tag{6}$$

Step 4: the CORSUS is computed as

$$\text{CORSUS} = \frac{\sum_{l=1}^{25}(\overline{z}_l\overline{w}_l + \underline{z}_l\underline{w}_l)}{\sum_{l=1}^{25}\overline{z}_l + \underline{z}_l}, \tag{7}$$

where $\underline{w}_l$ and $\overline{w}_l$ represent the mean of l-th fired MF in output.

5. Results Analysis

We constructed a framework for optimizing CORSUS assessment using an imprecise data set from the field and two input parameters and one output parameter. To make proper decisions, the IT2-FLC was used. Some statistical data analysis was successfully computed to monitor and access the sensitivity and accuracy of the suggested model on. The statistical techniques such as (a) (R^2) and (b) RMSE have been calculated and compared with the predicted and measured values of flexibility of the proposed model. The value of RMSE is defined by the following mathematical equation:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(a_i - b_i)^2}, \tag{8}$$

where a_i denotes the actual values of the output, b_i denotes the predicted values of the output, and n is the number of the points. In the following (9), we have calculated the determination coefficient (R^2) as

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(a_i - b_i)^2}{\sum_{i=1}^{n} b_i^2}. \tag{9}$$

The average of the squares of the errors can be measured by the value of MAPE. The very smaller values of MAPE make sure better performance. The MAPE has been calculated by using the (10) as

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|(a_i - b_i)|}{b_i} \times 100\%. \tag{10}$$

The performance of the proposed models' efficiency has been assessed using mean absolute error as a benchmark (MAE). The MAE can be determined using the following formula:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|a_i - b_i|. \tag{11}$$

Here, we have considered n number of data patterns from the industry data set, x_{pred_i} denotes the predicted value of one data point i and x_{obs_i} denotes the observed value of one data point i. The summary of statistical data analysis of output has been given in the Table 2.

6. Simulation Results and Discussion

There are two types of data collected from an industry (CEP and CFP) and we have used the IT2FIS. The IT2FS retrieves these data and, as an output, develops a sustainability performance model based on the data characteristics shown in Figure 7.

6.1. Managerial Inferences. The 2008 financial and economic crisis has led to worries about the impact of CEP and CFP. Banks' CSR policies have been questioned as a real strategic commitment to the major stakeholders in light of the controversies and scandals surrounding their participation in the crisis. These scandals and controversies have also brought to light shortcomings in several CSP areas. This study demonstrates that merely executing socially and environmentally conscious actions is inadequate to boost financial success; rather, management's disregard for such practices hurts financial performance. Instead, taking advantage of corporate sustainability performance benefits depends on the industry's ability to profit from such efforts, since properly managing CORSUS fosters stakeholder linkages that boost performance. For this reason, it is crucial for any industry to grasp that "being excellent" is not enough and that "how they need to be good" by striking the right balance between CEP and CFP is crucial. It is now necessary to have this degree of industrial mindfulness and awareness.

Table 2: Statistical parameters for sensitivity analysis of T2 and T1-FLS.

FIS	RMSE	R^2	MAPE
T1-FLS	0.115	0.952	4.823
T2-FLS	0.062	0.991	3.971

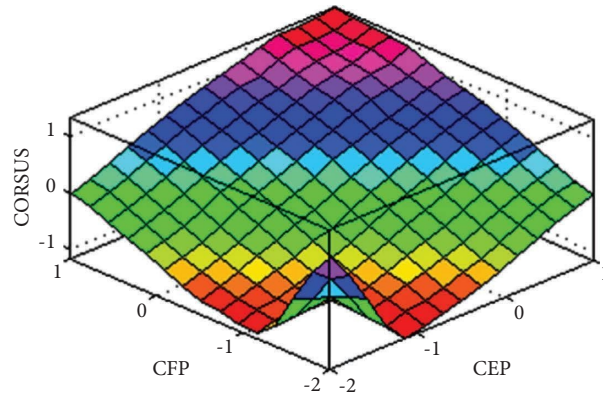


Figure 7: The diagram of CORSUS.

Remark 1. In this study, a type-2 FLS is introduced for attaining corporate sustainability. More recently the new type-3 FLSs have been presented with better ability in practical problems. For the future studies, the suggested approach can be developed by the use of type-3 FLSs, new risk analysis techniques, and new optimization methods [11, 30–37].

7. Conclusions

For the purpose of evaluating corporate sustainable performance (CSP) assessment, we have created and formulated an interval type-2 fuzzy logic control technique. The input parameters (CEP and CFP) and an output parameter were helped to track by the IT2FLC (CORSUS). In order to foresee a range of corporate environmental and financial performance assessment circumstances for an industry, this study contains a safe and reliable approach for drawing conclusions. Because it is more difficult and complex to make judgments from the perspective of the industry,

including significant uncertainty, we have taken for the implementation of CSP with the aid of this specific IT2FLC by employing Principal components analysis (PCA). The prediction of different economic, social, and environmental factors aids in CORSUS optimization during an industry's CSP monitoring system. The authors claim that the recommended IT2FLC strategy is a brand-new and distinctive way to build and regulate the CSP that must be met in order to provide a successful workplace output.

Nomenclature

WaC:	Water consumption
SS:	Suspended solids
OIW:	Ordinary industrial waste
TM:	Toxic metals
GHG:	Greenhouse gas
VOC:	Volatile organic compound
SO_2:	Sulphur dioxide
NO_2:	Nitrogen oxides
ROA:	Return on assets
HIW:	Hazardous industrial waste
CFP:	Corporate financial performance
PE:	Price-earnings
ROI:	Return on investment
ROE:	Return on equity
OM:	operating margin
GS:	Growth in sales
ROS:	Return on sales
DP:	Days payable
CEP:	Corporate environmental performance
CORSUS:	Corporate sustainable performance
CSP:	Corporate sustainability performance
EI:	Eco-innovation
LCA:	Life Cycle Assessment.

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 study was financially supported by the key project of the National Social Science Foundation (Grant no. 18AJY013); the Social Science Program of Hainan Province (Grant no. HNSK(YB)19-10); and the Program of the Ministry of Culture and Tourism (Grant no. TYETP201552).

References

[1] M. Tan-Mullins and G. Mohan, "The potential of corporate environmental responsibility of Chinese state-owned enterprises in Africa," *Environment, Development and Sustainability*, vol. 15, pp. 265–284, 2013.

[2] S. Basu, M. Roy, and P. Pal, "Corporate greening in a large developing economy: pollution prevention strategies," *Environment, Development and Sustainability*, vol. 21, pp. 1603–1633, 2019.

[3] Y. A. Ünvan and C. Ergenç, "Financial performance analysis with the fuzzy COPRAS and entropy-COPRAS approaches," *Computational Economics*, vol. 59, no. 4, pp. 1577–1605, 2022.

[4] R. Venugopal, C. Veeramani, and S. A. Edalatpanah, "Analysis of fuzzy DEMATEL approach for financial ratio performance evaluation of NASDAQ exchange," in *Proceedings of the International Conference on Data Science and Applications*, pp. 637–648, Springer, Singapore, November 2022.

[5] Ž. Stević, D. K. Das, R. Tešić, M. Vidas, and D. Vojinović, "Objective criticism and negative conclusions on using the fuzzy SWARA method in multi-criteria decision making," *Mathematics*, vol. 10, p. 635, 2022.

[6] I. M. Hezam, A. R. Mishra, P. Rani et al., "A hybrid intuitionistic fuzzy-MEREC-RS-DNMA method for assessing the alternative fuel vehicles with sustainability perspectives," *Sustainability*, vol. 14, no. 9, p. 5463, 2022.

[7] Y. Chen and V. Sivakumar, "Investigation of finance industry on risk awareness model and digital economic growth," *Annals of Operations Research*, pp. 1–22, 2021.

[8] X. T. Lei, Q. Y. Xu, and C. Z. Jin, "Nature of property right and the motives for holding cash: empirical evidence from Chinese listed companies," *Managerial and Decision Economics*, vol. 43, no. 5, pp. 1482–1500, 2022.

[9] Y. Wu and W. Zhu, "The role of CSR engagement in customer-company identification and behavioral intention during the COVID-19 pandemic," *Frontiers in Psychology*, vol. 12, p. 721410, 2021.

[10] N. Anderson, K. Potočnik, and J. Zhou, "Innovation and creativity in organizations: a state-of-the-science review, prospective commentary, and guiding framework," *Journal of Management*, vol. 40, no. 5, pp. 1297–1333, 2014.

[11] C. De Lucia, P. Balena, M. R. Stufano Melone, and D. Borri, "Policy, entrepreneurship, creativity and sustainability: the case of 'principi attivi' ('Active ingredients') in apulia region (southern Italy)," *Journal of Cleaner Production*, vol. 135, pp. 1461–1473, 2016.

[12] C. Lane and D. Lup, "Cooking under fire: managing multilevel tensions between creativity and innovation in haute cuisine," *Industry & Innovation*, vol. 22, no. 8, pp. 654–676, 2015.

[13] B. K. B. Joo, G. N. McLean, and B. Yang, "Creativity and human resource development: an integrative literature review and a conceptual framework for future research," *Human Resource Development Review*, vol. 12, no. 4, pp. 390–421, 2013.

[14] S. Paramanathan, C. Farrukh, R. Phaal, and D. Probert, "Implementing industrial sustainability: the research issues in technology management," *R & D Management*, vol. 34, no. 5, pp. 527–537, 2004.

[15] N.. Roome, "Business strategy, R&D management and environmental imperatives," *R & D Management*, vol. 24, no. 1, pp. 065–082, 1994.

[16] P. H. Andersen and H. Kragh, "Managing creativity in business market relationships," *Industrial Marketing Management*, vol. 42, no. 1, pp. 82–85, 2013.

[17] A. Moldavska and T. Welo, "A Holistic approach to corporate sustainability assessment: incorporating sustainable development goals into sustainable manufacturing performance evaluation," *Journal of Manufacturing Systems*, vol. 50, pp. 53–68, 2019.

[18] A. G. Besné, D. Luna, A. Cobos, D. Lameiras, H. Ortiz-Moreno, and L. P. Güereca, "A methodological framework of eco-efficiency based on fuzzy logic and Life Cycle Assessment applied to a Mexican SME," *Environmental Impact Assessment Review*, vol. 68, pp. 38–48, 2018.

[19] S. Mustafa, A. A. Bajwa, and S. Iqbal, "A new fuzzy grach model to forecast stock market technical analysis," *Operational Research in Engineering Sciences: Theory and Applications*, vol. 5, no. 1, pp. 185–204, 2022.

[20] M. R. Gharib, "Comparison of robust optimal QFT controller with TFC and MFC controller in a multi-input multi-output system," *Reports in Mechanical Engineering*, vol. 1, no. 1, pp. 151–161, 2020.

[21] V. Lukovac and M. Popović, "Fuzzy Delphi approach to defining a cycle for assessing the performance of military drivers," *Decision Making: Applications in Management and Engineering*, vol. 1, no. 1, pp. 67–81, 2018.

[22] W. Zheng, X. Liu, and L. Yin, "Sentence representation method based on multi-layer semantic network," *Applied Sciences*, vol. 11, no. 3, p. 1316, 2021.

[23] W. Zheng, L. Yin, X. Chen, Z. Ma, S. Liu, and B. Yang, "Knowledge base graph embedding module design for Visual question answering model," *Pattern Recognition*, vol. 120, Article ID 108153, 2021.

[24] T. Rasham, M. S. Shabbir, P. Agarwal, and S. Momani, "On a pair of fuzzy dominated mappings on closed ball in the multiplicative metric space with applications," *Fuzzy Sets and Systems*, vol. 437, no. 2022, pp. 81–96, 2022.

[25] E. Zambrano-Serrano, S. Bekiros, M. A. Platas-Garza et al., "On chaos and projective synchronization of a fractional difference map with no equilibria using a fuzzy-based state feedback control," *Physica A: Statistical Mechanics and Its Applications*, vol. 578, Article ID 126100, 2021.

[26] D. K. Jana, S. Roy, S. Bhattacharjee, P. Dostal, and S. Roy, "Saw dust-derived activated carbon in different impregnation ratios and its application in de-fluoridation of waste water using it2flc and rsm," *Biomass Conv. Bioref*, 2021.

[27] S. Bhattacharjee, K. A. Alattas, F. F. M. El-Sousy et al., "A type-2 fuzzy logic approach for forecasting of effluent quality parameters of wastewater treatment," *Mathematical Problems in Engineering*, vol. 2022, Article ID 1965157, 10 pages, 2022.

[28] P. Wicher, F. Zapletal, and R. Lenort, "Sustainability performance assessment of industrial corporation using Fuzzy Analytic Network Process," *Journal of Cleaner Production*, vol. 241, Article ID 118132, 2019.

[29] M. Pislaru, I. V. Herghiligiu, and I. B. Robu, "Corporate sustainable performance assessment based on fuzzy logic," *Journal of Cleaner Production*, vol. 223, pp. 998–1013, 2019.

[30] B. A. Hennessey and T. M. Amabile, *Annual Review of Psychology*, vol. 61, no. 1, pp. 569–598, 2010.

[31] J. K. Hall, G. A. Daneke, and M. J. Lenox, "Sustainable development and entrepreneurship: past contributions and future directions," *Journal of Business Venturing*, vol. 25, no. 5, pp. 439–448, 2010.

[32] J. Klewitz and E. G. Hansen, "Sustainability-oriented innovation of SMEs: a systematic review," *Journal of Cleaner Production*, vol. 65, pp. 57–75, 2014.

[33] S. Kristiansen, "Social networks and business success: the role of subcultures in an African context," *The American Journal of Economics and Sociology*, vol. 63, no. 5, pp. 1149–1171, 2004.

[34] P. Marius, D. Trandabat, and A. Trandabat, "Assessment of corporate environmental performance based on fuzzy approach," *APCBEE Procedia*, vol. 5, pp. 368–372, 2013.

[35] A. H. Rahdari, "Developing a fuzzy corporate performance rating system: a petrochemical industry case study," *Journal of Cleaner Production*, vol. 131, pp. 421–434, 2016.

[36] S. Rajak and S. Vinodh, "Application of fuzzy logic for social sustainability performance evaluation: a case study of an Indian automotive component manufacturing organization," *Journal of Cleaner Production*, vol. 108, pp. 1184–1192, 2015.

[37] A. Kumar and R. Anbanandam, "Assessment of environmental and social sustainability performance of the freight transportation industry: an index-based approach," *Transport Policy*, vol. 124, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2023, Article ID 5143620, 9 pages
https://doi.org/10.1155/2023/5143620

Research Article

Nonlinear Energy Optimization in the Wireless Sensor Network through NN-LEACH

Avinash Bhagat,[1] **Manmohan Sharma**,[1] **Ajay Shriram Kushwaha**,[2] **Shilpa Sharma**,[3] **and Hussien Sobahi Mohammed**[4]

[1]*School of Computer Application, Lovely Professional University, Phagwara, Punjab, India*
[2]*School of CS & IT, Jain (Deemed-to-be-University), Bangalore, Karnataka, India*
[3]*School of Computer Science & Engineering, Lovely Professional University, Phagwara, Punjab, India*
[4]*University of Gezira, Wad Medani, Sudan*

Correspondence should be addressed to Hussien Sobahi Mohammed; hussiensobahi@uofg.edu.sd

Received 27 June 2022; Revised 20 August 2022; Accepted 6 April 2023; Published 30 April 2023

Academic Editor: A. M. Bastos Pereira

Researchers have developed a range of methods and strategies to decrease wireless sensor network energy consumption. Mote clustering is one of the competent topological control approaches to boost the networks' energy efficiency, scalability, and performance. Energy is dissipated during the cluster creation, cluster head selection, routing from the head and base stations of the cluster, and data aggregation—clustering and routing emphasis on the stability and the longevity of the network. This research work provides the optimization technique for the wireless sensor network to optimize the energy through NN-LEACH. The main goal is to extend network life and reduce power consumption by clustering and routing sensor nodes using the two-step NN-LEACH protocol, which is suggested. An additional goal might be to establish the appropriate course of action for the suggested approach for this network.

1. Introduction to Wireless Sensor Networks

Wireless sensor networks (WSNs) offer huge potential to improve people's lives and study the environment using sensor nodes. The key obstacle in using these motes is independent and low-powered devices with small batteries. According to the NSF, new technologies lower motes and sensor arrays' cost, size, and weight. Integration approaches reduce distance barriers by increasing performance and longevity while decreasing size [1]. WSNs are typically hidden and rely on small batteries for power. Changing batteries is not an option. Managing the power requirements of a wireless sensor network is difficult. Contrary to popular belief, energy consumption is not determining network efficiency [2, 3]. Figure 1 shows the example of the wireless sensor network [4].

Researchers focused on layer-based components, expecting that changes in one layer would immediately influence the whole system, but it did not work. Many energy-saving models consider sending and receiving data, ignoring other factors. Most contemporary energy reduction models ignore rest parameters [5].

2. Literature Review

The energy-efficient routing algorithm is categorized as communication architecture, network structure, reliable routing, and topology-based routing [6]. These protocols might be flat or hierarchical [7]. In the WSN, flat routing algorithms do comparable responsibilities. Smaller networks frequently use these networks. Algorithms such as SPIN and directed diffusion are examples of balanced routing algorithms. The cluster head rotates to balance the energy consumed. The hierarchical routing algorithm's architecture is efficient and scalable. Each network mote receives the data via the SPIN protocol. The data of each node are comparable

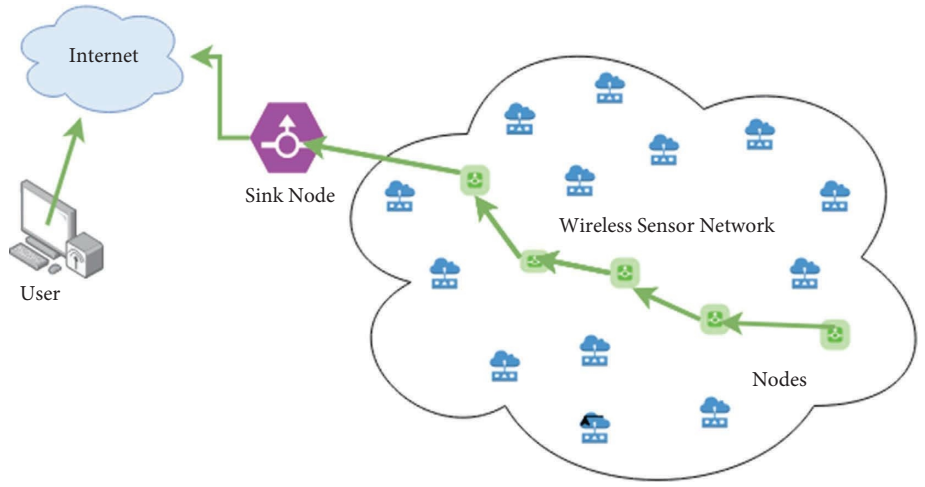


Figure 1: Example of wireless sensor network [4].

Table 1: Analysis of successors of LEACH [16].

Year	LEACH variants	Cluster types	Level of complexity	Delay in transmission	Energy efficiency	Overhead	Scalability
2000	LEACH	Distributed	Low	Less	Moderate	Higher	Lower
2002	LEACH-C single hop	Centralized	Low	Less	Highly efficient	Less	Lower
2002	LEACH DCHS	Distributed	Moderate	Less	Highly efficient	Higher	Lower
2003	LEACH B	Distributed	Moderate	Less	Highly efficient	Higher	Lower
2005	LEACH S	Distributed	High	Less	Very high	Higher	Moderate
2006	LEACH B- TL	Distributed	Lower	Less	Highly efficient	Less	Lower
2006	LEACH M	Distributed	High	Less	Highly efficient	Higher	Higher
2007	LEACH E	Distributed	Extreme	Less	Highly efficient	Higher	Lower
2008	LEACH-TB	Distributed	Extreme	Less	Moderate	Higher	Moderate
2008	LEACH-ME	Distributed	Extreme	Higher	Moderate	Higher	Higher
2010	LEACH-U	Distributed	Extreme	Less	Highly efficient	Less	Lower
2010	LEACH-C multihop	Distributed	Extreme	Higher	Highly efficient	Higher	Lower
2010	LEACH D	Distributed	High	Less	Very high	Higher	Very high
2011	LEACH-GA	Distributed	High	Less	Highly efficient	Higher	Lower
2011	LEACH-FZ	Distributed	Extreme	Higher	Highly efficient	Higher	Higher
2012	LEACH-FL	Distributed	Extreme	Less	Lower	Less	Higher
2012	LEACH-MR	Distributed	High	Higher	Highly efficient	Less	Lower
2012	LEACH-CELL	Distributed	Extreme	Less	Moderate	Very high	Very high
2013	LEACH-EP	Distributed	Extreme	Less	Very high	Higher	Lower
2013	LEACH-I	Distributed	High	Less	Highly efficient	Moderate	Lower
2014	LEACH-SAGA	Distributed	High	Less	Highly efficient	Moderate	Higher
2015	LEACH-V	Distributed	High	Less	Very high	Higher	Lower
2019	LEACH-MG	Distributed	Complex	Less	Moderate	Very high	Very high
2020	LEACH-MW	Adaptive	Low	Less	Lower	Very high	Very high
2021	LEACH-ESO	Distributed	Extreme	Less	Highly efficient	Moderate	Lower
2022	LEACH-IACO	Adaptive	Extreme	Less	Highly efficient	Very high	Very high

to those of its neighbor's. This protocol disseminates information to all motes when the user does not require data to be transferred between motes.

In 2000, Heinzelman et al. [8] introduced the first hierarchical method, LEACH ("low energy adaptive clustering hierarchy"). A strategy for geographically segmenting a system into smaller cells is suggested by Naghibi and Barati [9]. Each cell can contain both single-hop and multihop cells. A novel EGPRM is being used to collect data from sensor nodes using two portable sinks. EECS [10], HEED [11], PEGASIS [12], TEEN [13], LEACH-IACO [14], and T-LEACH [15] are all based on LEACH. LEACH and its derivatives aim to improve energy efficiency through coverage, data aggregation, data protection minimal latency, resilience, and scalability, and the main goal of these algorithms is to save energy. Table 1 compares various LEACH variants.

3. Network Model

The nodes are chosen at random. The starting energy of a node is E_0. The MS moves along the y-axis. We know that

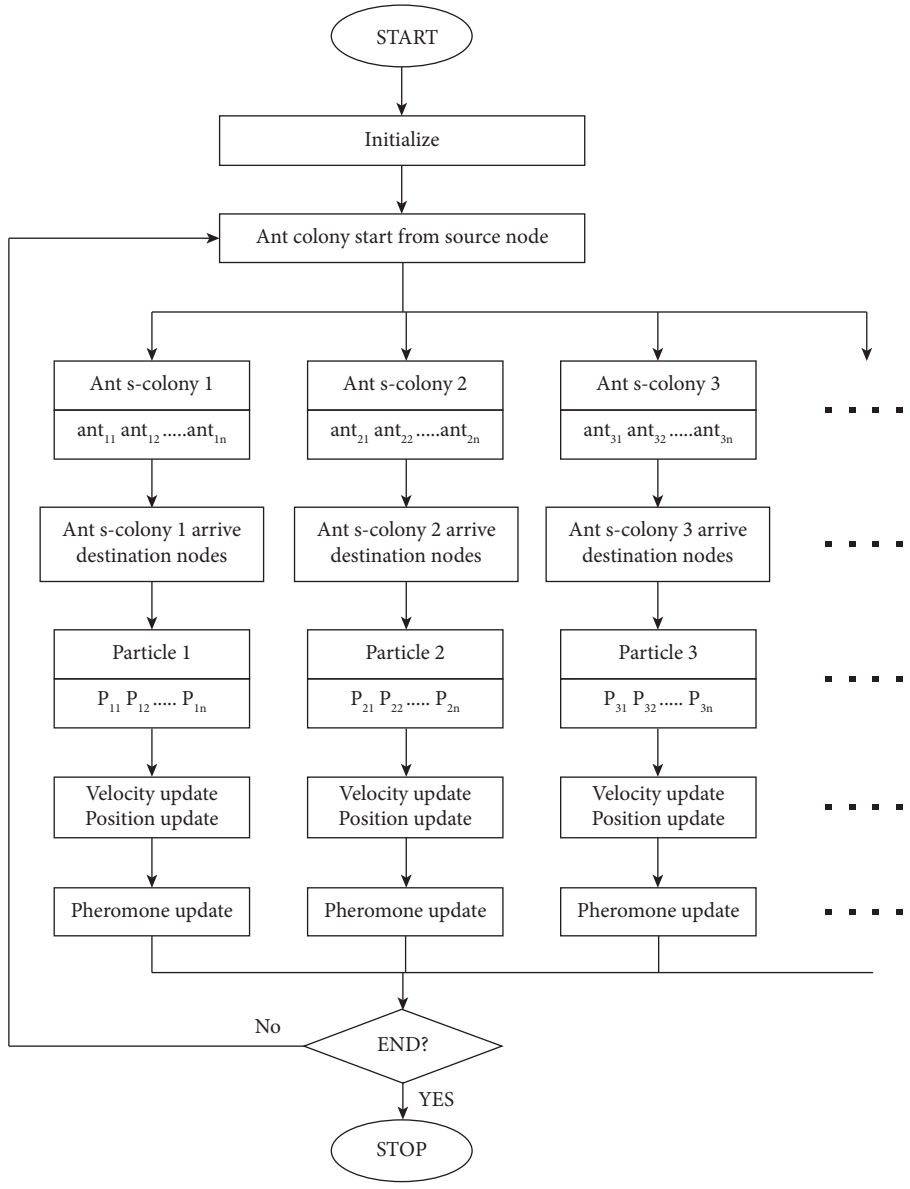


Figure 2: Flowchart of hybrid ACO/PSO.

the base station has unlimited power and that sensor nodes and sinks are everywhere. Now, we have setup and steady phases.

3.1. Step-Up Phase. The setup phase is divided into three stages: the Task Presentation (TP), the Selection of cluster head (CH), and rendezvous node (RN). Clustering, i.e., creating clusters and election of the cluster head are carried out during the next part of the setup phase [17, 18]. After the cluster formation, the last stage is known as scheduling (S). In the entire scheduling phase, the message is disseminated from the cluster head to every cluster member. Every node itself arranges their organization in the period of transmission.

3.2. Task Ordination. In this phase, firstly, RNs are selected. Initially, all nodes are assumed to be normal nodes. Every hub themselves choose whether they meet RN condition or not. To become a rendezvous hub, the hubs must fulfil a requirement. The situation representing RN is specified below the following equation:

$$\frac{y_w}{2}(1 + \mathrm{Rx}) <= y_y <= \frac{y_w}{2}(1 - \mathrm{Rx}). \quad (1)$$

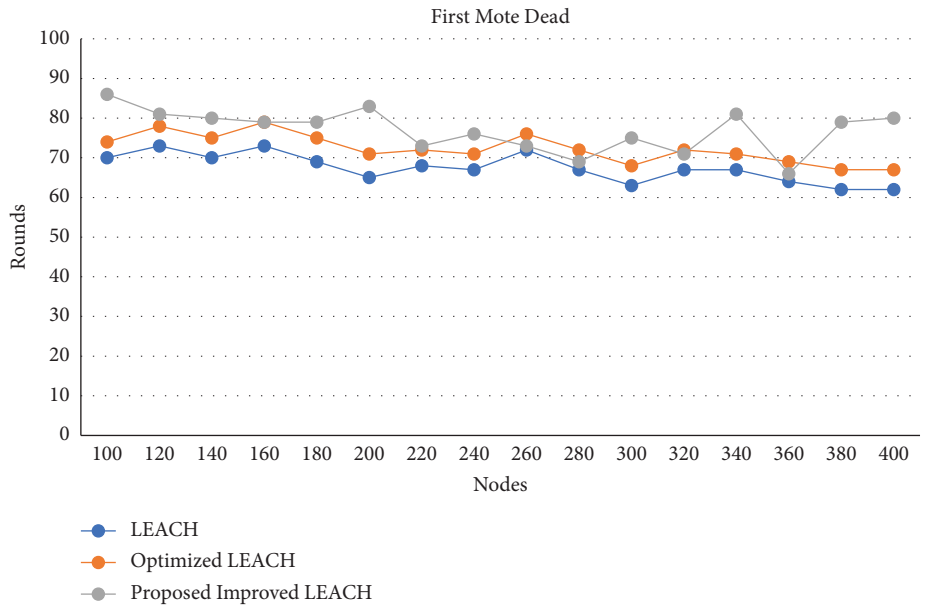
Following points are needed to be considered to get the cluster head using the method of HNN:

(a) Prepare weights $T_{xy} = \sum_{c=0}^{M-1} i_x^c i_y^c \, x \neq y$, where i_x^c is element of class c exemplar

(b) Put on input to the desired outputs $z = i$

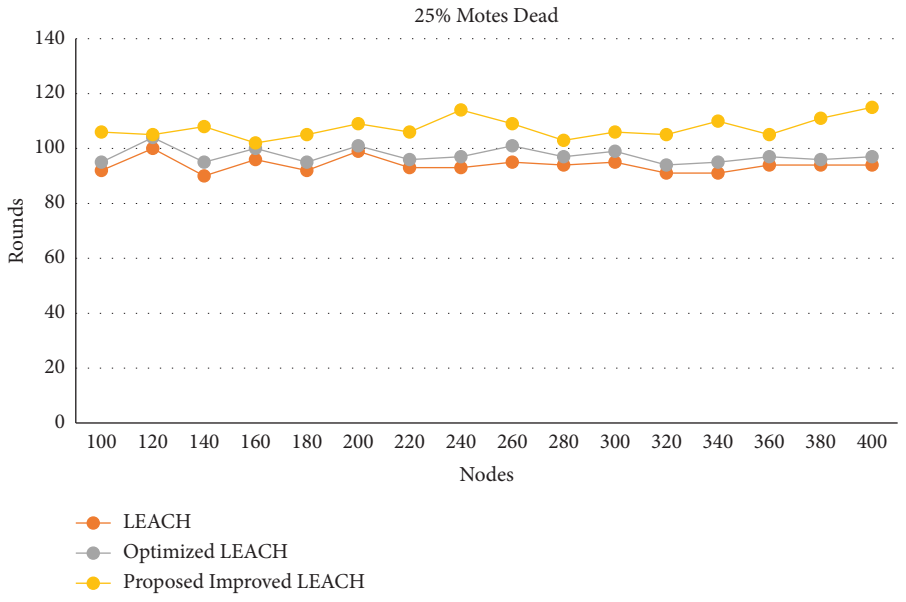
Table 2: Final results for 1st, 10th, and all motes.

Motes	Number of rounds								
	LEACH [8]			Optimized LEACH [17]			Proposed improved LEACH		
	1st mote dead	25% motes dead	All motes dead	1st mote dead	25% motes dead	All motes dead	1st mote dead	25% motes dead	All motes dead
100	70	92	222	74	95	232	86	106	241
120	73	100	222	78	104	233	81	105	243
140	70	90	220	75	95	233	80	108	244
160	73	96	224	79	100	234	79	102	243
180	69	92	224	75	95	234	79	105	242
200	65	99	223	71	101	233	83	109	244
220	68	93	224	72	96	233	73	106	244
240	67	93	222	71	97	232	76	114	242
260	72	95	223	76	101	235	73	109	243
280	67	94	224	72	97	235	69	103	244
300	63	95	222	68	99	234	75	106	243
320	67	91	223	72	94	233	71	105	243
340	67	91	224	71	95	235	81	110	244
360	64	94	225	69	97	234	66	105	245
380	62	94	225	67	96	234	79	111	244
400	62	94	224	67	97	235	80	115	245

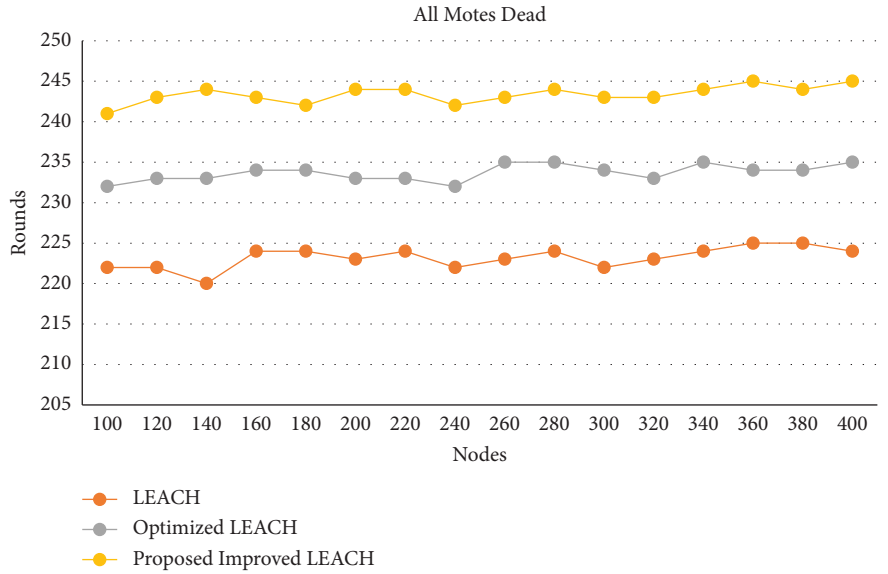


(a)

Figure 3: Continued.

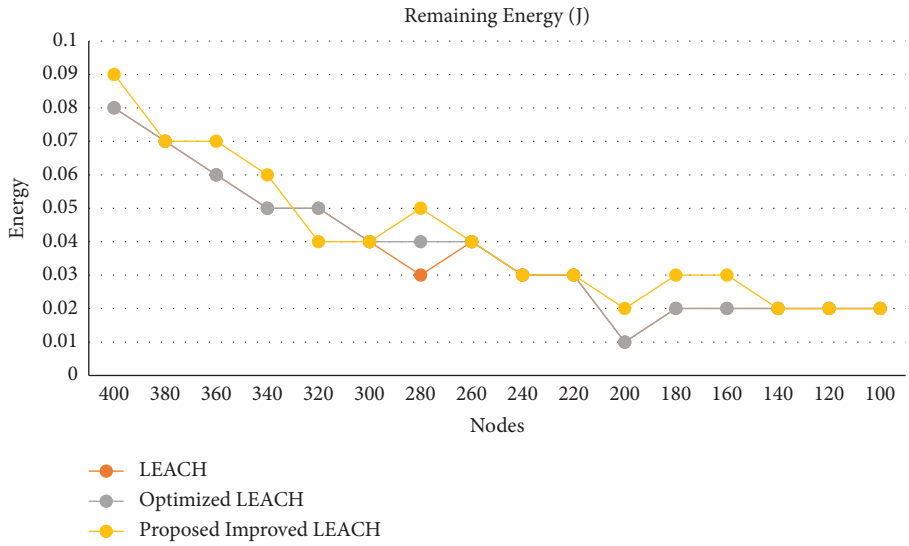


(b)

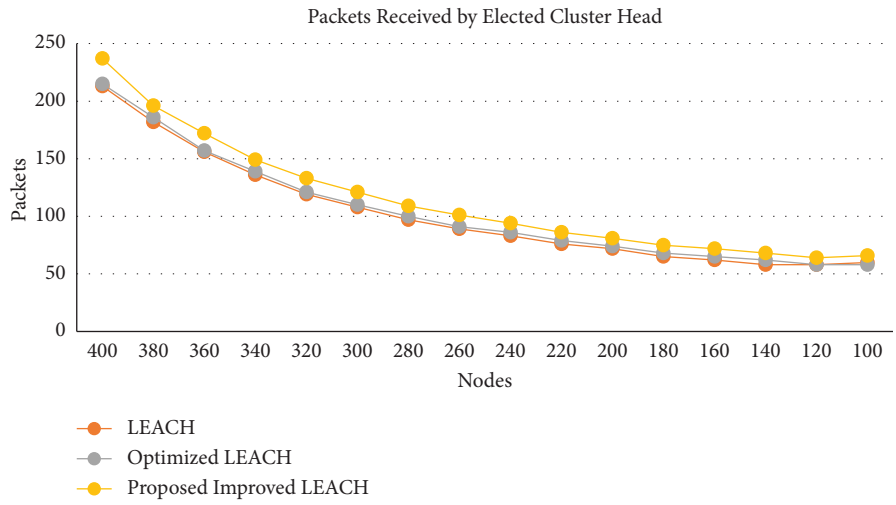


(c)

Figure 3: Continued.

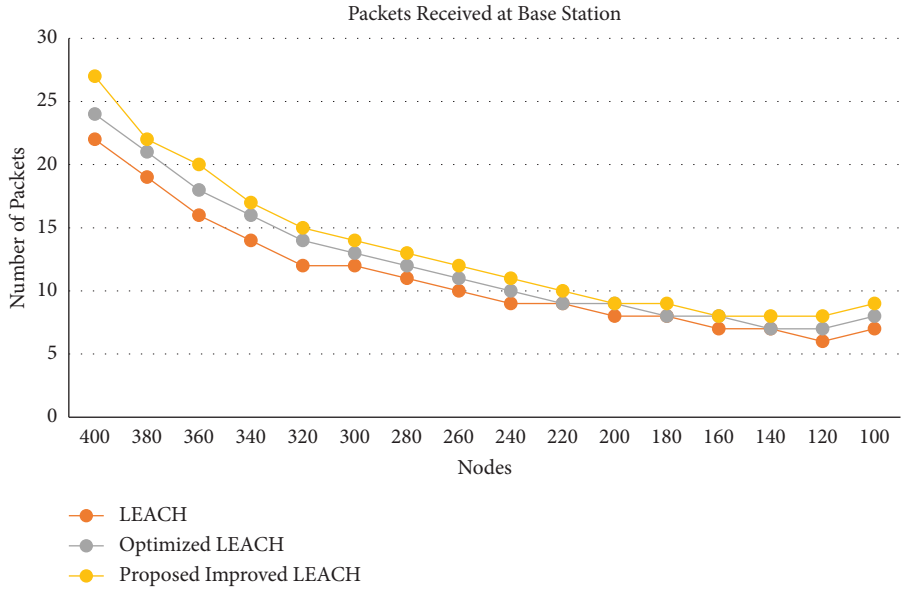


(d)



(e)

Figure 3: Continued.



(f)

Figure 3: Comparing LEACH, optimized and improved LEACH: (a) first node dead, (b) 25% nodes dead, (c) all nodes' dead, (d) residual energy, (e) packets received by the elected cluster head, and (f) packets received at the base station.

Table 3: Simulations results for energy remaining, data/communication packets received by the elected cluster head and finally packets received at the base station.

Motes	LEACH [8]			Optimized LEACH [16]			Proposed improved LEACH		
	Energy remained (J)	Packets received by elected CH	Packets received at the BS	Energy remained (J)	Packets received by elected CH	Packets received at the BS	Energy remained (J)	Packets received by elected CH	Packets received at the BS
100	0.08	213	22	0.08	215	24	0.09	237	27
120	0.07	182	19	0.07	186	21	0.07	196	22
140	0.06	156	16	0.06	157	18	0.07	172	20
160	0.05	136	14	0.05	139	16	0.06	149	17
180	0.05	119	12	0.05	121	14	0.04	133	15
200	0.04	108	12	0.04	110	13	0.04	121	14
220	0.03	97	11	0.04	100	12	0.05	109	13
240	0.04	89	10	0.04	91	11	0.04	101	12
260	0.03	83	9	0.03	86	10	0.03	94	11
280	0.03	76	9	0.03	79	9	0.03	86	10
300	0.01	72	8	0.01	74	9	0.02	81	9
320	0.02	65	8	0.02	68	8	0.03	75	9
340	0.02	62	7	0.02	65	8	0.03	72	8
360	0.02	58	7	0.02	62	7	0.02	68	8
380	0.02	58	6	0.02	58	7	0.02	64	8
400	0.02	60	7	0.02	58	8	0.02	66	9

(c) Recapitulate unless the system converges $z_y^+ = f_h(\sum_{x=0}^{N-1} T_{xy} z_x^-$, where f_h is hard restraint

3.3. Ant Colony Optimization (ACO). WSN performance is mainly measured via routing. Routing is a method of moving data from one place to another. While routing, two ideas are defined [17]. The first is optimum routing, where the quickest path is found using various methods, and the second is internetwork, where packets are transmitted.

3.4. Proposed PSO (Particle Swarm Optimization). Consider a situation where birds are randomly put in an area with only a piece of food to seek. The birds do not know where food is. They just know food distance. The easiest way to obtain food is to follow the bird to it. A particle is a bird. The particles have a fitness function and a velocity that can be computed. Every particle updates the "best" two values, pbest and gbest. Lesser particles update their locations and velocity using equations (2) and (3).

$$\text{velocity}[] = \text{veell}[] + c1 * \text{ran}() * (\text{perbest}[] - \text{present}[]) + c2 * \text{ran}() * (\text{glbest}[] - \text{present}[]), \quad (2)$$

$$\text{pre}[] = \text{per}[] + \text{vel}[]. \quad (3)$$

3.5. Nondominated Sorting Genetic Algorithm. In the last several years, EAs (evolutionary algorithms) *t* have evolved to solve mono-objective, multi-even-handed, and multi-target enhancement concerns in a precise order [18]. Despite explicit efforts to combine various mono-target developmental and nondevelopmental calculations, there are few studies that include all three types of enhancement concerns. The Pareto technique is used to segment the incoming population into subpopulations [19, 20]. Figure 2 shows the flowchart of hybrid ACO/PSO.

4. Simulations

Simulation is carried out using MATLAB for easy results computation. The limit limiting the framework's display has been disregarded. For reproduction, we dissect findings by several nodes and then by region.

Case 1: the proposed NN-LEACH is compared with existing RZ-LEACH and ACO-RZLEACH based on the node scalability by a varying number of motes and adjusting of the threshold values of various parameters such as alive nodes, dead nodes, the packet sent to BS, packet sent to CHs, and energy remained.

Case 2: in this case, the adaptability issue is considered for the organization's differing size for the consistent number of hubs ($n = 100$). The presentation of the present framework has been contrasted and existing one for the accompanying local sizes, for example, 50 m by 50 m, 100 m by 100 m, . . ., to 500 m by 500 m against the number of rounds.

5. Results Comparison of Leach and Optimized Leach

The proposed hybrid model is compared to the basic LEACH [8] and the optimized LEACH [17] with node scalability. Table 2 shows the simulation model's initial settings. Motes placed randomly in the area of interest. As demonstrated in Figure 3, assuming that all of the nodes have the same initial energy level (0.3 J), the energy optimization of the enhanced LEACH is compared to LEACH and optimized LEACH. Packets sent from the cluster head to the base station and residual energy are observed after round numbers 100, 120, 140, . . ., 400 (0 : 3 J), and we compare their leftover energy in Figure 3. Table 3 shows the simulations results for energy remaining, data/communication packets and finally packets.

6. Conclusion

As shown in the graphs of Figure 3, LEACH has better outcomes in terms of initial, tenth, and whole nodes dead for the specified population. Proposed LEACH also has higher residual energy and sends more packets to the base station and cluster head. All of these factors lower energy usage and so extend network lifetime. The number of packets transmitted from the cluster head to the base station determines performance. The WSN lifetime can be balanced extended by the NN-LEACH algorithm that has been presented. In three different network settings, experimental results demonstrate that NN-LEACH surpasses its competitors GAECH, GCA, EAERP, and LEACH. Most real-time applications can benefit from the performance gain when the base station is removed from the network. In the best-case scenario, where the base station is not linked to the network, NN-LEACH exhibits a significantly longer lifetime than its equivalent. Additionally, we discovered that LEACH optimization reduces energy use by distributing it evenly among clusters by 5%.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. Liu, C. Sun, and Y. Lai, "A data transmission approach based on ant colony optimization and threshold proxy re-encryption in WSNs," *Journal of Artificial Intelligence and Technology*, 2021.

[2] E. Dixit and V. Jindal, "IEESEP: an intelligent energy efficient stable election routing protocol in air pollution monitoring WSNs," *Neural Computing & Applications*, vol. 34, no. 13, pp. 10989–11013, 2022.

[3] W. Kim, M. M. Umar, S. Khan, and M. A. Khan, "Novel scoring for energy-efficient routing in multi-sensored networks," *Sensors*, vol. 22, no. 4, p. 1673, 2022.

[4] M. El Barachi, A. Kadiwal, R. Glitho, F. Khendek, and R. Dssouli, "A presence-based architecture for the integration of the sensing capabilities of wireless sensor networks in the IP multimedia subsystem," in *Proceedings of the 2008 IEEE Wireless Communications and Networking Conference*, pp. 3116–3121, Las Vegas, NV, USA, March 2008.

[5] A. Yektaparast, F. Nabavi, and A. Sarmast, "An improvement on LEACH (Cell-LEACH)," in *Proceedings of the 14th International Conference on Advanced Communication Technology (ICACT)*, pp. 992–996, PyeongChang, Korea (South), February 2012.

[6] N. A. Pantazis, S. A. Nikolidakis, and D. D. Vergados, "Energy-efficient routing protocols in wireless sensor networks: a survey," *IEEE Commun. Surv. Tutorials*, vol. 15, no. 2, pp. 551–591, 2013.

[7] A. Choudhary, S. Kumar, and H. Sharma, "Study and analysis of hierarchical routing protocols in wireless sensor networks," *Applied Information Processing Systems: Proceedings of ICCET 2021*, pp. 461–474, Springer, Berlin, Germany, 2022.

[8] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless microsensor networks," in *Proceedings of the 33rd Annual*

Hawaii International Conference on System Sciences, vol. 1, p. 10, Maui, HI, USA, January 2000.

[9] M. Naghibi and H. Barati, "EGRPM: energy efficient geographic routing protocol based on mobile sink in wireless sensor networks," *Sustainable Computing: Informatics and Systems*, vol. 25, p. 100377, 2020.

[10] M. Ye, C. Li, G. Chen, and J. Wu, "EECS: an energy efficient clustering scheme in wireless sensor networks," in *Proceedings of the PCCC 2005. 24th IEEE International Performance, Computing, and Communications Conference*, pp. 535–540, Phoenix, AZ, USA, April 2005.

[11] C.-H. Lin and M.-J. Tsai, "A comment on 'HEED: a hybrid, energy-efficient, distributed clustering approach for ad hoc sensor networks," *IEEE Transactions on Mobile Computing*, vol. 5, no. 10, pp. 1471-1472, 2006.

[12] S. Lindsey and C. S. Raghavendra, "PEGASIS: power-efficient gathering in sensor information systems," in *Proceedings of the IEEE Aerospace Conference*, vol. 3, pp. 3–1125, Big Sky, MT, USA, March 2002.

[13] Y. Ge, S. Wang, and J. Ma, "Optimization on TEEN routing protocol in cognitive wireless sensor network," *EURASIP Journal on Wireless Communications and Networking*, vol. 2018, 2018.

[14] M. Soni, N. R. Nayak, A. Kalra, S. Degadwala, N. K. Singh, and S. Singh, "Energy efficient multi-tasking for edge computing using federated learning," *International Journal of Pervasive Computing and Communications*, 2022.

[15] J. Hong, J. Kook, S. Lee, D. Kwon, and S. Yi, "T-LEACH: the method of threshold-based cluster head replacement for wireless sensor networks," *Information Systems Frontiers*, vol. 11, no. 5, pp. 513–521, 2009.

[16] J. Bhola, M. Shabaz, G. Dhiman, S. Vimal, P. Subbulakshmi, and S. K. Soni, "Performance evaluation of multilayer clustering network using distributed energy efficient clustering with enhanced threshold protocol," *Wireless Personal Communications*, vol. 126, no. 3, pp. 2175–2189, 2021.

[17] A. Gupta and L. K. Awasthi, "Security issues in cross-organizational peer-to-peer applications and some solutions," in *Communications in Computer and Information Science*, pp. 422–433, Springer, Berlin, Germany, 2009.

[18] L. Sathish Kumar, S. Ahmad, S. Routray et al., "Modern energy optimization approach for efficient data communication in IoT-based wireless sensor networks," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 7901587, 13 pages, 2022.

[19] M. Shabaz and U. Garg, "Shabaz–urvashi link prediction (sulp): a novel approach to predict future friends in a social network," *Journal of Creative Communications*, vol. 16, no. 1, pp. 27–44, 2020.

[20] P. A. Deepshikha and Varsha, "Enhanced NN based RZ leach using hybrid ACO/PSO based routing for WSNs," in *Proceedings of the 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, pp. 1–7, Delhi, India, July 2017.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 8179642, 7 pages
https://doi.org/10.1155/2022/8179642

Research Article

Double-Connected Intuitionistic Space in Double Intuitionistic Topological Spaces

Asmaa Ghassob Raoof,[1] Taha H. Jassim,[1] and Ngiste Amare [2]

[1]*Department of Mathematics, College of Computer Sciences & Mathematics, Tikrit University, Tikrit, Iraq*
[2]*Department of Civil Engineering, Mizan Tepi University, Tepi, Ethiopia*

Correspondence should be addressed to Ngiste Amare; ngiste@mtu.edu.et

Received 11 June 2022; Accepted 28 June 2022; Published 13 December 2022

Academic Editor: Mukesh Soni

The structures of the families of fuzzy sets that arise out of various notions of openness and closeness in a double fuzzy topological space are explored. This research is to present a new portion of space (Double-connected intuitionistic space) in Double intuitionistic topological spaces. Through these concepts, we advance some of their characteristics and relate to themselves.

1. Introduction

Ajmal and Kohli [1] explain connectedness in fuzzy topological spaces. Yasry [2] investigated lectures in advanced topology. El-Hamed et al. [3] presented Double connected spaces. Kandil et al. [4] investigated on flour (intuitionistic) topological spaces. Kendal et al. [5] studied on flour (intuitionistic) compact space. The concept was used to define intuitionistic sets and the intuitionistic gradation of openness by Coker [6–8]. Atanassov and Stoeva [9] describe Intuitionistic fuzzy sets. Atanassov [10] studied more intuitionistic fuzzy sets. Ozcelik and Narli [11] introduced the concept of submaximal intuitionistic topological spaces. Tantawy et al. [12] researched soft connections with double spaces. Selma and Coker [13] examined the concept of connectedness in intuitionistic fuzzy special topological spaces. Levine [14, 15] introduced generalized closed sets in topology. The concept of a fuzzy set was presented by Zadeh in his classic paper from 1965 and 2002 [14, 16]. After that, we introduce a new class of sets in DITS, namely the Double connected set, the separated Double I sets, the strongly Double connected, Double CO connected I space, and the Double I component. We also presented several examples of each type and concluded that there are relationships between them that were presented through theorems.

2. Preliminaries

We remind the following definitions, which are needed in our effort:

Let $x \neq \varnothing$, as well as ζ and $\mathscr{L}$ be IS having the form $\zeta = \langle x.\zeta_1.\zeta_2\rangle.\mathscr{L} = \langle x.\mathscr{L}_1.\mathscr{L}_2\rangle$, respectively. Also, $\{\zeta_i: i \in I\}$ be an arbitrary family of I S in $\mathfrak{X}$, where $\zeta_i = \langle x.\zeta_i{}^{(1)}.\zeta_i{}^{(2)}\rangle$, afterward:

(1) $\widetilde{\varnothing} = \langle x.\varnothing.x\rangle$; $\widetilde{x} = \langle x.x.\varnothing\rangle$.

(2) $\zeta \subseteq \mathscr{L}$ iff $\zeta_1 \subseteq \mathscr{L}_1$ and $\zeta_2 \supseteq \mathscr{L}_2$.

(3) $\zeta^c = \langle x.\zeta_2.\zeta_1\rangle$.

(4) $\cup\zeta_i = \langle x.\cup\zeta_i^{(1)}.\cap\zeta_i^{(2)}\rangle.\cap\zeta_i = \langle x.\cap\zeta_i^{(1)}.\cup\zeta_i^{(2)}\rangle$ [7]. Let x be a nonempty set, an intuitionistic set $\mathfrak{B}$ (I S, for short) is an object having g the form $\mathfrak{B} = \langle x.\mathfrak{B}_1.\mathfrak{B}_2\rangle$, where $\mathfrak{B}_1$ and $\mathfrak{B}_2$ are disjoint subset of x. The set $\mathfrak{B}_1$ is called set of members of $\mathfrak{B}$, while $\mathfrak{B}_2$ is called set of nonmembers of B [7]. An intuitionistic topology (IT, for short) on a nonempty set x is a family T of IS in $\mathfrak{X}$ containing $\widetilde{\varnothing}.\widetilde{x}$ and closed under arbitrary unions and finitely intersections. The pair $(x.T)$ is called ITS [11]. Let $x \neq \varnothing$.

(1) A Double-set (D-set, for short) $\mathfrak{U}$ is an ordered pair $(\mathfrak{U}_1, \mathfrak{U}_2) \in \boldsymbol{p}(x) \times \boldsymbol{p}(x)$, such that $\mathfrak{U}_1 \subseteq \mathfrak{U}_2$.

(2) $D(x) = \{(\mathfrak{U}_1, \mathfrak{U}_2) \in \boldsymbol{p}(x) \times \boldsymbol{p}(x), \mathfrak{U}_1 \subseteq \mathfrak{U}_2\}$ is the family of all D-sets on x.
(3) The D-set $\tilde{x} = (x, x)$ is called the universal D-set, and the D-set $\tilde{\varnothing} = (\varnothing, \varnothing)$ is called the empty D-set.
(4) Let $\mathfrak{U} = (\mathfrak{U}_1, \mathfrak{U}_2)$; $\vartheta = (\vartheta_1; \vartheta_2) \in D(x)$:
(1) $(\mathfrak{U}^c) = (\mathfrak{U}_2^c, \mathfrak{U}_1^c)$, where $\mathfrak{U}^c$ is the complement of $\mathfrak{U}$.
(2) $\mathfrak{U} - \vartheta = (U_1 - \vartheta_2, \mathfrak{U}_2 - \vartheta_2)$ [4]. Let x be a nonempty set. The family η of D-sets in x is called a double topology on x if it satisfies the following axioms:
(a) $\varnothing, x \in \eta$.
(b) If $\mathfrak{U}, \vartheta \in \eta$, then $\mathfrak{U} \cap \vartheta \in \eta$,
(c) If $\{\mathfrak{U}z: z \in Z\} \subseteq \eta$, then $\cup_{z \in Z} \mathfrak{U}z \in \eta$. The pair (x, η) is called a DTS. Each element of η is called an open D-set in x. The complement of open D-set is called closed D-set [4]. Let x be a nonempty set:
(1) IN $(x) = \{\varnothing, x\}$ is a DTS, which is called indiscrete DTS.
(2) dis $(x) = \boldsymbol{p}(x)) \times \boldsymbol{p}(x)$ is a DTS, which is called discrete DTS [4]. Let $(\mathfrak{X}, \eta)$ be a DTS and $\partial \in D(\mathfrak{X})$. The double closure of $\mathfrak{U}$, denoted by cl η $(\mathfrak{U})$ defined by cl η $(\mathfrak{U}) = \cap \{\vartheta: \vartheta \in \eta_c$ and $\mathfrak{U} \subseteq \vartheta\}$ [2, 4]. Let $\mathscr{X}$ nonempty set, $w \in \mathscr{X}$ a fixed element in $\mathscr{X}$, and let $\mathscr{M} = \langle x.\mathscr{M}_1.\mathscr{M}_2\rangle$ be an intuitionistic set (IS, for short). The IS $\dot{\mathfrak{w}}$ defined by $\dot{\mathfrak{w}} = \langle x.\{\mathfrak{w}\}.\{\mathfrak{w}\}^c\rangle$ is called an intuitionistic point (I$\mathfrak{w}$ for short) in $\mathscr{X}$. The IS $\ddot{\mathfrak{w}} = \langle x.\varnothing.\{\mathfrak{w}\}^c\rangle$ is called a vanishing I point (VI$\mathfrak{w}$, for short) in $\mathscr{X}$. The IS $\dot{\mathfrak{w}}$ is said to be contained in $\mathscr{M}$ ($\dot{\mathfrak{w}} \in \mathscr{M}$ .for short) only if $\mathfrak{w} \in \mathscr{M}_1$, and similarly IS $\ddot{\mathfrak{w}}$ contained in $\mathscr{M}$ ($\ddot{\mathfrak{w}} \in \mathscr{M}$ for short) only if $w \notin \mathscr{M}_2$. For a given IS w in $\mathscr{X}$, we may write $\mathscr{M} = (\cup\{\mathfrak{w}: \dot{\mathfrak{w}} \in \mathscr{M}\}) \cup (\cup\{\ddot{\mathfrak{w}}: \ddot{\mathfrak{w}} \in \mathscr{M}\})$, and whenever $\mathscr{M}$ is not a proper IS (i.e., if $\mathscr{M}$ is not of the form $\mathscr{M} = \langle x.\mathscr{M}_1.\mathscr{M}_2\rangle$ where $\mathscr{M}_1 \cup \mathscr{M}_2 \neq \mathscr{X}$), then $\mathscr{M} = \cup\{\dot{\mathfrak{w}}: \dot{\mathfrak{w}} \in \mathscr{M}\}$ hold. In general, any IS $\mathscr{M}$ in $\mathscr{X}$ can be written in the form $\mathscr{M} = \dot{\mathscr{M}} \cup \ddot{\mathscr{M}}$ where $\dot{\mathscr{M}} = \cup\{\dot{\mathfrak{w}}: \dot{\mathfrak{w}} \in \mathscr{M}\}$, and $\ddot{\mathscr{M}} = \cup\{\ddot{\mathfrak{w}}: \ddot{\mathfrak{w}} \in \mathscr{M}\}$ [6]. A topological space $\mathscr{X}$ is connected if it cannot be written as $\mathscr{X} = \mathscr{X}_1 \cup \mathscr{X}_2$, where $\mathscr{X}_1$ and $\mathscr{X}_2$ are both open and $\mathscr{X}_1 \cap \mathscr{X}_2 = \varnothing$, otherwise $\mathscr{X}$ is called disconnected [2]. Let $(\mathscr{X}, \eta)$ be a DTS and Υ be a nonempty subset of $\mathscr{X}$. Then, $\eta_\Upsilon = \{q \cap \Upsilon: q \in \mathscr{M}$ and $\Upsilon = (\Upsilon, \Upsilon)\}$ is a double topology on Υ. The DTS $(\Upsilon, \eta_\Upsilon)$ is called a double topological subspace of $(\mathscr{X}, \eta)$ (DT-subspace, for short) [4]. Let $(\mathscr{X}, \eta)$ be a DTS and let $\mathfrak{g}, \mathfrak{h} \in D(\mathscr{X})$: $\mathfrak{g}, \mathfrak{h}$ are said to be separated double sets (separated D-sets, for short) if cl η $(\mathfrak{g}) \cap \mathfrak{h} = \varnothing$ and cl η $(\mathfrak{h}) \cap \mathfrak{g} = \varnothing$ [3]. Let $(\mathscr{X}, \eta)$ be a DTS, and let be a nonempty subset of $\mathscr{X}$. If there exist two nonempty separated D-sets $\mathfrak{g}, \mathfrak{h} \in D(\mathscr{X})$ such that $\mathfrak{g} \cup \mathfrak{h} = \mathfrak{N}$, then the D-sets $\mathfrak{g}$ and $\mathfrak{h}$ form a D-separation of $\mathfrak{N}$ and it is said to be double disconnected set (D-disconnected set, for short). Otherwise, $\mathfrak{N}$ is said to be double connected set (d-connected set, for short) [3]. The DTS $(\mathscr{X}, \Pi)$ is said to be (1) C_5- disconnected, if $(\mathscr{X}, \Pi)$ has a proper open and closed D-set in Π. (2) C_5-connected, if $(\mathscr{X}, \Pi)$ is not C_5-disconnected [13]. An intuitionistic fuzzy special topological space $(\mathscr{X}, \Pi)$ is said to be strongly connected if there exist nonempty IFSC, SS $\mathfrak{g}$, and $\mathfrak{h}$ in $\mathscr{X}$ such that $\mathfrak{h} \cap \mathfrak{g} = \varnothing$ [13]. Let $(\mathscr{X}, \Pi)$ be a DTS and $\mathfrak{y} \subseteq \mathscr{X}$ with $\Pi_h \in (\tilde{\mathfrak{y}})$. The double component of $\mathfrak{y}$ with respect to Π_h is the maximal of all D-connected subsets of $(\mathfrak{y}, \Pi)$ containing the D-point Π_h and denoted by $C(\tilde{\mathfrak{y}}, \Pi_h)$, i.e., $C(\mathfrak{y}, \Pi_h) = \cup\{\tilde{\mathfrak{Z}} \subseteq \tilde{\mathfrak{y}}: \Pi_h \in \tilde{\mathfrak{Z}}, \tilde{\mathfrak{Z}}$ is a D-connected set$\}$ [12].

3. Double-Connected Intuitionistic Space in DITS

In this section, we define new kinds of x is called Double connected I space, separated Double I sets, strongly Double connected, Double CO connected I space, and Double I component in Double intuitionistic topological spaces, and joined to other kinds of sets that are defined in this work.

We start this section by the following definitions:

Definition 1. Let x be a nonempty set.

(1) A Double intuitionistic set (Double I set, for short) is an ordered pair $(\mathscr{Q}) = (\langle x.\mathscr{Q}_1.\mathscr{Q}_2\rangle.\langle x.\mathscr{D}_1.\mathscr{D}_2\rangle) \in \mathbf{p}(x) \times \mathbf{p}(x)$ s.t $\mathscr{Q} \subseteq \mathscr{D}$.
(2) DI $(x) = \{(Q, D) \in p(x) \times p(x), Q \subseteq D\}$ is the family of all Double I sets on x.
(3) The Double I set $(\langle x.x.\varnothing\rangle, \langle x.x.\varnothing\rangle = (\tilde{x}, \tilde{x})$ is called the universal Double I set, and the Double I set $(\tilde{\varnothing}, \tilde{\varnothing}) = ((\langle x.\varnothing.x\rangle.\langle x.\varnothing.x\rangle)$ is called the empty Double I set.
(4) Let $(Q, D), (C, G) \in$ Double I (x): 1) $(Q, D)^c = (\mathscr{D}^c, \mathscr{Q}^c) = (\langle x.\mathscr{D}_1.\mathscr{D}_2\rangle^c.\langle x.\mathscr{Q}_1.\mathscr{Q}_2\rangle^c) = (\langle x.\mathscr{D}_2.\mathscr{D}_1\rangle.\langle x.\mathscr{Q}_2.\mathscr{Q}_1\rangle).2((\mathscr{Q}.\mathscr{D})/(\mathscr{Q}.\mathscr{D}))(\mathscr{C},) = ((\mathscr{Q}/\mathscr{G}), (\mathscr{D}/\mathscr{C}) = ((\langle x.\mathscr{Q}_1.\mathscr{Q}_2\rangle.\ \langle x.\mathscr{D}_1.\mathscr{D}_2\rangle)/\ (\langle x.\mathscr{C}_1.\mathscr{C}_2\rangle.\langle x.\mathscr{G}_1.\mathscr{G}_2\rangle)) = ((\langle x.\mathscr{Q}_1.\mathscr{Q}_2\rangle/\langle x.\mathscr{G}_1.\mathscr{G}_2\rangle), (\langle x.\mathscr{D}_1.\mathscr{D}_2\rangle.\langle x.\mathscr{C}_1.\mathscr{C}_2\rangle)$

Each element of ω is called a DIOS in x. The complement of DIOS is called DICS.

Now, we want to introduce the important theorem to construct the Double intuitionistic topological spaces.

Theorem 1. *Let $x \neq \varnothing$. The family T of all a Double intuitionistic open sets in x is Double intuitionistic topological spaces (DITS).*

Proof

(1) Let $(x.T)$ be intuitionistic topological spaces (ITS), then $\tilde{\varnothing} = \langle x.\varnothing.x\rangle$, $\tilde{x} = \langle x.x.\varnothing\rangle \in$ IT $\longrightarrow (\tilde{\varnothing}, \tilde{\varnothing}), (\tilde{x}, \tilde{x}) \in$ DITS
(2) Let $(\tau.\sigma), (\phi.\epsilon) \in$ DIT $\longrightarrow \tau, \sigma, \varphi, \varepsilon \in$ IT. Since IT is an intuitionistic topology; then, $\tau \cap \in$ IT and $\varphi \cap \varepsilon \in$ IT. Now, let $\mathscr{K} = (\tau.\sigma)$ and $\mathscr{W} = (\phi.\epsilon) \longrightarrow (\mathscr{K}.\mathscr{W}) = ((\tau, \phi), (\sigma.\epsilon)) \in$ DITS.

(3) Let $(\theta_{\mathfrak{r}}.\phi_{\mathfrak{r}})$ be a family of IS and $r \in R$ and $(\theta_{\mathfrak{r}}.\phi_{\mathfrak{r}}) \in$ DIT $\longrightarrow \theta_{\mathfrak{r}}.\phi_{\mathfrak{r}} \in$ ITS, since IT is intuitionistic topology, then $\cup_{\mathfrak{r}\in R}\Theta_{\mathfrak{r}} \in$ IT and $\cup_{\mathfrak{r}\in R}\phi_{\mathfrak{r}} \in$ IT. Thus, $\cup_{\mathfrak{r}\in R}(\Theta_{\mathfrak{r}}.\phi_{\mathfrak{r}}) \in$ DIT. Therefore, DITS is Double intuitionistic topological spaces. □

Definition 2. Let x nonempty set, $(\tilde{\mathfrak{b}}, \tilde{\mathfrak{b}}) \in xa$ fixed element in x, and let $(\mathcal{Q}.\mathcal{D}) = (\langle x.\mathcal{Q}_1.\mathcal{Q}_2\rangle.\langle x.\mathcal{D}_1.\mathcal{D}_2\rangle$ be an Double intuitionistic set (DIS).The DIS $(\tilde{\mathfrak{b}}, \tilde{\mathfrak{b}})$ defined by $(\tilde{\mathfrak{b}}, \tilde{\mathfrak{b}}) = (\langle x.\{\mathfrak{b}\}.\{\mathfrak{b}\}^c\rangle.\langle x.\{\mathfrak{b}\}.\{\mathfrak{b}\}^c\rangle)$ is called an Double intuitionistic point (DI$\mathfrak{b}$ for short) in $\mathcal{X}$. The DIS $(\tilde{\mathfrak{b}}.\tilde{\mathfrak{b}})$is said to be contained in $(\mathcal{Q}.\mathcal{D})$, if and only if $(\tilde{\mathfrak{b}}.\tilde{\mathfrak{b}}) \in (\mathcal{Q}_1.\mathcal{D}_1)$.

Definition 3. Let (x, ω) be a DITS, let $(\mathcal{L}, \digamma)$, $(\mathcal{I}, \Omega) \in$ DI (x); then, $(\mathcal{L}, \digamma)$, $(\mathcal{I}, \Omega)$ are said to be separated Double intuitionistic sets (separated Double I sets, for short) if cl $(\mathcal{L}, \digamma) \cap (\mathcal{I}, \Omega) = (\tilde{\varnothing}, \tilde{\varnothing})$; and $(\mathcal{L}, \digamma) \cap$ cl$(\mathcal{I}, \Omega) = (\tilde{\varnothing}, \tilde{\varnothing})$. Or (cl $(\mathcal{L}, \digamma) \cap (\mathcal{I}, \Omega)) \cup ((\mathcal{L}, \digamma) \cap$ cl $(\mathcal{I}, \Omega)) = (\tilde{\varnothing}, \tilde{\varnothing})$.

Proposition 1. *Let (x, ω) be a DITS and let $(\mathcal{L}, \digamma)$, $(\mathcal{I}, \Omega) \in DI\ (x)$, if $(\mathcal{L}, \digamma)$, $(\mathcal{I}, \Omega)$ are separated Double I sets, then $(\mathcal{L}, \digamma) \cap (\mathcal{I}, \Omega) = (\tilde{\varnothing}, \tilde{\varnothing})$.*

Proof. Let $(\mathcal{L}, \digamma)$, $(\mathcal{I}, \Omega)$ are separated Double I sets. By definition if cl $(\mathcal{L}, \digamma) \cap (\mathcal{I}, \Omega) = (\tilde{\varnothing}, \tilde{\varnothing})$ and $(\mathcal{L}, \digamma) \cap$ cl $(\mathcal{I}, \Omega) = (\tilde{\varnothing}, \tilde{\varnothing})$. But $(\mathcal{L}, \digamma) \subseteq$ cl $(\mathcal{L}, \digamma)$.So, $(\mathcal{L}, \digamma) \cap (\mathcal{I}, \Omega) = (\tilde{\varnothing}, \tilde{\varnothing})$.

The following example shows that the converse is not true: □

Example 1. Let $x = \{\gamma, \beta, \varepsilon, \Lambda\}$: $\omega = \{(\tilde{\varnothing}, \tilde{\varnothing}), (\tilde{\mathcal{X}}, \tilde{\mathcal{X}}), (x_1.x_2), (x_3.x_2), (x_4.x_2), (x_4.x_5), (x_4.x_3), (x_2.x_2)\}$, where $(x_1.x_2) = (\langle x.\{\varepsilon\}.\{\beta.\Lambda\}\rangle, \langle x.\{\gamma.\beta.\varepsilon\}.\{\Lambda\}\rangle)$, $(x_3.x_2) = (\langle x.\{\gamma.\beta\}.\{\varepsilon.\Lambda\}\rangle.\langle x.\{\gamma.\beta.\varepsilon\}.\{\Lambda\}\rangle$, $(x_4.x_2) = (\langle x.\varnothing.\{\beta.\varepsilon.\Lambda\}\rangle.\langle x.\{\gamma.\beta.\varepsilon\}.\{\Lambda\}\rangle)$. $(x_4.x_5) = (\langle x.\varnothing.\{\beta.\varepsilon.\Lambda\}\rangle.\ \langle x.\{\beta\}\ \{\varepsilon.\Lambda\}\rangle)$, $(x_4.x_3) = (\langle x.\varnothing.\{\beta.\varepsilon.\Lambda\}\rangle,\ \langle x.\{\gamma.\beta\}.\{\varepsilon.\Lambda\}\rangle)$ and $(x_2.x_2) = (\langle x.\{\gamma.\beta.\varepsilon\}.\{\Lambda\}\rangle,\ \langle x.\{\gamma.\beta.\varepsilon\}.\{\Lambda\}\rangle)$. $\omega^c = \{(\tilde{\varnothing}, \tilde{\varnothing}), (\tilde{\mathcal{X}}, \tilde{\mathcal{X}}), (x_2^c.x_1^c), (x_2^c.x_3^c), (x_2^c.x_4^c), (x_5^c.x_4^c), (x_3^c.x_4^c), (x_2^c.x_2^c)\}$, where$(x_2^c.x_1^c) = (\langle x.\{\Lambda\}.\{\gamma.\beta.\varepsilon\}\rangle, \langle x.\{\beta.\Lambda\}.\{\varepsilon\}\rangle)$, $(x_2^c.x_3^c) = (\langle x.\{\Lambda\}.\{\gamma.\beta.\varepsilon\}\rangle.\langle x.\{\varepsilon.\Lambda\}.\{\gamma.\beta\}\rangle)$. $(x_2^c.x_4^c) = (\langle x.\{\Lambda\}.\{\gamma.\beta.\varepsilon\}.\ \langle x.\{\beta.\varepsilon.\Lambda\}.\varnothing\rangle)$. $(x_5^c.x_4^c) = (\langle x.\{\varepsilon.\Lambda\}.\{\beta\}\rangle.\langle x.\{\beta.\varepsilon.\Lambda\}.\varnothing\rangle)$, $(x_3^c.x_4^c) = (\langle x.\{\varepsilon.\Lambda\}.\ \{\gamma.\beta\}\rangle, \langle x.\{\beta.\varepsilon.\Lambda\}.\varnothing\rangle)$ and $(x_2^c.x_2^c) = (\langle x.\{\Lambda\}.\{\gamma.\beta.\varepsilon\}\rangle, \langle x.\{\Lambda\}.\{\gamma.\beta.\varepsilon\}\rangle)$. Let $(x_6.x_5^c) = (\langle x.\{\Lambda\}.\{\gamma.\beta\}\rangle.\ \langle x.\{\varepsilon.\Lambda\}.\{\beta\}\rangle)$ and $(x_7.x_7) = (\langle x.\{\beta\}.\{\gamma.\varepsilon.\Lambda\}\rangle.\ \langle x.\{\beta\}.\{\gamma.\varepsilon.\Lambda\}\rangle)$*so cl* $(x_6.x_5^c) = (x_3^c.x_4^c)$. Hence $(x_3^c.x_4^c) \cap (x_7.x_7) = (\tilde{\varnothing}, \tilde{\varnothing})$, but cl $(x_6.x_5^c) \cap (x_7.x_7) \neq (\tilde{\varnothing}, \tilde{\varnothing})$. Therefore, $(x_6.x_5^c)$, $(x_7.x_7)$ is not separated Double I sets.

Proposition 2. *Any two Double I closed (Double I open) subset say (v, u), (s, l) of DITS, (x, ω) are separated Double I sets if and only if they are two disjoint.*

Proof. Since separated Double I sets are already two disjoint; then, we only need to show that the Double I sets (v, u), (s, l) are separated Double I sets, they are Double I closed and two disjoint, $(v, u) \cap (s, l) = (\tilde{\varnothing}, \tilde{\varnothing})$. Since cl $(v, u) = (v, u)$ and cl $(s, l) = (s, l) \longrightarrow$ cl $(v, u) \cap (s, l) =$ and $(v, u) \cap$ cl $(s, l) = (\tilde{\varnothing}, \tilde{\varnothing}) \longrightarrow (v, u)$, (s, l) are separated Double I sets.

Now, take (v, u), (s, l) are Double I open and two disjoint $\longrightarrow (v, u) \cap (s, l) = (\tilde{\varnothing}, \tilde{\varnothing})$. Since (v, u) and (s, l) are Double I open $\longrightarrow (v, u)^c$ and $(s, l)^c$ are Double I closed $\longrightarrow$ cl $(v, u)^c = (v, u)^c$, cl $(s, l)^c = (s, l)^c$ $[(v, u) \subseteq (s, l)^c$, $(s, l) \subseteq (v, u)^c] \longrightarrow$ cl $(v, u) \subseteq$ cl $(s, l)^c = (s, l)^c \longrightarrow$ cl $(v, u) \subseteq (s, l)^c$ and cl $(s, l) \subseteq$ cl $(v, u)^c = (v, u)^c \longrightarrow$ cl $(s, l) \subseteq (v, u)^c$. Hence, cl $(v, u) \cap (s, l) = (\tilde{\varnothing}, \tilde{\varnothing})$ and $(v, u) \cap$ cl $(s, l) = (\tilde{\varnothing}, \tilde{\varnothing})$. So, (v, u), (s, l) are separated Double I sets.

The following definition of Double-connected intuitionistic sets in DITS: □

Definition 4. Let (x, ω) be a DITS; let (k, Z) be a nonempty subset of $\mathcal{X}$. If there exist two nonempty separated Double I sets $(\mathcal{L}, \digamma)$, $(\mathcal{I}, \Omega) \in$ DI $(\mathcal{X})$ such that $(\mathcal{L}, \digamma) \cup (\mathcal{I}, \Omega) = (k, Z)$, then the Double I sets $(\mathcal{L}, \digamma)$ and $(\mathcal{I}, \Omega)$ form a Double separation of (k, Z) and it is said to be Double disconnected intuitionistic sets (Double disconnected I sets, for short) Otherwise, (k, Z) is said to be Double-connected intuitionistic sets (Double connected I sets, for short).

The following two examples one of them is Double connected I sets, and the other is not:

Example 2. Let $\mathcal{X} = \{\ell.\ q\}$; $\omega = \{(\tilde{\varnothing}, \tilde{\varnothing}), (\tilde{\mathcal{X}}, \tilde{\mathcal{X}}), (\tilde{\varnothing}.\xi_1), (\tilde{\varnothing}.\xi_2), (\tilde{\varnothing}.\xi_3), (\tilde{\varnothing}.\xi_3^c), (\tilde{\varnothing}.\xi_1^c), (\tilde{\varnothing}.\xi_2^c), (\tilde{\varnothing}.\xi_4), (\xi_1.\tilde{\mathcal{X}}), (\xi_3.\tilde{\mathcal{X}}), (\xi_2.\tilde{\mathcal{X}}), (\xi_3^c.\tilde{\mathcal{X}}), (\xi_1^c.\tilde{\mathcal{X}}), (\xi_2^c.\tilde{\mathcal{X}}), (\xi_4.\tilde{\mathcal{X}}), (\xi_3, \xi_1).(\xi_1, \xi_1).(\xi_3, \xi_3).(\xi_2, \xi_2).\ (\xi_3^c.\xi_3^c).(\xi_1^c.\xi_1^c).(\xi_2^c.\xi_2^c).\ (\xi_4, \xi_4).(\xi_3^c, \xi_2).(\xi_1^c, \xi_1).(\xi_1^c, \xi_2).(\xi_1^c.\ \xi_3^c).(\xi_1^c, \xi_4).(\xi_2^c, \xi_1).(\xi_2^c, \xi_3).\ (\xi_2^c, \xi_2).(\xi_2^c, \xi_4).(\xi_4, \xi_1).(\xi_4, \xi_2)\ .(\tilde{\varnothing}.\tilde{\mathcal{X}})\}$ where.$(\tilde{\varnothing}.\xi_1) = (\langle x.\varnothing.\mathcal{X}\rangle.\langle x.\{\ell\}.\varnothing\rangle)$. $(\tilde{\varnothing}.\xi_2) = (\langle x.\varnothing.\mathcal{X}\rangle.\langle x.\{q\}.\varnothing\rangle)$. $(\tilde{\varnothing}.\xi_3) = (\langle x.\varnothing.\mathcal{X}\rangle.\ \langle x.\{\ell\}.\{q\}\rangle)$. $(\tilde{\varnothing}.\xi_3^c) = (\langle x.\varnothing.\mathcal{X}\rangle.\ \langle x.\{q\}.\{\ell\}\rangle)$. $(\tilde{\varnothing}.\xi_1^c) = (\langle x.\varnothing.\mathcal{X}\rangle.\langle x.\varnothing.\{\ell\}\rangle)$. $(\tilde{\varnothing}.\xi_2^c) = (\langle x.\varnothing.\mathcal{X}\rangle.\ \langle x.\varnothing.\{q\}\rangle)$. $(\tilde{\varnothing}.\xi_4) = (\langle x.\varnothing.\mathcal{X}\rangle.\langle x.\varnothing.\ \varnothing\rangle)$. $(\xi_1.\tilde{\mathcal{X}}) = (\langle x.\{\ell\}.\varnothing\rangle.\langle x.\mathcal{X}.\varnothing\rangle)$. $(\xi_3.\tilde{\mathcal{X}}) = (\langle x.\{\ell\}.\ \{q\}\rangle.\langle x.\mathcal{X}.\varnothing\rangle)$, $(\xi_2.\tilde{\mathcal{X}}) = (\langle x.\{q\}.\varnothing\rangle.\langle x.\mathcal{X}.\varnothing\rangle)$, $(\xi_3^c.\tilde{\mathcal{X}}) = (\langle x.\{q\}.\{\ell\}\rangle.\ \langle x.\mathcal{X}.\varnothing\rangle)$. $(\xi_1^c.\tilde{\mathcal{X}}) = (\langle x.\varnothing.\{\ell\}\rangle.\langle x.\ \mathcal{X}.\varnothing\rangle)$, $(\xi_2^c.\tilde{\mathcal{X}}) = (\langle x.\varnothing.\{q\}\rangle.\langle x.\mathcal{X}.\varnothing\rangle)$, $(\xi_4.\tilde{\mathcal{X}}) = (\langle x.\varnothing.\varnothing\rangle.\langle x.\mathcal{X}.\varnothing\rangle)$. $(\xi_3,\ \xi_1) = (\langle x.\{\ell\}.\{q\}\rangle.\langle x.\{\ell\}.\varnothing\rangle)$. $(\xi_1, \xi_1) = (\langle x.\{\ell\}.\varnothing\rangle.\langle x.\ \{\ell\}\ .\varnothing\rangle)$, $(\xi_3, \xi_3) = (\langle x.\{\ell\}.\{q\}\rangle.\langle x.\{\ell\}.\{q\}\rangle)$, $(\xi_2, \xi_2) = (\langle x.\{q\}.\varnothing\rangle.\langle x.\{q\}.\varnothing\rangle)$. $(\xi_3^c.\xi_3^c) = (\langle x.\{q\}.\{\ell\}\rangle.\ \langle x.\{q\}.\{\ell\}\rangle)$,$(\xi_1^c.\xi_1^c) = (\langle x.\varnothing.\{\ell\}\rangle.\langle x.\varnothing.\{\ell\}\rangle)$, $(\xi_2^c.\xi_2^c) = (\langle x.\varnothing.\{q\}\rangle.\ \langle x.\varnothing.\{q\}\rangle)$, $(\xi_4, \xi_4) = (\langle x.\varnothing.\varnothing\rangle.\langle x.\varnothing.\varnothing\rangle)$. $(\xi_3^c, \xi_2) = (\langle x.\{q\}.\ \{\ell\}\rangle.\langle x.\{q\}.\varnothing\rangle)$. $(\xi_1^c, \xi_1) = (\langle x.\varnothing.\{\ell\}\rangle.\langle x.\{\ell\}.\varnothing\rangle)$. $(\xi_1^c,\ \xi_2) = (\langle x.\varnothing.\{\ell\}\rangle.\langle x.\{q\}.\varnothing\rangle)$, $(\xi_1^c.\xi_3^c) = (\langle x.\varnothing.\{\ell\}\rangle.\ \langle x.\{q\}.\{\ell\}\rangle)(\xi_1^c, \xi_4) = (\langle x.\varnothing.\{\ell\}\rangle.\langle x.\varnothing.\varnothing\rangle)$, $(\xi_2^c, \xi_1) = (\langle x.\varnothing.\{q\}\rangle.\langle x.\{\ell\}.\varnothing\rangle)$. $(\xi_2^c, \xi_3) = (\langle x.\varnothing.\{q\}\rangle.\langle x.\{\ell\}.\ \{q\}\rangle)$, $(\xi_2^c, \xi_2) = (\langle x.\varnothing.\{q\}\rangle.\langle x.\{q\}.\varnothing\rangle)$, $(\xi_2^c, \xi_4) = (\langle x.\varnothing.\{q\}\rangle.\ \langle x.\varnothing.\varnothing\rangle)$, $(\xi_4, \xi_1) = (\langle x.\varnothing.\varnothing\rangle.\langle x.\{\ell\}.\varnothing\rangle)$. $(\xi_4, \xi_2) = (\langle x.\varnothing.\varnothing\rangle.\langle x.\{q\}.\varnothing\rangle)$, and $(\tilde{\varnothing}.\tilde{\mathcal{X}}) = (\langle x.\varnothing.\mathcal{X}\rangle.\langle x.\mathcal{X}.\varnothing\rangle)$. Let $(\xi_3, \xi_3) =$ cl $(\xi_3, \xi_3) \cap (\tilde{\varnothing}.\xi_1^c) = (\tilde{\varnothing}, \tilde{\varnothing})$ and $(\xi_3, \xi_3) \cap$ *cl* $(\tilde{\varnothing}.\xi_1^c) = (\tilde{\varnothing}.\xi_2^c) \longrightarrow (\xi_3, \xi_3) \cap (\tilde{\varnothing}.\xi_1^c) = (\tilde{\varnothing}, \tilde{\varnothing})$, such that $(\xi_3, \xi_3) \cup (\tilde{\varnothing}.\xi_1^c) = (\xi_3, \xi_1)$. Therefore, $(\xi_3, \xi_3).(\tilde{\varnothing}.\xi_1^c)$ are not Double connected I sets.

Example 3. Let $\mathcal{X} = \{\mathfrak{e}, \mathfrak{d}, \mathfrak{g}\}$; $\omega = \{(\tilde{\varnothing}, \tilde{\varnothing}), (\tilde{\mathcal{X}}, \tilde{\mathcal{X}}), (\Theta_1.\Theta_2), (\Theta_2.\Theta_3), (\Theta_4.\Theta_5), (\tilde{\varnothing}.\Theta_4)\}$ where $(\Theta_1.\Theta_2) = (\langle x.\{\mathfrak{g}\}.\{\mathfrak{e}.\mathfrak{d}\}\rangle, \langle x.\{\mathfrak{e}.\mathfrak{g}\}.\{\mathfrak{d}\}\rangle)$, $(\Theta_2.\Theta_3) = (\langle x.\{\mathfrak{e}.\mathfrak{g}\}.\{\mathfrak{d}\}\rangle.\ \langle x.\{\mathfrak{e}.\mathfrak{g}\}.\varnothing\rangle)$. (

$\Theta_4.\Theta_5$) = ($\langle x.\{\varepsilon\}.\{\text{ɗ}.\text{ɤ}\}\rangle.\langle x.\{\varepsilon\}.\{\text{ɤ}\}\rangle$), and ($\widetilde{\varnothing}.\Theta_4$) = ($\langle x.\varnothing.\mathcal{X}\rangle.\langle x.\{\varepsilon\}.\{\text{ɗ}.\text{ɤ}\}\rangle$). $\omega^c = \{(\widetilde{\varnothing}, \widetilde{\varnothing}), (\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}}), (O_2^c.\Theta_1^c).(O_3^c.\Theta_2^c).(O_5^c.\Theta_4^c).(O_4^c.\widetilde{\mathcal{X}})\}$ where $(O_2^c.\Theta_1^c)$ = ($\langle x.\{\text{ɗ}\}.\{\varepsilon.\text{ɤ}\}\rangle, \langle x.\{\varepsilon.\text{ɗ}\}.\{\text{ɤ}\}\rangle$), $(O_3^c.\Theta_2^c) = (\langle x.\varnothing.\{\varepsilon.\text{ɤ}\}\rangle, \langle x.\{\text{ɗ}\}.\{\varepsilon.\text{ɤ}\}\rangle)$, $(O_5^c.\Theta_4^c) = (\langle x.\{\text{ɤ}\}.\{\varepsilon\}\rangle, \langle x.\{\text{ɗ}.\text{ɤ}\}.\{\varepsilon\}\rangle)$, and $(O_4^c.\widetilde{\mathcal{X}}) = (\langle x.\{\text{ɗ}.\text{ɤ}\}.\{\varepsilon\}\rangle. < x, \mathcal{X}, \varnothing >)$. Let $(\Theta_1.\Theta_2)$ and $(\Theta_2.\Theta_3)$. Then, $(O_4^c.\widetilde{\mathcal{X}}) = \text{cl}\ (\Theta_1.\Theta_2) \cap (\Theta_2.\Theta_3) = (\Theta_1.\Theta_3)$ and $(\Theta_1.\Theta_2) \cap cl\,(\Theta_2.\Theta_3) = (\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}}) \longrightarrow (\Theta_1.\Theta_2) \cap (\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}}) = (\Theta_1.\Theta_2)$ and $(\Theta_1.\Theta_2) \cup (\Theta_2.\Theta_3) = (\Theta_2.\Theta_3)$. Hence, $(\Theta_1\Theta_2).(\Theta_2.\Theta_3)$ are Double connected I sets.

Definition 5. Let (x, ω) be a DITS; If there exist two nonempty separated Double I sets $(\mathcal{L}, \mathfrak{F}), (\mathcal{I}, \Omega) \in \text{DI}\ (x)$ such that $(\mathcal{L}, \mathfrak{F}) \cup (\mathcal{I}, \Omega)) = (\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}})$; then, $(\mathcal{L}, \mathfrak{F})$ and $(\mathcal{I}, \Omega)$ are said to be Double I division for DITS (x, ω). (x, ω) is said to be Double disconnected intuitionistic space (Double disconnected I space, for short), if (x, ω) has a Double I division. Otherwise, (x, ω) is said to be Double-connected intuitionistic space (Double connected I space, for short).

Example 4. Let $x = \{40, 41, 42\}$; $\omega = \{(\widetilde{\varnothing}, \widetilde{\varnothing}), (\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}}), (\sigma_1.\sigma_1), (\sigma_2.\sigma_3), (\sigma_4.\sigma_5), (\sigma_1^c.\sigma_1^c)\}$ where $(\sigma_1.\sigma_1)$ = ($\langle x.\{40.42\}.\{41\}\rangle$, $\langle x.\{40.42\}.\{41\}\rangle$), $(\sigma_2.\sigma_3)$ = ($\langle x.\{41\}.\{42\}\rangle.\langle x.\{40.41\}.\varnothing\rangle$). $(\sigma_4.\sigma_5)$ = ($\langle x.\varnothing.\{41.42\}\rangle.\langle x.\{40\}.\{41\}\rangle$) and $(\sigma_1^c.\sigma_1^c)$ = ($\langle x.\{41\}.\{40.42\} > .\langle x.\{41\}.\{40.42\}\rangle$). $\phi^c = \{(\widetilde{\varnothing}, \widetilde{\varnothing}), (\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}}), (\sigma_1^c.\sigma_1^c).(\sigma_3^c.\sigma_2^c).(\sigma_5^c.\sigma_4^c).(\sigma_1.\sigma_1)\}$ where $(\sigma_1^c.\sigma_1^c)$ =($\langle x\{41\}\{4042\}\rangle, \langle x.\{41\}.\{40.42\}\rangle$), $(\sigma_3^c.\sigma_2^c)$ =($\langle x.\varnothing.\{40.41\}\rangle, \langle x.\{42\}.\{41\}\rangle$), $(\sigma_5^c.\sigma_4^c) = (\langle x.\{41\}.\{40\}\rangle, \langle x.\{41.42\}.\varnothing\rangle)$ and $(\sigma_1.\sigma_1) = (\langle x.\{40.42\}.\{41\}\rangle, \langle x.\{40.42\}.\{41\}\rangle)$. Let $(\sigma_1.\sigma_1) = \text{cl}\ (\sigma_4.\sigma_5) \cap (\sigma_4.\sigma_5) = (\sigma_4.\sigma_5)$ and $(\sigma_4.\sigma_5) \cap \text{cl}\,(\sigma_4.\sigma_5) = (\sigma_1.\sigma_1)$ and $(\sigma_4.\sigma_5) \cup (\sigma_4.\sigma_5) \neq (\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}})$. So, $(\sigma_4.\sigma_5)$ is Double connected I spaces.

Remark 1

(1) (x, IN) is Double connected I space always since the only Double I open sets are $(\widetilde{\varnothing}, \widetilde{\varnothing}), (\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}})$, and this I sets not make the I space is Double disconnected.

(2) $(\mathcal{X}, \text{dis})$ is Double disconnected I space if x contains more than two element, since there exist $(\mathcal{L}, \mathfrak{F}), \widetilde{\varnothing} \neq (\mathcal{L}, \mathfrak{F}) \subseteq x \longrightarrow (\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}}) = (\mathcal{L}, \mathfrak{F}) \cup (\mathcal{L}, \mathfrak{F})^c$, $[(\mathcal{L}, \mathfrak{F}), (\mathcal{L}, \mathfrak{F})^c] \in \text{dis}\ (\mathcal{L}, \mathfrak{F}) \cap (\mathcal{L}, \mathfrak{F})^c = (\widetilde{\varnothing}, \widetilde{\varnothing})$, $(\mathcal{L}, \mathfrak{F}) \neq (\widetilde{\varnothing}, \widetilde{\varnothing})$, and $(\mathcal{L}, \mathfrak{F})^c \neq (\widetilde{\varnothing}, \widetilde{\varnothing})$.

Theorem 2. *Let (x, ω) is Double connected I space, if f $(\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}})$ cannot be written as the union of two disjoint nonempty Double I closed sets.*

Proof. ($\Longrightarrow$) Suppose that (x, ω) is Double connected to prove that $(\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}}) \neq (\Omega, \Upsilon) \cup (\Upsilon, \Pi)$, (Ω, Υ) and (Υ, Π) are Double I closed set, cl $(\Omega, \Upsilon) \cap (\Upsilon, \Pi) = (\widetilde{\varnothing}, \widetilde{\varnothing})$ and $(\Omega, \Upsilon) \cap \text{cl}\ (\Upsilon, \Pi) = (\widetilde{\varnothing}, \widetilde{\varnothing})$. Let $(\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}}) = (\Omega, \Upsilon) \cup (\Upsilon, \Pi)$, (Ω, Υ) and (Υ, Π) are Double I closed set, cl $(\Omega, \Upsilon) \cap (\Upsilon, \Pi) = (\widetilde{\varnothing}, \widetilde{\varnothing})$ and $(\Omega, \Upsilon) \cap \text{cl}\ (\Upsilon, \Pi) = (\widetilde{\varnothing}, \widetilde{\varnothing}) \longrightarrow \text{cl}\ (\Omega, \Upsilon) \subseteq (\Upsilon, \Pi)^c = (\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}}) \backslash (\Upsilon, \Pi) \subseteq (\Upsilon, \Pi)$, but $(\Omega, \Upsilon) \subseteq \text{cl}\ (\Omega, \Upsilon)$, so that $(\Omega, \Upsilon) = (\Upsilon, \Pi)^c \wedge (\Upsilon, \Pi) = (\Omega, \Upsilon)^c \longrightarrow (\Omega, \Upsilon) \in \omega \wedge (\Upsilon, \Pi) \in \omega$ (since $(\Omega, \Upsilon) = (\Upsilon, \Pi)^c \wedge (\Upsilon, \Pi)$ is Double I closed set and $(\Upsilon, \Pi) = (\Omega, \Upsilon)^c \wedge (\Omega, \Upsilon)$ is Double I closed set) $\longrightarrow (\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}}) = (\Omega, \Upsilon) \cup (\Upsilon, \Pi)$ [$(\Omega, \Upsilon), (\Upsilon, \Pi) \in \omega$], cl $(\Omega, \Upsilon) \cap (\Upsilon, \Pi) = (\widetilde{\varnothing}, \widetilde{\varnothing})$ and $(\Omega, \Upsilon) \cap \text{cl}\ (\Upsilon, \Pi) = (\widetilde{\varnothing}, \widetilde{\varnothing}) \longrightarrow (x, \omega)$ is Double disconnected, which a contradiction (since (x, ω) is Double connected) $\longrightarrow (\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}}) \neq (\Omega, \Upsilon) \cup (\Upsilon, \Pi)$, (Ω, Υ) and (Υ, Π) are Double I closed set, cl $(\Omega, \Upsilon) \cap (\Upsilon, \Pi) = (\widetilde{\varnothing}, \widetilde{\varnothing})$ and $(\Omega, \Upsilon) \cap \text{cl}\ (\Upsilon, \Pi) = (\widetilde{\varnothing}, \widetilde{\varnothing})$. Hence, $(\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}})$ cannot be written as the union of two disjoint nonempty Double I closed sets ($\Longleftarrow$). Suppose that $(\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}}) \neq (\Omega, \Upsilon) \cup (\Upsilon, \Pi)$, (Ω, Υ), and (Υ, Π) are Double I closed set, cl $(\Omega, \Upsilon) \cap (\Upsilon, \Pi) = (\widetilde{\varnothing}, \widetilde{\varnothing})$ and $(\Omega, \Upsilon) \cap \text{cl}\ (\Upsilon, \Pi) = (\widetilde{\varnothing}, \widetilde{\varnothing})$ to prove that (x, ω) is Double connected I space. Let (x, ω) is Double disconnected I space $\longrightarrow (\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}}) = (\Omega, \Upsilon) \cup (\Upsilon, \Pi)$ [$(\Omega, \Upsilon), (\Upsilon, \Pi) \in \omega$], cl $(\Omega, \Upsilon) \cap (\Upsilon, \Pi) = (\widetilde{\varnothing}, \widetilde{\varnothing})$ and $(\Omega, \Upsilon) \cap \text{cl}\ (\Upsilon, \Pi) = (\widetilde{\varnothing}, \widetilde{\varnothing}) \longrightarrow (\Omega, \Upsilon) = (\Upsilon, \Pi)^c \wedge (\Upsilon, \Pi) = (\Omega, \Upsilon)^c \longrightarrow (\Omega, \Upsilon)$ and $(\Upsilon, \Pi)$are Double I closed set, which is a contradiction. Since the complement of every one of them is Double I open set and this contradiction with hypotheses. Therefore, (x, Ψ) is Double connected I space. □

Theorem 3. *Let (x, ω) be a DITS and let Ψ be a nonempty subset of x. Then, if (Ω, Υ) and (Υ, Π) are Double I closed sets in Ψ; then, (Ω, Υ) and (Υ, Π) are separated Double I sets in Ψ if and only if (Ω, Υ) and (Υ, Π) are separated Double I sets in x.*

Proof. cl $(\Omega, \Upsilon) \cap (\Upsilon, \Pi) = (\widetilde{\Psi}, \widetilde{\Psi}) \cap \text{cl}\ (\Omega, \Upsilon) \cap (\Upsilon, \Pi)$; $(\Upsilon, \Pi) \subseteq (\widetilde{\Psi}, \widetilde{\Psi}) = (\widetilde{\Psi}, \widetilde{\Psi}) \cap (\Upsilon, \Pi) \cap \text{cl}\ (\Omega, \Upsilon) = (\Upsilon, \Pi) \cap (\widetilde{\Psi}.\widetilde{\Psi}) \cap \text{cl}\ (\Omega, \Upsilon) = (\Upsilon, \Pi) \cap \text{cl}_{\omega\Psi}\ (\Omega, \Upsilon) = (\widetilde{\varnothing}, \widetilde{\varnothing})$. Similarly, we have cl $(\Upsilon, \Pi) \cap (\Omega, \Upsilon) = (\widetilde{\Psi}, \widetilde{\Psi}) \cap \text{cl}\ (\Upsilon, \Pi) \cap (\Omega, \Upsilon)$; $(\Omega, \Upsilon) \subseteq (\widetilde{\Psi}, \widetilde{\Psi}) = (\widetilde{\Psi}, \widetilde{\Psi}) \cap (\Omega, \Upsilon) \cap \text{cl}\ (\Upsilon, \Pi) = (\Omega, \Upsilon) \cap (\widetilde{\Psi}, \widetilde{\Psi}) \cap \text{cl}\ (\Upsilon, \Pi) = (\Omega, \Upsilon) \cap \text{cl}_{\omega\Psi}\ (\Upsilon, \Pi) = (\widetilde{\varnothing}, \widetilde{\varnothing})$.

Conversely, $\text{cl}_{\omega\Psi}\ (\Omega, \Upsilon) \cap (\Upsilon, \Pi) = (\widetilde{\Psi}, \widetilde{\Psi}) \cap \text{cl}\ (\Omega, \Upsilon) \cap (\Upsilon, \Pi) = (\widetilde{\Psi}, \widetilde{\Psi}) \cap [\text{cl}\,(\Omega, \Upsilon) \cap (\Upsilon, \Pi)] = (\widetilde{\Psi}, \widetilde{\Psi}) \cap (\widetilde{\varnothing}, \widetilde{\varnothing}) = (\widetilde{\varnothing}, \widetilde{\varnothing})$. Also, $\text{cl}_{\omega\Psi}\ (\Upsilon, \Pi) \cap (\Omega, \Upsilon) = (\widetilde{\Psi}, \widetilde{\Psi}) \cap \text{cl}\ (\Upsilon, \Pi) \cap (\Omega, \Upsilon) = (\widetilde{\Psi}, \widetilde{\Psi}) \cap [\text{cl}\,(\Upsilon, \Pi) \cap (\Omega, \Upsilon)] = (\widetilde{\Psi}, \widetilde{\Psi}) \cap (\widetilde{\varnothing}, \widetilde{\varnothing}) = (\widetilde{\varnothing}, \widetilde{\varnothing})$. □

Theorem 4. *(x, ω) is Double disconnected I space if f there exist a nonempty proper subset of x, which are both Double I open and Double I closed sets in x.*

Proof. Suppose that $(\mathfrak{l}, \mathfrak{f}), (\mathfrak{k}, \eta)$ are a nonempty proper subset of x, which are both Double I open and Double I closed sets to prove x is Double disconnected. Let $(\mathfrak{k}, \eta)^c = (\mathfrak{l}, \mathfrak{f})^c$; then, $(\mathfrak{k}, \eta) \neq (\widetilde{\varnothing}, \widetilde{\varnothing})$ and $(\mathfrak{l}, \mathfrak{f}) \neq (\widetilde{\varnothing}, \widetilde{\varnothing})$, Moreover, $(\mathfrak{l}, \mathfrak{f}) \cup (\mathfrak{k}, \eta) = (\widetilde{\mathcal{X}}, \widetilde{\mathcal{X}})$ and

$$(\mathfrak{l}, \mathfrak{f}) \cap (\mathfrak{k}, \eta) = (\widetilde{\varnothing}, \widetilde{\varnothing}). \tag{1}$$

Since $(\mathfrak{l}, \mathfrak{f})$ is Double I closed as well as Double I open, then $(\mathfrak{k}, \eta)$ is also Double I open and Double I closed subset of x, so $(\mathfrak{l}, \mathfrak{f}) = \text{cl}\ (\mathfrak{l}, \mathfrak{f})$ and $(\mathfrak{k}, \eta) = \text{cl}\ (\mathfrak{k}, \eta) \longrightarrow \text{cl}\,(\mathfrak{l}, \mathfrak{f}) \cap (\mathfrak{k}, \eta) = (\widetilde{\varnothing}, \widetilde{\varnothing})$ (from (1)) and $\text{cl}\ (\mathfrak{k}, \eta) \cap (\mathfrak{l}, \mathfrak{f}) = (\widetilde{\varnothing}, \widetilde{\varnothing})$. Hence, $\mathcal{X}$ is Double disconnected I spaces.

Conversely: suppose that x is Double disconnected, then there exist a nonempty subset $(\mathfrak{l}, \mathfrak{f}), (\mathfrak{k}, \eta)$ of x such that $(\mathfrak{l}, \mathfrak{f})$

$\cup$ $(\mathsf{k}, \eta) = (\tilde{\mathcal{X}}, \tilde{\mathcal{X}})$; $(\mathfrak{l}, \mathfrak{f}) \neq (\tilde{\varnothing}, \tilde{\varnothing})$, $(\mathsf{k}, \eta) \neq (\tilde{\varnothing}, \tilde{\varnothing})$ $\longrightarrow$ cl $(\mathfrak{l}, \mathfrak{f}) \cap (\mathsf{k}, \eta) = (\tilde{\varnothing}, \tilde{\varnothing})$ and $(\mathfrak{l}, \mathfrak{f}) \cap$ cl $(\mathsf{k}, \eta) = (\tilde{\varnothing}, \tilde{\varnothing})$.Since $(\mathfrak{l}, \mathfrak{f}) \subseteq$ cl $(\mathfrak{l}, \mathfrak{f})$ $\longrightarrow$ cl $(\mathfrak{l}, \mathfrak{f}) \cap (\mathsf{k}, \eta) = (\tilde{\varnothing}, \tilde{\varnothing})$ $\longrightarrow$ $(\mathfrak{l}, \mathfrak{f}) \cap (\mathsf{k}, \eta) = (\tilde{\varnothing}, \tilde{\varnothing})$. Hence, $(\mathfrak{l}, \mathfrak{f}) = (\mathsf{k}, \eta)^c$ (since $(\mathfrak{l}, \mathfrak{f}) \subseteq (\mathsf{k}, \eta)^c = ((\tilde{\mathcal{X}}, \tilde{\mathcal{X}}) \setminus (\mathsf{k}, \eta)) = ((\tilde{\mathcal{X}} \setminus \eta). (\tilde{\mathcal{X}} \setminus \mathsf{k})) = (\mathsf{k}, \eta)^c)$ and $(\mathsf{k}, \eta) \cup (\mathsf{k}, \eta)^c = (\tilde{\mathcal{X}}, \tilde{\mathcal{X}})$, $(\mathsf{k}, \eta) \subseteq ((\tilde{\mathcal{X}}, \tilde{\mathcal{X}}) \setminus (\mathfrak{l}, \mathfrak{f})) = ((\tilde{\mathcal{X}} \setminus \mathfrak{f}), (\tilde{\mathcal{X}} \setminus \mathfrak{l})) = (\mathfrak{l}, \mathfrak{f})^c$ $\longrightarrow$ $(\mathsf{k}, \eta) = (\mathfrak{l}, \mathfrak{f})^c$ is a proper subset of $\mathcal{X}$. Now $(\mathfrak{l}, \mathfrak{f}) \cup$ cl $(\mathsf{k}, \eta) = (\tilde{\mathcal{X}}, \tilde{\mathcal{X}})$. Also, $(\mathfrak{l}, \mathfrak{f}) \cap$ cl $(\mathsf{k}, \eta) = (\tilde{\varnothing}, \tilde{\varnothing})$ $\longrightarrow$ $(\mathsf{k}, \eta) =$ (cl $(\mathsf{k}, \eta))^c$. Similarly, cl $(\mathfrak{l}, \mathfrak{f}) \cap (\mathsf{k}, \eta) = (\tilde{\varnothing}, \tilde{\varnothing})$ $\longrightarrow$ $(\mathfrak{l}, \mathfrak{f}) =$ (cl $(\mathfrak{l}, \mathfrak{f}))^c$. □

Theorem 5. *Let $(\mathcal{X}, \omega)$ be a DITS and let $\mathcal{H}$ be a nonempty subset of $\mathcal{X}$ such that $(\mathcal{H}, \omega_{\mathcal{H}})$ is Double connected, if (Ω, Υ) and (Υ, Π) are separated Double I sets such that $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq (\Omega, \Upsilon) \cup (\Upsilon, \Pi)$, then $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq (\Omega, \Upsilon)$or $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq (\Upsilon, \Pi)$.*

Proof. Since $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq (\Omega, \Upsilon) \cup (\Upsilon, \Pi)$, we have cl $(\Omega, \Upsilon) \cap (\Upsilon, \Pi) = (\tilde{\varnothing}, \tilde{\varnothing})$, and $(\Omega, \Upsilon) \cap$ cl $(\Upsilon, \Pi) = (\tilde{\varnothing}, \tilde{\varnothing})$, then $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) = (\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap [(\Omega, \Upsilon) \cup (\Upsilon, \Pi)] = [(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Omega, \Upsilon)] \cup [(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Upsilon, \Pi)]$ ((by Theorem 4)) $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Omega, \Upsilon)$ and $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Upsilon, \Pi)$ are separated Double *I* sets of $\mathcal{H}$. Let $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Omega, \Upsilon)$ and $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Upsilon, \Pi)$ are nonempty $\longrightarrow$ $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Omega, \Upsilon) \neq (\tilde{\varnothing}, \tilde{\varnothing})$ and $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Upsilon, \Pi) \neq (\tilde{\varnothing}, \tilde{\varnothing})$ $\longrightarrow$ $[(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Omega, \Upsilon)] \cap$ cl$[(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Upsilon, \Pi)] = [(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Omega, \Upsilon)] \cap [$cl $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap$ cl $(\Upsilon, \Pi)] = [(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap$ cl $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}})] \cap [(\Omega, \Upsilon) \cap$ cl $(\Upsilon, \Pi)] = (\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap$ cl $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\tilde{\varnothing}, \tilde{\varnothing}) = (\tilde{\varnothing}, \tilde{\varnothing})$. Similarly, cl $[(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Omega, \Upsilon)] \cap [(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Upsilon, \Pi)] = (\tilde{\varnothing}, \tilde{\varnothing})$. *i.* e, $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Omega, \Upsilon)$ and $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Upsilon, \Pi)$ are separated Double *I* sets. So $(\mathcal{H}, \omega_{\mathcal{H}})$ is Double disconnected, which is a contradiction. Let $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Omega) = (\tilde{\varnothing}, \tilde{\varnothing})$ $\longrightarrow$ $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) = (\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Upsilon, \Pi)$ $\longrightarrow$ $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq (\Upsilon, \Pi)$. Let $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Upsilon, \Pi) = (\tilde{\varnothing}, \tilde{\varnothing})$ $\longrightarrow$ $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) = (\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Omega, \Upsilon)$ $\longrightarrow$ $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq (\Omega, \Upsilon)$. Therefore, $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq (\Omega, \Upsilon)$ or $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq (\Upsilon, \Pi)$. □

Theorem 6. *Let $(\mathcal{X}, \omega)$ be a DITS and let $\mathcal{H}$ be a nonempty subset of $\mathcal{X}$ such that $(\mathcal{H}, \omega_{\mathcal{H}})$ is Double connected and $\mathfrak{R}$ be a subset of $\mathcal{X}$, such that $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq (\tilde{\mathfrak{R}}, \tilde{\mathfrak{R}}) \subseteq cl(\tilde{\mathcal{H}}, \tilde{\mathcal{H}})$, then $(\mathfrak{R}, \omega_{\mathfrak{R}})$ is Double connected subspace of $(\mathcal{X}, \omega)$. In particular, $(cl\ (\tilde{\mathcal{H}}, \tilde{\mathcal{H}}), \omega_{cl}(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}))$ is Double connected subspace of $(\mathcal{X}, \omega)$, $\Upsilon)$ k, $\eta)$.*

Proof. Let $(\mathfrak{R}, \omega_{\mathfrak{R}})$is Double disconnected subspace of $(\mathcal{X}, \omega)$, then $(\tilde{\mathfrak{R}}, \tilde{\mathfrak{R}})$ has a Double separation $(\Omega, \Upsilon)$and (Υ, Π) such that cl $(\Omega, \Upsilon) \cap (\Upsilon, \Pi) = (\tilde{\varnothing}, \tilde{\varnothing})$ and $(\Omega, \Upsilon) \cap$ cl $(\Upsilon, \Pi) = (\tilde{\varnothing}, \tilde{\varnothing})$ and $(\Omega, \Upsilon) \cup (\Upsilon, \Pi) = (\tilde{\mathfrak{R}}, \tilde{\mathfrak{R}})$, we have $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq (\tilde{\mathfrak{R}}, \tilde{\mathfrak{R}})$ $\longrightarrow$ $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq (\Omega, \Upsilon) \cup (\Upsilon, \Pi)$ $\longrightarrow$ $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq (\Omega, \Upsilon)$ or $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq (\Upsilon, \Pi)$ (by Theorem 5). Let $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq (\Omega, \Upsilon)$ $\longrightarrow$ cl$(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \subseteq$cl (Ω, Υ) $\longrightarrow$ cl$(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Upsilon, \Pi) \subseteq$cl $(\Omega, \Upsilon) \cap (\Upsilon, \Pi)$ $\longrightarrow$ cl$(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Upsilon, \Pi) \subseteq (\tilde{\varnothing}, \tilde{\varnothing})$, but $(\tilde{\varnothing}, \tilde{\varnothing}) \subseteq$cl$(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \cap (\Upsilon, \Pi)$.i.e.,

$$\text{cl}(\tilde{\mathcal{H}}, \tilde{\mathcal{H}})\bullet \cap (\Upsilon, \Pi)(\tilde{\varnothing}, \tilde{\varnothing}). \tag{2}$$

Again $(\tilde{\mathfrak{R}}, \tilde{\mathfrak{R}}) \subseteq$ cl $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}})$ $\longrightarrow$ $(\Omega, \Upsilon) \cup (\Upsilon, \Pi) \subseteq$ cl $(\tilde{\mathcal{H}}, \tilde{\mathcal{H}})$ $\longrightarrow$ $(\Upsilon, \Pi) \subseteq$ cl.

$$(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) \longrightarrow (\Upsilon.\Pi) \cap \text{cl}(\tilde{\mathcal{H}}, \tilde{\mathcal{H}}) = (\Upsilon, \Pi). \tag{3}$$

From (2) and (3), we have $(\Upsilon, \Pi) = (\tilde{\varnothing}, \tilde{\varnothing})$, which is a contradiction. So, $(\mathfrak{R}, \omega_{\mathfrak{R}})$ is Double connected. Similarly, cl $(\tilde{\mathcal{H}} < i > < /i > \tilde{\mathcal{H}})$ is Double connected. □

Theorem 7. *Let $(\mathcal{X}, \omega)$ be a DITS, if $\left\{(\mathcal{Y}_\beta. \mathcal{V}_\beta). \omega_{(\mathcal{Y}_\beta. \mathcal{V}_\beta)}): \beta \in J\right\}$ is a family of nonempty Double connected subspace of $\mathcal{X}$, $\cap_{\beta \in J} (\mathcal{Y}_\beta. \mathcal{V}_\beta)$ is nonempty; then, $(\mathbb{Z}_1.\mathbb{Z}_2) = (\cup_{\beta \in J} (\mathcal{Y}_\beta. \mathcal{V}_\beta). \omega_{(\cup_{\beta \in J} (\mathcal{Y}_\beta. \mathcal{V}_\beta))}$ is Double connected subspace of $(\mathcal{X}, \omega)$.*

Proof. Let $\left\{(\mathcal{Y}_\beta. \mathcal{V}_\beta). \omega_{(\mathcal{Y}_\beta. \mathcal{V}_\beta)}): \beta \in J\right\}$ is Double connected subspace of $(\mathcal{X}, \omega)$, $\cap_{\beta \in J} (\mathcal{Y}_\beta. \mathcal{V}_\beta) \neq (\tilde{\varnothing}, \tilde{\varnothing})$ to show that $(\mathbb{Z}_1.\mathbb{Z}_2) = \cup_{\beta \in J} (\mathcal{Y}_\beta. \mathcal{V}_\beta)$ is Double connected, if possible suppose that $(\mathbb{Z}_1.\mathbb{Z}_2)$ is Double disconnected there exist two nonempty disjoint Double *I* open sets (π, μ), (ζ, ρ) such that $(\pi, \mu) \cap (\zeta, \rho) = (\tilde{\varnothing}, \tilde{\varnothing})$, $(\zeta, \rho) \cap (\mathbb{Z}_1.\mathbb{Z}_2) = (\tilde{\varnothing}, \tilde{\varnothing})$ $\longrightarrow$ $[(\pi, \mu) \cap (\mathbb{Z}_1.\mathbb{Z}_2)] \cap [(\zeta, \rho) \cap (\mathbb{Z}_1.\mathbb{Z}_2)] = (\tilde{\varnothing}, \tilde{\varnothing})$ and $[(\pi, \mu) \cap (\mathbb{Z}_1.\mathbb{Z}_2)] \cup [(\zeta, \rho) \cap (\mathbb{Z}_1.\mathbb{Z}_2)] = ((\mathbb{Z}_1.\mathbb{Z}_2) \longrightarrow = [(\pi, \mu) \cup (\zeta, \rho)] \cap (\mathbb{Z}_1.\mathbb{Z}_2) \longrightarrow (\mathbb{Z}_1.\mathbb{Z}_2) \subseteq [(\pi, \mu) \cup (\zeta, \rho)]$, which is a contradiction, now $\cap (\mathcal{Y}_\beta. \mathcal{V}_\beta) \neq$ .Choose a Double *I* point $(\tilde{\mathfrak{b}}. \tilde{\mathfrak{b}}) \in \cap_{\beta \in J} (\mathcal{Y}_\beta. \mathcal{V}_\beta)$ $\longrightarrow$ $(\tilde{\mathfrak{b}}. \tilde{\mathfrak{b}}) \in (\mathcal{Y}_\beta. \mathcal{V}_\beta)$ $\longrightarrow$ $(\tilde{\mathfrak{b}}. \tilde{\mathfrak{b}}) \in \cup_{\beta \in J} (\mathcal{Y}_\beta. \mathcal{V}_\beta)$ $\longrightarrow$ $(\tilde{\mathfrak{b}}. \tilde{\mathfrak{b}}) \in (\mathbb{Z}_1.\mathbb{Z}_2)$ $\longrightarrow$ $(\tilde{\mathfrak{b}}. \tilde{\mathfrak{b}}) \in [(\pi, \mu) \cup (\zeta, \rho)]$ $\longrightarrow$ $(\tilde{\mathfrak{b}}. \tilde{\mathfrak{b}}) \in (\pi, \mu)$ and $(\tilde{\mathfrak{b}}. \tilde{\mathfrak{b}}) \in (\zeta, \rho)$. Let $(\tilde{\mathfrak{b}}. \tilde{\mathfrak{b}}) \in (\pi, \mu)$ also $(\mathcal{Y}_\beta. \mathcal{V}_\beta) \subseteq \cup_{\beta \in J} (\mathcal{Y}_\beta. \mathcal{V}_\beta)$ $\longrightarrow$ $(\mathbb{Z}_1.\mathbb{Z}_2) \subseteq [(\pi, \mu) \cup (\zeta, \rho)]$ $\longrightarrow$ $(\mathcal{Y}_\beta. \mathcal{V}_\beta) \subseteq [(\pi, \mu) \cup (\zeta, \rho)]$ $\longrightarrow$ $(\mathcal{Y}_\beta. \mathcal{V}_\beta) \cap [(\pi, \mu) \cup (\zeta, \rho)] = (\mathcal{Y}_\beta. \mathcal{V}_\beta)$ $\longrightarrow$ $[(\mathcal{Y}_\beta. \mathcal{V}_\beta) \cap (\pi, \mu)] \cup [(\mathcal{Y}_\beta. \mathcal{V}_\beta) \cap (\zeta, \rho)] = (\mathcal{Y}_\beta. \mathcal{V}_\beta)$ $\longrightarrow$ $[(\mathcal{Y}_\beta. \mathcal{V}_\beta) \cap (\pi, \mu)] \cap [(\mathcal{Y}_\beta. \mathcal{V}_\beta) \cap (\zeta, \rho)] = (\tilde{\varnothing}, \tilde{\varnothing})$.Also, $(\mathcal{Y}_\beta. \mathcal{V}_\beta) \cap (\pi, \mu)$, $(\mathcal{Y}_\beta. \mathcal{V}_\beta) \cap (\zeta, \rho)$ are nonempty two disjoint Double *I* open set of $(\mathcal{Y}_\beta. \mathcal{V}_\beta)$ $\longrightarrow$ $(\mathcal{Y}_\beta. \mathcal{V}_\beta) \subseteq [(\pi, \mu) \cup (\zeta, \rho)]$, so $(\mathcal{Y}_\beta. \mathcal{V}_\beta)$ is Double connected either $(\mathcal{Y}_\beta. \mathcal{V}_\beta) \subseteq (\pi, \mu)$ or $(\mathcal{Y}_\beta. \mathcal{V}_\beta) \subseteq (\zeta, \rho)$ $\longrightarrow$ $(\tilde{\mathfrak{b}}. \tilde{\mathfrak{b}}) \in (\pi, \mu)$, $(\tilde{\mathfrak{b}}. \tilde{\mathfrak{b}}) \in (\mathcal{Y}_\beta. \mathcal{V}_\beta)$ $\longrightarrow$ $(\mathcal{Y}_\beta. \mathcal{V}_\beta) \subseteq (\pi, \mu)$ $\longrightarrow$ $\cup_{\beta \in J} (\mathcal{Y}_\beta. \mathcal{V}_\beta) \subseteq (\pi, \mu))$ $\longrightarrow$ $(\mathbb{Z}_1.\mathbb{Z}_2) \subseteq (\pi, \mu)$, $(\pi, \mu) \cap (\zeta, \rho) = (\tilde{\varnothing}, \tilde{\varnothing})$ $\longrightarrow$ $(\mathbb{Z}_1.\mathbb{Z}_2) \cap (\zeta, \rho) = (\tilde{\varnothing}, \tilde{\varnothing})$, which is a contradiction to the fact that $(\mathbb{Z}_1.\mathbb{Z}_2) \cap (\zeta, \rho) \neq (\tilde{\varnothing}, \tilde{\varnothing})$ and $(\tilde{\mathfrak{b}}. \tilde{\mathfrak{b}}) \in (\zeta, \rho)$, which a contradiction $\longrightarrow$ $(\mathbb{Z}_1.\mathbb{Z}_2)$ is Double connected subspace of $(\mathcal{X}, \omega)$. □

Theorem 8. *Let $(\mathcal{X}, \Psi\omega)$ be a DITS, then, if $(\mathcal{X}, \omega)$ is Double disconnected I space, then $(\mathcal{X}, (\delta_1.\delta_2)$ are disconnected I space.*

Proof. Suppose that $(\mathcal{X}, \omega)$ is Double disconnected, then there exist $[(\mathcal{G}_1.\mathcal{G}_2), (\mathcal{J}_1.\mathcal{J}_2)] \in \Psi$ such that $(\mathcal{G}_1.\mathcal{G}_2) \cap (\mathcal{J}_1.\mathcal{J}_2) = (\tilde{\varnothing}, \tilde{\varnothing})$ and $(\mathcal{G}_1.\mathcal{G}_2) \cup (\mathcal{J}_1.\mathcal{J}_2) = (\tilde{\mathcal{X}}, \tilde{\mathcal{X}})$. So, $\mathcal{G}_1 \cap \mathcal{J}_1 = \tilde{\varnothing}$, $\mathcal{G}_2 \cap \mathcal{J}_2 = \tilde{\varnothing}$ and $\mathcal{G}_1 \cup \mathcal{J}_1 = \tilde{\mathcal{X}}$, $\mathcal{G}_2 \cup \mathcal{J}_2 = \tilde{\mathcal{X}}$ .Hence, $(\mathcal{G}_1.\mathcal{G}_2) \cap (\mathcal{J}_1.\mathcal{J}_2) = (\tilde{\varnothing}, \tilde{\varnothing})$ and $(\mathcal{G}_1.\mathcal{G}_2) \cup (\mathcal{J}_1.\mathcal{J}_2) = (\tilde{\mathcal{X}}, \tilde{\mathcal{X}})$,$[(\mathcal{G}_1.\mathcal{G}_2), (\mathcal{J}_1.\mathcal{J}_2)] \in (\delta_1.\delta_2)$.Therefore, $(\mathcal{X}, \delta_1)$ and $(\mathcal{X}, \delta_2)$ are disconnected *I* space.

The following example shows that the converse is not true in general: □

Example 5. Let $\mathcal{X} = \{i.\mathcal{J}.\hbar\}$; $\delta_1 = \{\tilde{\varnothing}, \tilde{\mathcal{X}}, \mathcal{M}_1.\mathcal{M}_2\}$, where $\mathcal{M}_1 = \langle \text{x}.\{i.\mathcal{J}\}.\{\hbar\}\rangle$, $\mathcal{M}_2 = \langle \text{x}.\{\hbar\}.\{i.\mathcal{J}\}\rangle$, and $\delta_2 = \{\tilde{\varnothing} .\tilde{\mathcal{X}} . \mathcal{M}_3.\mathcal{M}_4\}$, where $\mathcal{M}_3 = \langle \text{x}.\{i.\hbar\}.\{\mathcal{J}\}\rangle$ and $\mathcal{M}_4 = \langle \text{x}.\{\mathcal{J}\}.\{i.\hbar\}\rangle$. Then, $(\mathcal{X}, \delta_1)$ and $(\mathcal{X}, \delta_2)$ are intuitionistic topological spaces and disconnected I space. Since $\omega = (\delta_1.\delta_2) = \{(\tilde{\varnothing}, \tilde{\varnothing}), (\tilde{\mathcal{X}}, \tilde{\mathcal{X}}), (\tilde{\varnothing} .\mathcal{M}_3), (\tilde{\varnothing} . \mathcal{M}_4), (\tilde{\varnothing} .\tilde{\mathcal{X}}),$

$(\mathscr{M}_1.\widetilde{\mathscr{X}})$, $(\mathscr{M}_2.\widetilde{\mathscr{X}})$, $(\mathscr{M}_2.\mathscr{M}_3)\}$, where $(\widetilde{\varnothing}.\mathscr{M}_3) = (\langle x.\varnothing.\mathscr{X}>.\langle x.\{i.\hbar\}.\{\jmath\}\rangle)$, $(\widetilde{\varnothing}.\mathscr{M}_4) = (\langle x.\varnothing.\mathscr{X}>.\langle x.\{\jmath\}.\{i.\hbar\}\rangle)$, $(\widetilde{\varnothing}.\widetilde{\mathscr{X}}) = (\langle x.\varnothing.\mathscr{X}>.<x,\mathscr{X},\varnothing>)$, $(\mathscr{M}_1.\widetilde{\mathscr{X}}) = (\langle x.\{i.\jmath\}.\{\hbar\}\rangle.<x,\mathscr{X},\varnothing>)$, $(\mathscr{M}_2.\widetilde{\mathscr{X}}) = (\langle x.\{\hbar\}.\{i.\jmath\}\rangle.<x,\mathscr{X},\varnothing>)$,and$(\mathscr{M}_2.\mathscr{M}_3) = (\langle x.\{\hbar\}.\{i.\jmath\}\rangle.\langle x.\{i.\hbar\}.\{\jmath\}\rangle$. $\omega^c = (\delta_2^c.\delta_1^c) = \{(\widetilde{\varnothing},\widetilde{\varnothing}),(\widetilde{\mathscr{X}},\widetilde{\mathscr{X}}),(\mathscr{M}_3^c.\widetilde{\mathscr{X}}).(\mathscr{M}_4^c.\widetilde{\mathscr{X}}).(\widetilde{\varnothing}.\widetilde{\mathscr{X}}).(\widetilde{\varnothing}.\mathscr{M}_1^c).(\widetilde{\varnothing}.\mathscr{M}_2^c).(\mathscr{M}_3^c.\mathscr{M}_2^c)\}$, where $(\mathscr{M}_3^c.\widetilde{\mathscr{X}}) = (\langle x.\{\jmath\}.\{i.\hbar\}\rangle.<x,\mathscr{X},\varnothing>)$, $(\mathscr{M}_4^c.\widetilde{\mathscr{X}}) = (\langle x.\{i.\hbar\}.\{\jmath\}\rangle,<x,\mathscr{X},\varnothing>)$, $(\widetilde{\varnothing}.\widetilde{\mathscr{X}}) = (\langle x.\varnothing.\mathscr{X}>.<x,\mathscr{X},\varnothing>)$, $(\widetilde{\varnothing}.\mathscr{M}_1^c) = (\langle x.\varnothing.\mathscr{X}>.\langle x.\{\hbar\}.\{i.\jmath\}\rangle)$, $(\widetilde{\varnothing}.\mathscr{M}_2^c) = \langle x.\varnothing.\mathscr{X}>.\langle x.\{i.\jmath\}.\{\hbar\}\rangle)$, and $(\mathscr{M}_3^c.\mathscr{M}_2^c) = (\langle x.\{\jmath\}.\{i.\hbar\}\rangle, \langle x.\{i.\jmath\}.\{\hbar\}\rangle)$. Therefore, (x, ω) is not Double disconnected I space.

Definition 6. Let (x, ω) be a DITS is said to be the following:

(1) Double CO disconnected, if (x, ω) has a proper Double I open and Double I closed sets in x.

(2) Double CO connected, if (x, ω) is not Double CO disconnected.

Definition 7. The DITS (x, ω) is said to be the following:

(1) Strongly Double connected I space, if there exist no nonempty Double I closed sets $[(\nu, \mu), (\eta, \epsilon)] \in x$. Such that $(\nu, \mu) \cap (\eta, \epsilon) = (\widetilde{\varnothing}\,\widetilde{\varnothing})$.

(2) Strongly Double disconnected I space, if (x, ω) is not strongly Double connected I space.

Proposition 3. *(x, ω) is strongly Double connected, if and only, if there exist no Double I open sets (ν, μ),(η, ϵ) in x such that $(\nu, \mu) \neq (\widetilde{\mathscr{X}}, \widetilde{\mathscr{X}}) \neq (\eta, \epsilon)$and $(\nu, \mu) \cup (\eta, \epsilon) = (\widetilde{\mathscr{X}}, \widetilde{\mathscr{X}})$.*

Proof. Let (ν, μ), (η, ϵ) be Double I open sets in x such that $(\nu, \mu) \neq (\widetilde{\mathscr{X}}, \widetilde{\mathscr{X}}) \neq (\eta, \epsilon)$. If we take $(k, t) = (\nu, \mu)^c$ and $(n, m) = (\eta, \epsilon)^c$, then (k, t) and (n, m) become Double I closed sets in x and $(k, t) \neq (\widetilde{\varnothing}, \widetilde{\varnothing}) \neq (n, m)$ and $(k, t) \cap (n, m) = (\widetilde{\varnothing}, \widetilde{\varnothing})$, which is a contradiction.

Conversely, it is obvious. □

Remark 2. Strongly Double connectedness does not imply Double CO connectedness, and Double CO connectedness does not imply Strongly Double connectedness.

Example 6. Let $x = \{\Omega.\tau.\mu\}$; $\omega = \{(\widetilde{\varnothing}, \widetilde{\varnothing}),(\widetilde{\mathscr{X}}, \widetilde{\mathscr{X}}),(\phi_1.\phi_2),(\phi_3.\phi_3), (\widetilde{\varnothing}.\phi_4),(\phi_1.\phi_1)\}$where $(\phi_1.\phi_2) = (\langle x.\{\Omega\}.\{\tau.\mu\}\rangle, \langle x.\{\Omega.\tau\}.\varnothing\rangle)$,$(\phi_3.\phi_3) = (\langle x.\{\tau.\mu\}.\{\Omega\}\rangle.\langle x.\{\tau.\mu\}.\{\Omega\}\rangle)$. $(\widetilde{\varnothing}.\phi_4) = (\langle x.\varnothing.\mathscr{X}>.\langle x.\{\tau\}.\{\Omega\}\rangle)$ and $(\phi_1.\phi_1) = (\langle x.\{\Omega\}.\{\tau.\mu\}\rangle.\langle x.\{\Omega\}.\{\tau.\mu\}\rangle)$. $\omega^c = \{(\widetilde{\varnothing}, \widetilde{\varnothing}), (\widetilde{\mathscr{X}}, \widetilde{\mathscr{X}}), (\phi_2^c.\phi_1^c).(\phi_1.\phi_1)(\phi_4^c.\widetilde{\mathscr{X}}), (\phi_3.\phi_3)\}$,where $(\phi_2^c.\phi_1^c) = (\langle x.\varnothing.\{\Omega.\tau\}\rangle, \langle x.\{\tau.\mu\}.\{\Omega\}\rangle)$, $(\phi_1.\phi_1) = (\langle x.\{\Omega\}.\{\tau.\mu\}\rangle.\langle x.\{\Omega\}.\{\tau.\mu\}\rangle).(\phi_4^c.\widetilde{\mathscr{X}}) = (\langle x.\{\Omega\}.\{\tau\}\rangle, <x,\mathscr{X},\varnothing>)$, and $(\phi_3.\phi_3) = (\langle x.\{\tau.\mu\}.\{\Omega\}\rangle.\langle x.\{\tau.\mu\}.\{\Omega\}\rangle)$. Therefore, (x, ω) is strongly Double connected, but is not Double CO connected, for there exist $(\phi_1.\phi_1)$, $(\phi_3.\phi_3)$ are both Double I open and Double I closed sets.

Example 7. Let $x = \{a.\mathfrak{b}.\mathcal{C}.Đ\}$; $\omega = \{(\widetilde{\varnothing}, \widetilde{\varnothing}), (\widetilde{\mathscr{X}}, \widetilde{\mathscr{X}}),(\curlyvee_1.\curlyvee_4), (\curlyvee_2.\curlyvee_3), (\curlyvee_3.\widetilde{\mathscr{X}}).(\curlyvee_2.\curlyvee_4), (\curlyvee_5.\curlyvee_1), (\curlyvee_4.\widetilde{\mathscr{X}}), (\curlyvee_5.\curlyvee_2), (\curlyvee_4.\curlyvee_4), (\curlyvee_5.\curlyvee_4),(\curlyvee_2.\widetilde{\mathscr{X}}), (\curlyvee_2.\curlyvee_2), (\curlyvee_5.\curlyvee_5), (\curlyvee_3.\curlyvee_3)\}$, where $(\curlyvee_1.\curlyvee_4) = (\langle x.\{\mathfrak{b}.\mathcal{C}\}.\{Đ\}\rangle.\langle x.\{a.\mathfrak{b}.\mathcal{C}\}.\varnothing\rangle)$, $(\curlyvee_2.\curlyvee_3) = (\langle x.\{a\}.\{\mathcal{C}\}\rangle.\langle x.\{a.Đ\}.\{\mathcal{C}\}\rangle)$, $(\curlyvee_3.\widetilde{\mathscr{X}}) = (\langle x.\{a.Đ\}.\{\mathcal{C}\}\rangle, <x,\mathscr{X},\varnothing>)$, $(\curlyvee_2.\curlyvee_4) = (\langle x.\{a\}.\{\mathcal{C}\}.\langle x.\{a.\mathfrak{b}.\mathcal{C}\}.\varnothing\rangle)$, $(\curlyvee_5.\curlyvee_1) = (\langle x.\varnothing.\{\mathcal{C}.Đ\}\rangle.\langle x.\{\mathfrak{b}.\mathcal{C}\}.\{Đ\}\rangle)$, $(\curlyvee_4.\widetilde{\mathscr{X}}) = (\langle x.\{a.\mathfrak{b}.\mathcal{C}\}.\varnothing\rangle.<x,\mathscr{X},\varnothing>)$, $(\curlyvee_5.\curlyvee_2) = (\langle x.\varnothing.\{\mathcal{C}.Đ\}\rangle.\langle x.\{a\}.\{\mathcal{C}\}\rangle)$, $(\curlyvee_4.\curlyvee_4) = (\langle x.\{a.\mathfrak{b}.\mathcal{C}\}.\varnothing\rangle.\langle x.\{a.\mathfrak{b}.\mathcal{C}\}.\varnothing\rangle)$, $(\curlyvee_5.\curlyvee_4) = (\langle x.\varnothing.\{\mathcal{C}.Đ\}\rangle.\langle x.\{a.\mathfrak{b}.\mathcal{C}\}.\varnothing\rangle), (\curlyvee_2.\widetilde{\mathscr{X}}) = (\langle x.\{a\}.\{\mathcal{C}\}\rangle.<x,\mathscr{X},\varnothing>)$, $(\curlyvee_2.\curlyvee_2) = (\langle x.\{a\}.\{\mathcal{C}\}\rangle.\langle x.\{a\}.\{\mathcal{C}\}\rangle)$, $(\curlyvee_5.\curlyvee_5) = (\langle x.\varnothing.\{\mathcal{C}.Đ\}\rangle.\langle x.\varnothing.\{\mathcal{C}.Đ\}\rangle)$, and $(\curlyvee_3.\curlyvee_3) = (\langle x.\{a.Đ\}.\{\mathcal{C}\}\rangle.\langle x.\{a.Đ\}.\{\mathcal{C}\}\rangle))$. $\omega^c = \{(\widetilde{\varnothing}, \widetilde{\varnothing}),(\widetilde{\mathscr{X}}, \widetilde{\mathscr{X}}), (\curlyvee_4^c.\curlyvee_1^c).(\curlyvee_3^c.\curlyvee_2^c).(\curlyvee_4^c.\curlyvee_2^c).(\curlyvee_1^c.\curlyvee_5^c).(\widetilde{\varnothing}.\curlyvee_4^c).(\curlyvee_2^c.\curlyvee_5^c).(\curlyvee_4^c.\curlyvee_4^c).(\curlyvee_4^c.\curlyvee_5^c).(\widetilde{\varnothing}.\curlyvee_2^c), (\curlyvee_2^c.\curlyvee_2^c).(\curlyvee_5^c.\curlyvee_5^c).(\curlyvee_3^c.\curlyvee_3^c).(\widetilde{\varnothing}.\curlyvee_3^c)\}$, where $(\curlyvee_4^c.\curlyvee_1^c) = (\langle x.\varnothing.\{a.\mathfrak{b}.\mathcal{C}\}\rangle.\langle x.\{Đ\}.\{\mathfrak{b}.\mathcal{C}\}\rangle)$, $(\curlyvee_3^c.\curlyvee_2^c) = (\langle x.\{\mathcal{C}\}.\{a.Đ\}\rangle.\langle x.\{\mathcal{C}\}.\{a\}\rangle)$, $(\curlyvee_4^c.\curlyvee_2^c) = (\langle x.\varnothing.\{a.\mathfrak{b}.\mathcal{C}\}\rangle.\langle x.\{\mathcal{C}\}.\{a\}\rangle)$, $(\curlyvee_1^c.\curlyvee_5^c) = \langle x.\{Đ\}.\{\mathfrak{b}.\mathcal{C}\}\rangle.\langle x.\{\mathcal{C}.Đ\}.\varnothing\rangle)$, $(\widetilde{\varnothing}.\curlyvee_4^c) = (\langle x.\varnothing.\mathscr{X}>.\langle x.\varnothing.\{a.\mathfrak{b}.\mathcal{C}\}\rangle)$, $(\curlyvee_2^c.\curlyvee_5^c) = (\langle x.\{\mathcal{C}\}.\{a\}\rangle.\langle x.\{\mathcal{C}.Đ\}.\varnothing\rangle)$, $(\curlyvee_4^c.\curlyvee_4^c) = (\langle x.\varnothing.\{a.\mathfrak{b}.\mathcal{C}\}\rangle.\langle x.\varnothing.\{a.\mathfrak{b}.\mathcal{C}\}\rangle$, $(\curlyvee_4^c.\curlyvee_5^c) = (\langle x.\varnothing.\{a.\mathfrak{b}.\mathcal{C}\}\rangle.(\langle x.\{\mathcal{C}.Đ\}\rangle, (\widetilde{\varnothing}.\curlyvee_2^c) = (\langle x.\varnothing.\mathscr{X}>.\langle x.\{\mathcal{C}\}.\{a\}\rangle)$, $(\curlyvee_2^c.\curlyvee_2^c) = (\langle x.\{\mathcal{C}\}.\{a\}\rangle.\langle x.\{\mathcal{C}\}.\{a\}\rangle)$, $(\curlyvee_5^c.\curlyvee_5^c) = (\langle x.\{\mathcal{C}.Đ\}.\varnothing\rangle.\langle x.\{\mathcal{C}.Đ\}.\varnothing\rangle)$, $(\curlyvee_3^c.\curlyvee_3^c) = (\langle x.\{\mathcal{C}\}.\{a.Đ\}\rangle.\langle x.\{\mathcal{C}\}.\{a.Đ\}\rangle)$, and $(\widetilde{\varnothing}.\curlyvee_3^c) = (\langle x.\varnothing.\mathscr{X}>.\langle x.\{\mathcal{C}\}.\{a.Đ\}\rangle)$. (x, ω) is Double CO connected, but it is not strongly Double connected, for there exist $[(\curlyvee_1.\curlyvee_4), (\curlyvee_3.\curlyvee_3)] \in \Psi$, and $(\curlyvee_1.\curlyvee_4) \cup (\curlyvee_3.\curlyvee_3) = (\widetilde{\mathscr{X}}, \widetilde{\mathscr{X}})$.

Definition 8. Let (x, ω) be a DITS and $(\bar{\mathrm{y}}, \bar{\mathrm{y}}) \subseteq (\widetilde{\mathscr{X}}, \widetilde{\mathscr{X}})$ with $(\widetilde{\mathfrak{b}}.\widetilde{\mathfrak{b}}) = (\langle x.\{\mathfrak{b}\}.\{\mathfrak{b}\}^c\rangle.\langle x.\{\mathfrak{b}\}.\{\mathfrak{b}\}^c\rangle) \in \mathrm{DI}\ \mathfrak{b}\ (y)$. The union of all Double connected subsets of $(\bar{\mathrm{y}}, \bar{\mathrm{y}})$ containing the Double I point $(\widetilde{\mathfrak{b}}.\widetilde{\mathfrak{b}})$ is called Double I component of y with respect to $(\widetilde{\mathfrak{b}}.\widetilde{\mathfrak{b}})$, denoted by $(\mathfrak{c}_1.\mathfrak{c}_2)\ [(\bar{\mathrm{y}}, \bar{\mathrm{y}}), (\widetilde{\mathfrak{b}}.\widetilde{\mathfrak{b}})]$. *i.e.*, $(\mathfrak{c}_1.\mathfrak{c}_2)\ [(\bar{\mathrm{y}}, \bar{\mathrm{y}}), (\widetilde{\mathfrak{b}}.\widetilde{\mathfrak{b}})] = \cup\ \{(\nu, \mu) \subseteq (\bar{\mathrm{y}}, \bar{\mathrm{y}}): (\widetilde{\mathfrak{b}}.\widetilde{\mathfrak{b}}) \in (\nu, \mu)$, and (ν, μ) is Double connected I set $\}$.

Example 8. Let $x = \{\mathfrak{E}.\mathfrak{p}.\mho\}$; $\omega = \{(\widetilde{\varnothing}, \widetilde{\varnothing}), (\widetilde{\mathscr{X}}, \widetilde{\mathscr{X}}), (\curlyvee_1.\curlyvee_1), (\curlyvee_1^c.\curlyvee_1^c)\}$, where $(\curlyvee_1.\curlyvee_1)(\langle x.\{\mathfrak{E}\}.\{\mathfrak{p}.\mho\}\rangle.\langle x.\{\mathfrak{E}\}.\{\mathfrak{p}.\mho\}\rangle)$, $(\curlyvee_1^c.\curlyvee_1^c) = (\langle x.\{\mathfrak{p}.\mho\}.\{\mathfrak{E}\}\rangle.\langle x.\{\mathfrak{p}.\mho\}.\{\mathfrak{E}\}\rangle)$. Let $(\curlyvee_1.\curlyvee_1) = (\widetilde{\mathfrak{E}}.\widetilde{\mathfrak{E}}) = (\langle x.\{\mathfrak{E}\}.\{\mathfrak{p}.\mho\}\rangle.\langle x.\{\mathfrak{E}\}.\{\mathfrak{p}.\mho\}\rangle) \subseteq x \longrightarrow (\curlyvee_1.\curlyvee_1)$ is Double connected. $(\curlyvee_2.\curlyvee_2) = (\widetilde{\mathfrak{p}}.\widetilde{\mathfrak{p}}) = (\langle x.\{\mathfrak{p}\}.\{\mathfrak{E}.\mho\}\rangle.\langle x.\{\mathfrak{p}\}.\{\mathfrak{E}.\mho\}\rangle) \subseteq x \longrightarrow (\curlyvee_2.\curlyvee_2)$ is do double connected. $(\curlyvee_3.\curlyvee_3) = (\widetilde{\mho}.\widetilde{\mho}) = ((\langle x.\{\mho\}.\{\mathfrak{E}.\mathfrak{p}\}\rangle.\langle x.\{\mho\}.\{\mathfrak{E}.\mathfrak{p}\}\rangle) \subseteq x \longrightarrow (\curlyvee_3.\curlyvee_3)$ is Double connected. $(\curlyvee_4.\curlyvee_4) = (\{\mathfrak{p}.\mho\}.\widetilde{\{\mathfrak{p}.\mho\}}) = (\langle x.\{\mathfrak{p}.\mho\}.\{\mathfrak{E}\}\rangle.\langle x.\{\mathfrak{p}.\mho\}.\{\mathfrak{E}\}\rangle)$, and subE; $x \longrightarrow (\curlyvee_4.\curlyvee_4)$ is Double connected. $(\curlyvee_5.\curlyvee_5) = (\{\mathfrak{E}.\mathfrak{p}\}.\widetilde{\{\mathfrak{E}.\mathfrak{p}\}}) = (\langle x.\{\mathfrak{E}.\mathfrak{p}\}.\{\mho\}\rangle.\langle x.\{\mathfrak{E}.\mathfrak{p}\}.\{\mho\}\rangle) \subseteq x \longrightarrow (\widetilde{\curlyvee_5.\curlyvee_5})$ is not Double connected. $(\curlyvee_6.\curlyvee_6) = (\{\mathfrak{E}.\mho\}.\widetilde{\{\mathfrak{E}.\mho\}}) = (\langle x.\{\mathfrak{E}.\mho\}.\{\mathfrak{p}\}\rangle.\langle x.\{\mathfrak{E}.\mho\}.\{\mathfrak{p}\}\rangle) \subseteq x \longrightarrow (\curlyvee_6.\curlyvee_6)$ is not Double connected.

Note that, $(\curlyvee_1.\curlyvee_1)$, $(\curlyvee_2.\curlyvee_2)$, $(\curlyvee_3.\curlyvee_3)$, and $(\curlyvee_4.\curlyvee_4)$ are Double connected to find the Double I component of any Double I point in x, so $(\curlyvee_1.\curlyvee_1) \subseteq (\curlyvee_1.\curlyvee_1)$ only. Hence, $(\curlyvee_1.\curlyvee_1)$ is Double I component.

$(\curlyvee_2.\curlyvee_2) \subseteq (\curlyvee_2.\curlyvee_2)$ and $(\curlyvee_4.\curlyvee_4)$, so $(\curlyvee_2.\curlyvee_2)$ is not Double I component $(\curlyvee_3.\curlyvee_3) \subseteq (\curlyvee_3.\curlyvee_3)$ and $(\curlyvee_4.\curlyvee_4)$, so $(\curlyvee_3.\curlyvee_3)$ is

not Double I component, $(\top_4.\top_4) \subseteq (\top_4.\top_4)$ only. Hence, $(\top_4.\top_4)$ is Double I component. Therefore, $(\top_1.\top_1)$, $(\top_4.\top_4)$ are Double I component $\longrightarrow$ $(\top_1.\top_1) \cup (\top_4.\top_4) = (\tilde{\mathcal{X}}, \tilde{\mathcal{X}})$.

Theorem 9. *Every Double I component of a DITS is Double I closed set.*

Proof. Let (x, ω)be a DITS and let $(\mathfrak{c}_1.\mathfrak{c}_2)$ $[(\tilde{\mathfrak{y}}, \tilde{\mathfrak{y}}), (\tilde{\mathfrak{b}}.\tilde{\mathfrak{b}})]$ be a Double I component of the DITS (x, ω) with respect to an arbitrary Double I point $(\tilde{\mathfrak{b}}.\tilde{\mathfrak{b}}) \in \mathrm{DI}\ \mathfrak{b}(x)$.Then, $(\mathfrak{c}_1.\mathfrak{c}_2)$ $[(\tilde{\mathfrak{y}}, \tilde{\mathfrak{y}}), (\tilde{\mathfrak{b}}.\tilde{\mathfrak{b}})]$ is Double connected subset of x (by Theorem 6 and Theorem 7), cl $[(\mathfrak{c}_1.\mathfrak{c}_2)$ $[(\tilde{\mathfrak{y}}, \tilde{\mathfrak{y}}), (\tilde{\mathfrak{b}}.\tilde{\mathfrak{b}})]]$ is Double connected subset of $(\tilde{\mathcal{X}}, \tilde{\mathcal{X}})$ containing $(\tilde{\mathfrak{b}}.\tilde{\mathfrak{b}})$, then cl $[(\mathfrak{c}_1.\mathfrak{c}_2)$ $[(\tilde{\mathfrak{y}}, \tilde{\mathfrak{y}}), (\tilde{\mathfrak{b}}.\tilde{\mathfrak{b}})]] \subseteq (\mathfrak{c}_1.\mathfrak{c}_2)$ $[(\tilde{\mathfrak{y}}, \tilde{\mathfrak{y}}), (\tilde{\mathfrak{b}}.\tilde{\mathfrak{b}})]$.But $(\mathfrak{c}_1.\mathfrak{c}_2)$ $[(\tilde{\mathfrak{y}}, \tilde{\mathfrak{y}}), (\tilde{\mathfrak{b}}.\tilde{\mathfrak{b}})] \subseteq$ cl $[(\mathfrak{c}_1.\mathfrak{c}_2)$ $[(\tilde{\mathfrak{y}}, \tilde{\mathfrak{y}}), (\tilde{\mathfrak{b}}.\tilde{\mathfrak{b}})]]$.Therefore, $(\mathfrak{c}_1.\mathfrak{c}_2)$ $[(\tilde{\mathfrak{y}}, \tilde{\mathfrak{y}}), (\tilde{\mathfrak{b}}.\tilde{\mathfrak{b}})] =$ cl $[(\mathfrak{c}_1.\mathfrak{c}_2)$ $[(\tilde{\mathfrak{y}}, \tilde{\mathfrak{y}}), (\tilde{\mathfrak{b}}.\tilde{\mathfrak{b}})]]$. This indicates the Double I component $(\mathfrak{c}_1.\mathfrak{c}_2)$ $[(\tilde{\mathfrak{y}}, \tilde{\mathfrak{y}}), (\tilde{\mathfrak{b}}.\tilde{\mathfrak{b}})]$ is Double I closed set. □

4. Conclusions

In this paper, we got the following next results:

(1) We have presented a new set of the following concepts: Double intuitionistic set (DIS) (resp., Double intuitionistic topological spaces (DITS)), Double I point, Double connected ΨI set, separated Double I sets, strongly Double connected, Double CO connected I space, and Double I component Ψ in DITS.

(2) Study the basic characteristics and qualities related to these types and the relationships between them; giving examples is incorrect.

Data Availability

The data underlying the results presented in the study are available within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] N. Ajmal and J. K. Kohli, "Connectedness in fuzzy topological spaces," *Fuzzy Sets and System*, vol. 31, pp. 369–388, 1989.

[2] A. Z. A. Yasry, *Lectures in Advanced Topology*, College of Science for Women, Baghdad, Iraq, 2013.

[3] O. A. El-Hamed, A. S. Kandil, S. A. Ali, and S. H. Shaliel, "Double connected spaces," *Journal of New Theory*, vol. 17, pp. 1–17, 2017.

[4] A. Kandil, O. A. E. Tantawy, and M. Wafaie, "On flou (intuitionistic) topological spaces," *Journal of Fuzzy Mathematics*, vol. 15, no. 2, pp. 1–23, 2007.

[5] A. Kendal, O. A. E. Tantawy, and M. Wafaie, "On flou (intuitionistic) compact space," *Journal of Fuzzy Mathematics*, vol. 17, no. 2, pp. 275–294, 2009.

[6] D. Coker, "An introduction to intuitionistic topological spaces," *Busefal*, vol. 81, pp. 51–56, 2000.

[7] D. Coker, "A note on Intuitionistic sets and Intuitionistic points," *Turkish Journal of Mathematics*, vol. 20, pp. 343–351, 1996.

[8] D. Coker, "An introduction to intuitionistic fuzzy topological spaces," *Fuzzy Sets and Systems*, vol. 88, no. 1, pp. 81–89, 1997.

[9] K. T. Atanassov and S. Stoeva, "Intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 20, no. 1, pp. 87–96, 1986.

[10] K. T. Atanassov, "More on intuitionistic fuzzy sets," *Fuzzy Sets and Systems*, vol. 33, no. 1, pp. 37–45, 1989.

[11] A. Z. Ozcelik and S. Narli, "On submaximality intuitionistic topological spaces," *International Journal of Mathematical and Computational Sciences*, vol. 1, no. 1, pp. 139–141, 2007.

[12] O. A. E. Tantawy, S. A. EI-Sheikh, and S. Hussien, "Soft connected of double spaces," *South Asian Journal of Mathametics*, vol. 6, no. 5, pp. 249–262, 2016.

[13] O. Selma and D. Coker, "On connectedness in intuitionistic fuzzy special topolgical spaces," *International Journal of Mathematics and Mathematical Sciences*, vol. 21, no. 1, pp. 33–40, 1998.

[14] T. K. Mondal and S. Samanta, "On intuitionistic gradation of openness," *Fuzzy Sets and Systems*, vol. 131, no. 3, pp. 323–336, 2002.

[15] N. Levine, "Generalized closed sets in topology," *Rendiconti del Circolo Matematico di Palermo*, vol. 19, no. 1, pp. 89–96, 1970.

[16] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338–353, 1965.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 8798208, 7 pages
https://doi.org/10.1155/2022/8798208

Research Article

An Evaluation Method of International Trade Growth Potential Based on Fuzzy Algorithm

Feifei Fan and Weiling Xu

Department of Management and Economics, Beihai Vocational College, Beihai, Guangxi, China

Correspondence should be addressed to Weiling Xu; xuweiling@bhzyxy.edu.cn

Received 6 June 2022; Revised 7 July 2022; Accepted 21 July 2022; Published 17 November 2022

Academic Editor: Mukesh Soni

The new development pattern is a strategic choice to reshape China's new advantages in international cooperation and competition. It is based on China's current development stage, changes in the environment, and other factors. Located deep within China's interior, Xinjiang is a crucial region for the development of the western region and the Belt and Road Initiative. Its agricultural resources are exported to numerous foreign nations. In the context of the new development pattern, this paper first examines the foreign trade issues in Xinjiang. Second, this paper analyzes the current state of Sino-Kazakhstani trade, in particular the current state of trade in agricultural products. In addition, the growth potential of Xinjiang's agricultural international trade is analyzed utilizing the gravity model and fuzzy algorithm. Finally, countermeasures for the mode of development of Xinjiang's international trade are proposed.

1. Introduction

Despite its remote location in the interior of China, the development of Xinjiang is essential to the success of both One Belt One Road (OBOR) and China's far western region. Its agricultural sector is thriving, and its products are exported to numerous nations. In response to the expansion of the separatist movement in Xinjiang, the United States' Central Asia Strategy, the Russian Eurasian Strategy, and Pan-Turkism and Pan-Islamism are converging in the Central Asia region [1, 2]. It has become more challenging for Xinjiang to develop its economy, especially its foreign trade, which must contend with a more complex domestic and international economic environment. This has made the growth of international trade particularly difficult. In the course of its development, Xinjiang has encountered problems such as an unbalanced industrial structure, a lack of market competitiveness, and severe ecological damage. These issues are a result of the province's unique geographical, resource, and policy advantages, which have enabled its rapid expansion of international trade. Improvements must be made to the manner in which the expansion of international trade is conducted [3–5].

China has gradually surpassed Russia to become Kazakhstan's largest trading partner since OBOR Initiative was proposed. In addition, China's trade volume with Kazakhstan has increased from less than $400 million when it first began to $20 billion in 2020, which is an increase of nearly 50 times. China has surpassed Kazakhstan as Kazakhstan's most important trading partner, both in terms of exports and imports. China has recently overtaken Kazakhstan [6]. The volume of goods that are exchanged between Kazakhstan and China represents approximately 16 percent of Kazakhstan's total trade volume. It is anticipated that the economic and trade relations between China and Kazakhstan will continue to expand strongly in the year 2020. According to data from Kazakhstan, it is anticipated that the volume of trade between the two nations will reach approximately 16 billion dollars in the year 2020. This represents a growth of 4 percent in comparison to the volume of trade in the previous year. For instance, Kazakhstan's exports to China totaled 9 billion dollars,

while the country's imports from China totaled 7 billion dollars [7].

OBOR and China's westward construction in Kazakhstan, which is the first overseas stop, both place a significant emphasis on the development of Kazakhstan's agricultural sector. In order to make significant headway in the construction of the Silk Road Economic Belt, it is essential to conduct research into the possibilities of commercial agricultural exchange between the Xinjiang region of China and Kazakhstan [8].

According to existing research, the trade in agricultural products between China and the OBOR countries can be divided into two main categories. A comprehensive analysis of the current state of trade and competition is the first step that must be taken [9]. Some academics believe that this initiative will result in a broadening and deepening of agricultural trade links between countries along the route [10]. In addition, it will lead to a wider space for cooperation and opportunities for countries along the route to grow their agricultural businesses. China and the countries that it trades with have a good complementarity in agricultural products; however, they face challenges such as inadequate infrastructure, low added value for exported agricultural products, and rising green trade barriers. Agricultural products complement each other well [11]. On the other hand, the empirical research conducted by others has led to the conclusion that the complementarity of China's agricultural trade with that of other countries along OBOR is less obvious than it actually is. Xinjiang, China, which is situated in the geographic center of OBOR, has attracted the attention of a number of academics who are curious about the current state of the agricultural products trade there as well as the competitive characteristics of the market. Some researchers have examined the competitiveness of China's and Kazakhstan's agricultural trade using the explicit comparative advantage index and the trade complementarity index [12]. These indices were used to measure the degree to which the two countries complement each other in trade. Fruits and vegetables, oilseeds and olive fruits, and textile fibers are all examples of areas in which the economies of the two countries are highly complementary to one another. According to the findings of a few researchers, the agricultural cooperation between China's Xinjiang and the countries of Central Asia has formed a strong complementarity, and there is a great deal of untapped potential for agricultural trade in this region.

In addition, a significant number of academic researchers have investigated the factors that may influence the success of bilateral trade between China and Kazakhstan. In research, the method of regression based on the gravitational model is the one most frequently employed. Using the gravity model, a number of academics have conducted an empirical analysis of the potential for agricultural product trade between China and OBOR countries [12, 13].

It is believed that factors such as distance, population size, economic size, and government policies in these countries influence agricultural trade between China and the countries along the trade route. There is a lower probability of economic exchange between countries that are geographically distant from one another exhibited a significant connection or link between the two. Using a model known as the stochastic frontier gravity model, some researchers estimate China's agricultural export efficiency and potential with Silk Road countries. Economic growth and population density are viewed as advantages for China's agricultural exports, while the distance from markets is viewed as a disadvantage and political stability is viewed as an advantage [14–17]. Political stability is also seen as a positive. Some of the factors that are preventing China from exporting its agricultural products include indexes, trade and related infrastructure quality indexes, as well as procedures for customs clearance, economic freedom, and free trade agreements. In order to investigate the ways in which China's OBOR Initiative policy variables affect agricultural trade with the five Central Asian countries, some academics have utilized a panel data model in their research. They believe that there will be an increase in the amount of agricultural goods that are traded between China and the United States as a direct result of the OBOR Initiative. The amount of agricultural value added the distance that separates the two countries, and whether or not the countries share a border will all have an impact on the agricultural product trade that occurs between them. A few studies apply the method of overall network analysis to investigate the factors that influence trade and to measure the trade index in order to estimate the amount of potential for trade. Some academics use a method called overall network analysis to better understand how countries along OBOR trade agricultural products. This method suggests that factors such as economic development, distance from one another, shared languages, and whether or not the countries involved have signed free trade agreements (FTAs) are significant [18].

A fuzzy comprehensive evaluation method is the name given to an all-encompassing evaluation strategy that is underpinned by fuzzy mathematics. This all-encompassing method of evaluation converts qualitative evaluation into quantitative evaluation by employing the membership degree theory of fuzzy mathematics [19, 20]. Another way to phrase this is to say that this method of evaluation uses fuzzy mathematics to make a general evaluation of things or objects that are restricted by many factors [21]. Due to the fact that this method produces clear outcomes and has a robust structure, it can be applied to the solution of a wide variety of nondeterministic issues in a manner that is both systematic and clear. The development potential of health tourism in forests was evaluated by some researchers using fuzzy algorithms, and they discovered that the majority of the development value in forest health tourism was comprised of environmental value, resource value, development and construction value, and other values related to the development and construction of the forest. It is not common practice to employ a fuzzy algorithm when estimating the potential expansion of international trade [22, 23].

Therefore, the focus of this paper will initially be on the obstacles that Xinjiang's international trade must overcome. In addition, it investigates the current state of trade between China and Kazakhstan, with a focus on agricultural goods

trade. In this study, a gravity model and fuzzy algorithm are employed to determine the extent to which Xinjiang's agricultural exports have the potential to increase in the future. The countermeasures for Xinjiang's mode of foreign trade development are presented as the last item on the agenda, but they are by no means the least important.

2. Trade Status

2.1. Deficiencies in Trade. Central Asia is the region with which Xinjiang has the most significant economic ties. Because of Xinjiang's unique geographical characteristics, the region's international trade can only develop in a westerly direction, which creates a space-related problem. In recent years, China's share of the Central Asian foreign trade market has been steadily declining as Japan, South Korea, Europe, and the United States have all sought a piece of the action in the region. China's share of the Central Asian international trade market is at an all-time low. Therefore, trade goods originating from Xinjiang must frequently undergo multiple transshipments. This not only increases transportation costs but also increases competition on the Central Asian foreign trade market, which is detrimental to Xinjiang's exports. Adaptations are required for development and progress to continue.

Despite the fact that Xinjiang contains a number of dry and even extremely dry regions, the landscape of the region is almost entirely desert. The Gobi Desert and other deserts cover a vast portion of the region, which receives low annual precipitation on average. Xinjiang's 82 counties (cities) are home to 53 distinct deserts, which both surround and divide the oasis. This makes surviving in the sparsely populated region difficult. More than forty percent of the land in Xinjiang consists of mountains, which have sparse vegetation and a high concentration of surface salt. These factors combine to make the region vulnerable to land salinization and produce an ecologically fragile environment. Xinjiang is home to a large number of coal mines, oil fields, and other types of energy resources; however, the mining techniques employed there are still quite harsh, which will inevitably harm the local ecosystem. In addition, because the local ecological environment is so fragile, it will be more challenging to restore it once it has been damaged. There is a substantial problem with transformation.

Entrepreneurs' innovative spirit is essential for reshaping the growth of international trade and enhancing the country's standing on the international stage. Due to the backward economic climate in Xinjiang, there is little room for technological innovation. This indicates that an increase in inputs of resources and production factors is the primary factor driving international trade. It would appear that the local government of Xinjiang is unable to provide adequate funding and support for the technological innovation of businesses, which would allow them to engage in competitive market behavior. In addition, it is challenging to find innovative talents in the field of international trade in Xinjiang, indicating that the region lacks the compound talents required for the expansion of international trade. Internally, there is a lack of motivation for technological innovation within the organization as a whole. To begin with, the majority of Xinjiang's international trade companies employ only basic processing techniques when shipping their wares to neighboring countries. This hinders the region's capacity for scientific and technological innovation. During the time of the industrial transfer, the majority of Xinjiang's industries were textiles, food processing, and other light industries. As a result, the capacity for technological innovation among Xinjiang's international trade enterprises was diminished.

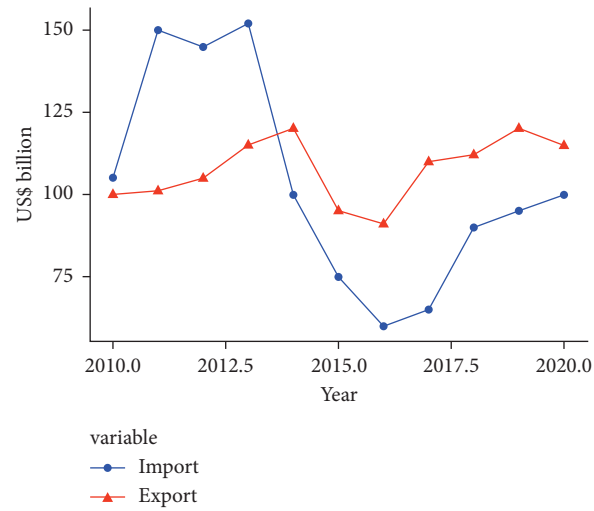


Figure 1: China-Kazakhstan import and export value in 2010–2020.

2.2. China-Kazakhstan Trade. The total volume of trade that took place between China and Kazakhstan from 2010 to 2020 is depicted in Figure 1. As a result of the ongoing trend toward globalization, the overall volume of imports and exports that occurs between the two countries has demonstrated an upward trend. Between the years 2012 and 2013, there was a sizeable rise in both China's exports to Kazakhstan and its imports from that country. During the years 2015 and 2016, the value of imports and exports experienced a significant decline, which was followed by an upward trend in the years that followed. Over the course of the past few years, the trade balance between China and Kazakhstan has shifted from one of a deficit to a surplus. Before 2013, China's trade balance was consistently negative throughout the country's history. After China reached a trade surplus in 2013, the country's economy continued to expand. The expansion of the One Belt One Road initiative is one of the most important factors.

Since the development of bilateral trade between China and Kazakhstan, the structure of trade products has remained relatively stable despite the fact that the trade goods used by both countries have changed and increased over time. Despite the trade imbalance between the two nations, this is the situation. According to the United Nations database, the Kazakh economy is highly dependent on Chinese imports. China was the recipient of 12 percent of Kazakhstan's total global exports in the year 2020. The bulk

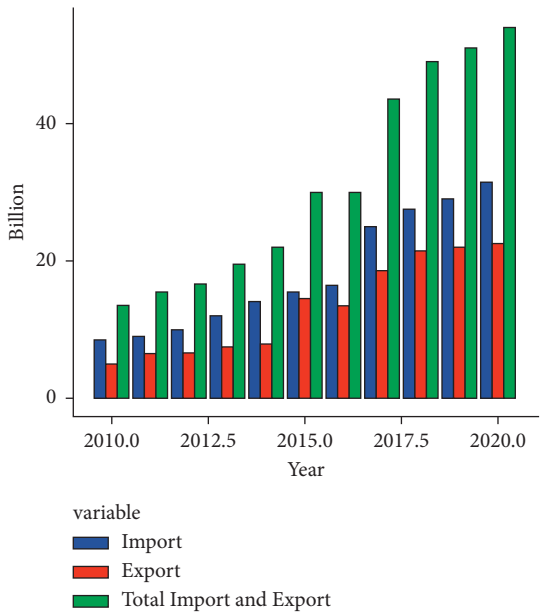


Figure 2: 2010–2020 China-Kazakhstan agricultural products trade situation.

of Kazakhstan's imports to China consists of mineral and metal products. The majority of China's exports to Kazakhstan, on the other hand, consist of textiles and mechanical and electrical goods.

Both China and Kazakhstan place a high priority on bilateral trade from the point of view of working together to develop policies governing bilateral trade. There have been 45 strategic agreements signed between the two nations ever since the establishment of the Shanghai Cooperation Organization. The Shanghai Cooperation Organization, which was initially established for political reasons but now serves primarily as a clearinghouse for financial and trade agreements, has signed twenty percent of all agreements. This makes it the organization responsible for one-fifth of all agreements. It is clear that the conditions are perfect for China and Kazakhstan's cooperation in their respective trade sectors at the bilateral level. As a direct consequence of this, the volume of trade between the two countries saw very little change as a direct result of the agreement.

In comparison to the ports and infrastructure of the other countries in the region, China and Kazakhstan have a greater number of ports, a greater variety of ports, and a more developed economy. There are currently seven ports between China and Kazakhstan, and this number was determined using a schedule that was approved by the State Council. Both Horgos and Alashankou are responsible for the majority of Xinjiang's cargo at the moment. When Alashankou was first established, the port's cargo volume was ahead of that of any other port in Greece; however, in 2014, Horgos surpassed it, and now Horgos is the largest port in Greece. From the perspective of cross-border and international cooperation, the joint production capacity cooperation that is part of the Silk Road Economic Belt has grown rapidly in recent years. The Khorgos Border Cooperation Center is the primary platform for cross-border cooperation between China and Kazakhstan. The center's primary focus is on promoting cross-border cooperation in the areas of e-commerce, logistics, and natural resources. Industrial parks, which are one of the future development trends of the economies of both China and Kazakhstan, are currently the most effective model for economic cooperation between the two countries. According to some incomplete data, there are currently operating in the world three industrial parks.

Agricultural goods will be traded between the Xinjiang region of China and Kazakhstan between 2010 and 2020, as shown in Figure 2. From 800 million yuan to approximately 3 billion yuan, the overall volume of import and export trade has increased by 2.98 times. This represents a 2.98-fold growth in the overall trade market. It also experienced a period of rapid expansion in the years that followed 2013. Throughout this time frame, China continued its process of opening up, and in 2016, it formalized a docking partnership with Kazakhstan. Recent years have witnessed a rise in the fervor of demand for trade in both directions. The amount of trade that occurs in Xinjiang, a key entry point to Kazakhstan, as well as the volume of imported and exported goods, has increased.

The volume of agricultural goods traded between China and Kazakhstan has increased consistently over the past several years. In recent years, China's importance as a trading partner for agricultural products exchanged between China and Kazakhstan has increased significantly. Since 2010, when it accounted for only 10 percent of Kazakhstan's total agricultural product trade volume, the volume of agricultural products traded between China and Kazakhstan has been increasing. Since then, this proportion has increased to approximately 25% of Kazakhstan's total agricultural product trade volume. This demonstrates how dependent Kazakhstan Stan is on the Chinese market, which in turn demonstrates that China is much more dependent on Kazakhstan's market than vice versa, despite China's agricultural product trade volume with Kazakhstan being significantly smaller.

3. Method

In order to evaluate the growth potential of China-Kazakhstan international trade, this chapter introduces the fuzzy evaluation method. The evaluation factor set and comment set are

$$\begin{aligned} U &= \{u_1, u_2, \cdots, u_m\}, \\ V &= \{v_1, v_2, \cdots, v_n\}, \end{aligned} \tag{1}$$

where u_m is the evaluation index, and v_i is the i-th evaluation result.

The weight distribution vector of each indicator u_m is

$$W = (w_1, w_2, \cdots, w_m), \tag{2}$$

where w_m is the weight corresponding to u_m.

Using the graded fuzzy subset, quantify the evaluated object for each factor u_m, determine the degree of

membership and then calculate the following fuzzy relationship matrix:

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}. \tag{3}$$

The calculation formula between $\mathbf{R}$, W and V is

$$V = W \cdot R = (w_1, w_2, \cdots, w_m) \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}. \tag{4}$$

In the vast majority of instances, membership degree is represented by a squishy vector. Compared to other types of evaluations, a fuzzy comprehensive evaluation can produce significantly more detailed and specific information. As a result, the evaluation criteria with the highest total scores are chosen for use in the final decision. It is difficult to obtain a satisfactory judgment matrix all at once due to the traditional AHP's imprecise evaluation of significant scales when determining the significance of indicators. This makes it more challenging to utilize. Due to this issue, a number of academics have developed a three-scale method that allows specialists to compare two factors without evaluating their relative significance. Consequently, it is significantly easier for specialists to compare relatively important factors. The accuracy and rationality of the evaluation are enhanced when a matrix is used to determine consistency. In this paper, we apply this methodology. The following is the expert scoring matrix:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \tag{5}$$
$$r_i = \sum_{j=1}^{n} a_{ij},$$

where α_{ij} is the importance scale value.

Next, we get the indirect matrix according to $\mathbf{A}$.

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix}, \tag{6}$$

$$b_{ij} = \begin{cases} \dfrac{r_i - r_j}{r_{\max} - r_{\min}} \dfrac{r_{\max} - r_{\min}}{r_{\min}} + 1, & r_i > r_j, \\ \left| \dfrac{r_i - r_j}{r_{\max} - r_{\min}} \dfrac{r_{\max} - r_{\min}}{r_{\min}} + 1 \right|^{-1}, & r_i < r_j. \end{cases}$$

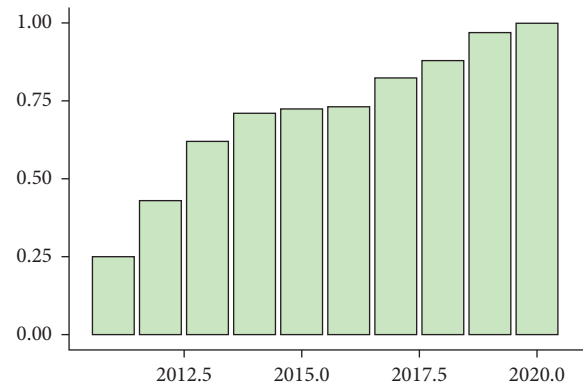


Figure 3: China-Kazakhstan bilateral trade level from 2010 to 2019.

Furthermore, according to the most transfer matrix calculation method in Reference [13], we obtain the quasioptimal transfer matrix of matrix B.

$$B' = \begin{bmatrix} b_{11}' & b_{12}' & \cdots & b_{1n}' \\ b_{21}' & b_{22}' & \cdots & b_{2n}' \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1}' & b_{n2}' & \cdots & b_{nn}' \end{bmatrix}, \tag{7}$$
$$l_p = (r - \hat{r})^2 + \lambda (y - \hat{y})^2,$$

Where a and b are parameters that control the weights, $\hat{y}$ is the true label, and $\hat{r}$ is the label of the rotation angle.

In terms of index selection, drawing on the gravity model and references 7, 12, 14, etc., this paper selects the trade volume of agricultural products (A), China-Kazakhstan nominal GDP (B), China-Kazakhstan spatial distance (C), China-Kazakhstan population size (D), and the same number of trade organizations (E).

4. Results

The trade gravity model can estimate the trade potential between the two parties, that is, the trade potential coefficient (Q) is obtained by comparing the actual value of the trade value with the simulated value. According to the usual classification of trade potential, $Q \geq 1.5$, $0.7 \leq Q \leq 1.5$, and $Q \leq 0.7$ belong to weak potential, medium potential, and huge potential, respectively.

To make it easier and more efficient for us to evaluate the level of bilateral trade that has occurred between China and Kazakhstan over the past decade, we will begin by calculating the level of bilateral trade between China and Kazakhstan from 2010 to 2019, as shown in Figure 3. The volume of trade between China and Kazakhstan will increase steadily from 2010 to 2020, as depicted in Figure 3. Among them is the fact that the level of bilateral trade entered a period of high cooperation in 2013 and subsequent years, which has continued to expand since then. The period from 2013 to 2016 was characterized by relatively slow development, whereas the period from 2017 to 2020 was characterized by relatively rapid development.

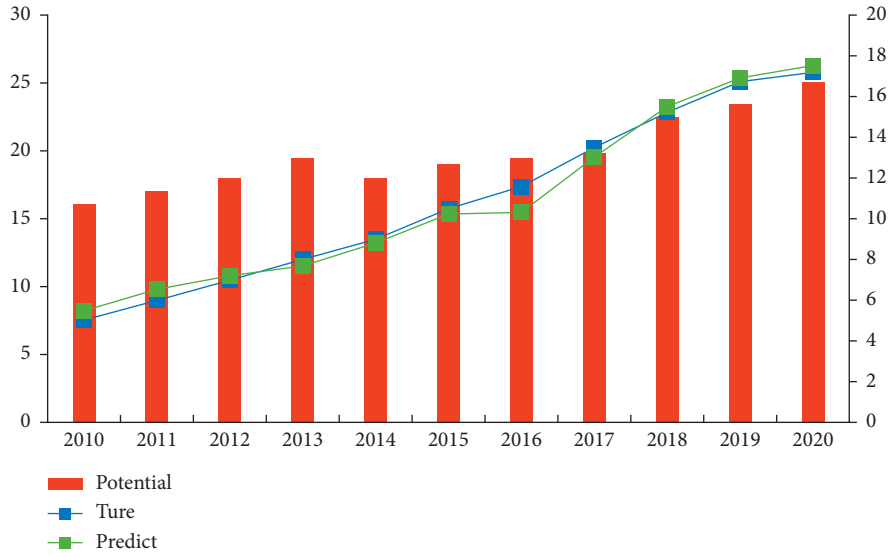


Figure 4: 2010–2020 potential calculation chart.

The fuzzy algorithm is employed to perform a simulation of the volume of bilaterally traded agricultural goods in order to determine the potential value depicted in Figure 4. The volume of trade between Xinjiang, China, and agricultural products from Kazakhstan is stable relative to their actual value, and the general trend is to increase gradually before increasing further. Between 2010 and 2013, the simulated value of the volume of bilateral trade increased gradually. The actual value of trade volume decreased slightly between 2013 and 2014, while the simulated value of bilateral agricultural trade increased slightly over the same period. It is possible that the problem lies in the model's variable selection. The condition of the global economy has a negligible impact on international trade patterns. Based on the simulated value of bilateral trade volume from 2014 to 2020, which indicates that there is ample room for growth in agricultural product trade between China's Xinjiang and Kazakhstan, there is ample room for growth in agricultural product trade between the two countries.

Using each of the three possible perspectives on the topic, this paper will assess the value of China and Kazakhstan's bilateral trade. Using each of the three possible perspectives, the purpose of this paper is to determine how much the trade between China and Kazakhstan is worth to both countries. Figure 5 illustrates the year-by-year evaluation value and the ranking for each year. Figure 5 depicts that the volume of trade between China and Kazakhstan is increasing at a rate comparable to the overall trend in bilateral trade. As trade between China and Kazakhstan continues to expand, it becomes more efficient. The nature and volume of trade between two nations have negative effects on economic growth. Overall, trade between China and Kazakhstan is growing, and it is reasonable to expect that bilateral trade and relations will continue to improve in the future. As the commercial relationship between China and Kazakhstan continues to develop, it is crucial that we

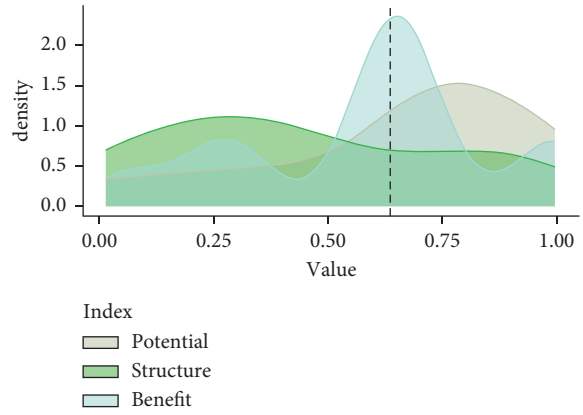


Figure 5: Comparison of China-Kazakhstan trade levels in 2020–2020 under different indicators.

place an emphasis on expanding the volume and value of trade, maximizing the benefits of trade, and streamlining trade structures.

5. Conclusion

Xinjiang is a crucial region for the development of the western region and the Belt and Road Initiative. It exports its agricultural resources to numerous foreign countries. This paper examines the foreign trade issues in Xinjiang within the context of China's new development pattern. This paper analyzes the current state of Sino-Kazakhstani trade, specifically the current state of agricultural product trade. Using the gravity model and fuzzy algorithm, the growth potential of Xinjiang's agricultural international trade is also analyzed. Finally, countermeasures are proposed for the mode of development of international trade in Xinjiang. In the future, we will not only be limited to Xinjiang trade, but will instead predict the trend of trade between China and other

countries in the world, and we will study the application of deep learning in cross-border trade.

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.

References

[1] R. Huang, T. Nie, Y. Zhu, and S Du, "Forecasting trade potential between China and the five central asian countries: under the background of belt and road initiative," *Computational Economics*, vol. 55, no. 4, pp. 1233–1247, 2020.

[2] X. Liu, Y. Chen, and X. Wang, "Research on China-Kazakhstan trade under "the belt and road initiative"—based on the perspective of factor endowment theory," in *Proceedings of the 2019 International Conference on Management Science and Industrial Economy (MSIE 2019)*, pp. 270–274, Atlantis Press, Guangzhou, China, December 2020.

[3] P. Allayarov, S. Arefin, and N. Nurmatov, "The factors affecting Kyrgyzstan's bilateral trade: a gravity-model approach," *The Journal of Asian Finance, Economics and Business*, vol. 5, no. 4, pp. 95–100, 2018.

[4] Z. Kembayev, "Development of China–Kazakhstan cooperation: building the Silk road of the twenty-first century?[J]," *Problems of Post-Communism*, vol. 67, no. 3, pp. 204–216, 2020.

[5] R. Rousseau, "Kazakhstan: continuous improvement or stalemate in its relations with China?" *Strategic Analysis*, vol. 37, no. 1, pp. 40–51, 2013.

[6] Y. Wang, P. Huang, Z. A. Khan, and F. Wei, "Potential of Kazakhstan's grain export trade," *Ciência Rural*, vol. 52, 2021.

[7] H. Xia, R. Liu, L. Zhao et al., "Characterization of ebinur lake virus and its human seroprevalence at the China–Kazakhstan border," *Frontiers in Microbiology*, vol. 10, p. 3111, 2019.

[8] G. Raballand and A. Andrésy, "Why should trade between Central Asia and China continue to expand?" *Asia Europe Journal*, vol. 5, no. 2, pp. 235–252, 2007.

[9] S. Y. Gelvig, "Promotion of Kazakhstan economic development under China-Kazakhstan infrastructure construction [J]," *Вестник КыргЃызско-РоссийскоЃо СлавянскоЃо университета*, vol. 20, no. 3, pp. 23–28, 2020.

[10] A Duisekina and Ж Е. Ашинова, "China-Kazakhstan: cooperation and innovation in education as part of the belt and road initiative," *Journal of Oriental Studies*, vol. 93, no. 2, pp. 26–33, 2020.

[11] J. Yang, Z. Liu, W. Wu, Z. An, L. Zhao, and Q. Sun, "Research on the control of slack flow of China-Kazakhstan crude oil pipeline[J]," *The International Journal of Multiphysics*, vol. 15, no. 4, pp. 409–436, 2021.

[12] F. Natale, A. Giovannini, L. Savini et al., "Network analysis of Italian cattle trade patterns and evaluation of risks for potential disease spread," *Preventive Veterinary Medicine*, vol. 92, no. 4, pp. 341–350, 2009.

[13] M. Bigras-Poulin, R. A. Thompson, M. Chriél, S. Mortensen, and M Greiner, "Network analysis of Danish cattle industry trade patterns as an evaluation of risk potential for disease spread," *Preventive Veterinary Medicine*, vol. 76, no. 1-2, pp. 11–39, 2006.

[14] A. R. B. Oglu and S. V. M. Oglu, "Estimation of potential locations of trade objects on the basis of fuzzy set theory," in *Proceedings of the International Conference on Intelligent and Fuzzy Systems*, pp. 228–237, Springer, Istanbul, Turkey, July 2020.

[15] L. Z. Lin, "A perceptual measure of trade shows using fuzzy quality deployment development," *Expert Systems with Applications*, vol. 37, no. 5, pp. 3921–3933, 2010.

[16] C. Alvarez, J. Corbal, and M. Valero, "Fuzzy memoization for floating-point multimedia applications," *IEEE Transactions on Computers*, vol. 54, no. 7, pp. 922–927, 2005.

[17] C. A. Poveda and A. R. Fayek, "Predicting and evaluating construction trades foremen performance: fuzzy logic approach," *Journal of Construction Engineering and Management*, vol. 135, no. 9, pp. 920–929, 2009.

[18] A. Demir, S. Shawkat, B. N. Majeed, and T. Budur, "Fuzzy AHP and VIKOR to select best location for bank investment: case study in Kurdistan Region of Iraq," , Springer, Cham, 2019pp. 485–510, Effective investments on capital markets.

[19] C. Emmanouilidis, A. Hunter, J. MacIntyre, and C. Cox, "Multiple-criteria genetic algorithms for feature selection in neuro-fuzzy modeling," in *Proceedings of the IJCNN'99. International Joint Conference on Neural Networks*, pp. 4387–4392, IEEE, Washington, DC, USA, August 2002.

[20] M. H. Haghighi, S. M. Mousavi, J. Antuchevičienė, and V. Mohagheghi, "A new analytical methodology to handle time-cost trade-off problem with considering quality loss cost under interval-valued fuzzy uncertainty[J]," *Technological and Economic Development of Economy*, vol. 25, no. 2, pp. 277–299, 2019.

[21] D. O. Ferraro, C. M. Ghersa, and G. A. Sznaider, "Evaluation of environmental impact indicators using fuzzy logic to assess the mixed cropping systems of the Inland Pampa, Argentina," *Agriculture, Ecosystems & Environment*, vol. 96, no. 1-3, pp. 1–18, 2003.

[22] L. A. Zadeh, "On fuzzy algorithms," *fuzzy sets, fuzzy logic, and fuzzy systems: selected papers by Lotfi A Zadeh*, pp. 127–147, China, 1996.

[23] R. J. Kuo, S. C. Chi, and S. S. Kao, "A decision support system for selecting convenience store location through integration of fuzzy AHP and artificial neural network," *Computers in Industry*, vol. 47, no. 2, pp. 199–214, 2002.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 7881636, 11 pages
https://doi.org/10.1155/2022/7881636

Research Article

A Method for Adjusting the Semantic Acceptability of English Corpora Based on the Kano Model

Xuyan Fan

School of Humanity & Science, Northeast Agricultural University, Harbin, Heilongjiang 150030, China

Correspondence should be addressed to Xuyan Fan; fanxuyan@neau.edu.cn

Received 16 May 2022; Revised 16 June 2022; Accepted 5 July 2022; Published 12 August 2022

Academic Editor: Mukesh Soni

The Semantic Accessibility Scale (SAS) is one of the criteria for systematically assessing the semantic readability of corpus texts. With the advent of the Internet, English language content has been widely distributed. This constitutes an adequate corpus for corpus research. However, how to assign and evaluate the semantic acceptability of English literary texts with the aid of corpus has become a hot topic of study for academics around the world. In this paper, we propose an analysis method for corpus semantic acceptance based on the Kano model. This method combines the Kano model with corpus semantic acceptance. Initially, the method identifies the initial corpus semantic acceptance demand items using the initial corpus semantic acceptance identification questionnaire. The Kano categories of each requirement item are then identified and filtered based on the second Kano questionnaire. In this paper, we propose a Kano-based method for corpus semantic acceptance requirement analysis and apply Kano theory to corpus semantic acceptance requirement analysis. First, the corpus semantic acceptance requirements are classified into corresponding Kano categories and filtered using a Kano survey; second, the initial weights of the corpus semantic acceptance requirements are determined using the coarse number method; and finally, the initial weights of the requirements are adjusted using the corresponding Kano adjustment coefficients to obtain the final weights of the corpus semantic acceptance requirements.

1. Introduction

The reform of vocabulary instruction, which is of crucial importance to the reform of the English curriculum, is the most difficult and crucial aspect of the English curriculum reform. As educators investigate the incorporation of corpora into English classrooms, vocabulary instruction within the context of corpora has evolved into a new model in this field. This research has dual purpose: first, corpus-based teaching models are more prevalent in higher education than in secondary schools; the second objective is to apply the functions and theories of the corpus to secondary-level English vocabulary instruction and to investigate the practicability and efficacy of the corpus teaching model.

The new English curriculum reform is currently being implemented in the education sector, which advocates student-centeredness, the improvement of students' comprehensive literacy and differentiation, and the integration of curriculum and modern information technology. The corpus approach exemplifies the connection between curriculum and practice advocated by the new curriculum reform, and its computer and corpus-assisted design satisfies the new requirement for integrating curriculum and modern technology.

Semantic Accessibility Scale (SAS) is a quantitative criterion used to measure the comprehensibility of a text; the study of how to analyze literary texts from a stylistic perspective has received attention from scholars at home and abroad [1, 2]. In natural language text analysis, the Automatic Text Summarization System (ATSS) is similar to this. The SAS of stylistics emphasizes the point-based approach to summarize the writing style of a writer by formally analyzing the content of randomly selected English texts. ATSS for natural language focuses on generalizing from the primary to the secondary, using a computer to automatically extract the main content of a text and generate a semantically coherent digest to cover the original text. Although SAS

source data are selected from literary texts and ATSS is selected from web texts, both emphasize the use of form to analyze and summarize texts and are comparable. With the development of the web, large amounts of literary texts have become electronic [3]. How to analyze and evaluate literary texts online with the help of research results in the field of natural language processing has become a new direction for computational linguistics research. In this paper, the electronic English literary texts are collectively referred to as natural language texts, and the evaluation of their comprehensibility can be carried out by the evaluation model of automatic digest system. At the same time, we try to build an online evaluation system of English text SAS and quantify it by using formulas, so that literary critics can analyze the acceptability of English text formally by computer.

Second, the corpus approach optimizes teaching resources and fosters the growth of teachers' instructional abilities. Traditional teaching resources include syllabi, textbooks, and exercise books. Language learning requires a large number of resources for reference and support. The limited number of resources available to teachers for lesson planning limits their ability to instruct. The corpus provides teachers and students with a substantial amount of authentic corpus, which complements and enhances the teaching resources of teachers. In accordance with their cognitive development, students acquire language knowledge and develop language skills in authentic corpus-generated contexts. The result is not only that learning becomes easier for students but also that teachers' instructional abilities are elevated to a new level. Again, the corpus approach enhances the learning styles of students [4]. Students find traditional vocabulary instruction dull and uninteresting, and they lack motivation to learn. Students are motivated to explore and acquire knowledge because the corpus-based vocabulary teaching method is intuitive and draws their attention. This study focuses on the combination of new methods, new theories, and new practices and emphasizes the feasibility and practicability of the corpus teaching method, which not only provides a new perspective for high school English vocabulary instruction but also serves as a useful reference.

In this paper, we apply Kano theory to the process of analyzing corpus semantic acceptance requirements. Additionally, we propose a method that is based on Kano for analyzing corpus semantic acceptance requirements. First, the corpus semantic acceptance requirements are categorized into corresponding Kano categories and filtered using Kano survey; second, the initial weights of the corpus semantic acceptance requirements are obtained by using the coarse number method; third, the initial weights of the requirements are adjusted according to the corresponding Kano adjustment coefficients to obtain the final weights of the corpus semantic acceptance requirements; and finally, the final weights of the corpus semantic acceptance requirements are presented. The most acceptable corpus is obtained through the utilization of a method known as joint analysis.

2. Related Work

2.1. Corpus and Corpus Linguistics

2.1.1. Definition of Corpus. The term "corpus" is derived from the Latin word "body," which usually refers to any segment or chapter in spoken or written language. The definition of "corpus" has been widely debated by linguists, and there is no unified definition in the linguistic community. There are several definitions given by early foreign linguists who have studied this area, and they have been accepted by the profession at various times. For example, Sinclair defines a corpus as "a collection of linguistic materials selected, classified, and ordered by special rules for use as a sample of language" [5]. Kennedy defines a corpus as "a collection of materials that can be used to analyze and describe a language. Kennedy defines a corpus as "any written text or speech transcription that can be used to analyze and describe language." The famous linguist Henou, in his book, also gives a unique insight into the use of a corpus for linguistic research, which contains a large number of written and spoken texts that are stored and processed with the help of computers for the purpose of application.

This branch in the field of applied linguistics started late in China, but some senior linguists in China have given their own views on the definition of corpus. The definition of corpus given by Saaty "refers to the combination of a large amount of linguistic materials existing in a computer and location retrieval management software [6]." A corpus is an electronic database with a certain capacity built according to the prescribed linguistic principles, using random sampling to gather continuous language usage texts and conversation fragments that occur in daily life.

2.1.2. Classification of Corpora. Nowadays, corpora are rapidly evolving, and different types of corpora are being developed to meet a variety of needs, making corpora classification difficult.

Corpora can be divided into two types based on their specialization: general-purpose corpora and specialized corpora. The general corpus is distinguished by its large capacity, broad range of topics, and high utility. The American National Corpus, the American Contemporary English Corpus, and other general corpora are commonly used. Specialized corpora reflect a specific field of language. For example, if we are studying business negotiation and contract signing in English, we will not be able to find relatively specialized knowledge in a general corpus but will need to use a specialized corpus of business English.

A parallel corpus is a bilingual or multilingual corpus that includes both the original and translated text. According to the translation direction, parallel corpus can be divided into one-way parallel corpus, two-way parallel corpus, and multiway parallel corpus [7]; analogical corpus refers to two or more corpora composed of different forms of variants of the same language or texts of different languages, which can be divided into monolingual analogical corpus and bilingual

or multilingual analogical corpus, for example, ECC (English Comparable Corpus), and so on.

With the application of corpora in language teaching in recent years, a family of corpora has gradually emerged as a learner corpus, i.e., a corpus composed of learners' own acquired works, e.g., the Chinese English Learner Corpus, the Chinese Student English Oral and Written Corpus, and so on.

2.1.3. History of the Corpus. The development of corpora can be divided into two stages in terms of the research on the corpus itself and the application of the corpus. The first stage is the traditional corpus stage. Between the nineteenth and twentieth centuries, a large number of dialect surveys were conducted in the United Kingdom and the United States for language collection, and the results of these surveys were gathered into a dialect corpus, which is the early traditional corpus. In China, evidence of the use of corpus as a research method emerged during the Ze-Shui [8], when the Western Han philosopher and linguist Hu and Zhang [9] used actual surveys to study language in a new way and produced the first monograph on the study of dialect vocabulary in China, "Trifling Xuan Emissary Jedi Language Explaining Dialects of Other Countries." The second stage of modern corpus began in the 1950s [10], when Brown University in the United States began to develop the first machine-readable corpus, and in 1964, Francis and Kucera completed the world's first machine-readable corpus, the Brown Corpus. All these corpora belong to the first generation of corpora. In the 1970s and 1980s, the second and third generation corpora such as the British National Corpus and the Longman English Corpus were rapidly established in the world. In China, the research on corpus started in the late 1970s and early 1980s, and the modern Chinese corpus in China also developed with the application of computers, and some large corpora such as the Chinese Language Resources Consortium and the International Chinese Language Resources Consortium were established. Nowadays, the construction and research of Chinese corpus is still going on in China [11].

2.1.4. Introduction to the COCA Corpus. Professor Mark of Brigham Young University pioneered the Corpus of Contemporary American English (COCA) at the beginning of the twenty-first century and successfully launched the platform in 2008. As one of the largest balanced corpora to date, it has made it possible for more English educators and learners to have access to original English knowledge and has generated a lot of buzz in the corpus academic and language education communities. In terms of size, the COCA corpus has about 360 million words (lexical entries), which makes it more competitive in terms of online corpus size; in terms of speed, the number of processing chips in COCA's servers is 8, each at 2.4 GHz, which ensures that users can find and extract the information they want within seconds. The obvious advantage of COCA is that all the corpus contained in the database are taken from the latest corpus in the past 20 years and updated twice a year, which truly guarantees the timeliness of the corpus. In the rapidly developing world of Internet, the free and universal nature of COCA is its important mark; traditional users are unable to search on the corpus according to their individual needs, such as querying word spacing, difficulty in determining the type of corpus and doing synonym analysis, and so on. The corpus perfectly solves a series of confusions and pioneeringly combines with Google or Yahoo search engines to provide users with more valuable data information.

2.2. Kano Model

2.2.1. Fundamentals of the Kano Model

(1) Theory of Kano Model. The Kano model was first proposed by Professor Noriaki Kano and his colleague Fumio in 1979. In 1984, Noriaki Kano formally proposed the complete structure of the Kano model, which defines three levels of customer needs: basic needs, desired needs, and arousal needs [12]. Traditionally, customer satisfaction varies with the degree of satisfaction of quality characteristics, i.e., customers are satisfied when quality characteristics are satisfied. However, there are some quality characteristics that will not only make customers feel satisfied but even dissatisfied when they are satisfied. Therefore, with the passage of time and more and more scholars' research on the Kano model, according to the relationship between the satisfaction of quality characteristics and customer satisfaction, scholars added two new quality characteristics to the original Kano model and classified the quality characteristics of products into a total of five types, which are charm quality, one-dimensional quality, essential quality, undifferentiated quality, and inverse quality [13]. Since customer needs are satisfied by certain quality characteristics, quality characteristics and customer needs are said to correspond to each other. Similarly, customer needs can be divided into charm needs, one-dimensional needs, essential needs, non-differentiated needs, and inverse needs, as shown in Figure 1.

As shown in Figure 1, the horizontal coordinate indicates the adequacy of the provided product attributes, and the more the axis moves to the right, the more the product attributes are adequate and exceed customer expectations, and the more the axis moves to the left, the more the product attributes are lacking; the vertical coordinate represents customer satisfaction, and customer satisfaction increases and decreases as the vertical coordinate increases and decreases. The following are the specific requirements for the five distinct needs depicted in Figure 1.

(a) *Attractive Requirement.* It is an unexpected requirement of the customer, which is a quality characteristic requirement that can bring surprise to the customer. This kind of requirement is not expected by customers, and providing such requirement can greatly enhance customer satisfaction. When this demand is not satisfied, customers will not feel dissatisfied; when this demand is satisfied, customers will feel very surprised. This kind of demand is generally obtained after market research and in-depth analysis of customer use of the product.

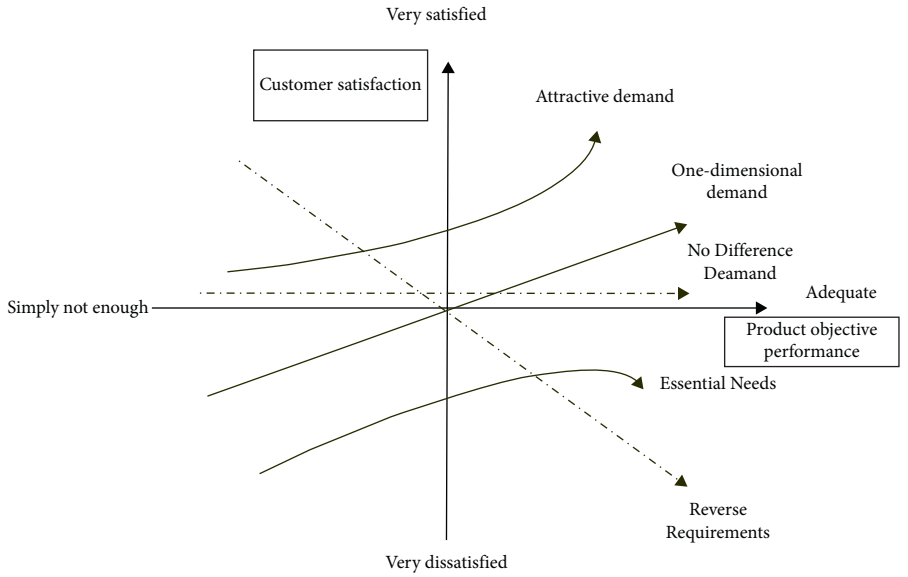


Figure 1: Kano model.

(b) *One-Dimensional Requirement*. The customer's requirement for the quality characteristics of the desired product is the definition of one-dimensional requirement, and the degree of customer satisfaction varies linearly with the degree of satisfaction of this requirement. When this requirement is not satisfied, the customer feels dissatisfied; when this requirement is satisfied, the customer feels satisfied, and the higher the degree of satisfaction is, the greater the customer's satisfaction will be. The questionnaire can be used to get the one-dimensional needs.

(c) *Must-Be Requirement*. It is a quality characteristic requirement that customers think the product should provide. When this requirement of the product is not satisfied, the customer will be extremely dissatisfied, but when this requirement is satisfied, the customer will not be very satisfied because the customer feels that the satisfaction of this requirement is the basic functional characteristic of the product. Essential needs are more basic needs that are invisible, and generally customers do not actively mention these types of needs.

(d) *Indifference Requirement*. It is a requirement that customers do not pay attention to. The satisfaction or non-fulfillment of such requirements and the degree of satisfaction do not affect the level of customer satisfaction.

(e) *Reverse Requirement*. The requirement that customers do not want to appear in the product is reverse requirement, and this requirement is inversely proportional to customer satisfaction.

(2) Advantages of Kano Model. The traditional method of measuring customer preference is to see how satisfied customers are with the quality characteristics. In the measurement, customer satisfaction is usually divided into 5–7 intervals from "very dissatisfied" to "very satisfied," and customers are asked to evaluate such quality characteristics to determine the level of customer satisfaction with the quality characteristics. One problem here is that the relationship between quality characteristics and customer satisfaction is not necessarily linear but may sometimes be non-linear. It is possible that the really important quality characteristics cannot be derived by the traditional linear way of thinking.

The Kano model can express the customer needs dynamically. While the general theory focuses on the analysis of customer preferences, the Kano model focuses on the analysis of the factors affecting customer satisfaction.

(a) By distinguishing the needs of different types of customers, the Kano model can find the quality characteristics that best improve customer satisfaction, which in turn allows companies to gain a detailed understanding of customer needs. It helps companies to find the focus of development before developing products and to know which quality features should be developed first with limited resources.

(b) The Kano model provides criteria for trade-offs between quality characteristics. If a company cannot satisfy all the needs of customers at the same time with limited technology or capital, it can make trade-offs between quality features in terms of their impact on customer satisfaction and customize the quality features that have a high impact on customer satisfaction first.

(c) The Kano model can be used to segment the market more clearly by the different demands of customers for product quality characteristics. For each market segment, the product is designed separately to produce the product that best meets each segment's needs.

(d) The Kano model enables companies to find attractive needs and highlight the highlights of their products, so that they can produce more attractive products and stand out from similar products.

2.2.2. Current Status of Kano Model Research. Zhao et al. [14] incorporated the Kano model in designing the robustness of products and combined the Kano model and Taguchi test to readjust the initial weights of product quality characteristics, and then combined with gray theory, the optimal combination of product design parameters was derived while considering customer satisfaction, making customers have a higher satisfaction with the designed products. Cecilia Garibay [15] chose to combine the Kano model and the QFD approach when evaluating digital books. In order to improve the service quality more effectively, the key factors affecting the improvement of the service quality of digital libraries were identified. Kano et al. [16], with the idea of classifying products and their service quality characteristics, proposed the idea of combining the Kano model with IPA (Importance-Performance Analysis), which can be used to analyze and identify the success factors of the company. In particular, the management strategies are effectively tailored to the success factors, allowing the improvement of the service quality of the company. Kurt Matzler et al. [17] combined Kano model and ANOVA for better market segmentation. The analysis is used to derive the evaluation of different key service quality attributes in various market segments, which helps companies to evaluate the service quality in a more targeted way, and the whole process is illustrated in detail with a case study. Requel [18] integrated the Kano model, fuzzy theory, and 2-element fuzzy-linguistic model to manage after-sales service and make it more effective. Vasilash [19] combined the fuzzy Kano model with QFD to incorporate customer emotions into the product design and to classify customer needs more accurately into the appropriate Kano types and applied this approach to the study of the attributes of the car form. Matzler and Hinterhuber [20] combined the Kano model with QFD and applied it to human environmental modification science. The questionnaire method was used to finally fit the prioritization of methods and applied to the design of workbenches to address the comfort of workbenches. Tan and Pawitra [21] combined the regression design approach with the Kano model to obtain the most desirable combination of design factors. The Kano model works better in reconciling the relationship between performance criteria and customer satisfaction, while identifying the key factors that affect customer satisfaction. Carmen [22] applied the Kano model to perceptual engineering using regression analysis and the Kano model to determine the relative importance of each emotional attribute that influences the purchase behavior of customers.

3. Kano Model-Based Approach to Customer Requirement Analysis

It has a recurrent neural network (RNN), which is developed from artificial neural network. Due to its special structural characteristics, it has a short-term memory function for time series. The recurrent unit of the neural network with feedback is very different from the feedforward neural network. In addition to receiving external input, it also receives the hidden state information of the previous moment, which can be regarded as a temporary memory, while previously due to the limitation of its structure, the feed-in neural network could only receive external input. Therefore, it cannot be used to process speech and other time-sequence signals closely related to the signal before and after. Recurrent neural networks have been widely used in video, speech, text, and other time-sequence related problems [22].

3.1. Process of Customer Demand Analysis Based on Kano Model. In this paper, we propose a Kano model-based customer demand analysis method; firstly, we use the Kano model to determine the category of customer demand and filter the customer demand items, then we combine the coarse number method to get the initial weight of customer demand items, then we assign different adjustment coefficients according to different Kano categories of customer demand items, and finally we get the final weight of customer demand items and its importance ranking; for the multiple level characteristics of key customer demand items, using the joint analysis method, we can finally get the combination of the most popular product parameters. The flow of this method is as follows (Figure 2).

Figure 2 shows the total number of questionnaires distributed in this study and their names on the left, the method applied in this paper on the right, and the main steps of this method in the middle.

(1) Identifying the product's initial customer needs: the product's initial customer needs were identified by reviewing relevant literature, conducting customer interviews, and disseminating the first initial customer needs identification questionnaire.

(2) Determining the Kano categories of customer demand items: the first questionnaire's customer demand items are designed, and the second Kano questionnaire is distributed to determine the Kano category corresponding to each customer demand item.

(3) Requirement screening based on the Kano model: screen the customer requirement items belonging to different Kano categories and keep only the A category requirements and O category requirements.

(4) Determining the initial weights of customer requirements: design and distribute the third importance questionnaire to the retained A and O requirements and calculate the initial weights of customer requirements using the coarse number method.

(5) Weight adjustment based on the Kano model: based on the different Kano categories of customer demand items, the initial weights of customer demand items are adjusted by selecting the corresponding adjustment coefficients, and the final weights of customer demand items are obtained.

(6) Determining the importance ranking of the customer demand items: the customer demand items

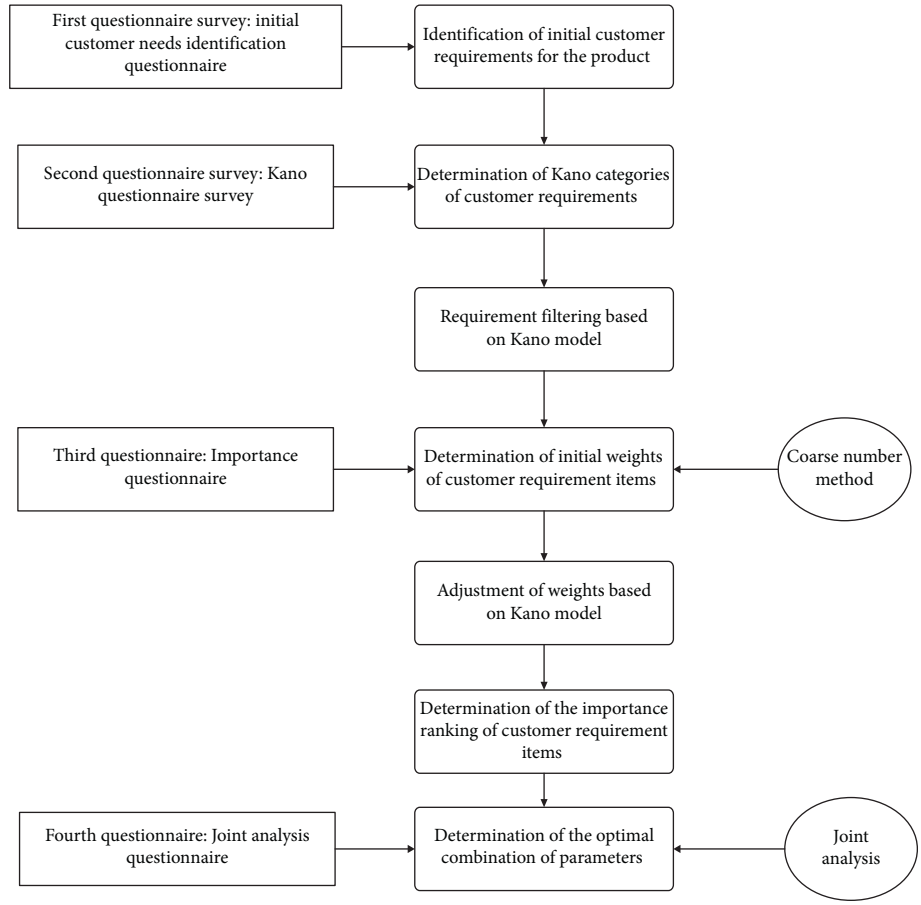


Figure 2: Customer requirement analysis process based on Kano model.

are ranked in the order from the largest to the smallest, and the importance ranking of the customer demand items is obtained.

(7) Determination of the optimal parameter combination: take the top six most important customer demand items and find the level corresponding to each customer demand item, produce a series of virtual products using the orthogonal experimental design method, distribute the fourth joint questionnaire to these virtual products, let consumers score these virtual products, and analyze the obtained results jointly to finally obtain the most popular product parameter combination.

3.2. Steps of Customer Demand Analysis Method Based on Kano Model

3.2.1. Determining the Kano Categories of Customer Demand Items. The application method proposed by Kurt Matzler is the one that is more commonly used today. The idea is that for each quality characteristic of the product, positive and negative questions are asked to determine the Kano category to which each quality characteristic belongs based on the Kano evaluation form [17]. The steps are as follows: first, the design and distribution of the Kano questionnaire are conducted, then obtain the initial customer demand items of the product, and design the Kano questionnaire according to the template given by the Kano theory, and the respondents are asked to score the two cases of providing or not providing the initial customer requirement items of the product. The form of the Kano questionnaire, as shown in Table 1, is as follows.

(2) Kano questionnaires were distributed, and the results were tallied corresponding to the Kano evaluation table [17] shown in Table 2 to obtain the Kano category to which each customer requirement item in each questionnaire belongs. A indicates charm quality, *o* indicates one-dimensional quality, *m* indicates must-have quality, *i* indicates undifferentiated quality, *r* indicates reverse requirement, and *q* indicates that there is a problem with the response (because Q represents that the respondents' answers to the forward and reverse questions are the same).

3.2.2. Determining the Kano Categories of Customer Requirements. After the statistics are completed, the data of each initial customer requirement item are summarized separately and the Kano category of each quality characteristic is obtained by applying the principle of "relative majority," as shown in Table 3. The "relative majority" means that among the five Kano categories, the Kano category with the most product quality characteristics is the Kano category of that quality characteristic.

Table 1: Kano questionnaire item format.

Customer needs	Question	Dislike	Tolerable	Do not care	Take for granted	Like
Ri	The product provides the requirement	1	2	3	4	5
	The product does not provide the requirement	1	2	3	4	5

Table 2: Kano evaluation form.

Positive problem	Reverse problem				
	Like	Take for granted	Do not care	Tolerable	Dislike
Like	Q	A	A	A	O
Take for granted	R	I	I	I	M
Do not care	R	I	I	I	M
Tolerable	R	I	I	I	M
Dislike	R	R	R	R	Q

In order to avoid misclassification of Kano categories, Berger [61] proposed to use the ratio of the coefficient of increasing satisfaction to the coefficient of eliminating dissatisfaction, i.e., "satisfaction/dissatisfaction ratio," to identify the Kano categories of product attributes with relatively similar percentages. The calculation formula is shown below.

$$\begin{aligned} \text{Increase the satisfaction factor} &= \frac{A+O}{A+O+M+I}, \\ \text{eliminate dissatisfaction coefficient} &= \frac{M+O}{A+O+M+I}, \end{aligned} \tag{1}$$

where A is the proportion of charm quality attributes in total attributes, O is the proportion of one-dimensional quality attributes in total attributes, M is the proportion of essential quality attributes in total attributes, and I is the proportion of undifferentiated quality attributes to the total attributes.

3.2.3. Requirement Screening Based on Kano Model. Customers who customize products pay different attention to different categories of customer demand items and treat them differently in order to improve customer satisfaction. The following are the demand screening principles. (1) For category A demand: if such customer demand is not satisfied in the final product, the customer will not be dissatisfied; however, if such demand is reflected in the final product, the customer will be very satisfied. As a result, this type of demand can greatly improve customer satisfaction, so it is critical to maintain this type of demand. (2) For O-type requirements: the more such requirements are provided, the more satisfied the customers will be; the fewer the requirements provided are, the less satisfied the customers will be. As a result, such requirements must be maintained in order to improve customer satisfaction. (3) For M-type requirements: if such requirements are satisfied in the final product, customers will not be satisfied because they believe this is what the product should have; however, if such requirements are not satisfied, customers will be very dissatisfied and will most likely stop using the product. As a result, such requirements must be met in the final physical product; they do not need to be included in the importance ranking of the customized product because they do not contribute significantly to customer satisfaction, but they must be met in the final physical product. (4) For class I requirements: these are requirements that customers do not care about, customers are satisfied when these requirements are met and dissatisfied when these requirements are not met, and such requirements have no impact on customer satisfaction. As a result, such requirements should be removed from the final product, allowing the company to reduce production costs. (5) For R-type requirements: such requirements are inversely proportional to customer satisfaction, and if the product meets such requirements, customer satisfaction will not only not increase but will actually decrease; thus, such requirements must be removed. Table 4 depicts the principles of customer requirement screening based on the Kano model.

3.2.4. Determination of Initial Weights of Customer Demand Items

(1) Design and Distribution of the Importance Questionnaire. The importance questionnaire of customer demand items was designed by dividing each customer demand item into five rating levels: very unimportant, not very important, average, relatively important, and very important, corresponding to the ratings: 1, 2, 3, 4, and 5, using the five-point Likert scale method to retain the A and O customer demands. The questionnaire will be distributed to those who meet the requirements, and the survey of importance questionnaire will be conducted.

(2) Calculation of Initial Weights. The recovered data are analyzed for customer needs by using the coarse number method. By applying the coarse counting method, the initial weight of each customer demand item can be obtained for later calculation. The introduction of the coarse counting method and the improved coarse counting method are given below.

3.2.5. Weight Adjustment Based on Kano Model. In the actual questionnaire survey, because customers think that the essential requirements are indispensable for the product, they tend to give high evaluation to the essential requirements in the importance evaluation, and their initial weights are generally higher; on the contrary, although the charm requirements can bring surprise to the customers, they do not think they are essential, so the initial weights of the charm requirements tend to be lower. Although the initial weight of must-have requirement is high, its contribution to improve customer satisfaction is small; however, the initial

Table 3: Quality characteristic categorization.

Quality characteristics of products	A	O	M	I	R	Total number of questionnaires	Quality characteristic categorization
A	204	99	35	14	0	352	A
B	56	253	32	11	0	352	B
.....	.....	.....	.....	.....	.....	.....	.....

Table 4: Principles of customer requirement screening based on the Kano model.

Customer requirement category	A	O	M	I	R
Screening principle	Keep	Keep	Remove	Remove	Remove

weight of charm requirement is different from its initial weight, but its contribution to improve customer satisfaction is large, and it can significantly improve customer satisfaction. Therefore, the initial weight of customer demand items does not really reflect the importance of customer demand items. When determining the final weight of customer demand items, not only the importance information given by customers but also the contribution of customer demand items to the improvement of customer degree should be taken into account, and both aspects should be considered comprehensively. Therefore, in this paper, we adjust the initial weight of customer demand items according to Kano's theory and use the adjusted weight as the final weight of customer demand items. The specific steps of the adjustment are shown as follows:

(1) Based on the Kano category of the customer demand item, the adjustment coefficient k is selected. The relationship between customer satisfaction and the degree of satisfaction of the customer demand item in Kano theory can be expressed by the following equation:

$$s = cp^k, \tag{2}$$

where s is the customer satisfaction, c is a constant, p is the degree of satisfaction of the customer demand item, and k is the adjustment factor in the Kano theory. For charm demand, $k > 1$; one-dimensional demand, $k = 1$; essential demand, $0 < k < 1$; undifferentiated demand, $k = 0$; and reverse demand, $k < 0$. The exact value of the adjustment coefficient can be determined by the researcher based on previous experience.

(2) Adjust the initial weights of customer requirements based on the k values to obtain the final weights of customer requirement items:

$$w_i' = \frac{w_i k_i}{\sum_{i=1}^{n} w_i k_i}. \tag{3}$$

Let w_i' be the final weight of customer demand item, tw_i be the initial weight of the customer demand item, and k be the corresponding adjustment factor.

3.2.6. Determination of the Ranking of the Importance of the Customer Requirement Items. The final weight of a customer requirement item takes into account both the customer's evaluation of the importance of the requirement item and the contribution of the requirement item to the improvement of customer satisfaction. Therefore, the final weight of a customer requirement item can reflect the importance of that requirement item to a certain extent. The importance ranking of customer demand items is obtained by ranking the items according to their final weights from the largest to the smallest. In practice, it is impossible to satisfy all customer needs when customizing products because of the limited production capacity of the company itself. In the custom production, the company can focus on the more important items according to the importance of customer requirements. Therefore, the importance ranking of customer requirements can provide some reference for customization companies so that they can meet customer requirements very well and improve customer satisfaction.

4. Experimental Results and Analysis

4.1. Data Evaluation. The Old Man and the Sea, a 1954 Nobel Prize-winning work by Ernest Hemingway, was extracted from the English corpus, and six groups of 30 pages, 50 pages, and 124 pages were selected to compare the evaluation results. The result is shown in Tables 5–7.

From the values in Table 8, it can be seen that when the sampling rate SR was 16%, 24%, 32%, 39%, and 52%, respectively, there was no significant difference between their respective SAS values and the full text SAS values, i.e., the systematic evaluation of the same text did not differ in evaluation due to different sampling rates.

4.2. Analysis of Key Quantitative Elements. Using the absolute values of satisfaction and dissatisfaction indices as horizontal and vertical coordinates, respectively, all the indicators are included to form a satisfaction matrix model (as shown in Figure 1). The center lines of the X-axis and Y-axis are the average values of satisfaction and dissatisfaction, respectively. The matrix is divided into four quadrants. Through the analysis of these quadrants, it is possible to quickly identify which quality attributes can effectively improve student satisfaction. In Figure 3, the first quadrant belongs to the area of high satisfaction and high dissatisfaction. The quality elements in this region are the most effective for improving students' satisfaction and reducing dissatisfaction and are the key concerns for improving the quality of course construction. The high charm attributes of teaching content setting (items 1, 2, and 3), teaching material

Table 5: Statistics of the 30-page group.

Extracted pages	Total sentences	Number of words	L	Polyphony number	H	Fog index
6	18	150	8.33	4	2.67	4.22
9	20	182	9.10	7	3.85	4.65
12	19	174	9.16	4	2.34	3.87
13	21	180	8.57	6	3,32	4.21
...	...	...	...	...	...	...
119	14	209	14.93	6	2.43	6.45
121	25	169	6.85	5	3.33	7.22
123	15	187	12.34	5	2.32	6.16
124	11	98	8.94	3	3.55	4.29

Table 6: Statistics of the 50-page group.

Extracted pages	Total sentences	Number of words	L	Polyphony number	H	Fog index
2	18	150	8.33	4	2.67	4.22
3	12	182	9.10	7	3.85	4.65
4	17	174	9.16	4	2.34	3.87
6	18	180	8.57	6	3,32	4.21
...	...	...	...	...	...	...
112	24	199	10.18	6	2.65	5.45
113	22	163	7.89	8	3.31	7.62
118	16	191	9.31	7	2.43	7.18
124	11	98	8.94	3	3.55	4.29

Table 7: Statistics of the 124-page group.

Extracted pages	Total sentences	Number of words	L	Polyphony number	H	Fog index
1	15	148	9.11	3	3.62	4.36
2	18	150	8.33	4	2.67	4.22
3	12	182	9.10	7	3.85	4.65
4	17	174	9.16	4	2.34	3.87
...	...	...	...	...	...	...
121	25	169	6.85	5	3.33	7.22
122	21	180	6.31	6	3.41	6.71
123	15	187	12.34	5	2.32	6.16
124	11	98	8.94	3	3.55	4.29

selection (items 4 and 5), and academic communication module setting (item 12) should be improved as a priority, which can effectively improve student demand satisfaction.

The second quadrant is characterized by low satisfaction and high dissatisfaction. Improving the quality elements in this area can significantly improve students' learning motivation, but it may have little effect on student satisfaction. Only item 13 (the Professional English module) is located in this quadrant, indicating that the Professional English section needs to be strengthened and improved.

The third quadrant is associated with low satisfaction and dissatisfaction, implying that these elements have little effect on increasing satisfaction and decreasing dissatisfaction. This quadrant's quality elements include 8, with item 10 (General Academic English module) having a low charm attribute. The reason for this is that graduate students generally have a stronger English foundation and a greater need for academic English. Item 11 (humanities and arts module) and item 8 (project-based teaching of interdisciplinary themes) are both appealing but less satisfying. This indicates that while students recognize that humanities and arts modules play an important role in improving their English literacy, due to increased academic pressure, they prefer to use their limited time to quickly improve their academic English application skills. The importance of interdisciplinary theme-based instruction is confirmed by research data, but achieving the desired results is difficult. The remaining four quality elements are all irrelevant. Cai Jigang recommends using general scientific humanities or superficial specialized knowledge texts as teaching materials in university academic English classes [14]. However, according to the findings of this survey, graduate students regard these teaching materials (item 6) as unimportant quality elements. Graduate students are more motivated than undergraduates to read and write academic English, and they want instructional materials that are relevant to their major, and instructional materials that account for differences in major categories (item 4) are seen as highly appealing quality elements. Students did not approve of the combination of flipped classroom, online independent learning outside the classroom, and classroom teaching (items 7 and 9) possibly because students had already

Table 8: Statistics of SAS values corresponding to different SRs.

Index	S1	W1	*P*1	*P*2	SR	Fog index	SAS
1	306	4433	0.16	0.17	0.16	7.29	0.14
2	451	6236	0.24	0.23	0.24	7.40	0.14
3	622	8298	0.33	0.31	0.32	7.12	0.13
4	730	10714	0.38	0.40	0.39	7.63	0.14
5	996	13967	0.52	0.53	0.52	7.16	0.14
6	1900	26587	1.00	1.00	1.00	7.13	0.14

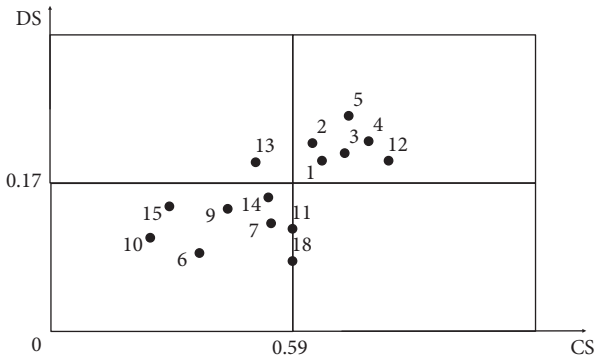


Figure 3: Satisfaction matrix model.

adapted to the traditional classroom lecture teaching mode, and the new teaching mode shifted some of the learning tasks to the classroom, increasing the pressure of independent learning and thus rejecting the new teaching mode. Furthermore, the FIF cloud learning platform's learning resources are too simple to meet students' learning needs. As a result, the enhancement of Project 9 should be prioritized in future curriculum development.

Items 14 and 15, which both involve evaluating teaching objectives, were also deemed irrelevant to quality, indicating that graduate students are more rational in their approach to their grades. While the method of grading is important, they are more concerned with the design of the course content and the selection of instructional materials, both of which are critical components in ensuring the quality of learning.

5. Conclusion

In this paper, we propose a Kano-based method for analyzing corpus semantic acceptance requirements and apply Kano theory to the corpus semantic acceptance requirement analysis. First, through Kano survey, the corpus semantic acceptance requirements are classified into corresponding Kano categories and filtered; second, the initial weights of the corpus semantic acceptance requirements are obtained by using the coarse number method, and then the initial weights of the requirements are adjusted according to the corresponding Kano adjustment coefficients to obtain the final weights of the corpus semantic acceptance requirements. Using a method of joint analysis, the most acceptable corpus is obtained. This paper's innovation is primarily reflected in the enhancement and expansion of the existing research method. The improvement is reflected in the adoption of the coarse number method to determine the initial weight of the corpus semantic acceptance requirements, which not only expresses the true sentiment of customers more precisely but also simplifies the data processing process by calculating the customer's score without adding an index. After applying Kano's theory to the corpus semantic acceptance requirement analysis to determine the importance ranking of the corpus semantic acceptance requirement items, a further extension is made, which reflects the extension. First, a series of virtual products were created using an orthogonal experimental design, then consumers were asked to rate these virtual products, and lastly, the most popular parameter combinations were determined using a conjoint analysis technique.

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 declares that there are no conflicts of interest.

References

[1] S. M. Davis, "Future Perfect," *Reading, Mass*, pp. 140–189, Addison-Wesley Publishing Company Inc, Boston, MA, USA, 1987.

[2] X. F. Shao, P. Q. Huang, and J. H. Ji, "A study of mass customization production model," *Industrial Engineering & Management*, vol. 2, pp. 13–17, 2001.

[3] H.-L. Wang, *Research on the Problem of Product Functional Fatigue*, Northeastern University, Boston, MA, USA, 2010.

[4] Z. Lu, L. Zhang, and C. Tang, "AHP judgment matrix generation algorithm based on preordering and upper rounding function," *Journal of Electronics*, vol. 36, no. 6, pp. 1247–1251, 2009.

[5] J. Xiao, C. S. Wang, and M. Zhou, "A comprehensive evaluation decision for urban power grid planning based on interval hierarchical analysis," *Chinese Journal of Electrical Engineering*, vol. 24, no. 4, pp. 50–57, 2004.

[6] L. Thomas, "Saaty. Decision-making with the AHP: why is the principal eigenvector necessary," *European Journal of Operational Research*, vol. 145, pp. 85–91, 2003.

[7] Z. Xu, "On consistency of the weighted geometric mean complex judgement matrix in AHP," *European Journal of Operational Research*, vol. 126, no. 3, pp. 683–687, 2000.

[8] X. U. Ze-shui, "Research on consistency of the complex interval judgement matrix in AHP," *Journal of Systems Science and Systems Engineering*, vol. 19, no. 2, pp. 159–163, 2000.

[9] Q.-G. Hu and P. Zhang, "Research on improved quality function unfolding based on group policy hierarchy analysis and fuzzy clustering theory," *Computer Integrated Manufacturing Systems*, vol. 13, no. 7, pp. 1374–1381, 2007.

[10] J. Xie, P. Song, and D. Liu, "Method of fuzzy quality function deployment in aircraft top-level design," *Chinese Journal of Mechanical Engineering*, vol. 40, no. 9, pp. 165–170, 2004.

[11] C. K. Kwong and H. Bai, "A fuzzy AHP approach to the determinstion of importance weights of customer requirements in quality function deployment," *Journal of Intelligent Manufacturing*, vol. 13, no. 5, pp. 367–377, 2002.

[12] Z.-Y. Kwan and N. G. Ching-yeol, "Customer requirements clustering analysis and customization optimization under

mass customization model," *Enterprise Management*, vol. 1, pp. 181–183, 2009.

[13] Q. Yang and W. Tang, "An optimal clustering method for determining customer requirements for product families," *Journal of Engineering Design*, vol. 20, no. 2, pp. 97–101, 2013.

[14] W. Y. Zhao, Z. G. Zhang, Z. He, and A. Natanzon, "Coarse number-a customer requirement analysis method," *Computer Integrated Manufacturing Systems*, vol. 17, no. 11, pp. 2493–2501, 2011.

[15] W. Xiong, *Quality Function Unfolding*, Chemical Industry Press, Beijing, China, 2005.

[16] N. Kano, K. Nishina, and K. Suzuki, "Attractive quality and must-be quality," *The Journal of Japanese Society for Quality Control*, vol. 14, no. 2, pp. 39–48, 1984.

[17] K. Matzler, H. H. Hinterhuber, and S. E. Bailom, "How to delight your customers," *The Journal of Product and Brand Management*, vol. 5, no. 2, pp. 6–18, 1996.

[18] Y. L. Long, *Research on Personalized Demand Acquisition Method Based on Kano Model*, Central South University, Hunan, China, 2011.

[19] G. S. Vasilash, "Attractive quality-getting it can help," *Production*, vol. 107, no. 1, p. 64, 1995.

[20] K. Matzler and H. H. Hinterhuber, "How to make product development projects more successful by integrating Kano's model of customer satisfaction into quality function deployment," *Technovation*, vol. 18, no. 1, pp. 25–38, 1998.

[21] K. C. Tan and T. A. Pawitra, "Integrating SERVQUAL and Kano's model into QFD for service excellence development," *Managing Service Quality: International Journal*, vol. 11, no. 6, pp. 418–430, 2001.

[22] Human-Computer Interaction, "Design and user experience, thematic area, HCI," in *Proceedings of the 2020, Held as Part of the 22nd International Conference*, HCII, Copenhagen, Denmark, July 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 1194752, 12 pages
https://doi.org/10.1155/2022/1194752

Research Article

Improved DV-Hop Algorithm Based on Swarm Intelligence for AI and IoT-Federated Applications in Industry 4.0

Lizhi Zhang

College of Engineering and Technology, Xi'an Siyuan University, Xi'an 710038, China

Correspondence should be addressed to Lizhi Zhang; zhanglizhi@mjwhedu.cn

Received 11 April 2022; Revised 14 June 2022; Accepted 22 June 2022; Published 11 August 2022

Academic Editor: Mukesh Soni

In the Internet of Things (IoT) ecosystem, localization is critical for tracking and monitoring targets via nodes. The distance vector-hop (DV-Hop) technique is a good choice for localizing neighborhood in IoT networks. The conventional DV-Hop algorithm is a distributed localization approach that does not consider the distribution of the nodes into full deliberation when calculating the hop count from the source to destined nodes. The transfer distance and node positions thus do not attain higher efficiency while ascertaining the distance between sources and destined nodes. The study aims to resolve the pitfalls in the traditional algorithm by making enhancements in controlling the original DV-Hop algorithm's hop count and transfer distance method by utilizing the particle swarm to estimate the node positions. Error rate in the distance between beacon nodes and unseen nodes is effectively reduced with the proposed technique that calculate error factors with corrections in a reversed fashion to revise hop counts. An escape factor is introduced to take control of updating particles' velocity in the system, and the inertia weight is defined by a piecewise function to enlarge search space. This mechanism increases the diversity of the particle populations and mitigates the tendency of estimations on node positions to be trapped into local optima under stationary state. Also, the improved DV-Hop algorithm described in the paper has a better convergence speed due to the presence of random inertia weight logarithmic method. Finally, the problem of premature convergence is also tackled as a variation factor is adopted in collaboration with a fitness function that affects the particles' movement range and assists in global convergence. The overall performance of improved DV-Hop is evaluated by statistical metrics and also compared with the traditional DV-Hop algorithm under simulated environment with the data collected from real-world scenarios. Industry 4.0 is fully dependent upon IoT and the count of hops is very important for deciding the routing from the source to destination for speedy transmission of data. The improved DV-hop algorithm can achieve better results and has reduced error rate by more than 30%. The DV-Hop algorithm plays an important role in IoT-enabled environment especially in Industry 4.0.

1. Introduction

Wireless sensing networks (WSNs) have a large number of small sensing nodes, which means they have less compute, storage, and transmission capability. Wireless sensor networks are less expensive, have lower power consumption, and are self-configurable sensor nodes [1]. By considering these parameters, WSNs are used in areas like healthcare, for smart transportation, home monitoring, military tracking, environmental monitoring, national security purpose, and indoor navigation [2]. The sensing nodes sense the location and accordingly disseminate the data. For this, the position of the destination node needs to be detected. The detection of position becomes difficult due to the fluctuations of signal and noise in the environment. Many difficulties have been faced for location analysis [3]. Localization methods used in WSNs are independent of the previous localization position. They rely on the position data of a few specific sensor nodes as well as some inter-network measurements. Using Global Positioning System (GPS), the accurate and precise location can be found. However, this technique is impractical due to its high cost, increased power consumption, unavailability of signal, and inefficient performance [4]. Hence, there are many methods proposed in the literature to locate the position of node by exchanging data between the nodes.

Many localization approaches have been presented as a result of the location estimation problem. It is primarily classified as a range-assisted and range-free method [4]. To compute location among neighboring sensing, range-based localization technique uses count of hops, distance, or angle information which requires a higher cost to calculate distance. The range-based technique is based on pattern matching or connectivity measurements. It is categorized as (1) time of arriving (TOA), (2) time difference of arriving (TDOA), (3) received signaling strength indicator (RSSI), and (4) angle of arriving (AOA) [2]. These approaches provide good location accuracy, but they necessitate the use of hardware for location calculation. In contrast, the range-free technique does not require any hardware. The cost is reduced here because it does not use any hardware and also the power consumption is less. To calculate distance between nodes, it uses hop counts and distance approximation algorithm [5].

The range-free approach is further classified as follows: centroid algorithm, amorphous, DV-Hop method, multidimensional scaling (MDS) method, and approximate point in triangulation (APIT) method. The DV-Hop technique is simple and easy to implement in a range-free localization algorithm. As a result, the DV-Hop algorithm is the most often employed [3]. The DV-Hop algorithm reduces localization errors. This algorithm has its localization function. Using this function, it requests an anchor node, which provides information of the node position. Sensor node arranges anchor node and calculates position. This technique gives better scalability and distribution. If sensor node distribution is not uniform, it affects the accuracy. Because of this, the algorithm gives poor localization accuracy [4, 6]. So, the improvement in the existing DV-Hop algorithm has been done. The main contributions of the paper are as follows:

(i) Improved DV-Hop algorithm is proposed which is hybrid of the DV-Hop algorithm and RSSI measurements to enhance the accuracy of localizing nodes.

(ii) An RSSI-based ranging mechanism is used that predicts one-hop distance.

(iii) Levenberg–Marquardt approach is used to compute node position.

(iv) Analysis of DV-Hop algorithm is performed on parameters like mean hop distance, hop count, and node coordinates. Also, error analysis has been done to achieve accuracy.

(v) To minimize the calculation error, the algorithm is improved.

(vi) The formula for mean hop distance, the hop count, and particle swarm optimization has improved. The improved DV-hop algorithm has achieved better results and has reduced the error rate by more than 30%.

Rest of the paper is organized as follows: In Section 2, state of the art of the existing work is discussed. In Section 3, existing working approach of DV-Hop algorithm is elaborated. In Section 4, detailed implementation details of the proposed improved DVH algorithm with the parameters introduced and steps of implementation are elaborated. In Section 5, detailed simulation and experimental result analysis are discussed. In Section 6, the conclusion of the research work is elaborated.

2. Related Works

Many authors have contributed on the same problem statement by offering solutions based on fuzzy-based approaches, evolutionary algorithms, swarm intelligence-based algorithms, and machine leaning-based schemes.

In [7], the authors have addressed the localization technique used to compute the place of nodes using a collection of nodes known as anchors. The density collections of these anchors would be increased or decreased due to various reasons such as maintenance, breakdown, and lifetime. The DV-Hop (DVH) technique is appropriate for the positioning of nodes that consists of a few neighbor anchors. However, the existing DVH-based technique has not taken into account the issue of anchor failure, which can occur during a localization operation. To solve this issue, the authors have proposed an online sequential DVH (OS-DVH) algorithm which is used to calculate the localization of nodes sequentially and to enhance the position accuracy of the nodes for multi-hop WSN. DVH method is used to process node localization using an optimized approach for estimating the average distance of hops between nodes. In [8], the authors have proposed an advanced DVH technique. The authors have tried to lessen the error rate for the DVH technique in two ways. At first, it is equivalent to communication radius when the gap between the two hops is less. These hops are known as sequential hops. The distance between the hops is computed by the shadowing structure. The unknown nodes indicate that the hop size started from beacon nodes; however, the distribution of nodes in WSN is not equal.

In [9], the authors have presented an improved DVH algorithm used to boost the accuracy of the DVH algorithm without enhancement in the computational complexity. The authors presented two different algorithms. Firstly, the presented algorithms use the K-mean approach followed by the repositioning of nodes. The second algorithm also uses the K-mean approach but is followed by cluster division localization. The first algorithm was not used frequently due to certain conditions in the applications during the repositioning of the nodes. Hence, this algorithm cannot achieve accuracy optimally. That is why the advanced algorithm was used to evaluate the distance among hops and have shown better enhancement in accuracy. In [10], the authors have examined the issue of hop count info among nodes having a large influence on the localization accuracy of the standard DVH method. As a result, an advanced method based on RSSI was proposed to overcome the problem. In [11], the authors have proposed to enhance the version of DVH algorithms such as quadratic DVH (QDVH) and unconstrained DVH (UDVH) algorithm for the greater

localization without the requirement of additional hardware for the measurement of range among nodes. The QDVH algorithm is utilized to reduce the error rate to generate higher localization, and the UDVH algorithm attains localization accuracy equivalent to QDVH. The proposed algorithms outperform than the existing DVH algorithm.

When applying the DVH algorithm to node position in WSN, the authors presented an iDVH algorithm in [12], which takes into account deprived precision localization. In this approach, the mean hop size of anchor nodes is enhanced by minimum mean square error (MSE) and is modified by the error factor. Then, the mean hop size among the unknown hops and anchor hop is improved through the dynamic load coefficient. In [13], the authors have discussed the two kinds of localization algorithms (LA) such as range-based LA and range-free LA. Range-based LA, on the other hand, has stringent hardware requirements, making it difficult to implement. In the case of range-free, however, it lowers the hardware cost. This practice is good for only known hops to evaluate accuracy. Hence, to evaluate better accuracy for unknown hops, a proposed RSSI-based DVH is proposed in this paper.

In [14], the authors have presented a DVH positioning algorithm to evaluate better positioning accuracy. The node position is determined based on the radio range of anchor nodes. In this paper, they have used the discovery probability technique for the evaluation of localization error. In [15], the authors have proposed the RMADVH (regular moving anchor DVH) algorithm to enhance the DVH algorithm based on the regular moving anchor node and RSSI range approach. The proposed localization algorithm utilizes a few anchor nodes, achieves distribution of hops equally, and reduces the hardware cost.

In [16], the authors have discussed the localization algorithm, which is of utmost proficiency because of its simplicity, low cost, and less complexity. But having the limitation is poor localization accuracy when the anchor nodes are reduced. Hence, this paper proposed a weighted DVH based on RSSI. In [17], the authors have proposed an improved DVH based on a dynamic anchor node-set (DANS DVH) to increase the accuracy of location. In this proposed algorithm, part of anchor nodes participates in localization, whereas in existing DVH algorithms complete nodes are applied. Binary particle swarm optimization (BPSO) algorithm is utilized to design DANS. In [18], the authors have a detailed analysis of the DVH algorithm which is not efficient in the evaluation of localization accuracy due to various constraints.

In order to overcome the drawbacks of existing approaches, this paper is introducing two new mechanisms such as the weighted averaged approach which is utilized to calculate each hop distance and estimation of unknown node distance and beacon nodes to estimate average of hop distance.

3. DV-Hop Algorithm

The hop counts operate as multipliers in most DV-Hop extensions' algorithms, applying to a normalized distance unit that represents the discrete form of real distance between nodes after quantification. When initializing the nodes, the minimum hop count of a beacon node gets initialized together with that of unseen nodes by simply checking connectivity according to the transmission range in the channel [19]. The current hop count is 1 if the two nodes can communicate or else is infinity (also, in implementations, mark the current hop count as 0 to indicate the nodes are identical). These connectivity-related meta-data are broadcast to the entire network, forming a topology model for future usage [20]. Then, iterate through the connection matrix using the minimum path method, updating the local hop count and sending it to the network. Finally, the number of hops to the destination node is calculated and logged as h_{ij}.

3.1. Mean Hop Distance. Firstly, after calculating the minimum hop count, the value is assigned to the nodes of the whole network through broadcasting, denoted as h_{ij}, and the value of correction nodes is calculated as given in

$$d_i = \frac{\sum_{j\neq i}\sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2}}{\sum_{j\neq i} h_{ij}},$$

$$k_{ij} = \left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2, \tag{1}$$

$$d_i = \frac{\sum_{j\neq i}\sqrt{k_{ij}}}{\sum_{j\neq i} h_{ij}}.$$

After the unseen node receives the transfer distance of the nearest beacon node, it takes the product of the transfer distance obtained from that node to estimate its value to each beacon node as shown in equation (2) [21]. Then, the known nodes and target nodes are calculated by using this algorithm.

$$D_i = d_i \times h_{ij}. \tag{2}$$

Unseen node coordinates: The distance between two points is as given in

$$\begin{cases} \sqrt{\mathrm{k}_{11}} = d_1, \\ \sqrt{\mathrm{k}_{22}} = d_2, \\ \vdots \\ \sqrt{\mathrm{k}_{33}} = d_n, \end{cases} \tag{3}$$

where (x, y) represents the distance between the unknown points.

Equation (3) uses the first $n-1$ terms to subtract the n-th term convertible least squares difference to estimate the unseen node coordinates as shown in

$$\mathbf{X} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{B}. \tag{4}$$

Among them,

$$\mathbf{A} = \begin{bmatrix} (x_1 - x_n) & M & (x_{n-1} - x_n) \\ (y_1 - y_n) & M & (y_{n-1} - y_n) \end{bmatrix},$$
$$\mathbf{B} = \begin{bmatrix} x_1^2 - x_n^2 + y_1^2 - y_n^2 + d_n^2 - d_1^2 \\ \vdots \\ x_{n-1}^2 - x_n^2 + y_{n-1}^2 - y_n^2 + d_n^2 - d_{n-1}^2 \end{bmatrix}, \quad (5)$$
$$\mathbf{X} = [x, y]^T.$$

3.2. DV-Hop Algorithm Error Analysis. Many methods use a correction factor to change the jump distance in order to increase the robustness [21], but the errors of the transfer distance are not weighted. Hence, the improvement is relatively large and has unstable positioning error rate. In other works, an error factor is used to improve the weight; however, since it did not arrange the problem into the early, middle, and later stage based on the particle swarm's properties, in the process of calculation, thus the global optimal position cannot be found. By assistance from some normalization and correction techniques on hop count and hop distance, the coordinates of nodes can be calculated more accurately [22].

3.2.1. Hop Count Error. The hop count is a number that represents the value of how many aliquots a number can be divided into. Previous methods assumed that the distribution of nodes was even enough to ignore disparities in distance per hop between node pairs. Thus, it simply counts every direct communication between two nodes as 1 hop, regardless of the actual distance. Therefore, the errors in hop count get accumulated in the routing process, which makes the final result inaccurate [23].

3.2.2. Error in Mean Hop Distance. The entire traditional algorithm resolves the routing paths as straight paths connecting node pairs, as opposed to polygonal lines in most cases when the node positioning strategy is generally considered random. Therefore, the current distance estimate model representing the airline distances rather than the actual routing paths would certainly produce large errors, especially when its propagation routes cannot be reduced to straight lines.

Maximum limit method for the error of point coordinates: From the formula of calculating the coordinates of points by the maximum likelihood method [24], the coordinates (x, y) of the unseen nodes are the intersections of the circles with the radius of $d_1, d_2, d_3, \ldots, d_m$ and $(x_1, y_1), \ldots, (x_m, y_m)$ as the center of the circle.

However, in practical applications, the circles would not intersect at one point in most cases, but intersect in a small area. Therefore, there certain error is obtained in the node coordinates that cannot be eliminated. Moreover, when the n^{th} equation is sequentially subtracted by the $(n-1)^{th}$ equation, a large iteration error is generated each time [25].

4. The Proposed Improved DVH Algorithm

According to the actual distribution of sensor nodes, the principle of shortening the gap between actual data and ideal conditions is adopted to minimize the calculation error as much as possible. When resolving the transition distance and number of transitions, consider the distribution of the sensor nodes in actual situations and then improve it according to the different distribution conditions by putting weight to the discrete raw data with the help of correction factor or error factor.

4.1. Improvement of the Mean Hop Distance. Calculating the average transfer distance by multiplying the transition count will directly affect the accuracy of the algorithm. In this paper, two variables are introduced as error factors to control the weight and average transfer distance of the beacon node.

First, this error factor calculates the transfer distance error of beacon nodes. The algorithm for most DV-Hop extensions adopts the principle of proximity which only accepts the first message sent by beacon nodes. The individualized data selection cannot match the random distribution situation of the actual nodes. It reduces the accuracy rate of the algorithm.

Then, this algorithm takes the measuring average error caused by the average node and the average numbers of transfer count of any single beacon node by using the iterative error generated by the traditional algorithm when calculating the distance between the nodes.

Equation (6) describes how to calculate measuring mean error of transfer distance of beacon nodes E.

$$E = \frac{1}{2}(e_1 + e_2). \quad (6)$$

In equation (6), e_1 and e_2 represent error rates.

$$e_1 = \left(d_i - D_{\text{avg}}\right)^2, \quad (7)$$

e_1 represents the error, which is generated by comparing the average transfer distance of the beacon node with the actual hop distance as shown in

$$D_{\text{avg}} = \frac{1}{n}\sum_{i=1}^{n} d_i, \quad (8)$$

D_{avg} represents the mean transfer distance of all beacon nodes. n represents the number of beacon nodes. d_i represents the mean transfer distance of beacon nodes.

The error rate of e_2 is represented by

$$e_2 = \frac{1}{n}\sum_{i \neq j}\left(\frac{d_{ij} - D_{ij}}{hop_{ij}}\right)^2, \quad (9)$$

e_2 represents mean transfer distance error of every hop. d_{ij} denotes the actual distance of the beacon nodes. hop_{ij} denotes the minimum transition count.

The distance of the beacon node is represented by

$$D_{ij} = d_i \times hop_{ij}, \tag{10}$$

where D_{ij} represents the estimated distance between beacon nodes.

A weighted average method is proposed to reduce the error by making the weighted value inversely proportional to the average error. The weights are calculated as shown in

$$w_i = \frac{(1/|E|) + (1/hop_i)}{\sum_{j=1}^{M}\left((1/|E|) + \left(1/hop_j\right)\right)}. \tag{11}$$

In equation (11), hop is the transition count between unseen nodes and the beacon node. M is the number of nodes.

Finally, the modified weighting method is used to calculate by using

$$d = \sum_{i=1}^{M} w_i d_i, \tag{12}$$

where d_i is the original transfer distance and d represents the improved transition distance.

4.2. Improvement of the Hop Count. The hop count is defined as a value that describes the amount of normalized distance unit. The conventional DV-Hop algorithm records the hop count as 1 hop abstractly regardless of actual hop distance. Obviously, this hypothesis is improper. In response to this problem, error factor and correction factor are introduced in this paper.

First, get the correction factor by broadcast

$$\xi_{\text{ij}} = 1 - L_{ij}^2, \tag{13}$$

where L_{ij} represents the error factor. Larger L_{ij} represents greater error as given in

$$L_{ij} = \frac{h_{ij} - H_{ij}}{h_{ij}}, \tag{14}$$

where h_{ij} represents the estimated transition count. H_{ij} represents the ideal hop count. The ideal hop count is calculated as actual distance d_{ij} as given in equation (15) which divide the communication radius R.

$$H_{ij} = \frac{d_i}{R}. \tag{15}$$

Then, calculate the improved hop count using

$$hop_{ij} = \left(1 - \frac{\left(h_{ij} - H_{ij}\right)^2}{h_{ij}^2}\right). \tag{16}$$

Finally, through the calculation of unseen nodes, an accurate location is found in

$$D_s = d \times hop_{ij}. \tag{17}$$

4.3. Improvement of the Particle Swarm Optimization. Particles in particle swarm algorithm not only act independently and move randomly, but also share information and cooperate among particles. The algorithm uses iteration search to find optimal solution. Then, we calculate the optimal position of the target node in the entire particle population.

The speed and position are updated as follows:

$$v_{id}^{k+1} = w * v_{id}^{k} + c_1 * r_1\left(p_{id}^{k} - x_{id}^{k}\right) + c_2 * r_2\left(p_{gd}^{k} - x_{id}^{k}\right), \tag{18}$$

$$x_{id}^{k+1} = \mathrm{x}_{id}^{k} + v_{id}^{k+1}. \tag{19}$$

In the common particle swarm algorithm, the particles will fall into a relatively stable stage which causes premature convergence of the particles. During this period, the convergence of particles will slow down so that the particles are difficult to achieve global optimal, thereby affecting the local minimum and increasing the difficulty of particle escape.

In this paper, the improvement has been made. Firstly, the particle velocity update equation is changed. This escape element is introduced to velocity update equation to disturb the particle learning strategy, thus escaping the local optimum. Then, the weight of the particle swarm algorithm is changed into a classification function classified according to the number of iterations, and different weights are calculated for different iteration times. Finally, the variation factor s is added to enhance the population diversity and reduce the probability of premature.

4.4. Improvement of the Particle Velocity Update Formula. To prevent particles from falling in local optimum in stable stage, this paper puts a premature flag to determine current position of particles by examining whether it is in the standard threshold. If the algorithm is in a normal state, it is optimized by the standard particle swarm algorithm. When the flag reaches the set threshold, it is judged that the particle enters the premature convergence at this time. At this time, a central learning strategy is applied to the particles, and an escape factor is defined to avoid premature aging of the particles. Let the particles escape the local optimum and continue to find the global optimal as in

$$\begin{aligned} v_{id}^{k+1} = {} & w * v_{id}^{k} + c_1 * r_1\left(p_{id}^{k} - x_{id}^{k}\right) + c_2 \\ & * r_2\left(p_{gd}^{k} - x_{id}^{k}\right) + c_3 * r_3\left(c_{\text{end}}^{k} - x_{id}^{k}\right). \end{aligned} \tag{20}$$

In equation (20), k is number of iterations, w is inertia weight, c_1 and c_2 are acceleration constants, p_{id} is the optimal position at which the particles are present, and p_{gd} is the optimal position of the global particles. r_1, r_2, r_3 are constants in the interval (0, 1), and x_{id} is the position vector of the particle.

Define $c_{\text{end}} = r_1 * p\text{best} + r_2 * p\text{gest}/2$ as the escape factor, the *pbest* is the optimal position in the transition process *i*, and *pgest* is the global counterpart.

4.5. Improvement of Weights. The inertia weight is usually used to control search ability of particle as shown in equation (21). Based on this theory, this paper proposes to set the weight as a piecewise function.

$$w = \begin{cases} w_{\max} - 0.1 \cdot (w_{\max} - w_{\min}) \cdot e^{\dfrac{k}{100}}, \\ w_{\max} - (w_{\max} - w_{\min}) \cdot \log\left(1.15 \cdot \dfrac{k}{100}\right) + 0.1A, \end{cases} \tag{21}$$

where $w_{\min}$ and $w_{\max}$ represent the minimum and maximum inertia weights. $A = 0.5 \cdot rand\ (\,)\,0$.

By gradually decreasing the inertial weight of the entire particle population with a random algorithm, the accuracy of the particle swarm algorithm is improved.

4.6. Addition of a Variation Factor. Variation factor is introduced to clearly show the fitness of particle between variation factors *s* and s_i. It is simpler to choose particles with relatively low fitness to continue iteration. This not only enlarges search space and increases the population diversity, but also clearly shows the distribution situation of new nodes, as represented in

$$s = s_i - a\left(s_{i1} - s_{i2}\right) + (1 - a)\left(s_i - pg\right), \tag{22}$$

where s_i represents the position of the current particle, s_{i1} and s_{i2} are any two particle positions except the whole population, and $s_{i1} \neq s_{i2}$.

4.7. Selection of Fitness Function. The distance error is expressed by

$$\begin{cases} \sqrt{\mathrm{k}_{11}} = d_1 + \varepsilon_1, \\ \sqrt{\mathrm{k}_{22}} = d_2 + \varepsilon_2, \\ \vdots \\ \sqrt{\mathrm{k}_{nn}} = d_n + \varepsilon_n. \end{cases} \tag{23}$$

When ε equals to its minimum, position of the optimal, the reciprocal of hop counts is introduced as weight to control error caused by estimated distance of numerical high hop in variation of fitness function. Fitness function is all shown in

$$fitness(x, y) = \sum_{i=1}^{n}\left(\frac{1}{h_i}\left|\sqrt{\mathrm{k}_{ij}} - d_i\right|\right). \tag{24}$$

Minimize the value of $fitness(x, y)$ by multiple iterations to improve positioning accuracy.

4.8. Steps of Proposed Improved DVH Algorithm

(1) Calculate hop count *h*.

(2) Calculate the error of hop count by

$$L_{ij} = \left(h_{ij} - H_{ij}\right)/h_{ij}. \tag{25}$$

(3) Calculate the correction factor by the reciprocal of error. Use correction factor to correct the hop count to get new hop count with less error as shown in

$$hop_{ij} = \left(1 - \frac{\left(h_{ij} - H_{ij}\right)^2}{h_{ij}^{\ 2}}\right)h_{ij}. \tag{26}$$

(5) Calculate mean transfer distance.

(6) Calculate weight according to error of hop count by using

$$hop_{ij} = \left(1 - \frac{\left(h_{ij} - H_{ij}\right)^2}{h_{ij}^{\ 2}}\right)h_{ij}. \tag{27}$$

(7) Use weight to correct transfer distance to get new transfer distance with less error by using

$$d = \sum_{i=1}^{M} w_i d_i. \tag{28}$$

(8) Distance between beacon node and unseen node is shown in

$$D = hd. \tag{29}$$

(9) Initialize *k*, *N*, c_1, c_2, $w_{\max}$, and $w_{\min}$.

(10) Update speed formula and calculate the optimum position of global particle.

(11) Get the final coordinate of node (x, y).

The optimized and improved location algorithm is as follows:

(1) The minimum hop count is calculated by the shortest path method.

(2) For each beacon node, first repeat steps (13)–(16) to correct the number of bars, and the result of the loop execution is the average hop distance.

(3) Use equations (6)–(9) to correct the mean transfer distance and combine the hop count to estimate the distance from each beacon node.

(4) Set relevant parameters.

(5) Initialize a certain number of particles in a given area.

(6) The variation factor *s* is generated by (22) and compared with s_i to preserve the less adaptable.

(7) Let $t = t + 1$, and update the velocity and position.

(8) The fitness value of each particle at its current position is calculated using the fitness function $fitness(x, y)$.

(9) First, compare the individual optimal values of individuals into *pbest*.

(10) Node positioning is based on the optimized particle swarm algorithm.

(11) Determine whether the iteration stop condition is reached, and if not return to step 8. Otherwise, the optimal coordinate solution is output.

(12) The optimal coordinate solution of the current output is used as the positioning coordinate of the unseen node.

5. Experimental Results and Analysis

5.1. Simulation and Result Analysis. The simulation was carried out on the MATLAB 2019b platform. The datasets used in this paper are the longitude and latitude of weather station and mobile terminals of the main campus of BUPT. The data selection range of weather station information in partial areas is from 99.228211E, 41.186073N to 117.404741E, 25.930456N, while the selection range of mobile terminal information in the main campus of BUPT is from 116.36188E, 39.969686N to 116.36806E, 39.969686N as shown in Figures 1 and 2. After normalization, the latitude and longitude are, respectively, mapped to the interval of (0, 1) in proportion to meet the input data requirements of the algorithm.

The advantages of this algorithm from the above results can be observed. The two pictures simulate sensor clusters with datasets of two real environments that shows sensor nodes are distributed in different places at random states. Red stars and black dots are used to represent beacon nodes and unseen nodes, respectively. After random shuffling of the dataset, the first few ones are selected as the beacon nodes. In each experiment, the beacon nodes obtained were fixed at the same scale.

5.2. Analysis of Experimental Results. Hop count and hop distance are shown in Figures 3 and 4.

Hops count and hop distance have an impact on maximum likelihood estimation. The suggested technique improves a variety of input parameters, including the number and distance of leaps transferred, as well as the positional fitting model that processes them. In terms of hop count, the hop count resolved by the original algorithm using simple linear division method cannot fit into certain scenarios very well, where the nodes are gathered partially while they are discretized globally. For instance, although the original DV-Hop works well with evenly distributed nodes, as the weather station dataset shown in Figure 1, it cannot handle datasets like mobile terminals inside buildings as shown in Figure 2 and can be sensitive to these changes. So, the improved DV-Hop algorithm can cover more cases with better precision by weighting the original maximum hop count, which can better reflect the actual routing path.

The actual node distribution is complicated, and the hops between nodes are completely different. To solve this problem further, the article solves the problem of dropping mean transfer distance information through broadcasting

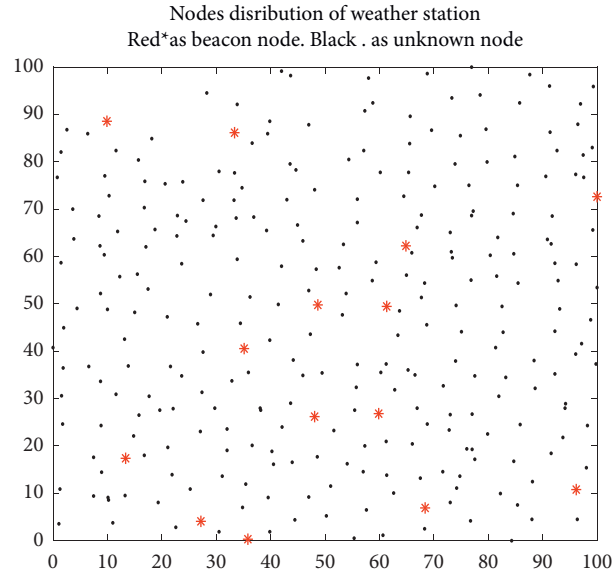


Figure 1: Weather stations in partial areas of China.

and introduces two error factors averaged algebraically. From Figures 3 and 4, when solving the hop count and hop distance, the average accuracy is higher; thus, it has higher robustness under different concentrations of node dataset. By comparison, it is found that the derivative of the traditional calculation is unstable, so serious deviation will occur in the operation. In other words, the distribution of nodes has high randomness in the process of selecting beacon nodes. This is because the two datasets have different properties on the distribution of nodes. It shows that the traditional original algorithm cannot effectively use the meta-information captured in the broadcast communication process. The improved DV-Hop algorithm aims to solve the error loss and utilizes more information of the broadcast mechanism, so that the accuracy of the estimated distance is steadily improved, and the impact is smoother as the nodes' distribution feature changes. Use particle swarm to improve the unseen node coordinate estimation. Figure 5 shows average error against transmission radius with 30 particles and 50 iterations, and Figure 6 shows average error with 30 particles and 100 iterations. Figure 7 shows average error with 100 particles and 100 iterations.

Figure 8 shows average error against beacon nodes with 30 particles and 50 iterations, and Figure 9 shows average error against beacon nodes with 30 particles and 100 iterations. Figure 10 shows average error against beacon nodes with 100 particles and 100 iterations.

The enhanced technique can estimate the unseen node with accuracy. The original technique uses the estimated distance as the radius and uses the coordinates of m beacon nodes as the center of the circle. The m circles are difficult to intersect at one point, leading to a certain error in the obtained node coordinates. After iteration, the error becomes larger, so the position accuracy is greatly affected. Unlike previous methods that employ the maximum likelihood estimate approach to determine node coordinates, the modified DV-Hop algorithm chooses to improve on the

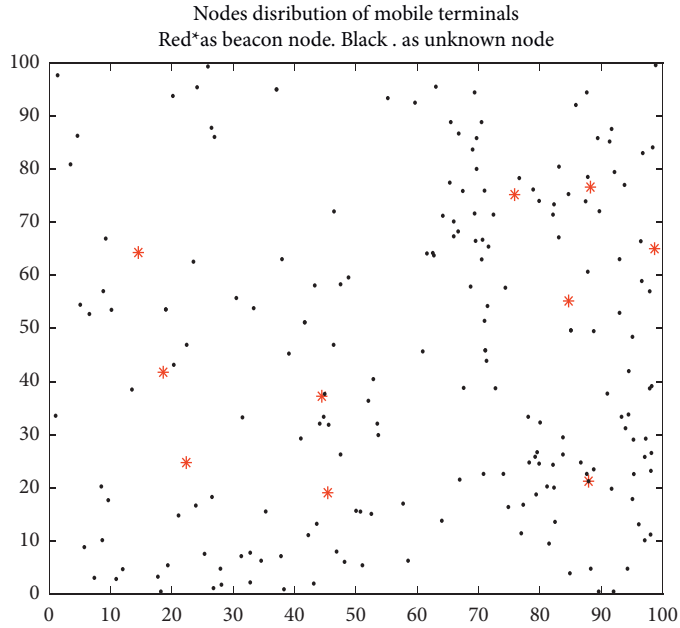


Figure 2: Mobile terminals of the main campus of BUPT.

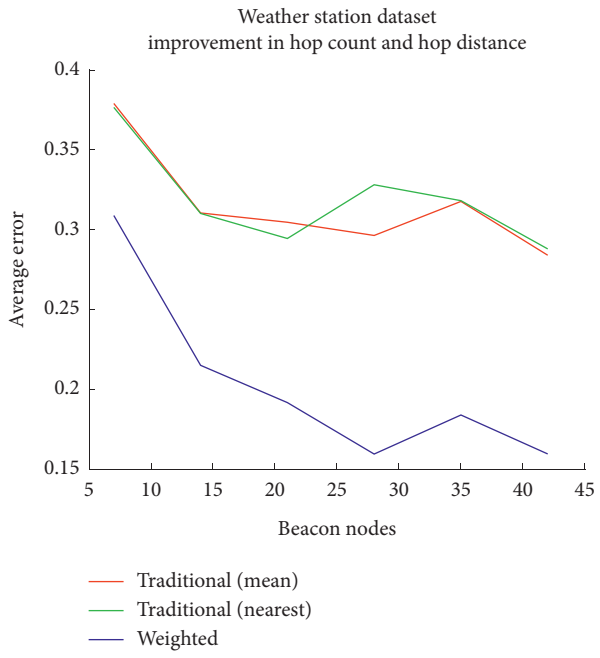


Figure 3: Hop count and hop distance on workstations.

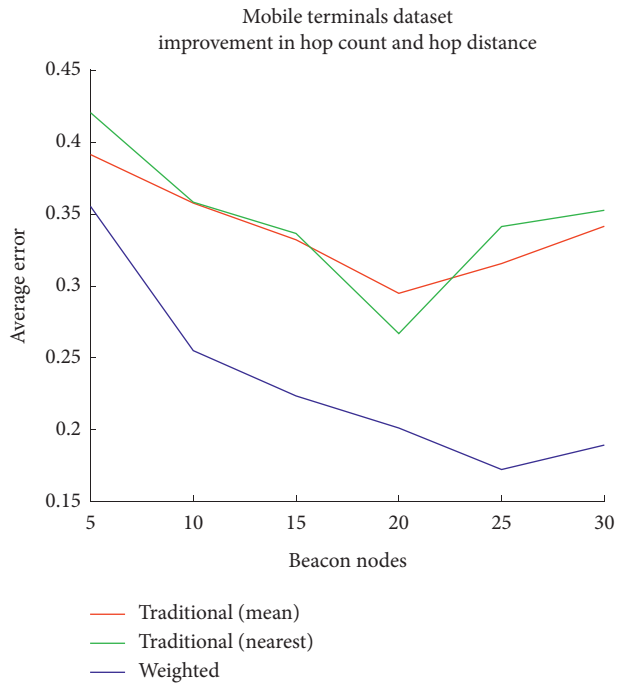


Figure 4: Hop count and hop distance on mobile terminals.

classic particle swarm algorithm. It incorporates escape factors to increase and manage particle update speed, preventing particles from escaping from the local optimum owing to premature status change, as well as finding the global optimum rapidly. Then, a piecewise function is introduced to calculate inertia weight in the iterative process, and weights are calculated according to the number of iterations.

The comparison between Figures 3 and 9 shows that this algorithm can effectively calculate the location of unseen nodes. By comparing Figures 5, 6, 8, and 9, it is concluded that the proposed algorithm optimizes the hop counts and hop distances. For example, within the communication range of 10, the average error of the experiment with 30 particles is 0.38, and the average error of the experiment with 100 particles is 0.24, so the improvement is more than 36%. The rest of Figures 6 to 10 shows the difference in average error when using different iteration times. After reducing the number of iterations from 100 to 50, the results of the algorithm did not change significantly; that is, analyzing its advantages, the number of iterations increases without causing an overreaction.

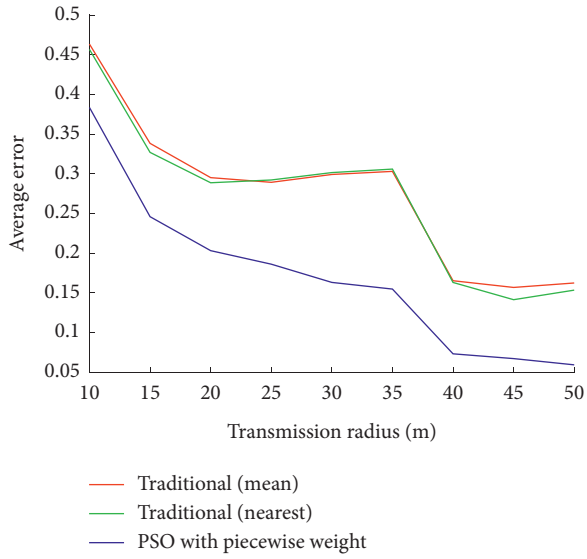


Figure 5: 30 particles and 50 iterations.

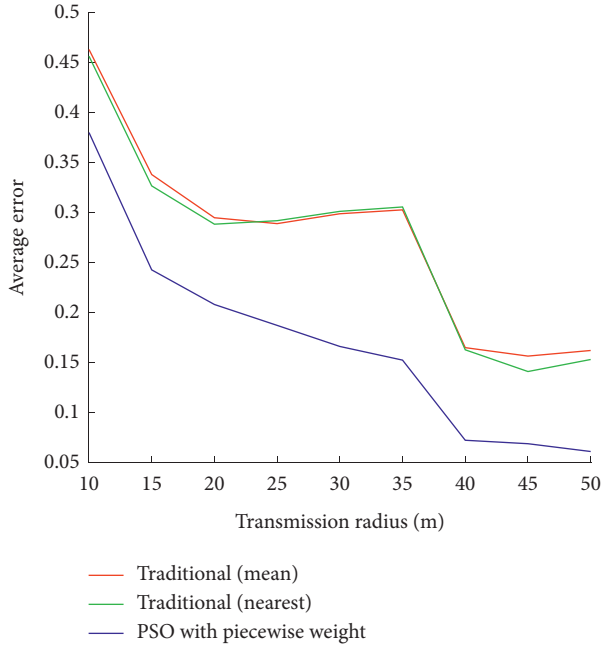


Figure 6: Average error for 30 particles and 100 iterations.

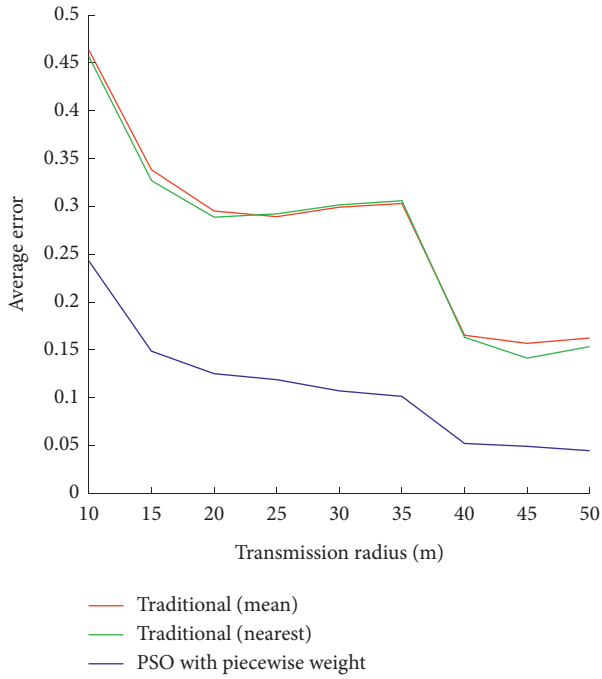


Figure 7: Average error for 100 particles and 100 iterations.

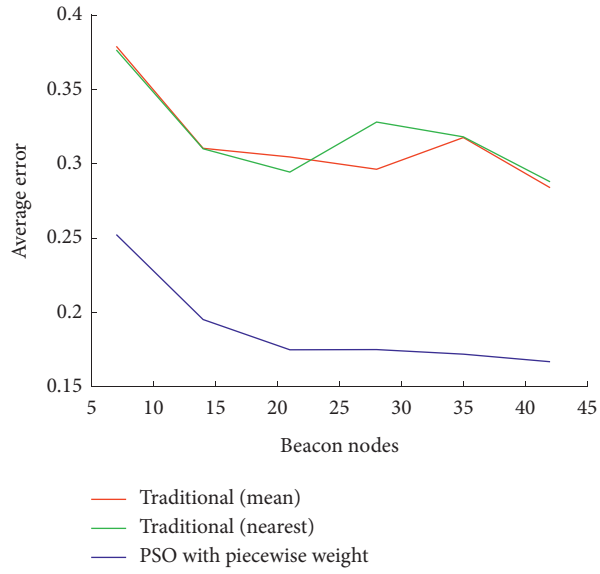


Figure 8: Average error for 30 particles and 50 iterations.

Figures 11 and 12 show the comparison of the two algorithms. The average error of this algorithm decreases more obviously with the change of communication range. Considering that the routing protocol used in this experiment is pre-static routing, the shortest path method is adopted. The communication range has little effect on the final routing path, so the main reason for the rapid improvement of efficiency in this part is the removal of invalid communication nodes, and the coordinate estimation accuracy of the node of this part is relatively close to a two-dimensional uniform distribution which can be found in experiments, especially if the mobile terminal dataset is used. After the communication range keeps increasing, the number of isolated nodes can be ignored. It can be observed that the two types of algorithms are in a stable state, but due to the improved DV-Hop algorithm using its reasonable correction of number of conversion and distance, the error rate of the results in any dataset all shows a monotonously decreasing nature. The traditional algorithms have a situation where the error rate increases as the communication radius increases. Although the performance of the communication range around 40 m that appeared on the weather station dataset has improved due to the local distribution

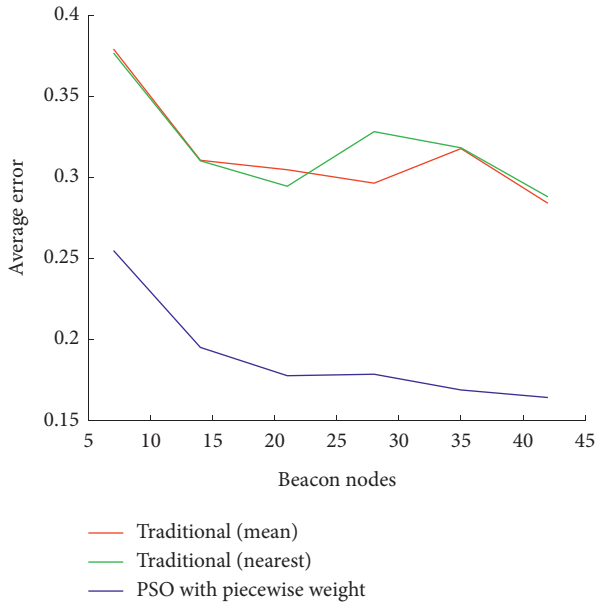


Figure 9: Average error for 30 particles and 100 iterations.

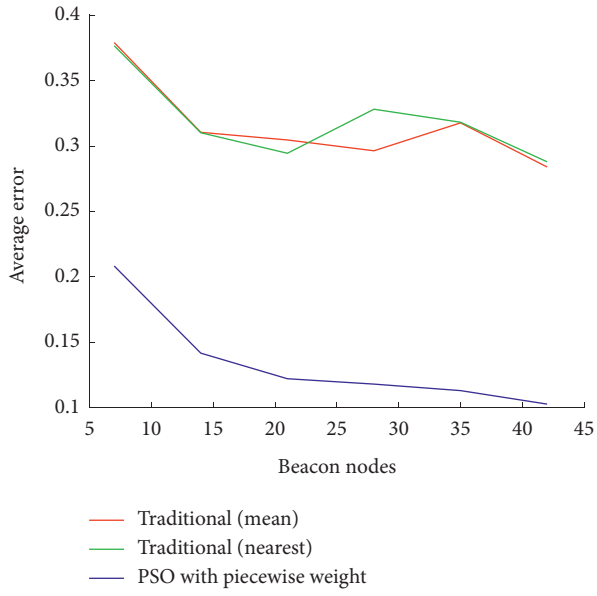


Figure 10: Average error for 100 particles and 100 iterations.

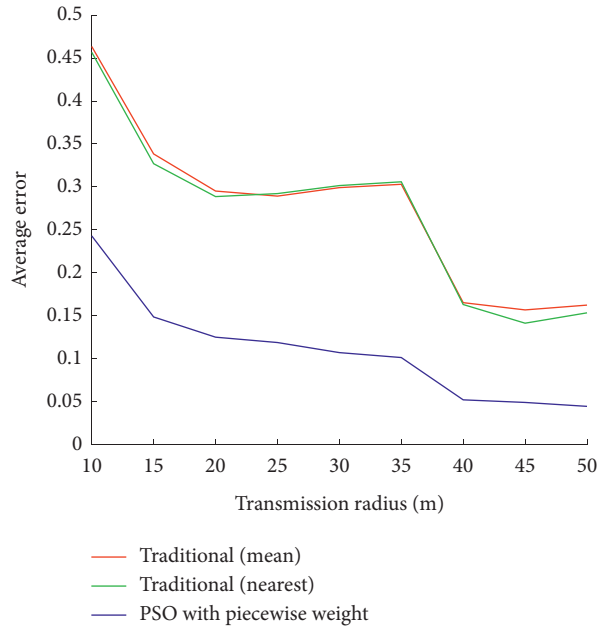


Figure 11: Influence of beacon node density on positioning error.

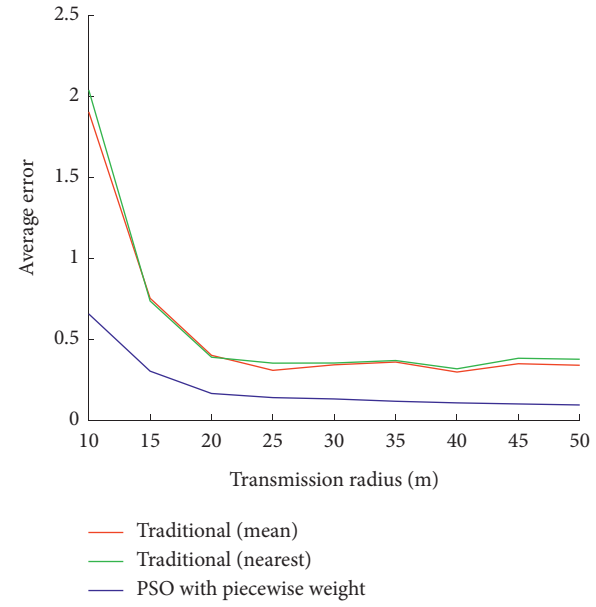


Figure 12: Coordinate estimation influenced by communication range.

characteristics of nodes, the error rate of the traditional algorithm remains stable at 0.5, which is close to unavailable on the mobile terminal dataset. In this situation, the improved DV-Hop algorithm's predicted coordinate error rate is 1/2 to 1/5 of the traditional algorithm's error rate. Influence of transmission radius on positioning error is shown in Figures 13 and 14.

Figures 13 and 14 show the variation trend of the average error calculated according to the different beacon nodes in different datasets. Moreover, for the traditional DV-Hop that uses maximum likelihood estimation, more beacon nodes mean longer estimation depth, which has a certain effect on the improvement of accuracy. Under large quantity of beacon nodes, traditional DV-Hop is restricted by the transfer distance acquisition under the proximity principle. As new beacon nodes are added to the global transfer distance value, the global transfer distance value changes greatly, resulting in the accuracy rate, and fluctuates up and down. The derived traditional algorithm basically effectively alleviates the instability problem by averaging the broadcasts of all beacon nodes, but it does not contribute to the overall accuracy rate, and there is still a serious error rate rebound. The improved DV-Hop algorithm can achieve better results in both datasets and has reduced error rate by more than

Figure 13: Coordinate estimation affected by beacon nodes.

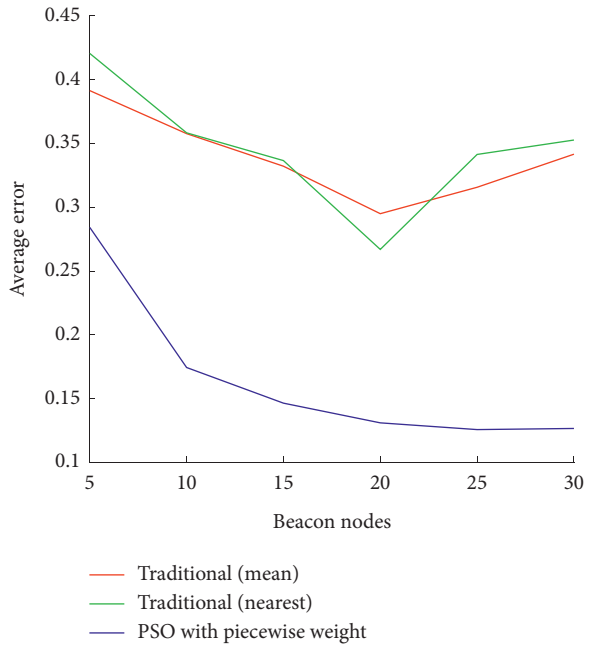


Figure 14: Average error influenced by beacon nodes.

30%. Until the beacon node accounts for 10% of all nodes, it has not entered the platform period. And as the proportion of beacon nodes increases, it shows a strict monotonic decrease, showing good robustness.

5.3. Discussion. In this paper, the improvement has been made in traditional DV-Hop algorithm. Firstly, the particle velocity update equation is changed. The escape element is introduced in velocity update equation to disturb the particle learning strategy, thus escaping the local optimum. Then, the weight of the particle swarm algorithm is changed into a classification function and classified according to the number of iterations. Different weights are calculated for different iteration times. Finally, the variation factor is added to enhance the population diversity and reduce the probability of premature convergence. To prevent particles from falling in local optimum in stable stage, this paper puts a premature flag to determine current position of particles by examining whether it is in the standard threshold. If the algorithm is in a normal state, it is optimized by the standard particle swarm algorithm. When the flag reaches the set threshold, it is judged that the particle enters the premature convergence at this time. The original DV-Hop works well with evenly distributed nodes, as the weather station dataset shown in Figure 1; it cannot handle datasets like mobile terminals inside buildings as shown in Figure 2 and can be sensitive to these changes. So, the improved DV-Hop algorithm can cover more cases with better precision by weighting the original maximum hop count, which can better reflect the actual routing path.

6. Conclusion

The study aims to propose an improved DV-Hop algorithm for IoT-enabled Industry 4.0 applications which make use of wireless communications, and hop count plays an important role. The improved DV-Hop improves the transfer distance method by using the advantages of particle swarm for the assessment of the node positions. Error rate in the distance between known and unseen nodes is optimized with the proposed technique that calculate error factors with corrections in a reversed fashion to revise hop counts. A new escape factor is devised to take control of updating particles' velocity in the system, and the inertia weight is defined by a piecewise function to enhance the search space. This mechanism increases the diversity of the particle populations and mitigates the tendency of particles' estimations on node positions to be trapped into local optima under stationary state. The improved DV-Hop algorithm described in the paper has a fast global convergence speed due to the presence of random inertia weight logarithmic method. The overall performance of improved DV-Hop is evaluated as shown in result section and is also compared with the traditional DV-Hop algorithm under simulated environment with the data collected from real-world scenarios. The DV-Hop algorithm plays an important role in IoT-enabled environment especially in Industry 4.0. In the future, we will propose a case study by using the improved DV-Hop algorithm.

Data Availability

The data are available for users on request.

Conflicts of Interest

The author declares no conflicts of interest with respect to this article.

Acknowledgments

We are thankful to all who have supported us to accomplish this research work.

References

[1] D. Xue, "Research of localization algorithm for wireless sensor network based on DV-Hop," *J Wireless Com Network*, vol. 218, 2019.

[2] S. Messous, H. Liouane, O. Hamam, and H. Hamam, "Improved recursive DV-hop localization algorithm with RSSI measurement for wireless sensor networks," *Sensors*, vol. 21, no. 12, p. 4152, 2021.

[3] S. Messous and H. Liouane, "Online sequential DV-Hop localization algorithm for wireless sensor networks," *Mobile Information Systems*, 2020.

[4] V. Kanwar and A. Kumar, "DV-Hop-based range-free localization algorithm for wireless sensor network using runner-root optimization," *The Journal of Supercomputing*, vol. 77, no. 3, pp. 3044–3061, 2021.

[5] M. Peyvandi and A. A. Pouyan, "An improved DV-Hop localization algorithm in wireless sensor networks," in *Proceedings of the 2015 Sig. Proc. & Int. Sys. Conf. (SPIS)*, pp. 153–158, IEEE, Wuhan, China, June 2018.

[6] Y. Hu, S. Liu, and Z. Zhang, "An advanced DV-hop algorithm," in *Proceedings of the 2017 1st Int. Conf. on Elec. Inst. & Info. Sys. (EIIS)*, pp. 1–5, IEEE, Harbin, China, June 2017.

[7] A. R. Kulaib, R. M. Shubair, and M. A. Al-Qutayri, "Improved DV-hop localization using node repositioning and clustering," in *Proceedings of the Int. Conf. on Comm. Signal Proc. and App*, pp. 1–6, IEEE, Sharjah, United Arab Emirates, February 2015.

[8] B. Mareschal, M. Kaur, V. Kharat, and S. S. Sakhare, "Convergence of smart technologies for digital transformation," *Tehnički glasnik*, vol. 15, pp. 1–IV, 2021.

[9] Y. Chen, X. Li, Y. Ding, J. Xu, and Z. Liu, "An improved DV-hop localization algorithm for wireless sensor networks," in *Proceedings of the 2018 13th IEEE Conf. on Ind. Elec. and App. (ICIEA)*, pp. 1831-1836, *IEEE, Wuhan, China*, June 2018.

[10] H. Xiao, H. Zhang, Z. Wang, and T. A. Gulliver, "An RSSI based DV-hop algorithm for wireless sensor networks," in *Proceedings of the 2017 IEEE Pacific Rim Conf. on Comm., Comp. and Sig. Proc.(PACRIM)*, pp. 1–6, Shanghai, China, September 2007.

[11] Y. Wang and Z. Wang, "Improved DV-hop localization algorithm based on dynamic anchor node set for wireless sensor networks," *IEEE Access*, vol. 7, pp. 124876–124890, 2019.

[12] L. Jing and L. Zhang, "Improved DV-Hop localization algorithm based on hop distance optimization," *Computer Times*, vol. 36, no. 4, pp. 945–951, 2016.

[13] M. Mehrabi, H. Taghdiri, and P. Taghdiri, "An improved DV-Hop localization algorithm based on evolutionary algorithms," *Telecommunication Systems*, vol. 64, no. 4, pp. 639–647, 2017.

[14] H. Dang and Z. Li, "An improved DV-Hop location algorithm for wireless sensor network," *Inst. Tech. and Sensor*, vol. 1, pp. 159–163, 2017.

[15] S. Xia, J. Zou, and X. Zhu, "An improvement on DV-Hop localization algorithm in wireless sensor networks," *Journal of Computer Applications*, vol. 35, no. 2, pp. 340–344, 2015.

[16] F. Zeng, W. Li, and X. Guo, "An improved DV-hop localization algorithm based on average hop and node distance optimization," in *Proceedings of the 2018 2nd IEEE Adv. Info. Mgmt. Comm., Elec. and Auto. Control Conf. (IMCEC)*, pp. 1336–1339, IEEE, Xi'an, China, May 2018.

[17] X. Li and X. Guo, "DV-Hop localization algorithm based on best mean hop distances and improved particle swarm," *Application Research of Computers*, vol. 1, no. 34, p. 12, 2017.

[18] A. Jadhav, M. Kaur, and F. Akter, "Evolution of software development effort and cost estimation techniques: five decades study using automated text mining approach," *Mathematical Problems in Engineering*, vol. 2022, Article ID 5782587, 17 pages, 2022.

[19] W. Kaur and M. Kaur, "A novel QACS automatic extraction algorithm for extracting information in blockchain-based systems," *IETE Journal of Research*, vol. 1, p. 13, 2022.

[20] Y. Sun, M. Cao, Y. Gao et al., "Uncertain data stream algorithm based on clustering RBF neural network," *Microprocessors and Microsystems*, vol. 81, no. 5, p. 103731, 2021.

[21] Y. Sun, Z. Song, and J. Zhao, "Research and improvement of particle swarm optimization in financial risk model," *Journal of Jilin University (Engineering and Technology Edition)*, vol. 38, no. 02, pp. 85–91, 2020.

[22] D. Xue, "Research of location algorithm of wireless sensor networks based on DV-HOP," *EURASIP Journal on Wireless Communications and Networking*, vol. 1, p. 2019, 2019.

[23] M. Kaur, A. Jadhav, and F. Akter, "Resource selection from edge-cloud for IIoT and blockchain-based applications in industry 4.0/5.0", security and communication networks," *Hindawi*, vol. 2022, Article ID 9314052, 2022.

[24] M. Kaur, S. Kadam, and N. Hannoon, "Multi-level parallel scheduling of dependent-tasks using graph-partitioning and hybrid approaches over edge-cloud," *Soft Computing*, vol. 26, no. 11, pp. 5347–5362, 2022.

[25] S. Qu, L. Xiong, and Z. Xiong, "Cross-layer congestion control of wireless sensor networks based on fuzzy sliding mode control," *Neural Computing & Applications*, vol. 32, no. 17, pp. 13505–13520, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 5298468, 7 pages
https://doi.org/10.1155/2022/5298468

Research Article

Fuzzy Logic and Machine Learning-Enabled Recommendation System to Predict Suitable Academic Program for Students

Tribhuwan Kumar,[1] **K. Sakthidasan Sankaran**,[2] **Mahyudin Ritonga**,[3] **Shazia Asif**,[4] **C. Sathiya Kumar**,[5] **Shoaib Mohammad**,[6] **Sudhakar Sengan**,[7] **and Evans Asenso**[8]

[1]*College of Science and Humanities at Sulail, Prince Sattam Bin Abdulaziz University, Al Kharj 11942, Saudi Arabia*
[2]*Department of ECE, Hindustan Institute of Technology and Science, Chennai, India*
[3]*Universitas Muhammadiyah Sumatera Barat, Padang, Indonesia*
[4]*Higher College, Sharjah, UAE*
[5]*Department of Computational Intelligence, School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, India*
[6]*School of Law, IMS Unison University, Dehradun, India*
[7]*Department of Computer Science and Engineering, PSN College of Engineering and Technology (Autonomous), Tirunelveli 627152, Tamil Nadu, India*
[8]*Department of Agricultural Engineering, University of Ghana, Accra, Ghana*

Correspondence should be addressed to Evans Asenso; easenso@ug.edu.gh

Received 10 May 2022; Accepted 22 July 2022; Published 11 August 2022

Academic Editor: Mukesh Soni

In recent years, educational data mining has gained a considerable lot of interest as a consequence of the large number of pedagogical content that can be gathered from a range of sources. This is because there is a lot of instructional information that can be obtained. The data mining tools collaborate with academics to improve students' learning strategies by analyzing, sifting through, and estimating components that are pertinent to students' characteristics or patterns of behavior. This is accomplished through the following steps: EDM is utilized in the vast majority of instances to develop the classification model, which then assigns a certain class to each student based on the known properties of the training dataset. Before putting the classification model into use, it is possible to utilize a test dataset to verify that the model is accurate. This article provides a description of a recommendation system that determines the most beneficial academic program for students by utilizing fuzzy logic and machine learning. The compilation of a student dataset has begun. It includes a total of 21 features and 1000 individual cases. The initial step is to employ the CFS attribute selection method. This methodology selects 15 of the initial set of 21 characteristics. Following the completion of the data gathering, it is put through various machine learning methods such fuzzy SVM, random forest, and C4.5. This methodology that has been offered makes predictions about the academic program that is best suitable for students.

1. Introduction

In recent years, EDM (educational data mining) has attracted a significant lot of interest as a consequence of the enormous amount of instructional material that can be obtained from a variety of sources. Primary purpose of EDM is to increase the efficacy of data mining models in order to secure and protect the large amount of educational data collected and to establish a safe learning environment for students. In this technique, many models for data management and analytics have been employed [1]. A number of prediction methods, including classification, regression, and latent component analysis, were also used to make the predictions.

The data mining tools work in collaboration with academics to enhance students' learning techniques by assessing, filtering, and estimating aspects that are relevant to students' traits or behavioral patterns [2]. Increase the

number of students who are placed, and the number of graduates who are hired is one of the most difficult goals for any educational institution to accomplish. EDM techniques such as clustering and visualization are displayed in Figure 1 as examples of popular EDM approaches.

The building of an ML (machine learning) model, which is a computationally demanding process, makes use of resampling and iterative classification algorithms. Through the use of machine learning approaches that combine optimal subset selection, traditional classifier flaws such as over-fitting and distributional demands on parameters may be removed. The original machine learning (ML) approaches in computer science were based on statistics rather than predicting group characteristics; instead, they started with an arbitrary group separator and tweaked it repeatedly until it met the requirements of the classification groups in question [3]. When the ML functions become unstable, it is a good idea to review the tuning variables as well as the individual ML functions.

In addition, because of the nonstatistical character of these approaches, the data may be presented in a number of formats, such as nominal data, in order to achieve the maximum possible classification precision. It is defined as a process that can be quantified and measured. With the help of machine learning, we are able to solve the classification issue in the context of collective characteristics [4]. These applications have benefited from the advancements in pattern analysis and machine learning (ML) during the last several decades, which have broadened the range of features that may be employed in these applications.

To minimize the amount of redundant characteristics, which are referred to as "overload" and "difficult processes," several ways have been implemented in order to reduce complexity. It is critical to pick characteristics that are most relevant to the issue at hand in order to learn more about it while also reducing the amount of processing necessary. In this work, we look at models that employ specific values to define subsets of features in order to enhance overall prediction efficiency. Clusters are collections of data elements that have been grouped together. To put it another way, a cluster is a collection of items that are all the same inside the cluster but are not related to any other clusters at all.

Clustering is one of the most essential UL statistical processes, and it is described here [5]. This preprocessing approach reduces the quantity of the data for useful clusters, which may subsequently be used in further in-depth studies in the DM model when they have been identified. Because the data are saved in a cluster format, which is a lossy compression approach, the file size is reduced as a result of the compression. Clustering is a challenging idea to categories since they overlap so much. Traditionally, clustering models such as hierarchical and partitional clustering have been divided into two groups. Understanding the subtle distinctions between clustering and supervised classification, on the other hand, is crucial. In supervised classification, prelabeled data patterns are utilized to make decisions about the data. In this procedure, unlabeled datasets are identified and clustered, with the objective being to estimate the labeling of those datasets while the clustering is taking place.

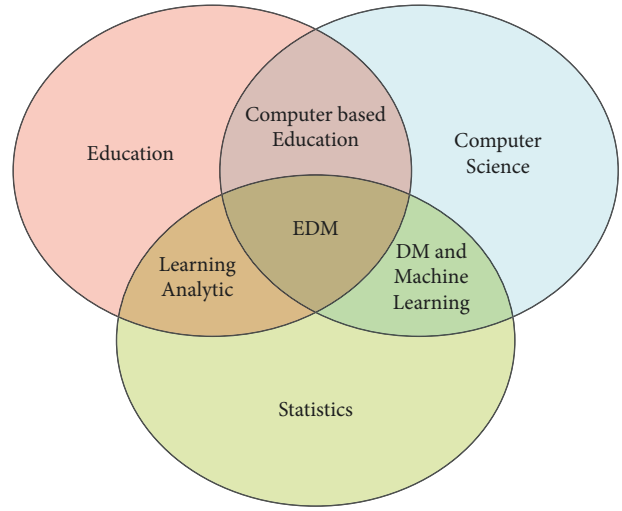


Figure 1: Educational data mining.

When it comes to the design of clusters that are used to develop a clustering solution, it is common to distinguish between partitional and hierarchical techniques. In addition, there is a distinction between "hard" and "soft" models, which is related to the way the items in the dataset are mapped into clusters and how the models are trained.

Data classification frequently necessitates the use of two sophisticated methods. This step of the learning process begins with an examination of a collection of training dataset samples to establish the classes that will be used. NN, rule-based methods, and data-driven techniques are examples of methodology. The technique for data classification is illustrated. Every instance is considered to be located in the previous class, which is defined as follows: it is sampled in the second phase, and this is done with the aid of an additional dataset that has been created to test the classification accuracy of a method. Once a suitable level of accuracy has been attained, this approach is used to categorize future data instances by assigning them a class label. Classification is, without a doubt, a critical component of the decision-making process. One of the classification ideas that is employed is the Bayesian technique [6, 7].

In the majority of cases, EDM is employed in the following areas: in order to create the classification model, assign a specific class to each student based on the known attributes of the training dataset. A test dataset may be used to ensure that the classification model is valid before deploying it. It is possible to use the categorization model to assign a new student to a certain class. It is possible to anticipate a student's academic progress using data from either online or offline sources. Content recommendations, course registration, and academic program enrolment for students are all included in this category.

Literature survey contains survey of various methods for educational data mining. Methodology section presents a fuzzy logic and machine learning-enabled recommendation system to predict suitable academic program for students. A student dataset is prepared. It is having 1000 instances and 21 attributes. First, CFS attribute selection algorithm is

applied. This algorithm selects 15 attributes out of original 21 attributes. Then, machine learning algorithms, namely, fuzzy SVM, random forest, and C4.5, are applied on the dataset. This proposed model predicts suitable academic program for students. Result section presents details related to dataset and accuracy achieved by the prediction model.

2. Literature Survey

Educational institutions may be able to solve these challenges through the use of EDM and learning abilities, which allow them to precisely anticipate the number of students who will graduate and be put in positions. DM, according to research conducted by Ji et al. [8], assists educators in the transformation of every model into data, such as learning objectives, learning actions, learning prioritization, participation and competition, functions, and accomplishment tendencies in a variety of learning events, according to the findings of their study. Recently, data mining and machine learning models in educational data have been employed to solve the issues raised above.

Two of the most widely used machine learning (ML) algorithms, known as predictive and descriptive analytics, are unsupervised learning (UL) and supervised learning (classification), which are two of the most popular ways to machine learning (ML) algorithms [9].

Both of these considerations are at the center of our study [10]. Students' performance in two disciplines, such as a Bachelor of Business Administration degree, should be evaluated in Finland and Spain. Then, examine what effects their performance in other areas, such as their motivation or priorities, should be evaluated in both countries. The primary objective of this study is to evaluate the effectiveness of a specific latent class model, Bayesian profile regression, in identifying students who are more likely to fail their classes. It is feasible to build student profiles that are associated with the highest levels of academic risk based on the performance, motivation, and resiliency of the kids. The data for this study are gathered via online questions submitted by undergraduate students at an Italian institution. The data are collected in real time and utilized to build a sample of data for the study.

Education at the next level is developed and implemented through the use of EDM and learning analytics [11]. For the goal of creating learning approaches, it provides a systematic paradigm for collecting, calculating, reporting, and operating on digitalized data. By incorporating EDM and LA into the educational setting, instructors may come up with innovative solutions to the problem of interaction. It is a virtual educational platform that serves to bridge the gap that exists between instructors and learners. LMS is an abbreviation for learning management system. With the aid of this technology, students and teachers may communicate more effectively with one another. It enables both instructors and students to share knowledge and answer issues and concerns in a safe environment.

DM and data analytics were the focus of Ray et al. [12] research in the educational business, where they examined how to handle the data collected. EDM and LA approaches may be used to handle enormous amounts of data by both commercial and noncommercial organizations. EDM and LA also give a complete analysis of how the role of shareholders in PG-level educational institutions is influenced by their findings. A brief description of how these models may be implemented, as well as how students' learning processes can be analyzed, and how they can be utilized to provide complete feedback, is also given in this part. These models finally have an impact on the administrative ideas that are acceptable for all stakeholders in the educational process.

The NB classification has been used to develop a model for predicting student dropout, which has just been published by Hegde et al. [13] in the R programming language. The next stage will be to take a more in-depth look at why students fail or succeed in their first year, and whether or not they will be dropped. As previously noted, there are a variety of factors that might lead to a kid being dismissed from school without warning. The capacity to predict whether or not a kid would drop out of school is tremendously valuable to company owners.

When dealing with large student databases, a number of data preparation techniques have been employed in order to generate the students' marks in accordance with the assessment modules. Student grades have been fine-tuned in the data preparation stage before being utilized to extract the categorical component from the data. As a result, there are no unique grades for each of the courses that have been recorded.

Following that, an investigation of the preprocessing procedures of EDM data was conducted. It is generally acknowledged that educational information should not be saved in the same way as other types of data since there are so many variables, such as diverse data sources, applications, and human error. As a result, the coursework estimation ratio has been utilized to investigate alternative module assessment approaches while simultaneously producing transcription data for students enrolled in the course. The coursework assessment ratios have been demonstrated to have an effect on classifiers that use radio frequency technology (CARs).

The academic performance of a university's students is typically used to determine the excellence of the institution. When it comes to EDM applications, popular and successful one is the prediction of students' grade point averages and educational achievement, which is classified in the EDM forecast as "excellent," "very good," "good," "moderate," and so on and so forth. This form of prediction may be used to identify the most deserving candidates for scholarship grants at a number of institutions. For the undergraduate level, subsequent elements such as grade point average (GPA) and academic efficiency have been researched in the literature. According to the researchers, about 300 pupils were utilized to predict the final grades of students in the faculty of computer systems and software engineering [14]. Using multivariate analytic models, the significance of a feature has been sampled to determine its relative importance. According to experts, a student's ability to succeed may be predicted by their family's support. Furthermore, the outcome is unaffected by the pupils' level of interest.

Asshraf et al. [15] conducted an investigation of the performance of 210 undergraduate students by analyzing their data. The qualities are used to make predictions about students' grades. Finally, the research demonstrates that it is possible to predict a student's graduation performance in a final semester of university using their preuniversity grades as well as the marks from their first- and second-year courses with greater accuracy using their preuniversity grades and the marks from their first- and second-year courses.

Pradeep et al. [16] predicted the dropout rate for bachelor students using a sample of students enrolled in the technology program. When using the Weka tool, the attribute selection algorithms help to limit the influence of the characteristics that are employed. Postenrolment characteristics like attendance, paying attention in class, and grades received are usually stated as determining factors in choosing a university. For reliable prediction of academic accomplishment, it is not required to include other information such as age, gender, or religion in the questionnaire.

The goal of this study is to determine whether or not hyper-parameter optimization can be employed effectively in educational environments, which is the subject of this study [17]. When it comes to forecasting student, achievement based on the role performed by the online learning environment, automatic machine learning has shown to be particularly successful. In order to obtain visible and intelligible results in the search space at the same time, both rule-based and tree-based approaches are employed. The final point to mention is that an impressive number of findings indicate the ability of auto ML systems to achieve extraordinary results. It is demonstrated that the DM model may be utilized to analyze educational data in the field of education [18]. When forecasting the number of students who will drop out of college or university, a researcher uses a classifier that is specifically designed for this purpose. To assess whether or not a student falls into the dropout group, it is necessary to have detailed information on their curriculum. Students who have enrolled in university throughout the preceding few decades have been able to access authentic academic data about themselves. When dealing with irregular data, it is necessary to preprocess the raw data.

The usage of decision support systems (DSS) is a vital element in the EDM process (DSS). The goal of an application is to make the process of making a decision easier. For example, giving comments and sending alerts, planning events, making ideas, and improving course materials are all instances of this sort of work. Although DSS is primarily intended for use by teachers, it is not capable of being used successfully by students, managers, or researchers.

New models of decision-making have been proposed in recent studies, including the following. Participants, on the other hand, are the primary emphasis of these recommendations. Instructors, for example, could provide courses to students. RS methods such as associative/content-based filtering, association rules, and hybrid systems are common examples of RS techniques. These have been widely employed in the field of business data management (EDM). Suggestions are also provided via the use of discovery techniques. To give an example, Vialardi-Sacón et al. [19] employed a performance forecasting tool to generate ideas for their clients. This model is used to forecast a student's performance in each class and to provide course suggestions based on that performance estimate.

The fundamental purpose of this study's research is to install a classifier based on locally produced student characteristics as soon as possible [20]. After that, we employ preprocessed data to determine student characteristics, which are subsequently used in the FS as well as in future learning and evaluation via Weka, among other things. It has been possible to construct predictor tools by utilizing the NB classification model, which is extremely accurate. It is consequently possible to utilize the tool as a sampling tool since it may anticipate students' roles in following studies based on its qualities.

3. Methodology

This section presents a fuzzy logic and machine learning-enabled recommendation system to predict suitable academic program for students as shown in Figure 2. A student dataset is prepared. It is having 1000 instances and 21 attributes. First, CFS attribute selection algorithm is applied. This algorithm selects 15 attributes out of original 21 attributes. Then, machine learning algorithms, namely, fuzzy SVM, random forest, and C4.5, are applied on the dataset. This proposed model predicts suitable academic program for students.

When determining the value of a subset of characteristics, CFS takes into account both the unique ability of each characteristic and the degree of duplication that occurs between the characteristics. Characteristics with a high relationship to the class, but a low correlation with the other criteria, are considered for selection [21].

The categorizing approach may be used in a supervised or especially unsupervised setting, depending on the situation. This is a well-known truth in the business community. This is the source of the fact that support vector networks are supervised learning standards. An SVM may be used to build feature points or attribute states by projecting them onto nonlinear hyperplanes and planar projections. The performance of SVMs is significantly influenced by the use of Gaussian kernels, data variance and standard deviation, and kernel selection procedures, among other factors. When using fuzzy SVM, each training point corresponds to a single class with pinpoint accuracy. When dealing with data that are stochastic or probabilistic in nature, it is necessary to collect prelearning data on the datasets themselves. In this section, we will take a look at a number of different stochastic connections [22].

Using random forests for classification and regression problems may prove to be beneficial in the near future. Regression methods are used in order to anticipate the outcomes of each decision tree that is generated during the process of training the decision trees. It has a low standard deviation when used to create predictions and is capable of quickly combining numerous bits of information. Because of the cryptic nature of random forest categorization, the general public was first skeptical of its efficacy. On the contrary, it has shown superior performance in a prediction-based challenge [23].

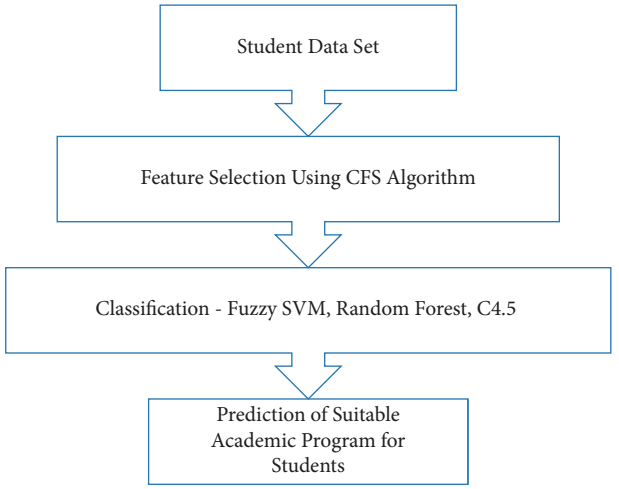


Figure 2: Fuzzy logic and machine learning-enabled recommendation system to predict suitable academic program for students.

A decision tree is a kind of classification algorithm that is often employed in machine learning models because of the speed and accuracy they provide. Tree trimming may be accomplished in a number of ways with the help of this method, which is quite adaptable. Following the process of pruning, only a small number of discoveries are produced that are easily understandable. Some scientists feel that overfitting might be done by the removal of trees. C4.5 method iterative classification continues until the data are classified as accurately as possible by creating pure leaf nodes, at which point the operation is completed. C4.5 method iterative classification allows for the most accurate possible results to be obtained from training data without the need for extraneous rules that merely identify a certain behavior [24].

4. Results and Discussion

A dataset of 1000 students has been created with 21 attributes like age, sex, university name, city, country, father's education, mother's education, family size, health status, travel time, family relationship, study hours, extra activities, hobbies, alcohol consumption, course name, course description, skills, difficulty level, results, and attendance.

First features are selected using correlation feature selection (CFS) algorithm. Then, these 15 attributes are selected, age, father's education, mother's education, health status, travel time, study hours, extra activities, hobbies, alcohol consumption, course name, course description, skills, difficulty level, results, and attendance.

For performance comparison, three parameters, accuracy, sensitivity, and specificity, are used.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{TN} + \text{FP} + \text{FN})},$$

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP} + \text{FN})}, \quad (1)$$

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN} + \text{FP})},$$

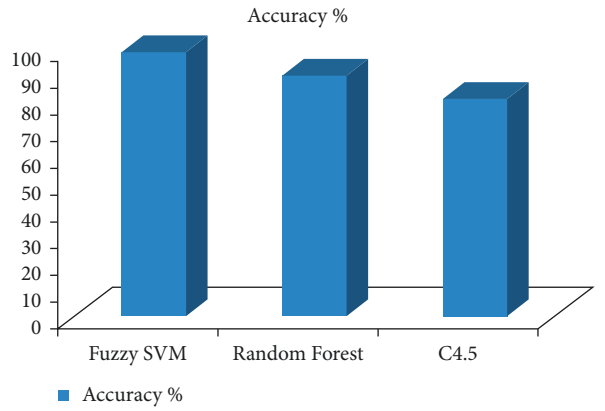


Figure 3: Accuracy of machine learning techniques for recommendation system to predict suitable academic program for students.

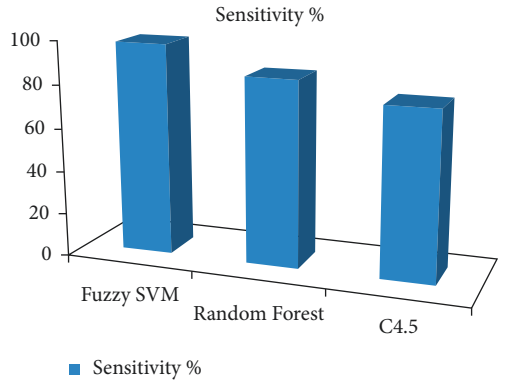


Figure 4: Sensitivity of machine learning techniques for recommendation system to predict suitable academic program for students.

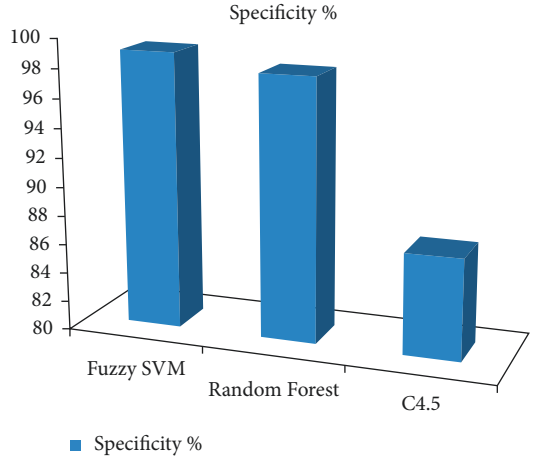


Figure 5: Specificity of machine learning techniques for recommendation system to predict suitable academic program for students.

where TP = true positive, TN = true negative, FP = false positive, and FN = false negative.

Results of different machine learning predictors are shown in Figures 3–5. Accuracy of fuzzy SVM is better than random forest and C4.5 algorithm.

5. Conclusion

As a result of the large quantity of instructional content that can be collected from many sources, educational data mining has garnered a great deal of attention over the last few years. Student learning approaches are improved via the use of data mining technologies, which work in partnership with academics to identify features that are significant to students' qualities or behavioral patterns and then analyze, filter, and estimate those aspects. It is most often used to develop the classification model, which then assigns a particular class to each student on the basis of known qualities from a training dataset. Before implementing a classification model, it is possible to test it on a test dataset to check that it is valid. It is described in this study how a recommendation system that employs fuzzy logic and machine learning may be used to find the optimum academic program to be followed by students. A student data collection is currently being compiled. It is made up of 1000 occurrences and 21 attributes, in total. After that, the CFS attribute selection process is used. This approach selects 15 characteristics from the original 21 characteristics. Following data gathering, machine learning algorithms such as fuzzy SVM, random forest, and C4.5 are used to the data. This proposed model predicts the best suited academic program for students based on their individual needs.

Data Availability

The data shall be made available on request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] R. S. Baker and P. S. Inventado, *Educational Data Mining and Learning Analytics*Learning analytics Springer, New York, NY, 2014.

[2] B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance," *International Journal of Advanced Computer Science and Applications*, vol. 2, no. 6, pp. 63–69, 2012.

[3] V. Hemamalini, S. Rajarajeswari, S. Nachiyappan et al., "Food quality inspection and grading using efficient image segmentation and machine learning-based system," *Journal of Food Quality*, vol. 2022, pp. 1–6, 2022.

[4] A. Gupta and L. K. Awasthi, "Peer-to-peer networks and computation: current trends and future perspectives," *Computing and Informatics*, vol. 30, no. 3, pp. 559–594, 2011, http://www.cai2.sk/ojs/index.php/cai/article/view/184.

[5] M. Shabaz and U. Garg, "Predicting future diseases based on existing health status using link prediction," *World Journal of Engineering*, vol. 19, no. 1, pp. 29–32, ahead-of-print (Issue ahead-of-print). Emerald, 2021.

[6] M. M. Ezz, "Advisory system for student enrollment in university based on variety of machine learning algorithms," *International Journal of Computing Academic Research (IJCAR)*, vol. 4, no. 2, pp. 34–45, 2015.

[7] M. Shabaz and U. Garg, "Shabaz–urvashi link prediction (sulp): a novel approach to predict future friends in a social network," in *Journal of Creative Communications*vol. 16, no. Issue 1, pp. 27–44, SAGE Publications, 2020.

[8] H. Ji, K. Park, J. Jo, and H. Lim, "Mining students activities from a computer supported collaborative learning system based on peer to peer network," *Peer-to-Peer Networking and Applications*, vol. 9, no. 3, pp. 465–476, 2016.

[9] C. Wu, P. Lu, F. Xu, J. Duan, X. Hua, and M. Shabaz, "The prediction models of anaphylactic disease," *Informatics in Medicine Unlocked*, vol. 24, p. 100535, 2021.

[10] M. d. M. Camacho-Miñano, C. del Campo, E. Urquía-Grande, D. Pascual-Ezama, M. Akpinar, and C. Rivero, "Solving the mystery about the factors conditioning higher education students' assessment: Finland versus Spain," *Education + Training*, vol. 62, no. 6, pp. 617–630, 2020.

[11] A. Van Barneveld, K. E. Arnold, and J. P. Campbell, "Analytics in higher education: establishing a common language," *EDUCAUSE learning initiative*, vol. 1, no. 1, 2012.

[12] S. Ray and M. Saeed, *Applications of educational data mining and learning analytics tools in handling big data in higher education. International Conference on Applications of Big Data Analytics*, , pp. 135–160, geeksforgeek, 2018.

[13] V. Hegde and P. P. Prageeth, "Higher education student dropout prediction and analysis through educational data mining," *International Conference on Inventive Systems and Control (ICISC)*, pp. 694–699, 2018.

[14] S. Sembiring, M. Zarlis, D. Hartama, S. Ramliana, and E. Wani, "Prediction of Student Academic Performance by an Application of Data Mining Techniques," *International Conference on Management and Artificial Intelligence*, pp. 110–114, 2011.

[15] M. Ashraf, M. Zaman, and M. Ahmed, "An intelligent prediction system forEducational data mining based on ensemble and filtering approaches," *Procedia Computer Science*, vol. 167, pp. 1471–1483, 2020.

[16] A. Pradeep and J. Thomas, "Predicting college students dropout using EDM techniques," *International Journal of Computer Application*, vol. 123, no. 5, pp. 26–34, 2015.

[17] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Implementing AutoML in educational data mining for prediction tasks," *Applied Sciences*, vol. 10, no. 1, pp. 90–117, 2019.

[18] M. Utari, B. Warsito, and R. Kusumaningrum, "Implementation of Data Mining for Drop-Out Prediction Using Random Forest Method," in *Proceedings of the International Conferenceon Information and Communication Technology*, pp. 1–5, Yogyakarta, Indonesia, June2020.

[19] C. Vialardi-Sacín, L. Shaftr, J. Braver, and A. Ortigosa, "Recommendation in higher education using data mining techniques," *Journal of Educational Data Mining*, pp. 190–199, 2009.

[20] A. A. Rimi, A. A. Ibrahim, and O. Bayat, "Developing classifier forthe prediction of students' performance using data mining ClassificationTechniques," *AURUM Mühendislik Sistemleri ve Mimarlik Dergisi*, vol. 4, no. 1, pp. 73–91, 2020.

[21] M. Aggarwal, "Performance analysis of different feature selection methods in intrusion detection," *International Journal of Scientific & Technology Research*, vol. 2, pp. 225–231, 2013.

[22] M. Nilashi, H. Ahmadi, A. A. Manaf et al., "Coronary heart disease diagnosis through self-organizing map and fuzzy support vector machine with incremental updates," *International Journal of Fuzzy Systems*, vol. 22, no. 4, pp. 1376–1388, 2020.

[23] S. Jayaprakash, S. Krishnan, and V. Jaiganesh, "Predicting students academic performance using an improved random forest classifier," in *Procedings of the 2020 International Conference on Emerging Smart Computing and Informatics*, pp. 238–243, Pune, India, March 2020.

[24] L. I. P. Aji and A. Sunyoto, "An implementation of C4.5 classification algorithm to analyze student's performance," in *Proceedings of the 2020 3rd International Conference on Information and Communications Technology*, pp. 126–130, Yogyakarta, Indonesia, November 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 2754965, 8 pages
https://doi.org/10.1155/2022/2754965

Research Article

Monitoring and Prediction of Highway Foundation Settlement Based on Particle Swarm Optimization and Support Vector Machine

Rui Yang and ShengLi Yuan

Department of Civil Engineering and Architecture, Xinxiang University, Xinxiang 453000, China

Correspondence should be addressed to Rui Yang; yangrui01@xxu.edu.cn

Received 24 May 2022; Accepted 21 July 2022; Published 10 August 2022

Academic Editor: Mukesh Soni

Highway construction has always been an important strategy in China's construction projects. However, because the soil in the construction area belongs to the soft soil zone, there will often be large vertical deformation in the construction process, which will seriously affect the engineering quality, so the highway FS (foundation settlement) prediction is particularly important. In order to improve the accuracy of highway stability prediction and ensure the safety of highway engineering, a prediction model based on PSO_SVM (support vector machine for particle swarm optimization) is proposed. By using the particle velocity and its position in the PSO algorithm to correspond to the kernel function parameters and penalty factors of the parameters in the model, the optimal parameters are found and substituted into the SVM prediction model to obtain the PSO_SVM. The results show that the MAD of section A# and section B # of PSO_SVM is 0.8991 and 1.3027 for different monitoring points. *Conclusion*. PSO_SVM has a strong learning and generalization ability, high prediction accuracy, stability, and adaptability, and can reflect the overall change information of highway FS data, which has practical application value.

1. Introduction

In the process of subgrade excavation of the project, excavators and bulldozers are used together, and the construction is carried out in different layers. In case of any change in the geological conditions of the soil layer during the construction, it shall be reported to the supervising engineer in time to facilitate the timely adjustment of the slope ratio. During mechanical excavation, attention shall be paid to the protection of underground pipelines, cables, and other structures. When the excavation is close to the slope, the construction contractor shall stop using manual excavation to properly excavate and trim the slope. Where the subgrade and slope are connected, earthwork shall be reserved in advance, which is conducive to the stability of the slope in the later stage. In the process of mechanical excavation, the design requirements shall be strictly followed to ensure that the excavation depth meets the design requirements and avoid over excavation. In addition, backfill materials shall be used in a timely manner for backfilling and compaction to ensure that the subgrade bearing capacity meets the requirements. With the increase of expressway construction scale, the problem of soft soil is becoming more and more serious, which has become one of the important factors affecting the stability of highway opening in operation period [1]. In the process of expressway construction with soft soil subgrade, in order to ensure the construction quality of expressway subgrade, the stability of pavement during the postconstruction operation period, and the safety of vehicles, it is necessary to predict the final settlement of soft soil subgrade, so as to provide reference for determining the subgrade filling scheme. In a period after construction, monitoring and data prediction of soft soil FS (foundation settlement) can improve the stability and bearing capacity of this section of highway subgrade, so as to reduce the occurrence of accidents.

Wang et al. [2] apply the numerical method to FS prediction; Li et al. [3] apply grey theory to FS prediction; Li et al. [4, 5] put forward several methods to calculate the settlement of sand drain foundation. The results show that the settlement of sand drain foundation caused by lateral deformation must be paid attention to. In reference [6], the grey theory is used to predict FS in unequal distance, and the corresponding GM (1, 1) model is established. Kim et al. [7] propose to optimize the parabolic settlement prediction model by using real code accelerated genetic algorithm. Zhou et al. [8] put forward a method to predict FS by using the momentum BP algorithm. Hu et al. [9] used the traditional three-layer BP network model to study this problem, and achieved satisfactory results. Liu et al. [10] apply wavelet time series model to subway settlement prediction, and prove that this model has the characteristics of gross error detection and robust interference. Chen [11] combines wavelet denoising and grey-time series model to predict the surface of subway lines, and good prediction results have been achieved.

With the passage of time, the ground buildings will have different degrees of settlement and deformation. At present, the numerical simulation method, the grey prediction model, and the artificial neural network for predicting the surface subsidence and deformation have their own defects. SVM (support vector machine) has good mathematical properties and shows good generalization ability, but its performance depends on the parameters of the learning algorithm. However, there is no fixed method for selecting SVM parameters, which can only be selected through experimental comparison. Therefore, the parameter determination of SVM has always been a hot issue in research [12, 13]. Therefore, a highway FS prediction method based on PSO_SVM (SVM based on particle swarm optimization) is proposed. Among them, PSO is used to select the optimal combination of SVM training parameters. According to the optimized SVM parameters, a vector machine model is established. Finally, the performance of the model is evaluated, and good results are achieved.

2. Research Method

2.1. On-Site FS Monitoring of Highway. The landform, geological conditions, and hydrological conditions along the roads in mountainous areas are complex and uncertain. The subgrade is filled and excavated frequently, and there are many deep excavation and high filling sections, which make the physical and mechanical properties of the subgrade very different and cause uneven settlement of subgrade and instability. This requires in-depth study of engineering geological conditions to find more targeted survey methods and means. Complete and accurate engineering geological data for the design and construction of mountain highways were provided. Therefore, it is necessary to fully understand the engineering geological conditions of the area before building mountain roads. Whether the roadbed can run stably for a long time is closely related to the geological structure. Highway construction is largely influenced by the strength of soil permeability, soil stress of foundation soil layer, change of groundwater level, and soil thickness. In addition, climatic conditions are also an important factor leading to subgrade settlement. In the process of highway subgrade filling construction, in order to speed up the construction progress, the on-site constructors did not carefully observe the highway subgrade and filling height, and did not fully consider the FS problem, which caused the foundation capacity to fail to meet the standard requirements, resulting in FS and cracks.

The stable observation of soft soil foundation settlement during construction and operation is an important link in highway construction. In the process of embankment filling, the load acting on the pressed rubble composite subgrade will gradually increase, which will lead to consolidation settlement of the composite subgrade. For this kind of situation, we can use the method of burying section pipes on the surface of composite subgrade or the bottom of embankment along the cross section and using inclinometer to test the section settlement deformation of composite subgrade during embankment filling, so as to deal with the settlement problem by controlling the load.

During the widening construction of expressway subgrade, the original load on the old subgrade was changed, resulting in the second settlement deformation of the old subgrade. At the same time, the settlement of newly built subgrade before the end of construction is called the first settlement only under the action of its own weight; the settlement of newly built subgrade is called the second settlement [14].

Due to the difference of physical properties and construction technology of different fills, it is not yet possible to form a unified result. At present, the research on differential settlement of widening subgrade usually adopts the back analysis method, that is, the data obtained by settlement monitoring are analyzed, the empirical formula is obtained, and then the change law of FS with time is inferred, and the calculation formula of main consolidation settlement is determined by analysis, such as the following formula:

$$S_c = \sum_{i=n}^{n} \frac{e_{0i} - e_{1i}}{1 + e_{0i}} \Delta h_i, \tag{1}$$

where n is the number of layers calculated by FS; Δh_i is the thickness of the i-th layer after subgrade stratification, generally 0.5~1.0 m; e_{0i} is the relevant void ratio between subgrade stratification and the dead weight stability of the i-layer subgrade soil; and e_{1i} is the relevant void ratio between subgrade stratification and i-layer subgrade soil after additional load and dead weight stability.

When carrying out road subgrade filling, the filling speed should be strictly controlled, and the measured value of surface settlement should be taken as a reference to accurately judge the FS trend, so as to have a clear understanding of the objects and time range required for preloading and unloading, and thus have an accurate grasp of the construction time of the pavement [15, 16]. According to the measured data, the surface uplift state is analyzed. Finally, the underground horizontal displacement meter should be used to measure the underground soil layers, so as to obtain

the specific displacement, and through the obtained displacement, the damaged position of the soil can be analyzed, so as to ensure that the subgrade will not settle.

For sections with poor geological conditions, the method of improving soil quality can be selected. Construction workers generally choose drainage consolidation method, which can effectively improve the soil hardness of subgrade.

2.2. Overview of PSO and SVM

2.2.1. PSO Principle. The basic principle of PSO is as follows:

Assuming that there are n particles flying at a constant speed in D-dimensional space, population $X = (X_1, X_2, \cdots, X_n)$, and then $X_i = (x_{i1}, x_{i2}, \cdots, x_{iD})$ is the position of the i-th particle and $V_i = (v_{i1}, v_{i2}, \cdots, v_{iD})$ is the velocity of the i-th particle.

According to the number of target pictures, the fitness value corresponding to X_i is calculated, $P_i = (p_{i1}, p_{i2}, \cdots, p_{iD})$ is its individual extreme value, and $P_g = (p_{g1}, p_{g2}, \cdots, p_{gD})$ is the global extreme value of the population.

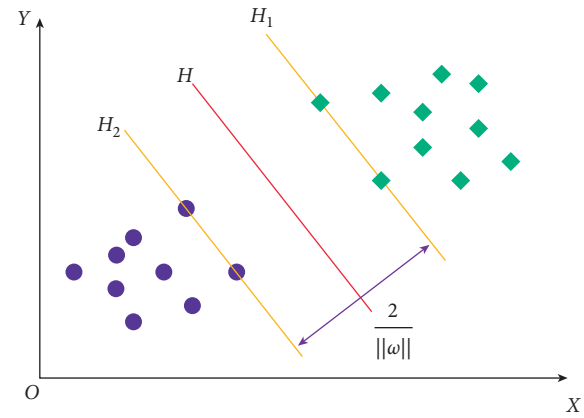


Figure 1: Optimal hyperplane.

The updated formula of the velocity and position of the particle is as follows:

$$\begin{aligned} v_{ij}(t+1) &= wv_{ij}(t) + c_1 r_1\left(p_{ij}(t) - x_{ij}(t)\right) + c_2 r_2\left(p_{gi}(t) - x_{ij}(t)\right), \\ x_{ij}(t+1) &= x_{ij}(t) + v_{ij}(t+1), \end{aligned} \tag{2}$$

where w is the inertia weight, $i = 1, 2, \cdots, n$; $j = 1, 2, \cdots, d$.

2.2.2. SVM Principle. SVM is a new general learning method developed on the basis of statistical learning theory. It is based on the principle of structural risk minimization, and has strong learning ability and generalization ability. It can solve the problems of small sample, nonlinearity, local minima, etc., so as to make effective classification.

Main idea of SVM: this theory adopts the structural risk minimization criterion, maps the input vector to a high-dimensional feature space through preselected nonlinear mapping, and constructs the optimal decision function in this space [17].

Assume a training sample set $(x_i, y_i)_i^N$, in which the input data $x_i \in R^n$ and the output data $y \in R$ construct the optimal linear function in the high-dimensional feature space:

$$f(x) = w^T \varphi(x) + b, \tag{3}$$

where w is the weight and b is the bias term.

The optimal classification surface can not only correctly separate the two types of training samples, but also requires the maximum classification interval (Figure 1).

Support vector machine has been widely used in pattern recognition, regression estimation, probability density estimation, and other neighborhoods. However, the emergence of support vector machines has promoted the rapid development of kernel-based learning methods, which enable researchers to efficiently analyze nonlinear relationships. This high efficiency can only be obtained by the linear algorithm.

Support vectors are sparse, which only accounts for a small part of the training samples. This feature is of great significance for solving large-scale problems. The performance of SVM mainly depends on the model selection. Different kernel functions and parameters can be selected to construct different SVM, and the results and generalization ability are also different [18].

Polynomial kernel function:

$$K(x, y) = (x \cdot y + 1)^d. \tag{4}$$

For a given training sample, the dimension of the system depends on the degree d of the polynomial, so we can control the dimension of the system by choosing the appropriate value d.

Sigmoid kernel function:

$$K(x, y) = \tanh(v(x, y) + a). \tag{5}$$

The S-shaped function adopts the hyperbolic tangent function tanh, which can satisfy the Mercer condition if and only if v, a takes an appropriate value.

Kernel determines the generalization ability of SVM. The essence of SVM is to transform linear inseparable variables into linear separable variables by selecting kernel functions, and then calculating them [19–22]. In the above process, the selection of kernel parameters will also have a great influence. Therefore, we must pay attention to the selection of kernel function and related parameters.

2.3. Establishment of PSO_SVM for Highway Monitoring and Prediction of FS.

FS is one of the main monitoring contents. Because of the complex nature of soft soil, there are too

many assumptions in the research of settlement theory, and the selection of various parameters will also be different from the original soil layer, which is inconsistent with the graded sand loading of road foundation treatment in the process of reclamation, and the accuracy of settlement calculation is not enough, so it is difficult to be widely used in engineering. Engineering experience shows that this method is accurate and reliable, and the form and development trend of curve meet the law of FS quantity development with time.

Under the influence of external load, excess pore water pressure is formed in saturated soil. In this case, the theory can be used for calculation. However, in practice, with the change of time and depth, the consolidation coefficient of soft soil with nonlinear characteristics and high compressibility will change greatly. Obviously, this theory is not applicable to solve it.

Soft soil has the characteristics of high compression, low strength, low permeability, etc. The engineering geological conditions along this high-grade highway change greatly, and the basic characteristics of the foundation are difficult to accurately grasp due to the disturbance of the soil caused by the construction. The deformation is mainly caused by normal stress, which will only make the volume of soil shrink and compact, and will not lead to soil destruction. At this time, the air content in the soil is very small, so the proportion of its compression in the total compression of the soil is not large. Except in some cases, it is necessary to consider the compression of closed gas, which can generally be ignored.

Because of the difference of embankment height and uneven settlement of foundation, uneven deformation of pavement occurs. Uneven settlement of subgrade exceeds a certain limit, which will lead to the functional and structural damage of pavement, and make the highway unable to meet the design requirements. With the increase of filling height, load, and time, the pore water in foundation soil is gradually discharged, the excess pore water pressure is gradually dissipated, and the soil is gradually compacted to produce volume compression deformation, and it enters the elastic-plastic state. At this time, the settlement rate of soil increases rapidly.

According to the statistical theory, SVM determines the regression function by minimizing the following target numbers:

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*),$$
$$s.t.\begin{cases} y_i - w\phi(x) - b \le \varepsilon + \xi_i, \\ w\phi(x) + b - y_i \le \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \ge 0, \end{cases} \tag{6}$$

where ξ_i, ξ_i^* is a non-negative relaxation variable; C is a penalty factor, which is a compromise between empirical risk and model complexity; ε is an insensitive loss function parameter.

Due to the uneven distribution of the initial population, the PSO algorithm is prone to premature phenomenon in the learning process. In order to avoid this shortcoming, chaotic particle swarm optimization based on logistic equation can be used. When the particle falls into the local optimum, the chaotic disturbance is used to jump out of the local optimum. Therefore, the "subjective initiative" of particles is introduced to improve the "passive learning" characteristics of standard PSO, and the "active detection" step of particles is added.

Using the LODISTIC chaotic sequence equation, $T + 100$ random positions are formed in D-dimensional space. The formula is as follows:

$$c(t+1,d) = a(d) \times c(t,d) \times (1 - c(t,d)),$$
$$a(d) = 3.6 + 0.2 \times \text{rand}, \tag{7}$$
$$x(t,d) = l(d) \times c(t+100,d),$$

where t is the current iteration time, $l(d)$ is the length of the d dimension, and *rand* is a random number in [0, 1], $d = 1, 2, \cdots, D$.

Then, the Lagrange method is used to solve the dual optimization problem:

$$L = \frac{1}{2}\omega \times \omega + C\sum_{i=1}^{n}\left(\xi^i + \xi_i^*\right) - \sum_{i=1}^{n}\alpha_i\left[\xi^i + \varepsilon - y_i + f(x_i)\right] - \sum_{i=1}^{n}\alpha_i^*\left[\xi_i^* + \varepsilon - y_i + f(x_i)\right] - \sum_{i=1}^{n}(\xi_i\gamma_i + \xi_i^*\gamma_i^*), \tag{8}$$

in which $\alpha_i, \alpha_i^*, \gamma_i, \gamma_i^* \ge 0$ is Lagrange multiplier and ε is allowable error. L to minimize $\omega, b, \xi_i, \xi_i^*$ is to find the maximum $^i_\alpha, \alpha_i^*, \gamma_i, \gamma_i^* \ge 0$.

The fitness function of each particle is:

$$S(x) = \left(\sum_{i=1}^{N}\frac{y - y_i}{N}\right)^{1/2}, \tag{9}$$

where y_i represents the measured value of the i-th sample; y is the predicted value of the i-th sample; and $i = 1, 2, \cdots, N$ (N is the number of test samples) to calculate the fitness value of each particle.

For each particle, the optimal position of the individual is compared with the optimal position of the group, and if it is compared, it is replaced by the optimal position of the group; otherwise, the optimal position of the group is unchanged.

The estimated $\widehat{f}$ depends on two variables α, β, namely, $\widehat{f} = \widehat{f}(\alpha, \beta)$, where β is the structural parameter and α is other parameter that affects the performance.

Assuming that the training set $\{X, Y\}$ is divided into training set $\{X_t, Y_t\}$ and confirmation set $\{X_v, Y_v\}$, the errors of training and confirmation are defined as formula (7) and formula (8), respectively:

$$R_{train}(f) = \frac{1}{m_{\text{train}}} \sum_{x_i, y_i \in \{x_t, y_t\}} e(x_i, y_i, f(x_i, y_i)),$$
$$R_{\text{valid}}(f) = \frac{1}{m_{\text{valid}}} \sum_{x_i, y_i \in \{x_v, y_v\}} e(x_i, y_i, f(x_i, y_i)). \quad (10)$$

In the above formula, $e(\cdot)$ is the loss function.

Appropriate kernel function parameter g and penalty factor C are the keys to improve generalization ability and classification performance of SVM algorithm. However, PSO has few parameters and is easy to implement.

Therefore, searching for the best g, C through PSO can improve the accuracy of SVM prediction results. The flow of highway stability analysis based on PSO_SVM is shown in Figure 2.

3. Result Analysis

For highway construction projects, soft soil foundation treatment technology is an indispensable part. Soft soil foundation treatment technology not only occupies a very important position in the whole highway engineering construction but its construction quality has a great impact on the final quality of highway construction projects in China. Therefore, it is necessary to avoid the highway engineering construction project quality failing to meet the relevant standards caused by the soft soil foundation treatment technology to a great extent. In order to ensure the safety of people's travel to a certain extent, relevant personnel need to increase the research on soft soil foundation treatment technology in highway engineering construction. In this case, the settlement of expressway is mainly caused by the high load of soft soil subgrade. The soil used to fill the paving foundation has certain compressibility, and under the action of high load, the volume of the roadbed will be reduced, consolidated, and deformed, and then settlement will be formed. In spatial distribution, soft soil has the characteristics of discontinuity and large thickness difference. These characteristics of the soft soil layer determine that the secondary consolidation settlement of the soft foundation section of the expressway is relatively large, and the settlement difference of some sections is obvious, which is directly reflected in the large postconstruction settlement of some sections.

The soft soil in the soft foundation section of the expressway is composed of dark grey or black flowing plastic silt or peat soil, which contains a large amount of humus such as plant roots and stems. Besides the commonness of conventional soft soil, the expressway in the case also has the characteristics of high organic matter. In this chapter, the section settlement data of a highway in a case is used, and a comprehensive analysis is carried out according to the section detection and settlement indicators in the data.

The PSO_SVM is used to predict the FS quantity, which has no strict requirements for the measured data. Three sections (A#, B#, C#) are selected, and the settlement data of these sections are obtained. Based on the PSO_SVM, the prediction method is to write the corresponding program in MATLAB. After obtaining the settlement data of each section, the corresponding improved SVM model can be established, as shown in Figure 3, which shows the related results of settlement prediction curve.

As shown in Figure 3(c), the predicted values in the settlement prediction of three sections of pso\u SVM are almost the same as the actual values. Group C was the most stable. PSO_SVM, PSO, and SVM are used to make mathematical statistics on C# section data, and the required parameters are obtained by fitting. According to the corresponding model, the corresponding parameters are adopted to predict the subsequent settlement. The simulation of the three forecasting methods is shown in Figure 4.

It can be seen that the PSO_SVM is basically consistent with the measured values. Among them, the prediction error of PSO_SVM ranges from 0.2 cm to 1.6 cm. It is not difficult to see that the error of SVM is the largest, the error of PSO is slightly smaller, and the error of PSO_SVM is the smallest. PSO_SVM can simulate the settlement value more accurately, with the smallest error among various models, and can better meet the requirements of engineering accuracy.

The settlement expressions fitted by the above three models are used to predict the settlement of section A# after 60 days, as shown in Table 1.

Obviously, the biggest deviation is PSO, followed by SVM, and the predicted value of SVM is smaller than the measured settlement, which is quite harmful to the actual project. It is impossible to predict whether the settlement is too fast, and PSO_SVM can predict the settlement in the short term.

PSO_SVM can determine the undetermined parameters according to the observed samples, so as to more accurately predict and reflect the postconstruction settlement law of soft soil foundation. PSO_SVM can conveniently predict the postconstruction settlement at any time according to the existing field measured data, so as to judge whether the subgrade deformation is stable or not, and provide decision-making basis for the paving of permanent pavement and the maintenance of pavement.

In order to compare the accuracy between models, the prediction data error of each model and the overall prediction accuracy of the model are summarized together. In this article, MSE (mean square error), RSS (residual sum of squares), and MAD (mean absolute deviation) will be used to evaluate the prediction accuracy of each model one by one. The comparison and summary results of the prediction accuracy of each model of section A# and B# are shown in Figures 5 and 6 and Table 2, respectively.

It can be seen that for different monitoring points, the prediction accuracy of PSO_SVM is higher than that of the traditional SVM model and the PSO model, which shows that PSO_SVM has strong adaptability. The MAD of section A# and section B # of PSO_SVM is 0.8991 and 1.3027, which proves that PSO_SVM has high prediction accuracy and stability. Compared with the other two models, it has a great advantage in prediction accuracy, all indexes are better than the other two models, and the stability of the model is better.

In practice, a particle is often affected by many factors, and all these factors appear in an optimal form, so what

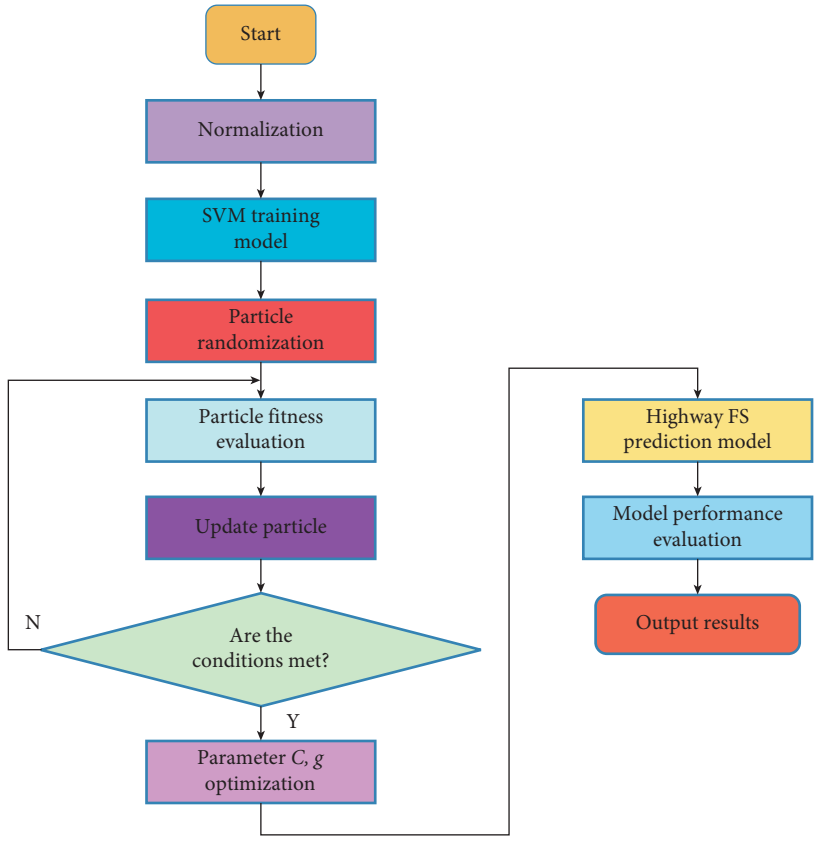


Figure 2: PSO_SVM highway FS prediction process.

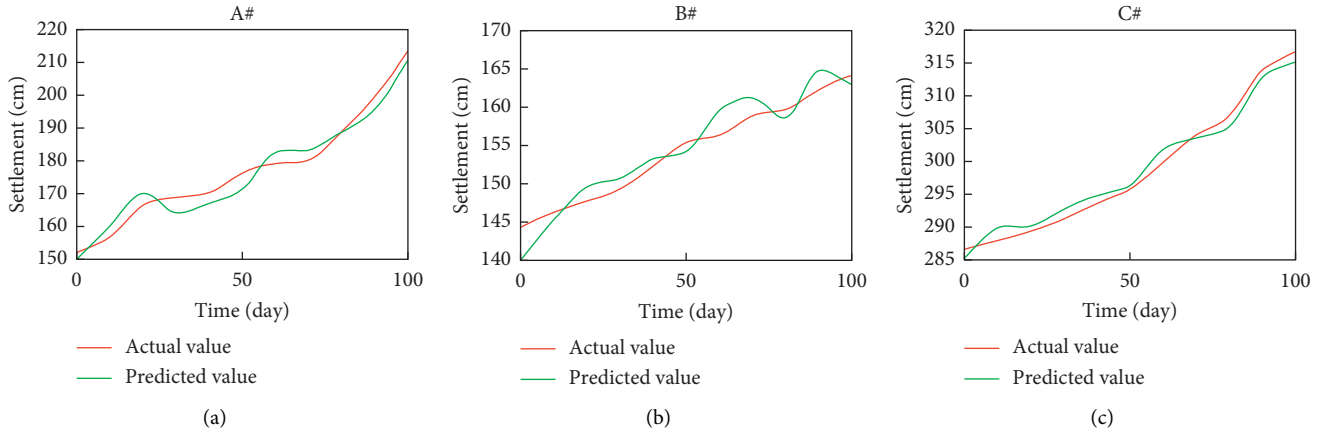


Figure 3: Settlement prediction of three sections by PSO_SVM.

happens is that the predicted value of the particle is less than the actual value. Otherwise, it is greater than the actual value. However, this kind of situation is generally difficult to achieve, and the usual situation is that some factors are in a good state, some are in a bad state, and most of them are in a medium state, so the predicted value is in good agreement with the actual value.

In a word, PSO_SVM has strong learning and generalization ability, high prediction accuracy, stability and adaptability, and can well reflect the overall change

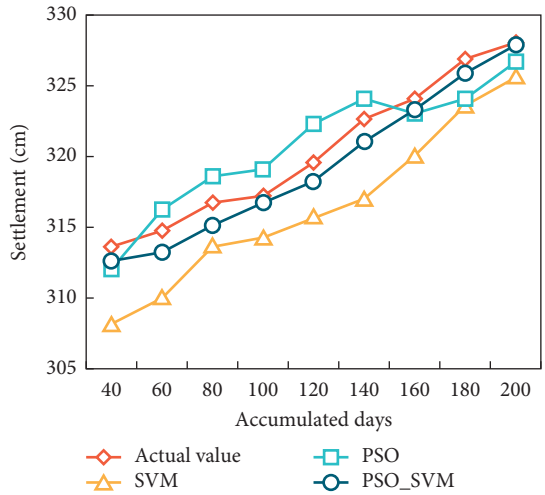


Figure 4: C# section settlement prediction analysis.

Table 1: Comparison of measured settlement values of section A# with predicted settlement values of different models.

Cumulative days	Measured settlement	SVM	PSO	PSO_SVM
260	346.8	340.1	343.8	344.9

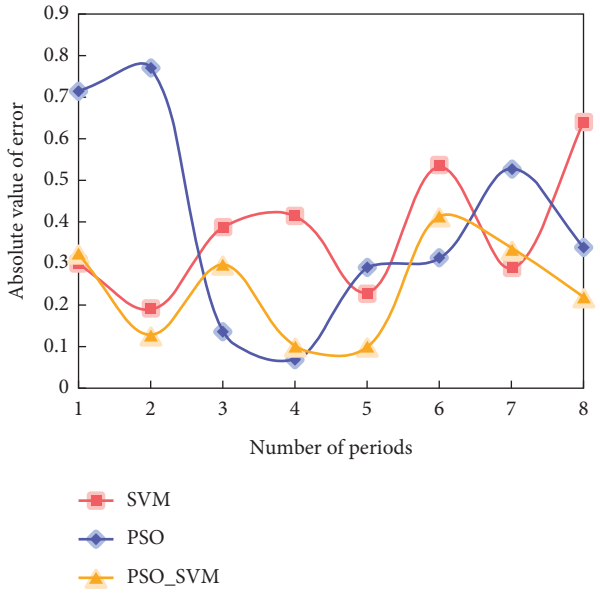


Figure 5: Error of prediction data of various models in section A#.

Table 2: Accuracy evaluation index of each model.

Section	Prediction model	MSE	RSS	MAD
A#	SVM	0.1227	0.8619	3.3284
	PSO	0.0812	0.5512	2.7146
	PSO_SVM	0.0123	0.0864	0.8991
B#	SVM	0.0714	0.4428	1.9681
	PSO	0.2886	1.7416	3.6687
	PSO_SVM	0.0311	0.1893	1.3027

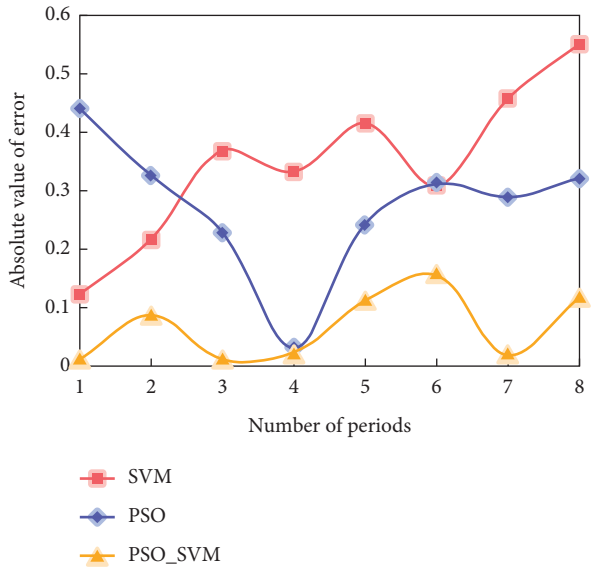


Figure 6: Error of prediction data of each model in section B#.

information of highway FS data, which has certain research significance and popularization value in practical engineering.

4. Conclusion

FS prediction of soft soil is a nonlinear and high-dimensional data processing problem, and SVM can solve this kind of data prediction problem well. Therefore, it is necessary to study the appropriate settlement prediction method for the settlement of soft soil subgrade, so as to guide the actual construction, make the settlement within the predictable range, and ensure that the difference between the predicted value and the measured value is small. It is our goal to improve the existing FS calculation and prediction methods, so that the predicted value of settlement is closer to the measured value. This article discusses the use of PSO algorithm to optimize the selection of three parameters of

SVM, namely, insensitive loss coefficient, penalty coefficient, and kernel parameter. Through specific tests, the results show that PSO_SVM is obviously superior to the other two models in prediction accuracy for different monitoring points, and the deformation data can better reflect the overall change information of highway FS data, which has certain popularization significance in practical engineering. However, how to accurately predict their settlement will be an important issue in highway construction. Therefore, the research of this article needs to further analyze and elaborate on this point.

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 interests.

References

[1] P. D. Long, "Prediction of piled raft foundation settlement—a case study," *Geotechnical Engineering*, vol. 47, no. 1, pp. 1–6, 2016.

[2] M. Wang, H. Sun, and R. Liu, "Denoising method and prediction technology of nuclear island foundation settlement," *Hedongli Gongcheng/Nuclear Power Engineering*, vol. 39, no. 3, pp. 56–61, 2018.

[3] X. Li, X. Liu, C. Z. Li, Z. Hu, G. Q. Shen, and Z. Huang, "Foundation pit displacement monitoring and prediction using least squares support vector machines based on multi-point measurement," *Structural Health Monitoring*, vol. 18, no. 3, pp. 715–724, 2019.

[4] Y. Li, K. Min, Y. Zhang, and L. Wen, "Prediction of the failure point settlement in rockfill dams based on spatial-temporal data and multiple-monitoring-point models," *Engineering Structures*, vol. 243, Article ID 112658, 2021.

[5] Y. G. Ejigu, F. N. Teferle, A. Klos, B. Janusz, and H. Addisu, "Monitoring and prediction of hurricane tracks using GPS tropospheric products," *GPS Solutions*, vol. 25, no. 2, pp. 1–15, 2021.

[6] Z. Yang and Z. Ge, "Monitoring and prediction of big process data with deep latent variable models and parallel computing," *Journal of Process Control*, vol. 92, pp. 19–34, 2020.

[7] S. H. Kim and H. G. Park, "Structure monitoring and long-term behavior prediction of wall-piloti structure," *Key Engineering Materials*, vol. 846, pp. 232–236, 2020.

[8] D. Zhou, Z. Li, J. Zhu, Z. Haichuan, and H. Lin, "State of health monitoring and remaining useful life prediction of lithium-ion batteries based on temporal convolutional network," *IEEE Access*, vol. 8, no. 99, p. 1, 2020.

[9] J. L. Hu, W. B. Wang, Z. Zhang, K. Zhang, K. Hua, and R Chen, "New monitoring program of transformer substation foundation settlement," *International Journal of Emerging Electric Power Systems*, vol. 20, no. 1, 2019.

[10] X. Liu, C. Zhao, Q. Zhang, C. Yang, and J. Zhang, "Characterizing and monitoring ground settlement of marine reclamation land of xiamen new airport, China with sentinel-1 SAR datasets," *Remote Sensing*, vol. 11, no. 5, p. 585, 2019.

[11] M. X. Chen, "Research on the surface subsidence status and residual settlement prediction of the goaf," *Journal of Railway Engineering Society*, vol. 35, no. 6, pp. 16–20, 2018.

[12] D. Hajializadeh, E. J. Obrien, and A. J. O'Connor, "Virtual structural health monitoring and remaining life prediction of steel bridges," *Canadian Journal of Civil Engineering*, vol. 44, no. 4, pp. 264–273, 2017.

[13] J. Zhao, Y. Liu, and M. Hu, "Optimisation algorithm for decision trees and the prediction of horizon displacement of landslides monitoring," *Journal of Engineering*, vol. 2018, no. 16, pp. 1698–1703, 2018.

[14] J. Sreekanth, H. Lau, and D. E. Pagendam, "Design of optimal groundwater monitoring well network using stochastic modeling and reduced-rank spatial prediction," *Water Resources Research*, vol. 53, no. 8, pp. 6821–6840, 2017.

[15] H. Yu and Y. Shangguan, "Settlement prediction of road soft foundation using a support vector machine (SVM) based on measured data," *MATEC Web of Conferences*, vol. 67, Article ID 07001, 2016.

[16] L. Jia, R. Liang, and K. Yao, "Research on the in-situ monitoring of post-construction settlement of bridge-subgrade transition section," *Journal of Railway Engineering Society*, vol. 34, no. 9, pp. 18–21, 2017.

[17] S. L. Yuan, "High-rise building deformation monitoring based on remote wireless sensor network," *IEEE Sensors Journal*, no. 99, p. 1, 2021.

[18] L. F. Wang, X. Z. Weng, Y. Li, Bw Guan, Zh Yao, and Q. Bo, "Study on the application of ultrasonic wave in foundation settlement monitoring," *Geotechnical Testing Journal*, vol. 42, no. 2, p. 20170296, 2019.

[19] Z. Yang, Y. Gao, and X. Fu, "A decision-making algorithm combining the aspect-based sentiment analysis and intuitionistic fuzzy-VIKOR for online hotel reservation," *Annals of Operations Research*, 2021.

[20] L. Liu, H. Moayedi, A. S. A. Rashid, S. S. A. Rahman, and H Nguyen, "Optimizing an ANN model with genetic algorithm (GA) predicting load-settlement behaviours of eco-friendly raft-pile foundation (ERP) system," *Engineering with Computers*, vol. 36, no. 1, pp. 421–433, 2020.

[21] Z. Yang, J. Chang, L. Huang, and A. Mardani, "Digital transformation solutions of entrepreneurial SMEs based on an information error-driven T-spherical fuzzy cloud algorithm," *International Journal of Information Management*, Article ID 102384, 2021.

[22] I. N. Shardakov, A. P. Shestakov, R. V. Tsvetkov, and V. Yepin, "The hydrostatic level method for continuous monitoring of building foundations," *Solid State Phenomena*, vol. 243, pp. 105–111, 2015.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 7798640, 7 pages
https://doi.org/10.1155/2022/7798640

Research Article

Analysis and Research on the Marketing Strategy of Agricultural Products Based on Artificial Intelligence

Wang Hongbing,[1,2] Gao Jing,[3] Kang Bohan,[2] Lyu Peng,[1] and Shi Yuxian [1]

[1]*Jeonju University, Jeollabuk-do, Jeonju 55069, Korea*
[2]*Shijiazhuang Information Engineering Vocational College, Shijiazhuang, Hebei 050000, China*
[3]*Hebei Transportation Investment Group Corporation, Shijiazhuang, Hebei 050000, China*

Correspondence should be addressed to Shi Yuxian; 631505020213@mails.cqjtu.edu.cn

Received 17 May 2022; Revised 22 June 2022; Accepted 4 July 2022; Published 30 July 2022

Academic Editor: Mukesh Soni

With the gradual development of artificial intelligence (AI), the traditional production, marketing, and management methods for agricultural products have undergone dramatic changes, necessitating a greater optimization of these methods. Agricultural product operators have begun incorporating AI technology into product production, marketing, and distribution processes. This article examines the current state of agricultural product management and then investigates the integration of production, marketing, and distribution using artificial intelligence. In addition, given the limitations of conventional methods for classifying agricultural products, this article presents a classification model that combines factor analysis with an enhanced support vector machine (SVM) based on genetic algorithms (GAs). The results of the experiments indicate that the improved method is capable of distinguishing agricultural product quality categories rapidly and precisely, significantly improving the classification accuracy of agricultural product quality, and being broadly applicable to the evaluation of agricultural product quality.

1. Introduction

Agriculture, a foundational industry in numerous nations, serves as the basis of their economies. Agriculture's growth is intricately intertwined with the production, processing, marketing, and distribution of agricultural products, as well as the logistics of transporting those goods. Important objectives include restructuring the agricultural supply side, optimizing agricultural mode and structure, promoting development of various agricultural industries, accelerating economic growth in agricultural counties and rural areas, increasing farmers' incomes, broadening the agricultural and rural economy, and cultivating family farms, farmers' cooperatives, and other new economic forms [1–3]. Due to the rapid development of artificial intelligence, it is necessary to investigate the distribution and marketing of agricultural products. This is crucial for the development of new agricultural market models and the advancement of the agricultural economy, among other things.

Self-media was coined in 2002 to describe the dissemination and sharing of factual news and other content via the Internet, social media platforms, and artificial intelligence-related technologies. Short video clips, in particular, have the potential to attract a large number of viewers in a short amount of time. Compared to traditional forms of media, Internet self-media comes in variety of shapes and sizes, and it is very easy to create people's interest in it. Amid rising living standards in their respective nations, consumers are becoming increasingly concerned with the quality of agricultural products [4, 5]. In recent years, an increasing number of consumers have opted to stay at home and make purchases online or on their mobile devices. Online sales of agricultural products have become increasingly popular with consumers over the past few years, according to retailers. In recent years, new methods for selling agricultural products, such as live e-commerce and short-form video marketing, have gained popularity. As a result of the rapid development of self-media platforms that enable customers to complete

agricultural product transactions online, consumption has become more convenient and comprehensive [6].

Thus, artificial intelligence (AI) has fundamentally altered the market structure for agricultural products. Traditional methods of selling agricultural products, such as physical stores and farmer's markets, are being phased out in favor of, among other things, community e-commerce, Douyin live broadcasts, and short video sales. In many communities, local governments are now selling and advertising agricultural products through self-media platforms, and some communities have even conducted live broadcasts with positive results, thereby creating new sales channels for local agricultural products. With increasing living standards comes a greater emphasis on product quality, the concept of eating well has become increasingly blurred [7, 8]. Consequently, a new type of agricultural production system is imaginable. Customers are interested in learning about the entire agricultural product lifecycle through artificial intelligence (AI) technology, and they want to participate in product design and development. As cloud computing and virtual reality technologies, such as virtual reality headsets, continue to advance, this demand should be met. With the help of these new technologies, it is possible to process agricultural products or display the entire growth process in front of customers, making it easier for them to feel and comprehend the entire process of agricultural production, harvesting, and processing while also bringing them closer to their desired product [9–14].

Numerous concepts for the classification of agricultural product quality have been generated by computer vision (CV) in artificial intelligence (AI) [15, 16], and some image-based detection methods have been successfully applied to the quality analysis of meats and fish, pizza, cheese, and bread products. However, this method's calculations are complex and its classification results are poor. According to the researchers, a machine vision system for automatically classifying fruits such as oranges, peaches, and apples can be used to estimate fruit quality online. Some researchers [17–19] have proposed a future-useful method for classifying and evaluating APT attack behavior based on stage characteristics. This technique could be used to distinguish the quality of agricultural products. It is possible to identify the most influential classification factors by summarizing and classifying agricultural products at the granular level; however, this process is time-consuming and challenging to implement. According to some researchers, cost-sensitive feedback neural networks were used to classify agricultural products, and the evaluation model was modified based on relevant agricultural product class data [20–23]. Even though the quality of agricultural products was not taken into account during the application process, the classification effect of this method was insufficient to differentiate them [24]. Other researchers have found that the geometric and morphological characteristics of potatoes can be used to distinguish between distinct quality levels [25]. A regression model relating potato quality to geometric feature parameters is constructed using linear regression analysis, and six potato invariant moment parameters are extracted and incorporated into the model [26, 27]. It is possible to grade with the assistance of a previously trained neural network.

To accurately classify agricultural products, which are influenced by a wide variety of factors, it is necessary to develop a comprehensive classification model that is nonlinear and applicable to a broad range of variables. This paper investigates the classification of agricultural product quality using GA-SVM. The model begins by compressing the dimensions of quality characteristics of agricultural products into smaller units. It improves the classification of the quality of agricultural products by decreasing the amount of input required by the classification model. GA's robust global optimization, parallelization, and efficient execution capabilities contribute to the algorithm's continued existence. Changing the construction of kernel functions and the parameters they employ is a challenge undertaking. In the following simulations, the accuracy and safety of the model described in this paper are demonstrated.

The structure of the article is as follows: Section 2 presents the current state of the market and relevant recommendations. Section 3 introduces the method proposed in this paper, Section 4 introduces the experimental results, and Section 5 summarizes the full text.

2. Marketing Status and the Improvement Strategy

2.1. The Current State of Marketing. As artificial intelligence (AI) technology continues to advance, more and more agricultural product producers are beginning to sell and distribute their products via mobile phones, while an increasing number of agricultural product producers are transitioning into the role of agricultural product operators (or agribusinesses). The development of agricultural products is changing as well, but there are still some issues with the way agricultural products are being operated today.

First and foremost, the main body of agricultural products management lacks a sufficient level of professional expertise and experience. Agribusinesses operate in geographically dispersed regions, and the traditional circulation of agricultural products necessitated the participation of a variety of links, ranging from agricultural product producers to local supply and marketing organizations to middlemen and local processing enterprises and the participation of different personnel at each of these links. With the advancement of artificial intelligence technology, more and more agricultural products are now being sold via mobile phones in real time with the role of the producer shifting to that of the operator in the previous years. In contrast to this, some farmers have a limited educational background and a lack of professional knowledge in areas such as e-commerce (product sales), business operations, and so on. They are also frequently not very knowledgeable about the in-depth operation of mobile phones. An extensive amount of agriculture is sold with the assistance of young children. The majority of agricultural product producers, as a result, have limited knowledge and experience with self-media technology. Furthermore, the distribution of marketing training

resources for agricultural products by supply and marketing agencies and local departments in different regions is inequitably distributed across the nation. Professional marketers and technicians are being actively recruited in some regions in order to provide self-media technology training and marketing expertise. In most cases, however, effective guidance and training are still lacking in the majority of the regions.

In addition, there is no integrated operation for managing agricultural products. Many regions have begun using self-media technology to manage agricultural products, but the vast majority of agricultural product producers only use self-media, such as mobile phone live broadcast software, for product sales and simple publicity because they lack the knowledge and experience necessary to run an integrated business. Using web-based media technologies, agricultural products can be produced, promoted, and distributed in addition to receiving after-sales support. When utilized in the manufacturing process, for instance, CV technology can enable both customers and employees to participate and observe concurrently. We-media technology can be used for a variety of marketing purposes, including, draining, promotion, and brand management. In terms of logistics, we perform product tracking and Internet of Things traceability via self-media, as well as C2B customer customization and reverse guidance for new product development via after-sales feedback and presales research. Agricultural product operators are unaware of the significance of integrated operations and are only concerned with the sales link. Regarding technology, the vast majority of self-media platforms are only responsible for the operation of one of their website's links. Coordination is lacking due to the absence of a comprehensive platform that can integrate and manage all links. Currently, there is an urgent market demand for the development of self-media platforms and software.

Third, the infrastructure for the management of agricultural products is relatively inefficient. Agricultural products are susceptible to damage, have a short shelf life, and are widely dispersed across production areas, which results in relatively high logistics and distribution costs. Particularly for fresh agricultural products, more stringent requirements for cold chain transportation technology have been imposed. Any delay in distribution of agricultural products such as vegetables, meat, and poultry will result in a significant increase in the product's imposed cost, which will be exorbitant. Despite the constant improvement and optimization of China's logistics infrastructure, the vast majority of agricultural products are still produced in rural areas, some of which are even remote and underdeveloped. It is challenging to establish a centralized warehouse and optimize transportation nodes in order to increase the efficiency and decrease the cost of logistics distribution question. Although there are many orders for agricultural products, they are few and dispersed, which increases logistical and transportation costs, which will ultimately be reflected in the selling price, making it difficult to generate a profit in the initial stages of the marketing cycle.

2.2. Strategy. Machine learning and artificial intelligence (AI) are transforming nearly every aspect of human life at a breakneck pace. Using artificial intelligence technology for agricultural product marketing is more efficient and convenient than using traditional marketing channels. Producers of local agricultural products have begun using artificial intelligence technology to broadcast live broadcasts in order to combat low sales and widespread support for many agricultural products. As a result of self-media platforms powered by artificial intelligence, consumers are becoming increasingly adept at identifying their own agricultural products and brand names (AI). As a result of the use of self-media technology in agriculture, progress in refrigeration and cold chain logistical systems has accelerated at the same time. Because of the urgency with which fresh agricultural products must be received, the cold chain storage requirements for these products are more stringent. Farmers who broadcast their products live not only increase the number of customers but they also help to open up new avenues for economic development, such as tourism and entertainment. The integration of artificial intelligence into agricultural product production, processing, sales promotion, and cold chain distribution logistics as well as research into agricultural product marketing strategies in an AI environment is therefore critical to the future of the industry.

In the context of agricultural product marketing, we will first develop a model for self-media analysis. An agricultural product practitioner can become an expert in marketing operations by employing descriptive analysis methods and various statistical analysis methods for the data generated by the operation. This is achieved by employing a model based on Douyin live broadcast and WeChat and Weibo video blog promotion. Using consumer preferences and behavior analysis in tandem with operational analysis, it is possible to create agricultural products and live broadcast activities for individual consumption. The marketing process of the farming industry is investigated using statistical methods, primarily hypothesis testing to determine whether the effect of activities has met expectations, time series analysis to forecast the sale of agricultural products, and regression analysis to guide the farming industry's production, inventory, and management decisions. A comprehensive analysis is done in the field of affiliate marketing. When using descriptive analysis to create daily traffic channels, hot items, slow-moving items, and inventory warnings, it is essential to be as specific as possible.

Second, design a system for the distribution and logistics of agricultural commodities using artificial intelligence. To design a suitable planning algorithm or an artificial intelligence model algorithm based on the findings of this research into the agricultural product logistics distribution mode, it is necessary to conduct research into the agricultural product logistics distribution mode to develop a suitable problem model. Comparing the developed algorithm model to the current agricultural product logistics and transportation system improves the algorithm's performance. The ultimate objective of the mathematical model for agricultural product logistics and distribution is to deliver

the corresponding products on time in response to the real-time needs of actual consumers and intermediary physical stores and then to establish optimal distribution routes to minimize logistics and distribution costs as much as possible. Due to the development and implementation of artificial intelligence algorithms, agricultural products can be delivered to consumers in the shortest amount of time possible. This affords consumers the opportunity to receive services of higher quality. The O2O Internet of Things (IoT) and the self-media platform, which are both available on the O2O platform, enable consumers to monitor agricultural products in real time. When product quality issues are identified, it is possible to identify the source of the issue and make timely returns and exchanges, which aids in resolving product after-sales issues and effectively excavates potential customers.

Finally, an agricultural product display system should be constructed. Because the product has no physical characteristics, it is difficult to locate it when shopping online. Self-media technology has the potential to significantly alter this circumstance. With the goals of reducing the distance between consumers and agricultural products and increasing consumer recognition of agricultural products, a system for the visualization of agricultural products is being developed using both the big data cloud platform and the self-media live broadcast platform. In addition, the Internet of Things, 5G technology, and virtual reality technology are being utilized to realize cloud adoption and cloud interaction among agricultural product consumers, enabling them to engage in immersive experiential consumption. Using We-media and 5G technology, it is possible to develop visual agriculture and stream agricultural processes such as processing, production, harvesting, and distribution of agricultural products over the Internet in order to increase efficiency. This will help promote rural tourism and the arts and entertainment industries in the surrounding area. In the modern era, rural economic forms are becoming more diverse, and new media agricultural product operation small family teams can help accelerate the development of new types of agricultural businesses, such as small family farms and Internet workshops.

3. Method

First, factor analysis is introduced. Simply put that factor analysis is to find common factors hidden behind multiple variables with commonality. Suppose there are N samples, P indicators, $\mathbf{X} = (x_1, x_2, \ldots, x_N)^T$ is a random vector. Then, the factor model is

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F + \cdots + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F + \cdots + a_{2m}F_m + \varepsilon_2 \\ \vdots \\ X_n = a_{n1}F_1 + a_{n2}F_2 + \cdots + a_{mm}F_m + \varepsilon_n \end{cases}, \tag{1}$$

where $F = (F_1, F_2, \ldots, F_m)^T$ denotes the common factor, a_{ij} is the factor loading, and $A = a_{ij}$ is the factor loading matrix. The matrix form of the above model is

$$\mathbf{X} = \mathbf{AF} + \varepsilon, \tag{2}$$

where $a^{(l)}$ represents the activate node, $b^{(l)}$ is the parameter matrix, and $b_0^{(l)}$ is the bias value.

GA is a random global search and optimization technique that imitates the natural biological evolution mechanism. It is essentially a fast, parallel, global search technique with a high degree of adaptability for solving problems. Using the principle of survival of the fittest, the GA operation continuously generates a near-optimal solution from a population of potential solutions. In each new generation, individual selection is employed in the reconstruction method to generate a new approximate solution based on the individual's fitness in the problem domain and natural genetics. As a result of this process, the population evolves and becomes more adaptable to the environment than its predecessors, just as nature does.

Suppose $x_A^{(t)}$ and $x_B^{(t)}$ are two random variables of A and B, t is the current time, then

$$\begin{aligned} x_A^{(t+1)} &= \alpha x_B^{(t)} + (1-\alpha)x_A^{(t)}, \\ x_B^{(t+1)} &= \alpha x_A^{(t)} + (1-\alpha)x_B^{(t)}, \end{aligned} \tag{3}$$

where α is the coefficient of variation. Then, we have the adaptable concept after crossover and mutation operators as follows:

$$P_c = \begin{cases} P_{c1} - \dfrac{(P_{c1} - P_{c2})(f' - f_{\text{avg}})}{f_{\max} - f_{\text{avg}}}, & f' \geq f_{\text{avg}}, \\ P_{c1}, & f' < f_{\text{avg}}, \end{cases}$$
$$P_m = \begin{cases} P_{m1} - \dfrac{(P_{m1} - P_{m2})(f - f_{\text{avg}})}{f_{\max} - f_{\text{avg}}}, & f \geq f_{\text{avg}}, \\ P_{m1}, & f < f_{\text{avg}}, \end{cases} \tag{4}$$

where P_{c1} and P_{c2} are the crossover rates, P_{m1} and P_{m1} are the mutation rates, $f_{\max}$ is the maximum fitness value, and f_{avg} is the average fitness value.

The core of SVM is the search for an optimal hyperplane for feature separation. It employs a maximum margin hyperplane to map vectors into a space with a high dimension. The data are separated by a hyperplane, which is flanked by two parallel hyperplanes. To maximize the distance between two parallel hyperplanes, the hyperplane that separates them is utilized. The greater the distance or separation between parallel hyperplanes, the smaller the overall classification error.

The SVM solves the classification problem by solving the optimization problem as follows:

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\varepsilon_i, \tag{5}$$

$$s.t. \begin{cases} y_i(wx_i + b) \geq 1 - \varepsilon_i \\ \varepsilon > 0 \end{cases}, \tag{6}$$

where ε_i is the slack variable introduced and C is the penalty factor.

To solve (6), we have

$$\max w(\beta) = \sum_{i=1}^{n} \beta_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_i \beta_j y_i y_j K\left(x_i, x_j\right),$$

$$\text{s.t.} \begin{cases} \sum_{i=1}^{n} \beta_i \beta_j = 0 \\ 0 \le \beta_i \le C \end{cases}, \tag{7}$$

where β_i represents the Lagrange multiplier and $K(x_i, x_j)$ is the kernel function. Then, we have

$$f(x) = \operatorname{sign}\left[\sum_{i=1}^{n} \beta_i j_i K(x_i, x) + b\right]. \tag{8}$$

Traditional SVM parameters are generated at random, and the classification accuracy is unstable. This paper uses GA to improve the selection of SVM parameters because of its strong optimization abilities. The following is a list of the steps involved.

Step 1 (data preparation). To eliminate the dimensional difference between the original variables, normalize the training and test samples.

Step 2 (population coding and initialization). Several initial populations are built, and the parameters of the penalty function and kernel function are binary-coded.

Step 3 (fitness function and compute). After the chromosomes in the population have been decoded, the fitness function can be calculated using the test sample set's accuracy in predicting C and o.

Step 4 Make a decision. To move on to step 5, choose the best parameter combination for output by evaluating whether the optimization process meets the genetic algorithm's termination conditions. Alternatively, you can use techniques like crossover and mutation to create new populations and kick off a new genetic era.

Step 5 Train the SVM model with the optimal parameter combination.

Step 6 To determine the classification accuracy, run a classification prediction on the test set.

4. Results

Classification of agricultural product quality is a complex nonlinear problem that is influenced and governed by a large number of unpredictability variables. Numerous factors influence the quality of agricultural products, and the factors that influence the quality of different types of agricultural products are also diverse. These factors can be classified broadly into five categories: shape, size, color, texture, and defect severity. In this paper, an apple is used as an illustration, and four eigenvalues, including the average diameter of the largest cross section, circularity, the area ratio of red area, and defect area, are selected as the discriminant factors for categorizing the fruit.

Table 1: KMO and Bartlett Spherical Test.

KMP sample test		0.873
Bartlett spherical test	Chi-square value	237.42
	Degree of freedom	20
	p	0.0001

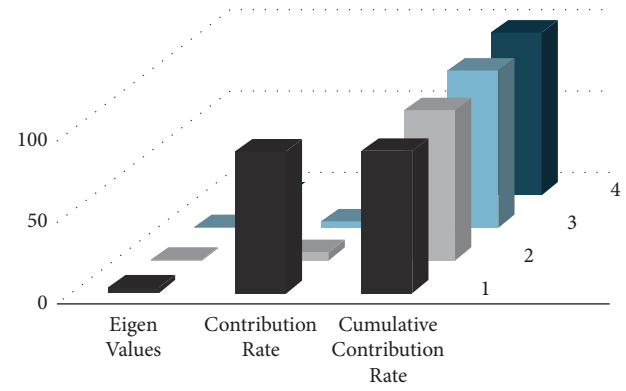


Figure 1: Total variance explain.

Literature [13] indicates where the 30 data sets used in the model originated. Table 1 displays the results of a Pearson correlation analysis and a factor analysis test conducted on the four eigenvalues. The results indicate that there is a clear correlation between the four eigenvalues and that the eigenvalues contain overlapping information, making them suitable for factor analysis. The results of the two preceding tests indicate that factor analysis can be used to process these four eigenvalues in order to achieve the desired dimensionality reduction.

The interpretation of the total variance of the common factor is shown in Figure 1.

As depicted in Figure 1, the first two common factors are extracted to reduce information loss and improve classification precision. This is due to the fact that the first two common factors can explain 96.7 percent of the original index variable's information, which is the case for the first two common factors.

In this article, four levels of Fuji apple quality are described. In this experiment, 30 data sets were used as training samples and 7 data sets were used as test samples, for a total of 37 data sets. Using the GA-SVM agricultural product quality classification model developed through factor analysis, 200 evolutions of the model's parameters yielded a stable iterative value of the best fitness. This is illustrated in Figure 2.

According to Figure 2, we set the penalty function as 10.6 and the kernel function parameter as 1.95, and the training classification result is shown in Figure 3.

The results on the test set are shown in Figure 4.

The above graph demonstrates that our proposed method for classifying the quality of agricultural products is extremely precise.

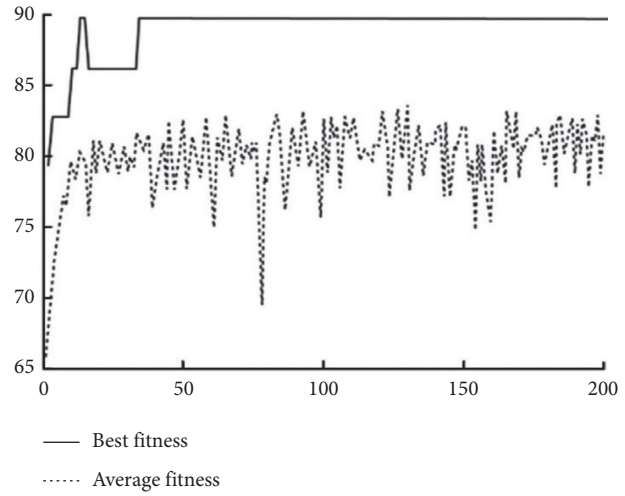


Figure 2: Fitness curves of GA-SVM.

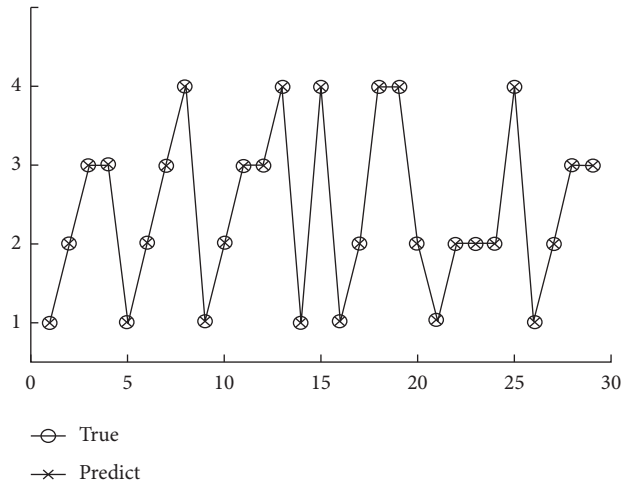


Figure 3: Classification prediction results on the training set.

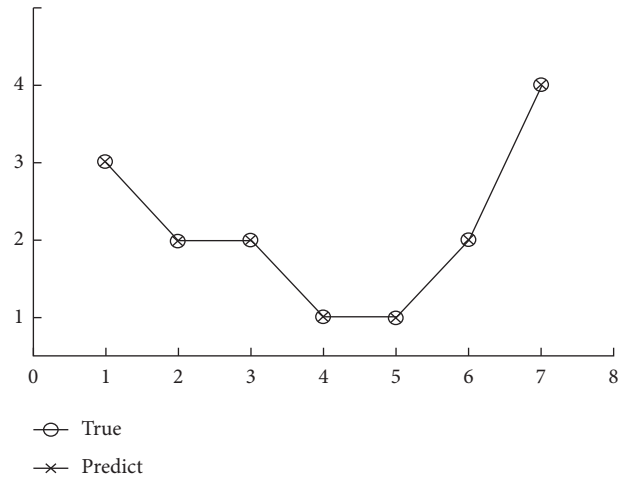


Figure 4: Classification prediction results on the testing set.

Figure 5 depicts the results of a comparison between the proposed method and SVM, BPNN, and decision tree (DS)

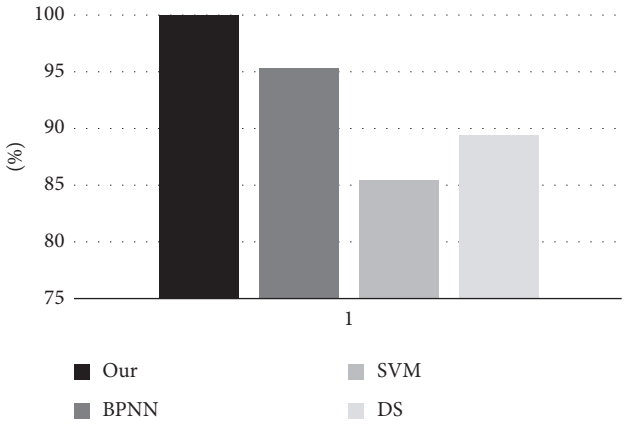


Figure 5: Comparison of classification performance of agricultural products by different methods.

to demonstrate that the proposed method is superior to the other studied methods. This paper has a classification accuracy rate of 100 percent, which is significantly higher than the classification accuracy rates of the other three tested models. Using factor analysis, one can demonstrate that the GA-SVM model for classifying the quality of agricultural products is safe and reliable. It can also be demonstrated that the model is capable of producing rapid and accurate agricultural product quality classifications.

5. Conclusion

Over the past few decades, there has been a significant shift in the production, marketing, and management of agricultural products, and these approaches must be further optimized in light of the rise of artificial intelligence (AI). As a result of AI's rapid advancements, many agricultural producers have become agricultural product operators, and agricultural product operators have begun to implement AI technology in product production, marketing, and distribution. In this study, agriculture product management is analyzed, and AI technology is employed to investigate how to integrate production, marketing, and distribution. In addition, this study provides a classification model for agricultural products that combines factor analysis with an improved SVM based on a GA. It was discovered that the improved method can rapidly and accurately identify quality categories of agricultural products, significantly improve classification accuracy, and can be widely used to evaluate agricultural product quality.

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] K. Drachal and M. Pawłowski, "A review of the applications of genetic algorithms to forecasting prices of commodities," *Economies*, vol. 9, no. 1, p. 6, 2021.

[2] G. Ren, Y. Sun, M. Li, J. Ning, and Z Zhang, "Cognitive spectroscopy for evaluating Chinese black tea grades (*Camellia sinensis*): near-infrared spectroscopy and evolutionary algorithms," *Journal of the Science of Food and Agriculture*, vol. 100, no. 10, pp. 3950–3959, 2020.

[3] O. Devos, G. Downey, and L. Duponchel, "Simultaneous data pre-processing and SVM classification model selection based on a parallel genetic algorithm applied to spectroscopic data of olive oils," *Food Chemistry*, vol. 148, pp. 124–130, 2014.

[4] K. Drachal and M. Pawłowski, "A review of the applications of genetic algorithms to forecasting prices of commodities," *Economies*, vol. 9, no. 1, p. 6, 2021.

[5] M. O. Özorhan, İ H. Toroslu, and O. T. Şehitoğlu, "A strength-biased prediction model for forecasting exchange rates using support vector machines and genetic algorithms," *Soft Computing*, vol. 21, no. 22, pp. 6653–6671, 2017.

[6] B. Sang, "Application of genetic algorithm and BP neural network in supply chain finance under information sharing," *Journal of Computational and Applied Mathematics*, vol. 384, Article ID 113170, 2021.

[7] J. S. Chou, M. Y. Cheng, Y. W. Wu, and A. D. Pham, "Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification," *Expert Systems with Applications*, vol. 41, no. 8, pp. 3955–3964, 2014.

[8] Z. Lin, S. Chen, H. Liang, and Zhang, "Analysis of capital flow in commodity futures market based on SVM[J]," *International Journal of Economics and Finance*, vol. 10, no. 8, pp. 1–28, 2018.

[9] K. Nagasubramanian, S. Jones, S Sarkar, and A. K. Singh, "Hyperspectral band selection using genetic algorithm and support vector machines for early identification of charcoal rot disease in soybean stems," *Plant Methods*, vol. 14, no. 1, pp. 1–13, 2018.

[10] C. L. Dunis, S. D. Likothanassis, A. S. Karathanasopoulos, G. S. Sermpinis, and K. A Theofilatos, "A hybrid genetic algorithm-support vector machine approach in the task of forecasting and trading," *Journal of Asset Management*, vol. 14, no. 1, pp. 52–71, 2013.

[11] Z. Alameer, M. A. Elaziz, A. A. Ewees, H. Ye, and Z Jianhua, "Forecasting copper prices using hybrid adaptive neuro-fuzzy inference system and genetic algorithms," *Natural Resources Research*, vol. 28, no. 4, pp. 1385–1401, 2019.

[12] Y. X. Pu and C. L. Yu, "Automatic identification of crop diseases image based on the dual coding genetic algorithm of support vector machine," *Guizhou Agricultural Sciences*, vol. 7, pp. 187–194, 2013.

[13] J. Ruan, H. Jiang, X. Li, Y. Shi, F. T. S. Chan, and W Rao, "A granular GA-SVM predictor for big data in agricultural cyber-physical systems," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 12, pp. 6510–6521, 2019.

[14] T. Guoxiang, Q. Ming, W. Xuan, and L. Jiake, "A Parameter Selection of Support Vector Machine with Genetic Algorithm for Citrus Quality classification," in *Proceedings of the 2011 6th International Conference on Computer Science & Education (ICCSE)*, pp. 386–390, IEEE, Singapore, August 2011.

[15] P. Zhang, Q. Guo, S. Zhang, and H. H. Wang, "Pattern mining model based on improved neural network and modified genetic algorithm for cloud mobile networks," *Cluster Computing*, vol. 22, no. 2, pp. 1–10, 2019.

[16] Q. Wu and C. Peng, "A least squares support vector machine optimized by cloud-based evolutionary algorithm for wind power generation prediction," *Energies*, vol. 9, no. 8, p. 585, 2016.

[17] M. A. Nanda, K. B. Seminar, M. Solahudin, A. Maddu, and D. Nandika, "Implementation of genetic algorithm (ga) for hyperparameter optimization in a termite detection system," in *Proceedings of the 2nd International Conference on Graphics and Signal Processing*, pp. 100–104, New York, NY, USA, October 2018.

[18] M. Shibang, "Nondestructive determination of kiwifruit SSC using visible/near-infrared spectroscopy with genetic algorithm," *Journal of Engineering Science and Technology Review*, vol. 14, no. 1, pp. 100–106, 2021.

[19] H. Chung and K. S. Shin, "Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction," *Neural Computing & Applications*, vol. 32, no. 12, pp. 7897–7914, 2020.

[20] H. Ocak, "A medical decision support system based on support vector machines and the genetic algorithm for the evaluation of fetal well-being," *Journal of Medical Systems*, vol. 37, no. 2, p. 9913, 2013.

[21] Y. Y. Hong, J. V. Taylar, and A. C. Fajardo, "Locational marginal price forecasting using deep learning network optimized by mapping-based genetic algorithm," *IEEE Access*, vol. 8, Article ID 91975, 2020.

[22] M. H. Tseng, S. J. Chen, G. H. Hwang, and M. Y. Shen, "A genetic algorithm rule-based approach for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 63, no. 2, pp. 202–212, 2008.

[23] A. Bouraoui, S. Jamoussi, and Y. BenAyed, "A multi-objective genetic algorithm for simultaneous model and feature selection for support vector machines," *Artificial Intelligence Review*, vol. 50, no. 2, pp. 261–281, 2018.

[24] A. Raghuvanshi, U. K. Singh, G. S. Sajja, H. Pallathadka, and E. Asenso, M. Kamal, A. Singh, and K. Phasinam, Intrusion detection using machine learning for risk mitigation in IoT-enabled smart irrigation in smart farming," *Journal of Food Quality*, pp. 2022–8, Article ID 3955514, 2022.

[25] S. Sarkar, A. Lohani, and J. Maiti, "Genetic Algorithm-Based Association Rule Mining Approach towards Rule Generation of Occupational accidents," in *Proceedings of the International Conference on Computational Intelligence, Communications, and Business Analytics*, pp. 517–530, Springer, New York, NY, USA, March 2017.

[26] S. Liu, H. Tai, Q. Ding, D. Li, L. Xu, and Y. Wei, "A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction," *Mathematical and Computer Modelling*, vol. 58, no. 3-4, pp. 458–465, 2013.

[27] K. Y. Chen, "Forecasting systems reliability based on support vector regression with genetic algorithms," *Reliability Engineering & System Safety*, vol. 92, no. 4, pp. 423–432, 2007.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 7200657, 9 pages
https://doi.org/10.1155/2022/7200657

Research Article

Local Defogging Algorithm for Improving Visual Impact in Image Based on Multiobjective Optimization

Qiuju Lu

Department of Informationization Construction and Management, Shaanxi University of Technology, Hanzhong 723000, China

Correspondence should be addressed to Qiuju Lu; luqiuju@mjcedu.cn

Received 27 April 2022; Revised 12 June 2022; Accepted 22 June 2022; Published 30 July 2022

Academic Editor: Mukesh Soni

The preprocessing of images is required for many applications based on industry, social, and academic requirements. Researchers have developed a number of techniques to improve the visual effect of images and appropriately interpret visual effects. The accuracy of visuals is important in cyber security, military organization, police organizations, and forensics to detect the true story from the pictures. They search for evidence by digging deep into the network in search of evidence. If visuals are not clear, preprocessing of images is not done correctly, then it may lead to wrong interpretations. This paper proposes an image local defogging technique based on multiobjective optimization to improve the visual effect of the image as well as the information entropy. The multiobjective function is selected to establish the image reconstruction model based on multiple objectives. The model is utilized to reconstruct a single image to moderate the impact of noise and other interference factors in the original image. The color constancy model and effective detail intensity model are also devised for image enhancement to get the visual details. The atmospheric light value and transmittance are evaluated using a physical model of atmospheric scattering, and the guided filter is used to maximize the transmittance of a single image and improve the efficiency of image defogging. The dark channel priority method is used to realize the local defogging of a single image and to design the local defogging algorithm. Experiments verify the optimization effect of the proposed algorithm in terms of information entropy and container network interface (CNI) value. The tone restoration degree is good, and it improves the overall image quality. The image defogging effect of the proposed algorithm is verified with respect to subjective and objective levels to check the efficacy of the proposed multiobjective model.

1. Introduction

Ordinary optical imaging of outdoor scenes is often affected by foggy weather [1]. In foggy weather, the light reflected from the object's surface will be scattered under the influence of suspended particles in the atmosphere before reaching the imaging equipment [2]. The degree of scattering is proportional to the type, size, shape, and degree of coagulation of dispersed particulates in the air, which means the fog concentration is proportional to light wavelength. [3]. When the fog is thick, the contrast of the image obtained by the imaging equipment is low, and the color is biased to gray white resulting in unclear identification of objects in the image [4]. This directly affects the normal work of most automatic image systems based on computer vision algorithms such as transportation, outdoor monitoring, and terrain detection system [5]. Therefore, the research of image defogging algorithm has an imperative need to apply this algorithm in real-time applications [6].

Many authors have presented their respective works in the area of image processing for defogging the images. The authors proposed an environmental scattering model and retinex-based infrared image defogging technique in [7]. The atmospheric scattering model is used to rebuild the infrared image based on the resemblances between infrared and visible images in the degradation process in foggy conditions. Based on the difference in radiation, infrared images can differentiate targets from their environments, which works effectively in all weather conditions whereas textual details are provided by visual images. However, after defogging and restoration, the image often has the characteristics of low contrast and unclear details which are not conducive to direct observation by human eyes. In this case, to improve the contrast of the defogged image, the method

employs the use of retinex model. The test results reveal that after algorithm processing, the signal-to-noise ratio (SNR) of infrared images can be improved, and this work can find a balance between calculation processing speed and defogging process, laying the groundwork for later embedded platforms to realize real-time video defogging. Both infrared and visible images process the intensity of the gray level in the image. In [8], authors propose a single-image defogging algorithm based on deep learning. Using a convolutional network (CNN), defogging is realized by learning the mapping relationship between YUV (Y represents brightness and UV represents chromaticity) channels of foggy image and clear image. The network structure consists of two identical feature modules mainly including multiscale convolution, convolution, and jump connection structure. The experimental results show that the image restored by the algorithm has a good processing effect irrespective of synthetic fog image data set or natural fog image data set, but the container network interface value is low indicating that there is less effective information in the image.

In [9], authors propose an image dehazing method based on the second-order variation model. First, the atmospheric light value and the original transmittance map of the foggy image are estimated by the dark primary color prior method, and then, the nonlinear diffusion model is applied to the image. In the solution of the transmittance map, it is combined with the second-order variational model, the Hessian matrix variational model, the total generalized variational model, and the total curvature variational model. The results show that the edge of the image after processing is well maintained, and the image noise can be suppressed but the degree of tone restoration and the visual effect of the image are not good.

In [10], dehazing models (H-LV model, H-HMV model, H-TGV model, and H-TCV model) are used to augment the image quality through defogging process. In order to improve the computational efficiency, the corresponding exchange direction multiplier algorithms are designed for the four models. By introducing auxiliary variables, the Lagrangian multipliers are continuously updated and iterated until the energy equation converges, and finally, the dehazed image is the output. Finally, the LIVE image defogging image database is used to verify the method experimentally.

In [11], the local aerial elements of the image are extracted and then resortation is performed using the multiscale fusion method for fog-free image. In foggy images, each cluster becomes a line in RGB space, which the method may utilize to reconstruct digital elevation models and fog-free images. When dealing with cloudy images under complex sensing situations, these solutions are not ideal. Some researchers used convolution neural networks (CNN) to train foggy images. In [12], a defogging network of images was built using CNN. Fog-free images were extracted using the atmospheric scattering model. But this model was built based on a prior assumption like in traditional methods. In [13], a gated fusion defogging network (GFN) was proposed which is built through the sequence of operations such as contrast enhancement, gamma correction, and white balance. But the implementations of these operations were complicated [14].

To progress the visual effect of the image, a local dehazing algorithm for a single image constructed on a multiobjective optimization algorithm is suggested in this paper to address the limitations in the existing literature. By analyzing the verification results at the subjective and objective levels, it can be seen that the images processed by the proposed algorithm have high sharpness and contrast and are superior to other contrast algorithms in both subjective and objective evaluations. The major contributions of the paper are as follows:

(i) To boost the visual effect and information entropy of an image, an image local defogging algorithm based on multiobjective optimization is proposed in this paper.

(ii) The proposed model is utilized to reconstruct a single image to shrink the effect of noise and other meddling factors in the original image. The color constancy model and effective detail intensity model are also used for image enhancement to get the better visual details.

(iii) According to the physical model of atmospheric scattering, the atmospheric light value and transmittance are estimated. The guided filter technique is used to increase transmittance and the effectiveness of image defogging by maximizing the transmittance of a single image.

(iv) The dark channel priority method is used to realize the local defogging of a single image and to design the local defogging algorithm.

(v) The proposed model improves the overall image quality. The image defogging effect of the proposed algorithm is verified with respect to subjective and objective levels to check the efficacy of the proposed multiobjective model.

The rest of the paper is organized as follows: in Section 2, the single image reconstruction method based on a multiobjective optimization algorithm is elaborated. In Section 3, the model for image enhancement is discussed. In Section 4, local dehazing algorithm for a single image is described. In Section 5, experimental results and analysis is discussed in detail using the sampling method and evaluation indicators (objective and subjective quality evaluation). The last section 6 concludes the research study.

2. Single Image Reconstruction Model Based on Multiobjective Optimization

2.1. Selection of Multiobjective Function. The process of image defogging begins with determining the optimization target. The ultimate image processing effect is determined by the optimization target chosen. The task of image processing as well as the final image quality must be taken into account while choosing optimization objectives. The objectives of image processing differ due to the various tasks of image processing, in which an impact on the optimization targets has been chosen. The single image reconstruction processing optimization problem has many optimization objectives,

making it a multi-objective optimization problem. In general, there will be conflicts between multiple targets in image processing, and there is no optimal solution for all targets at the same time. The improvement of one target performance is often at the cost of reducing the performance of one or more other targets [15].

Considering the above problems, in order to effectively suppress noise in image reconstruction, the norm of the image is selected as the first objective function as shown in the following equation:

$$\delta(l) = \sum_{k=1}^{N} \|a(k) + b(k)\|^2, \tag{1}$$

where l represents the peak value of the image; $a(k)$ represents the structural component of the image; $b(k)$ represents the texture component of the image; N represents the number of image pixels; k represents the gamma.

Through the optimization of the objective function, the reconstructed image has the smallest peak globally, thereby reducing the interference of noise and effectively suppressing the artifacts caused by noise interference.

According to the characteristics of the reconstructed image itself, each subimage is not in a continuous gray level, but only in a limited number of gray frequency bands. The smooth function [16] is chosen as the second objective function, as indicated in equation (2), to further remove the influence of statistical noise in the iterative process and the influence of statistical noise included in the original data, and to increase the quality of image reconstruction.

$$\begin{aligned}\delta(r) &= \frac{1}{2}\sum_{k\in K}^{N}\left[h_1\gamma(k) + h_2\gamma(k-1)\right] \times \eta(k) \\ &= \frac{1}{2}h^T Dh \quad ,\end{aligned} \tag{2}$$

where h_1 and h_2 represent the set of pixels belonging to the same subimage; D represent the smoothing matrix; $\eta(k)$ represent the image intensity.

By minimizing the smoothing function represented in equation (2), there is a minimum nonuniformity between the pixels of the same subimage. On the one hand, to further strengthen the suppression of various noises and improve the quality of image reconstruction, the joint action with other objective functions is taken to increase the accuracy of image reconstruction, and the results are improved.

In order to further effectively recover the target image from the noisy image and improve the image reconstruction effect, the log-likelihood function is used as the third objective function of the reconstruction model is shown in the following equation:

$$\delta(m) = \frac{\min\left[f_z(x)\right]}{\phi_i \Delta \max P_x + \phi_j \Delta \min P_x}, \tag{3}$$

where ϕ_i and ϕ_j represent the noisy smooth block and the noisy nonsmooth block, respectively; $f_z(x)$ represents the log-likelihood estimation of the image block; $\Delta \max P_x$ represents the maximum variance estimate; and $\Delta \min P_y$ represents the minimum variance estimate.

2.2. Single Image Reconstruction Model. Considering the similarity of the objective functions, $\delta(l)$, $\delta(r)$, and $\delta(m)$, and the consistency of the optimization process, the above optimization objective functions can be combined into one objective function as shown in the following equation:

$$\delta_k^2 = \sum_{i,j=1}^{N} \left[\delta(l) + \delta(r) + \delta(m)\right]^{ij}. \tag{4}$$

Finally, in order to control the accuracy of image reconstruction, in the reconstruction process, the following constraints are set as given in the following equation:

$$G(h) = \sum_{i=1}^{N}\sum_{j=1}^{M}\left(x_{ij} + y_{ij}\right) - \mu = 0, \tag{5}$$

where μ represents a preset constant used to control the accuracy of reconstruction.

So far, a single image reconstruction model S_k based on multiobjective optimization can be obtained as given in the following equation:

$$S_k = \begin{array}{l}\max \delta(l), \\ \min \delta(r), \\ \min \delta(m), \\ s, t, h(\delta) = 0.\end{array} \tag{6}$$

To summarize, the reconstruction of a single image based on a multiobjective optimization algorithm [17] can get better reconstruction results and reduce the influence of noise and other interference factors in the original image on the later image defogging effect.

3. Single Image Enhancements

The image noise can be effectively suppressed by a single image reconstruction model constructed under the constraints of multiple objective functions. In order to further improve the image dehazing effect, color preservation effect and detail enhancement effect should also be considered [18]. As a result, image enhancement is realized, and the image visual effect is further adjusted starting from the two levels of color consistency and effective detail intensity.

3.1. Color Constancy Model. Color constancy is an adaptive characteristic of the human visual system (HVS), which means that the HVS can recognize the true color of an object under a wide range of different color lighting conditions. Under typical conditions, the human eye's perception of the color of an object's surface is closely tied to the object's reflection qualities but has nothing to do with the spectral features that reach the human eye. That is, when the external lighting conditions change, the human eye's color perception can still remain relatively unchanged to a certain extent, exhibiting color constancy, as shown in Figure 1. In fact, the human eye can recognize the reflected color of the object surface even when the signals of different colors of the object surface to the retina are the same under different color lighting conditions. Imaging

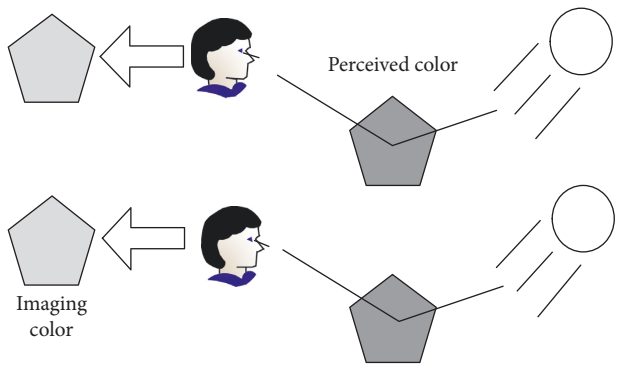


Figure 1: Color constancy model.

systems, unlike HVS, are unable to respond spectrally to changing lighting conditions, resulting in a color cast in the acquired image, or a shift in the entire image's color range.

Under weather conditions such as fog and haze, the scattering of atmospheric particles leads to low contrast and color distortion in scene imaging. Therefore, the color constancy algorithm can be well applied to image dehazing. The nonphysical model method avoids solving the atmospheric scattering model by treating the ambient light component of the model as the illuminance component and estimating the reflection image independently of illuminance.

3.2. Effective Detail Intensity Model. The image after dehazing enhancement should have higher sharpness, which is manifested image with high edge dissimilarity. However, the dissimilarity of the image is not as high as possible, and the edge information after filtering out noise and the influence of the Halo effect exists. The Halo effect occurs at the edge of the object, especially at the sudden change of the depth of the field, which is manifested as widening and brightening at the edge of the object. The presence of the Halo effect has a significant impact on the dehazed image's visual impression. According to the characteristics of Halo, the bright channel of the image is defined as

$$T(s) = \left(\frac{K}{G_s} + 1\right) \cdot \lambda^s, \tag{7}$$

where G_s represents the color channel of the image; λ^s represents the 7×7 small neighborhood-centered pixel point s.

The Canny operator is used to detect the image's edge, and the detected edge images are combined together to get the aggregate intensity detail of the image, which characterizes the intensity of the dehazed image after noise filtering, as shown in the following equation:

$$Z(s) = c_n(k) - \hat{h}(z)\partial(k), \tag{8}$$

where $c_n(k)$ represents the correlation of noise scale; $\hat{h}(z)$ represents Gaussian noise; $\partial(k)$ represents image signal-to-noise ratio.

Since the dehazed image may have the Halo effect for a pixel s' that exists in the set of all edge pixels of the image. The summation is performed of the small neighborhood values of the corresponding pixel points in the bright channels of the image as given in the following equation:

$$\theta_t = \sum_{k=1}^{N} h(x, y) e^t dt, \tag{9}$$

where e^t the image luminance channel.

According to equation (9), the brightness value of a single image is obtained to achieve image detail enhancement.

4. Local Dehazing Algorithm for Single Image

According to the color constancy model and the effective detail intensity model discussed in Section 3, the enhancement processing of the color and details of a single image is realized. Based on this, the local dehazing of a single image is studied.

4.1. Atmospheric Scattering Model. The reflected light of the target is absorbed and scattered by suspended particles in the environment in the presence of fog, haze, and other media. Simultaneously, external light, such as sunshine, is scattered by the scattering medium in the atmosphere, resulting in background light, making the camera's brightness dim and fuzzy. The atmospheric scattering model, as shown in Figure 2, is used to represent this physical process.

According to Figure 2, the atmospheric scattering physical model is represented as shown in the following equation:

$$U(n) = \frac{I(n)h(n)}{F_i(1 - w_i)}, \tag{10}$$

where $U(n)$ is the observed hazy image; $I(n)$ is the natural haze-free image; F_i and w_i represent the transmittance and atmospheric light values, respectively. Single image dehazing aims to recover a clear image from a hazy image. In this process, the atmospheric light value w_i and transmittance F_i need to be estimated, and their expressions are given by the following equations:

$$w_i = \ln\sqrt{1 - \Delta p} \times \cdot p_i, \tag{11}$$

$$F_i = \frac{x + x'}{D_t} \times \theta_i, \tag{12}$$

where p_i is the average brightness of the sky area; Δp is the variation coefficient of the average brightness of the sky area; D_t is the Halo phenomenon generated after transmission.

Assuming that the transmittance has local area consistency, equation (12) is improved, and the new transmittance expression can be obtained from the following equation:

$$F_i' = \left(\sum_{i=1}^{N} x_i(t)\right) \times h(t) \times \theta_i, \tag{13}$$

where $h(t)$ is the local atmospheric light; $x_i(t)$ is the initial transmission map.

In reality, even in sunny weather, the atmosphere cannot contain any particles. When viewing distant objects, fog still exists, that is, there is a phenomenon of spatial perspective. If

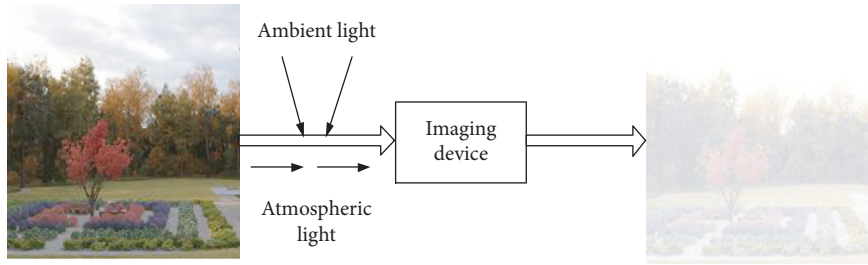


Figure 2: Physical model of atmospheric scattering.

removing the fog completely will make the image look unreal and the image depth will be lost, therefore, a constant ϕ is introduced in equation (13), $0 < \phi < 1$. Equation (13) can be further rewritten as the following equation:

$$F_i' = \int_{t_0}^{t} |h(t)| \mathrm{d}t \times \theta_i. \tag{14}$$

The advantage of this correction is that part of the fog covering distant objects can be retained, making the restored image more realistic and more in line with the characteristics of the human eye. The value of ϕ can be changed according to the actual situation. In this paper, the fixed value of ϕ is 0.95.

4.2. Improved Transmittance Optimization Based on Guided Filter. In the study of local dehazing of a single image, the ultimate goal is to remove the fog in the foggy image, so it can be regarded as a noise source. To remove the noise in the image, an appropriate image filter must be used in conjunction with the objective function. The median filter [13], mean filter [14], Gaussian filter [15], bilateral filter [16], and other image filters are common. These filters are widely used in image dehazing. In this paper, a new display image filter, namely image-guided filter, is selected in image dehazing. A local linear model is used to create the guiding filter. It works on the premise of limiting the filtered output image to the composition of a predefined guiding image. The guiding filter is similar to the bilateral filter and can maintain the edge information of the image. Moreover, because the guidance filter is more prominent in maintaining the edge information than the bilateral filter, its time complexity is low. It can increase the efficacy of image defogging.

For the guiding image q_1 and the final output image q_2, the guiding filter must finally have a local linear relationship, that is, one of the guiding images q_1, pixel s obtains the final output image q_2 through linear transformation in its domain ψ_k, which can be expressed as shown in the following equation:

$$q_2 = \psi_k b_k + \lambda_k, \tag{15}$$

where b_k and λ_k both represent linear coefficients. Obviously, equation (15) is a local linear model. According to this local linear model, it can well explain the good image edge retention of the guided filter, that is, for the edge of the object in the image. Because the gray level on both sides of the edge of the image object changes greatly, the gradient operator is usually used to detect the edge of the image, and the local linear model can be obtained according to the following equation:

$$\nabla q_2 = \nabla q_1 \times e^t. \tag{16}$$

Equation (16) indicates that as long as the edge is detected in the guide image q_1, the final output image q_2 after the guide filter processing must have the corresponding edge in the guide image q_1.

In order to determine the linear coefficients b_k and λ_k in this local linear model, the input filtered image must be restricted, and the final output image q_2 is represented by subtracting some unnecessary content (noise) from the input image and is as given in the following equation:

$$q_2 = q_1 - \hat{h}(z). \tag{17}$$

Solve the cost function in a local field ψ_k. The purpose of the cost function is to minimize the variance between the filtered input image q_1 and the final output image q_2 as much as possible. It is also necessary to maintain the local linear relationship between the guide image q_1 and the final output image q_2. Its expression is given in the following equation:

$$\beta(q_1, q_2) = \frac{D^2 \times \sum_{i=0}^{N} q_1(t) \times \psi_k}{D^2 + \sum_{i=0}^{N} q_2(t) \times \psi_k}. \tag{18}$$

In equation (18), D^2 is the regularization parameter used to constrain ψ_k. Since the guided filter is a linear model, the best solution to the cost function obtained by using the linear model is to use univariate linear regression [17], and the final linear coefficient is given by the following equations:

$$b_k = \psi_k \cdot [\beta(q_1, q_2)] \times \varpi_k, \tag{19}$$

$$\lambda_k = |P| \times \frac{q_1(t) + q_2(t)}{P(q_1, q_2)} \times \tau_k + \varphi^2, \tag{20}$$

where ϖ_k and τ_k are the expectation and variance of all pixels of the guide image q_1 in the neighborhood ψ_k, respectively; φ^2 represents the expectation of the input image in the neighborhood ψ_k; $|P|$ represents the total number of pixels in the neighborhood ψ_k of the guide image q_1.

Since the linear coefficients b_k and λ_k in the guided filter expression are obtained, the output of the guided filter can be calculated, but a pixel point s is included by multiple ψ_k neighborhoods, and different neighborhoods make the final output different. So here the expectation is calculated for all

the output results of the ψ_k neighborhood containing the pixel point s, and its expression is shown in the following equation:

$$\psi_k(s) = \arg\min\left\|\overline{b_k} + \overline{\lambda_k}\right\|^2, \tag{21}$$

where $\overline{b_k}$ and $\overline{\lambda_k}$, respectively, represent the average linear coefficient of all the pixels s in the neighborhood of ψ_k.

4.3. Local Dehazing of Single Image Based on Dark Channel Priority. The dark channel priority dehazing method [18] is a dehazing method for outdoor natural sceneries that is based on the dark channel priority law. The dark channel priority law considers that most of the fog-free outdoor natural scene images are processed by the dark channel priority. The brightness of the pixels will be close to zero, and if there are a large number of bright pixels in the dark channel image, this brightness should come from the fog in the air or the sky [19, 20]. For the fogged original image, the initial transmission map $x_i(t)$ and air color value $C(t)$ will be able to be obtained from the result of the dark channel prioritization. The higher the brightness in the transmission map, the better the passability of the scene color, which can also be understood as the closer to the viewpoint. Because the block calculation method is used in the dark channel processing, the initial transmission image has obvious squares, and at the same time, it cannot well conform to the geometric edge of the original image. Therefore, the dark channel priority method turns the process of transmission image optimization into the smallest one. The process of transforming the value function is shown in the following equation:

$$\vartheta(x) = \frac{\log x_i(t) \times C(t)}{1 - \rho^2}. \tag{22}$$

For the optimization of this quadratic function, it can be converted into the solution of a linear system of equations, and the optimized transmission map $x_i'(t)$ can be obtained by the following equation:

$$x_i'(t) = \exp\left\{\left\|O_i\right\| - L^2\right\}. \tag{23}$$

Combined with the transmittance optimization results in Section 4.2, the transmittance optimization of a single image is realized. So far, the design of the local dehazing algorithm for a single image is completed.

5. Experimental Results and Analysis

The experimental investigation is carried out to validate the effectiveness of a single image local defogging technique based on a multiobjective optimization algorithm. This experiment is implemented on an ordinary PC with windows 10 operating system, core 2 Dual Core 2.8 GHz CPU, and 8 GB system memory. The suggested approach is compared against a single image defogging algorithm based on deep learning and an infrared image defogging algorithm based on the atmospheric scattering model and retinex.

Table 1: Information entropy comparison results.

Image type	The proposed algorithm	Atmospheric scattering models and retinex	Deep learning
Indoor image 1	7.53	6.87	6.75
Indoor image 2	7.29	6.73	6.94
Indoor image 3	7.34	6.25	7.01
Indoor image 4	7.12	6.34	6.86
Outdoor image 1	7.07	5.96	6.31
Outdoor image 2	6.93	5.52	6.07
Outdoor image 3	6.99	5.21	5.91
Outdoor image 4	6.85	5.30	6.03

Table 2: CNI comparison results.

Image type	The proposed algorithm	Atmospheric scattering models and retinex	Deep learning
Indoor image 1	0.95	0.87	0.82
Indoor image 2	0.97	0.83	0.79
Indoor image 3	0.90	0.89	0.83
Indoor image 4	0.89	0.79	0.75
Outdoor image 1	0.88	0.72	0.65
Outdoor image 2	0.86	0.75	0.70
Outdoor image 3	0.82	0.69	0.67
Outdoor image 4	0.75	0.65	0.62

5.1. Experimental Sample Set. The experimental samples from the RESIDE dataset are considered. It contains both indoor (ITS) and outdoor (OTS) synthetic training sets which are not labeled real hazy images (URHI). To create the training set, 6000 synthetic hazy images were randomly selected as labeled data, among which 3000 images from ITS, 3000 images from OTS, and 4000 images from URHI were collected. RESIDE dataset contains a test subset namely a synthetic test set (SOTS). It has 500 pairs of fog and sharp images for indoor and outdoor sets. To demonstrate the usefulness of the suggested approach, it is tested on the RESIDE dataset.

5.2. Evaluation Indicators. The image dehazing effect is analyzed from two aspects: subjective evaluation and objective evaluation. The objective evaluation index includes information entropy, CNI, and the degree of tone restoration, and the subjective evaluation index refers to the visual effect of image dehazing.

(1) Information entropy: it signifies the image's information average value. The more the information entropy, the more clear is the image and has better quality.

(2) CNI is an indicator that is often used to assess the reliability and naturalness of an image. It is mainly used to judge the dehazing image, and the range is 0-1. The image nature is better when the CNI value is close to 1.

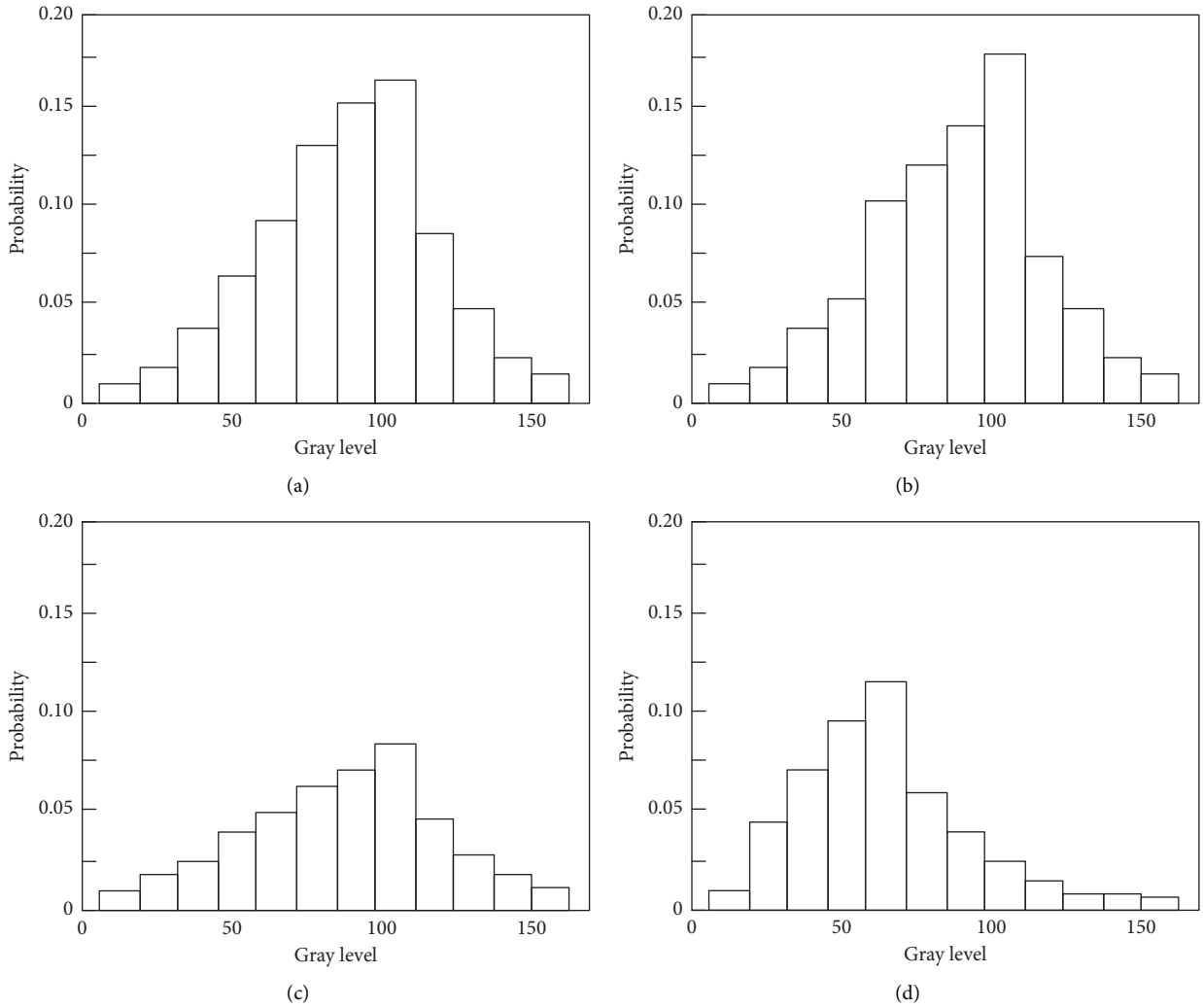


Figure 3: Comparison results of hue reduction degree. (a) Original image histogram. (b) Histogram of proposed algorithm. (c) Histogram of atmospheric and retinex Model. (d) Histogram of deep learning algorithm.

(3) The degree of tone restoration after the image is dehazed is represented by the histogram.

5.3. Objective Quality Evaluation

5.3.1. Information Entropy. The image dehazing effects of the proposed approach, the infrared image dehazing algorithm based on the atmospheric scattering model and retinex, and the single image dehazing technique based on deep learning are compared using the information entropy as the evaluation index. The results are shown in Table 1.

In aspects of image defogging objective quality, Table 1 compares the four defogging algorithms' image defogging effects in terms of objective quality. Table 1 shows that the suggested technique has a higher information entropy than the infrared image defogging algorithm based on the atmospheric scattering model and retinex model. The single image defogging algorithm based on depth learning shows that the proposed algorithm is relative to the two traditional algorithms. The defogging result has more information and clearer image. Both indoor and outdoor images have a good defogging effect.

5.3.2. CNI. The proposed algorithm, infrared image defogging algorithm based on atmospheric scattering model, retinex, and single image defogging algorithm based on depth learning are compared using CNI as the evaluation index. The results are shown in Table 2.

The proposed algorithm has a higher CNI value than the infrared image defogging algorithm based on atmospheric scattering model and retinex, as well as the single image defogging technique based on depth learning, as shown in Table 2. The highest value is 0.97, and the CNI value is greater than 0.80. The CNI value of the two traditional algorithms is low; especially, the CNI value of deep learning algorithm has a lot of room to improve. According to the above analysis, it can be concluded that the fidelity and naturalness of the image after defogging with the proposed algorithm are better, and the image is more natural.

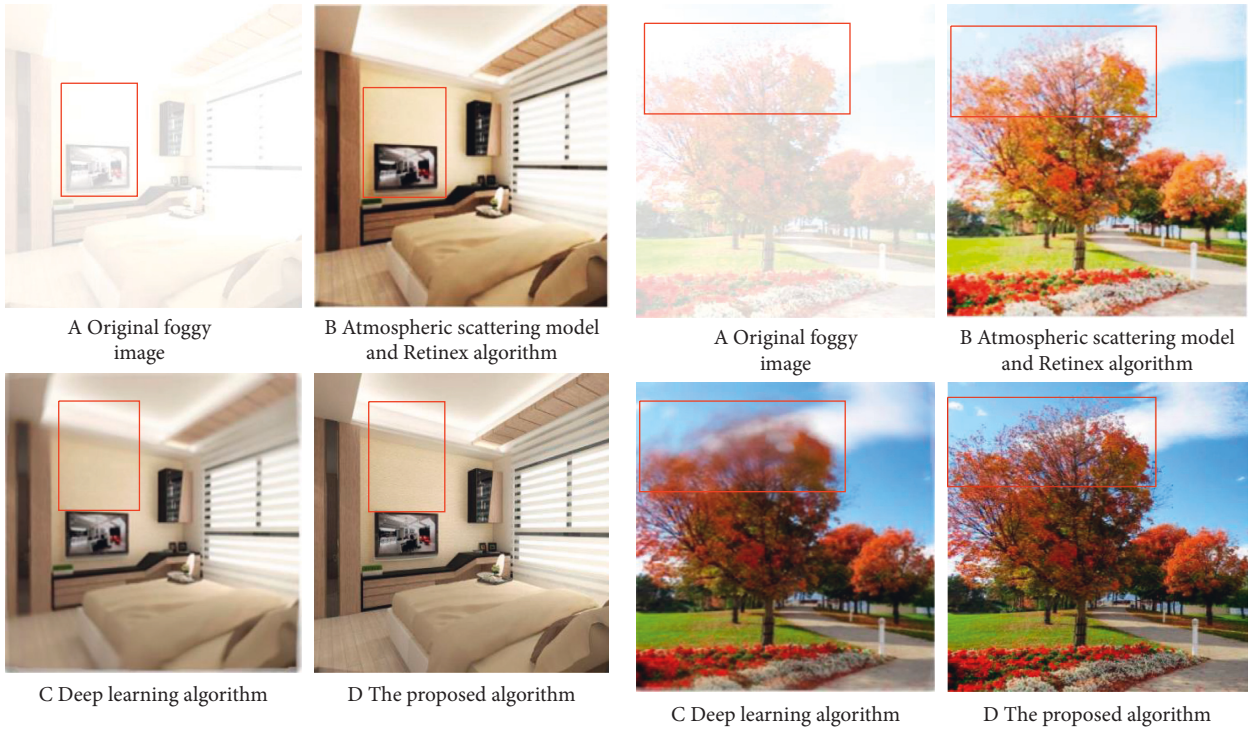


Figure 4: Comparison results of image defogging effect. (a) Defogging effect of indoor foggy image. (b) Defogging effect of outdoor foggy image.

5.3.3. Hue Reduction Degree. The image defogging effects of the proposed algorithm, the infrared image defogging algorithm based on atmospheric scattering model, retinex, and the single image defogging technique based on depth learning are compared using the hue restoration degree as the evaluation index.

The image histogram shifts to the right as a result of the fog effect on the image. An effective image defogging algorithm should return the image to its original appearance. In other words, the shape of the histogram of the original image and the defogging image should be generally consistent. Figure 3 shows the histogram of the original foggy image, the defogging image of the atmospheric scattering model and retinex algorithm, the defogging image of the depth learning algorithm, and the color component of the defogging image of the proposed algorithm.

Figure 3 shows that, when compared to the histograms in Figures 3(c) and 3(d), the histogram in Figure 3(b) maintains the shape of the original image histogram better. As a result, it can be stated that after defogging using the proposed methodology, the image tone restoration degree is higher, which has substantial benefits over the traditional algorithm.

5.4. Subjective Quality Evaluation. The above experiments have evaluated the image defogging effects of the three algorithms from an objective point of view. The image defogging effects of the proposed technique, the infrared image defogging algorithm based on atmospheric scattering model and retinex, and the single image defogging technique based on deep learning will next be compared from a subjective point of view. Figure 4 shows the effect of image defogging.

By analyzing the defogging results of the image in Figure 4, it can be found that the proposed algorithm has improved the image color, contrast, and clarity and maintained the real color of the image. Image saturation and color distortion are issues with the processing outputs of an infrared image defogging technique based on an atmospheric scattering model and retinex. The processing results of a single deep learning-based image defogging system exhibit issues with image detail loss and low image definition. By comprehensively analyzing the subjective and objective evaluation results of the above defogging algorithm, the effectiveness of the proposed algorithm is obtained. The defogging results are more in line with the visual characteristics of human eyes.

6. Conclusion

In this paper, a single image local defogging algorithm based on multiobjective optimization is proposed to improve the visual impacts of the image for better interpretations. The image reconstruction is realized by multiobjective optimization to consider multiple factors for improving the visual impacts of the images. By estimating the atmospheric light value, and improved transmittance, the transmittance of a single image is optimized. The dark channel priority method is also used to realize the local defogging of a single image. When compared to existing state-of-the-art methods, a

significant number of experimental results suggest that the proposed technique has good defogging performance. The proposed method can process normal and foggy images as well.

Data Availability

The data can be made available on valid request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Z. Tufail, K. Khurshid, A. Salman, and K. Khurshid, "Optimisation of transmission map for improved image defogging," *IET Image Processing*, vol. 13, no. 7, pp. 1161–1169, 2019.

[2] R. Ahmed, S. J. Oh, M. U. Mehmood et al., "Computer vision and photosensor based hybrid control strategy for a two-axis solar tracker –Day lighting application," *Solar Energy*, vol. 224, no. 2, pp. 175–183, 2021.

[3] V. Monga, Y. Li, and Y. C. Eldar, "Algorithm unrolling: interpretable, efficient deep learning for signal and image processing," *IEEE Signal Processing Magazine*, vol. 38, no. 2, pp. 18–44, 2021.

[4] J. Močkoř and P. Hurtík, "Approximations of fuzzy soft sets by fuzzy soft relations with image processing application," *Soft Computing*, vol. 25, no. 10, pp. 6915–6925, 2021.

[5] H. W. Dong and J. CheN, "Infrared image defogging algorithm based on atmospheric scattering model and Retinex," *Infrared Tech. journal*, vol. 41, no. 4, pp. 347–356, 2019.

[6] J. T. Zhao, "Single-image defogging algorithm based on deep learning," *Laser & Opto. Elecs. Progress journal*, vol. 56, no. 11, pp. 146–153, 2019.

[7] Y. Alharbi, P. Ge, and H. Wang, *An Image Dehazing Method Based on an Improved Retinex Theory*, Atlantis Press, Amsterdam, Netherlands, 2016.

[8] A. Dabrowska and S. Koziel, "Rapid multi-objective optimization of compact microwave components," *Applied Computational Electromagnetics Society Journal*, vol. 35, no. 11, pp. 1344-1345, 2021.

[9] K. Navdeep, V. Singh, A. Rani, and S. Goyal, "An improved hyper smoothing function based edge detection algorithm for noisy images," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 5, pp. 6325–6335, 2020.

[10] J. Dahl and D. Hyun, "Neural networks as an applied tool for ultrasound beam forming and image reconstruction," *Journal of the Acoustical Society of America*, vol. 148, no. 4, p. 2446, 2020.

[11] S. He, Z. Chen, F. Wang, and M. Wang, "Integrated image defogging network based on improved atmospheric scattering model and attention feature fusion," *Earth Science Informatics*, vol. 14, no. 4, pp. 2037–2048, 2021.

[12] T. Dong, G. Zhao, J. Wu, Y. Ye, and Y. Shen, "Efficient traffic video dehazing using adaptive dark channel prior and spatial–temporal correlations," *Sensors*, vol. 19, no. 7, p. 1593, 2019.

[13] M. Ju, C. Ding, C. A. Guo, W. Ren, and D. Tao, "IDRLP: image dehazing using region line prior," *IEEE Transactions on Image Processing*, vol. 30, pp. 9043–9057, 2021.

[14] J. Du, J. Zhang, Z. Zhang, W. Tan, S. Song, and H. Zhou, "Rcanet: image recovery network with channel attention group for image dehazing," in *Proceedings of the International Conference On Smart Multimedia*, pp. 330–337, San Diego, CA, USA, December 2020.

[15] A. Muthukrishnan, J. Charles Rajesh kumar, D. Vinod Kumar, and M. Kanagaraj, "Internet of image things-discrete wavelet transform and gabor wavelet transform based image enhancement resolution technique for IoT satellite applications," *Cognitive Systems Research*, vol. 57, no. 10, pp. 46–53, 2019.

[16] G. Song and W.-J. Pan, "Real-time rendering algorithm of aerial scene based on atmospheric scattering model," *Computer Simulation*, vol. 38, no. 8, pp. 43–47, 2021, 322.

[17] M. Kaur and S Kadam, "Bio-inspired workflow scheduling on HPC platforms," *Tehnical Journal*, vol. 15, no. 1, pp. 60–68, 2021.

[18] F. C. Pinheiro and C. G. Lopes, "A low-complexity nonlinear least mean squares filter based on a decomposable volterra model," *IEEE Transactions on Signal Processing*, vol. 67, no. 21, pp. 5463–5478, 2019.

[19] A. S. Ajith and T. Latha, "Neutrosophic data formation using Gaussian filter based costas coding for wireless communication systems," *Cognitive Systems Research*, vol. 57, no. 10, pp. 115–125, 2019.

[20] M. Kaur, S. Kadam, and N. Hannoon, "Multi-level parallel scheduling of dependent-tasks using graph-partitioning and hybrid approaches over edge-cloud," *Soft Computing*, vol. 26, no. 11, pp. 5347–5362, 2022.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 5302826, 9 pages
https://doi.org/10.1155/2022/5302826

Research Article

State Evaluation Method of Distribution Equipment Based on Health Index in Big Data Environment

Fei Xue, Xutao Li, Xiaoli Wang, Chao Wang, Hongqiang Li, and Di Zhang

Electric Power Research Institute of State Grid Ningxia Electric Power Co., Ltd., Yinchuan 750001, China

Correspondence should be addressed to Chao Wang; chaowang311@live.com

Received 2 June 2022; Revised 23 June 2022; Accepted 27 June 2022; Published 14 July 2022

Academic Editor: Mukesh Soni

Aiming at the problems of time-consuming and low accuracy in the existing state evaluation methods of distribution equipment, a state evaluation method of distribution equipment based on health index in big data environment is proposed. Firstly, in order to optimize the time-consuming of big data analysis on large-scale and distributed clusters, a distribution equipment condition monitoring data platform in big data environment is designed, and a hive based relational online analysis method (ROLAP) is proposed. Secondly, the health index (HI) is introduced as the evaluation index to evaluate the health status of distribution equipment. According to the different influence degree of different fault factors on the equipment status, a comprehensive multifactor fault rate correction model is obtained, and the method based on success flow is used to solve the model to improve the accuracy of state evaluation. Finally, experiments show that when the data volume of distribution equipment is 60 GB, the time of the proposed method is only 30.0 s, which is far lower than 73.6 s and 82.5 s of the comparison method. The evaluation accuracy of the proposed method is 95.1%, while the evaluation accuracy of the comparison method is only 82.4% and 73.1%, respectively. Therefore, the proposed method can effectively improve the efficiency of distribution equipment condition evaluation.

1. Introduction

As an important part of smart grid construction, distribution network undertakes the responsibility of directly undertaking intermittent loads such as end users, distributed power generation, and electric vehicles. However, due to its complex operating conditions, uneven equipment level, and complex network structure, it is impossible to form a reliable structure similar to power generation and transmission system [1, 2]. Research shows that more than 85% of power failures occur on the distribution network side [3–5]. Therefore, the condition evaluation of distribution equipment is not only helpful to improve the efficiency of daily management of power system but also helpful to the construction of smart grid and the application of modern distribution technology [6–9].

British power grid researchers compared the operation status of power grid equipment with human health and proposed the concept of electrical equipment health index by using the relevant theoretical knowledge of human health index (HI) and analogy with the deterioration law of electrical equipment [10–12]. For the fear that the extended service will increase the failure rate of the power grid, damage the reliability of system power supply, and have an adverse impact on users, it is necessary to formulate a more reliable maintenance plan of power grid equipment based on it. Subsequently, this concept was introduced into the research of power supply reliability, asset management, and many other fields, which is helpful to the risk control of power grid and avoid the defects of blind maintenance and excess maintenance of equipment. To a certain extent, it reduces the operation and maintenance cost, optimizes the resource allocation, and improves the security of power system operation and power supply reliability [13–15].

Traditional power grid data processing technology can achieve good results for small and medium-sized data, such as power grid computing, P2P computing, and cluster computing. The online monitoring data of power equipment has the characteristics of large amount of data, various types of values, multiple changes, and low value density [16, 17].

As a new computing model framework, cloud computing is more suitable for processing massive online monitoring data of distribution equipment [18–20].

At present, although China has carried out relevant research on the health status evaluation of distribution equipment, most of the domestic research is focused on high-voltage transformers. At present, distribution equipment health evaluation not only lacks systematic theories and tools but also its research has not been extended from the equipment level to the network level. How to evaluate the health status of a large number of distribution equipment and complex real-time dynamic distribution network is not only a new topic in the development of modern power grid but also a great challenge in the development of modern smart grid. Reference [21] proposed an equipment fault prediction method based on a similar density array, which uses logistic fast minimum error entropy algorithm to predict equipment fault risk considering weather factors. Aiming at the complex mechanism of secondary equipment fault in intelligent substation, reference [22] combines analytic hierarchy process and antientropy theory to propose a state evaluation method of secondary equipment based on weighting method. Reference [23] uses entropy weight method to normalize each index, calculate the index weight, and finally form a comprehensive evaluation model. Reference [24] proposes a state sensing method of intelligent substation secondary system based on FCE and deep convolution neural network (DCNN), but training and optimization will occupy more resources. According to the standard and expert experience, the condition evaluation standard is scored in reference [25], and the state is divided into three levels in the form of health index and maintenance task. Reference [26] comprehensively considered hardware information and human factors combined with a fuzzy iterative method and weighted expert database and proposed a multi-source information fusion state evaluation model. Reference [27] established the distribution network health index model, proposed the combination of network equivalence method and goal-oriented method, and used the GO method in reliability analysis to analyze and solve the model.

However, in the context of big data, the above methods often have the problem of low accuracy, and big data management also has the problems of time-consuming, poor storage performance, and low analysis efficiency. Aiming at the problems existing in the above methods, a state evaluation method of distribution equipment based on health index in big data environment is proposed. The innovations are as follows:

(1) A relational online analysis method based on hive is proposed, and a distribution equipment condition monitoring data platform in the big data environment is designed, which improves the efficiency of the method when performing big data analysis on large-scale and distributed clusters.

(2) HI is introduced as a good or bad index to evaluate the health status of distribution equipment, and the method based on success flow is used to solve the model, which improves the accuracy of status evaluation.

2. Design of Distribution Equipment Condition Evaluation Architecture under Big Data

Traditional equipment condition evaluation mostly adopts traditional data storage and analysis methods, which cannot carry out complex analysis on the collected data. The proposed method integrates distributed data storage and big data analysis, which brings a new idea to the state evaluation of distribution equipment. The proposed framework is shown in Figure 1.

2.1. Data Acquisition. The data acquisition layer collects the data through state access controller (CAC) and sensors, and transmits the collected data to the state access network shutdown (CAG) through web services. Due to the complexity of data source types, it is necessary to use Sqoop, an open-source tool, to ETL (extract, transform, and load) the required data, and then use a unified structure to store the data after data association and aggregation. After completing the tasks of query, calculation, and statistical analysis, Sqoop can also export the analyzed results to the external relational database my SQL for users to view. Due to the huge amount and complex types of distribution equipment data, Sqoop needs to be used to extract, convert, and load. By cross sharing the data of multiple information management systems, analyzing massive multi-source heterogeneous data, mining, and analyzing the state quantity indicators with strong correlation, such as equipment defects, equipment faults, and equipment key health status

2.2. Data Storage. The data storage layer integrates HDFS and MySQL. Big data with unified and standardized status monitoring is stored in the distributed file system HDFS. MySQL is mainly used to store the model information of distribution equipment condition monitoring and manage Hive metadata.

2.3. Data Analysis. In the data analysis layer, a distributed ROLAP analysis method for distribution equipment condition monitoring data is designed. ROLAP service supports larger user groups and data volume, which is often used in occasions with high requirements for these capacities. Hive is a kind of distributed ROLAP services. It completes operations such as roll up and drill down through the MapReduce decomposition task. Parallel computing mainly depends on MapReduce architecture. Figure 2 is MapReduce data analysis flow chart.

Taking the big data of distribution equipment as the research object, the MapReduce architecture is analyzed and processed according to the following steps:

Step 1. The big data of distribution equipment is randomly divided into several data blocks such as split0, split1, and split2. Each data block is assigned to the Map node in the

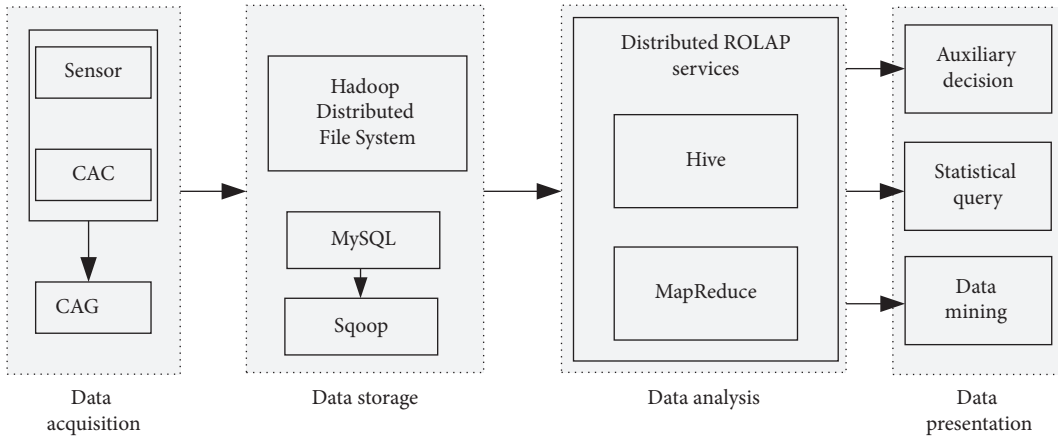


Figure 1: Architecture of distribution equipment condition monitoring platform under big data.

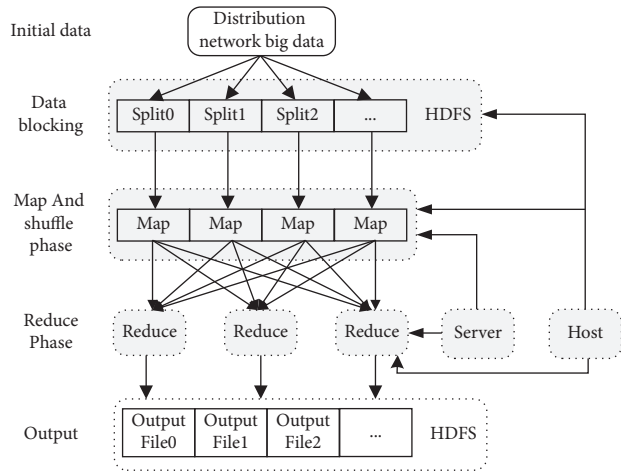


Figure 2: MapReduce data analysis flow chart.

form of < key1, value1 > key value pairs for parallel task processing, and these Map tasks are carried out at the same time;

Step 2. After the map task is completed, many intermediate results in the form of < key2, value2 > key value pairs are generated. The intermediate results are sent to the shuffle process for aggregation and processing. The key value pairs with the same key value are formed into a cluster and transmitted to the Reduce node in the form of key value pairs < key2, value2 >;

Step 3. Start the reduce task. According to the key value pairs < key2, value2 > passed from the Shuffle process to the Reduce node, the final sorting operation is carried out for the key value pairs with the same key value to form the key value pair result of < key3, value3 >;

Step 4. The calculation results of the reduce node are summarized and output as the final results.

In order to better realize computing localization, the slave node data node of HDFS and the slave node task tracker of MapReduce are fused and bound to a slave node, which is convenient to directly read data locally for analysis and processing and complete the specific interpretation of the data within the data processing time.

2.4. Data Presentation. The data presentation layer integrates the functions of statistical query, auxiliary decision-making, data mining, and so on. At the same time, it also provides the analysis, evaluation and prediction functions of various distribution equipment status information and provides condition monitoring data for other related systems.

3. State Evaluation Method of Distribution Equipment

3.1. Comprehensive Multifactor Failure Rate Correction Model. Distribution network equipment is mainly divided into five categories: overhead line, cable, distribution transformer, disconnector, and circuit breaker, and the failure rate of the i_{th} category equipment is expressed by $\lambda_i\,(1 \le i \le 5)$. Through the different influence degree of different fault factors on the equipment state, combined with the operation state and working environment of the equipment, the correction coefficient matrix of various equipment fault rates is obtained [25]. The weights of three types of fault factors and the correction coefficients of equipment faults are shown in Table 1.

The weight matrix of various equipment fault factors obtained by analytic hierarchy process is shown in the following equation:

$$W = \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{31} & w_{33} \end{bmatrix}, \tag{1}$$

where w_{ij} represents the weight of class i_{th} equipment failure caused by the j_{th} factor, $1 \le j \le 3$.

The failure rate correction coefficient matrix of various equipment is

Table 1: Weight and correction coefficient of equipment failure factors.

Failure factors	Weight	Correction factor
Heavy load or overload	w_1	m_1
Strong wind or heavy rain	w_2	m_2
Lightning strike	w_3	m_3

$$S = \begin{bmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{31} & s_{33} \end{bmatrix}, \tag{2}$$

where s_{ij} is the correction factor.

3.1.1. Overload Correction. Generally, the equipment is allowed to operate under heavy load and overload for a short time, but the longer and higher the degree of heavy load and overload operation, the higher the equipment failure rate. When the equipment load rate does not exceed the rated load rate o, the visual equipment failure rate is zero. The definition L_P is used to characterize the overload degree of the equipment. At the same time, combined with the definition of definite integral and exponential function, the overload correction coefficient is calculated as follows:

$$\begin{aligned} L_p &= \begin{cases} 0, & L_R \le s, \\ L_R - s, & L_R > s, \end{cases} \\ u_{i1} &= \frac{(L_M - s)e^{L_p} - 1}{\int_0^{L_M - s}\left(e^{L_p} - 1\right)\mathrm{d}L_p}, \end{aligned} \tag{3}$$

where L_R is the real-time load rate, L_M is the maximum allowable load rate, and u_{i1} is the overload correction factor.

3.1.2. Failure Factors of Strong Wind and Heavy Rain. The correction factor for the fault factors of heavy wind and heavy rain of distribution equipment is u_{i2}, which can be obtained according to the average precipitation data of recent years counted by the local meteorological department. The calculation formula is as follows:

$$\begin{aligned} u_{i2} &= \frac{12\overline{H}_m}{\sum_{i=1}^{12}\overline{H}_i}, \\ \overline{H}_m &= \frac{\sum_{i=k}^{k+n}\overline{H}_{mj}}{n}, \end{aligned} \tag{4}$$

where m represents the month of the year to be evaluated, and $\overline{H}_m$, $\overline{H}_i$, and $\overline{H}_{mj}$ represent the monthly average precipitation.

3.1.3. Lightning Strike Correction. The correction factor of meteorological failure factors of lightning strike of distribution equipment is u_{i3}, which can be obtained according to the average lightning strike data in recent years counted by the local meteorological department. The calculation formula is as follows:

$$\begin{aligned} u_{i3} &= \frac{12\overline{V}_m}{\sum_{i=1}^{12}\overline{V}_i}, \\ \overline{V}_m &= \frac{\sum_{i=k}^{k+n}\overline{V}_{mj}}{n}, \end{aligned} \tag{5}$$

where $\overline{V}_m$ refers to the average monthly lightning stroke times in the m_{th} month of the year to be evaluated, $\overline{V}_i$ refers to the average monthly lightning stroke times in the i_{th} month of the year to be counted, and $\overline{V}_{mj}$ refers to the average monthly lightning stroke times in the m_{th} month of the j_{th} year of the year to be counted.

According to the equipment failure factor weight matrix and equipment failure rate correction coefficient, a comprehensive multi factor failure rate correction matrix $\mathbf{V}$ can be obtained:

$$\begin{aligned} \mathbf{V} &= \mathbf{W} \cdot \mathbf{S} \\ &= (v_1, v_2, v_3), \end{aligned} \tag{6}$$

where v_i represents the comprehensive correction coefficient.

3.2. Equipment Health Index Model. The actual equivalent service life of matching equipment is calculated according to the service age fallback theory, and the Weibull distribution parameters are estimated according to the collected data to obtain the equipment failure rate λ_i. Then, using the weight matrix and correction matrix of equipment fault factors, the correction coefficient of corresponding equipment is calculated, and then corrected. The real-time health index of the equipment can be expressed as follows:

$$HI_n = \frac{1}{E} \times \ln\frac{v_i \times \lambda_i}{K}, \tag{7}$$

where E is the curvature coefficient, K is the proportion coefficient, and v_i is the comprehensive correction coefficient of distribution equipment failure rate. Because the statistical data of power companies in different regions are different, it is difficult to obtain the actual common failure rate data. The scale coefficient and curvature coefficient in this paper are determined by reference [26]. According to the above formula and the λ_u and $\lambda_{\min}$ of various equipment, the values of proportion coefficient K and curvature coefficient E of transformer, cable, overhead line, and other equipment can be obtained, and then the functional relationship between equipment HI and failure rate can be obtained.

3.3. Solving Model Based on Success Flow Method. GO can well reflect the logic and functional relationship between components and systems. It is an analysis method based on the system structure diagram and aimed at the successful

operation of the system. It is mainly used for system simulation.

When the GO method is introduced into the reliability evaluation of distribution network, the fault rate data of equipment are preprocessed first, so that the fault rate data of all equipment appear in the form of probability. Then, with the help of success oriented technology, each user can delete various fault combinations according to the state dependency of the equipment and obtain the reliability index of the equipment. When the equipment failure rate is λ, the successful operation probability of the equipment is converted to $P = 1 - (\lambda/N)$.

The health index and failure probability satisfy a kind of exponential relationship, that is, the failure probability P_1 is

$$P_1 = e^{\left[-(HI-10/HI)^2\right]}. \tag{8}$$

The health index of components k is transformed by GO to obtain the probability of successful operation:

$$P_{ij,k} = f\left(HI_{ij,k}\right), \tag{9}$$

where $HI_{ij,k}$ is the health index of the element k in the j_{th} branch feeder of the i_{th} layer.

According to the obtained system hierarchy diagram, each branch feeder is analyzed in series from the lowest level, and an equivalent successful operation probability is used to replace the successful operation probability of all equipment on the same line:

$$P_{ij} = \prod_{k \in ij} P_{ij,k}. \tag{10}$$

If some important users use double circuit lines, parallel calculation shall be adopted, that is,

$$P_{ij}^1 = 1 - \left(1 - P_{ij,1}\right)\left(1 - P_{ij,2}\right), \tag{11}$$

wherein $P_{ij,1}$ is the successful operation probability of the first loop line in the double loop line, and $P_{ij,2}$ is the successful operation probability of the second loop line.

Put the calculated result of the i_{th} layer as an equivalent element M_i into the $i-1$ layer, as shown in Figure 3.

The probability of successful operation of components M_i is calculated as follows:

$$P_{M_i} = \sum_{j=1}^{n}\left(a_{ij}, P_{ij}\right), \tag{12}$$

where P_{M_i} is the successful operation probability of equivalent components, a_{ij} is the proportion of the j_{th} branch feeder in the total load of the i_{th} layer, and $\sum_{j=1}^{n} a_{ij} = 1$.

Repeat the above process until it reaches the top level of the network to obtain the successful operation probability P_{M_i}, which is inversely transformed into the health index. The transformation formula is

$$HI_{\text{total}} = \left(\frac{10}{\left(1 + \sqrt{-\ln\left(1 - P_{M_i}\right)}\right)}\right). \tag{13}$$

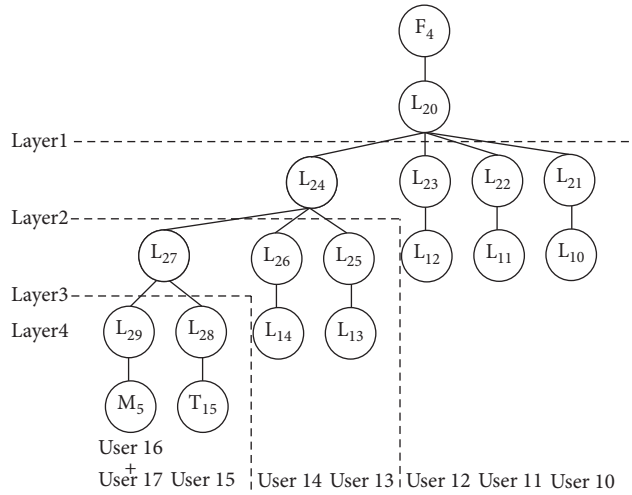


Figure 3: IEEE RBTS-Bus2 system feeder 4 system upper layer equivalent diagram.

4. Experiment and Analysis

4.1. Construction and Configuration of Big Data Platform. The main process of building and configuring the big data platform of distribution equipment is as follows:

(1) Install Ubuntu on the server to complete the implementation of virtual Linux under Windows environment, and determine the parameter configuration of virtual machine according to the data analysis requirements of power distribution equipment;

(2) Count up and down on Ubuntu, install JK, and configure Java Home and other relevant parameters as needed to establish a Java environment;

(3) Download and install Hadoop and create platform users;

(4) Configure SSH to ensure shared access between nodes and ensure security during remote management;

(5) Modify the configuration file one by one, including modifying the global configuration file (conf/core site. xml), yarn configuration file (conf/yarn site. XML), and creating and modifying the MapReduce configuration file (conf/mapred site. xml) to complete the configuration of the analysis platform environment;

(6) Format the HDFS system, start the platform environment, ensure the normal operation of each node of the platform through multiple tests, load the data to be analyzed, update the storage and output directories, and complete the integration of data flow and the platform. Figure 4 shows the construction and configuration process of big data platform.

After the above main steps, the distribution equipment big data analysis platform has been preliminarily built. The main software version parameters used are as follows:

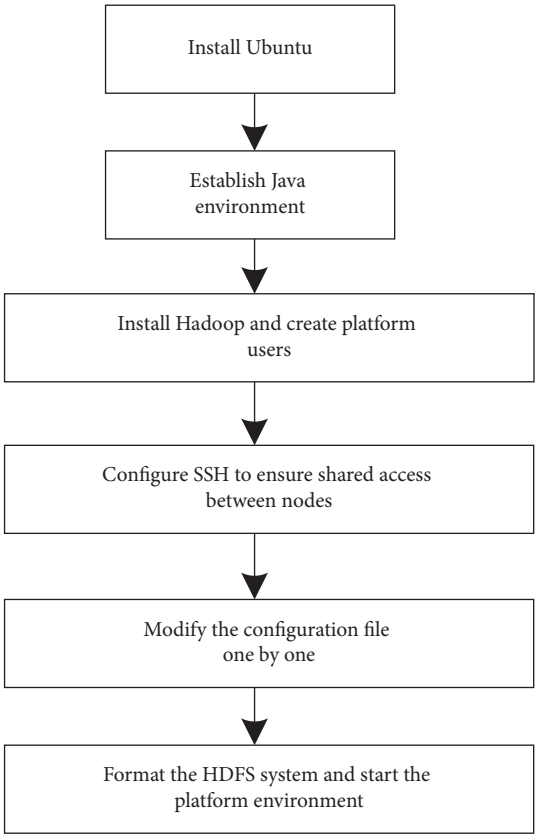


Figure 4: Construction and configuration process of big data platform.

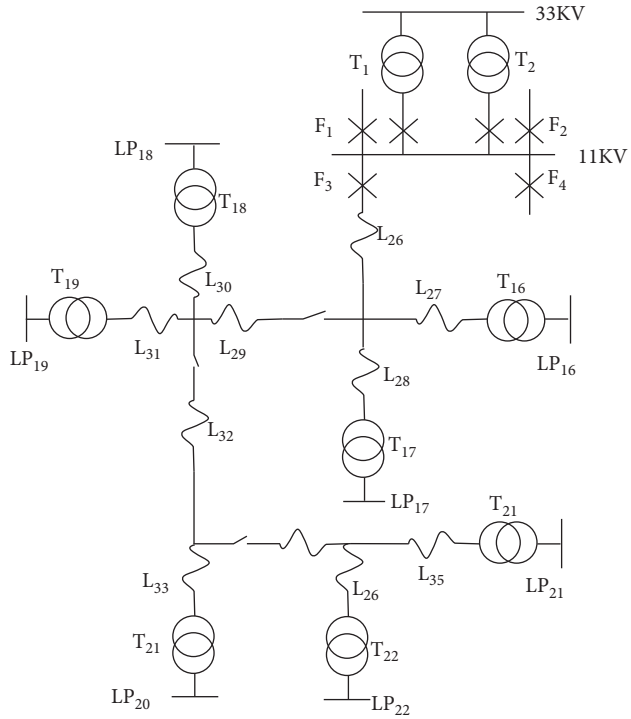


Figure 5: Systems architecture of IEEE RBTS-Bus2 feeder 4.

operating system: Ubuntu 12.04; Hadoop version: hadoop-2.7.2; JDK version: jidk-7u79-jinux-i586.

4.2. Structure Diagram of Simulation System. The simulated radial distribution system structure IEEE RBTS-Bus2 feeder 4 system is shown in Figure 5. The distribution network includes 12 distribution lines, 19 equipment such as 7 transformers, 1 circuit breaker, and 7 terminal load points. The failure rate of circuit breaker is 0.002 times/year and that of distribution transformer is 0.015 times/(a.km). The failure rate of each load point is shown in Table 2. Among them, load points 16 and 22 are commercial loads, with an average load of 0.559 MW. Load points 17–19 are residential load, with an average load of 0.492 MW. Load points 20 and 21 are secondary loads of government departments, with an average load of 0.587 MW. The length of lines 26, 31, 33, and 36 is 0.9 km, the length of lines 27, 29, 32, and 35 is 0.75 km, and the length of lines 28, 30, and 34 is 0.65 km.

4.3. Influence of Device Health Index Change on Network Health Index. Select the equipment on the same layer and different layers, respectively, and simulate the state of the equipment on each layer from good to bad (the corresponding health index changes from 1 to 8). Figure 6 shows the impact of equipment degradation on the network health index at the same layer, and Figure 7 shows the impact of equipment degradation at different layers on the network health index. Even if the health index of equipment is the same, different levels in the system have different effects on the overall health index of distribution network. The higher the level of equipment, the greater the contribution of its health state to the health index of distribution network. The reason is that the higher the level of equipment in the system, the more serious the loss caused by the failure. If equipment 26 (overhead line) fails, it will directly affect all load points connected with it, that is, it will have the greatest impact on the network. Compared with high-level equipment, the failure of low-level equipment T_{21} will only affect the load point LP_{21}. Therefore, the higher the level of equipment health in the network, the higher the impact on the overall network.

4.4. Comparison of Different Methods. Figure 8 shows the running time consumption of different methods when the amount of data changes. It can be seen that the time of abnormal detection of power equipment is different for different amounts of data. When the amount of data is 10 GB, Reference [26] takes 20.0 s, and Reference [27] takes 25.3 s. The time of the proposed state evaluation method of distribution equipment based on health index in big data environment is only 8.5 s. When the data volume is 60 GB, Reference [26] takes 73.6 s, the method of Reference [27] takes 82.5 s, and the time of the proposed method is only 30.0s. This is because this paper divides the index types, modifies the failure rate model by integrating multiple factors, designs the distribution equipment condition monitoring data platform under the big data environment, and proposes a relational online analysis method based on

Table 2: Reference values of failure rates at load points.

Line load point failure rate	16	17	18	19	20	21	22	Average value
Overhead line	0.241	0.245	0.239	0.257	0.245	0.249	0.238	0.240
Cable line	0.153	0.152	0.157	0.149	0.153	0.150	0.149	0.150

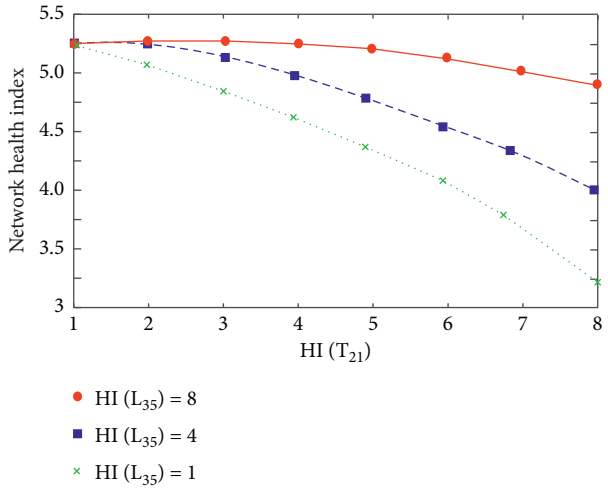


Figure 6: Effect of equipment deterioration on network health index at the same layer.

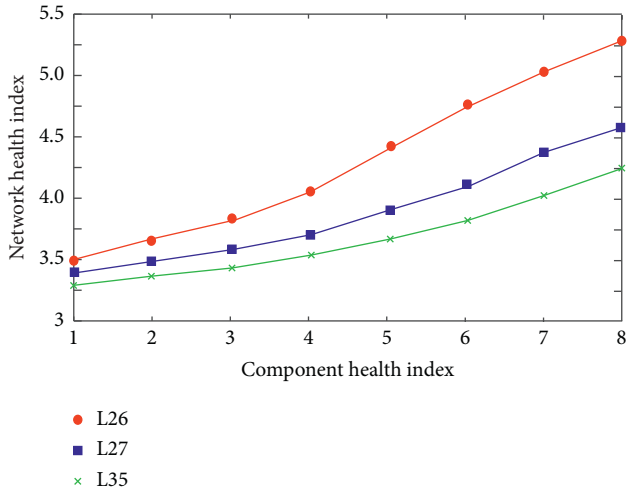


Figure 7: Effect of equipment degradation in different layers on network health index.

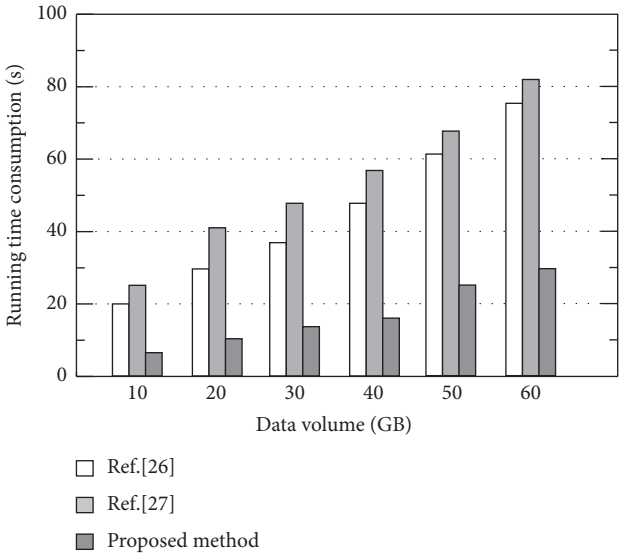


Figure 8: Running time consumption of different methods when the amount of data changes.

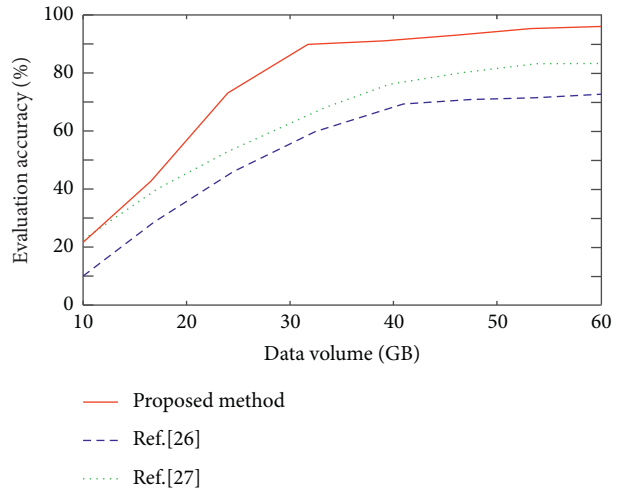


Figure 9: Evaluation accuracy of different methods when the amount of data changes.

Hive, which improves the analysis efficiency of the method when performing big data analysis on large-scale and distributed clusters, so the state evaluation takes less time.

The accuracy rate is the proportion of correctly evaluated samples to all samples. In order to verify the accuracy of the proposed method, Reference [26] method and Reference [27] method are compared with the proposed method. Figure 9 shows the comparison of evaluation accuracy under different methods. Under different data sets, this method has the highest evaluation accuracy. When the data volume of distribution equipment is 10 GB, the evaluation accuracy of the proposed method is 22.1%, the evaluation accuracy of Reference [26] method is 22.3%, and the evaluation accuracy of Reference [27] method is 10.0%. When the data volume of distribution equipment is 60 GB, the accuracy of the proposed method is 95.1%, the evaluation accuracy of Reference [26] method is 82.4%, and the evaluation accuracy of Reference [27] method is only 73.1%. The proposed method considers the influence of different fault factors on the state of equipment, introduces the health index as the index to evaluate the state of distribution equipment, and uses the

method based on success flow to solve the model, which improves the accuracy of state evaluation.

5. Conclusion

Aiming at the problems of time-consuming and low analysis efficiency of existing methods in distribution equipment condition evaluation under the background of big data, a distribution equipment condition evaluation method based on Hi under the environment of big data is proposed. The condition monitoring data platform of distribution equipment in big data environment is designed, and a relational online analysis method based on hive is proposed to improve the efficiency of big data analysis. Hi is introduced as the evaluation index of the health state of distribution equipment, the fault rate model is modified by multiple factors, and solved by GO method. This method can effectively improve the efficiency of distribution equipment condition evaluation.

Limited by the current laboratory hardware conditions and the scale of experimental data obtained, the data set used in the experiment only reaches the scale of GB. In the next step, we are going to carry out experimental research on parallel analysis of TB data. In addition, when establishing the distribution network health index model, this paper selects the widely used radial distribution network so that the follow-up research can take the ring network and other complex networks as the research object.

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.

References

[1] S. Li and J. Li, "Condition monitoring and diagnosis of power equipment: review and prospective," *High Voltage*, vol. 2, no. 2, pp. 82–91, 2017.

[2] X. Jiang and G. Sheng, "Research and application of big data analysis of power equipment condition," *High Voltage Engineering*, vol. 44, no. 4, pp. 1041–1050, 2018.

[3] J. Zhou, G. Luo, C. Hu, and Y. Chen, "A Classification Model of Power Equipment Defect Texts Based on Convolutional Neural network," *International Conference on Artificial Intelligence and Security*, Springer Cham, New York, NY, USA, 2019.

[4] V. N. Kuryanov, M. Sultanov, E. Kuryanova, and E. M. Skopova, "Mathematical model of the processes of restoration of power equipment in power systems by criterion of the index of technical condition," in *Proceedings of the Journal of Physics: Conference Series*, pp. 41–52, IOP Publishing, Bristol, UK, 2020.

[5] S. Oleg, K. Oleksii, and P. Serhii, "Informational and Analytical System for Diagnostics of the Electric Power Equipment Condition," in *Proceedings of the 2020 IEEE 7th International Conference on Energy Smart Systems (ESS)*, pp. 105–110, IEEE, Kyiv, Ukraine, May 2020.

[6] P. Agoris, S. Meijer, E. Gulski, and J. J. Smit, "Evaluation of On-Line Insulation Condition Assessment Techniques for Power transformers," in *Proceedings of the Second International Symposium on High Voltage and High Power Tests, Measurements and Certification of Electric Power Equipment*, pp. 1–6, Baile Herculane, Romania, October 2004.

[7] J. Qiu, H. Wang, D. Lin, B. He, W. Zhao, and W. Xu, "Nonparametric regression-based failure rate model for electric power equipment using lifecycle data," *IEEE Transactions on Smart Grid*, vol. 6, no. 2, pp. 955–964, 2015.

[8] W. Zhao and M. Cui, "Real-time health status evaluation for electric power equipment based on cloud model," *International Journal of Simulation and Process Modelling*, vol. 15, no. 1/2, p. 134, 2020.

[9] X. Zhang, M. Zhang, and Q. Hu, "Research on the application of artificial intelligence in operation and maintenance for power equipment," in *Proceedings of the IOP Conference Series: Earth and Environmental Science*, pp. 206–214, IOP Publishing, Bristol, UK, 2020.

[10] Y. Liu, G. Li, K. Gao et al., "Fundamental frame to draft guid for condition maintenance of electric power equipment," *Power System Technology*, vol. 27, no. 6, pp. 64–67, 2003.

[11] N. L. Fantana and L. Pettersson, "Condition-based evaluation: a new platform for power equipment life management," *ABB Review*, vol. 3, no. 4, pp. 45–54, 2000.

[12] H. Liu, Y. Wang, L. Zhou, C. Yufeng, and D. Xiuming, "An Optimization Method of Maintenance Strategy for Power equipment," in *Proceedings of the 2016 International Conference on Condition Monitoring and Diagnosis (CMD)*, pp. 940–943, IEEE, Xi'an, China, September 2016.

[13] H. Okubo, "Recent Progress and Future Perspective on Condition Monitoring and Diagnostic Techniques for Power Equipment in Japan," in *Proceedings of the 2008 International Conference on Condition Monitoring and Diagnosis*, pp. 13–18, IEEE, Beijing, China, April 2008.

[14] X. Tian, C. Li, and B. Zhao, "A novel classification model SA-MPCNN for power equipment defect text," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 20, no. 6, pp. 1–21, 2021.

[15] E. A. L. Vianna, A. R. Abaide, L. N. Canha, and V. Miranda, "Substations SF6 circuit breakers: reliability evaluation based on equipment condition," *Electric Power Systems Research*, vol. 142, no. 2, pp. 36–46, 2017.

[16] S. Zhang, X. Chen, S. Zhang, and H. Xiuzhen, "The Comparison between Cloud Computing and Grid computing," in *Proceedings of the 2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*, pp. 540–549, IEEE, Taiyuan, October 2010.

[17] D. S. Markovic, D. Zivkovic, I. Branovic, R. Popovic, and D. Cvetkovic, "Smart power grid and cloud computing," *Renewable and Sustainable Energy Reviews*, vol. 24, no. 6, pp. 566–577, 2013.

[18] S. Bera, S. Misra, and J. J. P. C. Rodrigues, "Cloud computing applications for smart grid: a survey," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 5, pp. 1477–1494, 2015.

[19] S. Rusitschka, K. Eger, and C. Gerdes, "Smart Grid Data Cloud: A Model for Utilizing Cloud Computing in the Smart Grid domain," in *Proceedings of the 2010 First IEEE International Conference on Smart Grid Communications*, pp. 483–488, IEEE, Gaithersburg, MD, USA, October 2010.

[20] X. Duan, C. Long, S. Feng, R. Luo, Y. Gao, and Y. Sun, "Status Evaluation of Power Secondary Equipment Based on Big Data of monitoring," in *Proceedings of the 2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, pp. 944–949, IEEE, Chengdu, China, May 2019.

[21] W. Mo, Z. Xu, K. Zhou, L. Luan, W. Fan, and H. Liu, "Similar-density-array-based equipment outage prediction method for distribution network considering weather factors," in *Proceedings of the 2021 IEEE 10th data driven control and learning systems conference (DDCLS)*, pp. 801–806, IEEE, Suzhou, China, May 2021.

[22] J. Zhou, D. Nan, J. Tan, L. Zhang, Q. Zhao, and X. Ma, "State Evaluation of Secondary Equipment Based on Combination Weighting and Cloud model," in *Proceedings of the 2020 Chinese Automation Congress (CAC)*, pp. 6528–6532, IEEE, Shanghai, China, November 2020.

[23] Y. Cui, L. Luan, Y. Liu, W. Mo, and H. Wang, "State Assessment on Distribution Network Equipment Oriented by Big Data Visualization," in *Proceedings of the 2018 IEEE International Conference on High Voltage Engineering and Application (ICHVE)*, pp. 1–4, IEEE, Athens, Greece, September 2018.

[24] H. Li, J. Zhu, and L. Luo, "State Perception Method of Intelligent Substation Secondary System Based on FCE and DCNN," in *Proceedings of the PURPLE MOUNTAIN FORUM 2019-International Forum on Smart Grid Protection and Control*, 2020.

[25] W. Luejai, T. Suwanasri, and C. Suwanasri, "Condition Evaluation of High Voltage Transmission Line in Thailand," in *Proceedings of the International Symposium on High Voltage Engineering*, pp. 171–183, Springer Cham, New York, NY, USA, November 2019.

[26] X. Yang and R. Wang, "Research into Power Distribution Network Condition Evaluation Method Based on Equipment Health Index," in *Proceedings of the 2nd International Conference on Electrical and Electronic Engineering*, pp. 78–84, Atlantis Press, Amsterdam, Netherlands, July 2019.

[27] N. Wang and F. Zhao, "An assessment of the condition of distribution network equipment based on large data fuzzy decision-making," *Energies*, vol. 13, no. 1, p. 197, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 4255835, 11 pages
https://doi.org/10.1155/2022/4255835

Research Article

Swarm Intelligence Algorithms for Optimal Scheduling for Cloud-Based Fuzzy Systems

Lulwah AlSuwaidan,[1] **Shakir Khan**,[1] **Riyad Almakki**,[1] **Abdul Rauf Baig**,[1] **Partha Sarkar**,[2] **and Alaa E. S. Ahmed**[1,3]

[1]*College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh, Saudi Arabia*
[2]*Department of Electronics and Communication Engineering, National Institute of Technology, Durgapur, West Bengal, India*
[3]*Electrical Engineering Department, Faculty of Engineering at Shoubra, Benha University, Cairo, Egypt*

Correspondence should be addressed to Shakir Khan; sgkhan@imamu.edu.sa

Received 29 April 2022; Accepted 7 June 2022; Published 14 July 2022

Academic Editor: Mukesh Soni

A fuzzy cloud resource scheduling model with time-cost constraints is built using fuzzy triangular numbers to represent uncertain task execution time. Task scheduling reduces total time and cost spent on a project. It connects virtual machines and functions. Particle swarm optimization (HPO) is used to plan cloud resources (HSOA). The approach uses orthogonal particle swarm initialization to increase the quality of the initial particle exploration, rerandomization to regulate the particle search range, and real-time updating of inertia weights to control particle speed. The suggested problem model and optimization approach are evaluated using random simulation data provided by the CloudSim simulation platform. Less overall execution time and a lower cost are shown to have fast convergence and solution capabilities in experiments.

1. Introduction

Cloud resource scheduling is the core content of cloud computing. In order to minimize the completion time, literature [1] established a corresponding cloud computing scheduling model. Reference [2] enables cloud service providers to obtain maximum benefits under the premise of ensuring service quality in the scheduling process. Reference [3] established a real-time scheduling system in order to minimize energy consumption. These studies used a deterministic execution time. However, due to the unpredictability of task execution, the execution time can only be an estimated value before the task execution is completed, which leads to uncertainty in the execution time of the task. Cases with time uncertainty are considered. When solving practical problems, mathematical models can be divided into three categories.

The fuzzy mathematical model is required to handle the uncertain problem that has to be estimated, as does the mathematical model of randomness, the mathematical model of randomness, and the mathematical model of fuzziness. Reference [4] looked at the temporal complexity of illness detection using a fuzzy genetic system, while [5] looked at task scheduling in the cloud using a fuzzy neural network method. In this research, triangular fuzzy numbers [6] are used to represent the unknown task execution time, and a fuzzy cloud resource scheduling model is developed. The issue of cloud resource scheduling is NP-complete [7, 8], and no efficient polynomial method exists. Scholars in the United States and overseas frequently employ clever optimization techniques to tackle the challenge of cloud resource scheduling. Genetic algorithm (GA) [9], particle swarm optimization (PSO) [10], and ant colony optimization (ACO) [11] are examples of intelligent optimization techniques. Because the PSO method has no overlap and mutation in processing and implementation as compared to the GA algorithm, it is quicker at handling cloud resource scheduling challenges. The PSO method is easier to acquire the initial solution than the ACO algorithm and can more effectively optimize the cloud resource scheduling issue.

Because the PSO approach in a distributed environment has a quicker computation speed and shorter processing time, this research employs an optimization technique based on particle swarm to tackle the cloud resource scheduling problem. The convergence accuracy of the PSO method when searching the optimal solution is low, it is simple to slip into the local optimal solution, and premature convergence difficulties are common. Reference [12] discusses the fuzzy rules and statistical analysis and the Fuzzy Unordered Rule Induction Algorithm (FURIA) has been assessed for ASD behaviors detection on cloud.

In view of the aforementioned defects and disadvantages, this paper proposes a hybrid particle swarm optimization approach (rerandomization inertia weight orthogonal initialization particle swarm optimization, RIOPSO).

Using the rerandomization strategy, the particle swarm may extensively explore the solution space and avoid premature particles [13]. The concept of real-time updating of inertia weight [14] is used to govern the speed of particles in the search process to prevent sliding into a local optimum. In this work, an orthogonal matrix is also used to start the particle swarm [15] so that the particle swarm may attain an ordered starting solution, which is more efficient while searching the solution space. The three optimization methodologies listed above are employed in this study to increase the quality of the solution produced by the PSO algorithm in the search process, as well as the particle's search ability, culminating in the optimal solution. Different related security issues related to data privacy and consistency at cloud are discussed in [16].

2. Related Work

2.1. Multiobjective Problem. The multiobjective problem (MOP) is often due to the conflict between multiple objectives, and it is impossible to obtain the optimal solution that makes each objective reach the optimal state [17]. To solve this problem, the solution algorithm is divided into the three following ways:

(a) Multiobjective evolutionary algorithm based on Pareto domination [18]. Due to the increase of objective functions, the optimal solution set is sometimes too large, which leads to the problem of message overflow when solving multiobjective problems with this method [19].

(b) Multiobjective evolutionary algorithms based on performance indicators [20]. When using the performance index as the reference information for the evolution of the algorithm, the calculation of the performance index is too complicated, resulting in a long running time [21].

(c) Decomposition-based multiobjective evolutionary algorithm [22]. The solution of multiobjective problems is transformed into the collaborative solution of multiple single objectives, and the idea of decomposition is introduced to simplify complex multiobjective problems. The algorithm has high solution efficiency and better solution set performance. The multiobjective evolutionary algorithm based on decomposition (MOEA/D) proposed by Pratap and Zaidi [23] is particularly effective.

Compared with the other two algorithms, the multiobjective evolutionary algorithm based on decomposition has stronger search ability for dealing with combinatorial optimization problems. The optimization goals of cloud resource scheduling include minimizing the total completion time, minimizing resource consumption, and satisfying QoS. Therefore, this paper uses the decomposition idea of the MOEA/D algorithm and uses the compromise model to decompose the weight of the objective evaluation function under the constraints of time and cost. According to the weight ratio, the objective function of time and cost is optimized synchronously.

2.2. Algorithms and Performance Indicators. In order to verify that the RIOPSO algorithm proposed in this paper can guarantee the convergence and diversity when solving the multiobjective cloud resource scheduling problem, this paper will use the NSGA-I algorithm [24], NSGA-II algorithm [25], NSGA-III algorithm [9], and MOEA/D algorithm [26], which are four mainstream algorithms, in comparison with the RIOPSO algorithm proposed in this paper.

With the multiobjective optimization algorithm proposed, how to evaluate the pros and cons of the algorithm has also become an important research direction. When the multiobjective optimization algorithm solves multiobjective problems, the performance indicators can be used to quantify the performance of the algorithm. Commonly used performance indicators can be divided into three categories: accuracy measurement indicators, diversity measurement indicators, and comprehensive measurement indicators. When the above algorithm solves the multiobjective cloud resource scheduling problem, two comprehensive metrics are used, the inverted generational distance (IGD) [27] and the hypervolume (HV) [28] metrics; the accuracy metric, the coverage-metric (C-Metric) [29] index, quantifies the performance of the algorithm, and the above algorithms are compared and evaluated through the quantified performance. This paper uses a fuzzy mathematical model to solve the multiobjective cloud resource scheduling problem with uncertain execution time, and an optimization algorithm is proposed for how to find the best scheduling scheme efficiently.

3. Fuzzy Cloud Resource Scheduling Problem

3.1. Cloud Resource Scheduling Model. Tasks must be executed on virtual machines according to the feasibility algorithm in cloud resource scheduling, and each task can only be completed on one virtual machine; however, virtual machines can execute numerous tasks as needed. There are now m jobs to do. There are n virtual machines; therefore TASK = Task1, Task2, and Task m. VM = Vm1, Vm2, and

Vm n, where n is the number of virtual machines and m is the number of jobs.

The mapping link between virtual machines and tasks may be represented since there is a one-to-many relationship between them in cloud resource scheduling. For example, each job can only be completed on one virtual machine, and each virtual machine can execute numerous tasks. Some fundamental concepts are used in the cloud resource scheduling model. A scheduling scheme Rk is depicted. The execution time of job I on virtual machine j is represented by 1 time ij, and its calculation algorithm is as follows:

$$\text{time}_{ij} = \frac{\text{taskSize}_i}{vm\text{Speed}_j}. \tag{1}$$

$$vm\text{Time}_j = \Sigma i = \sum_{i=0}^{m} m\text{time}_{ij} \times X_{ij}, vm\text{Cost}_j = \text{time}_{ij} \times \text{cost}_j, \text{Time}(R_k) = \max_{j=1}^{n} nvm\text{Time}_j. \tag{2}$$

Definition 1. Cost (R_k) represents the total execution cost of the scheduling scheme R_k. The total execution cost of the scheduling scheme is the sum of the costs consumed by all task executions. The calculation equations are as follows:

$$\text{Cost}(R_k) = \sum_{j=1}^{n} vm\,\text{Cost}_j. \tag{3}$$

$$r\text{TIME}(R_k) = \frac{\text{Time}(R_k) - \text{Time}_{\min}}{\text{Time}_{\max} - \text{Time}_{\min}}. \tag{4}$$

$$r\text{COST}(R_k) = \frac{\text{Cost}(R_k) - \text{Cost}_{\min}}{\text{Cost}_{\max} - \text{Cost}_{\min}}. \tag{5}$$

In the above equation, $\text{Cost}_{\min}$ and $\text{Cost}_{\max}$ are the minimum cost and highest cost required for task execution. Since only the time factor or cost factor is considered, the consideration factor is too single, and a solution that satisfies both cloud resource providers and cloud resource consumers cannot be obtained. References [30, 31] mentioned that cloud resource providers want to provide services at a lower cost, while cloud resource consumers want to perform tasks in a faster time. Therefore, through comprehensive consideration, in order to achieve the goal of satisfying both cloud resource providers and cloud resource consumers, a time-cost constraint is proposed, and the time factor (4) is put forward as well as cost factor (5) to convert the multiobjective problem into a single-objective problem as in equation (6). The evaluation function is

$$\text{eval}(R_k) = t \times r\text{TIME}(R_k) + c \times r\text{COST}(R_k). \tag{6}$$

In the above equation, t represents the time factor, and c represents the cost factor. By changing the values of t and c, this paper determines the proportion of the time factor and cost factor in the scheduling scheme and further controls the impact of time and cost on the evaluation function. The cloud resource scheduling model under the constraint of time-cost condition is

$$P = \min\{\text{eval}(R_k)\}. \tag{7}$$

According to the scheduling model P, each scheduling scheme is evaluated, and the minimum evaluation function value is obtained, and the corresponding scheduling scheme is the optimal scheduling scheme.

3.2. Fuzzy Cloud Resource Scheduling Model. According to equation (1), what is obtained is the determined time when the task is executed on the virtual machine. However, due to the uncertainty of task execution in the actual execution process, this paper uses triangular fuzzy numbers to represent the uncertainty of task execution time. The fuzzy range of task execution time is given by triangular fuzzy numbers, and the upper and lower ranges of the fluctuation range of task execution time on the virtual machine are determined by equations (8) and (9). Fuzzy lower bound on task execution time is

$$t^L = \text{time}_{ij} \times (\alpha_1 + (1 - \alpha_1) \times \text{Rand}()). \tag{8}$$

Fuzzy upper bound on task execution time is

$$t^R = time_{ij} \times (1 + (\alpha_2 - 1) \times Ran\,d()). \tag{9}$$

In the above equation, Rand() represents a random number in (0, 1), and α_1 and α_2 represent the fuzzy coefficients of the upper and lower bounds, respectively. After blurring the execution time of the task, it is expressed as ProTime~ = (t^L, t^M, t^R), where t^L represents the optimistic time of the task execution, t^M represents the most probable time of the task execution, and t^R represents the pessimistic time of the task execution.

According to the linear characteristics and decomposability of triangular fuzzy numbers, the corresponding evaluation function value is calculated using the fuzzy execution time, and equation (7) is converted into the following equation:

$$\min\{\text{eval}(R_k)\} = \min\{P\} = \min\{P^L, P^M, P^R\}. \tag{10}$$

Reference [32] proposed a discriminant rule to determine who has better fuzzy performance by calculating the mean and standard deviation of triangular fuzzy numbers. A fuzzy number is considered to be ranked higher if it has a higher mean and a lower standard deviation. According to the above method, the determined cloud resource scheduling model can be transformed into a fuzzy cloud resource scheduling model with uncertain task execution time. The goal of scheduling is transformed into the calculation of the mean and standard deviation of the evaluation function, and then, according to the addition and multiplication of triangular fuzzy numbers, the characteristics of triangular fuzzy numbers are used to obtain the cloud resource scheduling under uncertain task execution time. The problem model is shown in equation (11). min{eval(Rk)} = min{P̃} = min{P^L, P^M, P^R} =

$$\min\{P\eta + \partial P\ \mu\}, \partial \in [0, 1]. \tag{11}$$

In the above equation, $P\eta$ is the average value of the fuzzy numbers P, $P\mu$ is the standard deviation, and ∂ is the weighting factor for uncertainty.

4. RIOPSO Algorithm

4.1. Analysis of PSO Algorithm. The PSO algorithm is an evolutionary system that models bird foraging behaviour. It iteratively looks for the global optimum with the individual optimum to arrive at the final outcome. To tackle the cloud resource scheduling problem, this work employs the particle swarm optimization technique with an unpredictable execution time. The value of the evaluation function is used to determine the best scheduling strategy. The parameters that must be employed in the PSO method are defined in Table 1. The particles in the PSO algorithm may imitate the bird's foraging activity and be optimized in G-dimensional space. In the particle swarm, there are N particles, and the location of the ith particle is

$$X_i = \left(x_{i,1}, x_{i,2}, \ldots\ldots, x_{i,n}\right). \tag{12}$$

The speed is

$$V_i = \left(v_{i,1}, v_{i,2} \ldots\ldots, v_{i,n}\right). \tag{13}$$

The particle's position is updated by updating the velocity in Table 1 in the iterative process.

In the process of optimization, the particle swarm finds the individual optimum and the global optimum by comparing the value of the evaluation function and then iteratively updates the velocity and position. In the kth iteration, the individual optimum value of the ith particle is pBest_i, and the global optimal value is gBest_k. The velocity update equation of particle i is

$$\begin{aligned} V_{k+1} = {} & \omega \cdot V_k + C_1 \times \text{Rand}() \times \left(p\text{Best}_i - X_k\right) \\ & + C_2 \times \text{Rand}() \times \left(g\text{Best}_k - X_k\right). \end{aligned} \tag{14}$$

The displacement update equation of the particle is

$$X_{k+1} = X_k + V_{k+1}. \tag{15}$$

To use the PSO algorithm to solve the problem of cloud resource scheduling, firstly, the particles in the PSO algorithm correspond to the tasks and virtual machines in the cloud resource scheduling. The dimension of particle solution space depends on the number of tasks in cloud resource scheduling. The value range of each dimension is the number of virtual machines allocated in cloud resource scheduling; that is, the solution of each dimension corresponds to the number of virtual machines. In the process of each iterative update of the particle, the position of the particle will evolve accordingly, and a new evolution particle will be generated, resulting in a new correspondence between the task and the virtual machine, as well as a new scheduling plan. For a cloud resource scheduling with 8 tasks and 3 virtual machines, in the PSO algorithm used, the corresponding particles in the particle swarm have 8 dimensions, and the value of each dimension can be 0, 1, or 2.

Table 1: List of parameters.

Parameter	Meaning
X_i	The position of the particle
V_i	Particle search speed
ω	Inertia weight value
C_1	Influence factor of individual extreme value
C_2	Influence factor of global extreme value
pBest	Individual extreme value (individual optimum)
gBest	Global extremum (entirely optimal)

Table 2: Representation of a particle.

Task	VM	Task	VM
1	1	5	0
2	0	6	2
3	2	7	1
4	1	8	1

Note. task = 1 and VM = 0 mean that task 1 is executed on virtual machine 0.

For the correspondence between tasks and virtual machines, Table 2 is one possible representation of particles.

4.2. Orthogonal Initialization. In the PSO algorithm, the optimization process of particles needs to be carried out through iteration. The initial state of the particle swarm has a direct impact on the subsequent optimization process. In the initial stage of particle search, the more orderly the initial solution is, the better it is for the subsequent particle iterative search. When the population is initialized, it is necessary to ensure that the particles are evenly distributed in the solution space as much as possible, which makes it necessary to satisfy the solution that the particles have all directions in the initialization stage. In the PSO algorithm, the random initialization of the population does not guarantee that the particles can be uniformly distributed in the solution space, and it is possible that the particles are concentrated in one area, or the similarity of the particles is too high, which is not conducive to the subsequent optimization process. In this paper, an orthogonal matrix is used to initialize the population, so that the entire population is uniformly distributed in the feasible solution space.

When the system has elements (Ele), and each element has levels (Lev), a total of Lev^{Ele} combinations will be generated. If all the Lev^{Ele} combinations are tested in the experiment, when Lev and Ele are very large, many similar combinations will be produced. Therefore, constructing an orthogonal matrix $G = L_{\text{row}}(\text{Lev}^{\text{Ele}})$ can screen out the uniform distribution in initializing the population.

The particles in the solution space are used for experiments with fewer combinations, resulting in a smaller particle population. This uses fewer particles but achieves better results. Among them, row represents the total number of groups of horizontal combinations, and row is much smaller than Lev^{Ele}.

For the cloud resource scheduling problem where the number of tasks is 4 and the number of virtual machines is 3, the method of orthogonal initialization particle population is

Table 3: Orthogonal initial population.

Task1	Task2	Task3	Task4
1	1	1	1
1	2	2	2
1	3	3	3
2	1	2	3
2	2	3	1
2	3	1	2
3	1	3	2
3	2	1	3
3	3	2	1

used as an example. If each task is executed on each virtual machine, then $3^4 = 81$ experiments are required. However, if an orthogonal initialization design is used, only 9 experiments are required to find a representative solution. Moreover, with the increase of the number of virtual machines and tasks, the orthogonal initialization can show the characteristics of "evenly dispersed, neat, and comparable." Table 3 shows the allocation scheme of cloud resource scheduling with the use of orthogonal matrices, the number of initialized virtual machines is 3, and the number of tasks is 4. It can be seen from Table 3 that, using the orthogonal matrix to initialize the population, only 9 initial solutions can be uniformly distributed in the solution space. Each row represents a scheduling scheme, the number of columns represents the number of tasks, and each number represents the number of the virtual machines selected to execute the task. For building an orthogonal matrix, first need to determine the base column, then build the nonessential column based on the base column and finally merge the base column and the nonessential column.

It is the number of columns of the orthogonal matrix that needs to be constructed, that is, the total number of tasks that need to be executed as mentioned in Algorithm 1. The basic column J satisfies (16), and the following is the pseudocode for creating an orthogonal matrix.

$$\frac{\mathrm{Lev}^{j} - 1}{\mathrm{Lev} - 1} \geq \mathrm{Ele}. \tag{16}$$

After creating basic columns and nonbasic columns, a complete orthogonal matrix has been constructed, but, in the end, if you want to get a matrix suitable for particle swarm initialization, you need to choose columns for the created matrix. Algorithm 2 gets the final matrix.

This constructs the orthogonal matrix used to initialize the population.

4.3. Rerandomization. According to the shortcoming that the PSO algorithm is easy to fall into the local optimum, the rerandomization method that can make the particle jump out of the local optimum is used to optimize the particle optimization ability, so that the particle can explore a larger range in the solution space and the optimization efficiency is higher. The particles are iteratively updated by calculating the variance to ensure that the particles can obtain better solutions. According to equation (17), the variance value suitable for particle swarm is obtained, and then the particle swarm is iteratively updated according to the obtained variance.

$$\mathrm{variance}(k) = \frac{Ae^{-l+M/s}}{1 + e^{-l+M/s}}. \tag{17}$$

In equation (17), k represents the current iteration number, A represents the effective initial value of rerandomization, and S represents the slope, which controls the search range of particles. In the search process, the first part is called large-scale search, that is, extensive search. At this time, the slope of the variance curve is large, so that particles can search randomly in the search space far from the global optimal particle gBest. The second part is called small-range search, that is, fine search. At this time, the slope of the variance curve is small, and the particles are randomly searched around the global optimal particle gBest. The combination of the two parts can make the final particle converge to the optimal solution, so that the particle can search in the global scope, jump out of the local optimum, and search for the optimal solution. M is the number of iterations corresponding to the midpoint of the slope of the variance curve, and the two parts are divided by the midpoint M, which determines the search time of the broad search and the fine search.

4.4. Updating Inertia Weights. The particle's convergence speed impacts whether the particle can converge to the present optimal state throughout the search phase, while it is looking for optimization in the solution space. The particle's inertia weight value is an essential element in the PSO algorithm. The particle's convergence rate is regulated by altering the particle's inertia weight in each iteration procedure. When the particle's inertia weight is large, the particle search range expands, the convergence rate slows, and it is difficult to get an accurate solution; when the inertia weight is small, the particle search range shrinks, the convergence rate accelerates, and it is easy to fall into the local extreme value. This study employs the approach of real-time updating of inertia weight, along with rerandomization, to allow the particle to explore the complete solution space and manage the particle's convergence rate, according to this characteristic in the particle convergence process. The inertia weight value is changed in each iteration based on the change in the particle's evaluation function. When the evaluation function value of the particle after the iteration becomes smaller, the inertia weight value of the particle will increase or remain unchanged; if the evaluation function value of the particle becomes larger, the inertia weight value of the particle will decrease. Equation (18) is the analogous equation, and the ith particle's inertia weight value is updated in real time.

$$\omega_i(k) = \frac{1}{1 + e^{-\Delta j_i(l)/s}}. \tag{18}$$

In the above equation, $\omega_i(k)$ represents the inertia weight value of the current ith particle, the value range is (0,

```
(1)  For k = 1 to J do
(2)  Begin
(3)  j = Lev^(k−1) − 1/Lev − 1 + 1
(4)  For i = 1to Lev^j do
(5)  a_(i,j) = i − 1/Lev^(j−k)mod Lev
(6)  End
(7)  Step 2 Build the non-base column
(8)  For k = 2 to Jdo
(9)  Begin
(10) j = Lev^(k−1) − 1/Lev − 1 + 1
(11) For s = 1 to j − 1 do
(12) For t = 1to Lev − 1 E
(13) a_(j+(s−1)(Lev−1)+t) = (as × t + aj)mod Lev
(14) End
(15) Step 3 Add 1 to ai, j
(16) 1 ≤ i ≤ row and 1 ≤ j ≤ Ele
```

Algorithm 1: Step 1: build the base column [10].

1), S represents the range of the expected evaluation function value, $\Delta J_i(k)$ represents the difference between the current evaluation value of the particle and the previous one. The difference between the status evaluation values and velocity after re-randomization are updated by inertia weights, i.e., calculated by incorporating equation (16) into equation (21).

Equation (14) is changed into

$$\begin{aligned} V_{k+1} = {} & \omega_i(k) \times V_k + C_1 \times \text{Rand}() \times \left(p\text{Best}_i - X_k\right) \\ & + C_2 \times \text{Rand}() \times \left(g\text{Best}_k - X_k\right) \\ & + \text{variance}(k) \times \text{Rand}() \times \left(p\text{Best}_i - X_k\right). \end{aligned} \tag{19}$$

The particle's position update equation is

$$X_{k+1} = X_k + V_{k+1}. \tag{20}$$

In equation (19), variance(k) is the variance in rerandomization, and its value is determined by equation (17). $\omega_i(k)$ is the inertia weight value of each particle in each update, and the value is determined by equation (18).

5. Simulation Experiments

5.1. Data Generation and Parameter Selection. Simulation tests are carried out using the cloud computing simulation platform CloudSim in order to examine and compare the performances of the models and methods provided in this research. The experiment's outcomes are influenced by both time and expense. Both the time factor t and the cost factor c in equation (6) are set to 0.5 in this work. The dataset in this work is created using the data creation approach described in literature [33], with job sizes and virtual machine processing speeds ranging between [3000, 130000] and [300, 1300]. The execution time and cost of the job on the virtual machine may be calculated using the calculation technique in Section 2.1, based on the size of the acquired task, the virtual machine's processing speed, processing speed for velocity, location, and inertia of the particle swarm in the RIOPSO algorithm. The weight value is repeatedly changed in response to changes in the value of the evaluation function. The values of the individual learning factor C1 and the group learning factor C2 have an influence on the experimental findings when the parameters are established. When $C1 = 1$ and $C2 = 0$, the current particle is only affected by its individual optimal value pBest, and the learning ability of the current global optimal particle is 0; when $C1 = 0$ and $C2 = 1$, the current particle is only affected by its individual optimal value pBest, and the learning ability of the current global optimal particle is 0; and when $C1 = 0$ and $C2 = 1$, the current particle is only affected by its individual optimal value pBest, and the learning ability of the current global optimal particle has no effect on the updating of the global ideal gBest, and the learning ability of this particle individual is 0. The individual learning factor C1 and the group learning factor C2 are both set to 0.5 in this research. Table 4 shows the parameter settings used in this experiment. Except for the different algorithm enhancement methodologies employed in this article, the same dataset and experimental settings were used to address the cloud resource scheduling problem in the same model and environment.

5.2. Evaluation and Analysis of Algorithm Performance Index. The algorithms NSGA-I, NSGA-II, NSGA-III [10], and MOEA/D and the RIOPSO algorithm proposed in this paper are used to solve the problem of cloud resource scheduling.

(1) Algorithm NSGA-I. In contrast to the genetic algorithm, the nondominated sorting genetic algorithm stratifies individuals based on their dominance relationship before applying the selection operator.

(2) Algorithm NSGA-II. The fast nondominated multiobjective optimization method with elite retention strategy uses a crowding degree comparison

(1) Select the smallest $J1$ fulfilling
(2) $\mathrm{Lev}_1^{j_1} - 1/\mathrm{Lev}_1 - 1 \geq \mathrm{Ele}$
(3) If
(4) $\mathrm{Lev}_1^{j_1} - 1/\mathrm{Lev}_1 - 1 = \mathrm{Ele}$
(5) then $\mathrm{Ele}\prime = \mathrm{Ele}$
(6) Else $\mathrm{Ele}\prime = \mathrm{Lev}_1^{j_1} - 1/\mathrm{Lev}_1 - 1$
(7) Execute Algorithm 1 to construct the orthogonal
(8) array
(9) $L_{\mathrm{Lev}_1^{j_1}}(\mathrm{Lev}_1^{\mathrm{Ele}\prime})$
(10) Delete the last $\mathrm{Ele}\prime - \mathrm{Ele}$ columns of
(11) $L_{\mathrm{Lev}_1^{j_1}}(\mathrm{Lev}_1^{\mathrm{Ele}\prime})$
(12) To get $\mathrm{Lev}_1^{\mathrm{Ele}\prime}$ where $\mathrm{row}_1 = \mathrm{Lev}_1^{j_1} 1$

Algorithm 2: The final matrix.

approach and is a Pareto optimum solution optimization technique.

(3) Algorithm NSGA-III. A reference point selection operation based on the NSGA-II algorithm is presented to replace the crowding distance selection operation in the NSGA-II algorithm.

(4) Algorithm MOEA/D. It is a multiobjective evolutionary method based on decomposition.

(5) The IGD, HV, and C-Metric are calculated at this moment to measure the performance of each method while applying the above techniques to solve the cloud resource scheduling problem.

   (a) IGD is the comprehensive performance evaluation index of the algorithm, which calculates the minimum distance between the solution on the Pareto front and the solution obtained by the algorithm. The smaller the value of IGD, the more uniform the solution distribution obtained by the algorithm and the better the convergence.
   (b) HV measures the size of the solution space covered by the nondominated solution. The larger the value of HV, the higher the quality of the solution of the algorithm.
   (c) The C-Metric is calculated using the Pareto solution set's dominance, and the performance metric uses two sets of Pareto fronts to determine convergence performance. C(A, B) = 1 indicates that A dominates all people in B, whereas C(A, B) = 0 indicates that A does not dominate individuals in B. The performance comparison results of the five algorithms are shown in Table 5. The outcomes of performance measures are reported as mean values. Table 5 shows that, in the IGD performance comparison, the RIOPSO algorithm has the least IGD value, showing that the RIOPSO method achieves the lowest IGD value and improvement of optimization is shown in Table 6.

The distribution of the solution set is uniform, which can make the population better converge to the approximate Pareto front, and the comprehensive performance is better.

Table 4: Setting of experimental parameters.

Parameter	Value
Task size	3 000~130 000
Number of tasks	[100, 300]
Number of virtual machines	5, 10, 15, 20
Virtual machine processing speed	300~1 300
Virtual machine memory (Ram)/Mb	2 048
Virtual machine bandwidth/(MB/S)	10 000
Number of data centers	1
Number of hosts	1

It can be seen from the hypervolume (HV) index that the RIOPSO algorithm has the largest target area space between the nondominated solution and the reference point and also has the best solution performance compared to the other four algorithms. From the perspective of C-Metric coverage, the dominance relationship in the RIOPSO algorithm is less than those of other algorithms. This is because the RIOPSO algorithm uses the orthogonal initialization strategy and obtains a higher-quality solution set at the initial stage, achieving convergence the optimal with better solution. In summary, using the improved strategy of the algorithm in this paper can improve the quality of the solution and make the overall performance of the algorithm better when solving the cloud resource scheduling problem.

5.3. Model Comparison. Due to the asynchrony and uncertainty of task execution in the process of cloud resource scheduling in real life, this paper uses the triangular fuzzy method to deal with the problem model so that the results are closer to reality. The following paper compares the experimental results for both the deterministic and indeterminate execution times. When each determined execution time is represented by triangular fuzziness, it is necessary to determine the left and right range of the number after the fuzzy execution time, and set the parameters $\alpha 1 = 0.9$ and $\alpha 2 = 1.2$ in equations (8) and (9). The same task size is compared by the mean value of the evaluation function. Figure 1 shows the comparison results between the deterministic model and the fuzzy model. It can be seen from the image that, using the fuzzy model, the value

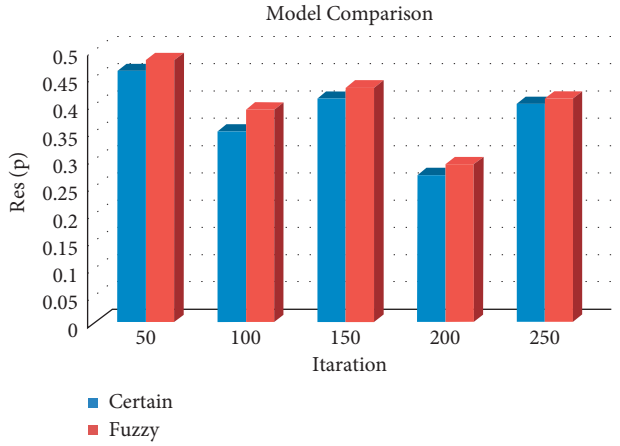


Figure 1: Model comparison.

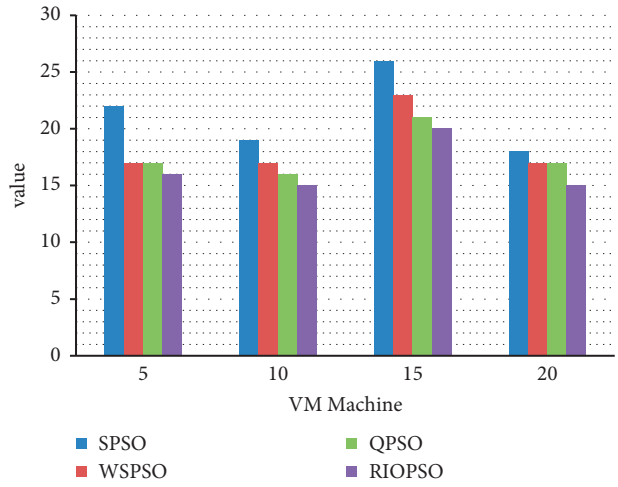


Figure 2: Optimization scales.

of the evaluation function will be higher than the value of the evaluation function under the deterministic model, which means that it is necessary to consider uncertain factors. If these uncertain factors are ignored, this will lead to the deviation between the actual effect and the theoretical estimated effect and reduce the efficiency of the system.

5.4. Algorithm Performance Comparison. The RIOPSO algorithm provided in this study is compared to three other algorithms in order to examine the unique advantages of the enhanced approach proposed in this paper. The SPSO (rerandomization particle swarm optimization) technique, for example, employs a rerandomization strategy [34] without altering the problem's beginning circumstances. To initialize, the random approach is employed while grouping. The SWPSO (rerandomization inertia weight particle swarm optimization) algorithm combines rerandomization and real-time update of inertia weight [35]. The OPSO (orthogonal initialization particle swarm optimization) algorithm optimizes the PSO algorithm using the orthogonal initialization swarm approach [31]. The RIOPSO method provided in this study is the same as that discussed before in order to more intuitively comprehend the efficiency and superior optimization capabilities of the algorithm proposed in this work.

The SPSO, SWPSO, and OPSO algorithms that are obtained are compared and examined. When the job scale is 100, 200, or 300 and the number of virtual machines is 5, optimization tests are performed and recorded for each of the four methods. The number of iterations and the value of the evaluation function related to the algorithm are shown by the horizontal and vertical coordinates, respectively.

The initial evaluation functions of the OPSO and RIOPSO algorithms are found to be superior to those of the other two methods. This is because the quality of the first solution is high, and the particle swarm's search efficiency is correspondingly increased, proving orthogonality. However, the iterative curve shows that the OPSO method may not be able to get a better result than the other three algorithms since the particles may fall into a local optimum condition while looking for the best solution. The SWPSO method and the RIOPSO algorithm, in comparison to the SPSO algorithm, boost particle search speed control by managing the particle search range, allowing particles to leap out of the local optimum throughout the search process and better converge to the global optimum solution. The graphic also shows that the RIOPSO algorithm, which combines the three aforementioned techniques, may achieve better answers in less iterations than the other three algorithms in large-scale or small-scale problems, reducing the algorithm's current iteration time. As a result, the RIOPSO algorithm has a rapid optimization speed, a strong optimization capability, and a good optimization impact.

Multiple experiments were run on four distinct datasets to demonstrate the RIOPSO algorithm's capacity to handle the same number of jobs but with varying numbers of virtual machines. The task and virtual machine numbers were (100, 5), (100, 10), (100, 15), and (100, 20), respectively, and the experimental results were averaged as shown in Figure 2. When the number of tasks is 100, the abscissa indicates the number of virtual machines which matches the number of tasks, and the ordinate represents the value of the evaluation function. Figure 2 illustrates that when the number of tasks is the same but the number of virtual machines varies, RIOPSO may discover a better solution and offer a better scheduling scheme than the other three algorithms.

According to the evaluation function value of each technique in Figure 2, the relative difference percent images of the evaluation function value of the SPSO algorithm, SWPSO algorithm, and OPSO algorithm are larger than that of the RIOPSO algorithm. As shown in Figure 3, the abscissa represents the number of virtual machines, while the ordinate represents the relative difference percent. When compared to the other three algorithms, the RIOPSO technique has the greatest optimization capability, regardless of the size of the virtual machine, indicating the RIOPSO algorithm's usefulness. When the time factor t and the cost factor c are both 0.5, the four following strategies are utilized to perform repeated tests on scheduling models of varying work sizes. There are an average of 5 virtual machines in use. Figure 4 depicts the overall execution time of the tasks performed by the four algorithms, whereas Figure 5 depicts

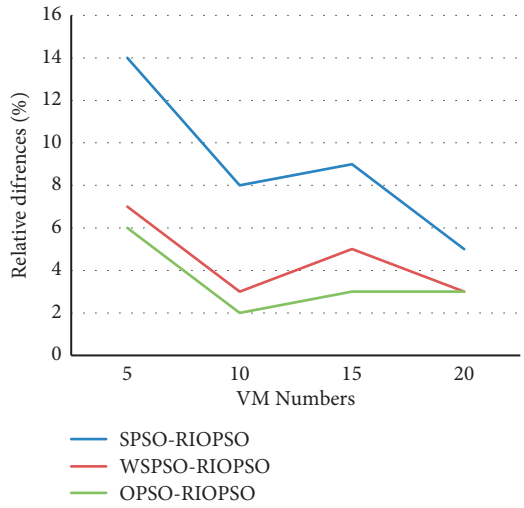


Figure 3: Improvement in optimization.

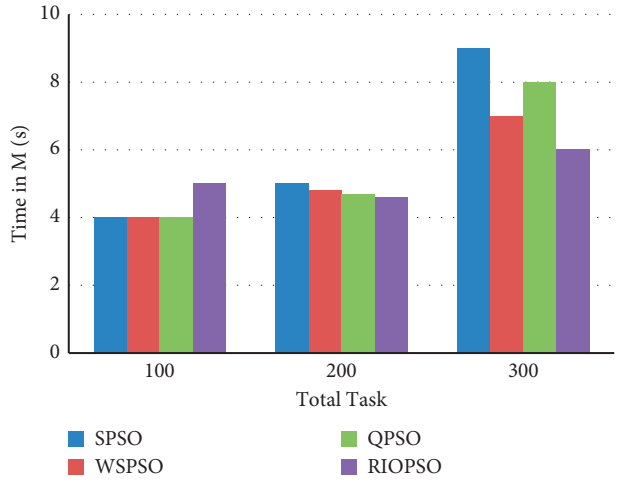


Figure 4: Time comparison.

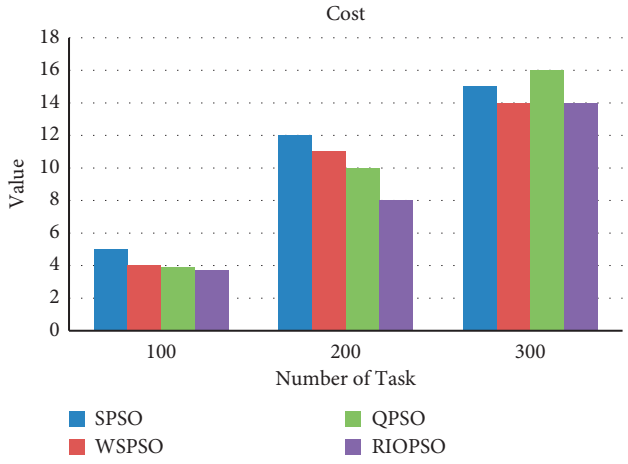


Figure 5: Cost comparison.

Table 5: Optimization scale values.

VM number	SPSO [14]	WSPSO [14]	QPSO [18]	RIOPSO [18]
5	22	17	17	16
10	19	17	16	15
15	26	23	21	20
20	18	17	17	15

Table 6: Improvement in optimization.

VM number	SPSO-RIOPSO	WSPSO-RIOPSO	OPSO-RIOPSO
5	14	7	6
10	8	3	2
15	9	5	3
20	5	3	3

Table 7: Time comparison.

Total tasks	SPSO [14]	WSPSO [14]	QPSO [18]	RIOPSO [18]
100	4	4	4	5
200	5	4.8	4.7	4.6
300	9	7	8	6

Table 8: Cost comparison.

Total tasks	SPSO [14]	WSPSO [14]	QPSO [18]	RIOPSO [18]
100	5	4	3.9	3.7
200	12	11	10	8
300	15	14	16	14

the total cost of task execution. The abscissas depict the task scale, while the ordinates depict the total execution time and cost of the tasks under the specified scheduling scheme. Comparison with respect to time is shown in Table 7 for these algorithms and cost comparison is depicted in Table 8.

The RIOPSO method may produce the least value of both the total time of task execution and the total cost of task operation, as shown in the figure. The twin goal of getting the shortest execution time and the cheapest execution cost is obtained. According to the results of the aforementioned trials, the RIOPSO algorithm described in this work outperforms the other three algorithms in terms of optimization ability and efficiency, and, in terms of identifying the best cloud resource scheduling scheme, it can reduce task execution time and cost, which enhances cloud resource scheduling's overall performance.

6. Conclusion

The main goal of this paper is to reduce the total completion time and total execution cost of the task, while also taking into account the impact of uncertain factors on task execution time, using a fuzzy cloud resource scheduling model, and propose an improved hybrid PSO algorithm for cloud resource scheduling. Three viable methodologies are employed in the RIOPSO algorithm suggested in this paper:

orthogonal initialization of particle swarms, rerandomization of particles, and updating of inertia weights. The three methodologies have shown promising results in terms of identifying the best solution, leaping out of the local optimum, and enhancing particle convergence. To bring cloud resource scheduling closer to actual applications, fuzzy models are used. The task execution time is reduced and the execution cost is reduced when the RIOPSO algorithm is used to optimize the fuzzy cloud resource scheduling. When compared to single SPSO, OPSO, and SWPSO algorithms, RIOPSO produces superior results that are more in line with real-world applications.

Data Availability

The data can be obtained 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 Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University for funding this work through Research Group no. RG-21-07-09.

References

[1] D. Wu, "Cloud computing task scheduling policy based on improved particle swarm optimization," in *Proceedings of the 2018 International Conference on Virtual Reality and Intelligent Systems*, pp. 99–101, (ICVRIS), Hunan, China, August 2018.

[2] X. Xiangqian Song, L. Jieping Wang, and J. Wang, "Job scheduling based on ant colony optimization in cloud computing," in *Proceedings of the 2011 International Conference on Computer Science and Service System*, pp. 3309–3312, (CSSS), Nanjing, China, June 2011.

[3] Y. Dai, Y. Lou, and X. Lu, "A task scheduling algorithm based on genetic algorithm and ant colony optimization algorithm with multi-QoS constraints in cloud computing," in *Proceedings of the 2015 7th International Conference on Intelligent Human-Machine Systems and Cybernetics*, pp. 428–431, Hangzhou, China, August 2015.

[4] H. Al-Zoubi, "Efficient task scheduling for applications on clouds," in *Proceedings of the 2019 6th IEEE International Conference on Cyber Security and Cloud Computing (CSCloud)/2019 5th IEEE International Conference on Edge Computing and Scalable Cloud*, pp. 10–13, (EdgeCom), Paris, France, June 2019.

[5] Y. Cui and Z. Xiaoqing, "Workflow tasks scheduling optimization based on genetic algorithm in clouds," in *Proceedings of the 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis*, pp. 6–10, ICCCBDA), Chengdu, China, April 2018.

[6] S. Liu and Y. Yin, "Task scheduling in cloud computing based on improved discrete particle swarm optimization," in *Proceedings of the 2019 2nd International Conference on Information Systems and Computer Aided Education*, pp. 594–597, (ICISCAE), Dalian, China, September 2019.

[7] A. Ragmani and A. E. Omri, N. Abghour, K. Moussaid, N. Abghour, K. Moussaid, and M. Rida, An intelligent scheduling algorithm for energy efficiency in cloud environment based on artificial bee colony," in *Proceedings of the 2017 3rd International Conference of Cloud Computing Technologies and Applications*, pp. 1–8, (CloudTech), Rabat, Morocco, October 2017.

[8] S. Khan, A. S. Al-Mogren, and M. F. AlAjmi, "Using cloud computing to improve network operations and management," in *Proceedings of the 2015 5th National Symposium on Information Technology: Towards New Smart World*, pp. 1–6, (NSITNSW), Riyadh, Saudi Arabia, February 2015.

[9] Z.-H. Huo, P. Wang, S.-J. Zhang, D. Wang, and Z. Kong, "A two-step multi-objective optimization frame-work for microgrid scheduling problem based on cloud-edge computing," in *Proceedings of the 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2)*, pp. 2764–2768, Wuhan, China, November 2020.

[10] V. Vinothina, "Scheduling scientific workflow tasks in cloud using swarm intelligence," in *Proceedings of the 2017 IEEE International Conference on Current Trends in Advanced Computing*, pp. 1–5, (ICCTAC), Bangalore, India, March 2017.

[11] R. Sharma and M. Bharti, "Mapping of tasks to resources maintaining fairness using swarm optimization in cloud environment," in *Proceedings of the 3rd International Conference on Reliability, Infocom Technologies and Optimization*, pp. 1–6, Noida, India, October 2014.

[12] S. Khan and M. Alshara, "Fuzzy data mining utilization to classify kids with autism," *IJCSNS*, vol. 19, no. 2, p. 147, 2019.

[13] D. Gabi, A. Zainal, A. S. Ismail, and Z. Zakaria, "Scalability-Aware scheduling optimization algorithm for multi-objective cloud task scheduling problem," in *Proceedings of the 2017 6th ICT International Student Project Conference*, pp. 1–6, (ICT-ISPC), Johor, Malaysia, May 2017.

[14] M. Arora, V. Kumar, and M. Dave, "Task scheduling in cloud infrastructure using optimization technique genetic algorithm," in *Proceedings of the 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability*, pp. 788–793, WorldS4), London, UK, July 2020.

[15] P. K. Chauhan, P. Jaglan, and P. Dabas, "An intensify deadline aware credit based cloud task scheduling," in *Proceedings of the International Conference on Computing, Communication and Automation*, pp. 1267–1270, (ICCCA), Greater Noida, India, April 2016.

[16] S. Khan and M. F. AlA, "A review on security concerns in cloud computing and their solutions," *International Journal of Computer Science and Network Security*, vol. 19, no. 2, pp. 9–15, 2019.

[17] S. Antony, S. Antony, A. S. A. Beegom, and M. S. Rajasree, "Task scheduling algorithm with fault tolerance for cloud," in *Proceedings of the 2012 International Conference on Computing Sciences*, pp. 180–182, Phagwara, India, September 2012.

[18] P. Han, C. Du, and J. Chen, "A DEA based hybrid algorithm for Bi-objective task scheduling in cloud computing," in *Proceedings of the 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems*, pp. 63–67, (CCIS), Nanjing, China, November 2018.

[19] V. Chavan, K. Dhole, and P. R. Kaveri, "Dynamic selection of job scheduling policies for performance improvement in cloud computing," in *Proceedings of the 2016 3rd International*

Conference on Computing for Sustainable Global Development, pp. 379–382, (INDIACom), New Delhi, India, March 2016.

[20] M. De, A. Kundu, and S. Guha, "Advanced cloud based task scheduling architecture to optimize performance in data-center," in *Proceedings of the 2020 IEEE 1st International Conference for Convergence in Engineering*, pp. 139–143, (ICCE), Kolkata, India, September 2020.

[21] P. Naithani, "Genetic algorithm based scheduling to reduce energy consumption in cloud," in *Proceedings of the 2018 Fifth International Conference on Parallel, Distributed and Grid Computing*, pp. 616–620, (PDGC), Solan, India, December 2018.

[22] C. He, Y. Yang, and B. Hong, "Cloud task scheduling based on policy gradient algorithm in heterogeneous cloud data center for energy consumption optimization," in *Proceedings of the 2020 International Conference on Internet of Things and Intelligent Applications*, pp. 1–5, (ITIA), Zhenjiang, China, November 2020.

[23] R. Pratap and T. Zaidi, "Comparative study of task scheduling algorithms through Cloudsim," in *Proceedings of the 2018 7th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)*, pp. 397–400, ICRITO), Noida, India, August 2018.

[24] M. Mtshali, H. Kobo, S. Dlamini, M. Adigun, and P. Mudali, "Multi-objective optimization approach for task scheduling in fog computing," in *Proceedings of the 2019 International Conference on Advances in Big Data, Computing and Data Communication Systems*, pp. 1–6, (icABCD), Winterton, South Africa, August 2019.

[25] F. Alhaidari, T. Balharith, and E. Al-Yahyan, "Comparative analysis for task scheduling algorithms on cloud computing," in *Proceedings of the 2019 International Conference on Computer and Information Sciences*, pp. 1–6, (ICCIS), Sakaka, Saudi Arabia, April 2019.

[26] X. Wang, "Optimization of server scheduling based on cloud platform," in *Proceedings of the 2020 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)*, pp. 587–590, Vientiane, Laos, January 2020.

[27] D. Yi, "Dynamic binary translation cache optimization algorithm in cloud computing environment," in *Proceedings of the 2021 Global Reliability and Prognostics and Health Management*, pp. 1–5, (PHM-Nanjing), Nanjing, China, October 2021.

[28] S. Kitagami, T. Ogino, T. Suganuma, and N. Shiratori, "Proposal of a multi-agent based flexible IoT edge computing architecture harmonizing its control with cloud computing," in *Proceedings of the 2017 Fifth International Symposium on Computing and Networking*, pp. 223–229, (CANDAR), Aomori, Japan, November 2017.

[29] R.-M. Chen and S.-C. Shih-Che Huang, "Particle swarm optimization for scheduling problems by curve controlling based global communication topology," in *Proceedings of the 2015 12th International Conference on Fuzzy Systems and Knowledge Discovery*, pp. 1716–1720, (FSKD), Zhangjiajie, August 2015.

[30] R. Khanam, R. R. Kumar, and B. Kumari, "A novel approach for cloud service composition ensuring global QoS constraints optimization," *Communications and Informatics*, (ICACCI), in *Proceedings of the 2018 International Conference on Advances in Computing*, pp. 1695–1701, September 2018.

[31] M. F. AlAjmi, A. S. Head, and S. Khan, "Growing cloud computing efficiency," *International Journal of Advanced Computer Science and Applications*, vol. 3, no. 5, 2012.

[32] M. T. Alotaibi, M. S. Almalag, and K. Werntz, "Task scheduling in cloud computing environment using bumble bee mating algorithm," in *Proceedings of the IEEE Global Conference on Artificial Intelligence and Internet of Things (GCAIoT)*, pp. 01–06, Dubai, UAE, December 2020.

[33] M. L. Chatterjee, C. Lanham, and M. Zerrudo, "Job scheduling in cloud datacenters using enhanced particle swarm optimization," in *Proceedings of the 2nd International Conference for Convergence in Technology (I2CT)*, pp. 895–900, Mumbai, India, April 2017.

[34] H.-J. Jiao, J. Li, and J.-P. Li, "The cloud parameters specification and scheduling optimization on multidimensional qos constraints," in *Proceedings of the 2018 15th international computer conference on wavelet active media technology and information processing*, vol. 49, no. 8, pp. 22–26, (ICC-WAMTIP), Chengdu, China, December 2018.

[35] L. Zhai, Z. Yang, and W. Ji, "Understanding crowd intelligence in large-scale systems: a hierarchical binary particle swarm optimization approach," in *Proceedings of the 2020 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking*, pp. 728–735, ISPA/BDCloud/SocialCom/SustainCom), Exeter, UK, December 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 5084178, 9 pages
https://doi.org/10.1155/2022/5084178

Research Article

Multicriteria Deming Regressive African Buffalo Optimized Mapping for 3D NoC Architecture Design

Sushma G[1], **Aravindhan Alagarsamy**[2], **Lakshminarayanan Gopalakrishnan**[1] and **Aruna Rai Vadde**[3]

[1]*Department of Electronics and Communication Engineering, National Institute of Technology, Tiruchirappalli, India*
[2]*Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Andhra Pradesh, India*
[3]*Department of Electrical and Computer Engineering, Debre Tabor University, Debra Tabor, Ethiopia*

Correspondence should be addressed to Aruna Rai Vadde; dr.rai@dtu.edu.et

Received 30 May 2022; Revised 12 June 2022; Accepted 22 June 2022; Published 6 July 2022

Academic Editor: Mukesh Soni

Purpose. Networks-on-Chip (NoC) is a network-based communication between operating cores and intellectual property (IP) cores integrated on the same chip. An efficient design of NoC ensures high-speed data transfer and minimum essential connections in large-scale multicore, low power applications. *Design/methodology/approach*. A unique technique called Multicriteria Deming Regressive African Buffalo Optimized Mapping Weighted Directive Graph Theory (MDRABOMWDGT) is introduced for an efficient Network-on-Chip architecture design. The main aim of the proposed technique is to find efficient operating cores integrated on the chip with minimum time. Initially, a set of IP cores are listed with their connections from the benchmark dataset. Then Multicriteria Deming Regressive African Buffalo Optimization is applied to the topology for mapping strategy on 3D NoC, with communication metrics such as throughput, latency, and computation time. The optimization technique initializes the population of cores in search space. For each core in the network, communication metrics are measured; then Deming regression is applied to analyze multicriteria functions for minimizing computation time. Further, fitness is measured to get an optimal IP core for improving the performance of mapping in 3D NoC. *Findings*. Comprehensive experimental evaluation is conducted using a benchmark dataset, communication metrics are measured, and results show significant improvement in performance with respect to energy parameters compared to state-of-the-art works. *Originality/value*. The mapping strategy for 3D NOC is developed and the results are compared with the state-of-the-art techniques.

1. Introduction

Digital systems are constructed with a large number of cores on a single integrated circuit. Network-on-chips (NoCs) have emerged as a communication backbone for facilitating the high degree of integration in many-core chips. Placing multicores with several on-chip processing elements in one chip offers greater performance gains, but it faces many challenges. The key challenge is to offer efficient, low power, scalable, and reliable communication among these cores. Efficient 3D Networks-on-Chip (NoC) designs are required to improve the QoS-aware communication and low power demands of large-scale multicore applications.

With the continual technology development in VLSI, new integrated circuits (ICs) contain a large number of processing components on a single chip directing to communication deficiency and design complexity in the System-on-Chip. To handle this issue, a Network-on-Chip (NoC) has emerged as a novel model for providing efficient communication structural design. While considering the more intellectual property (IP) cores in the design of NoC, the performance of the system gets decreased. In such scenarios, flexibility and assurance are primary concerns for regular and irregular topologies. In general, the NoC regular and irregular topologies include certain limitations such as poor communication and higher network resource utilization. In addition, the area and performance are two

significant factors in heterogeneous irregular networks. However, the design of these networks with low power consumption for large-scale multicore applications is more difficult. Therefore, the optimization of architectural design in the on-chip network is essential to maximize the performance of the network by minimizing resource utilization.

Three-dimensional Network-on-Chip (3D NoC) architecture is achieved for addressing the on-chip communication delay next generation System-on-Chip (SoC) systems. 3D NoC is designed to obtain better performance and lower power consumption. 3D NoC includes multiple planar layers and vertical communication. The vertical communication of 3D NoC is costly and complex to manufacture. In addition, 3D architecture takes low power and occupies more area per chip floorplan. Hence, more efficient architectures should be designed. Based on this motivation, the Multicriteria Deming Regressive African Buffalo Optimized Mapping Weighted Directive Graph Theory (MDRABOMWDGT) technique is used for achieving an efficient 3D NoC design. The MDRABOM is designed with the implementation of the weighted directive graph theory and Multicriteria Deming Regressive African Buffalo Optimized Mapping Weighted Directive Graph Theory.

1.1. Paper Outline. The rest of this paper is structured into five different sections. Section 2 reviews the related works. Section 3 provides a brief description of the MDRABOM technique for solving the multicriteria 3D NoC design problem. Experimental assessment of the proposed and the existing algorithm is described with benchmark dataset in Section 4; subsequently the performance analysis is provided in Section 5. At last, the conclusion and future work are presented in Section 6.

2. Related Works

A Stochastic Multiobjective Pareto-Optimization Framework was introduced in [1] for designing automatic Network-on-Chip to reduce network latencies and power utilization. But the designed framework was not efficient to improve the quality of automated NoC designs for multicore systems. A self-adaptive chicken swarm optimization algorithm (SCSO) was developed in [2] for efficient mapping to minimize the power consumption of NoC. However, the designed algorithm failed to consider 3D topologies, and mapping with other performance metrics like area, the delay was not considered.

A knowledge-based memetic algorithm (KBMA) was designed in [3] for 3D NoC mapping with standard network topologies. But the algorithm was not effective for application mapping by adopting different metaheuristics algorithms. Butterfly Fat Tree (BFT) topologies were developed in [4] for improving the power and performance analysis of 3D Network-on-Chip design. The designed method minimizes the network latency, but the higher throughput was not achieved. A deterministic and scalable arbitration mechanism was introduced in [5] to minimize the average latency. But the higher network throughput was not achieved.

Evaluation of Wireless Network-on-Chip architecture design was introduced in [6] for 3D wireless NoC design to initiate data communication with minimum energy consumption. However, pumping liquid through the microchannels can cause high pressure drops causing structural instability in the chip. An adaptive thermal-aware routing (ATAR) method was introduced in [7] to reduce the peak on-chip temperature. But the model failed to solve the problem of the thermal-aware solution in 3D many-core systems.

A multiobjective design space exploration framework was introduced in [8] for Network-on-Chip power grid design. But the framework failed to use an efficient evolutionary technique to minimize the convergence time. The machine learning techniques were introduced in [9] for designing the NoC architecture components. Though the designed technique minimizes the latency and time to get optimized solutions were not minimized. The Simulated Allocation (SAL) algorithm was designed in [10] to minimize the communication power and latency during the mapping process. However, the designed algorithm failed to consider the network throughput.

A genetic algorithm and Monte Carlo simulation techniques were introduced in [11] for improving the reliability and minimizing the energy consumption of embedded systems implemented with Network-on-Chip (NoC). The designed algorithm minimizes the computational runtime, but the performance of latency and throughput was not analyzed. A Butterfly Fat Tree (BFT) based design with a zone-based routing policy was introduced in [12] to minimize the latency and power consumption. But the method failed to involve the mapping process of IP cores with different layers.

A bat mapping algorithm (BMAP) was designed in [13] for mapping the IP cores. The designed algorithm minimizes the latency and power consumption. But the other metrics such as area and delay were not considered during the mapping process. A heuristic application mapping algorithm was designed in [14] for mesh-based NoC design to reduce both total energy and runtime. But the mapping algorithm was not applicable for 3D mesh NoC design.

A hybrid Scalable-Minimized-Butterfly-Fat-Tree (H-SMBFT) topology was introduced in [15] for on-chip communication. The designed topology minimizes the latency and energy consumption, but the throughput analysis was not performed. The different mapping approaches were developed in [16] for NoC design implementation. However, efficient heuristic algorithms were not considered for fast NoC designs to minimize the execution time of the mapping process.

Gaussian-based optical NoCs design was performed in [17] to ensure accurate and reliable communication on-chip. The designed method increases the throughput and minimizes the latency, but the computation time was not minimized. A hardware-efficient WiNoC with honeycomb topology was introduced in [18] for minimizing the resources such as network cost, delay, and energy utilization. However, this method failed to concentrate on delay and high power consumption.

Bat mapping algorithm was introduced in [19] for mapping on NoC strategy. But the algorithm failed to resolve the multiobjective problems during the mapping process. The EsyTest strategy was introduced in [20] to reduce the impact of test procedures on the Networks-on-Chip and to minimize the execution time. However, BIST causes significant performance loss due to data dependencies [21].

Machine learning algorithm-based applications have been deployed in [22] for supporting the Internet of things (IoT) and web search engines without losing accuracy in order to satisfy human requests. A new latency model was proposed in [23] for the routers which characterize the network contentions among different traffic flows from sharing of network resources.

A methodology was proposed in [24] to generate an NoC architecture along with a scheduling technique customized for different DNNs. However, crossbar-based in-memory computing may significantly increase the volume of on-chip communication since the weights and activations are on-chip. A new optimization technique called the African Buffalo Optimization was proposed in [25].

A detailed survey of the work is done in [26] in last decade in the domain of application mapping. An improvised cluster based mapping with metaheuristic search algorithm called simulated annealing with tabu search (SAT) was proposed in [27] to analyze and optimize the power consumption of NoC-based systems.

Liquid state machine (LSM) is an attractive spiking neural network (SNN) introduced in [28] for Network-on-Chip- (NoC-) based neuromorphic platforms due to their biological characteristics and hardware efficiency. But the randomly connected topology of the liquid in LSM brings different dataflow and communication congestion on the NoC-based platform. A new low complexity heuristic algorithm, CastNet, was designed in [29] for the application mapping and bandwidth constrained routing algorithm for mesh-based NoC architectures aiming to minimize the energy consumption. However, mapping applications represented by the weighted task graphs onto the mesh architectures is an NP-hard problem.

3. Proposed Methodology

The proposed technique is designed with the implementation of weighted directive graph theory and Multicriteria Deming Regressive African Buffalo Optimized Mapping Weighted Directive Graph Theory. Figure 1 demonstrates the architecture of the proposed MDRABOM technique for 3D NoC architecture design. Initially, the numbers of IP cores $C_1, C_2, C_3, \ldots C_n$ are collected from the benchmark dataset. Intellectual property (IP) core also known as IP block is a reusable part of a cell or integrated circuit (IC) design. To design the integrated circuit for improving the performance of the system, the optimized cores are used. In order to examine the different metrics and choose the optimal cores based on the multicriteria optimization technique, Deming regression is used. The mapping is carried based on the multiple objective functions such as energy consumption, area, power, and delay to find the optimal IP core. This helps to achieve the energy and power-aware architecture design.

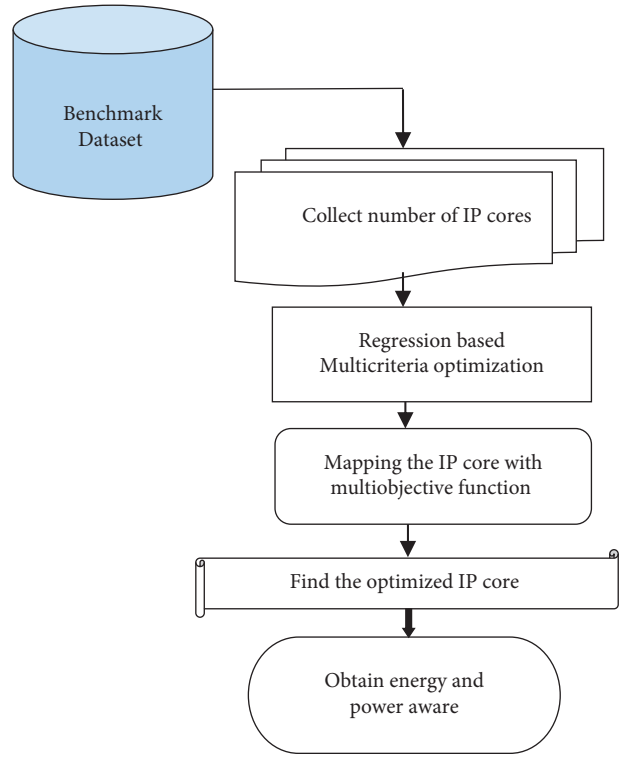


Figure 1: Flow process of the proposed MDRABOM technique.

The proposed technique is designed with the implementation of Multicriteria Deming Regressive African Buffalo Optimized Mapping Weighted Directive Graph Theory.

3.1. Key Contributions of the Proposed MDRABOM. Some of the key technical contributions of the MDRABOM are listed as follows.

(1) To improve the network throughput and latency, weighted directive graph theory is applied for NoC mapping. The mapping is performed based on the Deming Regressive African Buffalo Optimization technique to find the power and energy-efficient IP core from the population. The lesser number of cores in the graph minimizes the route length. This improves the transmission of the packet towards the destination and minimizes the latency.

(2) To minimize the computation time, a Deming regression is applied to an African Buffalo Optimization technique and fitness measure based on multicriteria function such as energy consumption, power, delay, and area. The regression analysis in the optimization technique finds the optimal IP core for 3D NoC design with minimum time.

(3) Finally, extensive simulations are carried out using a benchmark dataset to evaluate the performance of the MDRABOM technique and other optimization

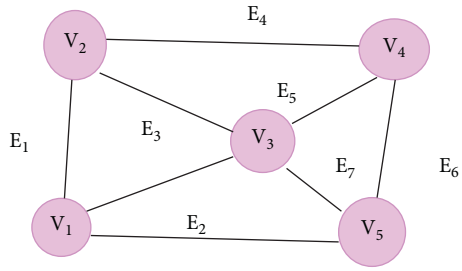


Figure 2: Undirected weighted directive graph.

works. The obtained result shows that the MDRABOM technique is highly efficient compared to the other methods.

3.2. Network Model. The proposed MDRABOM technique uses the weighted directive graph for NoC mapping. The weighted directed graph-based mapping is a mathematical model used to determine the relationship between the two variables.

From Figure 2 a weighted undirected graph $G = (V, E)$ where 'V' denotes the number of vertices (i.e., cores); 'E' stands for the links between the nodes. Figure 2 demonstrates the undirected weighted directive graph with five vertices (i.e., nodes) V_1, V_2, V_3, V_4, V_5 and seven edges $E_1, E_2, E_3, E_4, E_5, E_6, E_7$ (i.e., links). For each node in the network $V_1 \in N$, we have to find the multicriteria (i.e., multiobjective) functions such as energy consumption, area, power, and delay for mapping. The node with the best fitness has a higher weight than the others in the graph. The mapping is performed by applying the injective map function in the graph theory using the following mathematical formula:

$$F: V_i \longrightarrow V_j, \tag{1}$$

where F denotes a mapping function, and V_i, V_j denote a node in the network. The optimized resources are determined for applying multiobjective optimization of every node. Energy consumption of resources is φ_1, area φ_2, power φ_3, and delay φ_4 of node (V). By satisfying equation (2), optimal node selection is performed. In order to find neighboring node (i.e., minimum distance) and links, binary mapping has been applied.

$$P\left(V_i, V_j\right)\begin{cases} 1, & \arg\min\left\{\varphi_1, \varphi_2, \varphi_3, \varphi_4\right\}, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where $P(V_i, V_j)$ symbolizes mapping probability function among nodes. If mapping function returns '1', then optimal node is detected. Otherwise, mapping function returns '0'. In following sections, mathematical model of Multicriteria Deming Regressive African Buffalo Optimization is discussed.

3.3. Multicriteria Deming Regressive African Buffalo Optimization Model. In order to find the optimal node, Multicriteria Deming Regressive African Buffalo Optimization model is applied. The African Buffalo Optimization is a metaheuristic algorithm to identify the position of the best buffalo from the population. The proposed optimization was inspired by the behavior of buffalos in the African forests. The proposed optimization was inspired by the behavior of buffalos in the African forests. On the contrary to existing African Buffalo Optimization, the proposed MDRABOMWDGT technique uses multicriteria optimization and Deming regressive function. The multicriteria optimization considers the many objective functions, namely, energy consumption (φ_1), area (φ_2), power (φ_3), and delay (φ_4). Deming regression function is applied into an optimization technique to analyze the estimated factors such as energy consumption (φ_1), area (φ_2), power (φ_3), and delay (φ_4) of each IP core.

Randomly initialize the population of the IP cores (i.e., buffalos) in search space. For each IP core, the multiple objective functions are measured for evaluating the fitness. The multicriteria work is applied for identifying and comparing the different policy options by calculating their effects, performance, and impacts [30]. The multicriteria work provides a systematic approach for supporting complex decisions based on the predetermined criteria and objectives. After that, Deming regression function is utilized for examining the multiple objective functions. If the fitness of one buffalo is greater than the other, the position of buffalos' gets updated, and it finds the optimal. This process is repeated until the maximum iteration is reached. The optimization is more robust and efficient than the other optimization algorithm since it uses a few learning parameters resulting in guaranteeing the fast convergence rate. Therefore, the initialization process is expressed as follows:

$$C_i \in C_1, C_2, C_3, \ldots C_{\mathrm{n}}, \tag{3}$$

where C_i denotes IP cores. After the initialization in search space, the fitness is computed based on multiple objective functions of energy consumption (φ_1), area (φ_2), power (φ_3), and delay (φ_4).

Initially, the energy consumption is calculated as given below:

$$\varphi_1 = \frac{\Delta t_i}{r_j}. \tag{4}$$

From (4), φ_1 indicates energy consumed by module 'i', and 'Δt_{ij}' represents the temperature rise at module 'i' with respect to transfer resistance at module 'r_j'.

Area (φ_2) is defined as a total area model; 3D NoC is a sum of router/switch area (a_r), area of intellectual property (IP) cores (a_c), and area of on-chip global interconnects (a_g). The area is formulated as given below:

$$\varphi_2 = a_r + a_c + a_g, \tag{5}$$

$$a_r = n * \sum_{i=1}^{n_s} a_{ri}, \tag{6}$$

where n denotes the number of planes presented in 3D NoC, n_s denotes the number of switches in the 2D or 3D network,

and a_{ri} denotes the area of switch i. From (5), area of on-chip global interconnection 'a_g' is measured as follows:

$$a_g = n_L[f(r_w + q_w) + q_w]w_L, \tag{7}$$

where n_L denotes the number of links presented in the 3D networks, f represents flit size in bits, r_w denotes a wire width, q_w indicates a spacing between wires, and w_L denotes a wire length of the global interconnection in the on-chip network.

Power (φ_3) is considered as a global link power which is the sum of the three different power consumptions of 3DNoC. The global link power is estimated as given below:

$$p_g = p_s + p_t + p_c, \tag{8}$$

where p_g denotes a global link power, p_s denotes a power due to circuit switching, p_t denotes a short circuit power, and p_c denotes a static power.

Delay (φ_4) is measured based on three factors as a router, propagation delay due to link or channel, and serialization of packets. The overall delay is measured as given below:

$$D = A_{avg}R + d_p + s_d, \tag{9}$$

where A_{avg} denotes an average hop count, R denotes a router, d_p denotes a propagation delay due to link or channel, and s_d indicates a serialization of packets. The Deming regression function is applied into an optimization technique to analyze the estimated factors such as energy consumption (φ_1), area (φ_2), power (φ_3), and delay (φ_4) of each IP core. Deming regression is a machine learning technique to analyze the given input variables and find the best fit from the populations. The regression analysis is performed as given below:

$$\begin{aligned} Y_i &= \beta_0 + \beta_1[MC_k(C_i)] \\ MC_k(C_i) &\in \varphi_1, \varphi_2, \varphi_3, \varphi_4, \end{aligned} \tag{10}$$

where Y_i denotes an output of multiobjective estimation of cores C_i, β_0 and β_1 indicates the regression coefficients, and $MC_k(C_i)$ denotes multiobjective estimation of the cores, i.e., energy consumption (φ_1), area (φ_2), power (φ_3), and delay (φ_4). Based on the regression analysis, the node that has minimum energy consumption, area utilization, power consumption, and delay is chosen as the optimal from the population. After the analysis, the fitness is measured to find the best-fit IP core.

$$Q_F = \arg\min\{MC_k(C_i)\}, \tag{11}$$

where Q_F denotes a fitness function, and argmin denotes an argument of a minimum function. Based on the fitness value, the two processes such as exploration and exploitation are performed as given below:

$$x_k(t+1) = x_k + a_1[Q_{Fb} - E_k] + a_2[x_{bp}.k - E_k], \tag{12}$$

where $x_k(t+1)$ indicates an updated buffalos' exploitation of the 'k' th buffalo, x_k denotes a current position of the k' th buffalo, E_k indicates an exploration of the $E_k th$ buffalos', a_1 and a_2 are the learning parameters set values from 0.1 to 0.6,

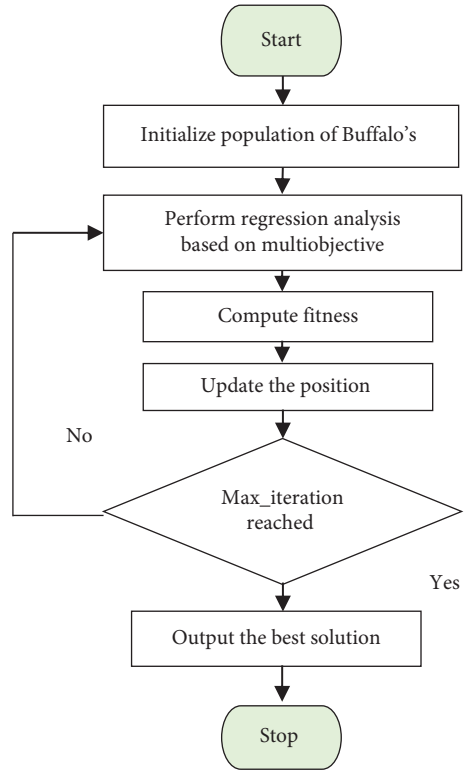


Figure 3: Flow diagram of Multicriteria Deming Regressive African Buffalo Optimization.

Q_{Fb} indicates the best fitness of buffalo's, and x_{bp} refers to the individual buffalo's best location. After that, the location of buffalos is updated as given below:

$$E_k(t+1) = \frac{[E_k + x_k]}{R}. \tag{13}$$

From (13), $E_k(t+1)$ indicates an updated location of buffalos, and R denotes a parameter value set as $\pm$ 0.5. If the convergence is not met, then go back to update buffalos; otherwise, stop the process.

Figure 3 illustrates the flow process of Multicriteria Deming Regressive African Buffalo Optimization for finding the optimal IP core. After finding the IP core, the mapping probability is done using equation (2) with the help of the injective mapping function. If the mapping function value is '1', then the optimal node is identified. Otherwise, the mapping function returns '0'. The entire process gets iterated until the maximum iteration is reached. As a result, the optimized neighboring core is selected for direct communication on 3D NoC.

The algorithmic process of the proposed MDRABOM technique is described as follows.

Algorithm 1 describes the step-by-step process of Multiobjective Deming Regressive African Buffalo Optimized Mapping to improve the efficiency of 3D NoC architecture design. Initially, the populations of IP cores are initialized in search space. After the initialization, the multicriteria function is measured for each IP core in the population. Then the Deming regression function is applied to analyze the multicriteria function. After the analysis, the

(1) Input: Benchmark dataset, Number of cores $C_i \in C_1, C_2, C_3, \ldots C_n$, Output: Find optimized IP core for NoC design
(2) Begin
(3) Step 1. Initialize the population of cores $C_i \in C_1, C_2, C_3, \ldots C_n$
(4) Step 2. For each core C_i
(5) Step 3. Compute multi-criteria function $MC_k(C_i) \in \varphi_1, \varphi_2, \varphi_3, \varphi_4$
(6) Step 4. Measure the fitness 'Q_F'
(7) Step 5. While ($t <$ Max_ iter)
(8) Step 6. If $(Q_F(C_i) < Q_F(C_j))$ then
(9) Step 7. Update buffalos' exploitation $x_k(t+1)$
(10) Step 8. Update the location of buffalos $E_k(t+1)$
(11) Step 9. End if
(12) Step 10. $t = t+1$
(13) Step 11. End while
(14) Step 12. Obtain the best solution
(15) Step 13. End For
(16) Step 14. Perform mapping $F: V_i \longrightarrow V_j$ based on probability $P(V_i, V_j)$ End.

Algorithm 1: Multiobjective Deming Regressive African Buffalo Optimized Mapping.

optimal one is selected through the fitness measure. If the fitness of the current core, i.e., $Q_F(C_i)$, is greater than the other $Q_F(C_j)$, the position of the 'i' th buffalo gets updated. Finally, the current best core is identified and mapped using the graphical model. This process gets repeated until it reaches the maximum iteration. After that, the mapping is performed based on the probability. In this way, an efficient mapping of cores in the 3D NoC architecture is performed with minimum time.

4. Experimental Settings

Experimental evaluation of proposed MDRABOM technique and two existing methods, namely, Stochastic Multiobjective Pareto-Optimization Framework [1] and SCSO [2] are implemented in Python. The algorithm processes are written by C++ code, and it is implemented in Python. In order to conduct the experiment, the MCNC Benchmark Netlists dataset is used [21]. The MCNC Benchmark Netlists are often used to perform the 3D Network-on-Chip architectures design. The benchmark circuits are in Yet Another Language (YAL) format. For the experimental consideration, the IP cores are taken from the benchmark dataset. The circuit description of the MCNC Benchmark dataset is given in Table 1.

Table 1 describes five circuit standards, namely, Apte, ami33, ami49, Xerox, and hp from the MCNC Benchmark Netlists. As shown in Table 1, different circuits comprise many IP cores.

5. Performance Results and Discussions

The performance discussion of the proposed MDRABOM technique and two existing methods, namely, Stochastic Multiobjective Pareto-Optimization Framework [1] and SCSO [2], are carried out with different parameters such as throughput, latency, and computation time. The performance results of the various parameters are discussed with help of a table and graphical representation.

Table 1: Standard MCNC benchmark circuits.

Circuits	Number of IP cores	Nets	I/O pad
Apte	9	97	73
ami33	33	123	42
ami49	49	408	22
Xerox	10	203	2
hp	11	83	45

Table 2: Comparison of throughput.

Circuits	Number of IP cores	Throughput (packets/cycles/IP block)		
		Proposed MDRABOMWDGT	SMPOF [1]	SCSO [2]
Apte	9	0.74	0.72	0.7
ami33	33	0.82	0.78	0.73
ami49	49	0.85	0.8	0.76
Xerox	10	0.76	0.73	0.7
Hp	11	0.79	0.74	0.72

5.1. Impact of Throughput. Throughput is the actual rate that the information (i.e., packets) is transferred between source-destination pairs in NoC. It also helps to perform the communication between the IP cores or blocks. The formula for calculating the throughput is given below:

$$T = \frac{\text{Packets transferred}}{\text{time}/IP\,\text{block}}, \tag{14}$$

where T denotes a throughput, and time is measured in terms of cycles. Therefore, the overall throughput is measured in the unit of the packets/cycles/IP block.

Table 2 describes the comparison of throughput for 5 different circuits, namely, Apte, ami33, ami49, Xerox, and hp taken from the MCNC Benchmark Netlists. The above table provides the performance of throughput using three different techniques, namely, MDRABOMWDG, Stochastic Multiobjective Pareto-Optimization Framework [1], and SCSO [2].

The observed results show that the proposed MDRABOMWDG offers higher performance when compared to

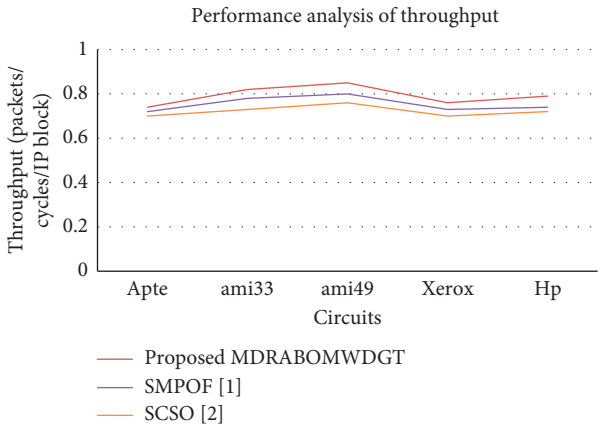


Figure 4: Performance analysis of throughput.

existing methods. Let us consider the Apte circuit with 9 IP cores. The throughput of the MDRABOMWDG is 0.74 packets/cycles/IP block.

The proposed MDRABOMWDG has obtained average percentage improvement for above circuits in throughput by 3.8% and 7% compared to SMPOF and SCSO, respectively.

The throughput of the conventional methods Stochastic Multiobjective Pareto-Optimization Framework [1] and SCSO [2] is 0.72 packets/cycles/IP block and 0.7 packets/cycles/IP block, respectively. The throughput is measured according to the packet injection rate. The packet injection rate is measured in terms of packets/cycles/IP block. Figure 4 shows the performance analysis of throughput for three different techniques, namely, MDRABOMWDG, Stochastic Multiobjective Pareto-Optimization Framework [1], and SCSO [2]. From the graphical representation, the throughput is measured for different circuits. The graphical plot indicates that the throughput of the MDRABOMWDG technique is higher than the other existing methods. The reason behind this improvement is to find the resource-optimized IP cores using the Multicriteria Deming Regressive African Buffalo Optimized Mapping technique. The resource-efficient IP cores are selected based on the multicriteria optimization technique such as energy consumption, area, power, and delay. The selected IP cores increase the packet transmission. Besides, the reduced number of average hop counts between source-destination pairs also improves the transmission by applying the weighted directive graph theory-based mapping.

5.2. Impact of Latency. The latency is referred to the average time taken to transmit the packets between source-destination pairs in NoC. Therefore, the latency is expressed as follows:

$$L = t_{avg}\,[\text{packet transmission}], \tag{15}$$

where L denotes a latency, and t_{avg} indicates an average time. The latency is measured in terms of clock cycles.

Table 3 reports the performance analysis of latency value for 5 various circuits, such as Apte, ami33, ami49, Xerox, and hp. The table provides the latency value of three different

Table 3: Comparison of latency.

Circuits	Number of IP cores	Latency (clock cycles)		
		Proposed MDRABOMWDG	SMPOF [1]	SCSO [2]
Apte	9	140	150	160
ami33	33	200	215	230
ami49	49	220	240	250
Xerox	10	160	170	180
Hp	11	175	180	200

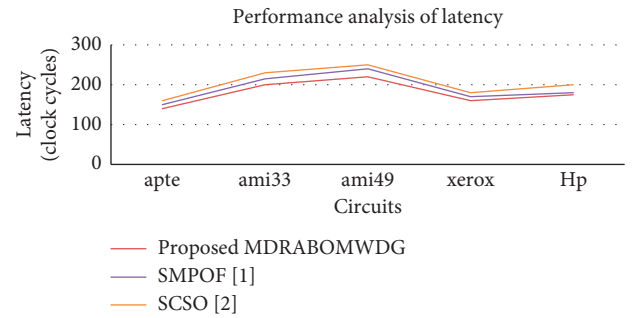


Figure 5: Performance analysis of latency.

methods as proposed MDRABOMWDG, Stochastic Multiobjective Pareto-Optimization Framework [1], and SCSO [2]. The latency of the data transfer using the proposed MDRABOM technique is minimized compared to existing techniques.

As shown in Table 3, the latency of the MDRABOMWDG technique is minimized when compared to other existing methods. This is due to the MDRABOMWDG technique that uses the weighted directive graph for NoC mapping. The mapping is done based on the Deming Regressive African Buffalo Optimization technique to find the optimal IP core for direct connection in the NoC design as shown in Figure 5. The lesser number of nodes in the weighted directive graph reduces the route length. This directs the transmission of the packet towards the destination with minimum latency. The proposed MDRABOMWDG has obtained average reduction in latency for above circuits by 12 and 24 clock cycles compared to SMPOF and SCSO, respectively.

5.3. Impact of Computation Time. The computation time is defined as the amount of time consumed by the algorithm to find the optimal core for efficient 3D NoC architectural design based on multiple objective functions. Therefore, the overall computation time is formulated as given below:

$$CT = [\text{end time} - \text{start time}]. \tag{16}$$

CT indicates a computation time. The computation time is measured in milliseconds (ms).

Table 4 reveals the performance results of the computation time for identifying the resource-optimized IP cores with respect to the number of circuits taken from the benchmark dataset. The observed results indicate that the

Table 4: Tabulation for computation time.

Circuits	Number of IP cores	Computation time (ms)		
		Proposed MDRABOMWDG	SMPOF [1]	SCSO [2]
Apte	9	11	14	16
ami33	33	22	25	28
ami49	49	28	31	33
Xerox	10	12	15	17
Hp	11	13	16	18

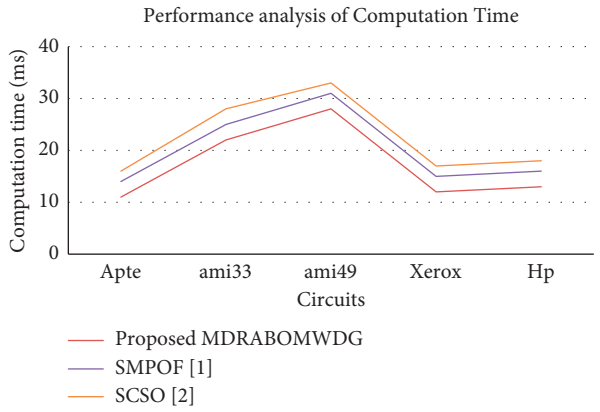


Figure 6: Performance analysis of computation time.

proposed MDRABOMWDG technique offers improved performance against the other two related approaches. This is proved through statistical analysis.

The experiment is conducted with Apte circuit and 9 IP cores, MDRABOMWDG technique has taken the 11*ms* of time to find the optimized core for 3D NoC design and the computation time of Stochastic Multiobjective Pareto-Optimization Framework [1] and SCSO [2] is observed to be 14 ms and 16 ms for finding the optimal IP core. Similarly, the other circuits are considered with different counts of IP core to conduct the experiment. Totally five runs are observed for each method with different counts of input IP cores as shown in Figure 6.

The computation time gets increased while increasing the number of cores taken as input. But comparatively, the proposed MDRABOMWDG technique minimizes the computation time. The appropriate reason is that the Multicriteria African Buffalo Optimization technique uses the Deming regression to analyze the multiple objective functions of each IP core. Based on the regression, the fitness is measured and identifies the best-fitted IP core with minimum time for 3D NoC design. The proposed MDRABOMWDG has obtained average reduction in computation time for above circuits by 3 and 5.2 ms compared to SMPOF and SCSO, respectively.

6. Conclusion and Future Work

6.1. Conclusion. In this paper, a novel MDRABOMWDG technique-based mapping approach was developed for 3D NoC design. The efficiency of the mapping approach is improved based on the multiple objective functions. The MDRABOMWDG technique is developed by the integration of a graph theory-based Multicriteria African Buffalo Optimization technique. Here multicriteria optimization problem is solved and performing the efficient IP core mapping on-chip that improves the throughput and reduces the latency of data communication. The optimization technique uses the Deming regression to analyze the various metrics and select the optimal cores from the population. This process helps to minimize the computation time of core mapping. The experimental evaluations are conducted to estimate the performance of the MDRABOMWDG technique over the two existing optimization methods. The experimental outcome illustrates that the MDRABOMWDG technique produces improved performance in minimization of latency and computation time and increases the throughput compared to the conventional methods.

6.2. Future Work. In future work we are planning to use Booksim 2.0 and Python simulator for energy parameters calculation and use evolutionary multiobjective optimization techniques for mapping both (regular and irregular) topologies, thereby achieving better results compared to the state-of-the-art methods.

Data Availability

The required data can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

Editing and writing were supported by Debre Tabor University, Ethiopia.

References

[1] T.J. Kao and W. Fink, "Stochastic multi-objective Pareto-optimization framework for fully automated ab initio network-on-chip design," *Journal of Systems Architecture*, Elsevier, vol. 103, , pp. 1–32, 2020.

[2] G. Peter and S. Bin Iderus, "Design of enhanced energy meter using GSM prepaid system and protective relays," *Materials Today Proceedings*, vol. 39, pp. 582–589, 2021.

[3] A. Alagarsamy, L. Gopalakrishnan, and S.-B. Ko, "KBMA: a knowledge-based multi-objective application mapping approach for 3D NoC," *IET Computers & Digital Techniques*, vol. 13, no. Issue 4, pp. 324–334, 2019.

[4] A. Sherine and G. Peter, "A novel biometric recognition system for FingerprintUsing polar harmonic transform," *International Journal of Pharmaceutical Research*, vol. 13, no. 01, (Crossref), Jan. 2021.

[5] M. Baharloo, K. Ahmad, M. Dolati, M.E PouyaShiri, and D. Rahmati, "Traffic-aware performance optimization in Real-time wireless network on chip," *Nano Communication Networks*, Elsevier, vol. 26, , pp. 1–16, 2020.

[6] R. Shamim, N. Singh and R. S Narde, Gonzalez-Hernandez, J.V Ganguly, S.G Venkatarman, and S. G. Kandlikar, Evaluation of wireless network-on-chip architectures with microchannel-based cooling in 3D multicore chips," *Sustainable Computing: Informatics and Systems*, Elsevier, vol. 21, , pp. 165–178, March 2019.

[7] A. Sherine, G. Peter, A. A. Stonier, K. Praghash, and V. Ganji, "CMY color spaced-based visual cryptography scheme for secret sharing of data," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–12, 2022.

[8] S. Manna, G. Mani, S. Ghildiyal et al., "Ant colony optimization tuned closed-loop optimal control intended for vehicle active suspension system," *IEEE Access*, vol. 10, pp. 53735–53745, 2022.

[9] Jefferson Silva, M. P. MárcioKreutz, and M. Da Costa-Abreu, "An investigation of latency prediction for NoC-based communication architectures using machine learning techniques," *The Journal of Supercomputing*, Springer, vol. 75, , p. 7573, 2019.

[10] S. P. Sathiyan, C. B. Pratap, A. A. Stonier et al., "Comprehensive assessment of electric vehicle development, deployment, and policy initiatives to reduce GHG emissions: opportunities and challenges," *IEEE Access*, vol. 10, pp. 53614–53639, 2022.

[11] W. Guan and C.A MiladGhorbaniMoghaddam, "Microprocessors and microsystems," *Quantifying the Impact of Uncertainty in Embedded Systems Mapping for NoC Based Architectures"*, Elsevier, vol. 80, , pp. 1–16, 2021.

[12] A. Bose and PrasunGhosal, "A low latency energy-efficient BFT based 3D NoC design with zone-based routing strategy," *Journal of Systems Architecture*, Elsevier, vol. 108, , p. 101738, 2020.

[13] N. A. Kumar, G. Shyni, G. Peter, A. A. Stonier, and V. Ganji, "Architecture of network-on-chip (NoC) for secure data routing using 4-H function of improved TACIT security algorithm," *Wireless Communications and Mobile Computing*, Article ID 4737569, 9 pages, 2022.

[14] P. K. Sharma, S. Biswas, and P. Mitra, "Energy efficient heuristic application mapping for 2-D mesh-based network-on-chip," *Microprocessors and Microsystems*, Elsevier, vol. 64, , pp. 88–100, 2019.

[15] G. Peter, A. Sherine, Y. Teekaraman, R. Kuppusamy, and A. Radhakrishnan, "Histogram shifting-based quick response steganography method for secure communication," *Wireless Communications and Mobile Computing*, pp. 1–11, 2022.

[16] W. Amin, S. Hussain, Anjum et al., "Performance evaluation of application mapping approaches for network-on-chip designs," *IEEE Access*, IEEE Access, vol. 8, pp. 63607–63631, , Volume8, 2020.

[17] T. Song, YiyuanXie, Y. Ye, Y. Du, B. Liu, and Y. Liu, "Gaussian-based optical networks-on-chip: performance analysis andoptimization," *Nano Communication Networks*, Elsevier, vol. 24, , pp. 1–13, 2020.

[18] P. K. R. Maddikunta, Q.-V. Pham, N. DeepaB et al., "Industry 5.0: a survey on enabling technologies and potential applications," *Journal of Industrial Information Integration*, vol. 26, p. 100257, 2022.

[19] B. Naresh Kumar Reddy and D.K.& B. VeenaVani, "Performance constrained multi-application network on-chip core mapping," *International Journal of Speech Technology*, Springer, vol. 22, , pp. Pages927–936, 2019.

[20] J. Wang and M. Ebrahimi, L. Huang, X. Xie, Q. Li, G. Li, and A. Jantsch, Efficient design-for-test approach for networks-on-chip," *IEEE Transactions on Computers, Volume68, Issue*, vol. 2, p. 198, 2019.

[21] G. Peter, K. Praghash, A. Sherine, and V. Ganji, "A combined PWM and AEM-based AC voltage controller for resistive loads," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–11, 2022.

[22] Y. Seyedeh, H. Mirmahaleh, and A. Masoud Rahmani, "DNN pruning and mapping on NoC-Based communication infrastructure," *Microelectronics Journal*, vol. 94, 2019.

[23] W. Gao, Z. Qian, and P. Zhou, "Reliability- and performance-driven mapping for regular 3D NoCs using a novel latency model and Simulated Allocation," *Integration*, vol. 65, 2019.

[24] S. K. Mandal, G. Krishnan, C. Chakrabarti, J.-S. Seo, and U. Y. Ogras, "A latency-optimized reconfigurable NoC for in-memory acceleration of DNNs," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 10, no. 3, pp. 362–375, Sept. 2020.

[25] J. Beneoluchi Odili, M. N. Mohmad Kahar, and S. Anwar, "African buffalo optimization: a swarm-intelligence technique," *Procedia Computer Science*, vol. 76, pp. 443–448, 2015.

[26] P. K. Sahu and S. Chattopadhyay, "A survey on application mapping strategies for network-on-chip design," *Journal of Systems Architecture*, vol. 59, no. 1, pp. 60–76, 2013.

[27] G. Peter, J. Livin, and A. Sherine, "Hybrid optimization algorithm based optimal resource allocation for cooperative cognitive radio network," *Array*, vol. 12, p. 100093, 2021.

[28] S. Li, S. Tian, Z. Kang et al., "A multi-objective LSM/NoC architecture co-design framework," *Journal of Systems Architecture*, vol. 116, 2021.

[29] R. P. Rangan, C. Maheswari, S. Vaisali et al., "Design, development and model analysis of lower extremity Exo-skeleton," *Medical Engineering & Physics*, vol. 106, p. 103830, 2022.

[30] A. K. Loganathan, A. Alexander Stonier, Y. Uma Maheswari, G. Peter, and T. Samraj Lawrence, "A real-time implementation of air audit system for compressors towards energy conservation: an industrial case study," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–12, 2022.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 5761815, 9 pages
https://doi.org/10.1155/2022/5761815

Research Article

Integrated Learning-Based Algorithm for Predicting Graduates' Employment Mental Health

Chen Dongrui,[1] **Wang Shengjie**,[2] **and Wen Kate**[2]

[1]*Wenzhou Polytechnic, Wenzhou 325200, Zhejiang, China*
[2]*Wenzhou University of Technology, Wenzhou 325200, Zhejiang, China*

Correspondence should be addressed to Chen Dongrui; 631605040113@mails.cqjtu.edu.cn

Received 23 April 2022; Revised 18 May 2022; Accepted 30 May 2022; Published 25 June 2022

Academic Editor: Mukesh Soni

Adaboost is a mental health prediction method that utilizes an integrated learning algorithm to address the current state of mental health issues among graduates in the workforce. The method first extracts the features of mental health test data, and after data cleaning and normalization, the data are mined and analyzed using a decision tree classifier. The Adaboost algorithm is then used to train the decision tree classifier for multiple iterations in order to improve its classification efficiency, and a mental health prognosis model is constructed. Using the model, 2780 students in the class of 2022 at a university were analyzed. The trial results demonstrated that the strategy was capable of identifying sensitive psychological disorders in a timely manner, providing a basis for making decisions and developing plans for mental health graduate students.

1. Introduction

With the rapid development of the economy and information technology in today's society, university graduates not only enjoy the convenience brought about by social progress and development but they also face a variety of pressures and negative information that are pervasive in contemporary society. As a result, the mental health of many university graduates is worrisome, and the proportion of graduates with psychological problems is rising, which has a negative effect on the healthy development of university graduates. The percentage of college graduates who suffer from psychological disorders is increasing, which has numerous negative effects on their healthy development [1, 2].

Workplace stress may result in mental health problems [3]. Currently, the number of graduates from Chinese colleges and universities is increasing annually, and employers' expectations of their employees' overall skillsets are also rising, putting graduates under greater pressure to find employment. After more than ten years of carefree education, graduates must adjust to a new social life and work stage. In addition to inner fear and anxiety about the future, they must also consider how to choose a future career path, how to develop the future prospects of the chosen industry, and how to apply for the position of their choice, etc. These are the issues that graduates must confront and consider, with many in the process of deciding and deliberating. Inevitably, graduates will experience anxiety, worry, indecision, and depression during the process of deciding and pondering, which will negatively impact their mental health. Second, many graduates will complete internships prior to employment. During this process, many graduates will perceive that the actual employment environment and job content are vastly different from their own expectations, resulting in a significant psychological gap that will create pessimism, negativity, and irritability regarding employment and future development, and negatively impact their psychological health. Again, college students are easily influenced by negative ideas prevalent in society and on the Internet, such as money worship and hedonism in the workplace, which results in unemployment and mental health issues. In addition, some college students will decide to start their own business if they lose certain employment opportunities or if they believe the employment unit is not optimal. However, because college students lack social and entrepreneurial experience, they are under a great deal of

stress as a result. Employment is a problem that every graduate must face, and it is also the segment most likely to cause mental health issues, requiring educators' focused attention.

Academic stress may result in mental health problems. In addition to selecting a career path, college grads must also decide whether or not to pursue further education. Since graduates are nearing the end of their academic careers, they face significant academic pressure that may lead to mental health issues. Additionally, they must complete their graduation project and defend their thesis. In the first place, graduates must complete academic requirements such as their graduation project and thesis. Many college students pay insufficient attention to their studies, and when they are confronted with a thesis at graduation, they have no idea where to begin. During the thesis writing process, people become anxious, unhappy, and confused, resulting in a great deal of stress and subsequent negative emotions and difficulties. Some students may be expelled from college if they are unable to earn the required number of credits or if they encounter genuine obstacles such as make-up exams and retakes in certain subjects before graduation. After graduating from university, many students choose to continue their education by enrolling in graduate school or studying abroad. Many students will experience confusion, pessimism, lack of confidence, depression, and other negative emotions, as well as doubts about their own decisions, which, if not alleviated in a timely manner, will be detrimental to their mental health [4–7].

Graduates are receiving training in mental health; however, it is insufficient to rely solely on professional counselors, as there are many graduates and fewer professional counselors, and many graduates are unwilling to confide their psychological problems to counselors. Therefore, it is necessary to use technical means to identify the mental health issues of students and to educate, guide, and treat them. In this context, the influence of mental disorders education research at universities has grown significantly. In data mining, algorithms are used to discover new information in large amounts of data. These algorithms can intelligently and efficiently analyze and process large amounts of data, and have become increasingly popular in the fields of intelligence retrieval, data analysis, and pattern recognition in recent years. Numerous researchers have shown a great deal of interest in the application of data mining techniques to the study of psychological issues raised by college students and to crisis early warning.

2. Related Work

2.1. Research on Heart Health. Yu Jiayuan [1, 2] investigated the use of rough sets integrating neural networks in the field of psychological assessment, applied rough sets to relevant data analysis, proposed a genetic algorithm-based fuzzy comprehensive evaluation technique, and conducted a useful analyses of the Likert scale's psychological measuring data; Yan Jie [3] applied the decision tree mining method to the study of college students' mental health data, and with the assistance of Clem, constructed a decision tree mining model using the C5.0 algorithm and the C5.0 algorithm for decision tree mining. The solution speed of decision tree is fast, and the classification accuracy is no less than the traditional BP neural network. The decision tree mining model was using C5.0 algorithm with the help of Clem entine12.0 platform to study university students' mental health; Gao et al. [8] used Hazy set theory to design the core set of factors and applied HMM model to predict college students' psychological crisis. In their study, Li et al. [9] employed the k-means cluster analysis method to examine students' performance. He [10] investigated the possibility of incorporating data mining into the investigation of the psychological difficulties of college students and employed the tree algorithm C4.5 was used to examine the data on psychological disorders among college students; Li [4] investigated and predicted the factors influencing college students' subhealth by the construction of a binary logistic incremental regression model and also a decision tree model; Huang [5] investigated and predicted the factors influencing college students' subhealth by establishing a binary logistic step the results of prior investigations guided the development and implementation of a college student the CART decision tree is used to create a psychological data mining model, the pattern recognition network (PRN) algorithm, and the algorithm for artificial neural networks (BP) [6]. For the reasons that classification techniques for mental data mining or connection analysis are used, these studies have been able to produce specific research results in their respective disciplines. The data from mental health checks are exceedingly complex, on the other hand, results in specific limitations in performance metrics such as classification accuracy and recall, as well as poor stability while processing diverse data sets. Following the principle of group optimization, we choose decision trees as the fundamental classifier, and we use the integrated learning algorithm Adaboost to develop a powerful classifier that we then use to data mining for college student mental health. Essentially, given a set of psychological training samples, different training sets are obtained by altering the distribution probability of each sample, and each training set is trained to obtain a basic classifier, which is then combined with other basic classifiers according to different weights in order to produce a strong classifier. In addition to avoiding overfitting problems that can arise with other methods, this method steadily lowers the upper bound of the training error rate as the number of iterations grows [7], the class label of even the classifier is massively changed as a result of this enhancement.

2.2. A Study on the Classification of Imbalanced Data. Despite the fact that data imbalance is common in tasks such as fraud detection [11] and medical diagnosis [12], Thabtah et al. [13] conducted an experimental investigation into the effect of varying category imbalance rates on classifier accuracy using a Bayesian approach. The bulk of the time, when training a typical classifier with category imbalance data, the classifier is biased towards the majority of the categories.

Research on data imbalance classification problems can be broadly divided into three groups, namely, data-level techniques, algorithm-level approaches, and cost-sensitive learning approaches. Data-level approaches are the most common. Undersampling, oversampling, and integrated sampling are all examples of data-level approaches, which are also known as resampling methods. In terms of algorithm-level strategies, Lee et al. [14] suggested fuzzy support vector machines are used to construct an overlap-sensitive margin (OSM) classifier and k-nearest neighbor (KNN) to class imbalance and classes overlap issues. Improving the classification performance of class imbalance data sets using an integration approach is also a popular solution. By combining the advantages of multiple classifiers, integration can produce better performance than using a single classifier. Chawla et al. [15] invented the SMOTEBoost, include SMOTE (synthetic minority oversampling technique) and adaptive enhancement (AdaBoost). Liu et al. [16] proposed easy-ensemble and balance-cascade. Cost-sensitive learning makes use of cost misclassification data and recognizes higher probability to identify the more profitable categories [14].

2.3. Integrated Learning. Integration learning compensates for the lack of robustness of a single model subject to large perturbations from sensitive samples by multiple weakly supervised models are being combined in order to produce a more complete strongly supervised model. Integrated learning is divided into two categories: bagging and boosting [17]. Where bagging is represented by the algorithm Random forest (RF) and boosting is characterized by adaptive weighting and represented by the algorithm Adaboost.

RF uses Bootstrap sampling to ensure that the training set of each decision tree is different and random, making the RF model less likely to fall into overfitting and has better noise immunity and a more stable prediction capability than a single model [15]. Eventually, the prediction results of all decision trees are averaged as the output of the model, but because the training set is usually unbalanced and small training errors are usually required for certain sample data with large contributions, certain decision trees with large contributions to the integrated model cannot play a greater role. Adaboost adaptively updates the weights of the training samples according to the performance of the weak learner error rate and finally integrates them through an ensemble strategy [16] to improve the accuracy of the model. Unlike RF, Adaboost has the same training set each round and is susceptible to perturbation by sensitive samples.

3. Method

3.1. Mental Health Prediction Model. To achieve these objectives, a psychological health forecasting model is developed. It reads in the instructional sample data, cleans and normalizes it, and then assigns initial weights to all samples; it then uses the C4.5 algorithm to train the sample set with the anyway D composed of all test weights as parameters to obtain a set of basic classifiers; it then uses the Adaboost algorithm to compute the graded error rate of basic classifiers and obtains the classifier weights as parameters once; and it then uses the Adaboost algorithm to obtain the classifier weights as parameters. After multiple training rounds, a significant classifier was obtained, which was then used to process mental health data, identify mental health variables, and provide accurate estimates. A model for predicting psychological well-being is depicted in Figure 1.

After data preprocessing, different sample sets are extracted and distributed with different weights and then entered into the C4.5 training classifier to obtain different classifications, after which the results of the classifications are summed up based on the various weights to obtain a final effective classifier, and the samples to be tested are input into the final strong classifier to obtain our predicted samples, as shown in Figure 1.

3.2. Key Techniques

3.2.1. Basic Classifier Generation Algorithm. This research uses the decision tree algorithm C4.5 as the fundamental classifier in order to generate classification algorithms. This model, also known as the rule-based inference model, establishes classification rules by learning from training samples and then categorizes new samples based on the classification rules established by the training samples. Compared to other classification models, it performs significantly better in non-numerical data processing and has the potential to reduce a substantial amount of data preprocessing effort. In this study, we construct decision trees using the C4.5 decision tree algorithm, the most popular decision tree technique. C4.5 is an enhanced decision tree method proposed by *J* Ross Quinlan on the basis of ID3, which is based on the theory of information and classifies data according to information entropy and information gain rate. C4.5 considers the rate of information gain as a criterion for selecting branching attributes, thereby overcoming the disadvantage of selecting characteristics with greater values based on information gain. Furthermore, C4.5 is able to effectively deal with incomplete and nondiscrete data.

Assuming that $S = \{X1, X2, ..., Xn\}$ represents the training sample set containing n student psychological test data samples, and attributeSet represents the attribute value data set of S, the steps of algorithm C4.5 Decision-Tree for constructing Decision-Tree are

(1) Create the Decision-Tree root node Root.

(2) In case all of the samples in S are members of the same category C, return Root as a leaf node and mark it as a member of the category C.

(3) If the attributeSet is empty or the number of remaining samples in S is less than the set threshold, return Root as a leaf node and mark Root as the class with the greatest number of categories among the samples contained in the node; otherwise, return Root as a leaf node and mark Root as the class with the greatest number of categories among the samples contained in the node.

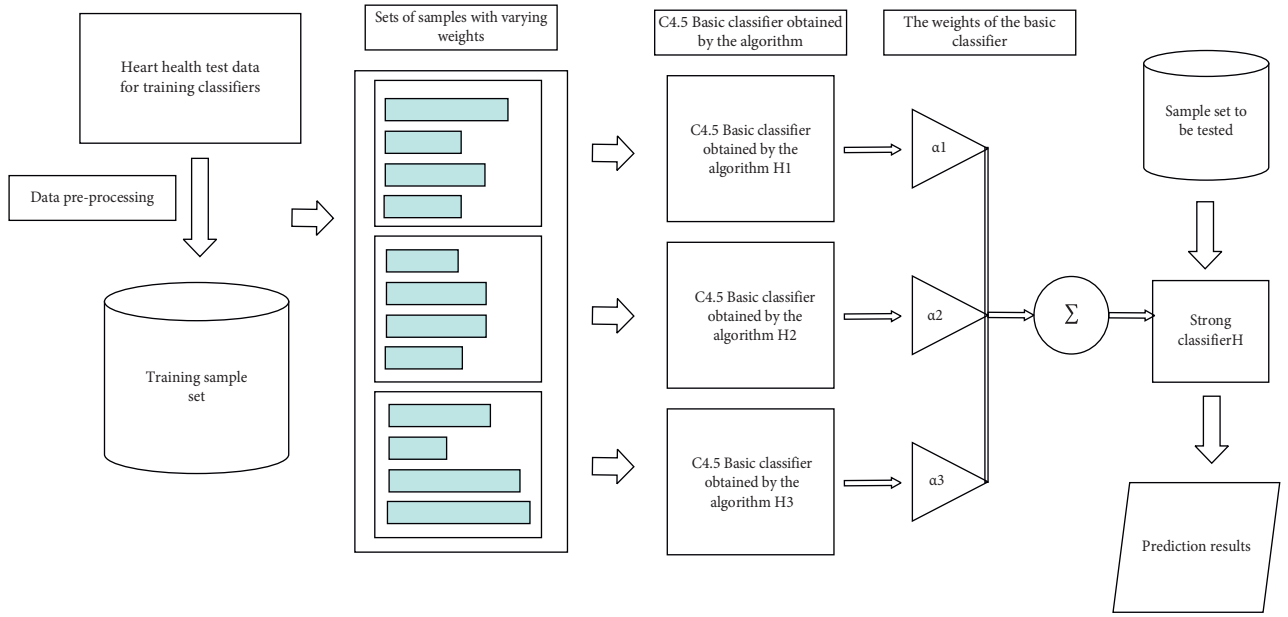


Figure 1: Mental health prediction model.

(4) Compute the gain ratio of the Gainratio(A) for each A ∈ attributeSet.

(5) The attribute A with the largest value of Gainratio (*A*) in the attributeSet is taken as the test attribute (best split attribute) Atest of Root.

(6) If the test attribute Atest is continuous, find the splitting threshold of the attribute.

(7) For each leaf node generated by a node Root based on Atest, check whether the sample subset S′ associated with this leaf node is empty before splitting the leaf node to generate a new leaf node and designating it as the class with the greatest number of classes in the samples contained within this node. If this is the case, the decision tree construction method C4.5 Decision-Tree (S′, S′, attributeSet) is conducted on this leaf node, and the splitting of the tree is carried out again.

(8) Calculate the classification error of each node and prune the Decision-Tree.

3.2.2. Building a Powerful Classifier by Adaboost. The Adaboost algorithm [17] is a modified Boosting algorithm that adapts the errors of weak classifiers (basic classifiers) obtained from weak learning to obtain an effective classifier with high degree of accuracy in classification. In this paper, the process of creating an effective classifier using Adaboost is as follows: a training sample is obtained after preprocessing the mental health test data, a series of basic classifiers are generated by repeatedly calling the C4.5 algorithm, and each basic classifier is assigned a certain weight according to its classification correctness, and then these classifiers are combined to obtain a strong classifier. When detecting new samples to be detected, the samples to be detected are given to each basic classifier in parallel for classification, the weights of the classifiers with the same classification result are added together, and the result with the highest weight is finally taken as the final output of the strong classifier. The steps for constructing a classifier using this algorithm are

In the first step, input a sample set containing N training samples of mental health data.

$$D = \{(x_1, y_1), \ldots, (x_i, y_i), \ldots, (x_N, y_N)\}, \quad x_I \in X,\ y_I \in Y. \tag{1}$$

X is sample set, Y is the sample category, $Y = \{-1, +1\}$, and each sample xi contains K-dimensional features $\{v_i\ (x_i), 1 \le j \le K\}$.

In the second step, the weight distribution of the sample set is initialized.

$$D_1 = \{w_1(x_1)\}, \quad w_1(x_1) = \frac{1}{N},\ i = 1, 2, \ldots, N. \tag{2}$$

In the third step, for any $t \in \{1, 2, ..., T\}$, T is the number of classifiers, a loop is executed.

(1) Using the samples with weight distribution D_t for learning, train a basic classifier h_j for each $v_i(x)$:

$$hj = \begin{cases} 1, & p_i v_i(x) < p_j \theta_j, \\ -1, & p_i v_i(x) \ge p_j \theta_j, \end{cases} \tag{3}$$

where θ_i is the threshold value, p_j is the bias value, and $p_j \in \{-1, 1\}$. Calculate the weighted.

Calculate the weighted error rate.

$$\varepsilon_j = \sum_{i=1}^{n} w_t(x_t)\left(h_j(x_i) \ne y_i\right). \tag{4}$$

The h_j corresponding to the minimum value of the error rate ε_j, and ε_t is used as the basic classifier h_t for this cycle.

(2) Calculate the weighting parameter α_t of h_t:

$$\alpha_t = \frac{1}{2}\ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right). \tag{5}$$

(3) Update the weight distribution using the weight α_t.

$$D_{t+1} = \{w_{t+1}(x_i)\}, \quad w_{t+1}(x_i) = \frac{w_t(x_i)e^{(-\alpha_t y_i h_t(x_i))}}{Z_t}. \tag{6}$$

Z_t as the normalization factor $Z_t = \sum w_t(x_i)e^{(-\alpha_t y_i h_t(x_i))}$.

In the fourth step, the final strong classifier H(x) is constructed using the T optimal basic classifiers obtained from T rounds of training.

$$H(x) = \sin\ g\left(\sum_{t=1} a_t h_t(x)\right). \tag{7}$$

3.3. Model Evaluation Methods. For measuring a model's performance, suitable evaluation metrics are required. Commonly employed classification model evaluation metrics may not be a good indicator of the performance of classification models for imbalanced data. Commonly employed evaluation metrics for unbalanced data classification models include specificity, G-mean, and CK [2]. Specificity measures the accuracy of classification for a small number of samples; G-mean and CK combine the accuracy of large and small categories to measure overall performance. As shown in Table 1, a confusion matrix of classification results can be formed for the dichotomous classification problem based on the combination of true and predicted categories.

Based on the confusion matrix, the accuracy, AUC value, specificity, G-mean and CK are used to evaluate the model in this paper.

(i) The accuracy rate is a measure of the percentage of correctly classified samples, i.e., the proportion of all correctly classified samples to the total number of samples, as shown in

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{8}$$

(ii) AUC value is calculated as the area of the ROC curve, and the ROC curve is a fit to the results of a model under different thresholds. Generally, the larger the AUC value, the better the prediction of the model.

(iii) The specificity is used to evaluate the correct rate of classification for a few classes of samples, as shown in

Table 1: Confusion matrix of binary classification results.

Real	Prediction	
	Positive	Negative
True	TP	TN
False	FP	FN

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \tag{9}$$

(iv) The geometric mean of both the accuracy in classification achieved by the majority and the accuracy in classification achieved by the minority is denoted by G-mean, with the goal of maximizing both classes while preserving a healthy balance between both the majority class classifier and the minorities class classification accuracy as illustrated in

$$\text{G} - \text{mean} = \sqrt{\frac{\text{TP} \times \text{TN}}{(\text{TN} + \text{FP})(\text{TP} + \text{FN})}}. \tag{10}$$

(v) CK, the Kappa Coefficient, can be used to evaluate the classification ability of the model, and the value range is [−1, 1], but usually [0, 1]. The larger the CK value is, the higher the consistency between the prediction result and the actual result, and the better the performance of the classification model, as shown in

$$\text{CK} = \frac{\text{Accuracy} - \text{CK}_c}{1 - \text{CK}_c}. \tag{11}$$

where accuracy is the accuracy rate; CK_c is shown in

$$\text{CK}_c = \frac{(\text{TP} + \text{FN})(\text{TP} + \text{FP}) + (\text{FP} + \text{TN})(\text{FN} + \text{TN})}{(\text{TP} + \text{FN} + \text{TN} + \text{FP})^2}. \tag{12}$$

4. Experimental Results and Analysis

The data used for the experiment in this paper were collected from the mental health test and personality test data of students in the class of 2022 at a university. These tests were administered online with the cooperation of the university's counseling center, and 2,780 valid questionnaires were collected. The mental health examination scale utilized the Symptom Self-Rating Scale SCL90, a commonly used outpatient examination scale for mental disorders and psychological disorders. The SCL90 scale has 90 entries, 10 factors were considered, including somatization, obsessive-compulsion symptoms, prosocial behavior, depression, anxiety, anger; terror; paranoia; psychoticism; and others, with each entry having 5 subscales. There are five options for every item: 1 (none), 2 (very mild), 3 (moderate), 4 (severe), and 5 (severe). The Eysenck Personality Questionnaire (EPQ), which consists of four scales: the Extraversion Scale (E), the Emotional Disposition Scale (N), the Psychological Quality Scale (P), and the Validity Scale, was used to

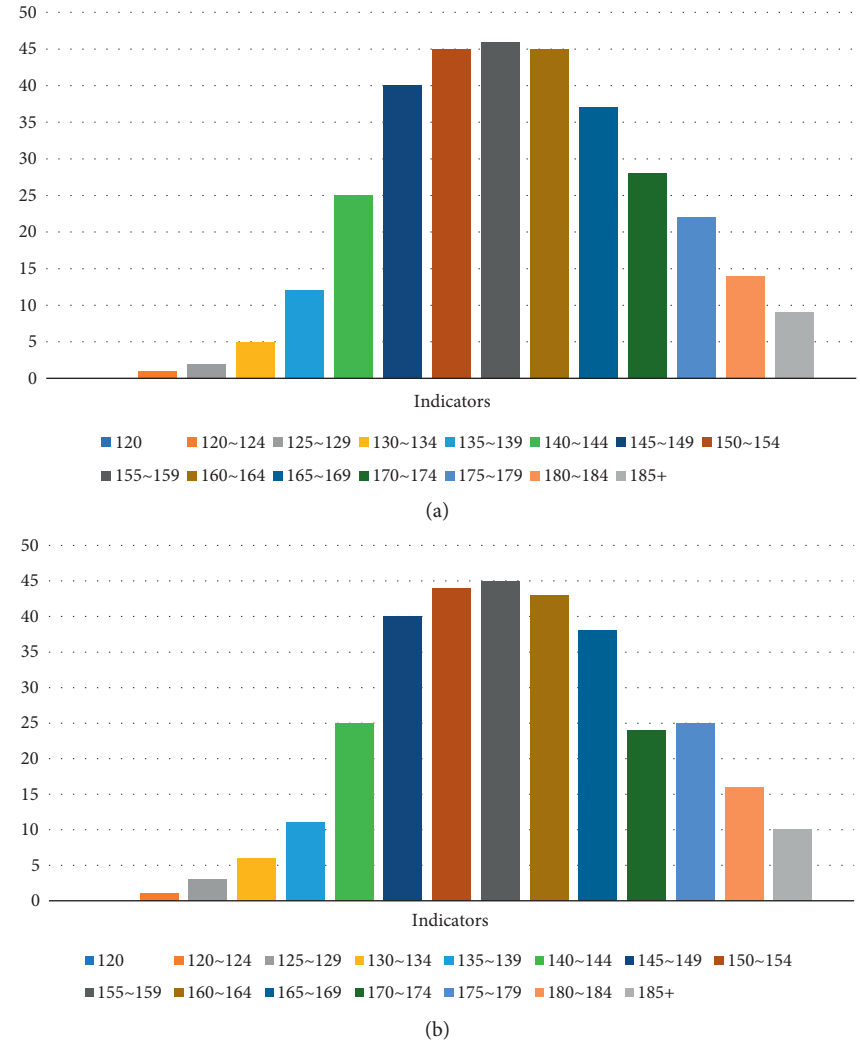


Figure 2: Data distribution map. (a) Before deleting data, (b) after deleting data.

administer the personality test (L). The first three scales, which are independent of one another, reflect the three distinct dimensions of human personality, whereas the L is a validity scale that reflects the hypothetical personality traits and the level of social simplicity and naiveté. In this study, the Chinese version of the Eysenck Personality Questionnaire Short Form Scale was utilized (EPQ-RSC).

Due to the fact that a small number of test scales may contain missing items or irregular filling, data cleansing is required prior to processing in order to verify data consistency, perform data normalization, eliminate invalid values, account for missing values, etc. Since the possible response options for the scale entries are discrete values, the most frequent value of that item is counted to fill in the missing values. In order to facilitate the algorithm for data processing, coded representations of all questionnaire entries' values were abstracted. The database management system SQL Server 2012 was used to store data, and the database tables "SCLTable" and "EPQTable" were created to store the symptom self-assessment scale and Eysenck personality questionnaire data, respectively. The relationship between the two tables was established by the "StudentID" attribute. The experiment was conducted on a Dell Inspiron 3650-D1838 using a Python-programmed algorithm. A training set and a test set were created from the test samples. There were 1,800 samples for training and 980 for testing. In order to ensure the generality of the algorithm, the hyperparameter settings in this paper are consistent with those in Reference 4.

Table 2: Some parameters of GBDT.

Parameter	Value
Learning_rate	0.1
Max_depth	4
Min_sample_split	4
N_estimators	100

The impact of erasing these records on other variables was examined through additional analysis. According to the results, the effect of deleting these records on the distribution of other variables was insignificant, so it was deemed acceptable to delete this portion of the data. As shown in

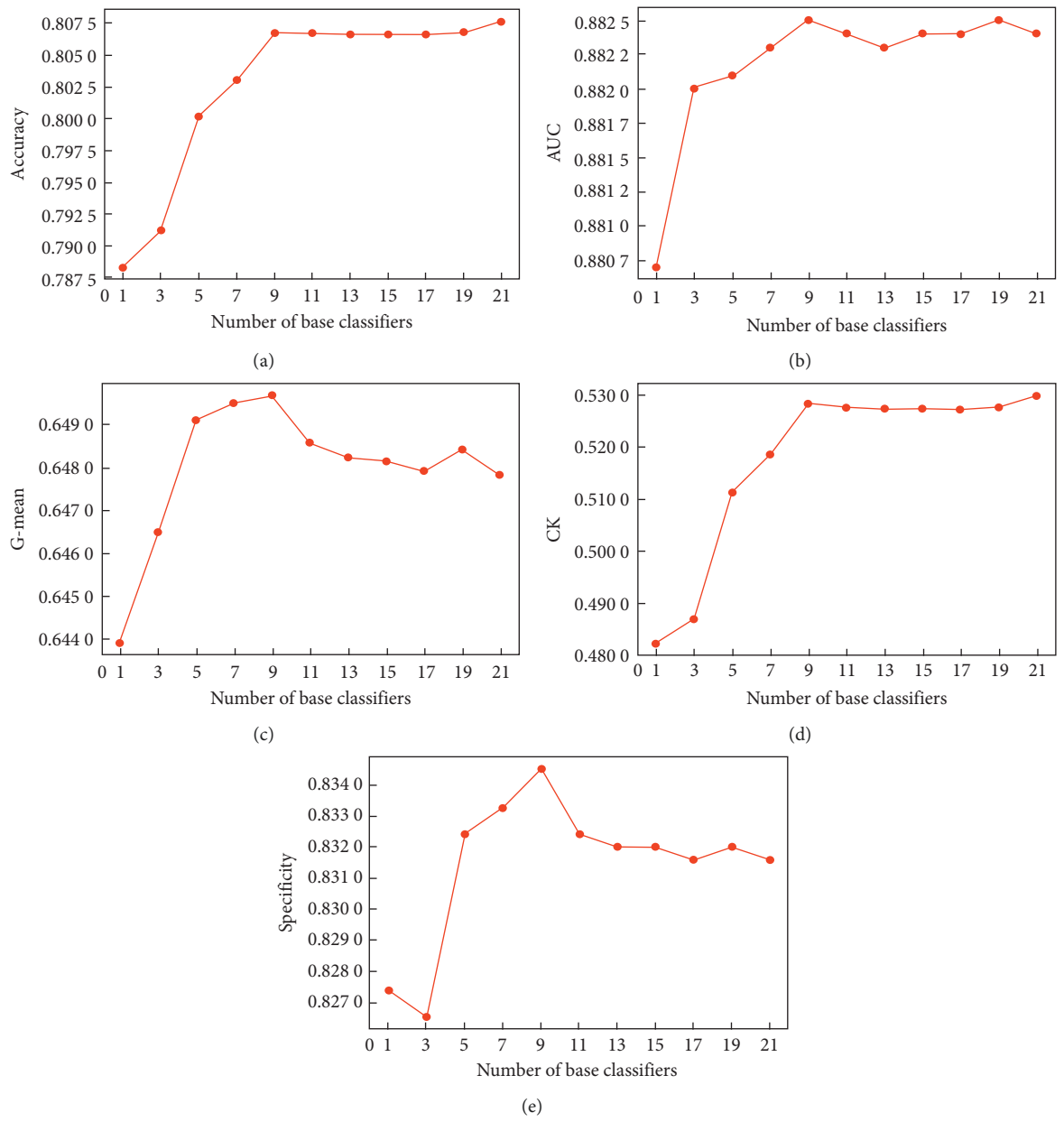


Figure 3: Evaluation indicators for various number-based classifiers.

Figure 2, there was no significant difference between the distribution of the variable age before and after the deletion of the data. In addition, the values of the age group were compared by calculating the weighted average instead of the mean and the standard deviation, and it was found that the differences were minimal.

The base classifier is GBDT, and the Gradient Boosting Classifier algorithm in "sklearn. Ensemble" is called, and some relevant parameters are obtained by the grid search algorithm, and the others are default values, as shown in Table 2. For the determination of the number of base classifiers T, the variation of the evaluation index of different number of base classifiers is experimentally explored, as shown in Figure 2. It is known that the model shows better results on each evaluation index when T is 9, so the hyper parameter T is set to 9.

For the determination of the number of base classifiers T, the variation of evaluation indexes with different numbers of base classifiers is experimentally explored, as shown in Figure 3. It can be seen that the model shows better results on each evaluation index when T is 9, so the hyper parameter T is set to 9.

After determining the data division and parameters, the performance of the model is evaluated. Numerous machine learning algorithms have been applied to medical prediction and have demonstrated good performance; these algorithms can be broadly categorized as either single models or integrated models. Single models consist primarily of Decision Tree (DT), Logistic Regression (LR), and K-Nearest Neighbor (KNN), among others. Also included is the more popular Artificial Neural Network (ANN). The majority of

Table 3: Using the SEER gastric cancer data set, we compared the performance of various algorithms.

Type	Algorithms	Accuracy	AUC	Specificity	G-mean	CK
Single model	DT [18]	0.780	0.667	0.472	0.406	0.329
	LR [19]	0.825	0.835	0.274	0.265	0.310
	KNN [20]	0.823	0.840	0.352	0.333	0.352
	ANN [21]	0.836	0.852	0.380	0.362	0.400
Integrated model	RF [22]	0.844	0.877	0.460	0.434	0.457
	Boosting [17]	0.843	0.876	0.463	0.436	0.457
	GBDT [23]	0.848	0.880	0.480	0.452	0.476
Data imbalance processing + integrated model	SMOTETomek + RF	0.827	0.867	0.666	0.578	0.502
	SMOTETomek + boosting	0.805	0.860	0.708	0.587	0.474
	SMOTETomek + GBDT	0.815	0.868	0.727	0.609	0.498
	BalancedRandomForest	0.759	0.851	0.819	0.608	0.432
	Easy-Ensemble	0.787	0.877	0.819	0.638	0.478
	Ours	0.808	0.883	0.835	0.650	0.528

the integration models are Random Forest (RF), AdaBoost, and Gradient Boosted Decision Tree (GBDT). The primary methods for processing unbalanced data are oversampling, downsampling, and integrated sampling. The SMOTETomek method of integrated sampling incorporates oversampling and undersampling and performs reasonably well. The integrated model demonstrates superior classification performance, and SMOTETomek in conjunction with the integrated model can effectively address the data imbalance issue. This article also compares the model to SMOTETomek in conjunction with the integrated model to address data imbalance. Table 3 compares the effectiveness of various algorithms utilizing the SEER Gastric Cancer Data set.

5. Conclusion

Modern college students are subjected to unprecedented levels of stress from their studies, jobs, and interpersonal relationships, and the resulting psychological issues are becoming increasingly prevalent, with occasional violent acts occurring as a result. In this study, we provide a method for predicting the mental health of graduates by employing the decision tree algorithm C4.5 as the primary classifier and the Adaboost classifier developed by the integrated learning algorithm as the secondary classifier. The experimental results of the algorithm demonstrate that the accuracy and recall of the classification indices of the integrated learning classifier are significantly improved as a result of employing this strategy. This algorithm may assist mental health counseling teachers, student management, and counselors in understanding students' psychological development, the main symptoms and personality traits of mental health problems, and focusing on students who may be suffering from mental health problems and providing them with more care, guidance, and psychological treatment to prevent the problems from occurring. In the future, we will introduce deep learning models to predict mental health.

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.

Acknowledgments

This study was supported by the WenzhouPolytechnicof 2021 Wenzhou Polytechnic's Key Projectupon Construction of CPC and Research of Ideological and Political Education: The Inquiry and Promotion of Vocational College Student's Positive Psychological Traits under the Perspective of Craftmanship (WZYDS202101), and General Scientific Research Project of Zhejiang Provincial Department of Education in 2021: Research on the Loss of Graduates of Wenzhou University and Countermeasures—Taking Wenzhou Institute of Technology as an Example (Y202148016).

References

[1] M. M. Hasan, M. R. Haque, and M. M. J. Kabir, "Breast cancer DiagnosisModels using PCA and different neural network architectures," in *Proceedings of the 2019 International Conference on Computer,Communication, Chemical, Materials and Electronic EngineeringIC4ME2*, July 2019.

[2] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[3] E. Ahmed, Y. Soumaya, and S. O. Mohamed, "Bidirectional handshaking LSTM for remaining useful life prediction," *Neurocomputing*, vol. 323, pp. 148–156, 2018.

[4] M. Li, *Analysis of College Students' Subhealth Status and Influencing Factors Based on Logistic and Decision Tree Models*, Beijing Jiaotong University, Beijing, 2015.

[5] Z. Huang, *The Application of Data Mining in the Prevention of Psychological Crisis of College Students*, Wuhan University of Science and Technology, Wuhan, 2015.

[6] Z. Fu, "Analysis on the effectiveness of AdaBoost," *Computer Research and Development*, no. 10, pp. 1747–1755, 2008.

[7] Y. Cao, Q.-G. Miao, J.-C. Liu, and G. Lin, "Progress and prospects of AdaBoost algorithm research," *Journal of Automation*, no. 6, pp. 745–758, 2013.

[8] G. Gao, L. Jiang, and Y. Ding, "Psychological crisis prediction of college students based on vague set and Hidden Markov," *Journal of Shanghai University of Technology*, vol. 2, no. 4, pp. 380–384, 2015.

[9] Z. Li, Y. Zhang, and H. Hu, "The application of cluster analysis in the management of college students' mental health," *Journal of Hubei Engineering College*, vol. 5, no. 6, pp. 53–57, 2014.

[10] G. He, *Research on the Application of Data Mining Technology in the Analysis of College Students' Psychological Problems*, Hebei University, Baoding, 2013.

[11] W. Kim, K. S. Kim, and R. W. Park, "Nomogram of naive BayesianModel for recurrence prediction of breast cancer," *HealthcareInformatics Research*, vol. 22, no. 2, pp. 89–94, 2016.

[12] W. Kim, K. S. Kim, J. E. Lee et al., "Development of novel BreastCancer recurrence prediction model using support VectorMachine," *Journal of Breast Cancer*, vol. 15, no. 2, pp. 230–238, 2012.

[13] C. M. Lynch, B. Abdollahi, J. D. Fuqua et al., "Prediction of LungCancer patient survival via supervised machine LearningClassification techniques," *International Journal of Medical Informatics*, vol. 108, pp. 1–8, 2017.

[14] Z. Zhao, B. Liang, X. Wang, and W. Lu, "Remaining useful life prediction of aircraft engine based on degradation pattern learning," *Reliability Engineering & System Safety*, vol. 164, pp. 74–83, 2017.

[15] E. Balaban, A. Saxena, S. Narasimhan et al., "Prognostic health-management system development for electromechanical actuators," *Journal of Aerospace Information Systems*, vol. 12, no. 3, pp. 329–344, 2015.

[16] R. Guo and Q. Gan, "Prognostics for a leaking hydraulic actuator based on the f-distribution particle filter," *IEEE Access*, vol. 5, pp. 22409–22420, 2017.

[17] M. Pecht and J. Cu, "Physics-of-failure-based prognostics for electronic products," *Transactions of the Institute of Measurement and Control*, vol. 31, no. 3, pp. 309–322, 2009.

[18] A. Henga, S. Zhang, A. C. C. Tan, and J. Mathew, "Rotating machinery prognostics: state of the art, challenges and opportunities," *Mechanical Systems and Signal Processing*, vol. 23, no. 3, pp. 724–739, 2009.

[19] F. Zhao, T. Zhigang, and Y. Zeng, "Uncertainty quantification in gear remaining useful life prediction through an integrated prognostics method," *IEEE Transactions on Reliability*, vol. 62, no. 1, pp. 146–159, 2013.

[20] R. Khelif, B. Chebel-Morello, S. Malinowski, E. Laajili, F. Fnaiech, and N. Zerhouni, "Direct remaining useful life estimation based on support vector regression," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 3, pp. 2276–2285, 2016.

[21] T. Berghout, L-H. Mouss, O. Kadri, L. Saïdi, and M. Benbouzid, "Aircraft engines remaining useful life prediction with an improved online sequential extreme learning machine," *Applied Sciences*, vol. 10, no. 3, p. 1062, 2020.

[22] Z. Tian, "An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring," *Journal of Intelligent Manufacturing*, vol. 23, no. 2, pp. 227–237, 2012.

[23] G. Liang, N. Li, F. Jia, Y. Lei, and J. Lin, "A recurrent neural-network based health indicator for remaining useful life prediction of bearings," *Neurocomputing*, vol. 240, no. C, pp. 98–109, 2017.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 5248338, 9 pages
https://doi.org/10.1155/2022/5248338

Research Article

A ResNet50-Based Approach to Detect Multiple Types of Knee Tears Using MRIs

Shilpa Sharma,[1] **Mohammad Umer**,[1] **Avinash Bhagat**,[2] **Jeevan Bala**,[1] **Punam Rattan**,[2] **and Abdul Wahab Rahmani**[3]

[1]*School of Computer Science Engineering, Lovely Professional University, Phagwara 144411, India*
[2]*School of Computer Applications, Lovely Professional University, Phagwara 144411, India*
[3]*Isteqlal Institute of Higher Education, Kabul, Afghanistan*

Correspondence should be addressed to Abdul Wahab Rahmani; ab.wahab.professor@isteqlal.edu.af

Received 13 April 2022; Revised 8 May 2022; Accepted 18 May 2022; Published 14 June 2022

Academic Editor: Mukesh Soni

The recommended tool for assessing knee injury is magnetic resonance imaging (MRI). However, knee MRI interpretation takes time and is vulnerable to clinical errors and inconsistency. A deep learning automated technique for reading knee MRI might help physicians identify high-risk patients and make diagnosis easier. In this study, we have proposed a deep learning-based model to detect ACL and meniscus tears and other knee abnormalities. At its core, this model is based on the ResNet50 transfer learning technique. In this paper, we have focused to present a ResNet50-based model for detecting different knee problems using MRIs. The best models for every option achieved the objectives that were probably similar. The models were developed using 18, 3, and 1 slice. These models' outcomes were rather startling. The AUC findings obtained with 1 slice per MRI exam were equivalent to those obtained with 18 and 3 slices and, in some cases, were significantly better. The dataset used in this model is from Stanford University. We trained this model in three different settings of MRI slices (18, 3, and 1). The best results that our models were able to achieve were when trained using 3 slices of each MRI sample. The area under the receiver operating characteristic curve, or AUC curve, values that our best models were able to achieve for detecting ACL, meniscus, and other knee abnormalities are 0.87, 0.82, and 0.90, respectively. The results of our models are comparable to some state-of-the-art models. These models are very fast and efficient to train and hence will be helpful to doctors for making an effective and fast diagnosis based on knee MRIs.

1. Introduction

Magnetic resonance imaging (MRI) exams are amongst the most prevalent diagnostic methods, with approximately 40 million MRIs performed in the United States each year [1]. MRIs performed on the knees can help in identifying abnormalities such as anterior cruciate ligament or ACL tears, meniscus tears, and other similar abnormalities. Deep learning methods are increasingly being used in the evaluation of medical images. A deep learning automated image analysis technology might help healthcare professionals to detect high-risk patients and make diagnostics simpler. Magnetic resonance imaging scans require significantly more time to conduct compared to CT scans, and patient satisfaction is a concern, which may be aggravated by the technology. Even when conducted by experienced radiologists, proper interpretation of knee MRI is time-consuming and prone to errors due to the number and complexity of images in each exam [2]. MRIs aid in the identification of numerous knee ailments as well as in differentiating between different kinds of knee tears. Magnetic resonance imaging of the knee can aid in the detection of anomalies such as anterior cruciate ligament (ACL) injury, meniscus tears, and other related conditions. Such models are incredibly quick and efficient to develop, which will enable clinicians to make effective and timely diagnoses based on knee MRIs. An automated system to detect knee abnormalities can prove to be very useful. In recent years, we have seen an increasing

application of deep learning models in medical image assessment. These models have been found to be better or at par with human experts at diagnosing diseases in most cases.

In this study, we propose a ResNet50-based model to detect various abnormalities in knees using MRI. MRIs are helpful in the diagnosis of various knee injuries and help in distinguishing different types of knee tears from each other [3]. The dataset used in this study is the MRNet dataset, provided by Stanford University, which contains 1370 MRI exams of the knees. The dataset contains three types of MRIs taken from three planes, namely, sagittal, coronal, and axial planes. This dataset contains MRIs for ACL tears, meniscus tears, and other knee abnormalities. To detect these tears, MRIs from all three planes are required [4]. This dataset can be approached in multiple ways. In this study, we created a base model using ResNet50 and then we trained it 27 times using different combinations of input data.

The rest of this paper is laid out as follows: The motivation for this work is in the following paragraph. Work related to this paper is described in Section 2. Methodology and materials used in this study are described in Section 3. Section 4 has the details regarding the results and the discussion based on these results. Section 5 contains the comparison of a few state-of-the-art models with our proposed model along with the comparison of all models employed in this study. Section 6 has the details regarding to the future scope of this study and its conclusions.

1.1. Motivation. The usage of deep learning in medical image assessment is growing day by day, but most of these models are used for analysis of brain, chest, and breast images [5]. Other parts of the human body do not have nearly as many research papers written about them as the ones mentioned above. One of the reasons for this is the nonavailability of data. We, in this study, decided to work with knee MRIs as the number of research papers published related to knee MRI analysis using deep learning is very few. Knees are essential in the human body as they allow us to walk. If someone's knee hurts, doctors usually prescribe an MRI. However, a nonexpert human cannot understand an MRI. An automated approach for detecting knee problems might be highly valuable. Deep learning methods have been widely used in medical image identification in recent years. Throughout many situations, computational models have been proven to be more effective than or on the level with human specialists at identifying illnesses.

Even experts sometimes face difficulties in interpreting these MRIs. Therefore, we decided to create a deep learning model that, just by giving the image of the MRI as input, would be able to tell if the MRI is normal or not. It would also give us an idea about how well a deep learning-based model performs compared to human experts.

2. Related Work

Due to the difficulty of the task for a person to do, the abundance of data, and the success of deep learning, there is already a large amount of research work done on employing deep learning for MRI interpretation [6–8]. A perfect prognosis of cardiac illness might save a person's life, whereas wrong predictions can lead to death. For example, to locate and categorize a brain tumor, deep learning techniques can be used. Breast cancer is one of the most common tumors in women. The use of MRI for early identification of breast cancer helps patients recover more quickly. Although the number of research studies published related to medical image analysis using deep learning is large, most of these research studies are focused toward brain tumors, chest diseases, etc. [9]. Knees and bones in general do not have nearly as many studies associated with them. This became one of the reasons why we decided to work on knee MRI analysis in this study.

Researchers in one study presented a thorough study on the use of deep learning for MRI image manipulation, which included a wide range of MRI image applications. This study describes the numerous issues that individuals and healthcare providers are encountering as a result of COVID-19. Several methods for controlling the influence of COVID-19 employing Internet of things are explored in this study [10]. One of the statements to emerge from this report was that there are many different deep learning architectures from which researchers can choose when to use deep learning to interpret MRI data.

In another study on knee MRIs, researchers developed a CNN that predicted the likelihood of having an abnormal exam when provided with a knee MRI sample [11]. A CNN model called MRNet was trained for each plane: axial, coronal, and sagittal, and for each form of knee injury, the probabilities from the model were blended to get a single probability.

ELNet, a CNN-based model designed for early knee MRI diagnosis, was presented in another study [12]. Unlike most techniques for knee MRI analysis, researchers trained ELNet from the ground up rather than utilizing a transfer learning method. The ELNet model is lightweight and performs as well as other models due to the merging of multislice normalization and Blur Pool operations.

Researchers calculated the likelihood of ACL injury, meniscus tear, and knee anomalies using the ResNet18, ResNet50, and ResNet152 models in another study. Two approaches were discussed in this study. The first process was to consider only the center slice of each MRI sample. This method assumed that the slices in the middle of the series would include more significant data than those at the beginning and end. The second technique used basic mathematical operations on a certain number of pictures to interpolate the number of slices to a specific number. The second approach seemed to produce better results than just utilizing center slices. Unpermitted consumers misappropriate multimedia material by distributing it on multiple web domains to earn more money fraudulently without any assistance of the original copyright owners [13].

Because of its quality of local connections and shared weights, a deep learning model like CNN can infer the representation of pictures [14]. Another research based on the MRNet dataset investigated two deep CNN models (VGG16 and VGG19) with a transfer learning strategy to

predict knee injury [15, 16]. Chest cancer is the most common intrusive cancer in women as well as the second leading reason of cancer mortality in women. It could be normal or abnormal [17].

3. Materials and Methods

3.1. Dataset. The dataset that we used in this study was provided by researchers from Stanford University and has been publicly made available since 2018 [18]. The dataset contained 1370 MRI exams from 1088 patients, which included 1,104 abnormal exams, with 319 ACL tears and 508 meniscal tears. These were examined and labeled by human experts. The dataset was split into a training set (1130), a validation set (120), and a test set (120). The test set is not publicly available. The dataset comprises 80.6% abnormal exams, 23.3% ACL (anterior cruciate ligament) tears, and 37.1% meniscus tears. The labels for each MRI were manually extracted from the patient's clinical files. The MRNet dataset, which was contributed by Stanford University and comprises 1370 MRI tests on knees, was used in the investigation. The collection comprises three different types of MRIs acquired in three different planes: sagittal, coronal, and axial. MRIs for ACL tears, meniscus tears, and other knee anomalies were included in this collection. Because patients who received an MRI were more likely to develop a knee injury, the dataset was heavily skewed in favor of classification of injuries. The organization of the dataset is shown in Figure 1.

Each exam in the dataset consists of multiple axial, sagittal, and coronal slices. The sagittal plane, which is on the *xz*-plane, splits the knee into the left and right halves. The coronal plane, which is on the *xy*-plane, splits the knee into front and rear halves, and the axial plane is parallel to the ground and splits the knee into the top and bottom sections. Figure 2 depicts these planes.

All exams in the dataset are of .npy format. The dataset contains labels for each exam in three different .csv files, namely, abnormal.csv, meniscus.csv, and acl.csv. There are separate labels for the validation set too. The shape of each file is (x, 256, 256) and (x, 256, 256) where x represents the number of slices of each exam. The number of slices in the dataset varied greatly, for example, for the same patient, an exam in the axial plane had 44 slices and 36 slices in the coronal and sagittal planes.

3.2. Data Preprocessing. Preprocessing is a word that refers to activities performed on images at the most fundamental level. The intensity of a picture is frequently specified as a matrix of image function values. Geometric transformations of pictures (e.g., rotation and scaling) are some of the preprocessing techniques. The purpose of preprocessing is to enhance the image information by removing undesired noise or boosting certain crucial picture elements for subsequent processing [19].

In this study, we initially applied some of the traditional preprocessing techniques like Gaussian blur to remove the noise from the MRI samples, a Laplacian filter to detect edges, and a sharpening filter. The resulting samples of these preprocessing techniques, when passed through our model, were not providing satisfactory performance. We tried different combinations of these techniques and used different kernels, but the results were close to the results we were getting without applying any preprocessing techniques, or in some rare cases, even worse than that. Therefore, we decided not to go through this route [20].

As discussed above, the MRI samples in our dataset had different number of slices for each sample. To fix this, we used interpolation of MRI samples. In simple words, the use of known points to guess unknown points is referred to as interpolation. Interpolation is frequently used in the picture field to change the size of an image. From the points in an old image matrix, the points in a new image matrix are calculated and added [21]. Various interpolation algorithms are employed. Here we have used a simple interpolation method called "zoom" which is included in the SciPy library. We have modified each MRI sample to contain 18 slices. Due to the computational limits of our computing hardware, the maximum number of slices that we included in our study was 18 slices per MRI exam, but this does not take away anything from the results of this study as in our observation: after using more than 3 slices, we were not getting any major differences in the results. Based on this observation, we decided to train our models using 3 and only 1 slice of each MRI sample. The results produced by using only 1 slice were pretty comparable to the results that we achieved by using 3 or 18 slices. We got the zoom values by dividing the required shape, i.e., 18,256,256 (in the case of 18 slices) by the original shape, i.e., x, 256, 256, where x represents the number of slices in the original MRI sample. We created one NumPy file for each type of MRI sample. For example, we created one big file for axial samples, one file for sagittal samples, and one file for coronal samples. This made it easier to pass all data into our model or do any modifications to it.

3.3. Training Model. We trained our models in three different settings:

(i) Using 18 slices of each MRI sample
(ii) Using 3 slices of each MRI sample
(iii) Using only 1 slice of each MRI sample

The rest of the setting was the same for every model. We built our model based on the ResNet50 model and used it as a feature extractor. We used the pretrained ImageNet weights. The input to this base model is of the shape x, 3, 256, 256, 3, 256, 256 where x denotes the number of slices. 3 denotes the number of channels, and 256, 256 denotes the shape of the 2D MRI slice. By passing this through our feature extractor, we get a tensor of shape x 2048, 8, 8. We used a global average pool layer to transform the data into the shape x 2048. We then finally reshaped the data into 2048, x, 1. We did this for both training and validation sets. Therefore, our final training data were of shape 1130, 2048, x, 1 and our final validation data were of shape 120, 2048, x, 1, where 1130 and 120 represent the number of samples in each set, respectively.

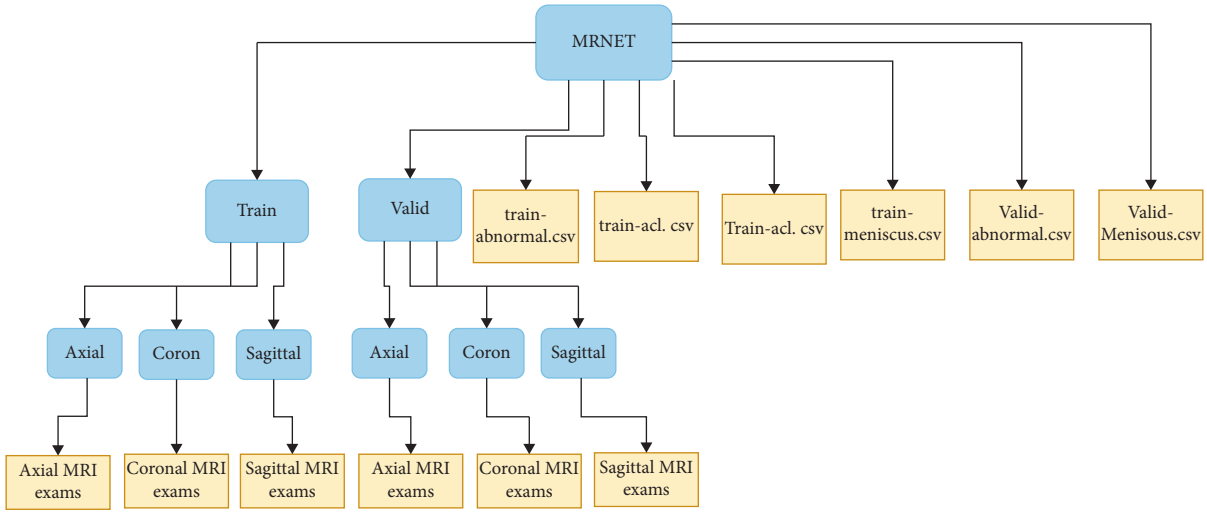


Figure 1: Organization of the MRNet dataset.

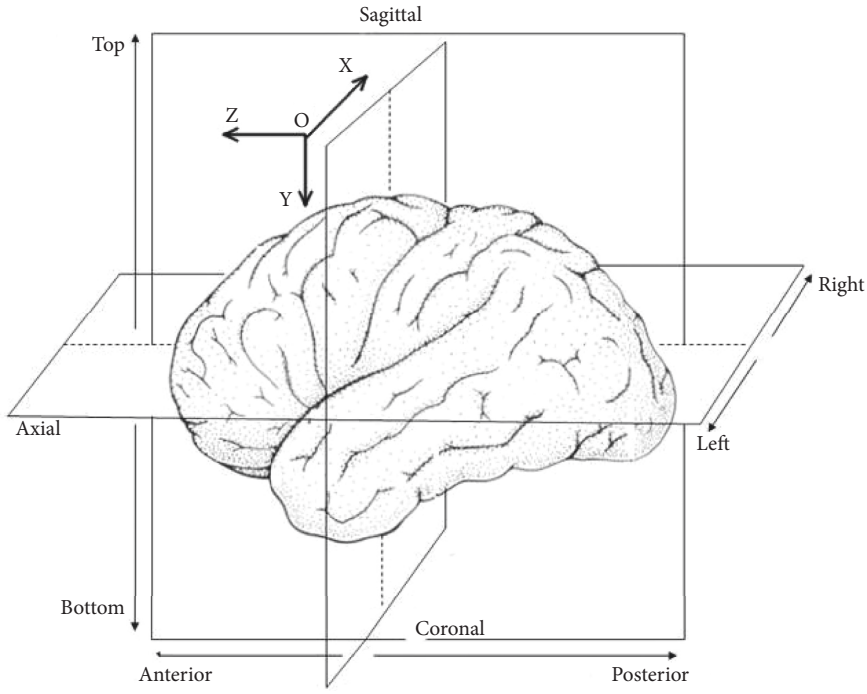


Figure 2: Depiction of various planes in our MRI's files [16].

We then passed these data into a batch normalization layer and a max pool layer [22]. The output of this max pool layer was passed through a flattened layer and a dense layer with 128 units and activation as ReLu. We then passed the output of this dense layer into another dense layer with 64 units and used a dropout layer with the dropout rate as 0.15. Finally, the output of this dense layer was passed to a final dense layer with units as 1 and activation as sigmoid to get output as 0 or 1, which indicated the prediction of our model [23]. We then compiled our model using Adam as the optimizer and binary cross entropy as loss. The architecture of our ResNet50-based CNN model is shown in Figure 3.

We also employed callbacks by using validation accuracy as a monitoring parameter for model checkpoint and model validation loss for early stopping. We used batch size 20 to train our model and trained our model over 50 epochs. We trained a total of 27 models, 9 using 18 slices for each MRI sample, 9 using 3 slices, and 9 using only 1 slice. To get 18 slices for each MRI sample, we used interpolation. For 3 slices, we got the middle slice and subtracted and added 2 to it to get the other 2 slices. For one slice, we selected the middle slice from each MRI sample. The base models were exactly the same for all these slices [24]. Since the dataset was largely unbalanced, there was some

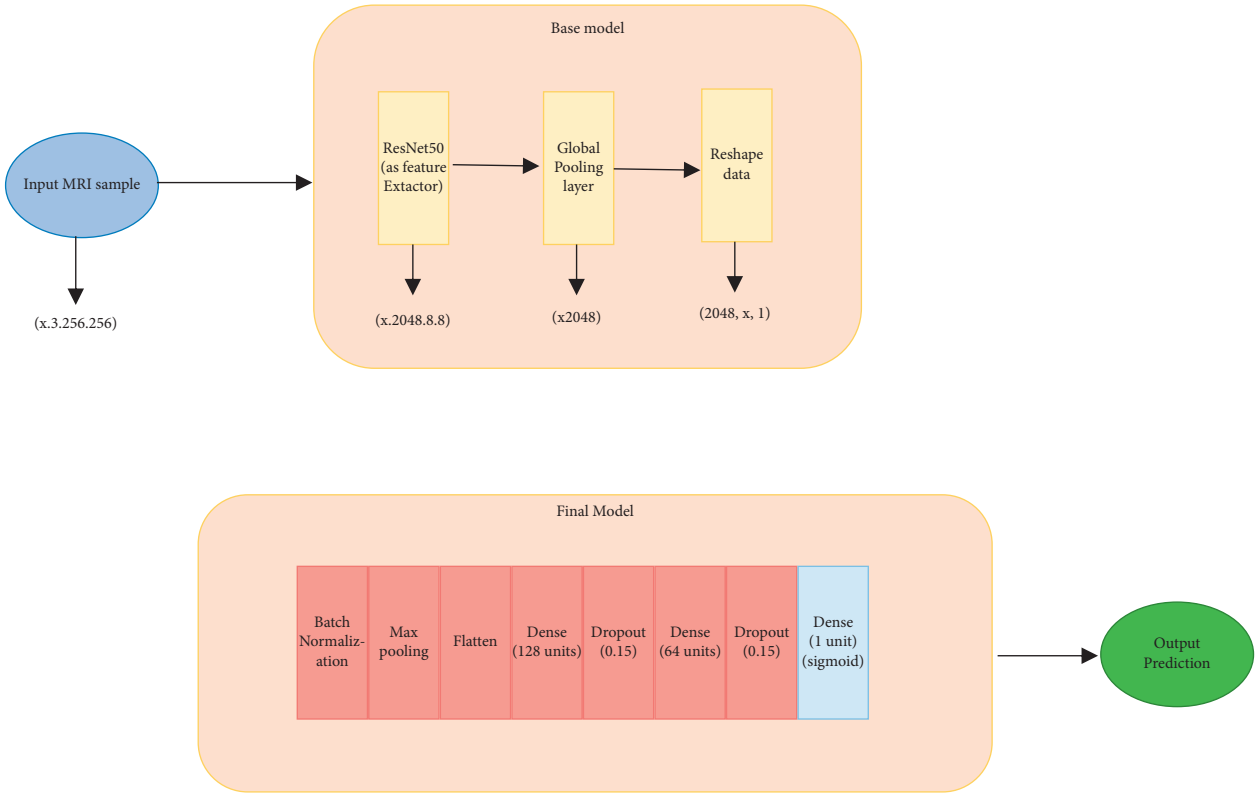


Figure 3: Architecture of our ResNet50-based model.

difference between the training and validation accuracies and losses between epochs [25].

4. Results and Discussion

The results of the best models for each setting were quite comparable to each other. We trained the models using 18, 3, and 1 slice. The results of these models were quite surprising. The AUC results using 1 slice per MRI exam were quite comparable to those using 18 and 3 slices and, in some cases, were even better. We also observed that by increasing the number of slices beyond 3, the results were not improving that much and, in some cases, were getting even worse.

After getting the AUC scores, we decided to get the accuracy scores of the models too and a similar trend was found to be present here as well. The best validation accuracy scores for abnormal and meniscus tears were obtained using 3 slices per exam, and for ACL tears, 18 slices per exam gave the best results. The accuracy scores of these models are mentioned in Table 1. The ROC-AUC graphs of the best models for every knee tear in different settings are shown in the following.

The best MRI plane in terms of model performance turned out to be the axial plane. One more thing to note is that the results of the models trained using sagittal planes were obtained close to the models trained using axial plane MRIs. For meniscus and abnormal MRIs, there was not a large difference between models trained using 1 slice and models trained using 18 slices as can be seen in Table 1. The accuracy scores in the table are quite similar to the AUC scores in a way that for meniscus and abnormal samples, models trained using 3 slices perform the best and for ACL samples, models trained with 18 slices perform better. There is also a trend that can be seen from the table and the AUC graphs that meniscus tears are best detected using sagittal plane MRI while ACL and abnormal samples are best detected using axial plane MRIs.

4.1. For Abnormal MRIs. For abnormal MRIs, the best results were achieved when the models were trained using axial plane MRIs. The AUC scores in all three settings were comparable, but the model with 3 slices with an AUC score of 0.90 and an accuracy score of 91.66% performed marginally better than the models with 18 and 1 slices, and it is shown in Figures 4–6.

4.2. For ACL Tears. In the case of ACL tears, the best results were also achieved when the models were trained using axial plane MRIs. Here, the results of all three settings were quite close, with the models trained using 18 slices performing marginally better than the models trained using 3 slices and 1 slices. The model trained using 18 slices achieved an AUC score of 0.92 compared to 0.87 and 0.89 of the models trained using 3 slices and 1 slice, respectively, and it is shown in Figures 7–9.

4.3. For Meniscus Tears. For meniscus tears, the best results were achieved when the models were trained using sagittal

Table 1: Accuracy scores of the best models in each setting.

Number of slices considered	ACL (percent)	Meniscus (percent)	Abnormal (percent)
18	**86.66 (axial)**	72.50 (axial)	90.00 (axial)
3	83.33 (axial)	**79.16 (sagittal)**	**91.66 (axial)**
1	82.50 (axial)	71.6 (sagittal)	90.00 (axial)

The bold values represent highest values.

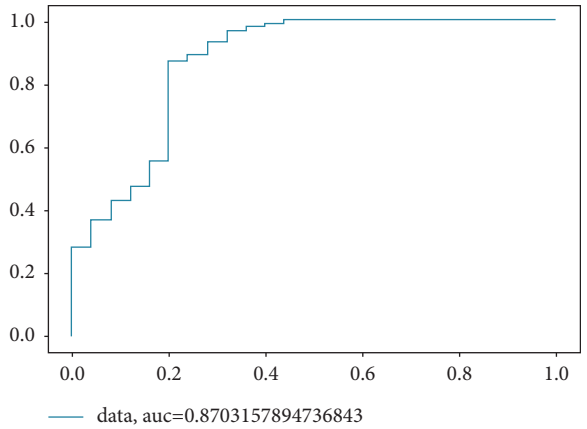


Figure 4: Using 18 slices of the axial plane.

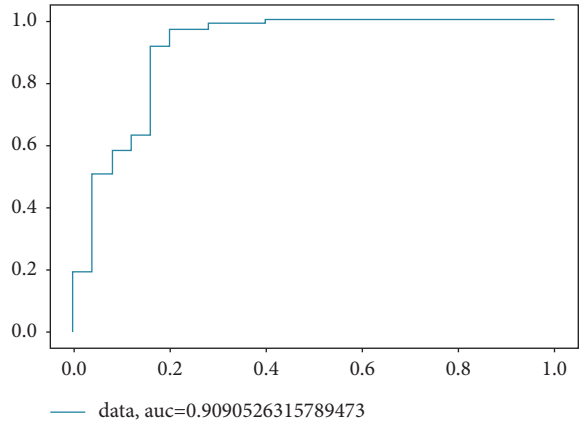


Figure 5: Using 3 slices of the axial plane.

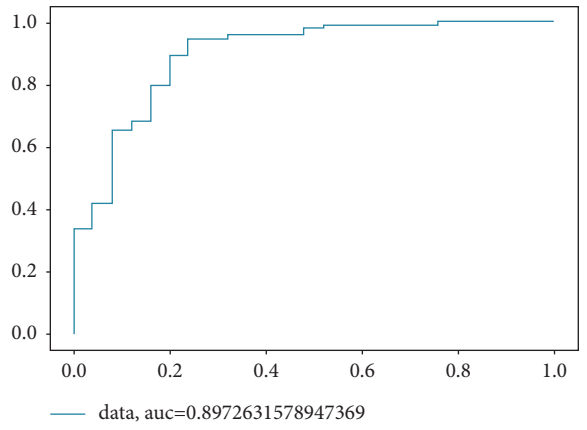


Figure 6: Using 1 slice of the axial plane.

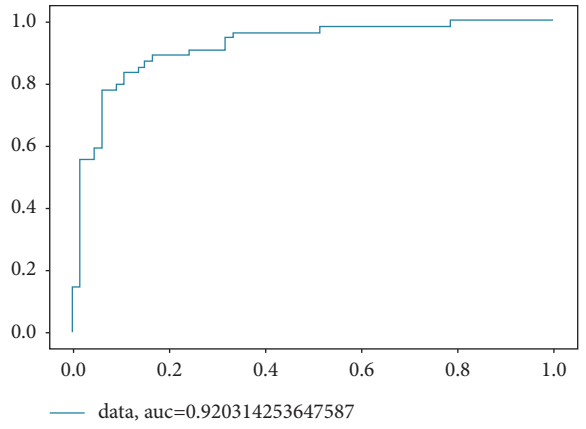


Figure 7: Using 18 slices of the axial plane.

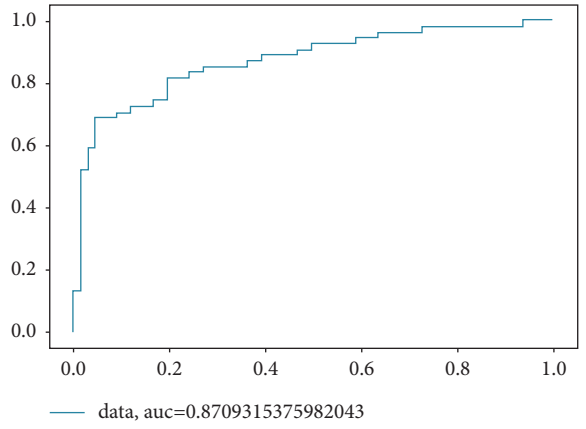


Figure 8: Using 3 slices of the axial plane.

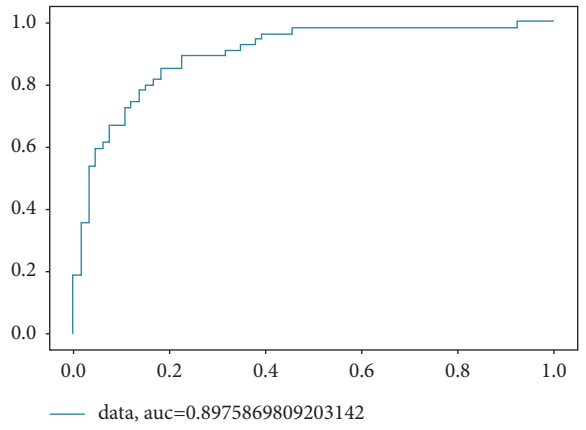


Figure 9: Using 1 slice of the axial plane.

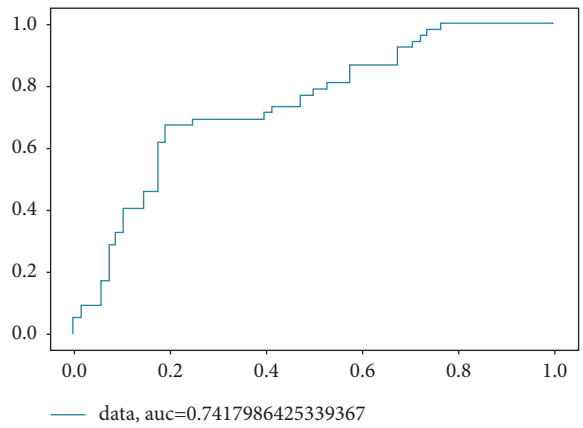


Figure 10: Using 18 slices of the sagittal plane.

plane MRIs. Meniscus tears were the only type of knee tears that we tested in our study that were best detected using sagittal plane MRIs. Although the models trained using sagittal plane MRIs were performing the best, the models trained using axial slices were very close or even as good in some cases. Here, the models trained using 3 slices outperformed the models trained using 18 and 1 slices by achieving an AUC score of 0.82 compared to 0.74 and 0.76 of 18- and 1-slice models respectively, and it is shown in Figures 10–12.

5. Comparison

A considerable amount of work is already done in the field of MRI analysis using deep learning, and even for the dataset that we considered in our study, some of the good work is already published. However, the others focused on improving the results by employing various techniques like preprocessing and hyperparameter tuning. Even if certain cutting-edge models outperform our model in some circumstances, our model is still far faster and more efficient to train. In industry, the ability and flexibility to efficiently train models are sometimes more important than small performance measures. However, ResNet has shown to be effective in a wide range of areas; one major downside is that larger networks typically need months of training, and finding them is almost impossible in practical use. We employed a rather simple model with ResNet50 as a base, and we focused on getting similar results to the state-of-the-art models by using as fewer slices of each MRI as possible. For 2 out of 3 types of tears (meniscus and abnormal samples), using only three slices for model training proved to be better than using 18 slices. Even though for ACL tears, the 18-slice models performed better and the difference in AUC and accuracy scores was not that large.

One study based on the MRNet dataset discovered that the ROC-AUC value for detecting a meniscus tear was 0.826, and it was 0.956 for detecting an ACL tear and 0.936 for detecting abnormalities [11]. To create a shared disturbance detection model in the coronal, sagittal, and axial planes, the researchers used a logistic regression-based ensemble learning strategy. Even though their model performed slightly better for abnormal and ACL tears, our model is way simpler than theirs and only uses 3 slices per MRI.

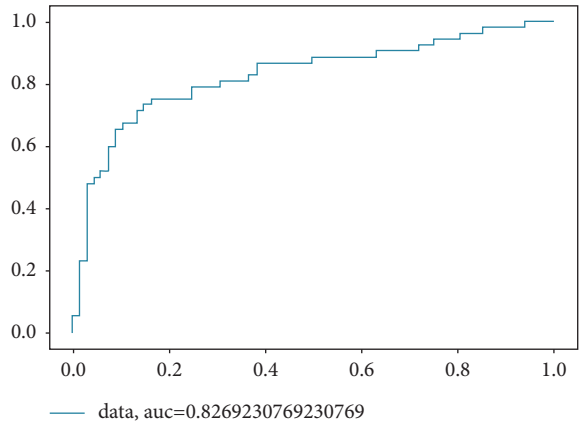


Figure 11: Using 3 slices of the sagittal plane.

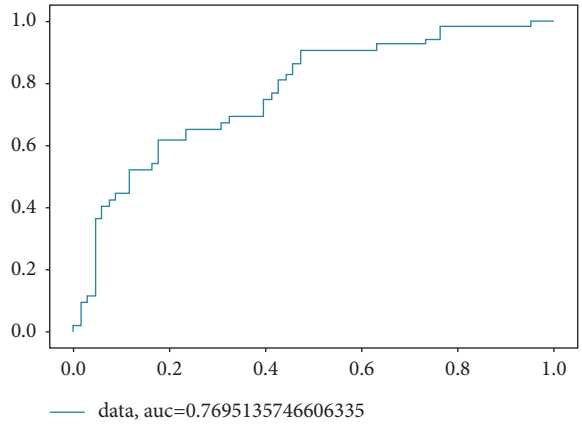


Figure 12: Using 1 slice of the sagittal plane.

In another study, images were first selected based on noise and other parameters that might aid in the detection of diseases. Then, a region of interest was located using CNN and denoising autoencoder, and finally for the diagnosis, a ResNet50 model was used [20]. The best accuracy scores this

Table 2: Comparison of all models employed in this study.

Number of slices	Tear type	Sagittal plane		Coronal plane		Axial plane	
		Accuracy	ROC-AUC	Accuracy	ROC-AUC	Accuracy	ROC-AUC
18	ACL	0.81	0.86	0.76	0.83	0.86	0.92
	Abnormal	0.89	0.75	0.87	0.78	0.90	0.87
	Meniscus	0.70	0.74	0.67	0.73	0.72	0.74
3	ACL	0.77	0.82	0.71	0.77	0.83	0.87
	Abnormal	0.90	0.89	0.86	0.78	0.91	0.90
	Meniscus	0.79	0.82	0.72	0.72	0.70	0.73
1	ACL	0.79	0.83	0.71	0.79	0.82	0.89
	Abnormal	0.89	0.83	0.83	0.76	0.90	0.89
	Meniscus	0.71	0.76	0.70	0.72	0.68	0.73

model was able to achieve for ACL, abnormal, and meniscus tears are 83.19%, 89.92%, and 77.12%, respectively. Our model performed better than this model for all three cases. Table 2 shows the comparison for all our models trained using different settings.

6. Conclusions and Future Work

We explored how to set up a deep learning model to enhance the detection efficiency on a dataset of knee MRI images in this study. We used deep learning on the MRNet dataset and trained a range of ResNet50-based deep learning models to suggest the likelihood of a knee injury on a given knee MRI. We used scans from three planes: axial, coronal, and sagittal, since doctors routinely view MRIs from several perspectives.

Our best performing models were trained using 3 slices for each MRI sample for meniscus and abnormal MRIs. For ACL tears, the model trained using 18 slices provided the best results. The difference however was not large. To get 18 slices for each MRI sample, we used interpolation. For 3 slices, we got the middle slice and subtracted and added 2 to it to get the other 2 slices. For one slice, we selected the middle slice from each MRI sample. The three-slice model performed equally well even when compared to some state-of-the-art models. Even though some state-of-the-art models performed marginally better than our model, in some cases, our model is way faster and more efficient to train. In industry, the capacity and flexibility to quickly train models are sometimes more essential than minor changes in performance. It is also worth noting that the models trained using axial plane MRIs consistently performed better than the models trained using the other two planes, especially for the detection of ACL and abnormal tears. For meniscus tears, models trained using sagittal planes were slightly better than models trained using axial plane MRIs.

Finally, coming to the future scope, ensemble approaches would probably improve the overall performance. Moreover, a multiclassification model can also be developed in the future, which will classify an MRI sample into ACL, abnormal, and meniscus tears. When using MRI images, to diagnose cancer, the location of the tumor site is crucial. Studies have shown that focusing on a specific part of the MRI for detecting tumors can yield better results than focusing on the whole image. The same concept can be employed here too.

Data Availability

The data will be made available on request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] N. Subhas, S. H. Patel, . N. . A. Obuchowski, and M. H. Jones, *Value of Knee MRI in the Diagnosis and Management of Knee Disorders*, Orthopedics, 2014.

[2] A. Kim, L. Khoury, M. Schweitzer et al., "Effect of Specialty and Experience on the Interpretation of Knee MRI Scans," *Bulletin of NYU hospital for joint diseases*, vol. 66, no. 4, 2008.

[3] Y. Kocabey, O. Tetik, W. Isbell, O. Atay, and D. Johnson, "The value of clinical examination versus magnetic resonance imaging in the diagnosis of meniscal tears and anterior cruciate ligament rupture," *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, vol. 20, no. 7, pp. 696–700, 2004.

[4] S. H. Hong, J.-Y. Choi, G. K. Lee, J.-A. Choi, H. W. Chung, and H. S. Kang, "Grading of anterior cruciate ligament injury," *Journal of Computer Assisted Tomography*, vol. 27, no. 5, pp. 814–819, 2003.

[5] G. Litjens, T. Kooi, B. E. Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[6] M. Mohammed, M. Ghani, N. Arunkumar, S. Mostafa, M. Abdullah, and M. Burhanuddin, "Trainable model for segmenting and identifying Nasopharyngeal carcinoma," *Computers & Electrical Engineering*, vol. 71, pp. 625–638, 2018.

[7] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of heart disease using a combination of machine learning and deep learning," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–11, Article ID 8387680, 2021.

[8] M. K. Abd Ghani, M. A. Mohammed, N. Arunkumar et al., "Decision-level fusion scheme for nasopharyngeal carcinoma identification using machine learning techniques," *Neural Computing & Applications*, vol. 32, no. 3, pp. 625–638, 2020.

[9] O. Obaid, M. Mohammed, M. Ghani, A. Mostafa, and F. Taha, "Evaluating the performance of machine learning techniques in the classification of Wisconsin Breast Cancer," *International Journal of Engineering & Technology*, vol. 7, pp. 160–166, 2018.

[10] F. Ajaz, M. Naseem, S. Sharma, M. Shabaz, and G. Dhiman, "COVID-19: challenges and its technological solutions using IoT," *Current Medical Imaging Formerly Current Medical Imaging Reviews*, vol. 18, no. 2, pp. 113–123, 2022.

[11] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 102–127, 2019.

[12] J. Liu, Y. Pan, M. Li, Z. Chen, L. Tang, and C. Lu, "Applications of deep learning to MRI images: a survey," *Big Data Mining and Analytics*, vol. 1, no. 1, 2019.

[13] C. Sharma, A. Bagga, R. Sobti, M. Shabaz, and R. Amin, "A robust image encrypted watermarking technique for neurodegenerative disorder diagnosis and its applications," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–14, Article ID 8081276, 2021.

[14] N. Bien, P. Rajpurkar, R. L. Ball et al., "Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet," *PLoS Medicine*, vol. 15, no. 11, 2018.

[15] C. H. Tsai, N. Kiryati, E. Konen, I. Eshed, and A. Mayer, "Knee injury detection using MRI with efficiently-layered network (ELNet)," *Third Conference on Medical Imaging with Deep Learning*, vol. 1, 2020.

[16] D. Azcona, K. McGuinness, and A. F. Smeaton, "A comparative study of existing and new deep learning methods for detecting knee injuries using the MRNet dataset," in *Proceedings of the 2020 International Conference on Intelligent Data Science Technologies and Applications (IDSTA)*, Valencia, Spain, 19-22 Oct. 2020.

[17] S. Chaudhury, N. Shelke, K. Sau, B. Prasanalakshmi, and M. Shabaz, "A novel approach to classifying breast cancer histopathology biopsy images using bilateral knowledge distillation and label smoothing regularization," *Computational and Mathematical Methods in Medicine*, vol. 2021, pp. 1–11, Article ID 4019358, 2021.

[18] R. Shrivastav, "Deep convolutional neural network-based knee injury classification using magnetic resonance imaging," in *Proceedings of the International Conference on Communication*, Bhubaneswar, India, October 16-18, 2020.

[19] Stanford university, *MRNet: A Dataset of Knee MRIs and Competition for Automated Knee MRI Interpretation*Stanford university, Serra Mall, Stanford, 2018, https://stanfordmlgroup.github.io/competitions/mrnet/.

[20] Brainvisa, "Extract connected components in order," 2022, https://brainvisa.info/axon/en/processes/ExtractComponentsInOrder.html.

[21] A. Gupta and L. K. Awasthi, "Peers-for-peers (P4P): an efficient and reliable fault-tolerance strategy for cycle-stealing P2P applications," *International Journal of Communication Networks and Distributed Systems*, vol. 6, no. 2, p. 202, 2007.

[22] D. S. Ushakov, O. O. Yushkevych, O. O. Yushkevych, N. L. Ovander, H. Y. Tkachuk, and V. H. Vyhovskyi, "The strategy of Thai medical services promotion at foreign markets and development of medical tourism," *GeoJournal of Tourism and Geosites*, vol. 27, no. 4, pp. 1429–1438, 2019.

[23] A. Gupta and L. K. Awasthi, "Peer enterprises: possibilities, challenges and some ideas towards their realization," in *Proceedings of the On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops*, pp. 1011–1020, Springer Berlin Heidelberg, Vilamoura, Portugal, November 25-30, 2007.

[24] S. Ioffe and C. Szegedy, "Batch normalization: accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, Lile, France, July 2015.

[25] F. H. Ali, "Can kara, "detection and classification of knee injuries from MR images using the MRNet dataset with progressively operating deep learning methods," *Machine Learning and Knowledge Extraction*, vol. 3, no. 4, pp. 1009–1029, 2021.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 4948190, 11 pages
https://doi.org/10.1155/2022/4948190

Research Article

Vampire Attack Mitigation and Network Performance Improvement Using Probabilistic Fuzzy Chain Set with Authentication Routing Protocol and Hybrid Clustering-Based Optimization in Wireless Sensor Network

Lulwah M. Alkwai,[1] Arwa Naser Mohammed Aledaily,[1] Shahad Almansour,[1] Shoayee Dlaim Alotaibi,[1] Kusum Yadav,[1] and Velmurugan Lingamuthu[2]

[1]*College of Computer Science and Engineering, University of Ha'il, Hail, Saudi Arabia*
[2]*Ambo University, Ambo, Ethiopia*

Correspondence should be addressed to Velmurugan Lingamuthu; velmurugan.lingamuthu@ambou.edu.et

Received 21 April 2022; Revised 17 May 2022; Accepted 25 May 2022; Published 14 June 2022

Academic Editor: Mukesh Soni

The most effective threat for wireless sensor networks (WSN) is Vampire attacks on sensor nodes as they can stretch the network connectivity among them and influence the network's energy, which can drain the network. Vampire attack has particular malicious nature of sensor nodes in which they can widely exploit features of combined routing protocol. Fuzzy rules and fuzzy sets are highly optimal techniques in mitigating the vampire attacks of the network, which can quantify the uncertain behaviour of sensor nodes. This study aims to propose a novel technique using a probabilistic fuzzy chain set with authentication-based routing protocol and hybrid clustering technique for data optimization of the network. The suggested approach here employs a fuzzy-based chain rule set to combat growing types of vampire assaults using probability formulas. The authentication routing protocol has increased network routing security. The proposed technique (PFCS-ARP_HC) has optimized the energy consumption of network. Simulation for this technique has been carried out using NS2 and experimental results show the performance of the proposed model in terms of throughput of 98%, packet delivery ratio of 89%, energy consumption of 67%, latency of 46% control overhead of 53%, and attack detection ratio of 87.9%.

1. Introduction

Wireless network is a physical infrastructure that allows computers, mobile phones, printers, and other devices to communicate with each other via a router. Ad hoc sensor networks are being organized in numerous areas such as the military, health care, and defence as a result of technical advancements [1]. These sensors measure environmental elements such as humidity, temperature, traffic surveillance, movement, noise, military, and management of agricultural land. These sensors are vulnerable to DoS assaults such as the vampire, directional, black hole, and selective forwarding attacks because they are battery-powered. Because vampire attacks are not protocol-specific, they're difficult to spot. Security is a significant study topic in WSNs (wireless sensor networks) [2]. Because routing is a trust-based operation among nodes, attackers have a good possibility of interfering with it. As a result, actions to safeguard WSN from security threats must be deployed. Security studies of these networks are undertaken separately because they are typically built without prior planning and are only used for a short period of time. DoS attacks are one of the most well-known types of network sensor attacks. DoS prevents radio from going into sleep mode, which would drain the battery completely [3].

Because of their distributed structure and positioning in remote places, these networks are exposed to a variety of

security vulnerabilities that can jeopardise their correct operation. WSNs with resource-constrained nodes are extremely vulnerable to a number of attacks due to their simplicity. Sensor devices' high resource constraints present significant hurdles to resource-hungry security systems. Because of the hardware limits, security methods must be exceedingly efficient. This is no trivial task [4]. Sensor networks' most valuable resource is energy. In terms of power, communication is very costly [5, 6]. The available communication channels for coordinating their attack. As a result, WSN is vulnerable to a number of attacks that could obstruct its operations and negate the benefits of using its services. If a sensor node is compromised, all vital material, data and code stored on that node can be retrieved. An attacker can quickly recover valuable information from packets that are sent. Finally, attacker can send bogus data into network [7, 8], possibly impersonating one of sensors, with goal of altering sensor readings or interrupting internal control data (Message Injection). Attackers may collaborate on a system attack by using a few MNs with similar or greater hardware capacity than genuine nodes [9]. These rogue nodes can be obtained either individually or by capturing and physically overwriting the memory of a few legitimate nodes. In rare cases, participating nodes may have high-quality communications links available to coordinate their attack. As a result, WSN is susceptible to a variety of assaults that might disrupt its operations and nullify the benefits of using its services [10–12].

Vampire attacks are tough to identify and avoid because they use protocol-compliant messages. Vampire attacks are not protocol-specific in the sense that they do not rely on individual routing protocols' design or implementation flaws. Effects of a vampire attack: (a) Vampire attacks are protocol-independent. (b) They do not cause any interruptions in instant availability. (c) Protocol-compliant communications are used by vampires. (d) Send a small amount of data while consuming the most energy. (g) Vampires do not alter or disturb established courses. Vampire assaults can be divided into two categories. There are two types of attacks: Carousel and Stretch [13–15].

The contribution of this study is as follows.

(i) To develop the novel technique in mitigating vampire attacks and improve the network performance using security-based routing protocol and clustering-based data optimization.

(ii) Here the data optimization has been carried out using hybrid clustering technique and vampire attack is mitigated using probabilistic fuzzy chain set with authentication-based routing protocol.

(iii) Simulation for this technique has been carried out using NS2 and experimental results shows the performance of the suggested model in terms of throughput, packet delivery ratio, energy consumption, control overhead, and attack detection ratio.

2. Related Works

Numerous works have been carried out in this area and enormous data are available on DoS attacks and their effects on ad hoc networks. At the routing protocol layer, the author [16] investigates resource depletion assaults, often known as "Vampire Attacks," which permanently disable networks by rapidly draining node battery power [17, 18]. They've included information on numerous vampire assaults and how they affect ad hoc networks. They've also listed a number of ways to protect the network from these threats. Our previous work [19] provides a quick overview of these two procedures as well as the consequences of vampire attacks on them. Some recent study in this topic includes numerous implementations and theoretical work, such as [20], which shows how to tolerate attack by using CH. In the event of a vampire assault, the Cluster Head intervenes and distributes the packet to its target without dropping it. As a result, even in the event of a vampire assault, the message is delivered successfully and reliably. The authors of [21] suggested how to use PLGP, an identifying malicious assault to lessen the vampire attack. The next study [22] looks into the energy issues and assaults that Ad Hoc WSNs and, eventually, future IoT face. It also provides a strategic method to dealing with energy threats that is resilient. The proposed approach in [23] aims to give a method that may be utilized to find vampire attack in WSN. A new trust routing motif is proposed in [24]. Multiagents gather multifactor data and come together to decide on the trust path. The degree of belief in the long-term conduct of alternative entities that is based on the nodes' previous experience is referred to as trust. CAWS and MES-1, a collection of algorithms put together by a researcher [25], were examined. CAWS (cellular automata-based security algorithms) entails key management and secure digital communication under cellular automata rules, requiring minimal memory and simple calculation. The author of [26] first aims to assess these vulnerabilities in terms of router layer battery reduction attacks. Second, it focuses on making changes to current routing protocols in order to prevent packet loss due to vampire attacks during packet forwarding. To protect wireless sensor networks from wormhole attacks, they propose a trust aware distance vector routing protocol (TAODV) in [27]. Their proposed approach was tested utilising experimental results in terms of enhanced PDR, end-to-end delay and node to destination variation. Ariadne, an on-demand routing system, and the LEACH protocol, intended to reduce battery use, are among the approaches and protocols used to resist these assaults. The LEACH methodology involves two phases: steady state and startup. The cluster is established in the steady-state phase based on threshold value, with each node calculating its value by picking a random number between 0 and 1 and broadcasting it to its neighbours. A node with a value less than threshold limit is nominated as CH and rest of nodes join as member nodes. Although the LEACH protocol helps to save energy, it does not ensure secure routing against vampire attacks. There are two phases to EWMA (Energy Weighted Monitoring Algorithm) [28]. The cryptographic keys can guard against active attacks from the outside, but they are unreliable in the case of passive attacks that compromise service quality as well as reliability. Optimization techniques based on the behaviour of lions during territorial defence and

takeover [29, 30], such as the Lion algorithm, were devised to minimise passive attacks and to optimise path selection. The nonlinear system identification solution is implemented. However, solutions to difficult situations are not taken into account. As a consequence, the authors in [31] presented the WOA, which is also an optimization system based on the behaviour of humpback whales. Because the location of the prey is unknown at beginning, the current solution, or prey, is treated as the best response and position is updated in each iteration, despite the fact that network nodes use more energy for calculation.

3. System Model

This section discusses the proposed design in mitigating vampire attacks of the sensor networks. Here the proposed module has been divided as three parts: first is hybrid clustering in which the nodes has been clustered based on energy level of energy; secondly analysing for vampire attacks by authenticating data transmission path through routing protocol and probabilistic fuzzy chain set with fuzzy rule sets. The overall proposed flow diagram is shown in Figure 1.

3.1. Hybrid Clustering of Nodes Based on the Energy Level. Consider n nodes in a field. Lifetime of node I is denoted by Li. Let the network lifetime, L, be time it takes for first node in network to exhaust its energy. The main goal is to increase L, which necessitates uniformly utilising the energy of all nodes. To avoid drawbacks of both static and dynamic clustering methods, designed hybrid method so that clustering is not done every round. To do so, CHs save their residual energy in their memory at the end of each setup step after clusters have formed. When a CH detects that its residual has fallen below ECH, it sets a specific bit in a data packet ready to be forwarded to BS.

3.1.1. Cluster Setup Phase. Nodes that will become CH will be chosen during the clustering setup step. CH is chosen based on delay duration of processing. The primary CH will be chosen when the delay time approaches zero. If a node's delay time is assessed to be lower, it has a better chance of becoming a CH and sending a broadcast packet. Packet size, node ID, node location, residual energy, and packet type are all included in the content of the packet. If a node gets packet type 2 from other nodes before delay timer expires, it will become a normal node "N." Additionally, the candidate CH will send messages to sensor nodes through broadcast. Messages from the CH will be verified by the sensor nodes [32–34]. If the CH has a limited amount of residual energy and is likely to expire soon, the nodes in the network will advise their neighbours to modify path and deliver data to another CH. Several nodes having the same number of neighbours play a vital role in ensuring that each node has a diverse number of neighbours. Each node delay time is estimated using

$$D_t(i) = \left(1 - \frac{E_{rem(i)}}{E_{avg}(i)}\right) * W_t + R_v, \quad (1)$$

where $D_t(i)$ is node i's delay time, $E_{\mathrm{rem}(i)}$ is node i's residual energy, W_t is primary CH's competition time, and R_v is the random value. When nodes have same residual energy, random value can help to reduce communication conflicts.

Algorithm for cluster setup phase (Algorithm 1):

3.1.2. Cluster Formation (CF). The CF function chooses best non-CH candidate to join CH. After selecting principal CH, it will broadcast packet type 3 to node Si while waiting for the packet to arrive. If node Si receives a message from node Sj, add it to CH_list of potential nodes and alter it to a non-CH state. In this situation, it will calculate distance between non-CHs and CH to select best nodes as its CH. The best CH selection is determined by a small number of neighbour nodes, a short distance, and a high residual energy. Furthermore, more number of nodes in every cluster has an impact on the network's performance.

Algorithm for CF (cluster formation) (Algorithm 2):

3.1.3. Transmission Phase. This phase transfers data over the network between CHs and CMs. A CH rotation technique and stacking implementation make up the transmission phase. In the following part, we'll go over the CH rotation approach and the layered implementation design in more detail. After selecting the principal CH, the transmission procedure is used to create a TDMA schedule that transmits a broadcast schedule to all CMs. Data will be sent from the CMs to primary CH, which will aggregate it. Aggregation operations are forwarded to BS after that.

3.1.4. Analysing the Malicious Activities of the Network. The network's nodes are responsible for not only delivering precise data, but also for updating new data in received data packet and sending it to next node. As a result, it is critical to confirm that data exchanged between nodes is secure. When a node is subjected to an aggressive or passive attack, it loses its trustworthiness. A carousal attack occurs when a rogue node causes a data packet to loop continuously, preventing it from reaching the BS or the destination node. The energy of the node depletes dramatically in a short amount of time as a result of this anomalous behaviour, resulting in a vampire attack. If BS does not receive requested data packet within specified time interval, BS requests that CH evaluate trust value of all its member nodes given by (2)

$$T^d(t) = \frac{P_{n1}(t)}{P_{n2(t)}}. \quad (2)$$

$T^d(t)$ is computed direct trust between $n1$ and $n2$ nodes. $P_{n1}(t)$ denotes packets that have been received. The total number of packets sent is $P_{n2}(t)$. The surrounding nodes' estimated trust are represented in (3)

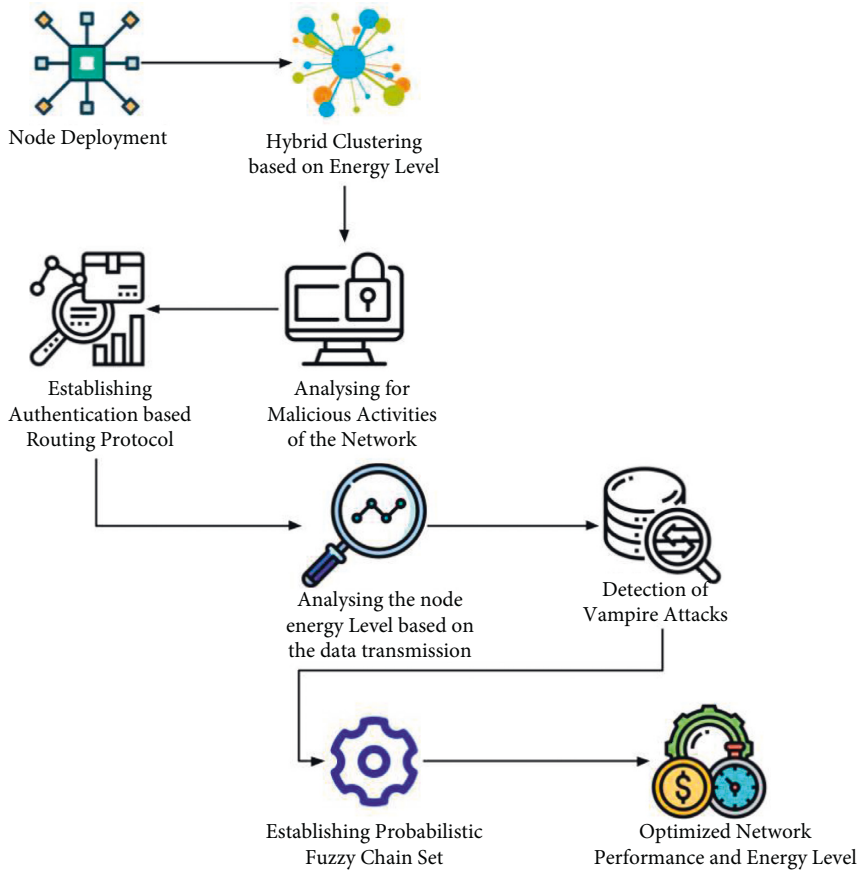


Figure 1: Overall proposed flow diagram.

$$T^t(t) = \frac{1}{k}\sum_{d=1}^{k} T^d(t), \tag{3}$$

$$T = \alpha T^d(t) + \beta T^I(t), \tag{4}$$

α, β values range from 0 to 1 such that $\alpha + \beta = 1$. Trust threshold is 0.99–1. If estimated value falls below this threshold, sensor is classified as an MN, as it experiences higher packet loss during transmission. For effective transmission of data to send packet along chosen path, the trust criteria are taken into account. If BS identifies an MN based on the trust degree calculation after the verification procedure, it sends an alarm message to CH with ID and location of detected MN. CHs have now blacklisted and isolated rogue node, as well as broadcasting its ID to all of its members.

3.1.5. Authentication-Based Routing Protocol. The sink is ready to receive a data packet from source/intermediate node k. It uses the selective authentication algorithm to determine whether or not to inspect the data packet. As a result, timer on the higher priority node is shorter. It would put up its own forwarder network and get ready to send data. The process is repeated by subsequent sensor nodes until data packet reaches sink.

Algorithm for Authentication-based routing protocol (Algorithm 3):

3.1.6. Probabilistic Fuzzy Chain Set. Three factors are chosen in the suggested work while considering the concept of vampire assault. Three trust factors, such as node's packet drop rate, percentage of battery drain, and the number of link requests initiated by the node, can be used to detect the vampire node. These are described as follows.

(1) Packet Drop Rate (PDR): Malicious nodes lose a substantial percentage of packets, whereas trustworthy nodes forward all of the packets they receive. The ratio of number of packets dropped to total number of packets is used to calculate this value.

(2) Battery Discharge Rate (BDR): This node parameter is dependent on the network's operation. An active node appears to be in use and uses more energy. Nodes with a malignant or selfish aim, on the other hand, are observed to be more active than other nodes.

(3) Number of link requests (NLR): The vampire node tries to connect to numerous nodes at the same time. As a result, the node will generate an unusually large number of link requests. The ith node initial position is given by (5)

(1) Input $(S_i, W_t, A, B, Th_v, RL_{\max}, a, b, R_v)$
(2) Output
(3) For each node S_i do
(4) Evaluate $R_c(i)$
(5) Nodes known their neighbours
(6) Broadcast packet type_1
(7) Evaluate number of neighbour nodes $NN(i, r)$
(8) Every node evaluates delay time $D_t(i)$
(9) Evaluate average energy of neighbour nodes $E_{avg}(i)$
(10) S S_i. Type = "N".
(11) If $S_i.D_t(i)$ = close to 0
(12) CountCH = countCH + 1
(13) S_i. Type = "CH."
(14) End
(15) While $S_i.D_t(i) \neq$ close to
(16) Broadcast packet type_2
(17) If $S_i.D_t > S_j.D_t$
(18) End
(19) S_i. Type = "N."
(20) While $S_i.W_t \neq 0$
(21) If $S_i.D_t < S_j.D_t$
(22) If S_i. Type = "CH."
(23) If $S_i.E < S_j.E$
(24) S_i. Type = N.
(25) Else S_i. Type = "CH."
(26) End While

Algorithm 1: Cluster setup phase.

(1) Input (S_i, S_j)
(2) Output (CH_list)
(3) For each node S_i do
(4) If S_i. Head = S_j. Head
(5) If S_i. Type = "N" && S_i. $E > 0$&&S_i. Type =* "Awake."
(6) Evaluate less distance from non-CH to CHs
(7) If $S_i.E >$ av_energy && S_i. Neighbour <S_j. Neighbour
(8) S_i. Type = "CH."
(9) CH_list store S_i
(10) Broadcast packet type 3
(11) Else S_i. Type = "CH."
(12) End for

Algorithm 2: CF (cluster formation).

$$f(x_i),\ x_i = \left(x_i^1, \ldots, x_i^D\right). \tag{5}$$

The matrices' dimension is D. For example, the position of node 1 is given by $x_{iD} = x_{1D} = (x_{11}, y_{12}, z_{13}, w_{14})$, position of node 2 is given by $Sx_{2D} = (x_{21}, y_{22}, z_{23}, w_{24})$, and so on. Fitness value is given by (6)

$$f(x) = \frac{1}{4}[T + E + D + d(i, j)]. \tag{6}$$

After determining the fitness of all nodes, the fitness with the lowest value, fbest, is regarded the best, and the fitness with the highest value, fworst, is measured worst. Masses of all surrounding nodes are determined in molecular dynamics. The atoms with the greatest mass have best fitness value, while atoms with smallest mass have poorest.

$$m_{i(t)} = \frac{M_i(t)}{\sum_{j=1}^{N} M_j(t)}, \tag{7}$$

where $M_i(t)$ and $M_j(t)$ signify the mass of ith and jth atoms (node); fbest(t) indicates the least fitness value; and fworst(t) denotes the highest fitness value. PFCS is denoted as a collection of four tuples PFCS = $(F_H, F_\kappa, F_n, F_d)$ with various encouragement factors of vampire attack as FI = f1, fe, f3, . . ., fm. The following steps make it easier to mitigate a vampire attack with PFCS.

(1) For removing the discrepancy that exists between generalisation and quality of approximations.
(2) Evaluate probability-oriented fuzzy membership value "$F_n(\lambda_{F_{n(j)}}(f_i))$" from matrix F_d using the

```
Vector<nsaddr.t>ARP:multi_shortest_path(nsaddr_t_sourfe_node,int packet.energy);
Bool exchange;
If (high ≥ low)}
While(time ≥ TIME_MAX&& listening (ACK = false){
If (exchange){
Int high, low, path.amount
Exchange = false
Path_amount
Multi_shortest_path(nsaddr.t source_node, int packet_energy)
High = path_amount, low = 1
Count(int timer); exchange = true
Send(route[high--]source_node)
}
}
Else
{
Path_amount
Send(route[low++] source_node)
Exchange = true
}
}
Else{
Int packet_energy)
High = path_amount;low = 1
Multi_shortest_path(nsaddr_t souce_node
exchange = true
}
}
Return
}
```

Algorithm 3: Authentication-based routing protocol.

uncertainty factor' $F_n(\chi_{i,j}^g)$ and fuzzy associativity factor $\lambda F_{n(i)}(f_i)$ as

$$F_n\left(\lambda_{F_{n(J)}}(f_i)\right) = \sum_{k=1}^{S_{i,j}} \lambda F_{n(j)}(f_i) * F_n\left(\chi_{i,j}^g\right). \tag{8}$$

(3) Finding of equivalent classes based on union and intersection operations utilising a probabilistic fuzzy theory technique, as shown in (2) and (3).

$$\text{Union}\left(F_{n(i)} \in F_{n(i)}(j) = \max_j \left\{F_N\left(\left(\lambda_{F_{n(j)}} f_i\right)\right\}_{i=1}^m,\right.$$
$$\text{Intersect}\left(F_{n(i)} \in F_{n(i)}(j) = \min_j \left\{F_N\left(\left(\lambda_{F_{n(j)}} f_i\right)\right\}_{i=1}^m\right.\right. \tag{9}$$

(4) Mean, higher, and lower degrees of approximations (π_M, π_G, π_L) evaluated for quantifying the degree of conformation that is estimated:

$$\pi_M = \min(\pi_L, \pi_G),$$
$$\pi_L = \kappa\left(F_d, \text{Interect F}_{n(i)}\right), \tag{10}$$
$$\pi_G = \kappa\left(\text{UnionF}_{n(i)}, F_d\right).$$

Finally, the crisp factor of PFCS is calculated using (7) as the ratio of divergence between lower as well as greater approximations to increasing count of entities inspected in universe utilized for decision making.

$$\beta = 1 - \frac{\sum \pi_G - \lambda_L}{|F_u|}. \tag{11}$$

(5) Then evaluation of inclusion degree is enabled using

$$\beta(M_a, M_b) = \frac{|\phi(M_a, M_b)|}{|\text{Supp}(M_b)|}. \tag{12}$$

3.1.7. Fuzzification. For FCS, three variables are used as input. The maximum and minimum values of the fuzzifier crisp input variable for calculating the eligibility index. Euclidean distance between each SN and BS is called distance to BS. The overall power accessible with the SN at that moment is known as Remnant Energy. The count of neighbouring nodes in vicinity of node under consideration for CH candidacy is known as node density. The FIS receives these crisp values (discrete values).

3.1.8. Fuzzy Rule Base. After fuzzification, the membership values are given into the rule base for IF-THEN situations. A value is derived by applying the fuzzy AND and OR operators to inputs. The aggregation method unites all of the

```
Input:n_round, deployment area, p, sing_p, initial_e, c_range, n_node, t_node
Output:R_VALUE = [CHs,alive_n]
(1) Initialization;
(2) For r = 1 to n_round
(3) For I = 1 to n
(4) If(mod(r,3) = = 1)
(5) If (S(i).energy ≤ 0)
(6) Alive = alive-1;
(7) SAY_HI_MESSAGE (ID,CR,SE)
(8) If r = 1
(9) Apply PFCS method to choose initial tentative CHs
(10) End if
(11) S(i).channel = evalfis([S(i).energy S(i).cound_tch],z)
(12) If r > 1
(13) End if
(14) If S(i) = best (channel])
(15) S(i).t_node = TCH;
(16) If (S(i).E < S(j).E)
(17) CONFIRM_TCH_MESSAGE(ID)
(18) S(i).type = N;
(19) MEMBER_JOIN_CH_MESSAGE(ID)
(20) Nodes with S(i).t_node = TCH will be confirmed final CH
(21) Transmit data ti the CH
(22) If (mod(r,3) = = 2) and previous CHs are alive
(23) Else choose best chance CH
(24) If (mode(r,3) = = 0)
(25) If (S(i).energy ≤ 0)
(26) For I = 1 to n
(27) Alive = alive-1;
(28) End if
(29) S(i).chance2 = evalfis([S(i).energy S(i).mch],z;
(30) SAY_HI_MESSAGE(ID,CR,SE)
(31) If (S(i).E < S(j).E)
(32) S(i).type = N;
(33) S(i).t.node = TCH;
(34) If (S(i) = best(chance2)
(35) AUTHORIZE_TCH_MESSAGE(ID)
(36) MEMBER_JOIN_CH_MESSAGE(ID)
(37) Nodes with S(i).t_node = TCH will be confirmed final CH
(38) End if
(39) Return R_VALUE
```

Algorithm 4: PFCS.

output after applying the 27 rules and a high value is determined from the gathered fuzzy set. Probabilistic chain rule was used, which is the most often used due of its properties, to generate the eligibility index using Fuzzy Logic.

3.1.9. Defuzzification. Center of area ($Z*$) method is utilized for defuzzification by (13)

$$Z* = \frac{\int \mu_A(x)x\mathrm{d}x}{\int \mu_A(x)\mathrm{d}x}. \tag{13}$$

After calculating the eligibility index for all nodes, the threshold (TH) is derived using the equation (9).

$$TH = \frac{\mathrm{Node}(i) \cdot P \times \mathrm{mean}[EI]}{1 - \mathrm{Node}(i) \cdot P \times \mathrm{mod}(r, (1/\mathrm{Node}(i) \cdot P))}. \tag{14}$$

For indiscriminate CH selection, every node in network creates a randomised number. If that value is less than computed TH, node will be assigned to CH role. CH's role is critical to the network's energy efficiency and rotated after each round to balance load among organized SN. Popt clusters are created in this manner. If some nodes remain after CF, they will join the cluster that is closest to them following CH recognition. Once all of the SN have been bound to clusters, the topology configuration process is complete. This method of weightage calculation enables for accurate values of criteria weights with the highest possible membership grade

Table 1: Comparative analysis of the proposed and existing techniques in vampire attack mitigation.

Parameters	TAODV	LEACH	WOA	PFCS-ARP_HC
Throughput	93	95	97.8	98
PDR	87.8	88	88.5	89
Energy consumption	63	65	74.5	75.5
Attack detection ratio	81.5	82	87.8	88
Latency	58	56	53.5	46
Control overhead	50	52.8	53	62

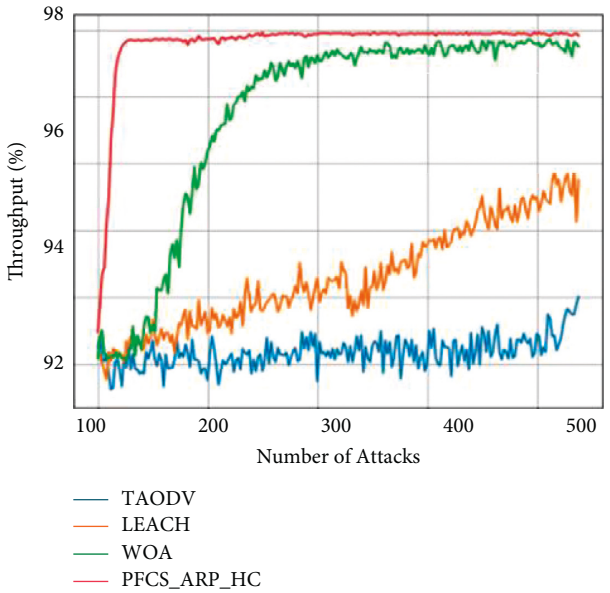


Figure 2: Comparative analysis of throughput.

for their membership function. Equation (10) is a linear programming model that enables the max–min weightage computation approach computationally.

$$z = \beta \longrightarrow \max,$$

$$\frac{w_i + \alpha_j^i - u_j}{\alpha_j^i} \geq \beta \text{ where } 1 \leq j \leq k,$$

$$\frac{\alpha_j^i + u_j - w_i}{\alpha_j^i} \geq \beta \text{ where } 1 \leq j \leq k, \tag{15}$$

$$\text{Such that } \sum_{j=1}^{k} w_i = 1,\ w_i \geq 0 \text{ with } \beta \in (0, 1).$$

The weightage set in question is made up of all strict and fuzzy sets with a separate high value intersected. The objective function is stated to be maximising in this scenario based on gain corresponding to result membership maximal grade. Furthermore, w_i is the weight value of the crisp parameter linked with the jth condition.

Algorithm for PFCS (Algorithm 4):

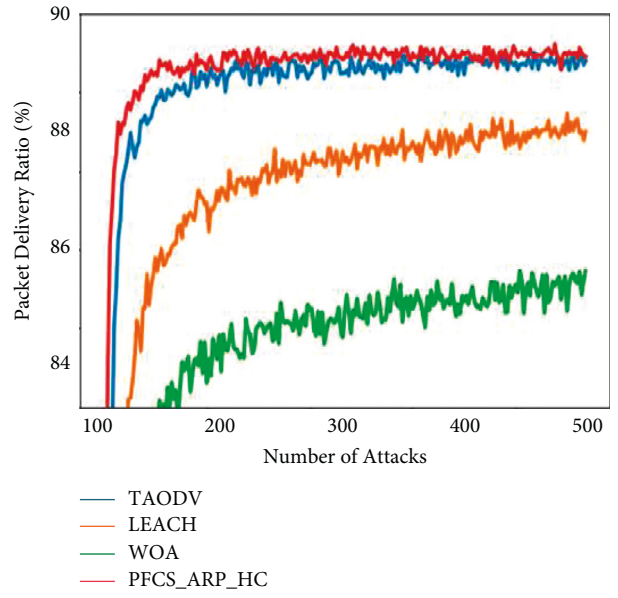


Figure 3: Comparative analysis of PDR.

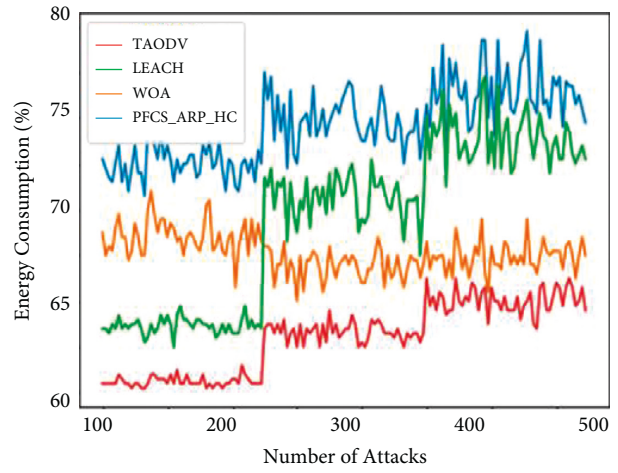


Figure 4: Comparative analysis of energy consumption.

4. Performance Analysis

The PFCS-ARP HC simulation configuration contains 100 mobile nodes that are randomly distributed within a terrain perimeter of 10001000 square metres. For validating the performance of PFCS-ARP HC, the pause period and simulation time were 20 seconds and 300 seconds, respectively. The PFCS-ARP HC simulation environment uses 802.11 as MAC protocol with 2 Mbps channel capacity of and Constant Bit Rate data source. Source and destination pairs of 20 and 50 mobile nodes are also used in the simulation investigation.

Table 1 shows comparative analysis for the proposed and existing techniques in minimizing the vampire attacks. Here the parameters compared are throughput, PDR, energy consumption, attack detection ratio, latency, and control overhead and compared with TAODV, LEACH, and WOA with the proposed PFCS-ARP_HC.

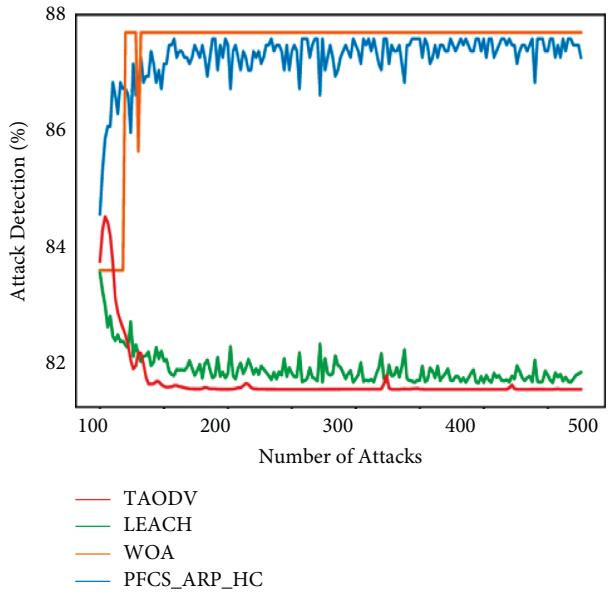


Figure 5: Comparative analysis of attack detection ratio.

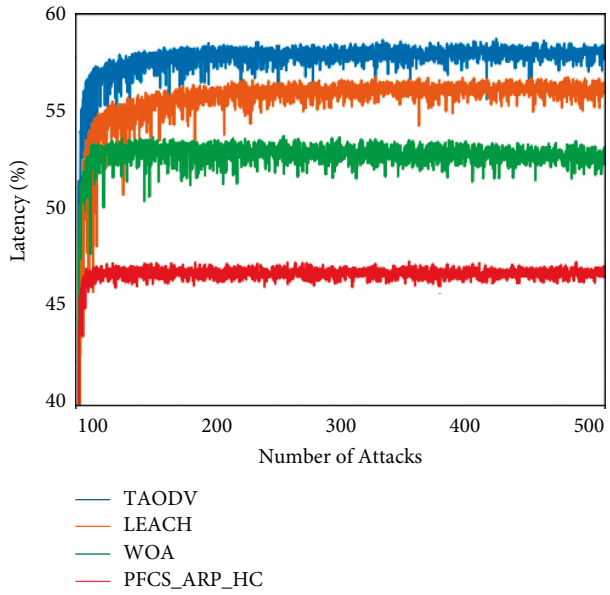


Figure 6: Comparative analysis of latency.

Figures 2–7 show comparative analysis of the proposed and existing techniques in vampire attack mitigation and improving the network performance. Comparative analysis has been carried out for throughput, PDR, energy consumption, attack detection ratio, latency, and control overhead. Here throughput attained by the proposed technique is 98% which enhanced when compared with TADOV, LEACH, and WOA. In terms of PDR, 89% has been obtained by the proposed technique; energy consumption by the proposed protocol is 75.5%, which is optimized when compared with existing protocol while data

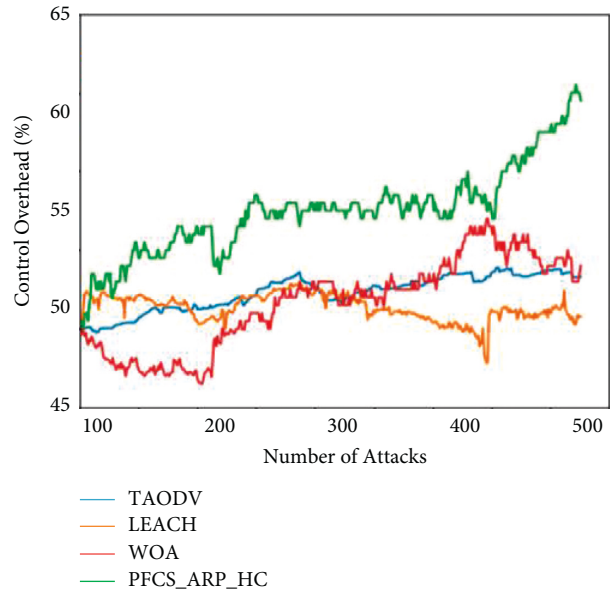


Figure 7: Comparative analysis of control overhead.

transmission. Since their energy consumption is high, the mitigation of vampire attacks by existing protocol is not efficient. Vampire attack detection ratio obtained by the proposed technique is 88% which is optimal than the existing techniques. Latency of the proposed technique in network is 46% and control overhead is 62% which is enhanced based on this comparative analysis.

5. Conclusion

This research offered a unique design strategy for minimising vampire attacks and improving network performance by utilising a security-based routing protocol and clustering-based data optimization. The data was optimised using a hybrid clustering approach, and the vampire attack was neutralised using a probabilistic fuzzy chain paired with an authentication-based routing protocol. The routing security of the network has been improved by authentication routing protocol. The performance can be improved by optimizing the data and network by hybrid-based sensor nodes clustering. Simulation for this technique has been carried out using NS2 and experimental results show that the performance of the proposed model in terms of throughput of 98%, packet delivery ratio of 89%, energy consumption of 67%, latency of 46% control overhead of 53%, and attack detection ratio of 87.9% [35].

Data Availability

The data used to support the findings of this study are available from the author upon request (kusumasyadav0@gmail.com).

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] C. Wang, R. S. Batth, P. Zhang, G. S. Aujla, Y. Duan, and L. Ren, "VNE solution for network differentiated QoS and security requirements: from the perspective of deep reinforcement learning," *Computing*, vol. 103, no. 6, pp. 1061–1083, 2021.

[2] B. Gao, T. Maekawa, D. Amagata, and T. Hara, "Detecting reinforcement learning-based Grey hole attack in Mobile wireless sensor networks," *IEICE - Transactions on Communications*, vol. E103.B, no. 5, pp. 504–516, 2020.

[3] D. Zhang, T. Zhang, and X. Liu, "Novel self-adaptive routing service algorithm for application in VANET," *Applied Intelligence*, vol. 49, no. 5, pp. 1866–1879, 2019.

[4] M. K. Shahzad, L. Nkenyereye, and S. M. R. Islam, "A fuzzy system based approach to extend network lifetime for enroute filtering schemes in WSNs," in *Proceedings of the 2019 11th International Conference on Computer and Automation Engineering - ICCAE 2019*, Perth, Australia, February 2019.

[5] G. Dhiman, K. K. Singh, A. Slowik et al., "EMoSOA: a new evolutionary multi-objective seagull optimization algorithm for global optimization," *International Journal of Machine Learning and Cybernetics*, vol. 12, no. 2, pp. 571–596, 2021.

[6] R. Kumar and G. Dhiman, "A comparative study of fuzzy optimization through fuzzy number," *International Journal of Modern Research*, vol. 1, no. 1, pp. 1–14, 2021.

[7] P. K. Vaishnav, S. Sharma, and P. Sharma, "Analytical review analysis for screening COVID-19 disease," *International Journal of Modern Research*, vol. 1, no. 1, pp. 22–29, 2021.

[8] V. K. Gupta, S. K. Shukla, and R. S. Rawat, "Crime tracking system and people's safety in India using machine learning approaches," *International Journal of Modern Research*, vol. 2, no. 1, pp. 1–7, 2022.

[9] C. Nalini and G. L. V. Prasad, "Secure routing protocol in WSN against vampire attacks," *International Journal of Modern Agriculture*, vol. 9, no. 4, pp. 1247–1253, 2020.

[10] V. Verma and V. K. Jha, "Detection and prevention of vampire attack for MANET," in *Nanoelectronics, Circuits and Communication Systems*, pp. 81–90, Springer, Singapore, 2021.

[11] T. Sharma, R. Nair, and S. Gomathi, "Breast cancer image classification using transfer learning and convolutional neural network," *International Journal of Modern Research*, vol. 2, no. 1, pp. 8–16, 2022.

[12] S. K. Shukla, V. K. Gupta, K. Joshi, A. Gupta, and M. K. Singh, "Self-aware execution environment model (SAE2) for the performance improvement of multicore systems," *International Journal of Modern Research*, vol. 2, no. 1, pp. 17–27, 2022.

[13] P. B. Srikaanth and V. Nagarajan, "Fuzzy rough set derived probabilistic variable precision-based mitigation technique for vampire attack in MANETs," *Wireless Personal Communications*, vol. 121, no. 1, pp. 1085–1101, 2021.

[14] M. S. Mekala, G. Dhiman, G. Srivastava et al., "A DRL-based service offloading approach using DAG for edge computational orchestration," *IEEE Transactions on Computational Social Systems*, 2022, in Press.

[15] B. Sumathy, A. Chakrabarty, S. Gupta et al., "Prediction of diabetic retinopathy using health records with machine learning classifiers and data science," *International Journal of Reliable and Quality E-Healthcare*, vol. 11, no. 2, pp. 1–16, 2022.

[16] I. S. R and J. J, "A secure routing scheme to mitigate attack in wireless adhoc sensor network," *Computers & Security*, vol. 103, Article ID 102197, 2021.

[17] W. Viriyasitavat, L. D. Xu, A. Sapsomboon, G. Dhiman, and D. Hoonsopon, "Building trust of Blockchain-based Internet-of-Thing services using public key infrastructure," *Enterprise Information Systems*, pp. 1–24. 2022, in Press.

[18] G. Dhiman, J. Rashid, J. Kim, S. Juneja, W. Viriyasitavat, and K. Gulati, "Privacy for healthcare data using the byzantine consensus method," *IETE Journal of Research*, pp. 1–12. 2022, in Press.

[19] A. A. Jasim, M. Y. I. Idris, S. Razalli Bin Azzuhri, N. R. Issa, M. T. Rahman, and M. F. b. Khyasudeen, "Energy-Efficient wireless sensor network with an unequal clustering protocol based on a balanced energy method (EEUCB)," *Sensors*, vol. 21, no. 3, p. 784, 2021.

[20] C. Lyu, X. Zhang, Z. Liu, and C.-H. Chi, "Selective authentication based geographic opportunistic routing in wireless sensor networks for Internet of Things against DoS attacks," *IEEE Access*, vol. 7, pp. 31068–31082, 2019.

[21] J. Zhou, "Efficient and secure routing protocol based on encryption and authentication for wireless sensor networks," *International Journal of Distributed Sensor Networks*, vol. 9, no. 4, Article ID 108968, 2013.

[22] P. K. Mishra and S. K. Verma, "FFMCP: feed-forward multi-clustering protocol using fuzzy logic for wireless sensor networks (WSNs)," *Energies*, vol. 14, no. 10, p. 2866, 2021.

[23] P. S. Mehra, M. N. Doja, and B. Alam, "Fuzzy based enhanced cluster head selection (FBECS) for WSN," *Journal of King Saud University Science*, vol. 32, no. 1, pp. 390–401, 2020.

[24] R. Isaac Sajan and J. Jasper, "Trust-based secure routing and the prevention of vampire attack in wireless ad hoc sensor network," *International Journal of Communication Systems*, vol. 33, no. 8, Article ID e4341, 2020.

[25] P. B. Srikaanth and V. Nagarajan, "Semi-Markov chain-based grey prediction-based mitigation scheme for vampire attacks in MANETs," *Cluster Computing*, vol. 22, no. 6, pp. 15541–15549, 2019.

[26] A. M. Desai and R. H. Jhaveri, "Secure routing in mobile ad hoc networks: a predictive approach," *International Journal of Information Technology*, vol. 11, no. 2, pp. 345–356, 2019.

[27] J. Yang, S. He, Y. Xu, L. Chen, and J. Ren, "A trusted routing scheme using blockchain and reinforcement learning for wireless sensor networks," *Sensors*, vol. 19, no. 4, p. 970, 2019.

[28] J. Jiang, X. Zhu, G. Han, M. Guizani, and L. Shu, "A dynamic trust evaluation and update mechanism based on C4.5 decision tree in underwater wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 9031–9040, 2020.

[29] S. Swain, B. Bhushan, G. Dhiman, and W. Viriyasitavat, "Appositeness of optimized and reliable machine learning for healthcare: a survey," *Archives of Computational Methods in Engineering*, vol. 1, 2022.

[30] R. Bharti, A. Khamparia, M. Shabaz, G. Dhiman, S. Pande, and P. Singh, "Prediction of heart disease using a combination of machine learning and deep learning," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 8387680, 2021.

[31] G. Sanjay Gandhi, K. Vikas, V. Ratnam, and K. Suresh Babu, "Grid clustering and fuzzy reinforcement-learning based energy-efficient data aggregation scheme for distributed WSN," *IET Communications*, vol. 14, no. 16, pp. 2840–2848, 2020.

[32] H. Upadhyay, S. Juneja, A. Juneja, G. Dhiman, and S. Kautish, "Evaluation of ergonomics-related disorders in online education using fuzzy AHP," *Computational Intelligence and Neuroscience*, vol. 2021, Article ID 2214971, 2021.

[33] G. Dhiman, K. K. Singh, M. Soni et al., "MOSOA: a new multi-objective seagull optimization algorithm," *Expert Systems with Applications*, vol. 167, Article ID 114150, 2021.

[34] H. Kaur, A. Rai, S. S. Bhatia, and G. Dhiman, "MOEPO: a novel multi-objective emperor penguin optimizer for global optimization: special application in ranking of cloud service providers," *Engineering Applications of Artificial Intelligence*, vol. 96, Article ID 104008, 2020.