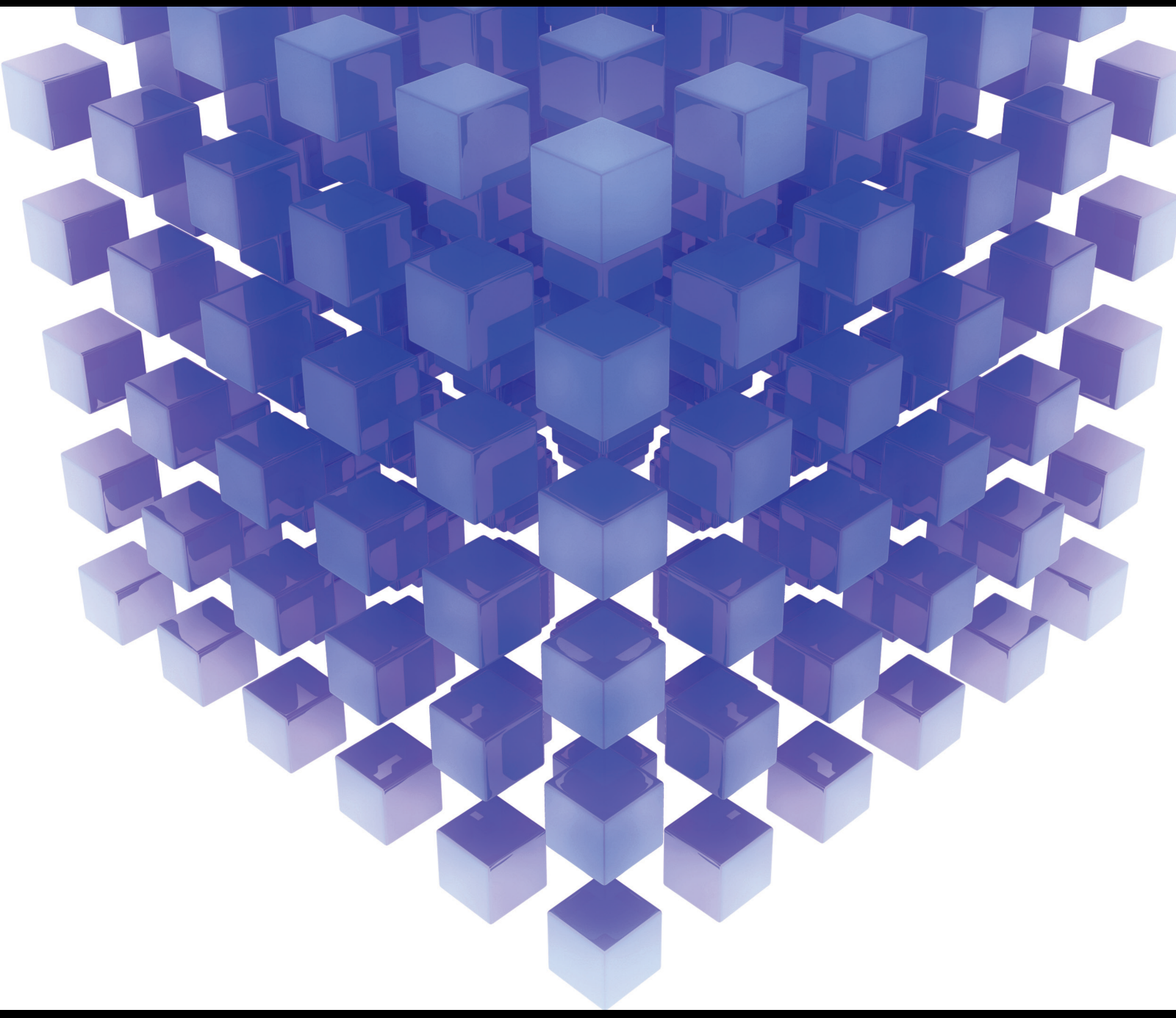
Mathematical Problems in Engineering

New Trends in Evolutionary Optimization for Big Data

Lead Guest Editor: Xiangtao Li
Guest Editors: Zhiqiang Ma, Jian Zhang, and Ka-Chun Wong



New Trends in Evolutionary Optimization for Big Data

Mathematical Problems in Engineering

New Trends in Evolutionary Optimization for Big Data

Lead Guest Editor: Xiangtao Li
Guest Editors: Zhiqiang Ma, Jian Zhang, and Ka-Chun Wong

Chief Editor

Academic Editors

Contents

Hindawi
Mathematical Problems in Engineering
Volume 2023, Article ID 9798161, 1 page
https://doi.org/10.1155/2023/9798161

Retraction

Retracted: Construction of a Multimedia-Based University Ideological and Political Big Data Cloud Service Teaching Resource Sharing Model

Mathematical Problems in Engineering

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Feng, W. Zhang, and S. Tsai, "Construction of a Multimedia-Based University Ideological and Political Big Data Cloud Service Teaching Resource Sharing Model," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9907630, 12 pages, 2021.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 4047508, 12 pages
https://doi.org/10.1155/2022/4047508

Research Article

Talking about the Innovative Application of Big Data in Enterprise Human Resources Performance Management

Dazhi Xu,[1,2] Tianyi Tu,[3] and Xiaoyong Xiao [2,4]

[1]*School of Public Administration, Central South University, Changsha 410083, Hunan, China*
[2]*College of Economics and Management, Hunan University of Arts and Science, Changde 415000, Hunan, China*
[3]*School of Computer and Electrical Engineering, Hunan University of Arts and Science, Changde 415000, Hunan, China*
[4]*Hunan Province Cooperative Innovation Center for the Construction and Development of Dongting Lake Ecological Economic Zone, Hunan University of Arts and Science, Changde 415000, Hunan, China*

Correspondence should be addressed to Xiaoyong Xiao; zxxy2001@huas.edu.cn

Received 20 January 2022; Revised 2 March 2022; Accepted 13 April 2022; Published 30 May 2022

Academic Editor: Xiangtao Li

With the advent of the wave of big data, data has become an important information asset. With data information, we have the ability to gain insight into market conditions and strengthen management. The use of big data technology to conduct statistics, analysis, and mining of massive information can make information assets play a very high value in corporate strategy formulation and accurate decision-making. However, the traditional performance management concepts and management methods have not adapted to the development requirements of the times, and there are various drawbacks. This paper analyzes the impact of big data on enterprise performance management. Combining the key aspects of the company's performance in the old data era, a new method of company performance management and the subsequent innovative development path are proposed. Recognition frameworks, employee similarity, FCM, and other assessment formulas are emphasized. From the employee's work attitude, workability and appearance, and daily performance, combined with the calculation of employee benefits, the evaluation is more clear; after the implementation of the material reward experiment, the attendance rate increased by 16.3%. At the same time, the workload was gradually refined, from quantitative to qualitative, and sales increased by 28.6%, maximizing the advantages of human resources and promoting the sustainable development of the enterprise.

1. Introduction

At present, the essence of competition among Chinese enterprises is professional competition. If enterprises want to benefit from fierce competition, they must pay attention to quality and give full play to the positive role of competitiveness in enterprise development. Therefore, companies need to strengthen human resource management. Through human resource management, companies can select and hire skilled workers, manage effective positions, continue to play their full role and abilities, and create new impetus for the development of the company. As far as corporate personnel management is concerned, performance management is an effective way to manage companies. In particular, in the era of big data, companies are facing a large amount of data and resource management performance pressure when conducting employee management. Enterprises need to innovate human resource performance management based on big data properties. The research in this paper well uses the data fusion technology to realize the design of the enterprise human resource performance management innovation system. At this stage, there are many problems in the practice of human resource performance management. HR departments still apply traditional human resource management ideas and lack of innovation is the problem. Therefore, it is expected that the research can promote companies to make full use of big data fusion technology to better implement human resources. Performance management creates favorable conditions for the company's management and sustainable development.

1.1. Review of Literature. Under big data, the creation of creative resources for managing corporate human resource performance is conducive to the growth of the corporate culture. Due to the far-reaching impact of big data, most business departments are very aware of its applications. In the era of big data and information technology, the number of applications continues to grow. Therefore, if we innovate in human resource performance management based on this method and use appropriate technology, we can reduce the problems caused by face-to-face communication and improve work efficiency. In this way, network technology can become the normal state of communication between employees, employees and executives, and employees and customers and further promote the development of a good corporate culture. Liu S believes that the performance evaluation of human resource management is of great significance for improving the standard of food businesses, so the application of gray system theory should be studied in depth. He first discussed the difficulties of performance evaluation of human management level in food enterprises. Secondly, the basic theory of gray system theory is deeply studied, and the corresponding mathematical model is deeply analyzed. The performance evaluation index system of the human resource management level of food enterprises is constructed again, and a corresponding evaluation is carried out with a food enterprise as an example. The results show that the gray system theory is an effective tool to evaluate the human resource management level of food enterprises [1]. Banimelhem studied the link between HR practices in the UAE's healthcare sector and employees' willingness to leave. This quantitative study used structural formula modeling (SEM) technology and the Moment Structure Analysis (AMOS) as a suite of software to analyze data and study the impact of HR behavior on staff turnover and intentions in healthcare services in the United Arab Emirates. The results of the study show that the human resources management (HRM) practices (recruitment and selection, performance evaluation, salary, and career development) of the healthcare sector have a significant relationship with the willingness to leave. However, his research is restricted to the public sector in the UAE, and it is therefore recommended to include the inclusion of the private health provision [2]. Big data has transformed research in many areas, including the fields of business such as marketing, finance, accounting, and the management of the supply network. However, the debate on big data analytics in human capital management is largely focused on the selection of job candidates. Hamilton considers how to solve important strategic human capital issues through big data analysis so that HR can improve overall company performance. Hamilton has also researched new data sources that can help evaluate workforce performance in real time, helping to identify knowledge stars who contribute to company performance and help strengthen the company's abilities. But for big data analysis to be successful in HR, it also needs to address regulatory and behavioral challenges, including data privacy issues and the General Data Privacy Protection Regulation (GDPR) in Europe. Hamilton finally discussed how big data analysis can promote human resources and strategic changes throughout the organization [3]. The purpose of Vargas' research is to show the relationship between the drivers of building knowledge and human capital for development (HC). These learning studies have a huge impact on the functioning of the organization. To accomplish this, Vargas developed a new relationship theory model and tested the model based on experience. He used data from companies in the Spanish biotechnology industry to quantitatively test the model. The results show that there is a close relationship between HC and the driving factors of knowledge creation (redundancy, trust, and autonomy) [4]. The primary focus of HRM studies is on "strategic HRM," i.e., the influence of HRM on company behavior. Not only are the cumulative results of Brewster's discussion of "leading research orthodoxy" disappointing in terms of their external validity, but their practical value is also limited. In addition, not only did it fail in its narrow corporate performance-oriented agenda, but its agenda principles also led to serious employee dissatisfaction and failure to deal with pressing global issues. To assess the contribution of mainstream research orthodoxy, Brewster analyzed the 16 most cited journal articles in the field of human resource management. The survey results found that US-centered research is dominant, so the transnational universality of the dominant research orthodoxy is questionable. Using cross-sectional data means that long-term impact cannot be measured. Brewster observed a lack of consensus on how to implement human resource management and company performance. The practical significance shows that, in order for human resource management to realize its potential for government, media, or charity organizations, human resource management must give up its limited scope and single-dimensional source of inspiration. The author not only pointed out the shortcomings of the dominant research orthodoxy in HRM but also pointed out how HRM has become more "centralized" by seeking participants who contribute to major issues in the world [5]. Otoo examines the mediating role of employee competence in the relationship between human resource management (HRM) practices and organizational performance. The comprehensive research model is developed by combining the main factors in the existing literature. Data was collected from 600 employees in selected hotels through a questionnaire survey, and structural formula modeling was used to test the validity of the model and hypothesis. The reliability and validity of these dimensions are established through confirmatory factor analysis. The results show that some human resource management practices affect organizational performance by improving employee competence. The research further shows that employee competence plays a mediating role between HRM practice and organizational performance [6]. Robert M identifies the inhibitors that affect the implementation phase of management innovation over time and addresses how they have evolved [7]. Szabo and Csontos use an exploratory research approach through a multilayered case study of a machinery company (observations, document analysis, and interviews). Szabo and Csontos found that to achieve the expected efficiency gains, management innovation must precede technological

innovation. Tight interorganizational networks and change agents are also required to catalyze these processes [8].

2. Plan for Rational Utilization of Human Resources

Literature analysis method: big data and "cloud computing" are hot topics in the current social sciences. Big data and human resource management are also deeply concerned by people in the industry. A large number of experts and scholars have extensively discussed this topic and have achieved considerable results. In the early stage of this research, I collected and read a lot of information about big data and its technology and human resources performance management through the Internet, related books, related literature, etc., had a wide understanding of the related concepts of big data and human resources performance management, summarize and analyze the previous research results, and accumulate abundant data for this research. Case analysis method: there are many related theories such as big data and human resource performance management. However, they must be introduced into practice to really play a role. Through the research and analysis of typical cases, this paper analyzes the current management status of the company and finds out the main problems of its human resources performance management. The performance management is innovated and optimized to achieve the purpose of improving the management level of the company.

2.1. Developing the Talents of Employees and Enhancing the Effectiveness of Cooperation between Members. If the overall quality of employees can meet the requirements of the business, business development will be carried out smoothly; otherwise, business development will be hindered. In order to build the complete quality of employees and business development to complement each other and develop simultaneously, the following tripartite work needs to be carried out, strengthen ideological and political education, emphasize concept learning, and guide employees to solid "four sobrieties," that is, hard-working business acumen, unacceptable anxiety, sense of collaboration, and new creativity. "Business is getting harder and harder to maintain;" if a company wants to continue to thrive in a highly competitive environment, it must have strong professional ethics, overcome challenges, and move forward boldly. "If there is no long-term anxiety, there must be urgency" which refers to the feeling of anxiety. If a company loses its sense of anxiety, it can only be arrogant; anxiety will cause people to focus on self-improvement, self-improvement, innovation, and development. The power of an individual is destined to be small. When the smaller forces unite into a unified group, its power becomes infinite until it becomes feasible and invincible. The ability of employees is limited. When most employees move to one place, it will have a significant impact on the team. Only by continuously improving the innovation ability of the enterprise and building an innovative enterprise can the economic viability of the enterprise be improved and the rapid development of the enterprise can be promoted.

To realize the effective integration of employee information, consider the identification framework. The recognition frame is usually represented by the mathematical symbol θ. The recognition frame is a complete set, which needs to contain all the known and wanted-to-know possibilities for a certain problem. Its structure depends on people's cognitive scope and knowledge level, and its structure needs to meet the following.

First, for a decision problem, any proposition about the decision is a subset of θ, and the elements are mutually exclusive. Therefore, the selection of the recognition framework should be rich enough to satisfy that any proposition we consider can correspond to a subset of θ; second, the introduction of data fusion and key technology research. Since the subsets are mutually exclusive, set the number of elements to n, and denote θ as

$$\theta = \{1, 2, \ldots, \theta_n\}. \tag{1}$$

When θ contains N elements, there are at most 2^n subsets in subset A, and the finite set formed by these 2^n subsets is the power set of θ, which is the hypothesis space formed by the recognition framework, denoted as

$$2\theta = \{\varphi, \{\theta_1\}, \{\theta_{,2}\}, \ldots \{\theta_1, \theta_2\}, \{\theta_1, \theta_3\} \ldots, \{\theta_1, \theta_2, \theta_3\}, \ldots \theta. \tag{2}$$

Among them, φ represents the empty set. Any subset A of the recognition framework corresponds to a proposition, which can be interpreted as follows: the answer to the question is in A, and the recognition framework is established:

$$\theta = \{1, 2, 3, 4, 5, 6\}. \tag{3}$$

Each subset A in the power set represents a proposition, which may be a hypothetical answer [9].

The basic possibility of allocation also has a significant impact on the data results.

The abbreviation of the Basic Probability Assignment function is BPA. BPA expresses the degree of trust of the evidence to the proposition. In evidence theory, whether the basic probability assignment is reasonable has a great influence on the result of the evidence combination.

The function $m(x)$ on the recognition frame θ is a mapping of $2\theta \longrightarrow [0, 1]$, and A is any subset of θ, denoted as $A \subseteq 2^\theta$, if the function $m(x)$ satisfies the following condition:

$$\begin{cases} m(\varphi) = 0, \\ \sum\limits_{A \subseteq 2^\theta} m(A) = 1. \end{cases} \tag{4}$$

At the same time, consider the use of the Belief Function: setting $A \subseteq 2\theta$ is any subset of the recognition frame θ, and the sum of the basic confidences corresponding to all the subsets in A is called the trust function Bel (A), namely:

$$\begin{aligned} &Bel\colon 2^\theta \longrightarrow [0, 1], \\ &Bel(A) = \sum_{B \subseteq A} m(B). \end{aligned} \tag{5}$$

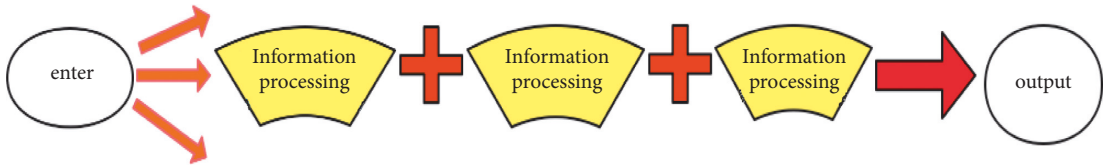


Figure 1: Information input integration analysis output flow chart.

Bel (A) is also called the reliability or confidence function of the event A value, which represents the confidence that the evidence is correct A; the trust value of the empty set is 0 [10]. The trust function is an estimate of the lower limit of the degree of trust in the imagination, so the trust function is sometimes called the lower limit function or the confidence function. It is easy to find the following definition:

$$\begin{aligned} \text{Bel}(\varphi) &= m(\varphi) = 0, \\ \text{Bel}(\theta) &= \sum_{B \subseteq \theta} m(B) = 1. \end{aligned} \tag{6}$$

Using a variety of big data fusion methods is more effective in the integration of employee information and the analysis and utilization of personnel and maximizes the value of personnel, as shown in Figure 1.

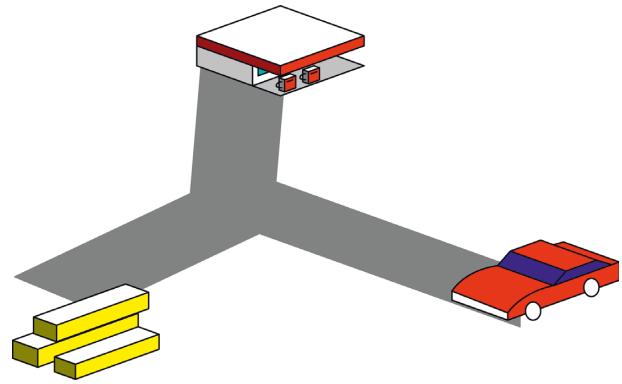


Figure 2: Employee incentives are like refueling cars and gold bullion rewards.

2.2. Coordinating the Corporate Management Mechanism to Realize the Great Optimization of the Corporate Human Resource Performance System. The personnel performance management of the business department evaluates the performance of each employee by formulating a unique evaluation system [11, 12]. Performance management aims to ensure the achievement of work goals. In the process of implementing employee performance management, corporate departments must go through several steps, such as setting performance goals and implementing performance models. In order to improve work efficiency and optimize the income distribution system, the personnel performance management of the business department should adhere to the following basic principles: (1) based on good management, conduct comprehensive performance planning and performance management for related departments; (2) strategies and policies applicable to all employees; (3) to give full play to the role of motivation, not only to motivate employees at the spiritual level but also to make employees feel tangible and substantial returns [13], as shown in Figure 2.

2.3. Management Talents Are Effective and the Evaluation Model Is Scientific and Objective. Improving and perfecting the staff appraisal system for performance is of great relevance in enhancing the scientific level of team management, effectively stimulating the initiative and creativity of staff and creating a new situation at work [14]. From this perspective, the following aspects of the performance appraisal system should be noted.

The standards of performance appraisal should reflect the specific characteristics of the department and position. There are significant variations between the various departments and positions in the company, and different performance appraisal standards need to be established to reflect or adapt to this difference. The appraisal standards match the department functions and job requirements and can reflect the differences in the positions and positions of employees. The different requirements of work tasks enable the performance evaluation and comparison of different job levels, different departments, and different types of staff to be achieved [15, 16]. Based on the evolution of the big data era, existing corporate human resource performance management should pay more attention to improving and enriching relevant database information to ensure that data resources are more accessible and more effective. The characteristics of innovative thinking and management strategies are shown in Figure 3.

Obtaining comprehensive decision-making requires a method to calculate the comprehensive influence of multiple pieces of evidence on each hypothesis in the identification framework and obtain the comprehensive trust level that makes the hypothesis valid under the action of multiple pieces of evidence [17]. BPA is the basis of the trust function and likelihood function. In practical applications, it often appears that for the same assumption or problem, the evidence comes from different data sources, so two or more different BPAs will be obtained. Therefore, in the future, the likelihood function and trust function can be better used to measure the credibility of the proposition. We need a synthesis rule that combines BPA from different data sources [18].

(1) Two evidence functions synthesize a certain proposition A. For all $A \subseteq \theta$, proposition A is for two mass functions $n_1, n_{2,}$ on θ on the same recognition frame, and their Dempster composition rule is

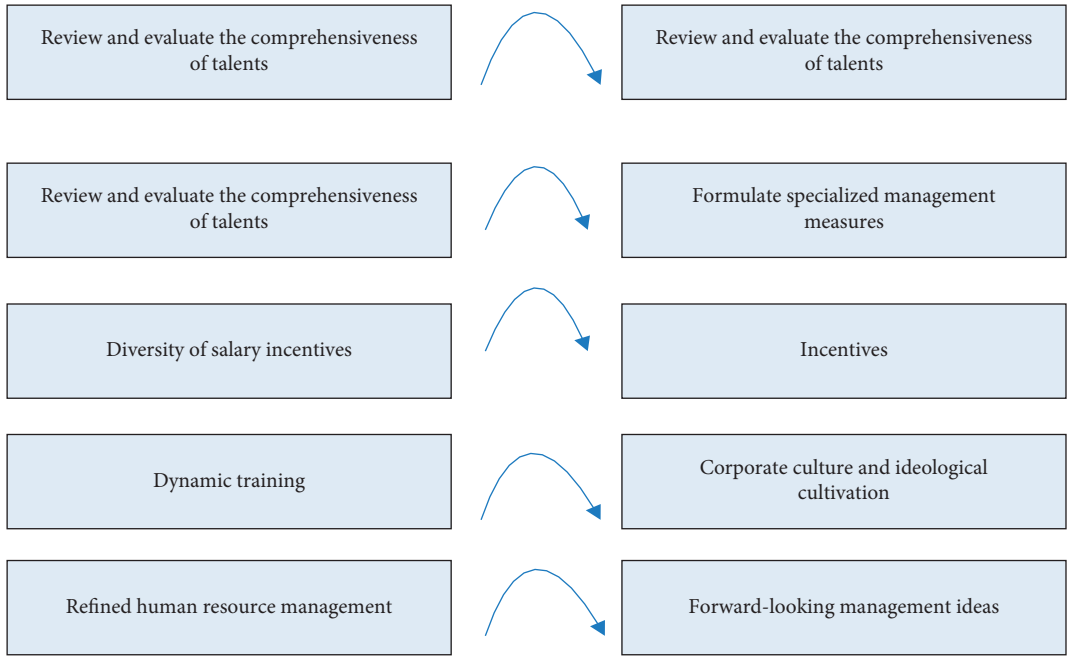


Figure 3: Framework diagram of innovative thinking and management strategy of employee human resource management.

$$n_1 \oplus n_2(A) = \frac{1}{k} \sum_{B \cap C = \theta} n_1 B \bullet n_2(C). \tag{7}$$

Among them, the symbol $\square$ represents the orthogonal sum [19]. In order to make the sum of the mass functions 1, K is the normalization constant:

$$K = \sum_{B \cap C \neq \theta} n_1(B) \bullet n_2(C) = 1 - \sum_{B \cap C = \varphi} n_1(B) \bullet n_2(C). \tag{8}$$

If $K = 1$, it means that the two pieces of evidence have a strong conflict, so there is no orthogonal sum; if $K \neq 1$, then the orthogonal sum of the BPA of the two pieces of evidence forms a new distribution function; if $K^{-1} = 0$, it is considered that n_1, n_2, is contradictory and there is no joint basic probability distribution function [17].

(2) Combination of multiple pieces of evidence. For all $A \subseteq \theta$, when it is necessary to process a limited number of mass functions $n_1, n_2, \ldots, n_m$ from multiple sources of evidence on the recognition framework θ, the same method can be used to orthogonally sum multiple basic probability distribution functions into a basic trust function. The Dempster composition rule is

$$\begin{aligned}(n_1 \oplus n_2 \oplus \cdots \oplus n_m)(A) &= \frac{1}{K} \sum A_1 \cap A_2 \ldots \cap A_m \\ &= {}_A n_1(A_1) \bullet n_2(A_2) \ldots n_m(A_m).\end{aligned} \tag{9}$$

Among them, normalized to a constant K is

$$\begin{aligned}K &= \sum_{A_1 \cap A_2 \cap \ldots \cap A_m \neq 0} n_1(A_1) \cdot n_2(A_2) \ldots n_m(A_m) \\ &= 1 - \sum A_1 \cap A_2 \cap \ldots \cap A_m = {}_\varphi n_1(A_1) \cdot n_2(A_2) \ldots n_m(A_n).\end{aligned} \tag{10}$$

In summary, the statistics of personnel data will be more accurate [20].

3. Specific Data Information of Experimental Research

3.1. Application of Big Data Technology. At present, people engaged in human resource performance management need to master a large amount of data as a whole. They can use the technology integration analysis in the era of big data to study the basic characteristics of a single TCP connection and the flow characteristics calculated in a 2-second window. We analyze and effectively use these technologies to apply to employee data collection and integration analysis [21, 22], as shown in Tables 1 and 2.

And it is necessary to update the actual content of performance management and also to update the management practice methods, effectively improve the creativity of all employees in performance management, and conduct accurate and basic information collection on business data [23]. Therefore, in this process, the company must update its database in time, not only to store and understand all the information related to the company's employees but also to make a paper record. There are also job types. In order to ensure that the company's employee office system is complete, when a problem occurs, the paper type of the document can be used to inquire in time to ensure the integrity of the work. In the era of big data, enterprise management can

Table 1: Basic characteristics of a single TCP connection.

Serial number	Field name	Describe	Value
1	Duration	Duration unit second	[0, 58329]
2	Protpcpl_type	Agreement type	TCP, UDP, ICMP
3	Service	Network service type of the target host	HTTP, Tel ent70 kinds
4	Src_bytes	Data volume from origin address to purpose of destination address	[0, 1379963888]
5	Dst_bytes	Amount of data from the origin address to the destination address	[0.1309937401]
6	Flag	The connection status is correct or incorrect	"OTH,".REJ'11 kinds
7	Land	If the source address and destination address of the data connection are a unified host or port, take 1; if it is other conditions, take 0	{0, 1}
8	Wrong_fragment	Number of faulty segments	{0, 37}
9	Urgent	Number of urgent data packets	{0, 14}

Table 2: Flow characteristics calculated in a time window of 2 seconds (Table 2 is reproduced from Bing Zhang et al. 2018 [under the Creative CommonslAttribution License/public domain]).

Serial number	Field name	Describe	Value
10	Count	The number of connections with the same target host as the current connection	[0, 511]
11	Srv_count	The number of connections with the same service as the current connection	[0, 511]
12	Serror_rate	The percentage of connections that have an "SYN" error in the connections that have the same target host as the current connection	[0.00, 1.00]
13	Srv_serror_rate	The percentage of connections that have "SYN" errors among connections that have the same service as the current connection	[0.00, 1.00]

borrow more advanced ways to build enterprise organizational structure, so as to achieve the effect of promoting human resource management [24]. As shown in Figure 4, the data layer uses SQL data mapping and SQL database.

In the era of big data, companies need to pay more attention to the human resources department, adjust the time management model, and speed up the pace. The personnel management system can store and record all the data components such as the workability of employees so that outsiders can know the information at any time. The interconnected system makes it easier for companies to navigate. Whether they are employees in other regions or even abroad, the system will restore employee information, which is convenient for real-time monitoring of the company and employee management [22]. Starting from the second layer, employees can understand their own data and the data of others through the system itself. When many employees see the difference between themselves and others, they will have comparative psychology, and they will work harder to diversify their choices. Estimation technology is quick and simple; the decision-making process is transparent; it creates an actively collaborative experience rather than adversarial.

3.2. Reasonably Choose the Main Body of Performance Appraisal. The resources for performance appraisal are multifaceted, and the core part of the evaluation should be multifaceted. It is important to check the detection target to avoid confusion, interaction, and influence. At present, human resource management performance evaluation lacks specific performance in the evaluation process, resulting in significant differences in operation management. To manage employee performance, many companies still use the old-fashioned method, using the evaluation method of the leader or manager, and adding many questions to get the evaluation result. At the same time, the true impact of this evaluation is irrelevant, and it has nothing to do with improving performance management performance. Generally speaking, the person being assessed and their leaders, colleagues, and subordinates should be investigated. Other evaluation topics can also be determined in combination with the specific standards of different positions. When the situation of two employees is very similar, how to divide their rank. For example, the project department has increased the satisfaction review of unit construction and the evaluation of the units participating in the project, the evaluation of employee performance, and the comprehensive evaluation of employee performance, professional ethics, and professional standards. And due to the processing and analysis of performance data, many top corporate auditors did not conduct in-depth research. They do not have a complete understanding of database organization. Employees should be able to score and evaluate leaders participating in the evaluation to promote outstanding elected leaders. As shown in Figure 5, the ABC of the Administration Department and the DEF Management Department of the Sales Department show the scores.

The employees can be evaluated from the following aspects as shown in Table 3.

The working attitude of the employees is generally good, but the work effect of the project is still uneven, including the need for more control over attendance.

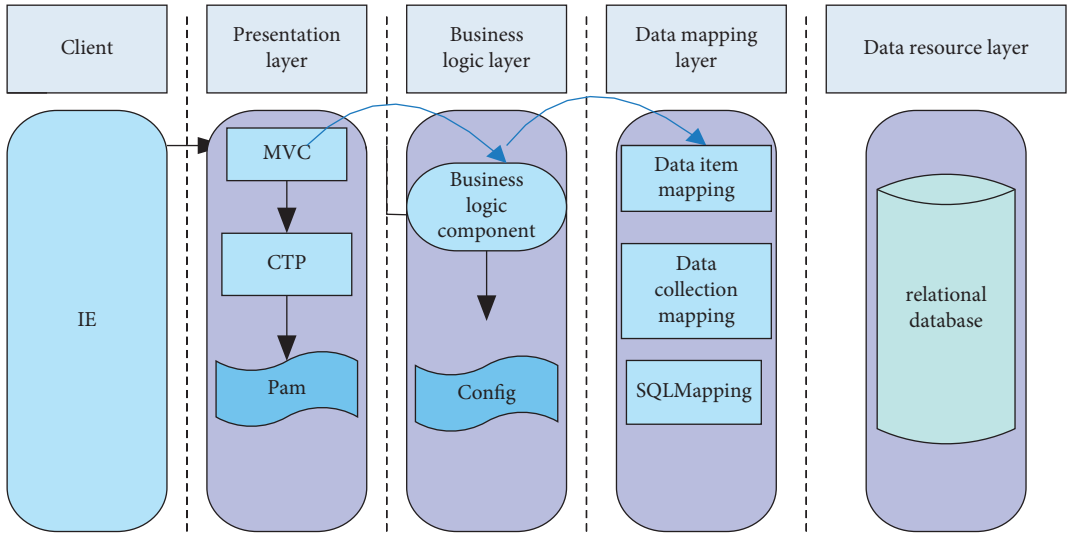


Figure 4: Architecture design diagram of each level of the application system.

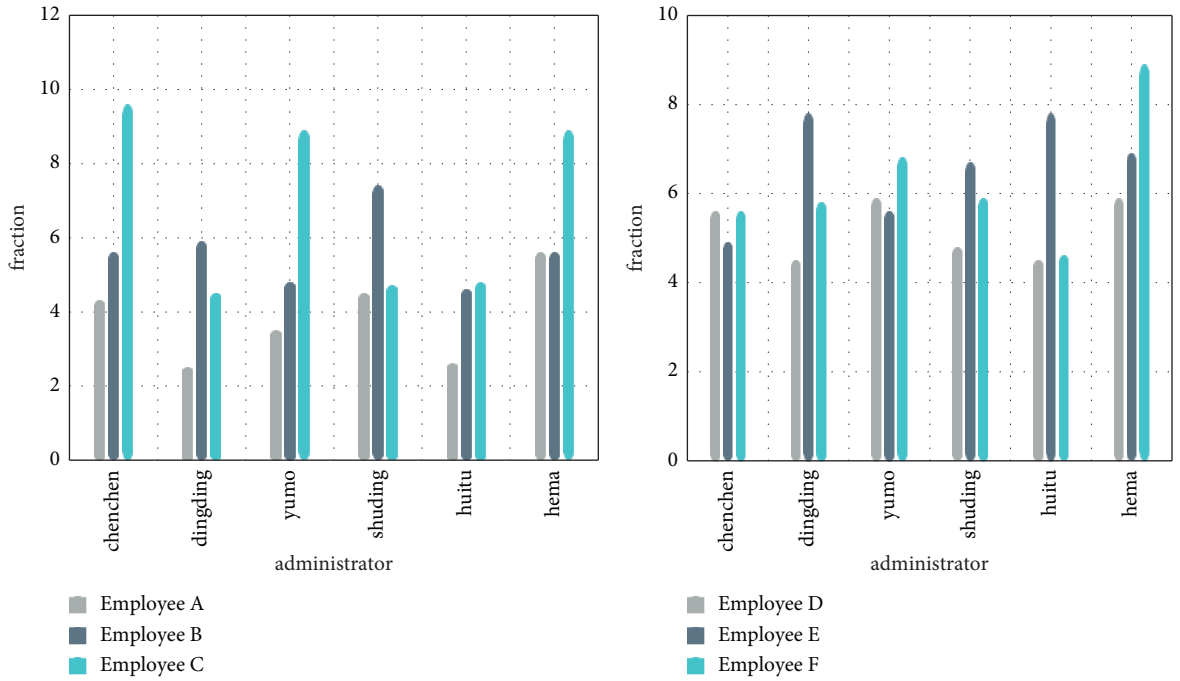


Figure 5: Anonymous evaluation of leaders by employees of the Administration Department and Sales Department.

3.3. Employee Differences. For how two similar employees can distinguish their differences, the similarity measurement can be considered. The similarity measurement is a measurement of the similarity between two objects in a clustering result. There are two measurement methods: the degree of dissimilarity expressed by the distance between objects and the similarity expressed by the correlation with the object.

Commonly used similarity measurement methods include Euclidean distance, Manhattan distance, and other calculation distance measurement methods, cosine similarity, correlation coefficient method, and other similarity measurement methods. At present, a denoising method based on Euclidean distance is commonly used. This method determines whether certain data belongs to noise data by setting a Euclidean distance threshold. The Euclidean distance can be calculated by the following formula:

$$d(x, v_i) = \sqrt{(x - v_i)^T (x + v_i)}. \tag{11}$$

Cosine similarity is also a common similarity measurement method. This method uses the cosine value formed between two samples as a measure of similarity. Therefore, cosine similarity pays more attention to the difference in direction. Its calculation formula is as follows:

Table 3: A hundred-point evaluation is carried out on the aspects of work attitude, workability, and appearance of employees.

Name	Work attitude	Work attitude	Daily behavior	Attendance	Manner and behavior
Dingding	95	90	90	88	52
Chongchong	81	99	56	22	20
Junjun	100	100	52	42	85
Zhouzhou	26	60	96	98	71
Qingqing	90	80	53	52	65

$$\theta = \operatorname{sim}(X, Y) = \cos\theta \frac{\overrightarrow{x}\,\overrightarrow{y}}{|x||y|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i=1}^{n} y_i^2}}. \quad (12)$$

The value range of cosine similarity is [−1, 1]. According to the definition of cosine value, when the cosine value is larger, the angle between them is smaller, and the two samples are more similar in this direction. On the contrary, it is the opposite, which is "cosine similarity."

But it only cares about the absolute distance between two objects, and it treats the differences between different attributes equally and cannot reflect some requirements in practical applications. In order to avoid the possible misjudgment of the method, this paper will use the cosine value of the angle between the point and the cluster center to weigh the Euclidean distance based on the cosine similarity. Assuming that the cluster center of a cluster after clustering is v_j, for any point $x_t^{(j)}$ in this cluster, the weighted Euclidean distance is

$$d_v\left(y_t, v_j\right) = \operatorname{sim}\left(y_t^{(j)}, v_j\right) \cdot \sqrt{\left(y_t^{(j)} - v_j\right)^t \left(y_t^{(j)} + v_j\right)}. \quad (13)$$

where $t = |v_j|$, $|v_j|$ represents the number of samples in a cluster with v_j as the cluster center, and $y_t^{(j)}$ represents all sample points in the cluster where the cluster center v_j is located.

The following describes the improved denoising algorithm based on the FCM algorithm.

First, set a Euclidean distance b as the threshold. Generally, b is the uniform value $\int$ of the weighted Euclidean distance from all sample points in the cluster to the cluster center. After the target denoising data is clustered,

$$d_v\left(y_t, v_j\right) \rangle b. \quad (14)$$

The sample point is a noise point, and delete it, when

$$d_v\left(y_t, v_j\right) \langle b. \quad (15)$$

Then, leave the sample:

$$\int = \frac{\sum_1^k \sum_1^t \mathrm{d}v\left(y_t^{(i)}, v_i\right)}{n}. \quad (16)$$

For the value of the number c of cluster centers, there is a method called the Elbow Method, which can provide a reference.

A variety of methods can integrate the sales staff information as shown in Table 4.

If you want to make the difference between the employees' work in the first half of the year more obvious, it can follow Figure 6 as usual.

The information integration of the same administrative staff is shown in Figure 7.

All the above information is integrated to get the company's annual income and the benefits generated by employees. Statistical evidence refers to the evidence obtained through statistical experiments. A set of evidence is determined by the statistical probability model $\{B_a | a \in \theta$. The given element a, B_a is a probability density function, and there are two assumptions. The first hypothesis is that the observation y determines a likelihood function that satisfies

$$Bl(a) = C \cdot B_a(y), \forall_a \in \theta. \quad (17)$$

The second assumption is that the likelihood function satisfies consistency; that is, for any focal element A_i in the recognition frame θ, there will always be $\mathrm{m}(A_i) > 0$ and $\sum_{i=1}^{\mathrm{r}} \mathrm{m}(A_i)$. Under these two assumptions, the likelihood function is obtained:

$$BI(A) = \frac{\max\{B_O(X): o \in A\}}{\max\{B_o(x): o \in \theta\}}. \quad (18)$$

Suppose $\theta^0 = \{o^{(1)}, o^{(2)}, \ldots, o^{(N)}\}$ is an ordered set of θ, which satisfies $B_O^i B_o^j$, $\forall 1 \le i < j \le N$. The corresponding basic probability distribution function is

$$m_x(A) = \begin{cases} \dfrac{B_O^{(K)}(X) - B_O^{(K+1)}(X)}{B_O^{(1)}(X)}, \forall A = \left\{O^{(1)}, O^{(2)}, \ldots, O(k)\right\}, 1 \le k \le N-1 \\ \dfrac{B_O^{(N)}(x)}{B_O^{(1)}(x)}, A = \theta = \theta^0 \end{cases}. \quad (19)$$

Based on the above algorithm, calculating the employees' annual income for the company from 2013 to 2021 is shown in Figure 8.

3.4. Other Assessment Factors. However, it focuses on analyzing the company's annual accounts and company profits, rather than analyzing the problems in the data. Some seemingly irrelevant data are ignored, such as diet and work environment. In fact, if this information is integrated and analyzed, combined with the monthly sales of employees, it will have important reference value for the performance appraisal of employees, as shown in Figure 9.

This can accurately describe the current performance of employees and help companies develop human resource management and effective practices. Human resource management is part of the daily operations of an enterprise. The relationship between employees and leaders and communication can ensure that the results of performance management are fully implemented. However, company

Table 4: Workload analysis of sales staff and achievement target completion rate.

Name	Attendance (%)	Phone volume	Project objectives	Actual completion rate (%)
Peng Yuyan	99	80	Forty thousand	80
Chen Daoming	100	89	Sixty thousand	90
Hu Ge	96	119	Fifty thousand	120
William chan	100	146	Eighty thousand	109

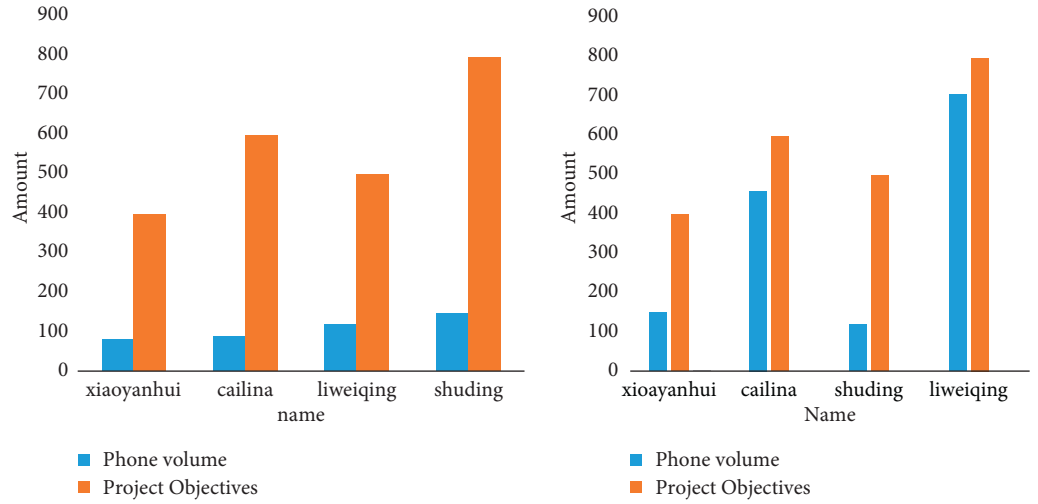


Figure 6: Comparative analysis of employees' performance before and after half a year.

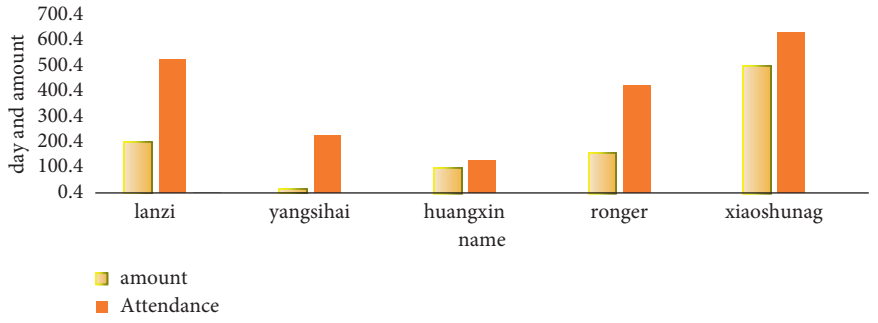


Figure 7: Work attendance of the Administration Department and the benefits generated.

managers' lack of understanding of employee performance management can easily lead to miscommunication between managers and ordinary employees. This will deprive the company of the opportunity to learn management and is not conducive to the development of the company.

4. Application of OA System in Performance Appraisal

4.1. OA System Analysis. Performance appraisal information must not only be connected to a computer to complete diagnostic calculations, nor is it a simple process of traditional diagnostic tools, but must reflect the performance appraisal level, strengthen daily performance collection, and use online tools to announce performance progress. The goal is to improve the transparency of evaluation, promote open, fair, and just evaluation, integrate the performance evaluation information system into the existing OA information management system, and effectively manage and quantify the use of scientific and modern technology. Figure 10 shows the performance comparison between the post-1980s and post-1990s and the post-1970s and zero-zero generations in recent years.

The second stage: OA system performance management is combined with all the company's goals in the execution system, plan execution system, daily operation, and various business systems so that automation is introduced into the system instead of becoming an independent distributed accounting system. At this time, performance management data and scores should be automatically generated, and performance should be presented efficiently and effectively.

Figure 8: The company's earnings in the past nine years.

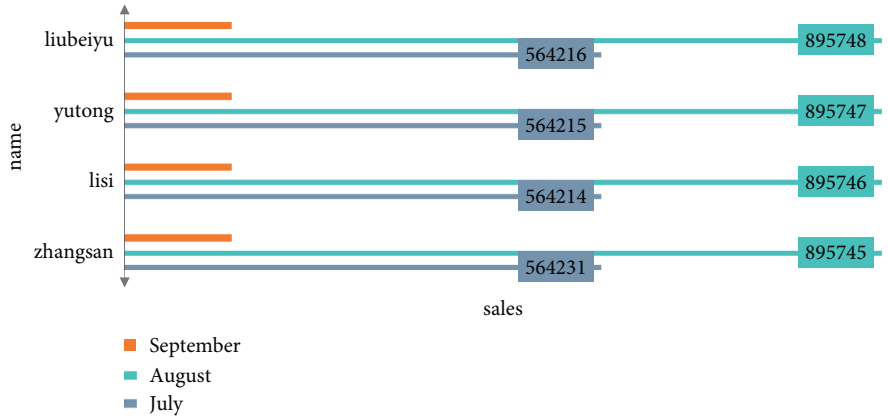


Figure 9: The benefits of employees to the company are detailed to the month.

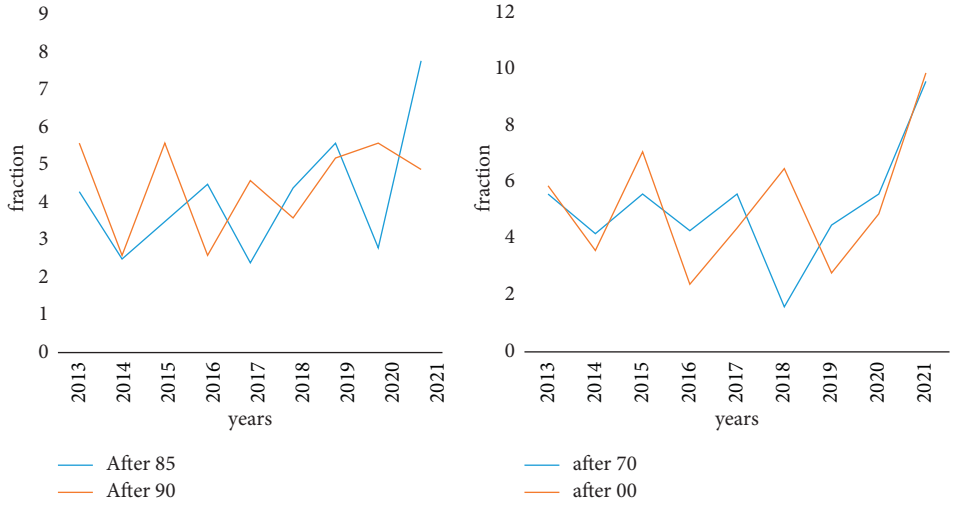


Figure 10: In recent years, the performance appraisal competitions of the post-1980s and post-1990s, and the post-1970s and post-2000s, encourage each other. The first stage: OA system performance management exists as an independent module, and relevant data can be displayed for scoring and evaluation.

Table 5: Project details and completion rate.

Serial number	Table name	Probability
1	Employee information form	100%
2	Project information Sheet	95% (Part of the information to be verified)
3	Attendance information form	99.3%
4	Log information table	89%

The best stage: OA system performance management can help realize the amoeba management style; that is, each subject can achieve organizational goals through self-motivation, self-efficacy, and organizational goals, as shown in Table 5.

The transaction volume of the project needs to change from quantity to qualitative change and reserve enough quantity, which requires employees to exert their own internal drive.

4.2. Value Analysis of Performance Appraisal Results. In terms of performance improvement, starting with performance appraisal results can improve performance appraisal results; in terms of rewards and punishments, performance appraisal results can be used as the basis for target bonus distribution, senior selection, and final elimination. In addition, for recruitment and dismissal, the results of a performance appraisal must be thoroughly analyzed. The same standard, first of all, is whether the best employees can be promoted and join them to judge their potential together. Newly recruited employees are arranged to require data analysis based on the comprehensive evaluation results of training and development. High-quality employees may become potential targets and top seeds for environmentally focused training and development. Moreover, the performance appraisal results can only reflect the unilateral ability of the employee and cannot explain the overall ability of the employee in many aspects.

5. Discussion

All employees are the executors of performance appraisal. Only when all employees fully participate in the implementation of performance appraisal can they combine personal career development planning with work goals and improve personal creativity and workability. By formulating performance appraisal standards, it is possible to develop high-level and lower-level interactions, allowing employees to fully participate. Based on extensive feedback hearings, department employees formulate performance appraisal standards and perform work analysis below to ensure that the organization's goals are achieved. Employee recognition and participation are increased by exchanging information and opinions. No matter how scientific and complicated the performance appraisal system is, if employees do not accept it, it will be put aside and will not work.

In the context of big data, innovative human resource management work in innovative companies is conducive to the development of employee equity and corporate benefits. Because the development of good performance and excellent business is inseparable from efficient and professional system management, and performance is one of the issues that every business person is very concerned about. Only by satisfying the special interests of employees and ensuring the fairness of business management and benefit distribution can employees devote more time and energy to the company, create a sense of company identity, and improve the management level of the company's employees.

6. Conclusion

In the aspect of performance feedback, the performance results are applied in many aspects to the recruitment, training, compensation, and labor relations management of human resources. A competency model is established, staff career planning is designed, and so on. Make human resource management more systematic and comprehensive. The design of the system should be simple and easy to implement. The testing requirements for general functions and other general functions can be indicated by thick lines. In principle, other points systems can be used without considering any errors or omissions to ensure eligibility. However, as an important performance indicator, it should be evaluated according to the importance of the component, and the value of the audit point should be set reasonably to show the focus of the evaluation. In addition, the human resources performance management information system has powerful computing and processing capabilities, replacing the manual data processing mode of the original performance team, which not only greatly reduces the work pressure of the performance team of the human resources department but also avoids human interference factors and reduces the error rate. It has been lowered to a lower level, making the assessment results more accurate and objective. The limitations of this study are reflected in the technical limitations, and the study is particularly suitable for development in the decision-making field. The article does not discuss the cost of building a performance management information system based on big data technology, which is divorced from the reality of the enterprise to a certain extent, and further research is needed on the cost and future benefits of the performance management information system.

Data Availability

No data were used to support this study.

Disclosure

The authors confirm that the content of the manuscript has not been published or submitted for publication elsewhere.

Conflicts of Interest

There are no potential conflicts of interest in our paper.

Authors' Contributions

All authors have seen the manuscript and approved to submit it to your journal.

Acknowledgments

This work was supported by the Social Science Foundation of Hunan Province of China (Grant no. 18YBA341).

References

[1] S. Liu, "Human resource management of Internet enterprises based on big data mobile information system," *Mobile Information Systems*, vol. 2021, no. 5, pp. 1–9, 2021.

[2] H. Banimelhem, H. M. Abu Elanain, and M. Hussain, "Impact of human resource management practices on employees' turnover intention in United Arab Emirates (UAE) health care services," *International Journal of Information Systems in the Service Sector*, vol. 10, no. 4, pp. 21–41, 2018.

[3] R. H. Hamilton, W. A. Sodeman, and C. M. Dalton, "The questions we ask: opportunities and challenges for using big data analytics to strategically manage human capital resources," *Business Horizons*, vol. 63, no. 1, pp. 85–95, 2020.

[4] N. Vargas, M. B. Lloria, and S. Roig-Dobon, "Main drivers of human capital, learning and performance," *The Journal of Technology Transfer*, vol. 41, no. 5, pp. 961–978, 2016.

[5] C. Brewster, P. N. Gooderham, and W. Mayrhofer, "Human resource management: the promise, the performance, the consequences," *Journal of Organizational Effectiveness: People and Performance*, vol. 3, no. 2, pp. 181–190, 2016.

[6] F. Otoo, "Human resource management (HRM) practices and organizational performance," *Employee Relations*, vol. 41, no. 5, pp. 949–970, 2019.

[7] M. Robert, P. Giuliani, A. Guilloton, and M Khallouk, "Management innovation: a dynamic analysis of the implementation phase over time," *Production Planning & Control*, vol. 30, no. 15, pp. 1219–1238, 2019.

[8] R. Z. Szabo and R. Csontos, "Efficient organisational renewal: the role of technological and management innovation," *International Journal of Business Management & Research*, vol. 9, no. 1, pp. 9–24, 2019.

[9] H. Shipton, P. Budhwar, P. Sparrow, and A Brown, "Human resource management, innovation and performance || human resource management and innovative behaviour: considering interactive," *Informal Learning Activities*, vol. 3, pp. 32–47, 2016.

[10] G. Zhai, Y. Yang, H. Wang, and S. Du, "Multi-attention fusion modeling for sentiment analysis of educational big data," *Big Data Mining and Analytics*, vol. 3, no. 4, pp. 311–319, 2020.

[11] S. B. Tsai, C. Y. Huang, C. K. Wang et al., "Using a mixed model to evaluate job satisfaction in high-tech industries," *PLoS One*, vol. 11, no. 5, Article ID e0154071, 2016.

[12] J. Zhang, J. S. Wang, W. Du, and X. H. Wang, "Big data analysis reveals the truth of lumbar fusion: gender differences," *The Spine Journal*, vol. 17, no. 5, pp. 754-755, 2017.

[13] S. Roy, D. Sarkar, and D. De, "Entropy-aware ambient IoT analytics on humanized music information fusion," *Journal of Ambient Intelligence and Humanized Computing*, vol. 11, no. 1, pp. 151–171, 2020.

[14] C. Esposito, A. Ca Stiglione, F. Palmieri et al., "Event-based sensor data exchange and fusion in the Internet of Things environments," *Journal of Parallel and Distributed Computing*, vol. 118, no. 2, pp. 328–343, 2018.

[15] T. J. Rato and M. S. Reis, "Optimal fusion of industrial data streams with different granularities," *Computers & Chemical Engineering*, vol. 130, no. 2, pp. 106564–106564.16, 2019.

[16] S. B. Tsai and K. Wang, "Using a novel method to evaluate the performance of human resources in green logistics enterprises," *Ecological Chemistry and Engineering S*, vol. 26, no. 4, pp. 629–640, 2019.

[17] P. Basanta-Val and L. Sánchez-Fernández, "Big-BOE: fusing Spanish official gazette with big data technology," *Big Data*, vol. 6, no. 2, pp. 124–138, 2018.

[18] G. Xiong, F. Zhu, X. Dong, F. Haisheng, H. Bin, and T. Teng, "A kind of novel ITS based on space-air-ground big-data," *IEEE Intelligent Transportation Systems Magazine*, vol. 8, no. 1, pp. 10–22, 2016.

[19] A. Daniel, K. Subburathinam, A. Paul, N. Rajkumar, and S. Rho, "Big autonomous vehicular data classifications: towards procuring intelligence in ITS," *Vehicular Communications*, vol. 9, pp. 306–312, 2017.

[20] L. Zhang, "Design of a sports culture data fusion system based on a data mining algorithm," *Personal and Ubiquitous Computing*, vol. 24, no. 1, pp. 75–86, 2020.

[21] W. Zhuo, Z. He, M. Zheng, B. Hu, and R. Wang, "Research on personalized image retrieval technology of video stream big data management model," *Multimedia Tools and Applications*, vol. 21, no. 2, pp. 1–18, 2021.

[22] S. B. Tsai, W. Wu, S. Ma, C. H. Wu, and B. Zhou, "Benchmarking, knowledge inertia, and knowledge performance in different network structures," *Enterprise Information Systems*, vol. 14, no. 5, pp. 641–660, 2019.

[23] E. E. Papalexakis, C. Faloutsos, and N. D. Sidiropoulos, "Tensors for data mining and data fusion: models, applications, and scalable algorithms," *ACM Transactions on Intelligent Systems and Technology*, vol. 8, no. 2, pp. 16.1–44, 2017.

[24] R. Patan and M. R. Ba Bu, "A novel performance aware real-time data handling for big data platforms on Lambda architecture," *International Journal of Computer Aided Engineering and Technology*, vol. 10, no. 4, pp. 418–430, 2018.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 9054209, 9 pages
https://doi.org/10.1155/2022/9054209

Research Article

Data-Driven Intelligent Risk System in the Process of Financial Audit

Tianheng Xie and Jianfang Zhang

Commercial College, Nantong Institute of Technology, Nantong 226001, Jiangsu, China

Correspondence should be addressed to Tianheng Xie; xietianheng@126.com

Received 12 January 2022; Revised 21 February 2022; Accepted 21 March 2022; Published 23 April 2022

Academic Editor: Xiangtao Li

Financial auditing is an important part of government control, and monitoring the risks of the financial system helps maintain the stability of a country's financial system. This paper aims to study the operation of data-driven intelligent risk analysis system in the process of financial auditing. When the risk level of the financial system reaches a certain warning point, it needs to be adjusted, conduct strict audits on high-risk institutions, strengthen cooperation with financial audit institutions, and make them more systematic. Risk control within a reasonable range will neither have a significant impact on the financial system nor increase the impact on the industry. This paper proposes to integrate the internal and establish an internal control system for the financial system.. If the internal control of the financial system cannot be sustained, financial innovation will cause great damage to financial institutions if the compensation is not good. Especially in terms of access strategy, it has made a strong contribution to economic innovation. The experimental data in this paper show that more than 25% of data-driven intelligent risk analysis systems have begun to be used in the process of financial auditing. At the same time, its utilization rate is significantly increasing with the development of technology. This paper introduces the relevant content of financial auditing, analyzes the auditing systems in different fields, integrates resources, taps on multisectoral needs, establishes a cross-departmental governance system, and improves financial stability and the efficiency of a wide range of social services.

1. Introduction

The development and deep integration of information technology, artificial intelligence, big data, and various industries is inseparable from the rapid development of network information technology. How to effectively integrate data analysis into the control project and how to manage an excellent big data analysis team and effectively empower big data analysis management and control project services, improve management and control efficiency, and provide more questions and control signs that the application management process faces are the main challenges. Big data point analysis is like a "double-edged sword." If used well, it will get twice the result with half the effort. Improper use will bring about the negative impact of progress and audit results on the entire project. How to effectively avoid and control these risks, reduce the possibility of danger, and increase the frequency of performing effective database checks are the problems that need to be solved at present. In the context of the integration of capital and data technology, the empowerment of blockchain, cloud computing, artificial intelligence, and other technologies is based on value guidelines and standards. The contribution of digital solutions and intelligence to improve compliance and performance governance is collectively referred to as the regulatory technology or compliance in technology. Regulatory technologies used in the field of supervision are called regulatory technologies, and those used in financial institutions are called compliance technologies.

Auditing financial innovation is an inexhaustible driving force for the development of financial institutions. It can help financial institutions grow better through innovation and meet the financial growth needs of financial services. New dangers are inevitable. We must pay full attention to

new dangers and take necessary precautions. This is the basic requirement of the "Basic Principles of Effective Banking Supervision." Rural credit cooperatives need to build a complete business innovation management system, strengthen risk management and control, combine financial innovation with hedging, and standardize the operating behavior of rural credit cooperatives. The contradiction between innovation and risk is externally integrated into economic growth and integrated into the comparison of benefits brought about by innovation and innovation. In the past, financial audits usually focused on the traditional error control and fraud investigations of individual financial institutions. I still do not understand the overall situation of the previous research. This paper starts with the factors that may affect the risk of the financial system. And after empirical research and statistical analysis, the relevant parameters are estimated, and hypotheses are tested. This will determine the internal control of liquidity, the proportion of nonperforming assets, the proportion of nonperforming assets in financial institutions, and which debt ratio indicators have a significant impact on systemic risks. A systematic description of the impact mechanism of these indicators and finally how financial control should deal with finances will be summarized.

For this research topic, many outstanding experts and scholars at home and abroad have discussed it. From the perspective of sociolinguistics, Chen et al.'s article aims to test whether the honorific and real names used by Chinese auditors to address clients in audit reports imply the risk of differential financial misstatement. Specifically, the author assumes that auditors use honorifics to indicate that they have a lower social status relative to the client, which leads to the loss of auditor independence, lower audit quality, and higher risk of financial misstatement. The authors used a data sample of manually coded names from the audit reports of Chinese listed companies from 2003 to 2012 to conduct research. The authors found that the financial misstatement of a company called a respectable name is much greater in terms of likelihood and scale than a company called a real name. In addition, compared with the auditor's consistent use of honorifics, the use of casual honorifics has a stronger positive correlation with misstatements. The author further shows that when the accounting firm is one of the top ten accounting firms in China, the positive correlation between the usage of honorifics and the risk of misreporting by clients weakens [1]. Lee and Izbicki mentioned that a key issue of modern statistics is how to make quick and reliable inferences on complex high-dimensional data. Although people are very interested in sparse technology, current methods cannot be extended to data with nonlinear structures. In this work, they proposed an orthogonal sequence estimator for predicting complex aggregated objects, such as natural images, galaxy spectra, trajectories, and movies. Our series of methods link the ideas of nuclear machine learning and Fourier methods. They extend the unknown regression of the data according to the characteristic function of the kernel-based operator, and we use the orthogonality of the basis relative to the basic data distribution P to speed up the calculation and parameter adjustment. If the kernel is selected appropriately, the feature function will adapt to the inherent geometry and dimensions of the data. They provide theoretical guarantees for radial kernels with varying bandwidths, and they relate the smoothness of the regression function with respect to P and the sparsity in the feature base [2]. Garg et al. mentioned in their article that, in the past few years, we have witnessed exponential growth in the computing and storage capabilities of smart devices, which has led to the popularization of an emerging technology called edge computing. Compared with traditional cloud computing-based infrastructures, edge computing end users can use computing and storage facilities nearby. In addition, with the widespread popularity of unmanned aerial vehicles (UAV), a large amount of information will be shared between edge devices and unmanned aerial vehicles in the next few years. In this case, traffic monitoring using drones and edge computing devices is expected to become an integral part of the next generation of intelligent transportation systems. However, monitoring requires uninterrupted data sharing, collaborative decision-making, and stable network formation. Edge computing supports data processing and analysis closer to the deployed machine (i.e., data source). Rather than simply storing data and missing opportunities to exploit it, edge devices can analyze the data to gain insights before acting on it. The use of drones can facilitate the transmission of data from the vehicle to the edge for real-time analysis, and the drone can act as an intermediate air node between the vehicle and the edge node [3].

The above three scholars have a comprehensive understanding of the financial audit and data-driven intelligent risk analysis chart system, but they did not consider the two well. These are some of the shortcomings of the above article. Therefore, in this paper, these issues will be supplemented and studied in-depth, with a view to better applying the data-driven intelligent risk analysis system to the financial audit process. At the same time, this paper also looks forward to the future development of this field.

2. Method of Data-Driven Intelligent Risk Analysis System in the Process of Financial Audit

2.1. Financial Audit. Audit big data analysis is an advanced technology [4], and the demand for data analysts is quite high. It not only requires other types of interdisciplinary knowledge and experience of mathematics, financial, and accounting graduate students but also often uses unknown and unusual technical means and so on. The auditor should be based on project history and personnel, configure requirements, budgets, schedules, and so on and analyze big data, identify process risks in advance and take control measures, and check whether there are potential hazards during the inspection process to avoid losses and consequences caused by poor risk control. The current financial audit needs to solve the limitations of traditional research and rectification and grasp the risks from a macroperspective grade. Therefore, macro-indicators of financial institutions are needed to predict the overall level of systemic risk to determine deposits. All

institutions have a remindering about the risk; if it is within the safety range, the problems arising during the investigation process will be properly resolved [5]. Take care of it and prevent it from expanding further; in the final analysis, buildings must be renovated, and districts must be rebuilt from scratch. And the unit fills up the gap, perfects the system, strengthens internal control, and streamlines management. Financial audit research can effectively guarantee the security of the national financial system.

2.2. Data-Driven Intelligent Risk Analysis System. Numerical examples and data-driven methods are similar in many behaviors [6]. But the purpose of planning is different. The reason is that the idea behind this is different, even with data management. But the digital example also emphasizes the preset logic. The logic of form creation must be written by the architect. Although the logic of solving problems in a data-driven approach is programming, a data-driven approach is like a documented design based on the relationships found in analysis and processing. Instead of creating is a logic formula 11 first, and then using that logic to solve the problem, start from a large amount of available information. Enter the data and find out the logic to solve the problem. Then use this logic as a solution. The data-driven approach not only sprouts in the architectural design industry but also has a lot of practice in the field of data-intensive urban planning. Roads, road networks, pedestrians, traffic, and so on in cities naturally have big data characteristics due to the high density of cities. In our country, traditional urban planning is often based on the judgment of the government and experts [7]. This top-down planning method cannot take into account the actual needs from multiple subjects, and the technical means of big data are expected to solve such problems and achieve the fullest use of urban resources. There are many precedents for such data-based decision-making abroad. For example, the research group at Brown University uses big data analysis to determine the best construction location for engineering facilities and can find construction sites that can meet the full utilization of resources through data such as the flow of people on campus and the distribution of main functions of the campus is student and school results. Big data analysis based on history can also help to discover specific laws or construction risks, which helps avoid traps for new projects. Information management functions [8, 9] include customer management, space management, collateral management, warehouse management, quota management, early warning management, information release, ledger management, query, report platform, system management, job transfer, and data maintenance. The credit business management system specifically includes the following information management functions and customer management. Loan customer creation and information maintenance include the creation of new customers in the system, the design and implementation of a financial risk early warning analysis system based on Zhajishi data statistics, inquiries about existing customer information, basic information, and financial information, as well as changes and maintenance manufacturing. The content of customer management can provide business personnel and business managers with detailed and timely customer information and key data. Available client types include corporate client management, corporate blacklist management, corporate senior client management, group-related management, personal client management, personal blacklist management, personal senior client management, joint insurance team management, joint client management, and financial reporting [10].

2.3. Algorithms for Data-Driven Intelligent Risk Analysis in the Financial Audit Process. In the process of financial audit, when the financial risk early warning analysis system is in the automatic operation mode [11], the system will execute according to the following definition list, as shown in formulas (1) and (2):

$$u = \sum_{i}^{i-1} (n - i) + 1, \tag{1}$$

$$k = \sum_{i}^{k-1} (n - i) + 1 - (x, y). \tag{2}$$

According to its resource library [12], it can be concluded that a properly configured data algorithm is as follows:

$$t = \frac{n(n-1)}{2} + \frac{(x-1)}{(y-1)}. \tag{3}$$

Through the system materials of the resource database and the quantitative relationship between the indicators defined by the above model [13], the preliminary algorithm of the audit process can be obtained:

$$\begin{aligned} F\left(a_x, a_y\right) &= \vee F_k\left(a_1 - a_2\right) = 1, \\ A &= \left\{a_1 \mid F_2\left(a_x, a_y\right) = 1\right\}. \end{aligned} \tag{4}$$

According to the above formula, combining formulas (1) and (3), we can get

$$\begin{aligned} z^2(k) &= h^2[x(k)] + w^j(k), \\ E\left[w^i(k) w^2(j)^1\right] &= R^i \ell(k - j). \end{aligned} \tag{5}$$

The early warning model of financial audit risk is stored in a standard format file [14, 15], which contains the decision tree structure of the model and the data definition of each node.

$$x(k \mid k = 1) = \theta(k-1) x^2(k-1 \mid k-1). \tag{6}$$

Using the decision information data, data processing and conversion processing are performed, and the source data are converted into model result data. Thus, we can obtain formulas (7) and (8):

$$S(k) = H^1(k) P(k \mid (k-1)) H^2(k)^1 + R(k)^2, \tag{7}$$

$$x(k \mid k) = E\left(x(k) \mid Z^i\right) = \sum_{j=0}^{m\left(k^i\right)} \beta\left(k^i\right) x^2. \tag{8}$$

Related indicators and corresponding safety thresholds [16] can also be calculated by the above method:

$$p^2(k \mid k) = P^2(k \mid l)|Z^{2,k} = \sum_{i=0}^{m^2 k} \beta_j\left(x_o^i, k\right), \tag{9}$$

$$\sum_{j=1}^{r} \eta_k(x_1, y_1) = 1 - \sum_{j=2} \nu(j, k_0). \tag{10}$$

3. Data-Driven Intelligent Risk Analysis System Experiment in the Process of Financial Audit

When financial auditing and fiscal supervision policies or related laws fail, especially when local financial regulations encourage financial institutions to take high returns with high risks, the possibility of systemic financial risks will increase. This should arouse the attention of my country's macroprudential departments and incorporate the evaluation results of the effectiveness of financial policies and financial regulatory laws into the content of macroprudential supervision. To ensure that relevant policies are implemented in place, do not pose a hidden danger to the security and stability of the financial system [2, 17]. In addition, fluctuations in monetary and fiscal regulatory policies will also have a negative impact on the stability of the financial system, especially when these policy objectives are different. For example, the overall goal of my country's central bank's monetary policy is to maintain the relative stability of the RMB currency in order to stimulate the growth of the national economy. The purpose of policy-based fiscal supervision is mainly to focus on the overall smooth operation of financial institutions and protect the legitimate rights and interests of investors. The difference in the policy objectives of the central bank and other financial regulatory agencies makes it difficult to coordinate their respective policies effectively, which easily threatens the security of the financial system and even the operation of the entire macroeconomic system. Therefore, national macroprudential regulators should pay attention to and evaluate the overall stability of the financial system and financial institutions (including banks and nonbank financial institutions), as well as systematic risk management methods and micromanagement vigilance. These changes assess their impact on systemic risk.

3.1. Experimental Strategies and Experimental Procedures. The experimental process of this paper is to determine the experimental objects first, then collect data samples, then process and analyze the data, and finally get the results. Many experts and scholars analyzed the necessity of coordination between government fiscal planning and fiscal supervision and suggested that government fiscal regulation should give full play to its independence and comprehensive benefits [18], and fiscal supervision agencies need to reintegrate existing responsibilities. Management and contract risk monitoring functions establish a comprehensive coordination framework for government financial planning and auditing. The government financial planning emphasizes the supervision function, and the financial supervision emphasizes the management function. Government fiscal planning must maintain its unique independence, objectivity, and impartiality and expose obvious risks of hidden dangers to the stability of the financial system: in terms of responsibilities, government fiscal control supervises its supervision and supervision and punishment. Financial control does not participate in the regulatory law. The business activities of audited financial institutions have a transcendence status and support their performance of supervisory functions; financial accounting transactions are adjusted in accordance with the overall state of the national economic development. Therefore, financial management and control should pay more attention to the overall understanding of the financial system and its working mechanism and grasp the stability from the source; in terms of goals, the government emphasizes financial control, systemic risks, and financial supervision. Financial audit should focus on the comprehensive external performance evaluation of finance, internal reasons, related control management procedures and risk capabilities, timely and effective analysis, and early warning of systemic risks; in terms of objectives, government auditing requires new government supervision, and the regulatory agency designates direct control. Intensify the accounting of the performance of the audit institutions' related responsibilities, especially the control of the financial responsibilities of the main persons in charge of the audit institutions. The financial audit risk management of banks is mostly nonempirical analysis methods and subjective estimation. The overall level of using information technology and expert systems to strengthen management and prevent risks is not high. As the process of global economic integration accelerates [19] and market competition is becoming increasingly international and fierce, commercial banks urgently need to strengthen their own professional knowledge and core capacity building, improve their ability to predict and respond to market risks, and improve their profitability. At the WTO level, it is to improve the competitiveness of commercial banks after joining the WTO.

3.2. Sample Collection. The model represents the linear relationship between each independent variable and the dependent variable (system risk). The difference is the internal control state of the independent variable, which has a linear relationship with the dependent variable [20, 21]. Because of the data given in the model, the internal control has a larger value than the dependent variable and other independent variables, so it is determined. Use modular division for linear relationships. Discuss two situations of type and model. The next step will be to test which of these two models is more suitable for this study to make a choice. Determine whether the coefficient is significantly increased due to the input of the estimated value according to the test judgment, and finally determine the significance level. If the test significance level is high, accept the null hypothesis and accept that there is no nonlinear feature. Otherwise, we should try to introduce nonlinear variables. After logging in, if it fails, please

continue to check if there is a better model that can replace the original model. This requires adding nonlinear variables from the calculation results. For example, perform the same operation under three conditions, and finally get the corresponding data theory [22]. In the regression analysis, the difference between the measured value and the predicted value of the regression equation is shown. The difference obeys normal distribution. But it will be called the rest of the model. Follow the standard normal distribution. The probability is that the rest of the standard of the test point is out of range. If the standard balance of a given test point is outside this range, it is regarded as an abnormal test below the confidence level and is not included in the regression line adjustment. Artificial balance is the difference between actual observations and regression estimates.

4. Data Analysis of Experimental Results

4.1. Summary of Experimental Results. At present, in the asset-liability structure of my country's financial institutions, especially commercial banks, the above-mentioned liabilities are all absorbed by deposits, and the asset business is also dominated by loans. The proportion of total assets exceeds total assets, and interest income is higher than total income. The spread reflects the typical characteristics of traditional banks. In financial auditing, you can pay more attention to assets and liabilities than the capital adequacy ratio so as to avoid excessive debt ratios causing excessive financial burdens and agency costs to financial institutions [23]. Taking into account the inevitable high debt ratio of banks, insurance financial institutions should pay more attention to the risk factors of asset-liability ratio. Taking this insurance financial institution as an example, its asset-liability ratio is generally slightly higher than that of normal business, but its operating activities and asset ratio are closer to normal business than banks. Therefore, in the financial audit, the asset-liability ratio of insurance financial institutions may be very high relative to the debt level, which may cause systemic risks. For banks, the focus should be on short-term liquidity levels. The risk early warning model is based on the business database and is divided into information entry, model entry, and source data entry. The risk warning model is stored in a standard format file, which contains the decision tree structure of the model and the data description of each node. The source data is the indicator data source in the model calculated through the access rules in the risk assessment system of financial institutions [24, 25]. The user queries and maintains warranty-related information in the system and establishes the relationship between warranty information and confirmation. Carry out customer evaluation; the user enters the customer evaluation information into the system, performs customer credit evaluation, and completes the credit evaluation form. General credit line application users enter individual customer credit line survey information in the system and fill in and submit a single customer general credit line survey application form. The loan job application user enters the job application information into the system, evaluates the collateral according to the work status [26], and fills in the loan job application form. Edit group associations, that is, users query group information in the system, add, modify, delete, and approve credit group associations.

Connect the information of popular companies to the bank's customer management system. Taking the financial industry as an example, the customer management system of local commercial banks has complete and accurate financial business operations, finance, and related company data [27]. The relevant H department of the company will be adjusted downwards. The internal evaluation process of the bank's credit and risk department can translate the core business management and financial impact of a financial company into a downgrade of the company, an increase in the provision for bad debts, and an increase in risk investment; the bank's asset-liability management process can reduce risks and increase funds. Rising costs have led to flat estimates of bank revenue, asset value, and liquidity. Banks that have completed the microstress test can use the above system to measure the impact of a decline in the company's financial and governance ratios on their forecasts. Capital adequacy ratio and liquidity risk extend this method to many large banks. The stability of Chinese financial system is measured through the durability test of multiple banks and the aggregation of macro indicators.

4.2. Experimental Data Analysis. Auditors should predetermine the risks inherent in the big data analysis process according to the background, personnel, budget, and time requirements of the specific project, control the potential risks in the audit implementation process, and avoid insufficient risk control and in situ effects. Regarding how to effectively carry out project risk control, the author believes that four aspects should be focused on: one is to focus on the refined management of the project and to do a good job in project feasibility studies; the second is to pay attention to the quality requirements; the third is the organization and employee coordination of development; the fourth is to organize the budget and progress of the completion of the project. Will the on-site feasibility study play an important role in all project requirements? During the management development process and in the data analysis before starting the project, the auditor should identify the factors that may affect the progress and results of the project during the audit project review, as well as the information system and business information audited by the audited unit. The plan requires personnel requirements and so on to check the impact of project data analysis results and to check decision-making to avoid half of the losses caused by project failure.

4.3. Analysis of Data-Driven Intelligent System in the Process of Financial Audit. In the financial audit process, many agency processes need to cooperate with each other to complete a complete job, and the specific collaboration process is shown in Figure 1.

From the perspective of financial auditing, there are differences in monthly data analysis. Through Figure 2, we can see the cause of the equipment data error problem and analyze it.

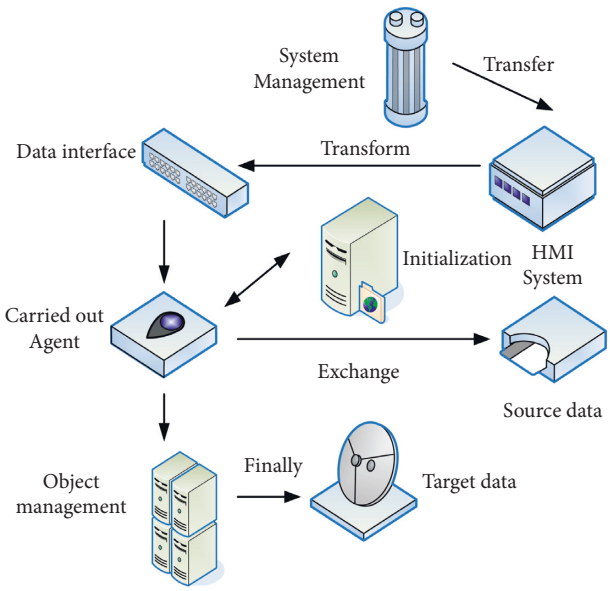


Figure 1: Flowchart of collaboration between multiple agents.

It can be seen from Table 1 that, through data processing, especially after intelligent processing, a high-precision fitting function of the speed can be obtained, and the results of the data fitting are as follows.

The fluctuation error of the same month's data (net consumption, etc.) exceeds 10%. A fluctuation error greater than 10% can easily cause the data link break strength index to fluctuate throughout the manufacturing process. Correspondingly, the fluctuation of the strong force index will increase the probability of the whole process being interrupted, which also means that there are many defects in the data statistics process, as shown in Figure 3.

Analyze the change trend between the data, and form the convergence trend shown in Table 2.

The time sequence of the financial risk early warning function based on data statistics of the samples involved in the experiment is shown in Figure 4. Through the sequence diagram, we can see that the arrangement of the device function of risk early warning is very necessary.

Certain related variables are sometimes involved in the model. Here, we will make specific explanations for some of these variables in order to better understand and analyze the conclusions brought about by the data, as shown in Table 3.

The relevant indicators of the risk warning function and the corresponding safety thresholds are listed in Table 4.

Different levels of the data layer will also produce different data classification results, as shown in Figure 5.

Preprocess and store structured data such as customer information and business data through technologies such as collection, cleaning, verification, conversion, and classification, and link all data to form a unified data view, as shown in Figure 6.

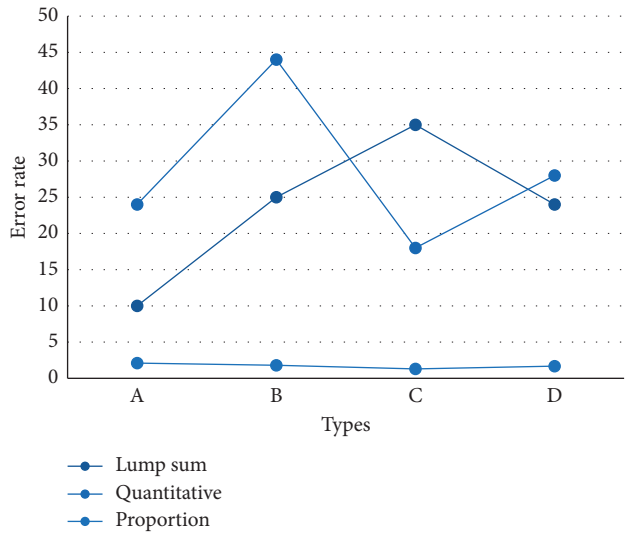


Figure 2: Data error analysis.

Table 1: Data fitting results.

Machinery code	Variety name	Rotating speed	Effectiveness	Total time
A025	001	520	87	31
A026	002	490	99	40
A027	003	610	106	54

Data are documents, visited magazines, emails, images, audio, video, and so on. Extract unstructured data such as natural language processing technology and machine learning technology. This can be seen in Table 5.

Table 2: Convergence curve table.

Number of iterations	Total actual consumption	Net usage	Net consumption	Back to spend
100	5	56	45	32
200	9	78	69	44
300	17	64	67	52

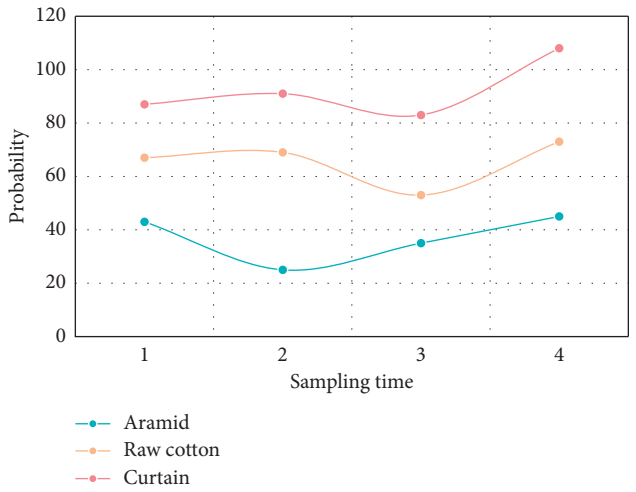


Figure 3: Data fragmentation impact.

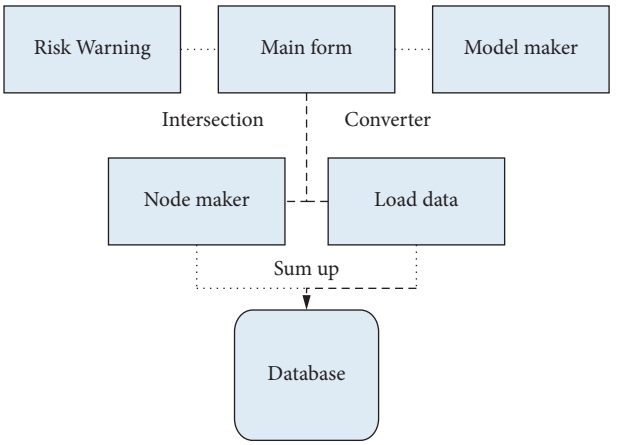


Figure 4: Timing diagram of risk warning function.

Table 3: Variable interpretation table.

Variable name	Nature	Significance
Risk	Dependent variable	Risk measurement
Liq	Dependent variable	Liquidity level
Inc	Independent variable	Debt level
Debt	Random variables	Proportion of unhealthy assets

Table 4: Regulatory index table.

Indicator type	First level indicator	Secondary indicators	Index value (%)	Risk level
Liquidity risk	Liquidity indicators	Debt dependence	25	6
Credit risk	Liquidity gap rate	Profit margin	30	5
Market risk	Nonperforming asset ratio	Cost to income ratio	33	3
Profitability	Correlation	Credit ratio	28	1

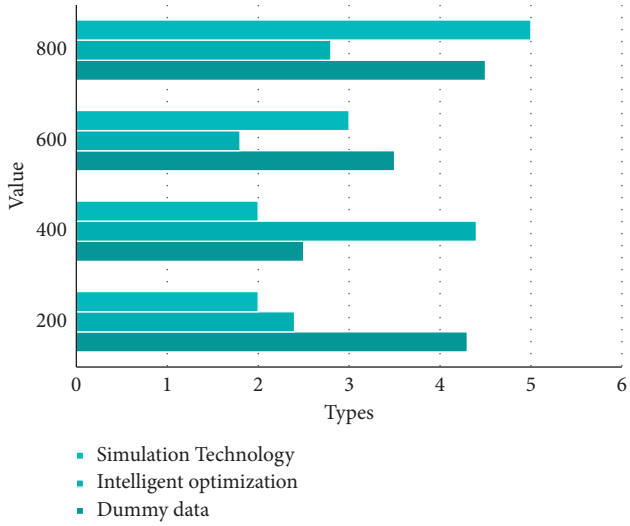


Figure 5: Schematic diagram of data tank problem.

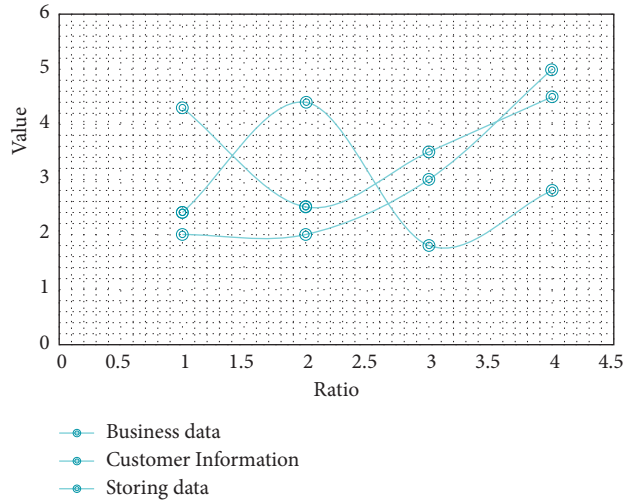


Figure 6: Data view.

Table 5: Structured data table.

Types	Skills	Ratio	Rank
Documentation	Natural language	1.2	A
Magazine	Structured data	2.01	A
Image	Machine learning	1.6	B

5. Conclusions

The calculation of the relationship between financial controlled entities based on multiple data sources mainly includes the calculation of the time weight of the relevant attributes of the entities. Calculate the weight of the data source and calculate the number of queries and each user at the same time for this application scenario. Because of each source, the calculation of the weight of the data source is based on the PageRank algorithm, using other data sources related to the data source to share the weight of the data source, which can adapt to the dimension, and the information is obtained through multiparty voting. It can better solve the problem of uneven data quality. Our country's fiscal treatment can conceal the design flaws of the United States, learn the new financial governance methods of the United States, and create efficient and diversified financial supervision methods suitable for my country's national conditions. It covers different areas of systems, resource integration, and multidepartmental requirements and establishes a cross-sectoral governance system to improve financial stability and the efficiency of a wide range of services to society.

Data Availability

No data were used to support this study.

Conflicts of Interest

The authors declare no conflicts of interest in this paper.

Authors' Contributions

All authors have seen the manuscript and approved to submit it to this journal.

Acknowledgments

This work was supported by the Training Program for Young and Middle-Aged Scientific Research Core Teacher of Nantong Institute of Technology: Humanities and Social Science Foundation (ZQNGG4080). Jiangsu Provincial 14th Five-year Business Administration Key Construction Discipline Project (SJYH2022-2/285); Training plan for young and middle-aged scientific research backbone of Nantong Institute of Technology: Humanities and Social Sciences Fund (ZQNGG4080).

References

[1] F. Chen, X. Du, S. Lai, and M. L. Z. Ma, "Does the use of honorific appellations in audit reports connote higher financial misstatement risk? Evidence from China," *Asian Review of Accounting*, vol. 26, no. 4, 2018.

[2] A. B. Lee and R. Izbicki, "A spectral series approach to high-dimensional nonparametric regression," *Electronic Journal of Statistics*, vol. 10, no. 1, pp. 423–463, 2016.

[3] S. Garg, A. Singh, S. Batra, N. Kumar, and L. T. Yang, "UAV-empowered edge computing environment for cyber-threat detection in smart vehicles," *IEEE Network*, vol. 32, no. 3, pp. 42–51, 2018.

[4] E. I. Vlahogianni, K. Kepaptsoglou, V. Tsetsos, and M. G. Karlaftis, "A real-time parking prediction system for smart cities," *Journal of Intelligent Transportation Systems*, vol. 20, no. 2, pp. 192–204, 2016.

[5] J. Wonglimpiyarat, "What is it about strategic implications of using financial models in the process of technology management?" *The Journal of High Technology Management Research*, vol. 30, no. 1, pp. 82–90, 2019.

[6] N. Demong, J. Lu, and F. K. Hussain, "An adaptive personalized property investment risk analysis method based on data-driven approach," *International Journal of Information Technology and Decision Making*, vol. 20, no. 2, pp. 1–36, 2021.

[7] G. Barta, "The increasing role of it auditors in financial audit: risks and intelligent answers," *Business, Management and Education*, vol. 16, no. 1, pp. 81–93, 2018.

[8] X. Liu, Y. Zhou, and W. Zongrun, "Can the development of a patient's condition be predicted through intelligent inquiry under the e-health business mode? Sequential feature map-based disease risk prediction upon features selected from cognitive diagnosis big data," *International Journal of Information Management*, vol. 50, no. Feb, pp. 463–486, 2020.

[9] V. Mahajan, C. Katrakazas, and C. Antoniou, "Crash risk estimation due to lane changing: a data-driven approach using naturalistic data," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–10, 2020.

[10] H. Sun, Z. Liu, G. Wang, W. LIAN, and J. MA, "Intelligent analys, is of medical big data based on deep learning," *IEEE Access*, vol. 7, 1 page, 2019.

[11] K. C. Finch, K. R. Snook, C. H. Duke, K.-W. Fu, and Z. T. H. Tse, "Public health implications of social media use during natural disasters, environmental disasters, and other environmental concerns," *Natural Hazards*, vol. 83, no. 1, pp. 729–760, 2016.

[12] D. Fu, "Guest editorial: unlocking unconventional reservoirs with data analytics, machine learning, and artificial intelligence," *Journal of Petroleum Technology*, vol. 71, no. 1, pp. 14-15, 2019.

[13] X. Zhao, D. Lord, and Y. Peng, "Examining network segmentation for traffic safety analysis with data-driven spectral analysis," *IEEE Access*, vol. 7, pp. 120744–120757, 2019.

[14] Y. Beebeejaun, A. Athithan, T. P. Copeland, M. S. Kamath, I. Sarris, and S. K. Sunkara, "Risk of breast cancer in women treated with ovarian stimulation drugs for infertility: a systematic review and meta-analysis," *Fertility and Sterility*, vol. 116, no. 1, pp. 198–207, 2021.

[15] S. He, J. Yang, M. He, D. Yan, S. Tang, and L. Rong, "The risk of future waves of COVID-19: modeling and data analysis," *Mathematical Biosciences and Engineering*, vol. 18, no. 5, pp. 5409–5426, 2021.

[16] A. Tairi, A. Elmouden, and M. Aboulouafa, "Soil erosion risk mapping using the analytical hierarchy process (AHP) and geographic information system in the tifnout-askaoun watershed, southern Morocco," *European Scientific Journal*, vol. 15, no. 30, pp. 1857–7881, 2019.

[17] F. Corman, S. Kraijema, M. Godjevac, and G. Lodewijkse, "Optimizing preventive maintenance policy: a data-driven application for a light rail braking system," *Proceedings of the Institution of Mechanical Engineers - Part O: Journal of Risk and Reliability*, vol. 231, no. 2, pp. 534–545, 2017.

[18] R. Coulter, Q. L. Han, L. Pan, and J. Zhang, Y. Xiang, Data-driven cyber security in perspective--intelligent traffic analysis," *IEEE Transactions on Cybernetics*, vol. 50, no. 7, pp. 3081–3093, 2019.

[19] I. A. Rakhmanenko and R. V. Meshcheryakov, "Identification features analysis in speech data using GMM-UBM speaker verification system," *Tr Spiiran*, vol. 3, no. 52, pp. 32–50, 2017.

[20] R. O. Sinnott and W. Voorsluys, "A scalable Cloud-based system for data-intensive spatial analysis," *International Journal on Software Tools for Technology Transfer: International Journal on Software Tools for Technology Transfer*, vol. 18, no. 6, pp. 587–605, 2016.

[21] A. Gomes, J. Kibilda, A. Farhang, R. Farrell, and L. A. DaSilva, "Multi-operator connectivity sharing for reliable networks: a data-driven risk analysis," *IEEE Transactions on Network and Service Management*, vol. 18, no. 3, pp. 2800–2811, 2021.

[22] T. S. Kuldova, "Imposter paranoia in the age of intelligent surveillance: policing outlaws, borders and undercover agents," *Journal of Extreme Anthropology*, vol. 4, no. 1, pp. 45–73, 2020.

[23] Y. Fang, Z. Shan, and W. Wang, "Modeling and key technologies of a data-driven smart city system," *IEEE Access*, vol. 9, pp. 91244–91258, 2021.

[24] J. Y. Yeh and C. H. Chen, "A machine learning approach to predict the success of crowdfunding fintech project," *Journal of Enterprise Information Management*, 2020.

[25] C. A. T. Romero, J. H. Ortiz, O. I. Khalaf, and A. R. Prado, "Web application commercial design for financial entities based on business intelligence," *Computers, Materials & Continua*, vol. 67, no. 3, pp. 3177–3188, 2021.

[26] E. Copland, D. Canoy, M. Nazarzadeh et al., "Antihypertensive treatment and risk of cancer: an individual participant data meta-analysis," *The Lancet Oncology*, vol. 22, no. 4, pp. 558–570, 2021.

[27] L. Ogiela, M. R. Ogiela, and H. Ko, "Intelligent data management and security in cloud computing," *Sensors*, vol. 20, no. 12, p. 3458, 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 5082544, 11 pages
https://doi.org/10.1155/2022/5082544

Research Article

A Study on the Design and Implementation of an Improved AdaBoost Optimization Mathematical Algorithm Based on Recognition of Packaging Bottles

Guozhu Liu [1] **and Sang-Bing Tsai** [2]

[1]*School of Art and Archaeology, Zhejiang University, Hangzhou 310028, China*
[2]*Regional Green Economy Development Research Center, School of Business, WUYI University, Nanping, China*

Correspondence should be addressed to Guozhu Liu; lgzzju@163.com

Received 11 January 2022; Accepted 17 February 2022; Published 22 March 2022

Academic Editor: Xiangtao Li

In this paper, a special design system is developed based on the design of the packaging bottle to achieve the effective acquisition of the image of the cross-section of the packaging bottle to be measured under the condition of limited space size, avoiding the distortion of the object to be measured. At the same time, the image of the area where the target packaging bottle is located is segmented, and the curve features are quickly determined and effectively matched with the template library to realize the recognition of the shape features of the bottle. In this paper, the design of the packaging bottle is first designed by mechanism design and 3D modeling, followed by rapid prototyping methods such as 3D printing, and the prototype is made for functional verification. Finally, the transmission speed and stability of the design system for packaging bottle recognition are improved through structural analysis and optimization methods. To realize the intelligent control of the packaging bottle transmission and identification system, the hardware control circuit is designed and the relevant intelligent control program is prepared based on the embedded system so that the packaging bottles in the transmission process can be quickly and accurately positioned and identified. An improved AdaBoost algorithm is proposed for packaging bottle detection. In the process of algorithm learning, the Haar features are too large and time-consuming, and the training sample is cropped to remove the sample edge pixels, which effectively reduces the number of features, thus reducing the computation. The proposed optical flow method is used to obtain the motion region in the video image as the region of interest, and the canny operator is used in the region of interest for edge detection, and the region of interest is filtered by the edge energy to exclude the noninterest region. Finally, the AdaBoost algorithm is used to detect the region of interest, which reduces the detection area and decreases the detection time. The improved AdaBoost algorithm has a high accuracy improvement over the traditional AdaBoost algorithm for the recognition of various packaging bottles with relatively suitable training set samples, and the system recognition time has reached the requirements of industrial recognition.

1. Introduction

In recent years, thanks to the rapid progress of the economy, the production and consumption of packaging bottle products have increased year by year, and the packaging bottle products industry has developed rapidly. With the increase of packaging bottle consumption, nondegradable packaging bottle products have become an important source of white pollution, causing damage to the environment while also wasting resources to a certain extent. To alleviate the pollution of the environment, effective measures must be taken to identify the use of waste packaging bottles [1]. Most of the packaging bottles are made of petroleum as raw material, in fact, according to the correct way to refine 1 ton of waste packaging bottles, you can get about 600 kg of fuel, so the packaging bottles contain a lot of renewable resources. Waste packaging bottles are also figuratively called the second oil field. Each ton of packaging bottles identified for

use can avoid the generation of about 1.5–2 tons of carbon dioxide emissions. The significance of the identification of waste packaging bottles is extraordinary. It not only brings good economic benefits and promotes resource recycling but is also one of the most effective ways to reduce carbon dioxide emissions [2]. At present, the promotion of the packaging bottle identification device has only just started. The packaging bottle identification system's function design is not perfect, leading to people accepting that the degree is not high. Most of the packaging bottle identification systems will use infrared scanning barcode results as the basis of discrimination, as the need to identify the bottle has the barcode end of the bottle towards upward delivery and requires the barcode to maintain integrity [3, 4]. The AdaBoost algorithm uses the integral graph to extract features, obtains the classifier through training samples, and selects the best weak classifier as the optimal weak classifier; in the next training, more attention is paid to the misclassified samples, and the optimal weak classifier is selected at the same time. To improve the speed of packaging bottle detection, the strong classifiers with weak to strong detection abilities are cascaded, and the weaker detection ability is made stronger. The classifier can quickly exclude areas that are easily identified as nonpackaged bottles.

Based on the research and development of core technologies for packaging bottle identification, sorting, and logistics, a new intelligent packaging bottle identification device has been developed, including key core components such as an integrated and highly reliable packaging bottle logistics and sorting device, a packaging bottle barcode/2D code online identification and weighing device, and a material identification and image recognition system. Firstly, the overall optimization design of the bottle logistics device is carried out, and the double roller rotation method drives the radial rotation of the bottle, avoiding the operational drawback that the bottle side with a barcode must be placed in the identification bin facing upward [5]. Secondly, a set of devices for item acquisition and identification is designed based on a microprocessor, which highly integrates barcode identification, weighing identification, material identification, image identification, and other technologies to improve the identification mechanism and enhance the stability and reliability of the identification device. This paper improves the strong classifier based on the classical AdaBoost algorithm and applies different strong classifiers for different test samples to classify them, from the original single, fixed strong classifier to a dynamic strong classifier that can be changed according to the different test samples [6]. The AdaBoost algorithm is a machine learning method that combines multiple base classifiers according to the weighted error rate, so it has a good effect on classification problems. Because the AdaBoost algorithm has a solid theoretical foundation and high prediction accuracy, it is computationally efficient. It is relatively simple, so it has been widely used.

Accurate and timely detection of packaging bottle information from packaging bottle images is an important prerequisite for the application of packaging bottle detection technology in the field of identification. The packaging bottle detection technology is also vulnerable to the influence of external as well as its algorithm itself factors, thus leading to differences in detection results. The algorithms that researchers have been working on often require certain prerequisites to meet the requirements of people. The algorithm in question needs to use a classifier generated by sample training when performing packaging bottle detection. There is no standard training sample pool in the field of packaging bottle detection, so the size, number, and whether the sample set is standard will have some impact on the trained classifier and thus affect the final detection results. This paper analyzes the current packaging bottle detection algorithm and the AdaBoost algorithm and proposes an improved algorithm for the AdaBoost algorithm. Chapter 1 introduces the research content and background significance of this paper, comprehensively points out the problems and challenges faced by the current algorithm and explains the arrangement of all the chapters in this paper. Chapter 2 conducts a study of related work, mainly analyzes the current research status, and points out the differences of this paper's research. The third chapter is based on the design and implementation research of an improved AdaBoost algorithm for recognizing packaging bottles, which is carried out in three dimensions: construction of packaging bottle recognition model, improved packaging bottle recognition by AdaBoost algorithm, and design and implementation of a packaging bottle recognition system. The improved AdaBoost algorithm is proposed to use the optical flow method to determine the motion region as the region of interest and reduce the detection time. The algorithm is improved in terms of reducing the amount of computation of feature values and optimizing the selection of threshold values. Chapter 4 analyzes the research presented in this paper, conducts experiments using the improved algorithm, and compares the experimental results with other related packaging bottle detection algorithms. Chapter 5 summarizes the work done in this paper, identifies the shortcomings, and provides an outlook on the future development of packaging bottle detection and recognition technology.

2. Current Status of Research

The AdaBoost algorithm is a machine learning method that combines multiple base classifiers weighted by error rate, so it has a good effect on classification problems. Since the AdaBoost algorithm has a solid theoretical foundation, high prediction accuracy, and is relatively simple to operate, it has been widely used [7]. Kutsanedzie et al. proposed the combination of the AdaBoost algorithm and SVM, using the SVM algorithm as the base classifier and the embedded multiangle AdaBoost algorithm to classify the group blocks [8]. The empirical analysis shows that the experimental results of this combined algorithm are much better than those of the AdaBoost algorithm [9]. Vanhoeyveld et al. proposed an improved AdaBoost algorithm for multiclassification problems. The boosting algorithm can solve the binary classification problem well, but when dealing with the multiclassification problem, it converts the multiclassification

problem into multiple binary classifications for processing, and the exponential loss is large [10]. In this paper, we propose the forward stepwise accumulation model algorithm, which is a direct extension of the AdaBoost algorithm to multiclass cases, to minimize the exponential loss [11]. The main idea of the WB algorithm is that part of the sample is used to train the classification model, and the other part of the sample is used to modify the weights of the model according to the classification results of the model. The results are compared with AdaBoost, Boosting, Bagging, Random Forest, and SVM algorithms using empirical analysis with real data, and the improved WB algorithm is found to be more effective than the other algorithms.

Currently available techniques for packaging bottle inspection include feature-based, motion analysis-based, differential-based, and learning-based, among others. Each method has its outstanding aspects. Zhou et al. proposed a projection curve model matching-based method for packaging bottle detection, using projection completeness, the weighted sum of offset expectation, and variance of matching points relative to the model as similarity measures, and using packaging bottle model matching to complete packaging bottle detection [12]. Teng et al. proposed a packaging bottle detection algorithm based on a priori shape information and an active contour model, using color and edge information to remove shadows and extract packaging bottle contour; introducing a priori knowledge of packaging bottle shape, establishing an a priori shape model of packaging bottle with the implicit representation of level set symbols, and constructing an active contour energy construction function with this constraint, using the variational method to find its minimum value, using shape alignment and level set method The segmentation curve of the packaging bottle is evolved to obtain the contour of the packaging bottle and then complete the detection [13]. Divya and Sri proposed a combination of an improved 3D Markov model for bottle detection, using a Markov model combined with a hybrid Gaussian model, followed by Bayesian estimation and an iterative conditional model to complete bottle detection [14]. The improved AdaBoost algorithm, which obtains the binarized mask image by three-frame differencing, finds the connected domain, reconstructs the foreground mask map after denoising, and loads the classifier to detect the image to be inspected, effectively improving the detection efficiency of the algorithm [15]. The improved AdaBoost algorithm combined with the region of interest uses a hybrid Gaussian model to extract the region of interest from the image to be examined and then uses the AdaBoost algorithm to detect it, which greatly reduces the detection time of the algorithm.

When the AdaBoost algorithm performs the contour detection of the packaging bottle, it needs to use the sample training generated classifier. There is currently no standard training sample library in the field of packaging bottle contour detection. Therefore, the size, number, and standard of the sample set will have a certain impact on the trained classifier, thereby affecting the final detection result. The intelligent bottle identification part mainly refers to the bottle logistics device and works through the logistics device to achieve the axial transfer and rotation of bottles and other actions. According to the design requirements, based on the existing mature technology, using the mechanical design optimization method, the logistics device is developed for this project, which is integrated with the packaging bottle identification system, and the workflow of the logistics device is optimized so that it has fast and efficient, stable and reliable performance. Due to a variety of factors, the quality of the packaging bottle image will be degraded. Therefore, before the packaging bottle detection, this paper first introduces image preprocessing. For the problem that the image is too bright or too dark, the packaging bottle image is enhanced using histogram equalization; there is more or less noise in the image, and different filtering methods are used for different noises. This paper focuses on the detection of packaging bottles based on the AdaBoost algorithm [16, 17]. We classify the test sample set by the clustering method in advance. Then find the closest sample group for each test set sample, and calculate the similarity between them. According to each base classifier, the error rate of each sample group classification is given its corresponding weight. Combined with the similarity between the test sample and each group, the weighted combination constitutes the final dynamic strong classifier.

3. Design and Implementation of Packaging Bottle Recognition Based on Improved AdaBoost Algorithm

3.1. Package Bottle Identification Model Construction. The barcode scanner transmits the scanning information to the main controller, and the main controller stores the information temporarily, but it is not used as the basis for controlling the movement mechanism; the material recognition sensor transmits the material information to the main controller, and the main controller stores the information temporarily, and compares it with the database in the background server, and feeds the comparison result to the main controller, and uses it as the basis for controlling the main controller; the pressure sensor transmits the weighing information to the main controller, and the main controller stores the information temporarily [18]. After taking the image of the packaging bottle, the image is processed by the image recognition algorithm, obtains the image outline, and passes it to the main controller. Temporarily stores image outline in the main controller, and the main controller then compares it with the database in the backend server by comparing it with the main controller then compares with the database in the backend server and feeds the comparison result to the main controller as the control basis of the main control. The control flow chart of the packaging bottle recognition model is shown in Figure 1. The pressure sensor required by the intelligent recycling device for plastic packaging bottles is mainly used for weighing and detecting whether there is too much residual liquid in the plastic bottles, and feeding back the detection results to the main controller, which compares the detection results with the preset weight value, recycle or return plastic bottles.

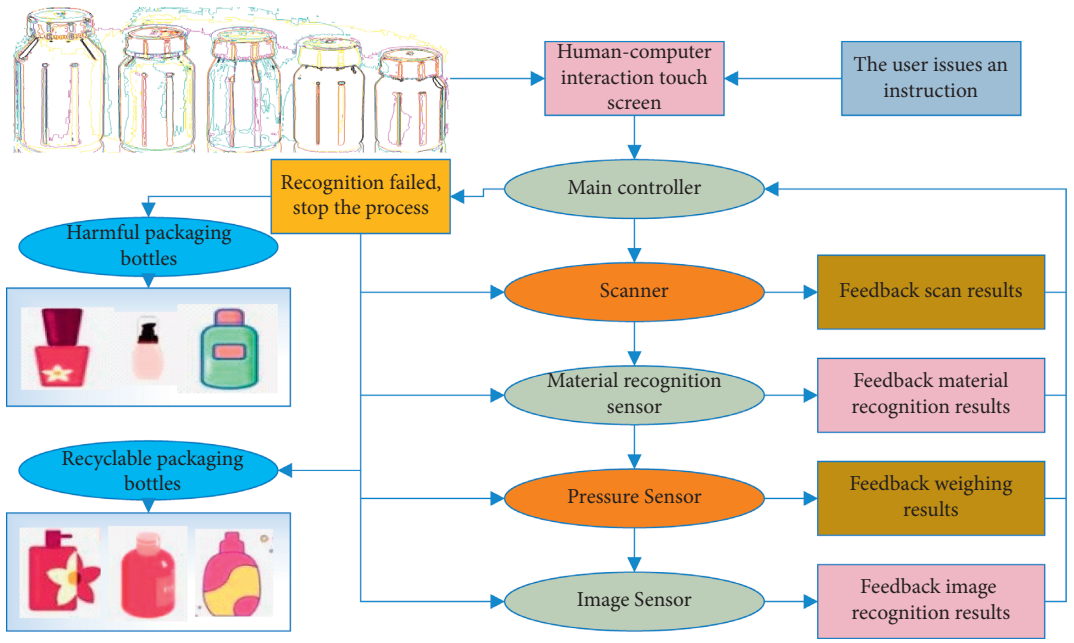


Figure 1: . Bottle identification model.

The material recognition sensor of the intelligent identification device is mainly used to distinguish the packaging bottles and cans, and the result of the recognition is fed back to the main controller for the sorting and recognition of the packaging bottles and cans. The data is uploaded at the same time, so the metal sensor with reliable performance is selected to meet the requirements. The selected material identification sensor is an inductive proximity sensor whose output is in the form of an NPN normally open. It is normally even in the normal state, and when the object is detected, the signal line outputs a negative voltage signal. The inductive proximity switch consists of a high-frequency oscillator, an amplifier circuit, and a switching circuit, and the material of the detected object must be metal [19, 20]. The high-frequency oscillator forms an electromagnetic field on the detection surface. When a metallic object reaches the induction area, the electromagnetic field causes an eddy current effect inside the metallic object, which absorbs the oscillator energy and makes the oscillator damp or even stop. The transformation of these two different states is converted by the postcircuit into a change of electrical signal in the internal circuit, which is then converted into a binary switching signal and finally amplified by the amplifier circuit and output.

When acquiring the ADC return value, we set the ruling group channel 5 of the specified ADC1 with a sampling time of 480 cycles. To obtain the value more accurately, the conversion value of channel 5 is obtained 100 times, and then the average value of the sampling is taken as the final ADC channel sampling value. According to the actual measurement of the pressure sensor, it was found that the voltage reading of the pressure sensor without the measured object was about 160 millivolts, and when the measured object was present, every 10 millivolts corresponded to about 2 grams of the bottle weight. Finally, the real weight of the acquired bottle is compared with the threshold value of 80 grams for bottles that are overweight to determine whether the packaging bottle is identified. When describing the characteristics of packaging bottles, different types of Haar features are selected, and the number of Haar features obtained is different; the size of packaging bottles is different, and the number of Haar features required is also different. For a 150 ∗ 150 bottle, there are more than 100,000 diagonal features and many more linear features. If the common Haar feature is used to describe the packaging bottle, about 800,000 features are required, that is, more than 800,000 calculations are required to obtain the feature value, and such a large amount of calculation greatly affects the speed of detection.

Image noise can be generated from external or internal sources. Most of the noise in images is random, and we usually use neighborhood averaging and median filtering to attenuate or eliminate the noise. Median filtering, as a nonlinear method to remove noise, can overcome the problem of blurred image details brought about by linear averaging filters. The principle is first to reorder the pixels in the neighborhood according to their gray value magnitude, and then to select the middle value of this ordered sequence as the output pixel value. The median is defined as a set of numbers x_1, x_2, x_3,..., x_n assuming their ordering as in the following equation:

$$Y(x) = \begin{cases} \sum_{i=}^{(n+2)/2} x_{i*2}, & n \subseteq (1, 3, 5 \ldots\ldots), \\ \dfrac{\sum_{i=}^{(n+2)/2} x_{i*2} + \sum_{i=}^{(n)/2} x_{i*2}}{2}, & n \subseteq (2, 4, 6 \ldots\ldots). \end{cases} \quad (1)$$

Y is called the median of x_1, x_2, x_3... the median of xn, and the domain of a point of a particular length or shape is called

a window. Let the input be $\{x_i\}$, then the filter output in one dimension is expressed as follows:

$$\begin{aligned} Y(x_i) &= y\{x_i\} \\ &= \min\left(y\{x_{i-a}, x_i, x_{i+a}\}, y\{x_{i-b}, x_i, x_{i+b}\}\right). \end{aligned} \tag{2}$$

The original image of the packaging bottle is obtained, and to improve the processing speed of the microprocessor, a large amount of color information contained in the image must be removed, and it is most appropriate to do the grayscale processing of the image by using the weighted average method. In the actual algorithm application, the three components of each point pixel are weighted with different weights for the weighted average. At the same time, the floating-point operation is turned into an integer operation, and the three components of blue, green, and red are multiplied by 512, 1024, and 256, respectively, and after accumulation, the integer is divided by 10000 to speed up the operation and grayscale as in the following equation:

$$y(m,n) = \frac{512R(m,n) + 1024G(m,n) + 256B(m,n)}{10000}. \tag{3}$$

3.2. AdaBoost Algorithm to Improve Packaging Bottle Recognition. Due to the complex background in the image of the packaging bottle contour detection process, nontarget moving objects are also counted as targets to be detected in the ROI region, which to a certain extent will generate the possibility of false detection. Therefore, before and after the detection of the region of interest by the AdaBoost algorithm, the motion region and the detection results are filtered to remove the regions of interest that are smaller before detection and the targets whose size and contour do not match the packaging bottle after detection according to a certain aspect ratio and the area size of the region of interest. The principle of the AdaBoost algorithm is to use nonlinear transformation to place the target space vector in the high-dimensional space and then generate the optimal classification plane in the new high-dimensional space to increase the interclass distance of samples of different categories.

Training is performed after cropping the training samples, which greatly reduces the training time while there is no significant decrease in accuracy [20]. The motion regions in the video images were acquired using the optical flow method, and the smallest outer rectangle of motion pixel points belonging to the same motion region was used as the region of interest, the region to be examined by the improved algorithm. The training samples were cropped to remove 1 to 2 pixels from the edges of the samples. N best weak classifiers are obtained after N iterations $h(x_1)$, $h(x_2)$...$h(x_n)$, which form the strong classifier according to the following equation:

$$f(x) = \sum_{i=1}^{N} a_i * h(x_i). \tag{4}$$

Due to the sharp angles of the packaging bottle contour, it is very favorable for edge detection. Using edge detection for the region of interest, the contour information of the region of interest is obtained, and again, a large number of noninteresting regions can be excluded. Using Gaussian smoothing filter convolution noise reduction, the gradient amplitude and direction are calculated as in the following equation:

$$M = \begin{vmatrix} 2 & 3 & 12 \\ 5 & 15 & 4 \\ 9 & 6 & 2 \end{vmatrix}. \tag{5}$$

The weight of each base classifier for each group is calculated by the error rate, where the error rate is the percentage of the number of wrongly classified samples to the total number of samples. The overall error rate we represent by the matrix $J(m * n)$ is shown in equation (6). Where $m * n$ denotes the error rate of the m-th base classifier for the n-th group classification [21].

$$J(m * n) = \begin{vmatrix} m_{11} & m_{12} & m_{1n} \\ m_{21} & m_{22} & m_{2n} \\ m_{n1} & m_{n1} & m_{nn} \end{vmatrix}. \tag{6}$$

The weight $w(n, i)$ of the n-th sample group to the i-th base classifier in the training sample can be calculated by equation (6), as shown in the following equation:

$$w(n,i) = \frac{\ln 1 - m_{ni}/m_{ni}}{2}. \tag{7}$$

By calculating the Euclidean distance between the test sample and the center of each training sample group $\{x_1, x_2 \ldots\ldots x_m\}$, the inverse of which is the corresponding similarity. x_{im} is the s-th attribute value of the centroid of the i-th sample group, x_{jm} is the s-th attribute value of the j-th test sample, $d(m, n)$ is the distance from the m-th sample to the n-th sample group, and $h(m, n)$ is the similarity between the m-th sample and the n-th sample group [21].

$$\begin{cases} d(m,n) = \sqrt{\sum_{i=1}^{m} (x_{im} - x_{in})^2 + \sum_{j=1}^{n} (x_{jm} - x_{jn})^2}, \\ h(m,n) = \dfrac{1}{d(m,n)}. \end{cases} \tag{8}$$

If a classifier is good at classifying a certain class of sample sets, then the classifier will also be good at classifying samples similar to such sample sets. We classify the test sample set by the clustering method beforehand, and then find the most similar sample groups in each test set and calculate the similarity between them. The base classifier is assigned a corresponding weight according to the error rate of each sample group, and when combined with the similarity of the test samples to each group, the weighted combination forms the final dynamic strong classifier. The final weight w_{mn} of the m-th test sample corresponding to the n-th base classifier can be obtained by the following equation:

$$w_{mn} = \sum_{k=1}^{N} h(m,k) * w(n,k). \tag{9}$$

The final dynamic strong classifier is obtained by weighting and combining multiple base classifiers.

$$G(x) = \max\left(0, \sum_{i=1}^{N} w_{mi} * h(x_i), 1\right). \tag{10}$$

Since the overall mean is unknown, we use the sample mean to estimate the overall mean. That is, in the test process, the difference between the sample means of the two-overall means is used to estimate the difference between the means of the two-overall means. If the two overall distributions obey a normal distribution, then

$$z(x,y) = \frac{x-y}{\sum_{i=1}^{N} \left(S(x_i) - S(y_i)\right)^2} \tag{11}$$

According to the $z(x, y)$-value table, the $z(x, y)$-values corresponding to the degree of freedom n-1 were found and compared with the calculated $z(x, y)$-values to determine whether the results were significant according to Table 1.

3.3. Design and Implementation of Packaging Bottle Identification System. The main function of the intelligent packaging bottle identification system is to identify the packaging bottle material, barcode information, weight, and the outline of the packaging bottle shape, to discern whether to identify the packaging bottle and at the same time provide the rebate basis by matching the comparison with the template in the backend server database. Now some intelligent packaging bottle recognition systems on the market to packaging bottle recognition, the function is refuted and not practical, resulting in human-computer interaction interface operation is too complex, does not meet the user operation convenient function demand. Therefore, simple operation and practical function are the main principles of this system design. Intelligent packaging bottle identification systems are mostly put in public areas with a large flow of people, and long-term unattended states. The Haar rectangular feature extracted by the AdaBoost algorithm is relatively simple, and it is easy to cause false detection and missed detection for nontarget interference and occlusion. At the same time, the adaptability of the AdaBoost algorithm is poor for bottles with different angles. At the same time, for the detection of the entire packaging bottle, the AdaBoost algorithm needs to use the sliding window to search and detect, including a large amount of useless information, which has a great impact on the speed of the algorithm.

The intelligent packaging bottle identification system embedded system adopts the modular design concept to simplify the design steps and make the design ideas clearer. According to the different functions of the system for modular design, including human-computer interaction module, machine control module, and intelligent identification module, there are three major parts. The human-machine interaction module is mainly realized by a touch-type all-in-one machine. The user can operate the all-in-one machine to complete the bottle-throwing. At the same time, the background staff can also view and understand the working status of the intelligent packaging bottle identification system; the motor control module includes the control of the motor and its driver, photoelectric switch, limit switch, and other related sensors; the identification module includes the barcode scanner for barcode identification and the metal sensor for material identification. The main controller identifies the input bottles by controlling each sensor and feeds the identification results back to the main controller for data interaction. The general design scheme of the system is shown in Figure 2.

The design divides the image recognition system into four parts according to the different functions, namely, an image acquisition module, an image recognition module, an image storage module, and an image display module. The working process of each functional module is as follows: the image acquisition module is mainly to complete the acquisition of image data through the image sensor device and provide the image data to the microprocessor for recognition processing; the image recognition module is for the microprocessor to recognize and process the acquired image data through the image recognition algorithm and output the recognition result; the image storage module is to cache the image data processed by the image recognition module. The image storage module caches the image data processed by the image recognition module and prepares it for transmission to the image display module; the image display module mainly displays the results of the image data after recognition processing, and the microprocessor then makes the recognition judgment.

4. Analysis of Results

4.1. Packaging Bottle Identification Model Simulation Analysis. Two photos were randomly selected from each of the packaging bottles in the database for training, and the other eight were used for recognition. Three algorithms were used for packaging bottle recognition, and the recognition rate was calculated. In the second experiment, 3 photos of each person in the data were randomly selected for training, and the other 7 were used for recognition. And so on, in the next 6 experiments, the photos of packaging bottles in the training set were gradually added 1 until the number of photos was 9. The recognition rate of the 3 recognition models for each experiment was counted as shown in Figure 3.

800 single-package bottle images and 500 multipackage bottle images were randomly selected, and they were detected using two recognition models, and then the results were simulated and analyzed. The simulation results are shown in Figure 4. For both algorithms, the detection rate of multipack bottles is lower than that of single-pack bottles, and the false detection rate of multipack bottles is higher than that of single-pack bottles. Because the background of multipack bottles is relatively more complex, the posture of the bottles is also varied, and the bottles may be missed if the face of the bottles is partially obscured, these factors lead to a lower detection performance of multipack bottles than that of single-pack bottles (see Figure 4).

Table 1: . $z\ (x, y)$-value and significance relationship table.

Serial number	$z\ (x, y)$-value	S Value	Significance level	Significant difference
1	$z\ (x, y) < 0$	$S < 0.05$	0.05	The difference is very significant
2	$z\ (x, y) = 0$	$S = 0.05$	0.08	Significant difference
3	$z\ (x, y) > 0$	$S > 0.05$	1.00	No obvious difference

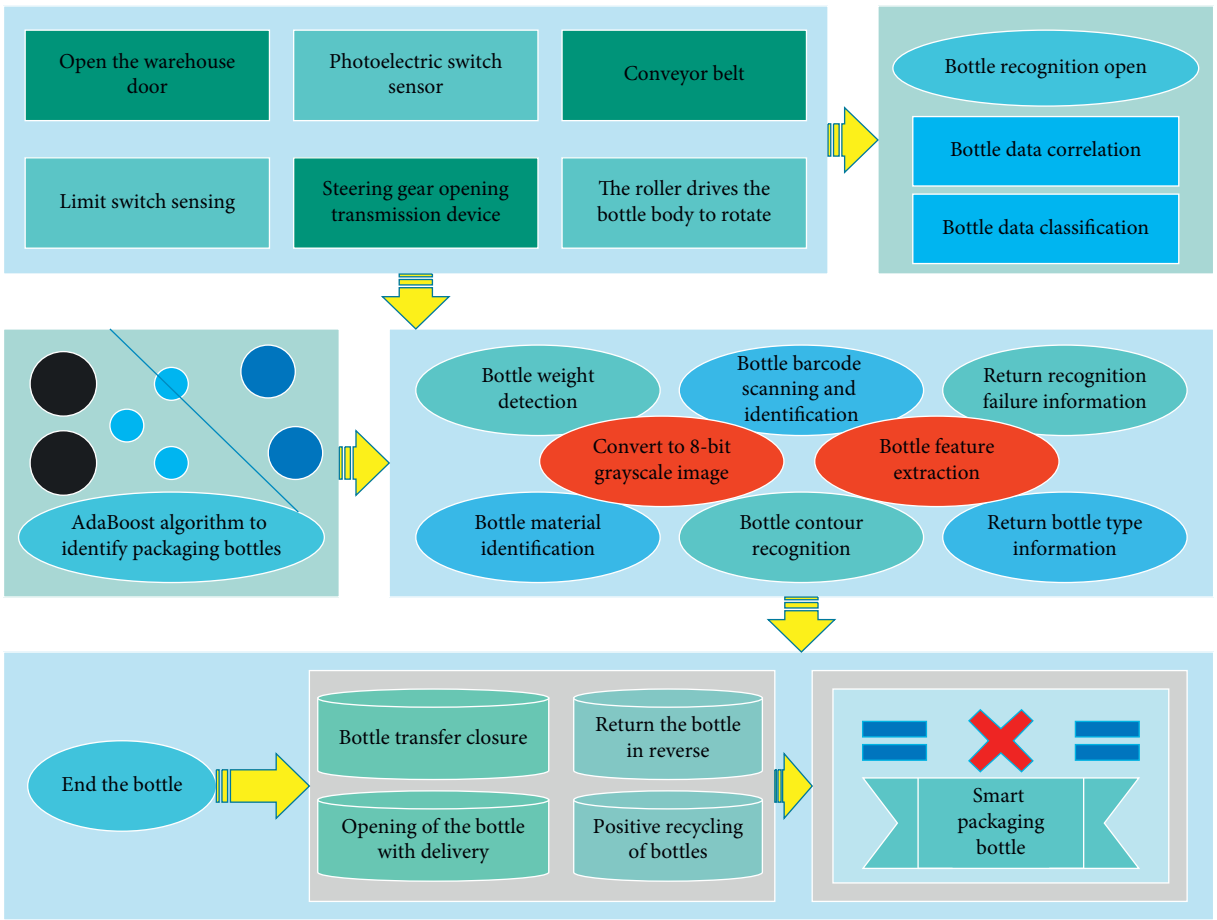


Figure 2: . System functional framework diagram.

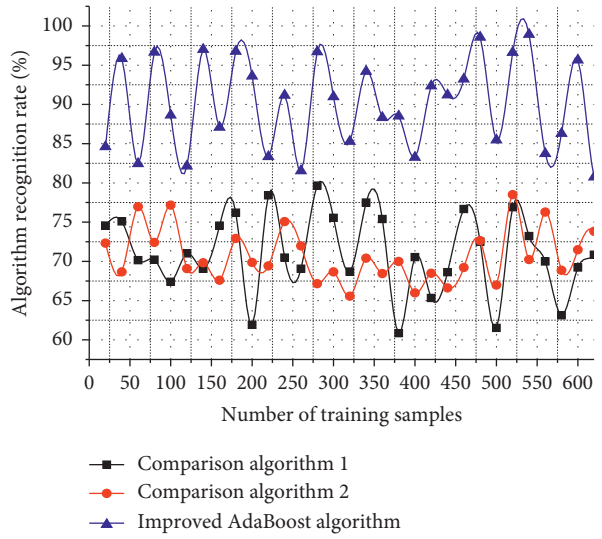


Figure 3: . The recognition rate of three recognition models.

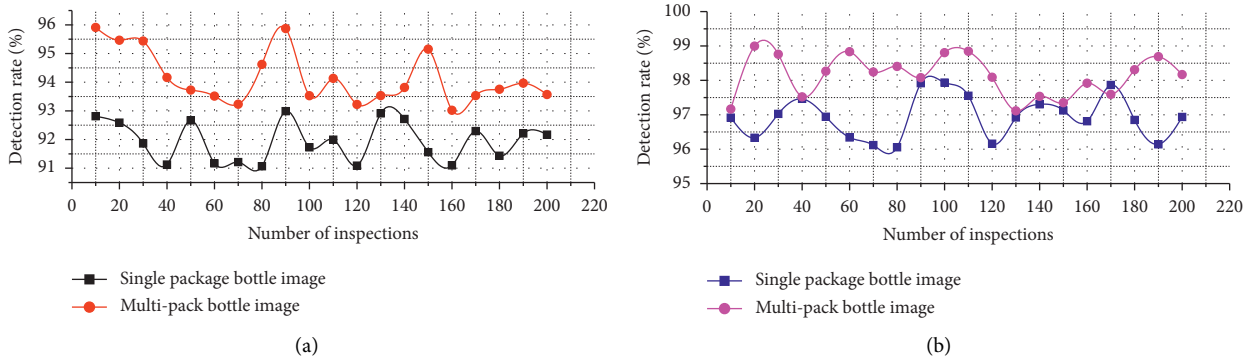


Figure 4: . Identification model simulation results. (a) Traditional AdaBoost algorithm. (b) Improved AdaBoost algorithm.

4.2. Performance Analysis of Packaging Bottle Recognition Algorithm. From Figure 5, it can be seen that the differential method is affected by the shadow of the packaging bottle itself in the static background, causing false detection and missed detection. In this paper, the algorithm has a leakage phenomenon for the contour detection of the packaging bottle. The reason is that in the training stage of the algorithm, the front face image of the positive sample bottle is mainly selected from common household bottles, and the algorithm cannot detect the special bottle contour very well (see Figure 5).

From Figure 6, it can be seen that the AdaBoost algorithm has a high false detection rate for packaging bottle contour detection in a complex background, which is more easily affected by the background. The algorithm in this paper shows better stability in this complex background. The experimental results show that the differential and traditional AdaBoost algorithms in the detection process will wrongly detect other motion targets in the complex background, resulting in false detection. The improved algorithm effectively removes the false targets after extracting the ROI and screening the ROI. Compared with the differential method and the traditional AdaBoost algorithm, the improved algorithm has a certain improvement in accuracy and a significant reduction in the false detection and missed detection rates (see Figure 6).

4.3. Packaging Bottle Identification System Actual Test Analysis. In terms of recognition accuracy, Figure 7 shows the relationship between the recognition accuracy and the number of training rounds of the traditional AdaBoost and the improved AdaBoost algorithms, and the experimental system is the packaging bottle recognition system designed in this paper. From Figure 7, we can see that the accuracy of both algorithms improves as the number of training rounds increases, but the improvement of this algorithm is more obvious with the same number of training rounds. Initially, the difference in accuracy between the two algorithms is about 4%, and the peak is reached at about 120 rounds, when the difference in accuracy between the two algorithms reaches 8%, after which the accuracy of the two algorithms stabilizes. The accuracy of the improved AdaBoost algorithm reaches 92.86%, which is about 9% higher than that of the traditional AdaBoost algorithm, and the practicality is improved (see Figure 7).

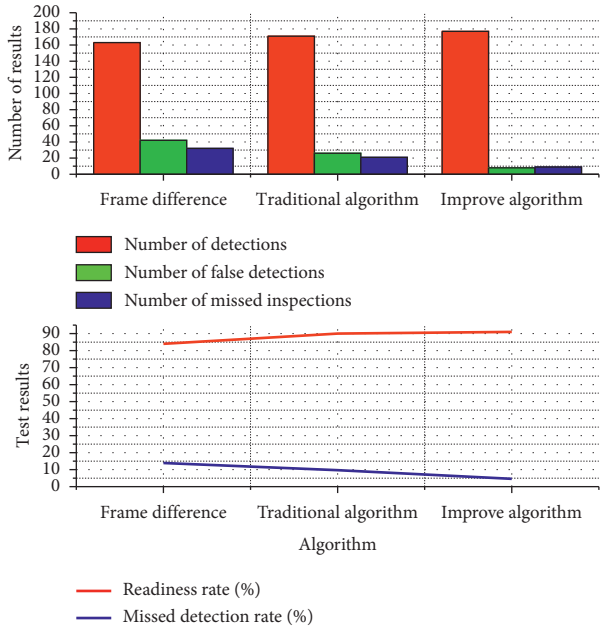


Figure 5: . Single background test results.

The purpose of the significance test in this paper is to verify whether the difference between the classification effect of the improved AdaBoost algorithm before and after is significant. The significance test was performed using the SPSS software, and the results are shown in Figure 8.

From the significance test of the error rate of sample 1 in Figure 8, it can be seen that the variances of the two samples are not equal, and at the significance level of 5.2%, $0.0 < p < 0.5$, indicating that there is a significant difference between the means of the two samples before and after the improvement of sample 1. From the significance test of the error rate of sample 2 in Figure 8, it is clear that the variances of the two samples are not equal, and at the significance level of 4.1%, $0.5 < p < 0.7$, indicating that there is a significant difference between the two sample means before and after the improvement of sample 2. In summary, there is a significant difference in the mean values of error rates derived from the AdaBoost algorithm before and after the

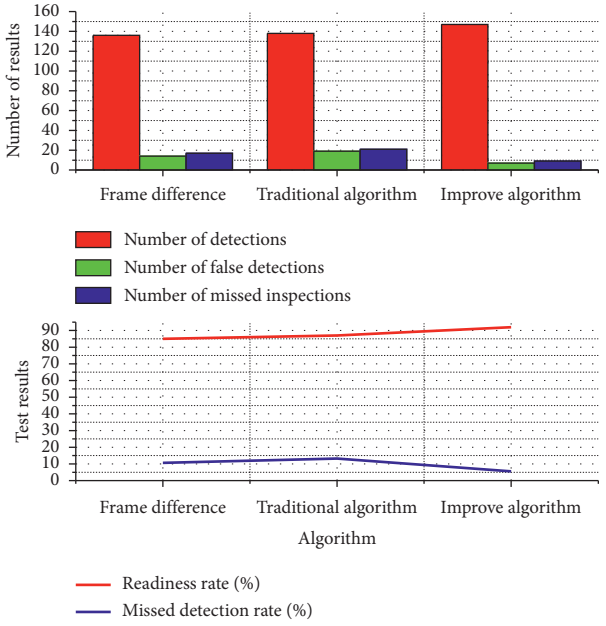


Figure 6: . Complex background detection results.

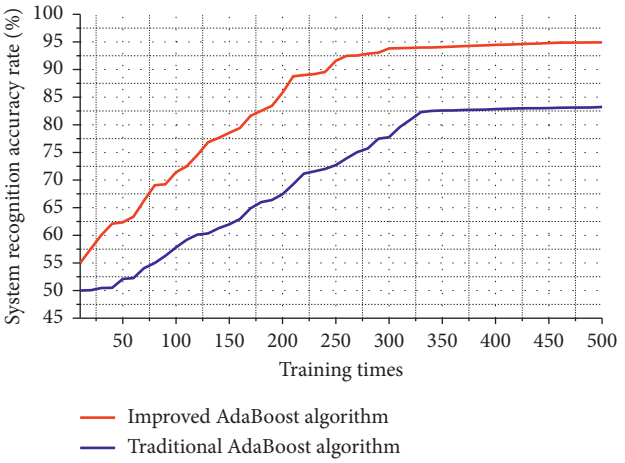


Figure 7: . System recognition accuracy rate.

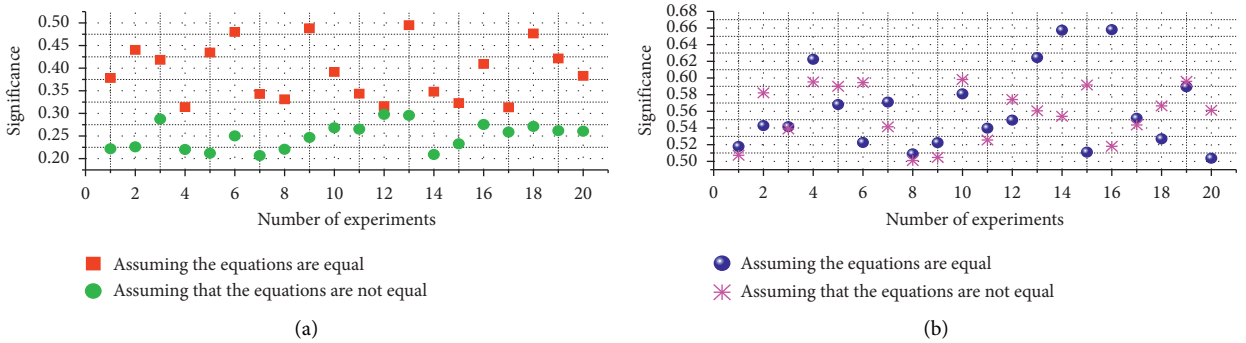


Figure 8: . Results of the systematic significance test: (a) sample 1 and (b) sample 2.

improvement, that is, the difference in the classification effect of the AdaBoost algorithm before and after the improvement is significant and the improvement is meaningful.

5. Conclusion

The research designs the design and implementation of packaging bottle recognition based on the improved AdaBoost algorithm. The innovative design and promotion of an intelligent packaging bottle recognition system are of great significance to improve the recognition rate of packaging bottles and standardize the recognition system. For the existing packaging bottle recognition methods, the multidimensional recognition methods of packaging bottle barcode and dimension code online recognition, weighing, material, and packaging bottle image recognition are proposed. Research into image recognition algorithms has also been launched. Through the research and application of preprocessing, edge detection, and image morphology, the clear outline of the packaging bottle in the image is obtained and matched with the outline template in the database to complete the image recognition process. Through a lot of experiments and tests, the intelligent recognition device can operate stably and reliably and can realize the effective intelligent recognition of waste packaging bottles, which has met the basic requirements of the design. For classification, this paper selects the improved AdaBoost algorithm, which can combine the similarity of test samples and each cluster into different strong classifiers according to different test samples, which improves the adaptability of strong classifiers to test samples and improves the overall accuracy. In this paper, the inverse of the Euclidean distance is used to calculate the similarity between test samples and clusters. Although the algorithm studied in this paper has achieved certain results, there are still many problems. Since the training samples are collected from the front face images of the packaging bottles, the detection of the contours of the packaging bottles on the side and back is not yet possible. Optimization for the AdaBoost algorithm sample training phase is effective, but there is further room. With the continuous improvement of artificial intelligence technology, we believe that packaging bottle contour detection technology will make great progress.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest with this study.

References

[1] M. R. Zarezadeh, M. Aboonajmi, and M. Ghasemi Varnamkhasti, "Fraud detection and quality assessment of olive oil using ultrasound," *Food Sciences and Nutrition*, vol. 9, no. 1, pp. 180–189, 2021.

[2] J. Peng, W. Zhu, W. Zhu et al., "Defect detection in code characters with complex backgrounds based on BBE," *Mathematical Biosciences and Engineering*, vol. 18, no. 4, pp. 3755–3780, 2021.

[3] Z. Mahmood, N. Bibi, M. Usman, U. Khan, and N. Muhammad, "Mobile cloud based-framework for sports applications," *Multidimensional Systems and Signal Processing*, vol. 30, no. 4, pp. 1991–2019, 2019.

[4] R. R. Karn, P. Kudva, and I. A. M. Elfadel, "Dynamic autoselection and autotuning of machine learning models for cloud network analytics," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 5, pp. 1052–1064, 2018.

[5] M. A. Ferrag, L. Maglaras, A. Derhab, and H. Janicke, "Authentication schemes for smart mobile devices: threat models, countermeasures, and open research issues," *Telecommunication Systems*, vol. 73, no. 2, pp. 317–348, 2020.

[6] D. Prasanna and M. Prabhakar, "An effiecient human tracking system using Haar-like and hog feature extraction," *Cluster Computing*, vol. 22, no. 2, pp. 2993–3000, 2019.

[7] V. E. Trujillo and M. K. Hinders, "Container monitoring with infrared catadioptric imaging and automatic intruder detection," *SN Applied Sciences*, vol. 1, no. 12, pp. 10–25, 2019.

[8] F. Y. H. Kutsanedzie, Z. Guo, and Q. Chen, "Advances in nondestructive methods for meat quality and safety monitoring," *Food Reviews International*, vol. 35, no. 6, pp. 536–562, 2019.

[9] C. Yáñez-Márquez, I. López-Yáñez, and M. Aldape-Pérez, "Theoretical foundations for the alpha-beta associative memories: 10 years of derived extensions, models, and applications," *Neural Processing Letters*, vol. 48, no. 2, pp. 811–847, 2018.

[10] J. Vanhoeyveld, D. Martens, and B. Peeters, "Customs fraud detection," *Pattern Analysis & Applications*, vol. 23, no. 3, pp. 1457–1477, 2020.

[11] W. Li, L. Cao, L. Yan, J. Liao, and Z. Wang, "Vacant parking slot detection and tracking during driving and parking with a standalone around view monitor," *Proceedings of the Institution of Mechanical Engineers - Part D: Journal of Automobile Engineering*, vol. 235, no. 6, pp. 1539–1551, 2021.

[12] G. Zhou, G. Zhang, and B. Xue, "A maximum-information-minimum-redundancy-based feature fusion framework for ship classification in moderate-resolution SAR image," *Sensors*, vol. 21, no. 2, pp. 519–600, 2021.

[13] J. Teng, D. Zhang, D.-J. Lee, and Y. Chou, "Recognition of Chinese food using convolutional neural network," *Multimedia Tools and Applications*, vol. 78, no. 9, pp. 11155–11172, 2019.

[14] V. Divya and R. L. Sri, "Docker-based intelligent fall detection using edge-fog cloud infrastructure," *IEEE Internet of Things Journal*, vol. 8, no. 10, pp. 8133–8144, 2020.

[15] S. Gharami, B. Prabadevi, and A. Bhimnath, "Semantic analysis - internet of things, study of past, present and future of IoT," *Electronic Government, an International Journal*, vol. 15, no. 2, pp. 144–165, 2019.

[16] G. AlRegib, M. Deriche, Z. Long et al., "Subsurface structure analysis using computational interpretation and learning: a visual signal processing perspective," *IEEE Signal Processing Magazine*, vol. 35, no. 2, pp. 82–98, 2018.

[17] S. P. Faustina Joan and S. Valli, "A survey on text information extraction from born-digital and scene text images," *Proceedings of the National Academy of Sciences, India - Section A: Physical Sciences*, vol. 89, no. 1, pp. 77–101, 2019.

[18] D. Masouros, S. Xydis, and D. Soudris, "Rusty: runtime interference-aware predictive monitoring for modern multi-

tenant systems," *IEEE Transactions on Parallel and Distributed Systems*, vol. 32, no. 1, pp. 184–198, 2020.

[19] D. A. S. Kumar, D. Ar, L. G. Babu, and G. Suresh, "Smart agriculture robo with leaf diseases detection using IOT," *European Journal of Molecular & Clinical Medicine*, vol. 7, no. 11, pp. 2462–2469, 2022.

[20] Y. Joshi, H. Mewada, and H. Mewada, "Object tracking in occlusion and contrast conditions using patch-wise sparse method," *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 5, pp. 295–306, 2020.

[21] D. YanLing, D. Ran, W. DongSheng, W. RuoYang, Y. ShaoJun, and N. Hui, "Research on islanding detection method of distributed photovoltaic power supply based on improved Adaboost algorithm," in *Proceedings of the 2020 IEEE Power & Energy Society General Meeting (PESGM), Shanghai, China*, August 2020.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 8373138, 14 pages
https://doi.org/10.1155/2022/8373138

Research Article

Multidepot Two-Echelon Vehicle Routing Problem for Earthwork Allocation Optimization

Qinglong Zhang, Naifu Deng, Yanwen Zhu, and Zhenping Huang

Department of Civil Engineering, School of Civil and Resource Engineering, University of Science and Technology Beijing, Beijing 100083, China

Correspondence should be addressed to Naifu Deng; nden8417@163.com

Received 9 December 2021; Accepted 6 January 2022; Published 29 January 2022

Academic Editor: Xiangtao Li

Prior to the construction of most engineering projects, earthwork is a complex and time-consuming task, requiring iterative operations in civil engineering. The effectiveness of earthworks determines the cost of many AEC (architecture, engineering, and construction) projects (e.g., road, embankment, railway, and slope engineering). As a result, creating effective earthwork planning is critical. The earthwork allocation problem is simplified in this study to the vehicle route problem (VRP), which is often studied in the field of transportation and logistics. An optimization model for the earthwork allocation path based on the modified genetic algorithm with a self-adaptive mechanism is developed to work out the global optimal hauling path for earthwork. The findings of the study are also used to shape the basic topographic shape of the Winter Olympic Skiing Course Project. Furthermore, a comparative study with the former methods is conducted to validate the performance of our proposed method on tackling such a multidepot two-echelon vehicle routing problem. Because of its flexibility, this optimization model is extremely compatible with various evolutionary methods in many fields, making future development viable and practicable.

1. Introduction

Earthwork allocation is a substantial and repetitive task required for the majority of AEC (architecture, engineering, and construction) projects [1, 2]. Earthwork is the process of leveling or shaping the ground in a target area by moving or handling the geological materials that make up the target area. This geological material allocation usually includes excavation, loading, handling, unloading, and compaction operations in different areas. Sometimes it may also include some intermediate steps, such as material mixing or processing [3, 4]. Because of the complexity of earthwork activities, the volume of work is large and often accounts for more than 50% of the total project cost [5]. In particular, among the above steps, transporting geotechnical materials from one location to another is in most cases the most expensive [6, 7]. Since earth distribution is usually a continuous and repeatable iterative process, finding an effective technical approach to rationalize the earth deployment path can yield significant economic benefits, such as reduced fuel consumption or carbon emissions [1, 8].

Earthwork allocation is essentially an extended application of the vehicle path problem (VRP). VRP is a combinatorial dynamic planning problem that seeks to utilize a fleet of vehicles, such as trucks, to serve a certain number of customers with different cargo requirements under certain constraints, such as the delivery time or load capacity of nominated vehicles [9, 10]. It primarily refers to a set of issues in which an ideal route for a fleet of vehicles based on one or more depots may be identified for a number of geographically separated consumers [11, 12]. Figure 1 represents the schematic diagram of a single-depot VRP. The vehicle route problem, first proposed by Dantzig and Ramser in 1959 [11], is a critical problem in transportation and logistics. Since its inception, the problem has piqued the interest of experts and researchers, and it has aided the development of transportation, aviation, navigation, communications, electric power, computer science, and other

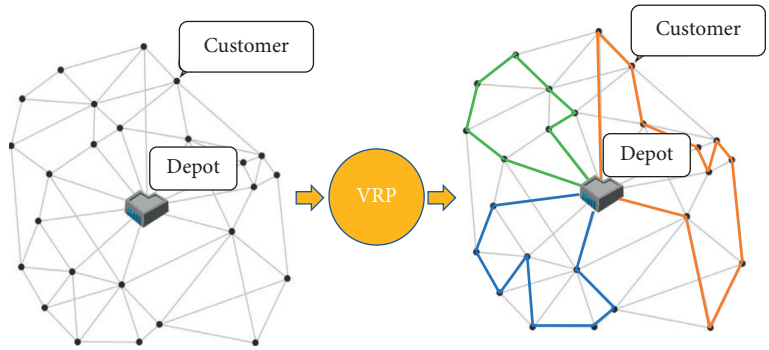


Figure 1: Schematic diagram of a single-depot VRP (also see in [6]).

fields in some ways. The VRP, on the other hand, is a nondeterministic polynomial hard (NP-hard) problem that cannot be solved in polynomial time. In other words, given the existing knowledge, it is impossible to find the exact answer to the problem, or it will take ages to find the exact solution. It is especially true when the task is large-scale. As a result, the current research has changed from finding the perfect answer using an exact algorithm to finding a better solution using a heuristic approach.

The frequently used heuristic algorithms are the genetic algorithm (GA) [13–16], ant colony algorithm (ACA) [17–19], particle swarm optimization algorithm (PSOA) [20–23], simulated annealing algorithm (SAA) [24–26], tabu search algorithm (TSA) [27–29], differential evolution algorithm (DEA) [30], and so on. ACA is adept at path planning problems, in particular, ACA shows great robustness in solving the traveling salesman problem [31]. However, ACA requires a large amount of computation because it usually requires all ants to choose the same route, which is the optimal line. In practical calculation, it is difficult to achieve this situation under a given number of cycles [32]. SAA is universal and robust, which is suitable for parallel processing and complex nonlinear optimization problem, nevertheless, it also relies on higher computation resources and longer computing time [33]. PSOA and TSA are the opposite, it has fast convergence speed, however, it is easy to produce premature convergence problem, especially in dealing with complex multipeak search problems [33, 34]. DEA and GA are evolutionary algorithms. DEA is modified from GA by improving the mutation operation to accelerate its convergence speed [35], however, fast convergence speed may lead to a premature problem when the initial population is small [36].

In this paper, GA is adopted to study the earthwork allocation path problem as it is one of the earliest algorithms to be applied to the field of transportation [37]. Furthermore, research demonstrates that GA is well-suited to handling NP-hard issues, including various variables, parameters, objectives, and weak connectivity across different areas [38, 39]. GA can also record numerous solutions at once, and the simultaneous optimization process for multiple solutions can be used to solve multiobjective optimization problems [39]. Furthermore, GA has a better universality and compatibility, which allows for further improvement when combined with other heuristic algorithms [40].

Therefore, this paper focuses on the vehicle route problem to refine the earth allocation problem based on the genetic algorithm. In reality, earth allocation may include multiple vehicle replenishment centers (depots), multiple cutting fields (distribution centers), and multiple filling fields (customers) distributed geographically in a certain range of areas. Before reaching its mileage restriction, each vehicle must return to any replenishment site to refuel. Furthermore, fuel usage varies depending on whether the truck is empty or loaded. Furthermore, each excavation area has a maximum amount of earthwork it can generate. Hence, an excavation area cannot contribute a limitless amount of earthwork to its neighboring filling regions. When the earthwork from their nearest excavation location is depleted, those filling regions must choose another suitable excavation place to furnish the earthwork. In light of the aforementioned restrictions, the goal of this work is to maximize the number of sent trucks and the overall fuel usage throughout the shipping distance. Accordingly, this paper made the following contributions:

(1) This paper constructs a complicated multidepot two-echelon vehicle routing problem (MD-TEVRP) and provides a many-to-many recursive pairing solution based on the genetic algorithm.

(2) A self-adaptive mechanism is designed to control the crossover and mutation rate to manipulate and maintain the diversity of the generated population, which can prevent the local convergence problem and provide robust performance.

(3) The proposed method demonstrates great performance in the case study and comparative study, which provided guidance for the construction of the skiing courses of the Beijing Winter Olympic Games Skiing Center in Yanqing, Beijing.

2. Model Formation

2.1. Problem Description. This paper assumes that the earthwork working sites are divided into 4 parts, i.e., the replenishment centers, the cutting fields, the filling fields, and the paths. The replenishment centers, denoted as $j \in J$, are defined as the depots where engineering vehicles (in this paper, dump trucks are mainly considered) can get refueled

or repaired. Each replenishment center is able to accommodate I_j dump trucks that can provide stable and nonstop hauling services at each dispatch, which means these dump trucks will not stop transporting earthworks from the cutting fields to the filling fields until they exhaust all fuel capacity. To this point, earthwork allocation can be regarded as an extension of the vehicle route problem (VRP) with the purpose of transporting the earthwork from several specific places to other different appointed locations. The planning key of earthwork allocation is to transport the required amount of earth volume from the cutting field $k \in K$ and distribute them according to the different earthwork needs of the filling fields $g \in G$. To simplify this problem, this paper defines the volume of earthwork to be excavated in the k^{th} cutting field as the number of fully loaded dump trucks b_k needed to transport away all the dredged earth. Similarly, the volume of earthwork needed in the g^{th} filling fields demands c_g fully loaded dump trucks to unload the earth. It is notable that the earthwork from all cutting fields should be not less than the earthwork needed in all filling fields, namely,

$$\sum_{k\in K} b_k \geq \sum_{g\in G} c_g, \quad \forall k \in K, \forall g \in G, \tag{1}$$

where b_k is the total number of fully loaded dump trucks needed to haul all earthwork in k^{th} cutting field, and c_g is the total number of fully loaded dump trucks needed to fill the g^{th} filling field.

As illustrated in Figure 2, an empty-laden truck is dispatched from the j^{th} replenishment centers to the k^{th} cutting field after fully loading with earth and rocks. This truck will move to the g^{th} filling field and unload the earth and rocks. Depending on its fuel capacity, this truck can either return to one of the replenishment centers to get refueled (in this case, this path for this truck is called one-way path) or head to one of the cutting fields to get loaded again and repeat the earthwork hauling work between the cutting fields and the filling fields (in this case, the path for this truck is called multiway path). In this paper, the maximum mileage of a truck is dynamic based on how often it is fully loaded during each dispatch, i.e., each truck has its own maximum fuel capacity l_{ij}, and thus, a fully loaded truck will consume more fuel at each unit hauling distance, and the more frequent it is fully loaded, the less mileage it can guarantee. Practically, earthwork allocation is repetitive, and thus, a multiway transportation path planning for each dump truck is indispensable. On the basis of not exceeding the mileage limit, it is required that a minimum number of dispatched trucks and a global minimum hauling distance are derived accordingly to achieve a minimum fuel consumption for the least carbon emission.

2.2. Objective Model and Constraints. According to the problem illustrated above and the schematic diagram shown in Figure 2, our objective is to find the shortest overall hauling distance and achieve the lowest fuel consumption. Therefore, the objective model can be formulated as follows:

$$\min Z = \sum_{j\in J}\sum_{i\in I_j}\sum_{k\in K}\sum_{g\in G} \left[\alpha\left(d_{ijk}X_{ijjk} + d_{igk}U_{ijgk} + d_{igj}V_{ijgj}\right) + \beta d_{ikg}Y_{ijkg}\right], \tag{2}$$

where d_{ijk} is the hauling distance of the i^{th} dump truck from the j^{th} replenishment center to the k^{th} cutting field, d_{ikg} is the hauling distance from the k$^{\text{th}}$ cutting field to the g^{th} filling field, d_{igk} is the hauling distance from the g^{th} filling field to the k^{th} cutting field, and d_{igj} is the hauling distance from the g^{th} filling field to the j^{th} replenishment center. α and β are the average fuel consumption per unit distance of each dump truck at no load and full load, respectively. X_{ijjk}, Y_{ijkg}, U_{ijgk}, and V_{ijgj} are binary decision variables.

Equation (2) is the objective function that consists of two parts. The first part is the total fuel consumption of all dump trucks at no load. To be specific, $\alpha d_{\text{ijk}}X_{\text{ijjk}}$ indicates the fuel consumption of the i^{th} dump truck at the j^{th} replenishment center dispatched from the j^{th} replenishment center to the k^{th} cutting field, $\alpha d_{\text{igk}}U_{\text{ijgk}}$ is the fuel consumption upon traveling from the g^{th} filling field back to the k^{th} cutting field at no load, and $\alpha d_{\text{igj}}V_{\text{ijgj}}$ represents the fuel consumption coming from the g^{th} filling field back to the j^{th} replenishment center at no load. The second part is the total fuel consumption of all dump trucks at full load, specifically referring to the sum fuel consumption hauling from the k^{th} cutting field to the g^{th} filling field.

In addition, the objective function in (1) should be subjected to the following constraints:

$$\sum_{i\in I_j}\sum_{k\in K} X_{ijjk} \leq a_j, \quad \forall i \in I_j, \forall j \in J, \forall k \in K, \tag{3}$$

$$\sum_{j\in J}\sum_{i\in I_j}\sum_{g\in G}\left(X_{ijjk} + U_{ijgk}\right) \geq b_k, \quad \forall i \in I_j, \forall j \in J, \forall k \in K, \forall g \in G, \tag{4}$$

$$\sum_{j\in J}\sum_{i\in I_j}\sum_{k\in K} Y_{ijkg} \geq c_g, \quad \forall i \in I_j, \forall j \in J, \forall k \in K, \forall g \in G, \tag{5}$$

$$\sum_{k\in K}\sum_{g\in G}\left[\left(d_{ijk} + d_{igk} + d_{ikj}\right)\alpha + d_{ikg}\beta\right] \leq l_{ij}, \quad \forall i \in I_j, \forall j \in J, \forall k \in K, \forall g \in G, \tag{6}$$

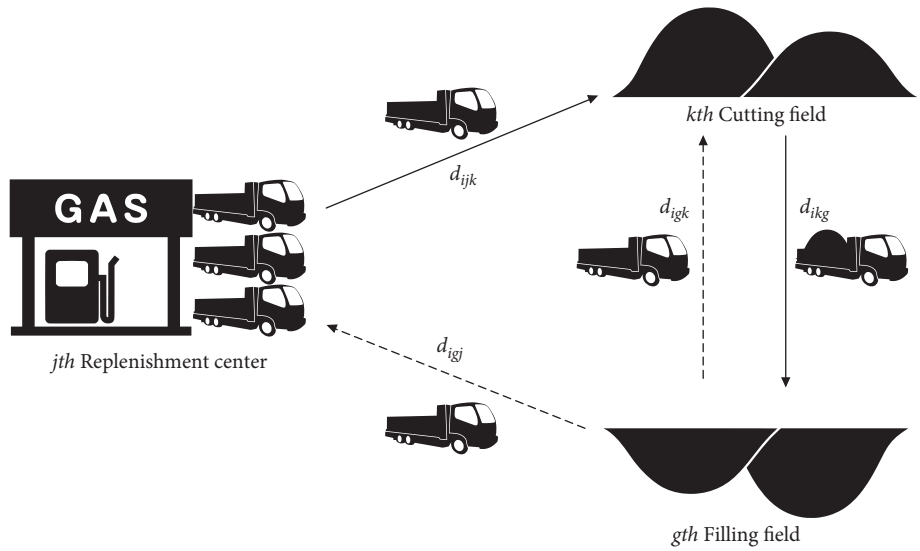


Figure 2: Schematic diagram of earthwork allocation (the dash lines indicate that the trucks have two options depending on their mileage limitation. They can either return to a replenishment center or head to a cutting field to reload).

$$\sum_{k\in K} b_k \geq \sum_{g\in G} c_g, \quad \forall k \in K, \forall g \in G, \tag{7}$$

$$X_{ijjk} \in \{0,1\}, \quad \forall i \in I_j, \forall j \in J, \forall k \in K, \tag{8}$$

$$Y_{ijkg} \in \{0,1\}, \quad \forall i \in I_j, \forall j \in J, \forall k \in K, \forall g \in G, \tag{9}$$

$$U_{ijgk} \in \{0,1\}, \quad \forall i \in I_j, \forall j \in J, \forall k \in K, \forall g \in G, \tag{10}$$

$$V_{ijgj} \in \{0,1\}, \quad \forall i \in I_j, \forall j \in J, \forall g \in G, \tag{11}$$

where a_j is the total number of available dump trucks at the j^{th} replenishment center. (3) represents that the dispatched dump trucks from each replenishment center should not exceed the maximum number of available dump trucks parked in each replenishment center. Equations (4) and (5) determine that the number of dump trucks sent to each cutting field and each filling field should satisfy the earthwork hauling need. Equation (6) indicates that the maximum hauling distance of each dump truck is constrained by its own maximum fuel capacity. Equation (7) indicates that the earth volume in all cutting fields must not be less than that in all filling fields, otherwise, the filling need will not be satisfied. Equations (8) to (11) are the decision variables to control the hauling path of each dump truck.

3. Adaptive Genetic Algorithm

The proposed earthwork allocation problem can be regarded as an MD-TEVRP, which is a many (dump trucks in the replenishment centers)-to-many (filling fields) matching problem, considering the transfer stations (cutting fields). The MD-TEVRP is an NP-hard question [41], and thus, this paper proposed an adaptive genetic algorithm to determine the optimal solution.

The genetic algorithm is a heuristic algorithm that is based on the "Survival of the Fittest" and "Natural Selection" theories. Figure 3 depicts the process and technique by which randomly created individuals compete with one another and form new generations using self-adaptive crossover and mutation strategies. When the system converges to a stable solution in the final phase, it signifies that the optimal solution has been discovered.

3.1. Generation of Initial Population. Based on the dynamic planning theory, the Floyd algorithm is used to calculate the shortest distance between the arbitrary points i and j in the original spatial weight matrix [42]. The Floyd algorithm seeks a third point k between the points i and j by comparing the distances from i to j directly and from i to j via k. If the latter distance is shorter, the path i-k-j is updated as the shortest distance between the points i and j. Therefore, the recursive process of the Floyd algorithm can be expressed as follows:

$$d_{ij}^{k} = \min\left(d_{ij}^{k-1}, d_{ik}^{k-1} + d_{jk}^{k-1}\right). \tag{12}$$

By deriving the shortest distance between each point via (4), the next step is to find the optimal path for the many-to-

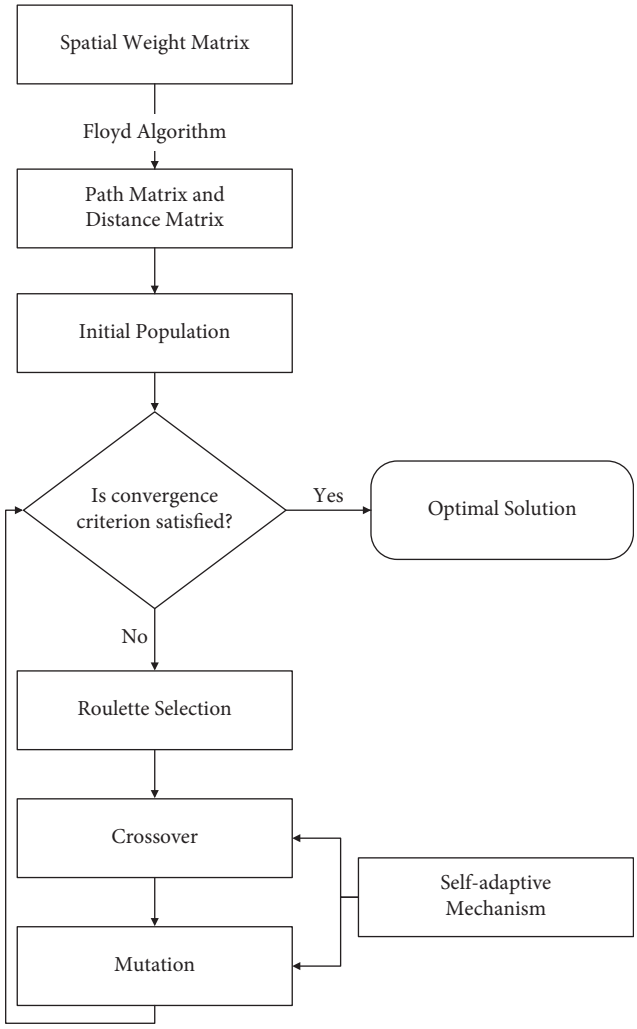


Figure 3: Flow chart of the adaptive genetic algorithm.

many problem described in Section 2. As depicted in Figure 3, the next step is to determine the coding of the initial population.

Assume that there are N units (chromosomes) in each generation, and each chromosome contains two parts: one is the feasible solution, while the other is the fitness value. Therefore, the information of a chromosome can be stored in a 1×2 Cell Array. According to the proposed model, the coding for the first element in the array can be designed as a $(\sum_{g \in G} c_g) \times 7$ matrix to store the information of a feasible solution. The formation of this matrix is generated by our proposed many-to-many algorithm, as seen in Table 1.

To be more specific, if the dump truck was assigned a one-way path because of its mileage limit, the one-way path solution can be expressed as a 1×7 matrix that contains the following information (see Table 2):

Similarly, a multiway path viable solution can be expressed in a $x \times 7$ matrix shown in Table 3. It is noticeable that the final retuning replenishment center should be the same as the departure one.

The second element of the array records the fitness value of each chromosome, which stores the sum of the total fuel consumption of all dispatched trucks in the corresponding feasible solution.

3.2. Roulette Select. Roulette selection is to determine some better individuals from the paternal chromosome based on the fitness value associated with the paternal chromosome. The smaller the fitness value, the more likely the corresponding chromosome is to be inherited. In addition, the elite strategy was adopted for the offspring generated after the above random operation, in which the best feasible solution of the parent gene was retained and the worst feasible solution of the offspring was replaced by the best feasible solution of the parent. The fitness value is defined as follows:

Table 1: Algorithm for generating initial population.

#pseudocode

Load node information;
Load distance information D and path information P from Floyd algorithm;
for each unit $n \in N$
  for each cutting field $k \in K$
    if $b_k - \sum_{j\in J}\sum_{i\in I_j}\sum_{g\in G}(X_{ijjk} + U_{ijgk}) \geq 0$
      if $a_j - \sum_{i\in I_j}\sum_{k\in K} X_{ijjk} \geq 0$
        Randomly select a replenishment center $j \in J$;
      end if
      if $c_g - \sum_{j\in J}\sum_{i\in I_j}\sum_{k\in K} Y_{jkg} \geq 0$
        Randomly select a filling field $g \in G$;
      End if
    End for
    Construct a one-way path solution S as shown in Table 1;
    While $\sum_{k\in K}\sum_{g\in G}[(d_{ijk} + d_{igk} + d_{ikj})\alpha + d_{ikg}\beta] < l_{ij}$
      $S(:,6) \leftarrow 0$, $S(:,2) \leftarrow x$ and append 1 row $\longrightarrow S$;
    End while
End for

Table 2: An example of a one-way path feasible solution.

1	0	2	4	6	8	1
Indexing the number of dispatched dump trucks	Indexing it is a one-way path solution	The serial number of the departure replenishment center	The serial number of the cutting field arrived at	The serial number of the filling field reached	The serial number of the returned replenishment center	The total distance the truck travels in the current dispatch

Table 3: An example of a multiway path feasible solution.

Indexing the number of dispatched dump trucks	Indexing how many runs the current truck travels	The serial number of the departure replenishment center	The serial number of the cutting field arrived at	The serial number of the filling field reached	0 if the dump trucks can continue shipment, otherwise indexing the serial number of the returned replenishment center	The distance the truck travels in each run
1	1	2	5	6	0	5
1	2	4	7	9	0	10
...						
1	x	3	2	5	2	14
2	1	2	5	6	0	5
2	2	4	7	9	0	10
...						
2	x	3	2	5	2	14
...						

$$f_n = \frac{1}{\sum_{j\in J}\sum_{i\in I_j}\sum_{k\in K}\sum_{g\in G}\left[\alpha\left(d_{\mathrm{ijk}}X_{\mathrm{ijjk}} + d_{\mathrm{igk}}U_{\mathrm{ijgk}} + d_{\mathrm{igj}}V_{\mathrm{ijgj}}\right) + \beta d_{\mathrm{ikg}}Y_{\mathrm{ijkg}}\right]}, \quad n \in N. \tag{13}$$

The probability of selecting each unit from the population is calculated as

$$p_n = \frac{f_n}{\sum_{n\in N} f_n}. \tag{14}$$

3.3. Crossover. In this model, the individual pairs of chromosomes can intersect with each other according to the crossover rate defined in (8), i.e., a certain rate of individual pairs can exchange their cutting fields and recalculate their overall mileage and fitness value thereafter. The crossover rate is self-adaptive according to the proportion p of the ratio of current optimal individuals to the whole population size. ω is the allowable peak crossover rate, and thus, $CR \in [0, \omega] \forall p \in (0, 1]$.

$$CR = \omega \log_{50}(50p). \tag{15}$$

Specifically, if the crossover rate is satisfied, the n^{th} unit will swap its cutting field with the $(n + N/2)^{th}$ unit. The algorithm is presented in Table 4.

Table 4: Algorithm for crossover.

#pseudocode

For each unit $n \in N/2$
if random $(0,1) \le CR$
$S(n,4) \rightleftarrows S(n+N/2,4)$;
end if
Recompute $f_n = 1/\sum_{j\in J}\sum_{i\in I_j}\sum_{k\in K}\sum_{g\in G}[\alpha(d_{ijk}X_{ijjk} + d_{igk}U_{ijgk} + d_{igj}V_{ijgj}) + \beta d_{ikg}Y_{ijkg}]$;
end for

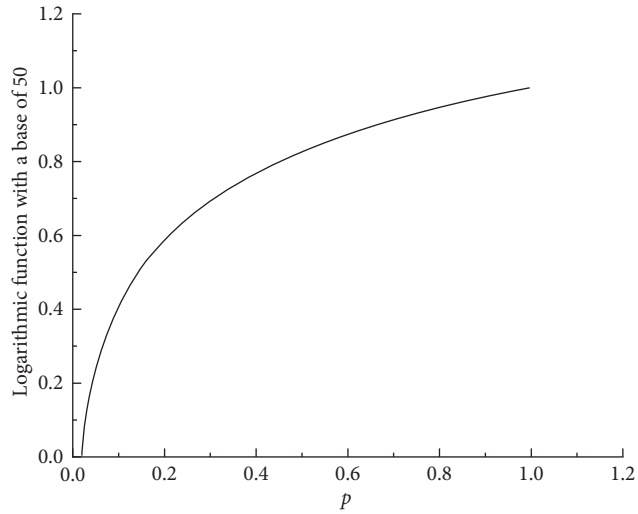


Figure 4: Logarithmic functions with a base of 50.

Table 5: Algorithm for mutation.

#pseudocode

For each unit $n \in N$
if random $(0,1) \le MR$
K_{remain} = find available cutting fields with remaining earth volume in K;
for $k_{\text{remain}} \in K_{\text{remain}}$
$S(n,4) \leftarrow k_{\text{remain}}$;
Recompute $f_n = 1/\sum_{j\in J}\sum_{i\in I_j}\sum_{k\in K}\sum_{g\in G}[\alpha(d_{\text{ijk}}X_{\text{ijjk}} + d_{\text{igk}}U_{\text{ijgk}} + d_{\text{igj}}V_{\text{ijgj}}) + \beta d_{\text{ikg}}Y_{\text{ijkg}}]$;
end for
$f_n^{\min k} = \min(f_n^{k_{\text{remain}}})$;
$S(n,4) \leftarrow \min k$;
end if
end for

Equation (8) is an increasing logarithmic function subjected to the proportion of the current optimal individuals. The shape of (8) is demonstrated in Figure 4. As indicated, as the number of optimal individual increases, the crossover rate will rise to increase the population diversity to avoid premature problem. It is noticeable that when p is larger than 0.02, the crossover rate and the mutation rate will become positive. p is the proportion of the current optimal individuals to the whole population. In this case, we assume that only when the proportion of optimal individuals is larger than 2%, the self-adaptive mechanism will work and start to manipulate the diversity of population. Otherwise, the algorithm will let the nature take its course.

3.4. Mutation. In the mutation stage, this paper mainly carries out the variation operations on the fourth element (cutting fields) of the first array. Specifically, this paper will replace the cutting field with another available one with the remaining earth volume required to be transported, which can achieve a shorter hauling distance from the replenishment centers to the filling fields on this single run. Therefore, it is necessary to ensure that the amount of earthwork in the cutting fields is not less than that in the filling fields when modeling, so that it can provide redundant earth volume in the cutting fields for variation. After mutation, the path length traveled by the dump trucks and the corresponding fitness value of the entire feasible solution will be updated again. The mutation rate is

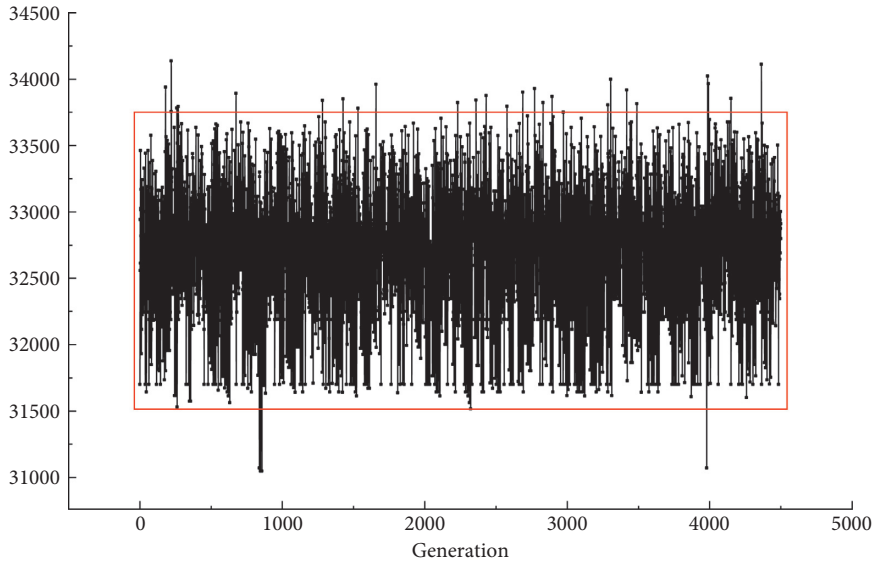


Figure 5: High diversity of optimal solutions at each generation.

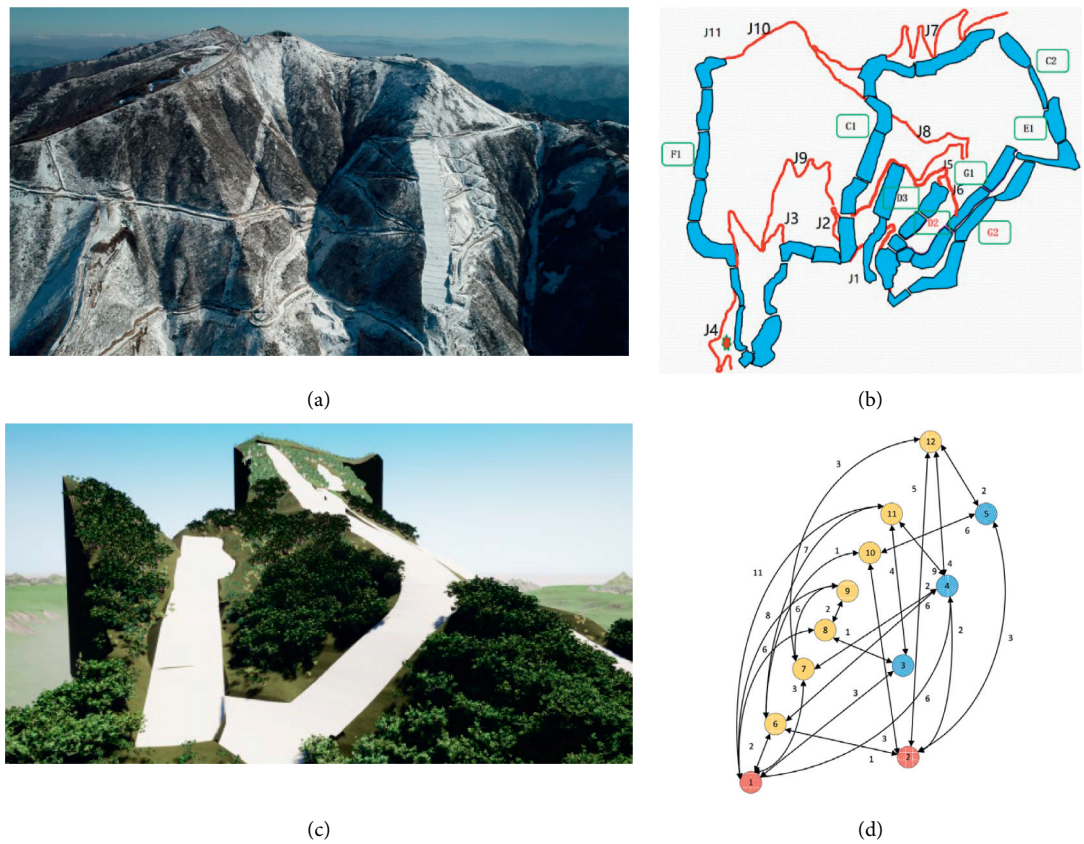


Figure 6: Skiing courses of the Beijing Winter Olympic Games Skiing Center in Yanqing District. (a) Real map of the skiing courses. (b) Distribution plan of the skiing courses. (c) BIM map of the G1 and D2 skiing courses. (d) Schematic diagram of the transportation network between the two depots, G1 and D2 skiing courses.

also controlled by a self-adaptive mechanism as (16), where χ is the maximum mutation rate, which ensures $MR \in [0, \chi] \forall p \in (0, 1]$.

$$MR = \chi \log_{50}(50p). \tag{16}$$

As displayed in Table 5, the mutation algorithm aims to find the best cutting field to achieve a lower fitness value for

Table 6: Node information.

	Replenishment centers		Cutting fields			Filling fields						
Node	1	2	3	4	5	6	7	8	9	10	11	12
Number of vehicles	40	30	835	785	732	280	260	330	300	250	334	224

Table 7: The spatial weight matrix.

	Destinations												
	0	—	3	6	—	2	3	6	8	—	11	—	0
	—	0	—	2	3	1	—	—	—	3	—	5	—
	3	—	0	—	-	-	-—	1	—	—	4	—	3
	6	2	—	0	—	6	2	—	—	—	9	4	6
	—	3	—	—	0	—	—	—	—	6	—	2	—
	2	1	—	6	—	0	—	—	—	1	7	—	2
Origins	3	—	—	2	—	—	0	—	6	—	—	3	3
	6	—	1	—	—	—	—	0	2	—	—	—	6
	8	—	—	—	—	—	6	2	0	—	—	—	8
	—	3	—	—	6	1	—	—	—	0	—	—	—
	11	—	4	9	—	7	—	—	—	—	0	—	11
	—	5	—	4	2	—	3	—	—	—	—	0	—
	0	—	3	6	—	2	3	6	8	—	11	—	0

each run. Once the mutation rate is satisfied, the algorithm will find all available cutting fields that can replace the original one and select the one with the least fitness value among all the alternatives.

The adaptive mechanism in the mutation stage is similar to that in the crossover stage, which aims to enlarge the diversity of the population and prevent a premature problem. As shown in Figure 5, during 4500 iterations, the optimal solution at each generation oscillates up and down greatly, ranging from 31,500 to 33,750, which means a high divergence population is maintained by this self-adaptive mechanism to avoid a premature problem. This adaptive mechanism provides more possibility to seek better solutions by jumping out local convergence.

3.5. Recursive Optimization. A better solution with the lower fitness value can be obtained by repeating sections 3.2 to 3.4 and iteratively updating the chromosomal information of each generation of the population until meeting the following convergence criterion:

$$\text{number of iterations} = \max(500, n_s), \tag{17}$$

where n_s is the number of iterations when the optimal solution remains unchanged for over 50 iterations.

4. Validation of the Proposed Algorithm

To validate the proposed algorithm, an earthwork allocation project on the G1 and D2 skiing courses of the Beijing Winter Olympic Games Skiing Center in 2022 is used. The schematic diagram of the original transportation network is shown in Figure 6, and the initial conditions of this case study are listed as follows:

(1) Overall excavation volume: 2352 m^3

(2) Overall filling volume: 1978 m^3

(3) Maximum carrying capacity of each dump truck: 10 m^3

(4) The transportation fuel consumption of empty load is 0.1 L/km

(5) The transportation fuel consumption of full loaded is 0.2 L/km

(6) The maximum capacity of fuel tank: 100 L

(7) Number of cutting fields: 3

(8) Number of filling fields: 7

(9) Number of replenishment centers: 2

The node information is listed in Table 6. The number of dump trucks in Table 6 provides the meanings, subject to the locations, respectively, as follows:

(1) Number of dump trucks each replenishment center can provide

(2) Number of dump trucks needed to transport all earth volume away from the cutting fields

(3) Number of dump trucks needed to transport the required earth volume to the filling fields

According to the distance information read from Figure 3, the spatial weight matrix can be constructed as Table 7.

According to equation (12), the distance information matrix and the path information matrix can be derived (see Tables 8 and 9).

The overall demand of all filling areas is 1978, and thus, the final optimal solution is a 1978×7 matrix. Each row represents the path information of a single run. Table 10 demonstrates the results of the optimal solution for the last 20 rows. It can be seen that 2 dump trucks are dispatched for multiple deliveries (e.g., the 63^{rd} dump truck has been assigned to carry out 12 transportation tasks from 1959 to 1970). Therefore, the optimal path can be read assisted by Table 9 generated from Floyd Algorithm. For instance, the 1959^{th} run travels from 1-4-11 as indicated in Table 10; however, the real hauling path from 1 to 11 is 1-6-2-4-11 by reading the path information matrix in Table 9. It is worth noting that 64 dump trucks are used to carry out the earthwork allocation task and achieve minimal fuel consumption so far.

Meanwhile, Figure 7 is the iterative optimizing process. The iteration operates 500 times in total and converges at the 440^{th} generation. The final optimal fuel consumption is 31352 L.

Table 8: The Floyd distance information matrix (1 = 5 km).

Node		Destinations											
Node		1	2	3	4	5	6	7	8	9	10	11	12
Origins	1	0	3	3	5	6	2	3	4	6	3	7	6
	2	3	0	6	2	3	1	4	7	9	2	8	5
	3	3	6	0	8	9	5	6	1	3	6	4	9
	4	5	2	8	0	5	3	2	9	8	4	9	4
	5	6	3	9	5	0	4	5	10	11	5	11	2
	6	2	1	5	3	4	0	5	6	8	1	7	6
	7	3	4	6	2	5	5	0	7	6	6	10	3
	8	4	7	1	9	10	6	7	0	2	7	5	10
	9	6	9	3	8	11	8	6	2	0	9	7	9
	10	3	2	6	4	5	1	6	7	9	0	8	7
	11	7	8	4	9	11	7	10	5	7	8	0	13
	12	6	5	9	4	2	6	3	10	9	7	13	0

Table 9: The Floyd path information matrix.

Node		Destinations											
Node		1	2	3	4	5	6	7	8	9	10	11	12
Origins	1	1	6	3	6	6	6	7	3	3	6	3	7
	2	6	2	6	4	5	6	4	6	6	6	6	12
	3	1	1	3	1	1	1	1	8	8	1	11	1
	4	2	2	2	4	2	2	7	2	7	2	11	12
	5	2	2	2	2	5	2	12	2	12	2	2	12
	6	1	2	1	2	2	6	1	1	1	10	11	2
	7	1	4	1	4	12	1	7	1	9	1	1	12
	8	3	3	3	3	3	3	3	8	9	3	3	3
	9	8	8	8	7	7	8	7	8	9	8	8	7
	10	6	6	6	6	6	6	6	6	6	10	6	6
	11	3	6	3	4	6	6	3	3	3	6	11	4
	12	7	2	7	4	5	2	7	7	7	2	4	12

Table 10: Optimal solution of the last 20 rows.

Indexing the number of dispatched dump trucks	Indexing how many runs of current truck travels	The serial number of the departure replenishment center	The serial number of the cutting field arrived at	The serial number of the filling field reached	0 if the dump trucks can continue shipment, otherwise indexing the serial number of the returned replenishment center	The distance the truck travels in each run
63	1959	1	4	11	0	14
63	1960	11	4	9	0	17
63	1961	9	4	9	0	16
63	1962	9	4	9	0	16
63	1963	9	4	9	0	16
63	1964	9	4	9	0	16
63	1965	9	4	9	0	16
63	1966	9	4	9	0	16
63	1967	9	4	9	0	16
63	1968	9	4	9	0	16
63	1969	9	4	9	0	16
63	1970	9	4	9	1	22
64	1971	1	4	9	0	13
64	1972	9	5	9	0	22
64	1973	9	5	9	0	22
64	1974	9	5	9	0	22
64	1975	9	5	9	0	22
64	1976	9	5	9	0	22
64	1977	9	5	9	0	22
64	1978	9	4	9	1	22

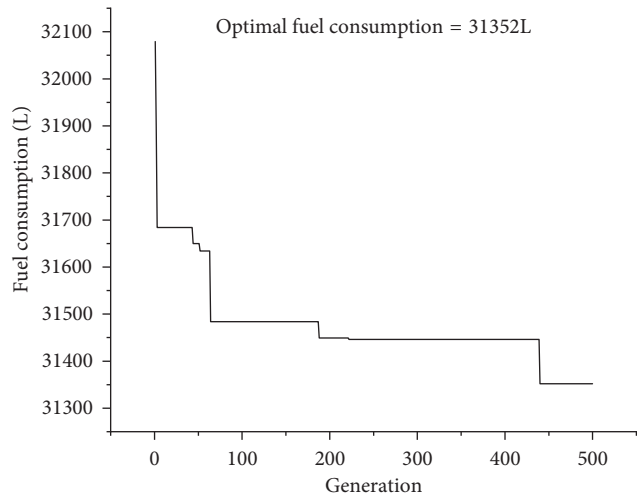


Figure 7: Recursive optimizing process of earthwork allocation.

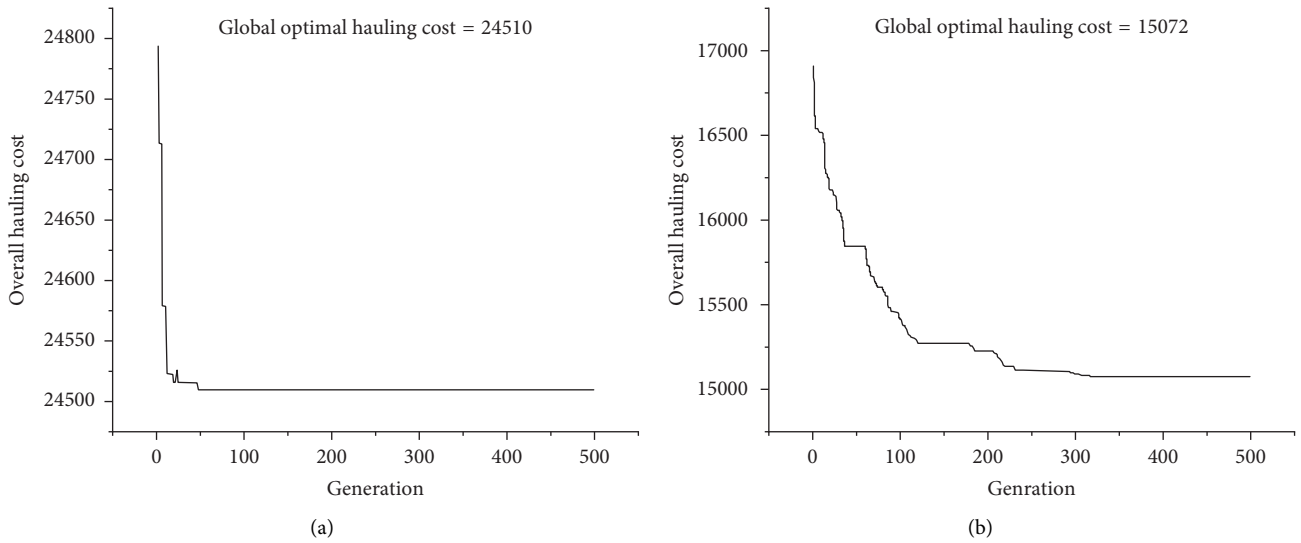


Figure 8: Comparative study between [6] and our method. (a) The former work with two-way path planning. (b) Our method with multiway path planning.

Table 11: Variables and explanation.

Variables	Explanation
J	Number of replenishment centers
I_j	Number of dump trucks in the j^{th} replenishment center
K	Number of cutting fields
G	Number of filling fields
d_{ijk}	The hauling distance of the i^{th} dump truck from the j^{th} replenishment center to the k^{th} cutting field
d_{ikg}	The hauling distance of the i^{th} dump truck from the k^{th} cutting field to the g^{th} filling field
d_{igk}	The hauling distance of the i^{th} dump truck from the g^{th} filling field to the k^{th} cutting field
d_{igj}	The hauling distance of the i^{th} dump truck from the g^{th} filling field to the j^{th} replenishment center
α	The average fuel consumption per unit distance of each dump truck at no load
β	The average fuel consumption per unit distance of each dump truck at full load
X_{ijjk}	Binary decision variable: whether the i^{th} dump truck at the j^{th} replenishment center travels from the j^{th} replenishment center to the k^{th} cutting field
Y_{ijkg}	Binary decision variable: whether the i^{th} dump truck at the j^{th} replenishment center travels from the k^{th} cutting field to the g^{th} filling field
U_{ijgk}	Binary decision variable: whether the i^{th} dump truck at the j^{th} replenishment center travels from the g^{th} filling field to the k^{th} cutting field

Table 11: Continued.

Variables	Explanation
V_{ijgj}	Binary decision variable: whether the i^{th} dump truck at the j^{th} replenishment center travels from the g^{th} filling field to the j^{th} replenishment center
a_j	Total number of available dump trucks at the j^{th} replenishment center
b_k	Total number of fully loaded dump trucks needed to haul all earthwork in the k^{th} cutting field
c_g	Total number of fully loaded dump trucks needed to fill the g^{th} filling field
l_{ij}	The maximum fuel capacity of the i^{th} dump truck at the j^{th} replenishment center
N	Population size
f_n	The fitness value of the n^{th} unit
p_n	The probability of selecting each unit from the population
ω	The allowable peak crossover rate
CR	The crossover rate
p	The ratio of current optimal individuals to the whole population size
χ	The maximum mutation rate
MR	The mutation rate
n_s	The number of iterations when the optimal solution remains unchanged for over 50 iterations

To further validate the performance of our proposed algorithm, a comparative study has been made by comparing our method with former work in [6]. The result in Figure 8 demonstrated that our method showed a dominant advantage over the past work, and roughly, a performance boost of 38% is achieved. The global optimal hauling cost in Figure 8 is related to fuel consumption. Thus, a slight change to the algorithm is made to accord with the optimization objective.

5. Conclusions

This paper constructs a complex earthwork allocation path-planning modeling based on MD-TEVRP by considering the multiple runs of dump trucks between the cutting fields and the filling fields. A well-designed genetic algorithm based on this model is proposed to achieve the following contributions:

(1) This paper provides a multiway transportation solution to a typical MD-TEVRP, which achieves many-to-many recursive pairing in earthwork allocation path planning.

(2) The proposed self-adaptive mechanism can manipulate the diversity of the generated populations to prevent the premature problem of GA and achieve a lower fitness value.

(3) The optimization method shows a dominant advantage over the past work by increasing approximately 38% of performance. The proposed method also provides guidance for the construction design of the Skiing Courses of Beijing Winter Olympic Games Skiing Center in Yanqing, Beijing.

The implication of variables is listed in Table 11.

Data Availability

The data presented in this study are available on request to the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. H. Kang, J. W. Seo, and K. G. Baik, "3D-GIS based earthwork planning system for productivity improvement," in *Proceedings of the American Society of Civil Engineers Construction Research Congress 2009*, pp. 151–160, Seattle, WA, USA, April 2009.

[2] Y. Ji, A. Borrmann, E. Rank, F. Seipp, and S. Ruzika, "Mathematical modeling of earthwork optimization problems," in *Proceedings of the International Conference on Computing in Civil and Building Engineering (ICCCBE)*, Nottingham, UK, January 2010.

[3] M. Parente, P. Cortez, and A. G. Correia, "An evolutionary multi-objective optimization system for earthworks," *Expert Systems with Applications*, vol. 42, pp. 6674–6685, 2015.

[4] R. Burdett, E. Kozan, and R. Kenley, "Block models for improved earthwork allocation planning in linear infrastructure construction," *Engineering Optimization*, vol. 47, no. 3, pp. 347–369, 2015.

[5] A. M. Ahmad, "Spline technique for modeling roadway profile to minimize earthwork cost," *Journal of Industrial and Management Optimization*, vol. 5, no. 2, pp. 275–283, 2009.

[6] N. Deng, X. Li, and Y. Su, "Optimization of earthwork allocation path as vehicle route problem based on genetic algorithm," in *Proceedings of the 2020 2nd International Conference on Civil Architecture and Energy Science (CAES 2020)*, Changchun, China, March 2020.

[7] K. Miao, X. Sun, and L. Li, "A roadbed earthwork allocation model based on ACO algorithm," *Applied Mechanics and Materials*, vol. 44-47, pp. 3483–3486, 2011.

[8] R. L. Burdett and E. Kozan, "An integrated approach for earthwork allocation, sequencing and routing," *European Journal of Operational Research*, vol. 238, no. 3, pp. 741–759, 2014.

[9] G. Desaulniers, J. Desrosiers, A. Erdmann, M. M. Solomon, and F. Soumis, "9. VRP with pickup and delivery," *The vehicle routing problem*, vol. 9, pp. 225–242, 2002.

[10] G. Kim, Y.-S. Ong, C. K. Heng, P. S. Tan, and N. A. Zhang, "City vehicle routing problem (city VRP): a review," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 1654–1666, 2015.

[11] G. B. Dantzig and J. H. Ramser, "The truck dispatching problem," *Management Science*, vol. 6, no. 1, pp. 80–91, 1959.

[12] E. D. Taillard, "A heuristic column generation method for the heterogeneous fleet VRP," *RAIRO - Operations Research*, vol. 33, no. 1, pp. 1–14, 1999.

[13] J. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Application to Biology. Control And Artificial Intelligence*, MIT Press, Cambridge, MA, USA, 1975.

[14] B. M. Baker and M. A. Ayechew, "A genetic algorithm for the vehicle routing problem," *Computers & Operations Research*, vol. 30, no. 5, pp. 787–800, 2003.

[15] A. K. M. Masum, M. Shahjalal, F. Faruque, and I. Sarker, "Solving the vehicle routing problem using genetic algorithm," *International Journal of Advanced Computer Science and Applications*, vol. 2, pp. 126–131, 2011.

[16] S.-I. Hwang, J.-H. Son, and S.-H. Lee, "Development of scheduling model for earth work using genetic algorithm," *KSCE Journal of Civil Engineering*, vol. 18, no. 6, pp. 1618–1624, 2014.

[17] R. Montemanni, L. M. Gambardella, A. E. Rizzoli, and A. V. Donati, "Ant colony system for a dynamic vehicle routing problem," *Journal of Combinatorial Optimization*, vol. 10, no. 4, pp. 327–343, 2005.

[18] M. Reed, A. Yiannakou, and R. Evering, "An ant colony algorithm for the multi-compartment vehicle routing problem," *Applied Soft Computing*, vol. 15, pp. 169–176, 2014.

[19] B. Yu, Z.-Z. Yang, and J.-X. Xie, "A parallel improved ant colony optimization for multi-depot vehicle routing problem," *Journal of the Operational Research Society*, vol. 62, no. 1, pp. 183–188, 2011.

[20] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proceedings of the MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, pp. 39–43, Nagoya, Japan, October 1995.

[21] Y. Marinakis, M. Marinaki, and G. Dounias, "A hybrid particle swarm optimization algorithm for the vehicle routing problem," *Engineering Applications of Artificial Intelligence*, vol. 23, no. 4, pp. 463–472, 2010.

[22] F. Belmecheri, C. Prins, F. Yalaoui, and L. Amodeo, "Particle swarm optimization algorithm for a vehicle routing problem with heterogeneous fleet, mixed backhauls, and time windows," *Journal of Intelligent Manufacturing*, vol. 24, no. 4, pp. 775–789, 2013.

[23] C. Xiutong and L. Lu, "Study on earthwork allocation method based on modified particle swarm optimization," *Journal of Hydroelectric Engineering*, vol. 29, pp. 68–72, 2010.

[24] M. Dorigo, "Optimization, Learning and Natural Algorithms," PhD Thesis, Politecnico di, Milano, Italy, 1992.

[25] I. H. Osman, "Metastrategy simulated annealing and tabu search algorithms for the vehicle routing problem," *Annals of Operations Research*, vol. 41, no. 4, pp. 421–451, 1993.

[26] L. Wei, Z. Zhang, D. Zhang, and S. C. H. Leung, "A simulated annealing algorithm for the capacitated vehicle routing problem with two-dimensional loading constraints," *European Journal of Operational Research*, vol. 265, no. 3, pp. 843–859, 2018.

[27] F. Glover, "Tabu search-Part I," *ORSA Journal on Computing*, vol. 1, no. 3, pp. 190–206, 1989.

[28] G. Barbarosoglu and D. Ozgur, "A tabu search algorithm for the vehicle routing problem," *Computers & Operations Research*, vol. 26, no. 3, pp. 255–270, 1999.

[29] D. Schermer, M. Moeini, and O. Wendt, "A hybrid VNS/Tabu search algorithm for solving the vehicle routing problem with drones and en route operations," *Computers & Operations Research*, vol. 109, pp. 134–158, 2019.

[30] X. Li and M. Yin, "Hybrid differential evolution with biogeography-based optimization for design of a reconfigurable antenna array with discrete phase shifters," *International Journal of Antennas and Propagation*, vol. 2011, Article ID 685629, 12 pages, 2011.

[31] Y. Xiao, J. Jiao, J. Pei, K. Zhou, and X. Yang, "A multi-strategy improved ant colony algorithm for solving traveling salesman problem," in *Proceedings of the 2018 5th International Conference on Advanced Composite Materials and Manufacturing Engineering*, Yunnan, China, June 2018.

[32] A. ElSaid, F. E. Jamiy, J. Higgins, B. Wild, and T. Desell, "Using ant colony optimization to optimize long short-term memory recurrent neural networks," in *Proceedings of the Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 13–20, Kyoto, Japan, July2018.

[33] H. Cai, X. Lu, T. Du, Y. Wang, S. Xia, and D. Zhang, "A survey of artificial intelligence algorithm in power system Applications," in *Proceedings of the 2019 IEEE 3rd International Electrical and Energy Conference (CIEEC)*, pp. 1902–1906, Beijing, China, September2019.

[34] X. Wang and K. Yang, "Economic load dispatch of renewable energy-based power systems with high penetration of large-scale hydropower station based on multi-agent glowworm swarm optimization," *Energy Strategy Reviews*, vol. 26, Article ID 100425, 2019.

[35] N. Ben Guedria, "An accelerated differential evolution algorithm with new operators for multi-damage detection in plate-like structures," *Applied Mathematical Modelling*, vol. 80, pp. 366–383, 2020.

[36] W. Deng, J. Xu, Y. Song, and H. Zhao, "Differential evolution algorithm with wavelet basis function and optimal mutation strategy for complex optimization problem," *Applied Soft Computing*, vol. 100, Article ID 106724, 2021.

[37] G. A. Vignaux and Z. Michalewicz, "A genetic algorithm for the linear transportation problem," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 2, pp. 445–452, 1991.

[38] X. Li and S. Ma, "Multi-objective memetic search algorithm for multi-objective permutation flow shop scheduling problem," *IEEE access*, vol. 4, pp. 2154–2165, 2016.

[39] X. Li, X. Zhang, M. Yin, and J. Wang, "A genetic algorithm for the distributed assembly permutation flowshop scheduling problem," in *Proceedings of the 2015 IEEE Congress on Evolutionary Computation (CEC)*, pp. 3096–3101, Sendai, Japan, May 2015.

[40] S. Roy and W. Crossley, "Hybrid multi-objective combinatorial optimization technique with improved compatibility between GA and gradient-based local search," in *Proceedings of the 12th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference and 14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, p. 5568, Indianapolis, IN, USA, September 2012.

[41] L. Zhou, R. Baldacci, D. Vigo, and X. Wang, "A multi-depot two-echelon vehicle routing problem with delivery options arising in the last mile distribution," *European Journal of Operational Research*, vol. 265, no. 2, pp. 765–778, 2018.

[42] R. W. Floyd, "Algorithm 97: shortest path," *Communications of the ACM*, vol. 5, no. 6, p. 345, 1962.

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 6032899, 11 pages
https://doi.org/10.1155/2022/6032899

Research Article

Network Design Algorithm Implementation for Resilient Transportation System under Continuous Risk Perturbation with Big Data Analysis

Hongxiao Wang,[1,2] **Qiang Li**,[3] **and Sang-Bing Tsai**[4]

[1]*College of Transportation Engineering, Chang'an University, Xi'an 710061, China*
[2]*Department of Mechanical and Traffic Engineering, Ordos Institute of Technology, Ordos 017010, China*
[3]*Public Security Bureau Traffic Management Detachment of Ordos, Ordos 017010, China*
[4]*Regional Green Economy Development Research Center, School of Business, WUYI University, Nanping, China*

Correspondence should be addressed to Hongxiao Wang; whxtc06@163.com

Received 27 October 2021; Revised 18 November 2021; Accepted 1 December 2021; Published 4 January 2022

Academic Editor: Xiangtao Li

With the rapid economic development and urbanization process accelerating, motor vehicle ownership in large cities is increasing year by year; urban traffic congestion, parking difficulties, and other problems are becoming increasingly serious; in ordinary daily life, continuous risk of disturbance, having a flexible transportation system network is more able to alleviate daily congestion in the city, and the main thing about flexible transportation network is its algorithm. It is worth noting that congestion in many cities is generally reflected in the main roads, while many secondary roads and branch roads are underutilized, and the limited road resources in cities are not fully utilized. As an economic and effective road traffic management measure, one-way traffic can balance the spatial and temporal distribution of traffic pressure within the road network, make full use of the existing urban road network capacity, and solve the traffic congestion problem. Therefore, it is of great theoretical and practical significance to develop a reasonable and scientific one-way traffic scheme according to the characteristics of traffic operation in different regions. Based on the fixed demand model, the influence of traffic demand changes is further considered, the lower-level model is designed as an elastic demand traffic distribution model, the excess demand method is used to transform the elastic demand problem into an equivalent fixed demand problem based on the extended network, and the artificial bee colony algorithm based on risk perturbation is designed to solve the two-level planning model. The case study gives a one-way traffic organization optimization scheme that integrates three factors, namely, the average load degree overload limit of arterial roads, the detour coefficient, and the number of on-street parking spaces on feeder roads, and performs sensitivity analysis on the demand scaling factor.

1. Introduction

With the rapid economic development, motor vehicle ownership in large cities has increased year by year, and traffic problems have become a common problem faced by countries around the world in the process of urban development. In a megacity, for example, motor vehicle ownership reached 6.084 million in 2018, an increase of 3% over the previous year; among them, private motor vehicle ownership reached 4.894 million, an increase of 2.9% over last year [1]. At present, traffic congestion and parking difficulties in large cities are becoming increasingly serious, mainly because the limited traffic resources cannot meet the growing traffic demand of the public and the consequent increase in travel costs, traffic accident rates are rising year by year, and many other problems not only reduce the overall operational efficiency of the urban transportation system but also restrict the rapid development of the economy and the city, which has continuous risk disturbance in terms of traffic. By the end of 2018, the total road mileage in its urban area was 6203 km, of which 390 km was urban expressways, 998 km was urban trunk roads, 632 km

was urban secondary roads [2], and 4183 km was branch roads and below; the total road area was 103.28 million m^2. the average traffic index in the central urban area during peak hours was 5.5. The road traffic congestion in the evening peak was higher than that in the morning peak, and the annual average road traffic index in the morning peak was 5.1, while that in the evening peak was 5.1. The average annual road traffic index is 5.1 in the morning peak and 6.0 in the evening peak, and the average duration of congestion (including severe congestion and moderate congestion) is 2 h 50 min. How to effectively improve urban traffic operations has gradually become a focus of widespread attention from the government and the community [3].

The application of big data is a very good tool for traffic management departments. The current solution for traffic congestion is to build new roads, which costs a lot. Based on big data, the method of rationally diverting vehicles and alleviating the pressure of traffic is proposed. To solve the traffic problem, the government has invested a lot of money in the improvement and expansion of roads. 35 km of new urban roads was added in its city in 2018, and the total mileage reached 6395 km. To solve the traffic congestion problem, governments at all levels have formulated many measures, such as motor vehicle restrictions and increased parking fees, which have alleviated traffic congestion to some extent, but have not fundamentally solved the traffic congestion problem. Traffic congestion is still serious during holidays and morning and evening rush hours, and the urban traffic problem needs to be solved urgently. Applying multiple traffic assignment models to urban road traffic network design is beneficial to improve the accuracy of investment decisions [4]. The study of traffic network design is based on system engineering as the starting point, and modern optimization methods are used to find the best investment plan for expanding or building new roads, maximize the effect of limited funds, and optimize certain indicators of the studied traffic network. In previous studies, experts and scholars have focused on proposing different types of traffic network indicators, considering more traffic network design constraints, or proposing more effective solution algorithms, but less on the relationship between traffic network design solutions and traveler behavior. Traveler behavior has a significant impact on the outcome of the design solution, and the adoption of a traffic assignment model that can more accurately describe the traffic distribution in the studied traffic network is a key point to ensure the success of traffic network design and urban transportation planning [5].

For urban traffic problems, a lot of researches have been conducted in the following two aspects: firstly, from the perspective of the flexible transportation system, rational planning of urban space to improve the supply of transportation facilities; secondly, from the perspective of demand management, balancing transportation supply and demand while reducing transportation demand. It has been proved that the investment in traffic infrastructure cannot fundamentally solve the traffic problem, and the increase of traffic supply will induce new travel demand, which will lead to a vicious circle of "increase supply-stimulate traffic demand-increase supply again," and road reconstruction and expansion and urban development and land use are closely related and cannot be increased indefinitely.

Therefore, a two-pronged approach is needed to solve the traffic problem, focusing on both infrastructure development and the flexible use of traffic demand management measures. At the same time, it is worth noting that congestion in many cities is generally reflected in the main roads, while many secondary and feeder roads are underutilized, and the limited road resources in cities are not fully utilized [6].

One-way traffic, as an economic and effective road traffic management measure, is one of the traffic organization methods often used in traffic microcirculation design, which can balance the spatial and temporal distribution of traffic pressure within the road network and improve the service level of bottleneck sections and intersections.

Improving the design of a flexible traffic system network, making the traffic network more flexible, can solve the problem of traffic congestion to a certain extent. Scientific and reasonable organization of one-way traffic can make full use of the existing urban road network capacity and effectively alleviate urban traffic congestion. To solve the traffic congestion problem, exploring scientific and comprehensive one-way traffic organization scheme optimization process has become a hot spot for scholars at home and abroad to study how to develop a reasonable one-way traffic organization optimization scheme to achieve the expected effect of alleviating traffic congestion and other problems, which has important theoretical and practical significance [7].

2. Related Work

With the famous deterministic user equilibrium (UE) model proposed in the literature, the study of traffic assignment has entered a completely new phase. A mathematical model of the user balance principle was proposed in the literature, and the model was successfully solved using the Frank-Wolfe algorithm in the literature and others. The deterministic user equilibrium model assumes that the traveler is fully aware of the road conditions and makes a completely correct path choice, which is not in line with reality. There is some error between the traveler's estimate of the path impedance and the actual value. Based on this, the literature proposes a stochastic user equilibrium (SUE) model, where the SUE still satisfies the user equilibrium principle, but the traveler selects the path according to the minimum impedance he/she perceives [8]. The literature proposed an equivalent mathematical planning model for SUE, which includes other models for path selection that satisfy the translation-invariant distribution in addition to the Logit and Probit models. The literature designed the Method of Successive Averages (MSAs) to successfully solve the SUE model, which is widely used in traffic assignment studies to subdivide travelers into four types according to the finite rationality of travelers' path selection and the variability of risk sensitivity coefficients, establish the prospect values of each type of travelers, and construct a stochastic user equilibrium model. The results of the algorithm show that the setting of the

cumulative prospect theory parameters has an important influence on the results of traffic assignment. The literature proposed a stochastic user balance model with exponential pheromone update strategy and solved the model using the successive averaging (MSA) and ant colony algorithm and finally compared the output results and algorithm sensitivity analysis. Literature [9] constructed a hierarchical Logit-based SUE model by considering the path selection and mode choice behaviors of travelers and designed an improved direction search algorithm to solve it.

The most basic one is the elastic demand-based user equilibrium model, which can be solved to obtain the roadway flow and OD demand satisfying the user equilibrium principle, and the OD demand must also satisfy the defined demand function, a multicategory, multicriteria transportation network equilibrium model considering elastic demand, and the demand function usually depends on the negative utility of all categories among ODs. An efficient algorithm for computing this model is proposed in the literature [10]. A multiuser stochastic user equilibrium model based on the elastic demand allocation criterion is constructed, an efficient path search algorithm is designed to solve it, and numerical simulations are performed by arithmetic examples. Someone constructs an elastic demand combination model with the coexistence of user balance and stochastic user balance and designs a diagonalization and MSA combination algorithm. For the elastic demand asymmetric stochastic user balance problem, a Variational Inequality (VI) model is constructed through Probit-based stochastic network loading, a Monte Carlo simulation method is proposed to solve the model, and a path-based Gradient Projection (GP) algorithm is used to solve the elastic demand traffic assignment problem of the large-scale road network, the elastic demand user equilibrium assignment problem is studied under the road section capacity constraint, the excess demand model is used to reconstruct the elastic demand model into a fixed demand model, an approximate solution is given, the three factors of path time and reliability are integrated, dynamic travel cost is generalized, and users are simulated to make a path with the criterion of the minimum estimated dynamic cost. The stochastic user equilibrium model is established under the elastic demand condition, and the continuous average algorithm (MSA) for solving the model is given. The convergence of the fluctuation of road section travel time with the change of road length under the elastic demand condition is studied [11], the travel process of vehicles in the road network is simulated by using MITSIM simulation, and the research results show that the road section travel time converges with the increase of road length under the elastic demand condition.

The characteristics and stability conditions of the dynamic evolution of traffic flow are analyzed, and the Logit-based SUE problem under elastic demand is studied, which not only models the route choice of travelers according to the SUE principle but also estimates the traffic demand by considering the influence of the expected travel time of the road section on the traffic demand, and constructs a two-layer planning model based on the Logit-SUE problem for The problem of difficulty in solving the elastic demand traffic assignment model is designed to achieve the user equilibrium state through adaptive adjustment in the iterations by an effective algorithm. The traffic paradox under different equilibrium conditions of fixed and elastic traffic demand, including user equilibrium and stochastic user equilibrium, is studied comprehensively, and the average travel [12] cost between OD pairs is chosen as the main index to discern whether there is a traffic paradox. Finally, the effects of travelers' perceptual errors and travel cost sensitivity on the occurrence of elastic transportation systems are analyzed.

3. Algorithm Design

Tourists are the main body of tourism and the basic element of the three major elements of tourism. Without tourists, natural tourism cannot be realized. Tourists, literally interpreted, are tourists; that is, people engaged in tourism activities. The topological method of super transportation network based on the urban road network, urban bus line network, and rail line network is used to combine the three networks into one through virtual interchange nodes and virtual interchange sections to form an urban super transportation network. A multinomial random path selection algorithm based on the super transportation network is proposed and its parameters are calibrated using the great likelihood method. In the upper model, the objective is to maximize the social benefits of the transportation system considering the construction cost of the hub [13]. In the lower layer model, the user balance model of the super network is used to realize the nonset count selection of passenger travel paths and traffic assignment distribution, and a forbidden search algorithm is proposed to solve the interchange hub layout optimization model. Traffic flow distribution is an important part of traffic network design and traffic management, which is to distribute the predicted traffic volume to each road section according to certain criteria based on the road network structure, impedance, and other conditions and, further, find the traffic volume and impedance of each road section and use it to evaluate the traffic network condition. According to whether the traffic demand varies with the degree of network congestion, it can be divided into fixed demand and elastic demand traffic allocation models. Elastic demand means that when congestion increases between an OD pair in the road network, some travelers will change their travel time or cancel their trips, and the traffic demand between ODs will decrease. Elastic demand traffic assignment is an extension of the study of fixed demand, and many research results have been obtained in recent years. The most basic one is the elastic demand user equilibrium (EDUE) model, which can be solved to obtain the roadway flow and OD demand satisfying the Wardrop user equilibrium principle, and the OD demand must also satisfy the defined demand function.

The two-level planning model constructed in the literature belongs to the nonlinear planning problem, which is the key problem in constructing the flexible transportation algorithm, and also a multiobjective optimization problem. Generally speaking, there is no unique optimal solution for

the multiobjective problem, and there is a game between multiple objectives, which cannot reach the optimum at the same time, so this paper transforms the multiobjective problem into a single-objective problem by weighting coefficients. The solution algorithms of two-level nonlinear programming mainly include descent algorithm, artificial bee colony algorithm, and genetic algorithm. The descent algorithm is often easy to get the local optimal solution, and the intelligent algorithm is more suitable for the solution of the two-level programming model [14]. Considering the characteristics of discrete decision variables in the upper-level model, this paper designs the artificial bee colony algorithm to solve the two-level planning model for one-way traffic organization optimization. The swarm algorithm is an intelligent algorithm that imitates the behavior of a bee colony to find a good food source. By comparing the advantages and disadvantages of the solution sets generated randomly and with objectives, the global optimal solution is finally obtained by an individual local search of the hiring bees, observation bees, and detection bees, and the convergence speed is faster. Compared with the genetic algorithm, the artificial bee colony algorithm is designed with neighborhood search and thus has a faster convergence speed and better local optimization search capability.

Using the real number coding method, firstly, the branch road sections in the one-way traffic organization area are numbered, B is the number of branch road sections, each section corresponds to a coding position, the coding value of the corresponding position is determined according to the direction of travel of the section, and the coding value range is {0, 1, −1}. All branch road sections in the network are sequentially encoded, forming a vector of length B. Figure 1 shows the encoding diagram of the solution [15].

In the swarm algorithm, a new solution is generated by a neighborhood search operation based on the current solution. In this paper, we propose a probabilistic reverse-order-and-swap neighborhood search strategy, which generates two sets of random numbers in the range of 1 to B without overlap, extracts two sequences based on the range of random numbers inside each solution, reverses the sequences separately, and then swaps the positions of the two sequences in the code. See Figure 1.

For any solution y generated by the domain search, it is brought into the lower layer to find out the traffic flow and load degree of the corresponding road section in equilibrium by the SUE model, and then, the objective function value of the upper layer problem is calculated. The upper layer problem is a minimization problem, and the fitness function is designed according to the objective function $Z(y)$, for any given coding sequence. corresponding to the solution y, the fitness value is defined as follows.

$$\text{Fitness}(y) = Z_{\max} - Z(y), \quad Z(y) > Z_{\max}. \tag{1}$$

The common food source selection strategies used in swarming algorithms are roulette selection and tournament selection methods.

The fitness value calculates the sharing probability of the food source under the selected strategy to ensure that the good food source is passed on with maximum probability. In this paper, the roulette selection method is used for the selection of a good food source, i.e., the optimal feasible solution, through which the strategy ensures that good individuals are passed on to the next generation with maximum probability without being lost in the evolutionary process and accelerates the convergence of the algorithm [16].

The basic steps for the solution of the artificial bee colony algorithm designed in this paper are given as follows.

Step 1: Initial solution generation

Randomly generate n initial solutions $\{Yi\}(i = 1, 2, 3, \ldots, n)$; let iter = 0 and neighborhood search count $Li = 0(i = 1, 2, 3, \ldots, n)$; set the upper limit of the number of iterations of neighborhood search Limit Num and the maximum number of iterations of the algorithm Max Iter Num.

Step 2: Hire bee stage

A neighborhood search is performed for each solution Yi, in turn, to generate a new solution, Yi, comparing the magnitude of the fitness function value of each solution in the original solution set X with that of the new solution generated by the corresponding search.

Step 3: Observation bee stage

The roulette selection method is used to randomly select one solution $_{Yi}$ for neighborhood search and generate a new solution Fitness(X) > Fitness(X), then replace Y with X, and update the solution in X so that $Li = 0$; conversely, let $Li = Li + 1$; repeat the process n times.

Step 4: Scouting bee stage

For any solution Yi, if the number of neighborhood searches Li = Limit Num, the old solution is discarded and a new solution Yi is randomly generated instead.

3.1. Resilient Equilibrium Model Solving Based on an Elastic Traffic Network Algorithm. The algorithms currently used to solve the SUE model based on discrete choice include Dial algorithm, iterative weighting (MSA), and Clark approximation. The Clark approximation is only applicable to the solution of the Probit model, and the Dial algorithm is only applicable to the solution of the Logit model. Sheffi and Powell first proposed the Method of Successive Averages (MSAs) in 1982 for solving the mathematical programming model of the SUE equivalent model [17].

Successive averaging is a cyclic allocation method between the balanced allocation method and the incremental allocation method, also known as the successive averaging method or the quadratic weighted averaging method. The basic idea of the algorithm is to gradually approach the equilibrium solution by adjusting the loading flow rate of each road segment. Each iteration is predetermined in steps, and the traffic flow is allocated by a random loading method (such as Dial algorithm), and the cycle is continued until the difference of the traffic volume allocated to each road section

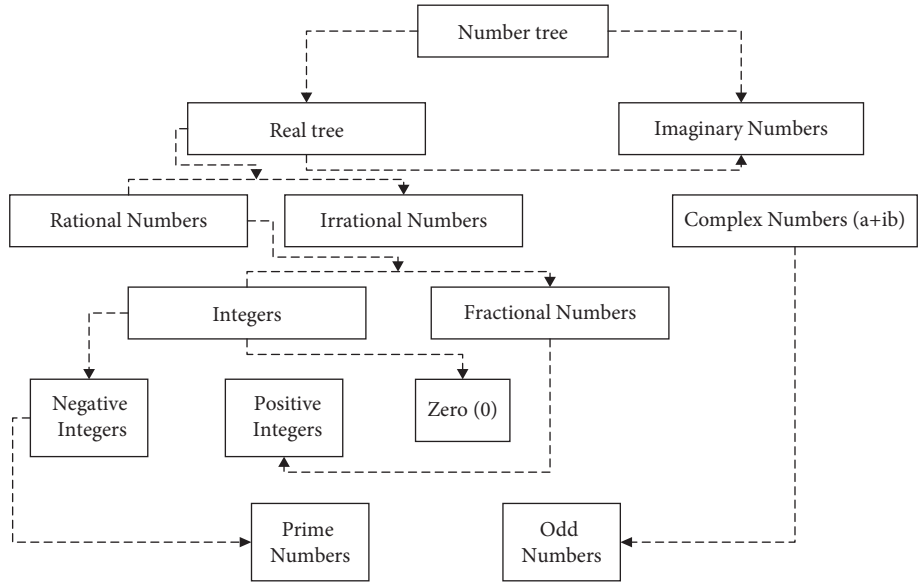


Figure 1: Principles of number sequence conversion.

in two iterations is less than the given error limit; then, the equilibrium solution is obtained. In response to the shortcomings of single path assignments, scholars have proposed random loading assignment methods, which are mainly divided into two types.

Decoding is a process of using a specific method to restore the digital to the content it represents or to convert electrical pulse signals, optical signals, radio waves, etc. into the information, data, etc. it represents. Decoding is the process by which the receiver restores the received symbols or codes to information, which corresponds to the encoding process.

Logit Method and Probit Method. To solve the Logit-SUE model proposed in this paper, the solution algorithm of the Logit model is highlighted here.

The Dial algorithm enables efficient implementation of the Logit model on the network in the following steps [18].

Step 1: Determine the effective paths and road sections. Calculate the minimum impedance from the starting point r to all nodes, denoted as $r(i)$; calculate the minimum impedance from all nodes to the end point s, denoted as $s(i)$; denote O_i as the set of section endpoints with the starting point I; define D_i as the set of section start points with the endpoint i.

$$L(i, j) = \frac{\delta y}{\delta x}, \sum_{i=i}^{n} X_j^2. \tag{2}$$

Step 2: Starting from the starting point r, $r(i)$ increases in order, and calculate the weight of each node leaving its data section according to the above formula:

$$w(i, j) = \sum_{i=1}^{n} \left(j_i - \overline{X}\right)^2. \tag{3}$$

The weight calculation stops when the end point s, i.e., $i = s$, is reached.

Step 3: Starting from the end point s, calculate the traffic volume of the road section in the ascending order of $s(j)$.

The computation stops when the starting point r, i.e., $j = r$, is reached.

It can be shown that the results obtained by the Dial algorithm are identical to the assignment results of the Logit model; i.e., the Dial algorithm and the Logit model are equivalent.

MSA algorithm is widely used, is simple to calculate, and is the closest to the equilibrium distribution method of a flow distribution method which is the classical algorithm for solving the SUE model; the solution steps can be summarized as follows.

Step 1. Initialization: Let 0, $Xa = A\Omega B$, get the initial impedance of each road section, and accordingly, the traffic flow between OD pairs is loaded randomly (this paper uses the Dial algorithm mentioned above) to get the initial road section flow 0ax, so that the number of iterations $n = 0$.

Step 2. Update the impedance of each road section, according to the current traffic volume of each road segment Xa.

Step 3. Search for the direction of descent: According to the new impedance obtained from Step 2, the Dial algorithm is invoked to randomly load the traffic flow between ODs to obtain the additional flow Fa of the road section.

Step 4. Update road traffic.

Step 5. Convergence conditions: If the formula meets the above requirements,

$$w(i,j) = \sum_{i=1}^{n} X_i Y_i. \quad (4)$$

3.2. Flexible Traffic Network Allocation Model

3.2.1. Elastic Demand Model Construction. Elastic demand refers to the variation of OD traffic demand with the degree of network congestion. Its formula is as follows:

$$\min Z_{\mathrm{EDUE}}(x,Q) = \sum_{i=1}^{n} X_i^2 \int \sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i^2 \int \sum_{i=1}^{n} X_i Y_i. \quad (5)$$

Usually, $D(rs)$ is a monotonically decreasing (at least not increasing) function of the travel resistance between OD pairs (r, s). In addition, the demand function is still online because the population size and car ownership of each node are limited functions. In summary, the EDUE model can be described by the following mathematical planning model:

$$\min Z_{\mathrm{EDUE}}(x,Q) = \sum_{i=1}^{n} X_i^2 \int \sum_{i=1}^{n} X_i Y_i - \sum_{i=1}^{n} X_i^2 \int \sum_{i=1}^{n} X_i Y_i. \quad (6)$$

Based on the above description, the change in the demand function to the excess demand function of the public in terms of travel can be derived, and the image of the function is as follows. Figure 2 shows changes in demand function to excess demand function.

In the elastic demand (EDUE) model, the traffic demand between OD pairs will be adjusted according to the change of travel impedance, and the process of network traffic equilibrium under the elastic demand condition is based on the degree of network congestion. Figure 3 shows the transformation diagram of the excess demand network. Supply and demand balance to determine the right demand for each OD pair.

3.2.2. Definition and Calculation Model of Road Network Resilience Index. To quantitatively analyze the resilience of the road network to disasters, the concept of the "road network resilience index" is proposed to characterize the ability of the road network to dissipate itself when it is partially damaged. When a path network is partially damaged, the efficiency of the network decreases or fails partially, a high resilience index of the network or nodes indicates that the network can absorb and dissipate the consequences better, and the network is resilient. The topological resilience of each node constitutes the topological resilience of the whole road network [19].

When the efficiency of a node of a network decreases or fails due to a disaster, the ability of its surrounding network to absorb and dissipate the effects of its failure is the node resilience index. The absorption capacity of all nodes constitutes the absorption capacity of the whole network. The road network resilience index is described from two aspects: (i) as a complex network, the network itself has a large impact on the resilience, so it is necessary to consider the topological characteristics of the road network; (ii) the traffic

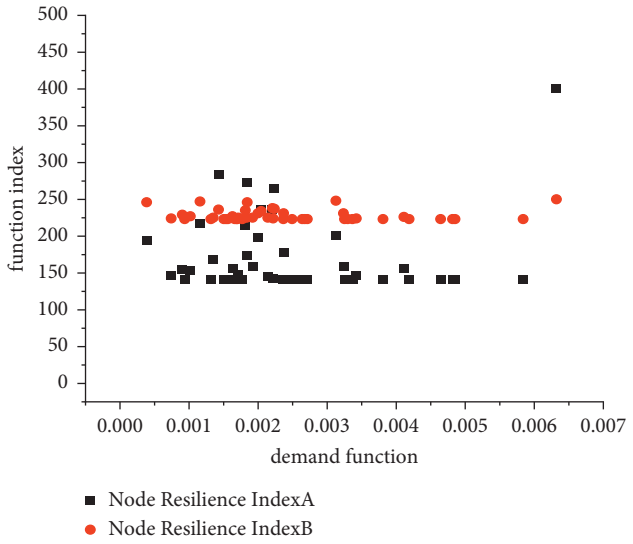


Figure 2: Changes in demand function to excess demand function.

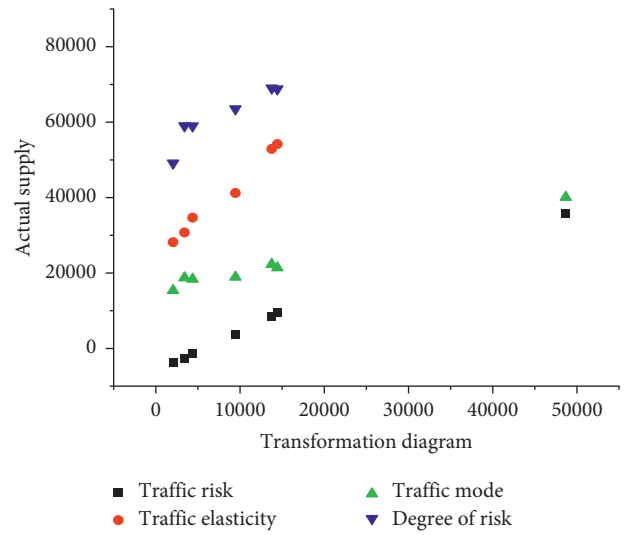


Figure 3: Transformation diagram of excess demand network.

characteristics of the road reflect the number and spatial and temporal distribution of traffic on the road, which is an important constraint on the resilience of the road network. The traffic characteristics on the road are often neglected in the previous studies of road network resilience [20]. See Figure 4.

Figure 4 displays the topology of the road network which is the "precondition" of the road network, which affects the resilience of the road network nodes to a certain extent. The topological methods commonly used to transform the topology of the actual road network are the principal method and the pairwise method. The topology of the path network is analyzed by using complex network theory, and the topological characteristics of the nodes are evaluated by using three indicators: node degree, intermediate degree, and near intermediate degree. The node degree is defined as the number of connections connected to the node, the intermediary intermediate degree reflects the importance of the

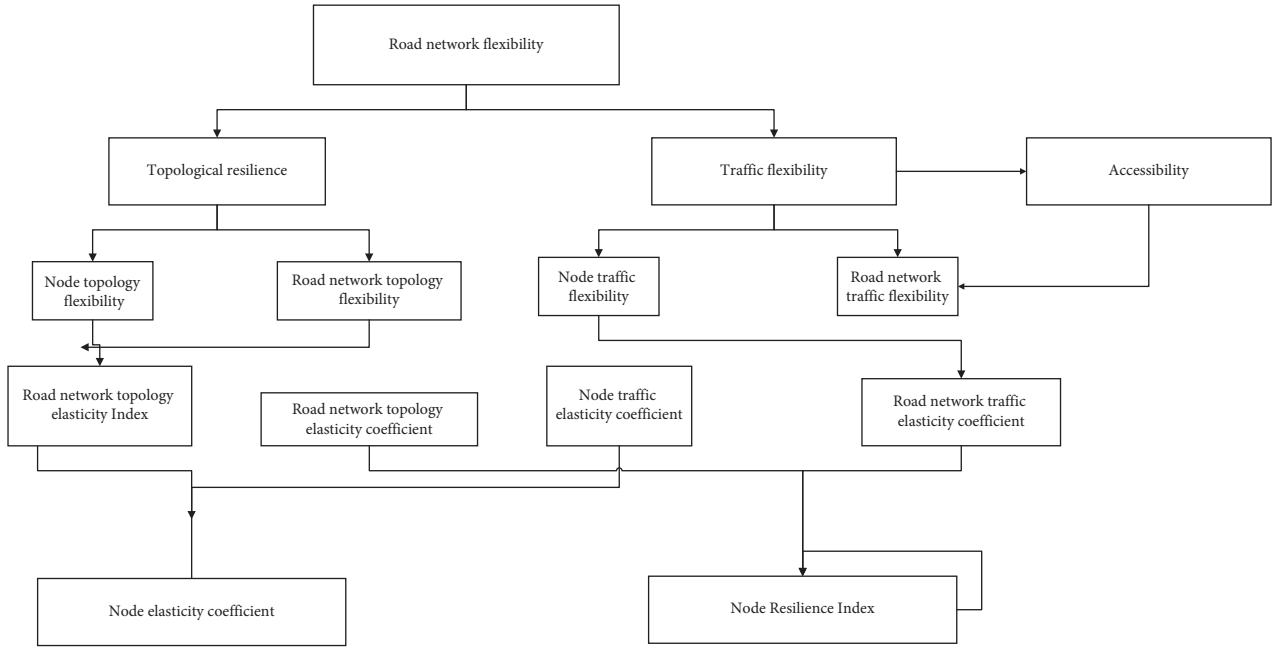


Figure 4: Conceptual components of the road network resilience index.

node in the network path selection, and the node with the highest degree of intermediary intermediate degree is the node with the highest number of shortest paths traversed in the network. Proximity centrality is the average shortest path between any point in the network and other nodes, which reflects the proximity of a node to other points on the network.

4. Model Solving Algorithm

Sensitivity analysis is a very important learning content in traffic network planning, especially in linear planning, but there are relatively few sensitivity studies applied to traffic network design, first of all, because of the complexity of traffic networks, and in many cases, it is extremely difficult to express them as nonlinear planning and let linear planning be alone. In addition, the uniqueness of the equilibrium path flow solution is not guaranteed, so it becomes very difficult to analyze the sensitivity of traffic equilibrium [21, 22]. However, it is a very effective research method in the descent method, and it is still quite effective for some specific equilibrium flow distribution models. Therefore, in recent years, more and more scholars have explored the sensitivity of traffic network equilibrium.

The nonnegative parameter 0 in the SUE model takes the value of 1 and the expansion cost function, which is the discount factor of the traffic manager's cost discounted to the expansion year during the expansion planning period. Here, we assume that the demand function is a constant 30, and the correlation coefficients for each road segment are shown in Table 1. The two-level planning problem is solved using the sensitivity analysis algorithm as developed in this paper. See Table 1.

Analysis of the Results. We know from Table 1 the actual situation; the equilibrium traffic volume of Sections 1, 4, and 5 is not going to change; that is, its equilibrium traffic volume and the derivative of the change in the capacity of the section are 0. Therefore, the circulation capacity of the model will change after the expansion of only Sections 2 and 3, which is consistent with the results obtained from the calculation of Table 2, so the algorithm is reasonable and effective. See Table 2.

Table 1: Parameters related to each road section.

Road network	1	2	3	4	5
t_a	5	6	9	56	5
c_a	35	54	77	87	44
y_a	6	7	3	66	68
K_a	8	5	77	78	9

4.1. Road Network Model and Parameter Setting. In this paper, a district is selected for the case study, and the arterial and branch roads in the region are abstracted. The road network topology is shown in Figure 5 with 34 nodes and 56 road sections. Among them, nodes 1, 2, 3, and 4 are the starting points. The dashed lines in the figure indicate the virtual road sections added between each OD pair. See Figure 5.

Among them, because the complexity of travel of the city's residents is much more than this, so this paper constructs more complex traffic network routes, with the support of algorithms, because the residents' travel needs are complex and diverse with multiplication. Figure 6 displays the road network topology index, so there are more risky perturbations, and more travel routes need to be constructed to have true simulation. The following is the distribution of the road network with real simulation. See Figure 6.

Table 2: Traffic network optimization diagram.

Road section	1	2	3	4	5
Section traffic volume	54	45	23	71	43
Circulation capacity after expansion	36	61	26	32	11
Elastic demand coefficient	44	29	37	35	23

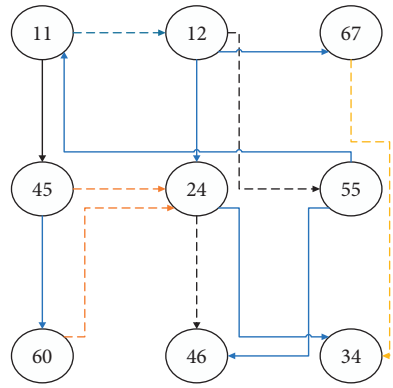


Figure 5: Road traffic network.

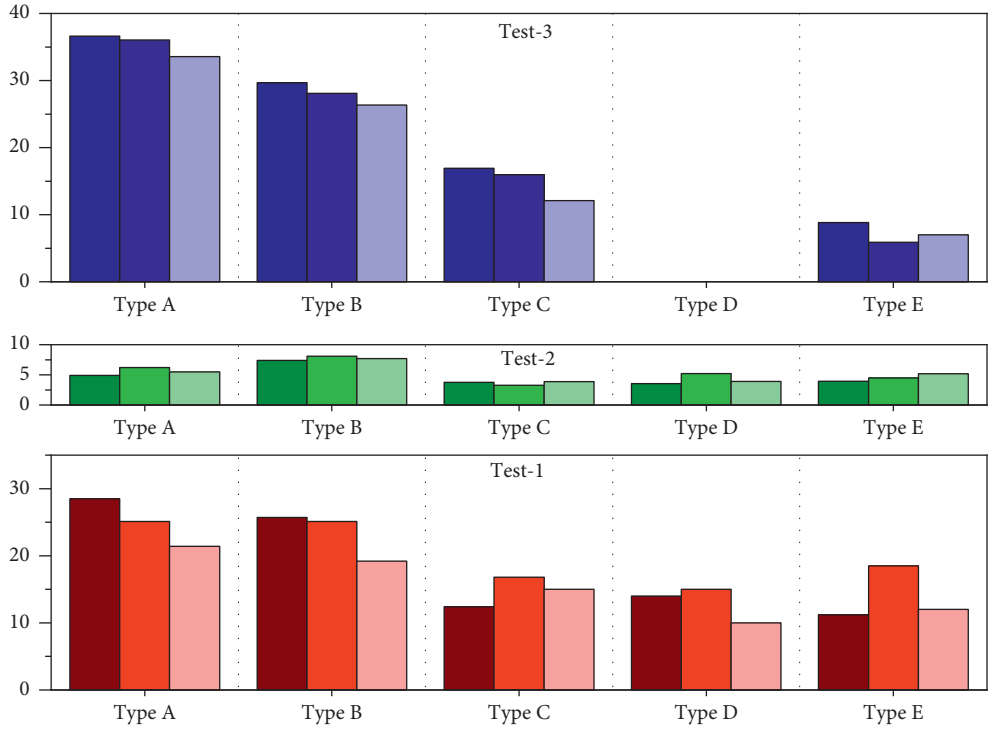


Figure 6: Road network topology index.

DCS subsystem is the core of the CBTC system to realize information interaction between subsystems, mainly transmitting control commands, train status, and other key information, ensuring bidirectional, safe, and reliable transmission of data, and playing an important role in the whole train control system. The following figure is the communication diagram of CBTC system subsystems; in the DCS system, there is mainly two-way information interaction between the following subsystems: ZC and VOBC, ZC and Cl, ZC and ATS, ZC and DSU, ATS and VOBC, ATS and Cl, ATS and DSU, and VOBC and Cl. VOBC communicates with ground equipment through a wireless access network, and the rest of the subsystems are based on the backbone network for data transmission. See Figure 7.

By comparing various transportation system operation diagrams and actual operation diagrams, it can be found that certain riskiness can cause changes in train operation status; i.e., Figure 7 shows that there are differences in the

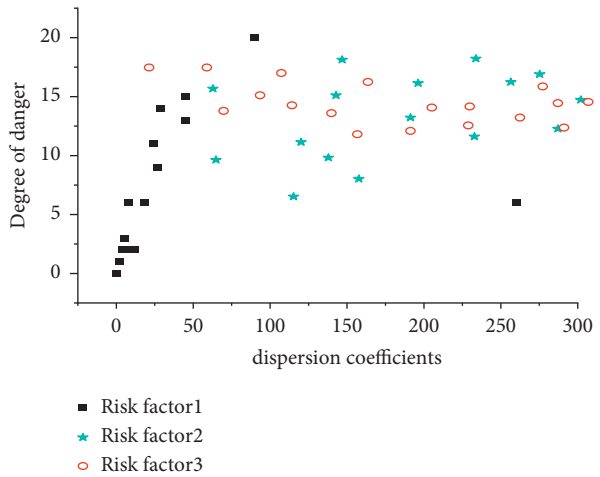


Figure 7: RMSE with different dispersion coefficients.

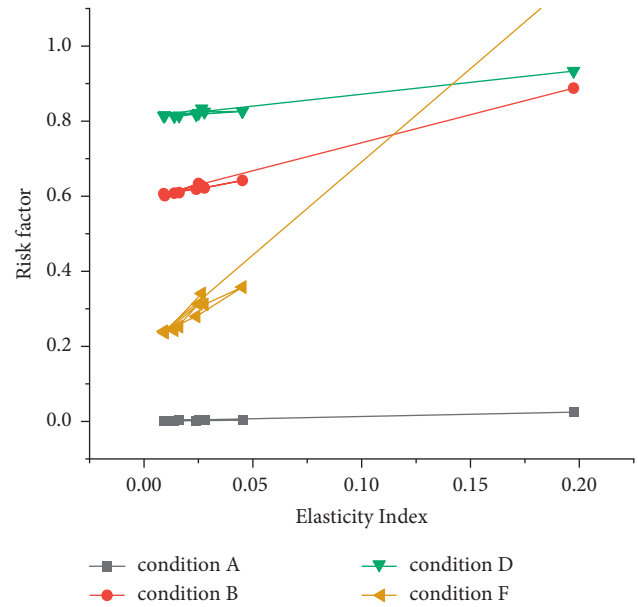


Figure 8: Comparison of trends in risk factors.

displacement of the same train traveling in the zone at the same moment before and after the system failure; plus, the driving mode of the train is different before and after the system failure, so there are also differences in the traveling speed of the same train at the same moment. Considering that the stopping time of the same train at the same station before and after the system failure may overlap, the above-mentioned displacement difference and the speed difference are zero; therefore, while considering the displacement and speed difference as indicators for evaluating the change of train operation performance, the train late time is also used as a measurement indicator.

4.2. Analysis of Experimental Results. This experiment uses the same traffic network as the numerical experiment above and gives the algorithm used to calculate and assign the roadway flow when the dispersion factor takes different values.

Figure 8 shows that when suboptimal congestion charging uses different lower-level models to characterize traveler behavior, the resulting roadway charging schemes are different, which indicates that when developing congestion charging policies, it is necessary to use more. See Figure 8.

Add the traffic distribution model that conforms to the traffic network condition of the city understudy, or consider the toll results obtained from multiple models to reasonably determine the final road section toll scheme. When the dispersion coefficient increases, the lower level of the kind of multiuser multicriteria stochastic user balance model of the path section toll value is more than a multiple of the model, which indicates that when the traveler travels a greater degree of dispersion, the amount of road section congestion toll is higher than in the model. The distribution results are similar to the fact that as the dispersion coefficient increases, there is an overall decreasing trend in the traffic of road sections and an overall increasing trend in the traffic of road sections. This is because a larger dispersion coefficient indicates that the travelers are more familiar with the path network, their travel choice is closer to the deterministic choice, the first term of the objective function dominates, and the travelers' travel is close to the result of user equilibrium. As the dispersion coefficient increases, the difference between the model and model section assignment results becomes smaller and smaller, which also reflects the fact that no increase leads to travelers' trips closer to deterministic choices. Regarding the complex transportation network G as a system containing n nodes and their interactions, there is a virtual field around each node, and any node in the network will be subject to the joint action of other nodes; thus, in the entire network, a data field is determined topologically, which is called network road resilience.

Similarly, regardless of the dispersion coefficients, the model assignment results are smaller for roadway segments and larger for roadway segments. This is because the assumption ignores the correlation between paths, and thus, the traffic assignment results for overlapping paths and overlapping path sections will be larger, while the overlapping relationship between paths is considered to adjust the negative utility of travel by paths, and the paths with a high degree of overlap are adjusted more and vice versa to a lesser extent. Thus, the model can overcome the characteristics of traditional models to some extent.

The above are the results of the resilience assessment of the system under two scenarios with riskiness, and the comparison reveals that the absorption capacity of the system resilience is comparable under both riskiness scenarios, indicating that the system has similar robustness in the face of both risks, and the adaptive capacity of the system is weaker under the first riskiness scenario, indicating that the first attack has a greater impact on the system; in addition, the relaxation time is greater than the initial system recovery time, indicating that the initial system recovery measures are effective and that scenario one is faster in human intervention to repair the system after a failure occurs, and therefore, the recovery is faster.

5. Conclusion

The urban rail transit system is closely related to the people's livelihood and is one of the key infrastructures to promote the efficient operation of the city and rapid economic development. In this paper, we analyze the serious information security situation faced by the transportation system and clarify that the research and application of continuous risk perturbation assessment methods are extremely urgent and important for the train control system. At present, there are relatively few studies on continuous risk perturbation of urban railways at home and abroad, and most of them only stay at the concept and framework level; although there are a lot of information security risk assessment methods of traditional IT systems and industrial control systems to draw on, given the unique properties of urban transportation systems, the relevant methods cannot be fully applied, so this paper focuses on the study of resilient transportation under continuous risk perturbation. Therefore, this paper focuses on the risk assessment method for resilient traffic under continuous risk perturbation. The resilient transportation method and its application play an important role in improving the information security capability of urban rail transportation.

In this paper, based on the detailed analysis of the structure and function of a typical urban transportation system one-by-one communication-based train operation control (CBTC) system, the concept of vulnerability index is proposed to assess the vulnerability of each device, scenario, and the whole system in CBTC; secondly, concerning the relationship between all devices in the resilient transportation system from the perspective of network structure, a complex network model is constructed, and a comprehensive information domain and physical domain CBTC system information security risk assessment method is proposed; finally, based on the resilience theory, the CBTC system information security resilience assessment model is proposed as an indicator, and there are model construction and algorithm implementation for quantitative analysis of column control system information security.

The attack tree model and the construction of the complex network model proposed in this paper are realized artificially. As the scale of the system increases and the connection between devices becomes more and more complex, the workload of model construction will increase greatly, and how to design and develop automatic model generation tools can be a subsequent research topic. The attack scenarios simulated in this paper are limited, and the attack process is determined by humans; how to combine the characteristics of the column control system to establish a perfect attack library and simulate the automated attack process is of great significance for the subsequent research related to the information security of the column control system. In the future, the algorithm of the article will mainly focus on improving the timeliness of the algorithm, so that congested roads can be diverted, to solve the phenomenon of traffic congestion.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This paper was not funded by any organization. The authors would like to acknowledge the Scientific Research Project of Colleges and Universities in Inner Mongolia Autonomous Region (NJZY21160).

References

[1] T. C. Sharkey, S. G. Nurre Pinkley, D. A. Eisenberg, and D. L. Alderson, "In search of network resilience: an optimization-based view," *Networks*, vol. 77, no. 2, pp. 225–254, 2021.

[2] D. Ivanov, B. Sokolov, and A. Dolgui, "The Ripple effect in supply chains: trade-off "efficiency-flexibility-resilience" in disruption management," *International Journal of Production Research*, vol. 52, no. 7, pp. 2154–2172, 2014.

[3] L. Zhang, G. Zeng, D. Li, H.-J. Huang, H. E. Stanley, and S. Havlin, "Scale-free resilience of real traffic jams," *Proceedings of the National Academy of Sciences*, vol. 116, no. 18, pp. 8673–8678, 2019.

[4] A. Javed and Y. Kohda, "Sustainable work opportunities for drivers' well-being: a case of careem as transportation network company," in *Proceedings of the International Conference on Applied Human Factors and Ergonomics*, pp. 70–77, Springer, New York, NY, USA, July 2021.

[5] M. Diao, H. Kong, and J. Zhao, "Impacts of transportation network companies on urban mobility," *Nature Sustainability*, vol. 4, no. 6, pp. 494–500, 2021.

[6] I. Dakic, K. Yang, M. Menendez, and J. Y. J. Chow, "On the design of an optimal flexible bus dispatching system with modular bus units: using the three-dimensional macroscopic fundamental diagram," *Transportation Research Part B: Methodological*, vol. 148, pp. 38–59, 2021.

[7] C. Li and B. Palanisamy, "Reversible spatio-temporal perturbation for protecting location privacy," *Computer Communications*, vol. 135, pp. 16–27, 2019.

[8] S. Mohammed, J. Fiaidhi, and M. Tang, "Towards using micro-services for transportation management systems," *LISS 2020: Proceedings of the 10th International Conference on Logistics, Informatics and Service Sciences*, Springer, Singapore, pp. 107–117, 2021.

[9] F. Goodarzian, D. Shishebori, H. Nasseri, and F. Dadvar, "A bi-objective production-distribution problem in a supply chain network under grey flexible conditions," *RAIRO-Operations Research*, vol. 55, pp. S1287–S1316, 2021.

[10] M. Nogal and D. Honfi, "Assessment of road traffic resilience assuming stochastic user behaviour," *Reliability Engineering & System Safety*, vol. 185, pp. 72–83, 2019.

[11] K. Govindan, M. Fattahi, and E. Keyvanshokooh, "Supply chain network design under uncertainty: a comprehensive review and future research directions," *European Journal of Operational Research*, vol. 263, no. 1, pp. 108–141, 2017.

[12] I. M. Cavalcante, E. M. Frazzon, F. A. Forcellini, and D. Ivanov, "A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing," *International Journal of Information Management*, vol. 49, pp. 86–97, 2019.

[13] N. U. Ibne Hossain, M. Nagahi, R. Jaradat, C. Shah, R. Buchanan, and M. Hamilton, "Modeling and assessing cyber resilience of smart grid using Bayesian network-based approach: a system of systems problem," *Journal of Computational Design and Engineering*, vol. 7, no. 3, pp. 352–366, 2020.

[14] G. Como, "On resilient control of dynamical flow networks," *Annual Reviews in Control*, vol. 43, pp. 80–90, 2017.

[15] F. O. Olowononi, D. B. Rawat, and C. Liu, "Resilient machine learning for networked cyber physical systems: a survey for machine learning security to securing machine learning for CPS," *IEEE Communications Surveys & Tutorials*, vol. 23, no. 1, pp. 524–552, 2020.

[16] D. Xue and S. Hirche, "Finite-time distributed topology design for optimal network resilience," *IET Control Theory & Applications*, vol. 13, no. 17, pp. 2792–2799, 2019.

[17] Y.-P. Fang and E. Zio, "An adaptive robust framework for the optimization of the resilience of interdependent infrastructures under natural hazards," *European Journal of Operational Research*, vol. 276, no. 3, pp. 1119–1136, 2019.

[18] T. N. Turan, J. H. Voeks, M. I. Chimowitz et al., "Rationale, design, and implementation of intensive risk factor treatment in the CREST2 trial," *Stroke*, vol. 51, no. 10, pp. 2960–2971, 2020.

[19] J. Lighter, M. Phillips, S. Hochman et al., "Obesity in patients younger than 60 years is a risk factor for COVID-19 hospital admission," *Clinical Infectious Diseases*, vol. 71, no. 15, pp. 896-897, 2020.

[20] D. Ivanov and B. Sokolov, "Simultaneous structural–operational control of supply chain dynamics and resilience," *Annals of Operations Research*, vol. 283, no. 1, pp. 1191–1210, 2019.

[21] Z. Lin, F. Long, Y. Yang, X. Chen, L. Xu, and M. Yang, "Serum ferritin as an independent risk factor for severity in COVID-19 patients," *Journal of Infection*, vol. 81, no. 4, pp. 647–679, 2020.

[22] D. Hutchison and J. P. G. Sterbenz, "Architecture and design for resilient networked systems," *Computer Communications*, vol. 131, pp. 13–21, 2018.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 8403025, 13 pages
https://doi.org/10.1155/2021/8403025

Research Article

Classification on Digital Pathological Images of Breast Cancer Based on Deep Features of Different Levels

Xin Li, HongBo Li, WenSheng Cui, ZhaoHui Cai, and MeiJuan Jia

School of Computer Science and Information Technology, Daqing Normal University, Daqing 163712, China

Correspondence should be addressed to HongBo Li; lhb20210401@163.com

Received 22 November 2021; Accepted 15 December 2021; Published 30 December 2021

Academic Editor: Xiangtao Li

Breast cancer is one of the primary causes of cancer death in the world and has a great impact on women's health. Generally, the majority of classification methods rely on the high-level feature. However, different levels of features may not be positively correlated for the final results of classification. Inspired by the recent widespread use of deep learning, this study proposes a novel method for classifying benign cancer and malignant breast cancer based on deep features. First, we design Sliding + Random and Sliding + Class Balance Random window slicing strategies for data preprocessing. The two strategies enhance the generalization of model and improve classification performance on minority classes. Second, feature extraction is based on the AlexNet model. We also discuss the influence of intermediate- and high-level features on classification results. Third, different levels of features are input into different machine-learning models for classification, and then, the best combination is chosen. The experimental results show that the data preprocessing of the Sliding + Class Balance Random window slicing strategy produces decent effectiveness on the BreaKHis dataset. The classification accuracy ranges from 83.57% to 88.69% at different magnifications. On this basis, combining intermediate- and high-level features with SVM has the best classification effect. The classification accuracy ranges from 85.30% to 88.76% at different magnifications. Compared with the latest results of F. A. Spanhol's team who provide BreaKHis data, the presented method shows better classification performance on image-level accuracy. We believe that the proposed method has promising good practical value and research significance.

1. Introduction

In recent years, the global prevalence of breast cancer (BC) has been gradually increased, and the affected organisms tend to be younger, gender-neutral, and have racial ambiguity, which has posed a huge threat to human beings' normal life. In 2018, the World Health Organization's International Agency estimated that there were 2.1 million new female cases of BC among women about 25 percent of all cancers. The number of female cases was far greater than any other cancers in both developed and developing countries [1]. BC is the leading cause of cancer deaths among women between 20 and 60 years old. Early diagnosis and treatment can effectively reduce the risk of diseases and prevent the progression of cancers [2]. The traditional diagnosis of BC includes breast mammography, breast B ultrasound, dynamic enhanced magnetic resonance, and pathological biopsy [3]. Normally, pathologists need to combine the feedback from medical equipment with their own diagnostic experience to test and analyze sample information for cancer diagnosis and treatment strategy. The diagnosis process is inefficient, costly, and subjective. Furthermore, due to the uneven distribution of pathologists and medical resources around the world, it is difficult to ensure timely and effective treatment for patients in remote areas and underdeveloped countries [4]. Therefore, an efficient, low-cost, and objective diagnosis method has important social significance and research value.

Digital pathological slicing scanner converts the pathological slicing of substance from photoelectric signal to digital signal and finally generates full-information and high-resolution digital image, namely whole slide imaging (WSI) [5]. Compared with traditional pathological slicing, it has the advantages of convenient preservation, less damage, remote diagnosis, and so on. WSI provides a basis for automatic classification and quantitative analysis of medical

pathological slicing [6, 7]. In pathological WSI, different diseases have different spatial pixel arrangements and cellular features. Cells of different shapes, sizes, and colors may also be of the same type. Therefore, automatic analysis of WSI by computer algorithm is a very challenging job. With the upgrade of software and hardware, machine-learning (ML) algorithm and deep learning (DL) algorithm have been widely used in the field of lesion region segmentation, location detection, and subtype classification as a digital pathology-aided diagnosis method. They also become one of the hot research fields in pattern recognition and artificial intelligence [8, 9]. In the automatic analysis of pathological images, the application of computer-aided diagnosis (CAD) technology has improved the accuracy and efficiency of disease diagnosis and grasps the development trend of disease more objectively. It also reduces the limitations of economic conditions, geographical environment, and medical foundation.

There are two types of BC: benign and malignant. The benign cancer is prone to be transformed into the malignant cancer in the early stage. The benign cancer mainly includes adenosis, phyllodes tumor, fibroadenoma, and tubular adenoma. The malignant cancer mainly includes papillary carcinoma, ductal carcinoma, lobular cancer, and mucinous cancer [10, 11]. Among them, ductal carcinoma and lobular cancer accounted for more than 95%.

In the traditional classification of the BC, doctors need to combine feedback from medical equipment with their own experience to diagnose the body, which is somewhat subjective and inefficient. In addition, with the increasing prevalence of BC and the lack of medical resources, problems such as missed diagnosis and misdiagnosis are prone to occur. Some ML algorithms can solve the above problems, but the model performance is not satisfactory and cannot provide scientific guidance and suggestions for clinical treatment. Therefore, researchers have conducted in-depth exploration and research on the CAD method based on the DL algorithm. Although great progress has been made, there are still some shortcomings in data preprocessing, feature extraction, unbalanced data analysis, and the selection of benchmark dataset. To solve the above problems, combining the correlation and difference between the ML and DL algorithms, we propose a novel BC's benign and malignant classification scheme based on deep features of different levels. Firstly, image preprocessing is performed on the BC 's pathological tissue images, including normalization and window slicing. Secondly, the input data images are trained based on the AlexNet model. Then, we discuss how to train the network model and choose appropriate nodes by extracting features. Finally, the trained network model extracts features by choosing node. Then, the extracted features are input into the ML model to classify. The experiment is performed using a fivefold cross-validation approach, the same folds released with the BreaKHis dataset. The experimental results show that the average classification accuracy and standard deviation of pathological tissue images at different magnifications (40×, 100×, 200×, and 400×) are 87.85%, 86.68%, 87.75%, and 85.30%, respectively. The average classification accuracy and standard deviation of patients at different magnifications (40×, 100×, 200×, and 400×) are 87.93%, 87.41%, 88.76%, and 85.55%, respectively. In summary, our approach has higher classification accuracy on the image-level accuracy and model stability than the latest classification results of Spanhol team from the BreaKHis image dataset in literature [12]. The following is a summary of the contribution of this article:

(1) To improve the generalization of the model, four data preprocessing strategies are used in this study. The four window slicing strategies are Sliding, Random, Sliding + Random, and Sliding + Class Balance Random. The experimental results show that the preprocessing strategy of Sliding + Class Balance Random window slicing has the best classification results.

(2) As a second step, the intermediate- and high-level features are extracted by the AlexNet model. The experiments show that the intermediate- and high-level features have a better classification effect.

(3) Based on the previous steps, extracted features are input into a variety of ML models for training and evaluation. We classify BC pathological tissue images (BreaKHis), and SVM has great potential for classifying.

2. Related Work

The histopathological analysis is a highly specialized task. The effectiveness of diagnosis depends on pathologists' experience, attention, and fatigue. Traditional techniques such as MRI, ultrasound, and biopsy technology are used to detect and grade BC's lesions. Although biopsy techniques are time-consuming to diagnose, it is still one of the gold standards on diagnosis [13]. Common biopsy techniques include skin biopsy, fine-needle biopsy, core biopsy, and surgical biopsy. In the biopsy assay, biopsy tissue slicing samples are first obtained, and then, hematoxylin and eosin (H&E) staining is performed. At last, the pathologists use the equipment to analyze texture, morphology, and histological characteristics of the biopsy tissue to give the corresponding diagnostic results [14]. The pathologists focus, magnify, and scan the entire tissue under a high-power microscope. The procedure is time-consuming, repetitive, and subjective, so the diagnosis result may be very different.

With the advent of computer era, how to use computer algorithms to better assist pathologists in the simple and repeatable process on the pathologic diagnosis has always been a challenging task [15]. CAD allows pathologists to devote more energy and time on dealing with difficult disease. Therefore, a great deal of intensive research emerges from the CAD field, especially for the analysis of BC. Khamparia et al. [16] use a modified VGG model as a pretraining model on the DDSM dataset of X-ray, and the classification accuracy is 94.3%. Sharma et al. [17] use various evolutionary algorithms as the feature extractor for the Wisconsin breast cancer dataset on the UCI database. Then, they use an ML model as a classifier and get a maximum accuracy of 96.45% by the combination of BPSO

and SVM algorithms. Ha et al. [18] classify benign and malignant MRI images of 216 BC patients, which are provided by Columbia University Medical Center. In the study, they present a convolutional neural network (CNN) based on MRI features and finally achieve 70% accuracy. Cruz-Roa et al. [19] study the WSI tissue region in 162 patients with invasive ductal breast, which is provided by Pennsylvania University and New Jersey Cancer Institute. At first, the team transforms the WSI detection problem into a classification problem, so they reduce the difficulty on detection of disease areas. Then, a multilayer CNN is designed to classify the tissue images after slicing. Finally, the evaluation results achieve F1 score of 71.80% and a balanced accuracy of 84.23%. Fuzzy color histogram is manually extracted features and is input into random forest classifier for training. The results achieve 67.53% for F1 score and 78.74% for balanced accuracy. In the same way, the results of RGB histogram achieve 66.64% for F1 score and 77.24% for balanced accuracy. Anuranjeeta et al. [20] use morphological features to detect BC lesions and classify cells in pathological tissue images. Finally, the results achieve 85.7% accuracy and 0.884 AUC value on rotating forest classification model.

Spanhol et al. propose a series of studies on benign and malignant BC classification. The experiment is performed using a fivefold cross-validation approach, and the specific research contents are as follows:

(1) At the beginning of the series, Spanhol et al. [12] provide 7909 BC digital pathological tissue images from 82 patient cases (BreaKHis). They also design six different feature extractors to extract low-level features such as texture and edge. After then, ML algorithms such as decision tree (DT), random forest (RF), and support vector machines (SVMs) are used to classify for extracting features. Finally, a better classification result is the combination of the feature extractors PFTAS [21] and SVMs. The average accuracy of classification at different magnifications (40×, 100×, 200×, and 400×) ranges from 80% to 85%, and the standard deviation is about 5%.

(2) To further improve the classification performance, Spanhol et al. [22] use a novel strategy to obtain small pieces from the BreaKHis image data. After obtaining the data, these small pieces are input into CNN model for training. Image-level accuracy and patient-level accuracy are given as classification evaluation indicators. The final classification is decided by voting mechanism of ensemble learning. At last, on the patient-level accuracy and image-level accuracy, the average accuracy of classification ranges from 80% to 90%, with a large standard deviation.

(3) In [23], Spanhol uses the AlexNet model for the pretraining on BreaKHis dataset and extracting deep features. The features in different depths of layers are obtained and all are input into the logistic regression (LR) classifier to train and evaluate. Finally, the decaffeinated coffee method named by Spanhol et al. is combined with the method in literature [12]. The experiments show that the average accuracy at 200× magnifications achieves 86.3% on the patient-level accuracy and 84.2% on the image-level accuracy. However, compared with literature [12], the evaluation results of other magnifications significantly decreased.

(4) Recently, Spanhol et al. [24] apply the multi-instance principle to classify BC pathological cancer images. On the patient-level accuracy, the average accuracy is 92.1%, 89.1%, 87.2%, and 82.7%, respectively. Similarly, on image-level accuracy, the average accuracy at different magnifications (40×, 100×, 200×, and 400×) achieves 87.8%, 85.6%, 80.8%, and 82.9%, respectively. Compared with literature [23], the image-level accuracy decreases at 200×and 400×magnifications, while the standard deviation increases.

Though researchers have made great progress in the image classification problem of BC, there are still many deficiencies. The studies from the literature [16–20] use a small BC dataset. As a result, it leads to the inability for fitting model and objectively evaluating generalization ability. In [12], it is hard to obtain the effective classification features using the traditional manual feature extraction method. In [22], researchers adopted the random data acquisition method. This method leads to model instability. Because 1000 data pieces (64*64*3) are obtained from each data image (350*230*3), the model training is difficult and the data are redundant. It is difficult to provide a large enough global view of the data.

3. Materials and Methods

3.1. Dataset. In our work, we use the publicly available BreaKHis image dataset as the benchmark for classifying BC. The dataset contains 7909 pathological tissue images of benign and malignant BC in 82 groups of patients. The patients were invited to participate in research activities in the P&D Laboratory in Brazil during January to December in 2014. All information is anonymous to ensure patient privacy. Pathologists use H&E to stain and label the pathological tissue images. Samples are prepared using a standard paraffin method, including fixation, dehydration, removal, and correction. In the process of the digital transformation of BC tissue images, a Samsung digital color camera SCC-131AN is used to obtain digital tissue images of RGB channels with the magnifications of 40×, 100×, 200×, and 400×. Pathologists remove the worthless areas such as the black border and text annotation from the original image by clipping operation. Finally, the digital pathological tissue image of BC with $700 \times 460 \times 3$ pixels is obtained. Sample tissue images of magnifications (40×, 100×, 200×, and 400×) are shown in Figure 1.

BreaKHis dataset is divided into benign and malignant BC. Table 1 shows the distribution of them. Besides, the benign cancer and malignant cancer have different subtypes

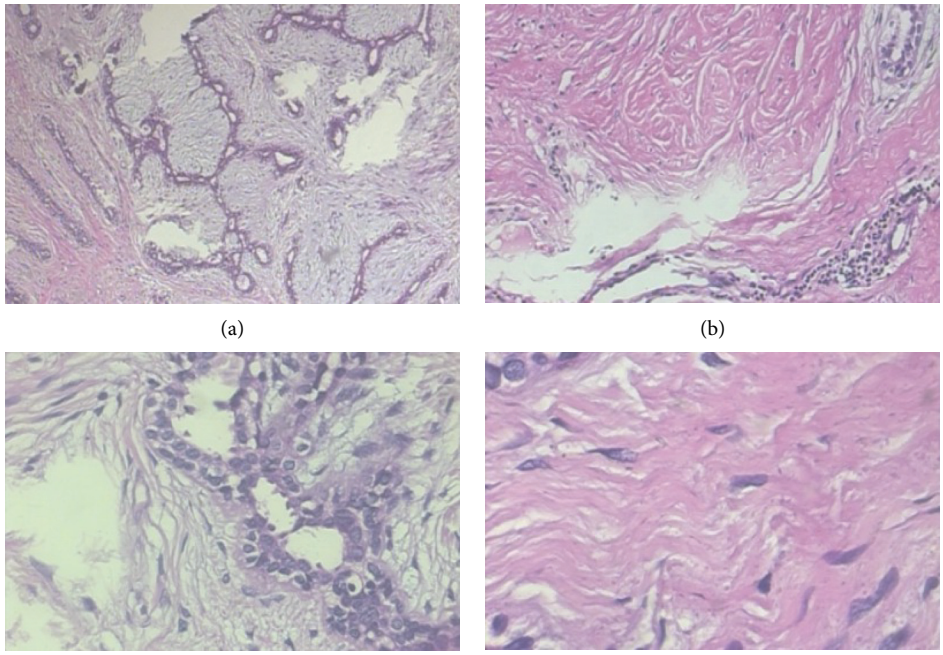


Figure 1: Digital pathological tissue images of breast cancer at different magnifications. (a) 40×, (b) 100×, (c) 200×, and (d) 400×.

Table 1: Distribution of benign cancer and malignant cancer.

Magnifications	Benign	Malignant	Total
40×	625	1370	1995
100×	644	1437	2081
200×	623	1390	2013
400×	588	1232	1820
Total	2480	5429	7909
Patients	24	58	82

Table 2: Distribution of subtypes of benign cancer.

Magnifications	A	F	TA	PT	Total
40×	114	253	109	149	625
100×	113	260	121	150	644
200×	111	264	108	140	623
400×	106	237	115	130	588
Total	444	1014	453	569	2480
Patients	4	10	3	7	24

Table 3: Distribution of subtypes of malignant cancer.

Magnifications	DC	LC	MC	PC	Total
40×	864	156	205	145	1370
100×	903	170	222	142	1437
200×	896	163	196	135	1390
400×	788	137	169	138	1232
Total	3451	626	792	560	5429
Patients	38	5	9	6	58

under a high-power microscope. The benign cancer includes four types: adenosis (A), fibroadenoma (F), tubular adenoma (TA), and phyllodes tumor (PT); the malignant cancer includes four types: ductal carcinoma (DC), lobular carcinoma (LC), mucinous carcinoma (MC), and papillary carcinoma (PC). The class distribution of benign and malignant subtypes is shown in Tables 2 and 3.

3.2. Methods. In our work, the normalization and window slicing operations are firstly used to preprocess the dataset. We design the training model based on the AlexNet model. The features of nodes in different depths are extracted by the model. The extracted features are input into the ML classifier for classification. Finally, we choose the best combination of feature extraction node and ML classifier. Four window slicing strategies (Sliding, Random, Sliding + Random, and Sliding + Class Balance Random) are used to obtain small pieces of data. Next, small pieces are input into the model for training and evaluation. The optimal window slicing strategy is proved by experimental results. At last, the best strategy is used as the basis of subsequent research. The method structure diagram is shown in Figure 2.

When the trained model is used to predict the whole image, the data are split into 15 small image sets with 128*128*3 pixel size by 128 pixel step length. The small image set is input into the model for prediction, and 15 decision results are obtained for each whole pathological image. The classification with the most votes is chosen by voting as the final prediction result.

3.2.1. Data Preprocessing. In this study, the normalization method is adopted to preprocess data, which can compress the data range from -1 and 1. This data processing method is not only simple and easy to implement, but also can improve the training efficiency and generalization of the model. The mathematical calculation formula is shown in the following equation:

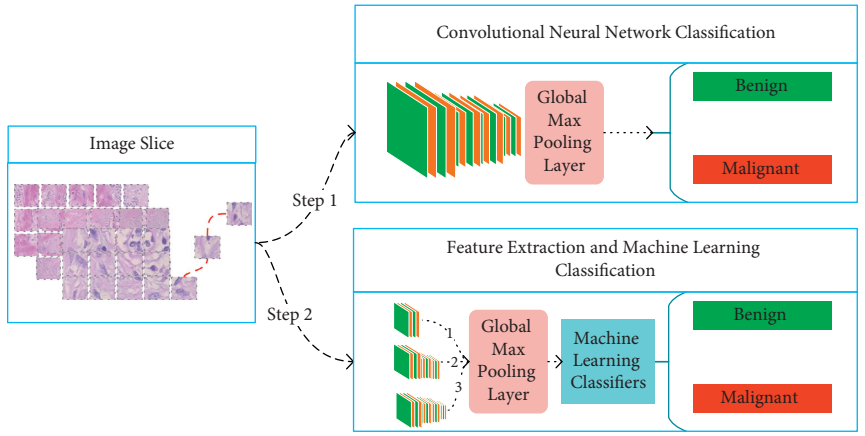


Figure 2: Method structure. Step1 is that the slicing images are input into CNN for benign-malignant classification. Step 2 obtains three different models for feature extraction according to the trained model of Step1 and combines the extracted features with the machine-learning model to classify benign cancer and malignant cancer.

$$X' = \left(\frac{X}{255} - 0.5\right) * 2. \quad (1)$$

Because there is no area of interest in the BreaKHis dataset, the experiments can be designed according to their own needs during the preprocessing by window slicing. In [22], the original data image is scaled from $700 \times 460 \times 3$ to $350 \times 230 \times 3$ in [25]. Four strategies are used to obtain small images. The specific division is shown in Table 4.

In Table 4, 32×32 and 64×64 pixel windows are used to obtain data images, and the sliding step is half of the corresponding window size, but it is hard to obtain a large global view of the data. When acquiring 1000 pieces of data by Random window slicing strategy, there is a lot of repeated information. Obviously, this preprocessing method can hardly provide effective data. It leads to underfitting and overfitting problems when the model is trained. To overcome this problem, we try four different window slicing strategies, including Sliding window slicing, Random window slicing, Sliding + Random window slicing, and Sliding + Class Balance Random window slicing. The experimental results show that Sliding window slicing strategy can guarantee model's fitting training, but cannot guarantee generalization training. Oppositely, Random window slicing can guarantee model's generalization training, but cannot guarantee the fitting training effectively. In our work, we have fully considered the advantages of the two strategies and proposed the Sliding + Random window slicing strategy. This idea not only guaranteed the model's fitting training, but also improved the generalization ability.

Due to the unbalanced distribution of classes, we improve the Random window slicing strategy by obtaining random minority class data with directivity. The specific window slicing strategy of the data is shown in Table 5. The size of various window slicing strategies is $128 \times 128 \times 3$ pixels, and the sliding step is128 pixels. This mode can ensure a large enough global view of data. Then, we use four data strategies to generate data for the model training. Sliding + Random window obtains 45 small pictures and can improve model fitting and generalization ability. To solve the problem of class imbalance, Sliding + Class Balance Random window gives a random data acquisition method based on the first three and transforms unbalanced data into balanced. Therefore, it improved the classification performance of the model by randomly acquiring benign minority classes with directivity.

Table 4: Data window slicing method of [22].

#	Patch size	Strategy	Number of patches
1	32×32	Sliding window	260
2	64×64	Sliding window	54
3	32×32	Random window	1000
4	64×64	Random window	1000

Table 5: Data window slicing mode.

#	Strategy	Number of patches
1	Sliding window	15
2	Random window	30
3	Sliding + Random window	45
4	Sliding + Class Balance Random window	Benign 45, malignant 22

3.2.2. Network Model Design and Construction. We use the Keras framework to design CNN based on the AlexNet model. In the 2012 CVPR Competition, Krizhevsky et al. [26] proposed the AlexNet model, which consists of multilayer convolution, pooling, and nonlinear mapping layers. The AlexNet model's network structure is shown in Figure3(a). After a large number of experiments, the network model with the best classification performance contains the following layers and parameters:

Input layer: the layer is the input of the network model and the output of the back propagation algorithm. The input layer image size is $128 \times 128 \times 3$.

Convolution layer: as an important structure for model feature extracting and learning, it is responsible for

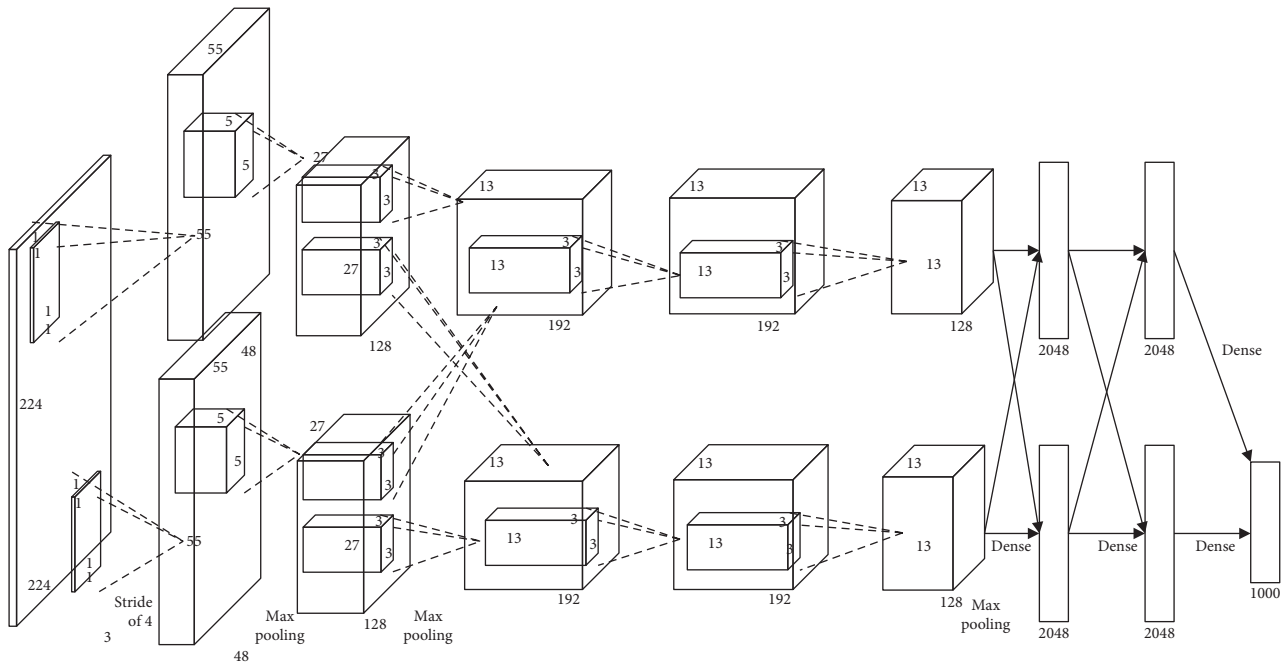


Figure 3: AlexNet model structure in [26].

gradually abstracting low-level features into high-level features. The model consists of three convolutional layers; the kernel size is 5×5, 5×5, and 3×3; the number of convolutional kernels is 16, 32, and 72; and the order of step is 3, 1, and 1. The zero filling method is adopted, and parameters of convolution kernel are initialized with the Gaussian distribution, and the bias is 0.

Pooling layer: the layer is responsible for reducing dimensionality of input features through down-sampling. Each convolutional layer is followed by a pooling layer. All pooling sizes are 2×2; step size is 2.

Nonlinear mapping layer: in this layer, low-level features are mapped to the high-level features to facilitate model classification. ReLU is used as the activation function in this process. Behind each convolution layer, a ReLU layer is added. The calculation formula of the ReLU function is shown as follows:

$$f(x) = \max(x, 0). \tag{2}$$

Global max pooling layer: in this layer, the maximum value of each feature map is output. The global max pooling layer can reduce fitting parameters and improve the generalization ability of the model.

Output layer: the layer is output of the model and input of the back propagation algorithm. The number of neurons in the output layer is set to 2, and softmax is used as the activation function.

The weight of the network layer is initialized by the Gaussian distribution, and the bias is initialized to 0. The structure of the CNN model and the configuration of the model's hyperparameter are shown in Tables 6 and 7.

Table 6: CNN model structure.

Layer	Type	Num kernels	Kernel size	Stride	Activation
0	Input	3	128×128	—	—
1	Convolution	16	5×5	3	ReLU
2	Max pool	—	2×2	2	—
3	Convolution	32	5×5	1	ReLU
4	Max pool	—	2×2	2	—
5	Convolution	72	3×3	1	ReLU
6	Max pool	—	2×2	2	—
7	Global max pool	—	—	—	—
8	Output	2	—	—	Softmax

Table 7: Hyperparameter settings for the CNN model.

Type	Values
Initial learning rate	$1e-4$
Learning rate schedule	Adam
Weight decay	$1e-5$
Train or test batch size	128
Loss function	Cross-entropy loss function
Number of epochs	100

3.2.3. Deep Feature Extraction and Machine-Learning Model Classification. According to the working principle of CNN, different convolutional layers have different feature extraction tasks. With the deepening of network layers, the low-level features are gradually abstracted to the high-level features. To fit data, CNN uses the high-level features to finish the corresponding learning tasks. Researches and experiments show that the classification effect of intermediate- and high-level features extracted by CNN is not worse than the high-level features; adopting an appropriate ML

classifier can improve the classification results. Therefore, we present a fusion method based on deep features and ML classifier for malignant-benign classification.

In Table 6, the trained CNN takes convolution + pooling layer as the splitting node. The models are split into three groups, and the global max pooling layer is added behind them to generate three new CNN models. Then, the data are input into the new CNN model for feature extraction, and the extracted features are input into ML classifiers for training. ML classifiers include SVM, LR, Gaussian naive Bayes (GNB), DT, RF, and multilayer feedforward neural network (MFNN). The feature extraction structure diagram is shown in Figure 4.

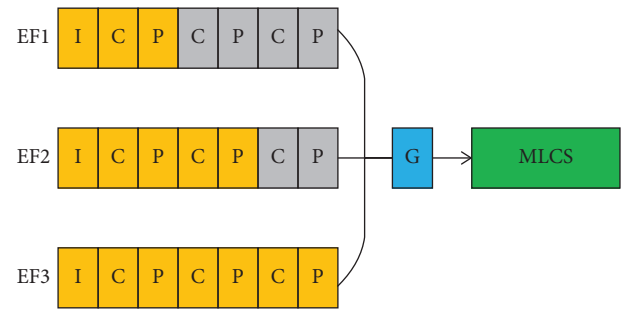


Figure 4: Feature extraction structure. EF1, EF2, and EF3, respectively, represent three different feature extraction methods. The orange area represents the effective feature extraction area, and the gray area represents the invalid area. I represents the input layer, C represents the convolution layer, P represents the max pooling layer, G represents the global max pooling layer, and MLCS represents the ML classifier set.

4. Results and Discussion

Firstly, patients used a fivefold stratified shuffle split cross-validation as the data splitting benchmark. This data splitting method is consistent in F. A. Spanhol team from literature [12, 22–24]. Secondly, the preprocessing operation normalizes the data and uses four window slicing strategies. Then, the preprocessed data are input into CNN for training and evaluation to obtain the best strategy. Once the optimal strategy is determined, the best feature extraction model nodes and choice should be discussed. Finally, after performing feature extraction, the extracted features are input into the ML model to classify. Figure 5 shows the main structure diagram of the experiment.

4.1. Experimental Environment. In this study, the experimental equipment is based on the Windows 10 system. The main hardware devices are 4-Core Xeon(R) W-2104 CPU @3.20 GHz and NVIDIA QuADro P2200 GPU with 4G video memory. Main software environments are PyChram 2021, Python 3.6.0, CUDA 9.0, CUDNN 7.0, TensorFlow GPU 2.2.0, Anaconda 3.5.0, and Keras GPU 2.4.3 version.

4.2. Evaluation Indicators. Two evaluation indicators, image-level accuracy and patient-level accuracy, are used to evaluate the classification performance of BC at different magnifications. The calculation formula of the image-level accuracy is shown as follows:

$$\text{image} - \text{level Accuracy} = \frac{\text{Number}_{\text{correct}}}{\text{Number}_{\text{image}}}, \tag{3}$$

where $\text{Number}_{\text{image}}$ represents the total number of all BC images. $\text{Number}_{\text{correct}}$ represents the number of correctly classified BC images. In the course of routine diagnosis, the pathologists need to evaluate the patient's overall cancer images to confirm health status. Therefore, patient-level accuracy is considered as an important evaluation indicator. The mathematical calculation formula is shown as follows:

$$\text{patient} - \text{level accuracy} = \frac{\sum_{p=1}^{Number_p}\left(\text{Number}_{\text{correct}}^{p}/\text{Number}_{\text{image}}^{p}\right)}{\text{Number}_{\text{patient}}}, \tag{4}$$

where $\text{Number}_{\text{patient}}$ represents the total number of patients. $\text{Number}_{\text{correct}}^{p}$ represents the number of correctly classified BC images of patient P, and $\text{Number}_{\text{image}}^{p}$ represents the total number of BC images of patient P. Due to the unbalanced distribution of the benign and malignant data, balanced accuracy (BAC) and F1 score are adopted to objectively evaluate the classification performance of the model. The mathematical calculation formulas of the two are shown as follows:

$$\text{balanced accuracy} = \frac{\text{recall} + \text{specificity}}{2}, \tag{5}$$

$$F1 = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}. \tag{6}$$

4.3. Experiment

4.3.1. Data Splitting. In this study, the dataset is split by the method of fivefold stratified shuffle split cross-validation. The cross-validation method can objectively evaluate the performance of the model and effectively avoid the evaluation result falling into the local optimal state. In the process of splitting the training set and the test set, 82 groups of patients are used as the splitting benchmark. 57 patients are in the training set of about 70%, while others are in the test set of about 30%. This data splitting method is the same as that of F. A. Spanhol team. When comparing the algorithm performance with it, the contribution of this study can be evaluated more objectively. The number of benign and malignant patients is 17 : 40 in the training set, while 7:18 in the test set. Both sets have similar class distribution, which is beneficial to objectively evaluate the model performance. The training set and the test set are shown in Table 8.

4.3.2. Convolutional Neural Network Classification. The data are normalized according to calculation formula (2). Then, we try to use four window slicing strategies to preprocess training data in Table 5. The preprocessed data are input into the model based on the CNN model for training and evaluation. As shown in Table 5, the data in Table 8 have

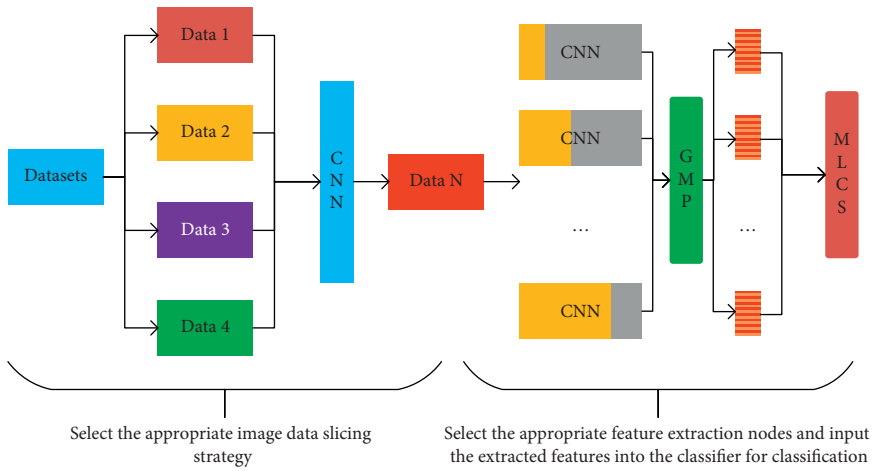


Figure 5: Experimental structure. Data 1 are obtained by Sliding window slicing. Data 2 are obtained by Random window slicing. Data 3 are obtained by Sliding + Random window slicing. Data 4 are obtained by Sliding + Class Balance Random window slicing. Data N are obtained by the best slicing strategy. The orange region represents the effective feature extraction region. The gray region represents the invalid feature extraction region. GMP represents the global max pooling layer. MLCS represents the abbreviation of ML classifier set.

been increased several times on different levels. The model structure of the CNN is shown in Table 6.

In the process of model evaluation, the small image of the test set is obtained by sliding in the test set image. The Sliding window is $128 \times 128 \times 3$ size, and the step is 128 pixels. The class of the large image is decided by voting. The voting rule is that the maximum number of image classes serves as the final prediction result for the whole image. Though other literature studies have not focused on classification problems from an imbalance view, we use F1 scores and BAC to evaluate model performance. The pathological tissue image classification results of BC at different magnifications are shown in Tables 9 and Tables 10.

As shown in Table 9 and Table 10, Sliding + Class Balance Random window slicing strategy has an excellent overall result on accuracy at different magnifications. Regarding the performance at image-level accuracy, the performance is more stable and has better evaluation results in F1 score and BAC accuracy. This experimental result shows that the presented method can alleviate the class imbalance problems. Regarding the performance at the patient-level accuracy, Sliding + Class Balance Random window slicing strategy is superior to other strategies in all magnifications. In a word, the Sliding + Class Balance Random window slicing strategy is more suitable for the following studies.

4.3.3. Feature Extraction and Machine-Learning Model Classification. For feature extraction, a well-trained CNN model takes convolution + pooling layer as feature extraction nodes after preprocessing of Sliding + Class Balance Random window slicing. For the ML model classification, the extracted features are input into the ML model for training and testing. Except for the MFNN, the hyperparameters of other classifiers (SVM, LR, GNB, DT, and RF) are all the default parameters in the Python-SkLearn ML Toolkit.

The combination of convolution and pooling layer is set standard for feature extraction nodes. Then, the global max pooling layer is added after it. There are three cases in the number of input layer neural networks in Figure 4(i). In the MFNN training process, the three cases share the model structure and the hyperparameter configuration as shown in Tables 11 and 12. The features are extracted from the CNN of different nodes on the trained model and input into ML for classification. The accuracy, F1 score, and BAC are shown in Tables 13–15.

Table 13 shows that most ML classifiers have a better classification effect at the second node's position at different magnifications. Compared with other ML classifiers, SVM has the best classification results. Regarding the performance at image-level accuracy, the mean and standard deviation at different magnifications achieve 87.85%, 86.68%, 87.75%, and 85.30%, respectively. The results are 2.81%, 1.02%, 0.78%, and 1.01% higher than the pure CNN (Sliding + Class Balance Random) in Table 9. Besides, regarding the performance at patient-level accuracy, the mean and standard deviation at different magnifications are 87.93%, 87.41%, 88.76%, and 85.55%, respectively. The results of SVM are 1.87%, 1.12%, 0.07%, and 1.98% higher than the pure CNN (Sliding + Class Balance Random) in Table 9.

In Table 14, compared with other classifiers, SVM has the highest F1 score at the second node. The mean and standard deviation at different magnifications achieve 91.12%, 93.30%, 92.54%, and 90.45%, respectively. The results of SVM are 1.13%, 1.37%, 0.51%, and 1.04% higher than the pure CNN (Sliding + Class Balance Random) in Table 10.

In Table 15, compared with other classifiers, SVM has the highest BAC score at the second node. The mean and standard deviation at different magnifications achieve 86.57%, 87.64%, 89.42%, and 85.31%, respectively. The results of SVM are 1.08%, 1.62%, 1.42%, and 1.56% higher than the BAC at different magnifications of image classification by a pure CNN (Sliding + Class Balance Random) in Table 10.

In summary, the intermediate- and high-level features are very important for SVM classification. By this means, the

Table 8: Number distribution of the dataset with different magnifications.

Fold	Data type	Magnifications				Count
		40×	100×	200×	400×	
1	Train	1365 (392 : 973)	1396 (410 : 986)	1394 (414 : 980)	1248 (381 : 867)	5403 (1597 : 3806)
	Test	630 (233 : 397)	685 (234 : 451)	619 (209 : 410)	572 (207 : 365)	2506 (883 : 1623)
2	Train	1405 (436 : 969)	1449 (450 : 999)	1417 (432 : 985)	1291 (425 : 866)	5562 (1743 : 3819)
	Test	590 (189 : 401)	632 (194 : 438)	596 (191 : 405)	529 (163 : 366)	2347 (737 : 1610)
3	Train	1304 (378 : 926)	1383 (389 : 994)	1343 (389 : 954)	1180 (348 : 832)	5210 (1504 : 3706)
	Test	691 (247 : 444)	698 (255 : 443)	670 (234 : 436)	640 (240 : 400)	2699 (976 : 1723)
4	Train	1439 (441 : 998)	1553 (473 : 1080)	1448 (439 : 1009)	1291 (418 : 873)	5731 (1771 : 3960)
	Test	556 (184 : 372)	528 (171 : 357)	565 (184 : 381)	529 (170 : 359)	2718 (709 : 1469)
5	Train	1422 (455 : 967)	1530 (491 : 1039)	1458 (452 : 1006)	1339 (440 : 899)	5749 (1838 : 3911)
	Test	573 (170 : 403)	551 (153 : 398)	555 (171 : 384)	481 (148 : 333)	2160 (642 : 1518)

"Train" represents the training dataset, and "Test" represents the test dataset, and the number in braces represents the number of benign cancer and malignant cancer, respectively.

Table 9: Image-level accuracy and patient-level accuracy evaluation results at different magnifications (/%).

Accuracy	Strategy	Magnifications			
		40×	100×	200×	400×
Image-level	Sliding	84.19 ± 2.27	83.80 ± 2.60	86.48 ± 2.80	83.93 ± 3.93
	Random	82.54 ± 2.81	84.41 ± 1.53	86.72 ± 2.77	82.74 ± 3.16
	Sliding + Random	84.50 ± 1.26	84.70 ± 1.25	86.95 ± 1.52	83.25 ± 2.47
	Sliding + class Balance Random	**85.04 ± 0.87**	**85.66 ± 1.07**	**86.97 ± 0.77**	**84.29 ± 2.36**
Patient-level	Sliding	85.70 ± 1.90	83.93 ± 2.12	87.73 ± 3.32	**83.70 ± 4.04**
	Random	83.73 ± 1.76	84.46 ± 2.52	88.31 ± 2.47	82.75 ± 3.35
	Sliding + Random	86.05 ± 2.17	86.03 ± 1.92	88.18 ± 3.31	83.31 ± 2.43
	Sliding + Class Balance Random	**86.06 ± 1.62**	**86.29 ± 2.85**	**88.69 ± 4.17**	83.57 ± 5.04

The optimal value of the evaluation result has been bolded.

Table 10: F1 score and BAC evaluation results at different magnifications (/%).

Evaluation	Strategy	Magnifications			
		40×	100×	200×	400×
F1 score	Sliding	88.61 ± 1.64	88.69 ± 0.63	90.30 ± 2.15	88.07 ± 3.03
	Random	87.38 ± 1.95	89.04 ± 0.96	91.13 ± 3.01	87.28 ± 2.40
	Sliding + Random	89.31 ± 2.23	90.07 ± 0.82	90.97 ± 2.50	87.76 ± 1.45
	Sliding + Class Balance Random	**89.99 ± 1.86**	**91.93 ± 0.71**	**92.03 ± 2.77**	**89.41 ± 4.31**
BAC	Sliding	83.58 ± 2.04	83.75 ± 2.02	86.47 ± 4.37	82.55 ± 5.24
	Random	81.45 ± 2.45	83.92 ± 3.56	86.99 ± 4.48	81.14 ± 4.33
	Sliding + Random	84.73 ± 3.00	85.86 ± 2.24	87.05 ± 4.50	82.07 ± 3.43
	Sliding + Class Balance Random	**85.49 ± 2.07**	**86.02 ± 3.31**	**88.00 ± 5.66**	**83.75 ± 6.28**

Table 11: MFNN model structure.

Layer	Layer type	Number of neurons	Activation
1	Input	(16, 32, 72)	—
2	Fully connected	128	ReLU
3	Dropout	—	—
4	Fully connected	64	ReLU
5	Dropout	—	—
6	Output	5	Softmax

Table 12: Hyperparameter settings for the MFNN model.

Type	Value
Initial learning rate	$1e-3$
Learning rate schedule	SGD
Weight decay	$1e-5$
Train or test batch size	512
Loss function	Cross-entropy loss function
Number of epochs	500

classification effect of BC pathological tissue images has been improved as a whole.

Literature [23] uses a fully connected layer to extract features and input them into LR. We compare the combination, which is different levels of features and SVM from Table 13 to the former. In Table 16, the results of location points 1, 2, and 3 are better than fc6, fc7, and fc8 on image-level accuracy. Similarly, the results perform better than, or at least comparable to, the former on patient-level accuracy. In particular, Location Point 2 has the best overall results. The experimental results show that the convolutional layer is better than the fully connected layer on feature extraction. Our experimental method is consistent with the comparative literature.

Table 13: Results of feature extraction nodes on image-level accuracy and patient-level accuracy (/%).

Accuracy	Point location	Classifier	Magnifications			
			40×	100×	200×	400×
Image-level	1	SVM	**83.98 ± 2.90**	**84.99 ± 2.91**	**84.66 ± 2.41**	**82.78 ± 4.89**
		LR	78.37 ± 2.72	81.03 ± 5.28	81.40 ± 7.19	80.04 ± 5.00
		GNB	61.96 ± 4.48	74.67 ± 6.82	78.38 ± 4.16	69.93 ± 6.29
		DT	80.60 ± 1.93	80.59 ± 3.29	84.02 ± 4.16	81.56 ± 3.73
		RF	80.43 ± 2.83	81.29 ± 2.93	84.48 ± 1.35	82.47 ± 3.61
		MFNN	79.56 ± 3.20	80.39 ± 4.76	81.76 ± 4.76	79.81 ± 7.76
	2	SVM	**87.85 ± 2.69**	**86.68 ± 2.28**	**87.75 ± 2.37**	**85.30 ± 4.41**
		LR	82.60 ± 3.34	82.48 ± 3.01	85.31 ± 2.20	82.89 ± 2.35
		GNB	81.17 ± 5.00	81.28 ± 3.30	85.91 ± 2.56	77.34 ± 6.61
		DT	80.88 ± 3.92	81.41 ± 2.33	85.33 ± 2.39	82.76 ± 2.90
		RF	80.60 ± 1.93	82.09 ± 2.82	85.68 ± 1.95	82.80 ± 3.9+
		MFNN	80.69 ± 3.15	82.94 ± 3.45	86.71 ± 3.14	83.34 ± 3.09
	3	SVM	**84.91 ± 1.58**	84.60 ± 1.78	**87.38 ± 3.52**	**84.78 ± 4.89**
		LR	81.00 ± 2.05	81.30 ± 1.30	85.23 ± 3.38	82.39 ± 4.94
		GNB	80.89 ± 3.88	81.36 ± 0.85	85.00 ± 3.15	80.48 ± 6.72
		DT	81.28 ± 2.40	80.94 ± 2.40	84.83 ± 2.46	81.89 ± 4.47
		RF	80.60 ± 1.93	81.31 ± 1.92	84.75 ± 2.51	82.43 ± 4.21
		MFNN	80.49 ± 2.85	**84.69 ± 0.86**	87.18 ± 3.63	82.80 ± 5.03
Patient-level	1	SVM	**81.93 ± 3.92**	80.96 ± 3.61	82.40 ± 2.37	81.17 ± 3.04
		LR	79.10 ± 3.91	80.81 ± 4.68	80.44 ± 5.33	78.21 ± 5.42
		GNB	68.16 ± 5.94	75.19 ± 3.35	78.94 ± 3.05	70.64 ± 6.80
		DT	79.81 ± 2.61	79.47 ± 3.72	83.63 ± 2.67	80.08 ± 3.13
		RF	80.67 ± 3.46	80.43 ± 3.17	**83.76 ± 4.16**	**81.50 ± 2.91**
		MFNN	80.34 ± 4.64	**81.06 ± 4.53**	81.92 ± 5.05	78.84 ± 3.72
	2	SVM	**87.93 ± 3.91**	**87.41 ± 3.26**	**88.76 ± 2.50**	**85.55 ± 4.03**
		LR	82.73 ± 2.84	82.76 ± 3.36	85.85 ± 2.90	81.60 ± 2.99
		GNB	78.65 ± 4.36	82.11 ± 3.92	84.47 ± 2.59	77.66 ± 5.07
		DT	81.57 ± 4.50	79.48 ± 3.32	85.15 ± 2.49	81.58 ± 2.84
		RF	80.96 ± 2.98	81.37 ± 3.55	84.54 ± 2.62	82.88 ± 2.90
		MFNN	83.32 ± 4.67	82.93 ± 4.72	87.64 ± 3.85	83.95 ± 2.63
	3	SVM	**83.91 ± 0.96**	83.49 ± 2.65	87.89 ± 3.46	82.28 ± 4.11
		LR	82.10 ± 0.97	82.44 ± 2.62	84.77 ± 3.75	81.10 ± 4.37
		GNB	82.35 ± 3.48	**84.64 ± 1.39**	83.93 ± 2.38	82.44 ± 5.14
		DT	80.82 ± 2.52	79.98 ± 2.54	84.97 ± 2.69	81.40 ± 4.12
		RF	80.72 ± 2.63	81.00 ± 2.77	84.42 ± 2.32	82.29 ± 3.92
		MFNN	82.58 ± 1.27	82.98 ± 2.83	**87.95 ± 3.93**	**82.73 ± 3.88**

Table 14: Extracted features from different feature extraction nodes on image-level F1 score (/%).

Point location	Classifier	Magnifications			
		40×	100×	200×	400×
1	SVM	**86.33 ± 2.05**	**87.17 ± 2.28**	87.76 ± 2.05	84.28 ± 1.92
	LR	82.80 ± 2.41	85.21 ± 5.32	84.93 ± 7.62	83.67 ± 4.42
	GNB	74.81 ± 3.25	80.91 ± 2.67	82.54 ± 4.03	73.57 ± 4.86
	DT	85.04 ± 1.42	85.69 ± 2.43	88.20 ± 1.27	86.08 ± 3.13
	RF	83.69 ± 3.59	86.19 ± 2.16	**88.52 ± 0.89**	82.71 ± 6.64
	MFNN	84.07 ± 2.74	85.15 ± 5.0	85.91 ± 4.43	**84.62 ± 4.00**
2	SVM	**91.12 ± 3.77**	**93.30 ± 4.71**	**92.54 ± 2.77**	**90.45 ± 3.41**
	LR	87.97 ± 2.10	88.00 ± 2.93	88.73 ± 1.85	86.21 ± 1.51
	GNB	86.61 ± 3.68	86.86 ± 2.95	89.62 ± 2.42	86.05 ± 5.27
	DT	85.84 ± 2.71	86.74 ± 1.60	89.45 ± 1.86	86.23 ± 2.08
	RF	84.21 ± 2.87	87.45 ± 2.00	90.76 ± 1.35	88.23 ± 5.47
	MFNN	86.58 ± 2.40	87.33 ± 3.15	90.19 ± 2.41	87.67 ± 2.74
3	SVM	**89.21 ± 0.95**	**90.98 ± 1.50**	**92.79 ± 2.66**	**89.83 ± 3.80**
	LR	87.34 ± 1.27	87.55 ± 1.10	88.69 ± 2.54	86.56 ± 3.85
	GNB	85.12 ± 2.58	87.68 ± 1.61	89.55 ± 2.45	86.01 ± 5.00
	DT	85.08 ± 1.71	86.27 ± 2.12	89.62 ± 3.91	86.23 ± 3.40
	RF	84.80 ± 3.06	87.09 ± 1.48	90.04 ± 1.64	86.67 ± 3.21
	MFNN	87.71 ± 1.40	88.98 ± 0.68	90.80 ± 2.75	85.91 ± 2.40

Table 15: Extracted features from different feature extraction nodes are input into the image-level BAC evaluation results in the ML model (/%).

Point location	Classifier	Magnifications			
		40×	100×	200×	400×
1	SVM	**80.07 ± 4.28**	**80.90 ± 3.45**	81.77 ± 3.31	80.24 ± 3.98
	LR	76.80 ± 3.66	79.74 ± 4.90	80.44 ± 4.88	78.75 ± 4.74
	GNB	51.83 ± 8.18	69.25 ± 4.68	82.54 ± 4.03	69.74 ± 5.36
	DT	77.64 ± 3.93	78.07 ± 5.01	82.27 ± 2.84	79.47 ± 3.78
	RF	79.20 ± 3.99	78.91 ± 4.52	**82.65 ± 2.23**	**80.39 ± 3.95**
	MFNN	77.64 ± 3.86	78.98 ± 5.46	80.18 ± 5.30	77.52 ± 4.37
2	SVM	**86.57 ± 4.06**	**87.64 ± 3.74**	**89.42 ± 3.31**	**85.31 ± 5.10**
	LR	82.68 ± 4.60	84.05 ± 3.79	86.82 ± 3.46	79.05 ± 3.46
	GNB	73.84 ± 6.93	82.81 ± 2.31	88.62 ± 2.42	80.74 ± 5.36
	DT	80.87 ± 5.54	79.50 ± 4.10	83.68 ± 3.51	79.63 ± 3.88
	RF	80.40 ± 4.40	80.67 ± 4.80	84.04 ± 2.97	80.81 ± 4.38
	MFNN	80.33 ± 4.56	81.54 ± 5.39	87.16 ± 5.02	80.93 ± 3.75
3	SVM	**83.21 ± 1.49**	84.74 ± 3.22	**87.88 ± 4.76**	**83.88 ± 5.56**
	LR	81.49 ± 1.73	82.80 ± 2.93	85.75 ± 4.96	80.37 ± 5.46
	GNB	76.45 ± 5.73	**87.68 ± 1.61**	87.54 ± 2.45	79.95 ± 7.65
	DT	79.66 ± 3.51	78.85 ± 3.60	82.76 ± 3.71	79.91 ± 5.17
	RF	79.65 ± 3.97	79.27 ± 3.16	83.29 ± 3.15	80.52 ± 4.96
	MFNN	81.97 ± 2.03	83.29 ± 3.11	86.94 ± 5.15	82.11 ± 5.14

Table 16: Combination with SVM is compared with [23] (/%).

Accuracy	Location point	Magnifications			
		40×	100×	200×	400×
Image-level	1	83.98 ± 2.90	84.99 ± 2.91	84.66 ± 2.41	82.78 ± 4.89
	2	**87.85 ± 2.69**	**86.68 ± 2.28**	**87.75 ± 2.37**	**85.30 ± 4.41**
	3	84.91 ± 1.58	84.60 ± 1.78	87.38 ± 3.52	84.78 ± 4.89
	fc6	83.00 ± 2.60	84.60 ± 5.00	84.00 ± 2.80	81.10 ± 3.90
	fc7	83.10 ± 2.10	83.30 ± 4.60	84.10 ± 1.50	81.60 ± 3.70
	fc8	83.20 ± 2.40	84.00 ± 4.90	83.40 ± 1.10	80.90 ± 3.70
Patient-level	1	81.93 ± 3.92	80.96 ± 3.61	82.40 ± 2.37	81.17 ± 3.04
	2	**87.93 ± 3.91**	**87.41 ± 3.26**	**88.76 ± 2.50**	**85.55 ± 4.03**
	3	83.91 ± 0.96	83.49 ± 2.65	87.89 ± 3.46	82.28 ± 4.11
	fc6	82.50 ± 8.60	83.60 ± 8.50	85.40 ± 5.20	81.10 ± 9.00
	fc7	83.40 ± 6.70	83.10 ± 8.40	86.00 ± 3.70	81.60 ± 8.60
	fc8	83.40 ± 6.90	83.80 ± 8.50	85.80 ± 3.50	80.70 ± 9.10

The fc6, fc7, and fc8 are the three top-most layers in [23].

4.3.4. Comparison with the Literature. To verify the performance of the proposed method, we compare the optimal combination with relevant literature. To prove the validity of the comparative data, the experimental methods are the same as those in the compared literature. We use a fivefold stratified shuffle split cross-validation method, and the ratio of the training set and the test set is 7 : 3. Table 17 and Table 18 show the results of the comparison on image-level accuracy and patient-level accuracy.

In Table 17, regarding the performance at image-level accuracy the classification results are higher than those models in the recent literatures at different magnifications. In Table 18, regarding the performance at patient-level accuracy, the results indicate that the proposed method performs better than, or at least comparable to, methods in terms of the quality of the classification from literature studies. Our method is not all superior to the optimal values of different magnifications compared with literature studies, but it is more accurate and stable in terms of overall model performance classification. In particular, the classification results on the patient-level accuracy are better than others at 200× magnification.

Table 17: Literature comparison results on image-level accuracy (/%).

Magnifications	40×	100×	200×	400×
Literature [12]	—	—	—	—
Literature [22]	85.60 ± 4.80	83.50 ± 3.90	82.70 ± 1.70	80.30 ± 2.90
Literature [23]	84.60 ± 2.90	84.80 ± 4.20	84.20 ± 1.70	81.60 ± 3.70
Literature [24]	87.80 ± 5.60	85.60 ± 4.30	80.80 ± 2.80	82.90 ± 4.10
Literature [27]	—	—	—	—
Our method	**87.85 ± 2.69**	**86.68 ± 2.28**	**87.75 ± 2.37**	**85.30 ± 4.41**

Table 18: Literature comparison results on patient-level accuracy (/%).

Magnifications	40×	100×	200×	400×
Literature [12]	81.60 ± 3.00	79.90 ± 5.40	85.10 ± 3.10	82.30 ± 3.80
Literature [22]	90.60 ± 6.70	88.40 ± 4.80	84.60 ± 4.20	**86.10 ± 6.20**
Literature [23]	84.60 ± 2.90	84.80 ± 4.20	84.20 ± 1.70	81.60 ± 3.70
Literature [24]	**92.10 ± 5.90**	**89.10 ± 5.20**	87.20 ± 4.30	82.70 ± 3.00
Literature [27]	83.08 ± 2.10	83.17 ± 3.50	84.63 ± 2.70	82.10 ± 4.40
Our method	87.93 ± 3.91	87.41 ± 3.26	**88.76 ± 2.50**	85.55 ± 4.03

5. Conclusions

In this study, we propose a novel classification method based on deep features of different levels to solve the BC classification problems. In the stage of data preprocessing, we present four slicing methods: Sliding window slicing, Random window slicing, Sliding + Random window slicing, and Sliding + Class Balance Random window slicing. The experimental results show that Sliding window slicing strategy can guarantee model's fitting training, and Random window slicing can enhance model's generalization training. Then, we combine the characteristics of both strategies and propose Sliding + Random window slicing strategy. This strategy performs well on model fitting ability and generalization ability. To overcome the classification problem caused by the unbalanced distribution of classes, we propose a Sliding + Class Balance Random window slicing. By comparing model performance, Sliding + Class Balance Random window slicing is the best data preprocessing strategy. In the stage of deep feature classification, features of different levels are combined with ML classifier, and the combination of intermediate- and high-level features with SVM has the best classification performance. The proposed method and some state-of-the-art method experiment on the BreaKHis breast cancer dataset. The experimental result shows that the proposed method can obtain better results than the ones reported in the relevant literature. It has been concluded that our method can be efficiently used for solving these problems due to its simplicity, reliability, and robustness. Although the method of benign and malignant BC has been presented, the classification of BC's subtypes has not been studied, and each subtype has a large imbalance ratio. In future research work, we aim to provide a DL algorithm for subtype classification.

Data Availability

The authors train and evaluate the method based on the BreaKHis datasets provided by F. A. Spanhol's team. This dataset is available at https://web.inf.ufpr.br/vri/databases/breast-cancer-histopathological-database-breakhis/

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the submission and publication of this manuscript.

Acknowledgments

This work was substantially supported by the Daqing Normal University Land Enterprise Application Cultivation Project under no. 19ZR16.

References

[1] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, 2018.

[2] K. P. Trayes and S. E. Cokenakes, "Breast cancer treatment," *American Family Physician*, vol. 104, no. 2, pp. 171–178, 2021.

[3] G. Murtaza, L. Shuib, A. W. Wahid et al., "Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges," *Artificial Intelligence Review*, vol. 53, no. 3, pp. 1655–1720, 2020.

[4] J. Huang, P. S. Chan, fnm Lok et al., "Global incidence and mortality of breast cancer: a trend analysis," *Aging*, vol. 13, no. 4, pp. 5748–5803, 2021.

[5] M. G. Hanna, V. E. Reuter, J. Samboy et al., "Implementation of digital pathology offers clinical and operational increase in efficiency and cost savings," *Archives of Pathology & Laboratory Medicine*, vol. 143, no. 12, pp. 1545–1555, 2019.

[6] J. A. Retamero, J. Aneiros-Fernandez, and R. G. del Moral, "Complete digital pathology for routine histopathology diagnosis in a multicenter hospital network," *Archives of Pathology & Laboratory Medicine*, vol. 144, no. 2, pp. 221–228, 2020.

[7] G. Litjens, T. Kooi, fnm Bejnordi et al., "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[8] D. Li, D. Wang, fnm Dong et al., "False-negative results of real-time reverse-transcriptase polymerase chain reaction for severe acute respiratory syndrome coronavirus 2: role of deep-

learning-based CT diagnosis and insights from two cases," *Korean Journal of Radiology*, vol. 21, no. 4, pp. 505–508, 2020.

[9] L. Cong, W. Feng, Z. Yao, X. Zhou, and W. Xiao, "Deep learning model as a new trend in computer-aided diagnosis of tumor pathology for lung cancer," *Journal of Cancer*, vol. 11, no. 12, pp. 3615–3622, 2020.

[10] A. Ruiz-Casado, A. Álvarez-Bustos, C. G. de Pedro, M. Méndez-Otero, and M. Romero-Elías, "Cancer-related fatigue in breast cancer survivors: a review," *Clinical Breast Cancer*, vol. 21, no. 1, pp. 10–25, 2021.

[11] M. Maleki, A. Mardani, M. Ghafourifard, and M. Vaismoradi, "Qualitative exploration of sexual life among breast cancer survivors at reproductive age," *BMC Women's Health*, vol. 21, no. 1, pp. 1–10, 2021.

[12] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1455–1462, 2015.

[13] A. Alba-Bernal, R. Lavado-Valenzuela, M. E. Domínguez-Recio et al., "Challenges and achievements of liquid biopsy technologies employed in early breast cancer," *EBioMedicine*, vol. 62, Article ID 103100, 2020.

[14] A. E. Giuliano, "The evolution of sentinel node biopsy for breast cancer: personal experience," *Breast Journal*, vol. 26, no. 1, pp. 17–21, 2020.

[15] J. Bai, R. Posner, T. Wang, C. Yang, and S. Nabavi, "Applying deep learning in digital breast tomosynthesis for automatic breast cancer detection: a review," *Medical Image Analysis*, vol. 71, Article ID 102049, 2021.

[16] A. Khamparia, S. Bharati, P. Podder et al., "Diagnosis of breast cancer based on modern mammography using hybrid transfer learning," *Multidimensional Systems and Signal Processing*, vol. 32, no. 9, pp. 747–765, 2021.

[17] M. Sharma, S. Gupta, P. Sharma, and D. Gupta, "Bio-inspired algorithms for diagnosis of breast cancer," *International Journal of Innovative Computing and Applications*, vol. 10, no. 3/4, pp. 164–174, 2019.

[18] R. Ha, S. Mutasa, N. Gupta et al., "Predicting breast cancer molecular subtype with MRI dataset utilizing convolutional neural network algorithm," *Journal of Digital Imaging*, vol. 32, no. 2, pp. 276–282, 2019.

[19] A. Cruz-Roa, A. Basavanhally, F. González et al., "Automatic detection of invasive ductalcarcinoma in whole slide images with convolutional neural networks," *Proceedings of SPIE-The International Society for Optical Engineering*, vol. 9041, pp. 139–144, 2014.

[20] A. Anuranjeeta, K. A. Tiwari, and S. Sharma, "Classification of histopathological images of breast cancerous and non cancerous cells based on morphological features," *Biomedical and Pharmacology Journal*, vol. 10, no. 1, pp. 353–366, 2017.

[21] L. P. Coelho, A. Ahmed, A. Arnold et al., "Structured literature image finder: extracting information from text and images in biomedical literature," *Linking Literature, Information, and Knowledge for Biology*, vol. 6004, pp. 23–32, 2010.

[22] F. A. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "Breast cancer histopathological image classification using convolutional neural networks," in *Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 2560–2567, Vancouver, Canada, July 2016.

[23] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, "Deep features for breast cancer histopathological image classification," *IEEE International Conference on Systems*, pp. 1868–1873, 2017.

[24] F. A. Spanhol, L. S. Oliveira, P. R. Cavalin, C. Petitjean, and L. Heutte, "Multiple instance learning for histopathological breast cancer image classification," *Expert Systems with Applications*, vol. 117, pp. 103–111, 2019.

[25] L. G. Hafemann, L. S. Oliveira, and P. Cavalin, "Forest species recognition using deep convolutional neural networks," in *Proceedings of the International Conference on Pattern Recognition(ICPR)*, pp. 1103–1107, Stockholm, Sweden, August 2014.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097–1105, 2012.

[27] N. Bayramoglu, J. Kannala, and J. Heikkila, "Deep learning for magnification independent breast cancer histopathology image classification," in *Proceedings of the International Conference on Pattern Recognition(ICPR)*, pp. 2441–2446, Cancun, Mexico, December 2017.

Hindawi
Mathematical Problems in Engineering
Volume 2023, Article ID 9798161, 1 page
https://doi.org/10.1155/2023/9798161

Retraction

Retracted: Construction of a Multimedia-Based University Ideological and Political Big Data Cloud Service Teaching Resource Sharing Model

Mathematical Problems in Engineering

Received 11 July 2023; Accepted 11 July 2023; Published 12 July 2023

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

(1) Discrepancies in scope
(2) Discrepancies in the description of the research reported
(3) Discrepancies between the availability of data and the research described
(4) Inappropriate citations
(5) Incoherent, meaningless and/or irrelevant content included in the article
(6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

[1] J. Feng, W. Zhang, and S. Tsai, "Construction of a Multimedia-Based University Ideological and Political Big Data Cloud Service Teaching Resource Sharing Model," *Mathematical Problems in Engineering*, vol. 2021, Article ID 9907630, 12 pages, 2021.

Hindawi
Mathematical Problems in Engineering
Volume 2021, Article ID 9907630, 12 pages
https://doi.org/10.1155/2021/9907630

Research Article

Construction of a Multimedia-Based University Ideological and Political Big Data Cloud Service Teaching Resource Sharing Model

Jian Feng,[1,2] **Weiliang Zhang**,[1] **and Sang-Bing Tsai**[3]

[1]*School of Marxism, Central South University, Hunan 410083, Changsha, China*
[2]*College of Marxism, Sichuan University of Science & Engineering, Sichuan 643000, Zigong, China*
[3]*Regional Green Economy Development Research Center, School of Business, Wuyi University, Nanping, China*

Correspondence should be addressed to Jian Feng; 2064307060@qq.com

Received 21 October 2021; Revised 9 November 2021; Accepted 23 November 2021; Published 29 December 2021

Academic Editor: Xiangtao Li

In this paper, the intelligent education cloud service platform is first constructed in view of the high cost consumption in the process of resource sharing scheduling in colleges and universities. Secondly, the hierarchical education resource sharing grid model is proposed. Specifically, according to the characteristics of the educational resource grid, the key factors affecting the performance of copy creation strategy are analysed, and a dynamic copy creation strategy is proposed. A multiresource equity distribution mechanism based on the concept of resource sharing equity is further proposed. The mechanism establishes a planning model according to the limited task resource demand and the amount of resources shared by the user at different times so that the global cumulative share vector of superior resources meets the dictionary order optimally. The simulation experiment shows that the grid sharing model proposed in this paper has better performance on the educational cloud service platform. The proposed resource allocation mechanism has achieved good results in ensuring the fair distribution of resources and ensuring high resource utilization when resource sharing users put forward multiple groups of time-changing resource demands.

1. Introduction

The multimedia-based education cloud service platform is a revolutionary driving force that promotes the sustainable evolution of high-quality educational resources, promotes the innovation of educational organization structure, promotes the transformation of education to service mode, and promotes the transformation of business processes [1–4]. Cooperate with the school's education and training plan to cultivate engineering and practical talents adapted to the society [5]. Online learning resources are the core and foundation of the online education process and have always been the focus of attention of experts, scholars, frontline teachers, and even learners [6]. In recent years, cloud computing is gradually penetrating the education field with its massive data storage and convenient network services. If cloud computing-related technologies and services are used to solve the problem of online learning resource sharing, a new idea can be explored to promote the effective sharing of online learning resources [7, 8].

Cloud computing is centralized computing and access on cluster servers in the network. Yao et al. [9] mainly studied how to achieve system integration and resource sharing and improve hardware utilization. However, there is a lack of research on massive data processing. Jiang et al. [10] proposed a hybrid architecture model to solve the problem of information islands in enterprises. But there is no introduction to resource integration and how to improve resource utilization. Zhou et al. [11] proposed a cloud-based virtual desktop system to study the use of virtual classrooms in the education field. Barreto et al. [12] proposed to integrate knowledge organization and knowledge management, and regard knowledge as a service. Martinez-Guijosa et al. [13] mainly studied the sharing and utilization of resources but did not solve the problem of unbalanced utilization of resources. Aldowah et al. [14] proposed to model the complex causal relationship in the system by constructing a system model and verify the result-factor relationship existing in the system model through model analysis. Liu [15] proposed a new teaching system that integrates web

services of courseware resources through the SOAP protocol. Al-Samarraie et al. [16] reviewed the transformation from traditional education information to education services under the cloud computing environment and the computing architecture of cloud education.

As the main application environment of the new generation of interconnection networks, the grid provides an effective means to solve the problem of distributed resource sharing [17]. In a grid system, creating a copy is an effective means to improve the quality of service of the system [18]. However, the current research on replica creation strategy is mainly concentrated in the grid system environment represented by the European data grid [19]. Compared with the European data grid, the educational resource grid has differences in the network environment such as raw data distribution, network bandwidth, and node storage capacity. First, this makes the existing copy creation strategy unable to be applied to the educational resource grid. Second, the resource-sharing environment with a single model can no longer meet the needs of various industries [20]. Resource balancing scheduling as an important part of network communication software technology has also become the focus of research in various fields [21–23]. However, the current shared resource balanced scheduling method generally has problems such as too long scheduling completion time and high cost consumption. Currently, Yimin et al. [24] proposed a cloud computing resource scheduling method based on the quantum particle swarm algorithm. This method analyses cloud computing resource scheduling. On this basis, a mathematical model of resource scheduling is constructed, the objective function is given, and the quantum particle swarm algorithm is used to solve it. This method can effectively improve the utilization of resource scheduling. However, it takes too long to complete in the scheduling process. Sun et al. [25] proposed a resource scheduling method based on resource weight and maximum resource utilization. This method analyses resources and obtains the resource weight ratio. Resource scheduling is performed according to the maximum resource utilization of the weight. This method can effectively improve the resource utilization rate of the entire system and quickly solve the load balancing problem. However, the cost required in the process of scheduling is relatively high.

In response to the above problems, we first built a smart education cloud service platform. Secondly, a hierarchical grid model of educational resource sharing is proposed. According to the characteristics of the educational resource grid, the key factors affecting the performance of the replica creation strategy are analysed, and a dynamic replica creation strategy is proposed. Finally, considering a cloud computing resource sharing system with multiple virtual machines, resource requirements for each task submitted by different users may be different. This paper designs a dynamic resource scheduling algorithm, a dynamic multiresource fair allocation algorithm in which users submit multiple sets of limited time-varying task requirements.

2. Overview of Cloud Services

2.1. SOA. Service-oriented architecture (SOA) is the methodology, software design model, and architecture model and the architecture of applications in a distributed environment [26]. The application function is service-oriented with a service-oriented design concept so that the application function is finally used in the form of a service. Services can communicate with each other, services can be reused, and services can operate independently of each other without affecting each other. Services can be reorganized and combined into new services to be used [27].

The core theme of SOA is service. Services are a bunch of processing methods. These processing operations are encapsulated to form a standard service, which is a service interface formed according to the definition of a service contract [28]. SOA is an extension of object-oriented design ideas. Component is an abstraction of the realization of specific functions in object-oriented design ideas. And service is to encapsulate components, and it is also an abstract concept for the realization of higher-level functions. Its specific functional implementation does not require waiting for the implementation processing state of other services to be acquired. It also does not need to rely on the response information context of other services, nor does it require additional information from service requesters. In this way, the service can achieve a high concurrency of request and response [29]. The SOA analysis process is shown in Figure 1.

2.2. Education Cloud. Education cloud [30–32] uses a new generation of basic equipment including large-scale servers and large-scale storage devices as the infrastructure. On this basis, a development and collaborative sharing educational knowledge resource management platform is established, and educational resources are pushed to the terminal through the platform. The education cloud service model can be divided into three levels: education cloud software as a service E-SaaS, education cloud platform as a service E-PaaS, and education cloud infrastructure as a service E-IaaS [33–35]. Education cloud is divided into three service types: IaaS, PaaS, and SaaS, as shown in Figure 2.

SaaS refers to providing educational application software as a service to end users. By deploying the service in the environment of the educational cloud service platform, educational cloud users can lease relevant educational services by using an educational cloud service purchase platform. They can also customize services to meet their needs from the cloud service platform without independent development and purchase.

PaaS refers to the use of software development platform as a service for purchasing users. The platform supports the development and use of a variety of programming languages, builds the running environment of the development software, and has runtime monitoring tools, business process development tools, secure and effective service middleware, and so on to support the safe and stable development and operation of the software.

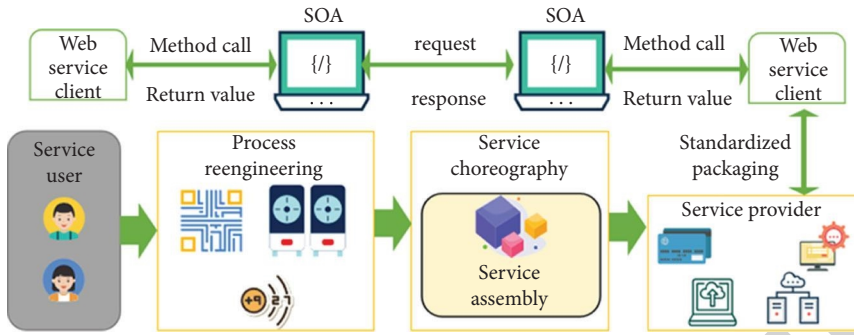


Figure 1: SOA analysis flowchart.

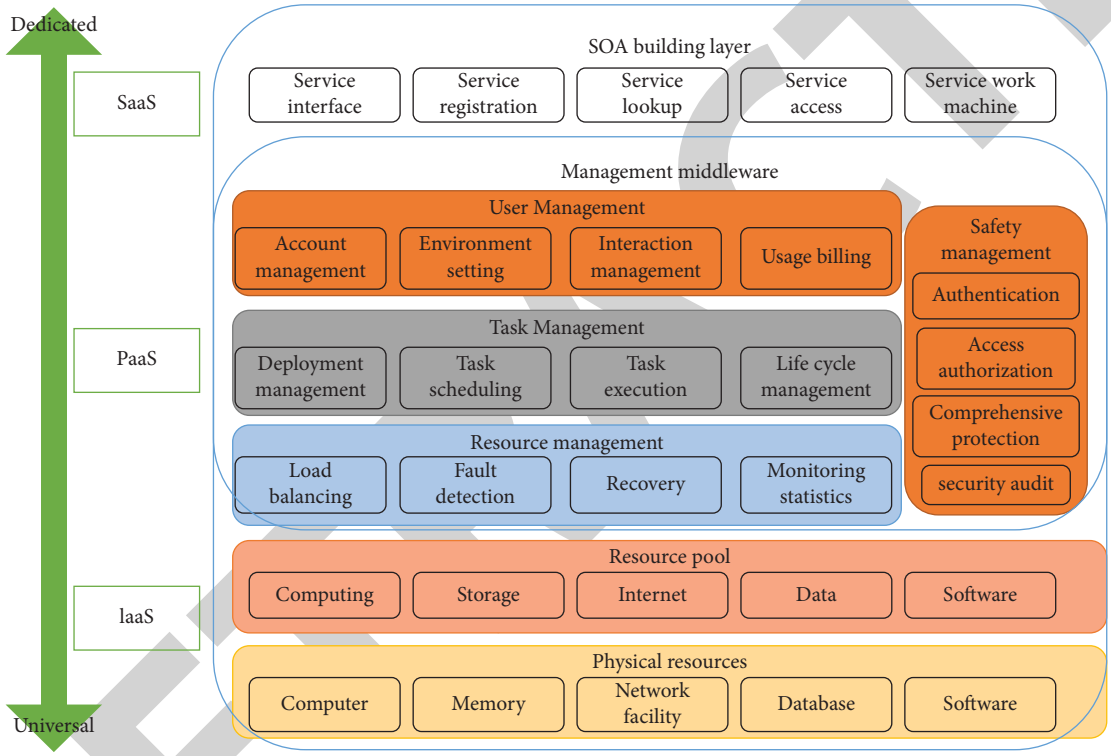


Figure 2: Architecture of educational cloud computing.

IaaS refers to the use of virtualization technology to virtualize servers, storage devices, networks, and other hardware devices into a computable and distributable pool of virtual resources. Users can use the resource pool service provided by the education cloud service platform by time, traffic, and other billing types. Because it is leased on demand, users can adjust their needs according to their own needs, and the resource pool resource management system will flexibly adjust resource allocation according to user needs.

3. Construction of Teaching Resource Sharing Model

3.1. Cloud Service Platform Based on Multimedia. This article establishes a low-cost, multitenant-oriented, scalable, and digital education cloud service platform. Utilizing the resource sharing characteristics of the education cloud service platform, the platform forms a collaborative and innovative digital education alliance based on the accumulation of massive resources and educational organizations. The platform provides decision-making services and collaborative cooperation services for the education alliance.

The layers of the platform are interrelated and independent of each other. The upper layer has implementation dependence characteristics on the lower layer, but the result of the lower layer needs to be delivered to the user by the upper layer. The service characteristics of each layer are different. IaaS provides virtualized IT resources as services to users, PaaS provides virtualized development platforms as services to users, and SaaS provides various teaching functions, teaching office functions, business functions, and decision-making functions as services to users, and the smart

education platform develops, deploys, and manages cloud services. The architecture of the multimedia-based cloud service platform is shown in Figure 3.

3.2. Hierarchical Educational Resource Sharing Grid Model. As a key technology in cloud computing and a new generation of Internet technology, grid has a huge advantage in realizing resource sharing. Resource sharing in this technological environment is a new product of the organic combination of traditional resources, network technology, and computer technology. Its essence is a series of technologies that are generated to meet the internal resource sharing and problem solving of virtual organizations.

The "five-layer hourglass structure" in the grid technology disperses the operation, management, and use functions of shared resources in five different levels. From bottom to top, they are Fabric, Connectivity, Resource, Collective, and Application. Based on this, combined with the three service forms of cloud computing, a layered architecture is adopted to design the network learning resource-sharing system model from the three levels of resource layer, management layer, and application layer, as shown in Figure 4.

The basic operations of educational resource grid users mainly include uploading, retrieving, and downloading data files. After introducing the copy mechanism in the educational resource grid, creating a reasonable copy can effectively increase the speed of user resource retrieval and download and shorten the response time for users to request resources.

In the educational resource grid, a data file can have multiple copies, and each copy has a physical file name (PF) and a logical file name (LF). Each copy corresponds to a unique physical file name, and the logical file name of the same copy is the same. That is, one logical file name corresponds to multiple physical file names. We denote the set of all physical file names corresponding to a logical file name as |PF|s.

When a user sends an access request to a data file, the copy mechanism selects the copy with the shortest response time for the user from multiple copies of the data file. Therefore, the response time of the copy to a specific request directly affects the performance of the copy mechanism. The response time is the sum of the time of the requesting node, the waiting time before transmission, the time of reading and writing files, and so on. The response time (RT) of the duplicate physical file name to a specific request can be calculated by the following formula:

$$\mathrm{RT}(\mathrm{PF}) = FT(\mathrm{PF}) + FA(\mathrm{PF}). \tag{1}$$

Among them, FT represents the time to transfer the physical file name of the file from its storage node to the requesting node. FA means the time it takes to access the file's physical file name, including the waiting time before transmission and the time to read and write the file.

$$FT(\mathrm{PF}) = \frac{\mathrm{PF.size}}{\alpha}. \tag{2}$$

Among them, PF. size represents the size of the copy and α is the minimum bandwidth of the transmission link between the copy storage node and the requesting node (see Figure 4).

Therefore, the network bandwidth between the replica storage node and the requesting node, and the size of the replica directly affects the response time of the replica. The factors that affect the FA include the physical file name, the load of the storage node, the read efficiency of the storage node, and the write efficiency of the requesting node. We have generally considered

$$FT(\mathrm{PF}) \gg FA(\mathrm{PF}). \tag{3}$$

The educational resource grid is composed of several nodes, the nodes form a hierarchical structure, and the nodes are denoted as N. N can contain multiple CUs and SUs. Among them, CU is a computing unit, used to process user's job requests; SU is a storage unit, used to store the original resources and copy resources of the system. Each node maintains a historical access record, and each item in the historical record is a triplet. (SN, PF, f) is used to represent the frequency with which the replica PF in the SU is accessed. (CN, LF, f) is used to represent the frequency of CN requesting data file LF.

When a user sends an access request to a data file, the copy mechanism selects a copy with the shortest response time for the user from multiple copies of the data file.

The file size of the copy PF is the physical file name's size, and the current access frequency is LF.f. The node is N1, and the node with the same LF as PF after PF is deleted is N2; then,

$$\mathrm{PF.c} = \mathrm{PF.size} * \frac{\mathrm{PF.f}}{BW(N_1, N_2)}. \tag{4}$$

Among them, BW is the minimum bandwidth of the transmission link between nodes N1 and N2.

Assume that the size of the logical file LF is LF.size. The current request frequency is LF.f. The node where the newly created LF copy is located is N1. The node of LF before the new copy is N2. So we can define

$$\mathrm{LF.r} = \mathrm{LF.size} * \frac{\mathrm{LF.f}}{BW(N_1, N_2)}. \tag{5}$$

The existing replica creation strategies are either designed for the special grid environment and cannot be applied to the grid system in the ordinary network environment, or consider few network parameters, which cannot well reflect the actual network environment, resulting in frequent establishment and deletion of replicas, and the implementation of some strategies is also limited. Aiming at the problem of sharing small and medium-sized educational resources in the general network environment, this paper puts forward a hierarchical educational resource grid model and defines the function of each node. Then, based on this model, a dynamic replica creation strategy is proposed. This strategy not only fully considers the factors such as network bandwidth, file size, and request frequency when creating replicas but also solves the problem of frequent establishment and deletion of replicas caused by the small storage capacity of grid nodes.

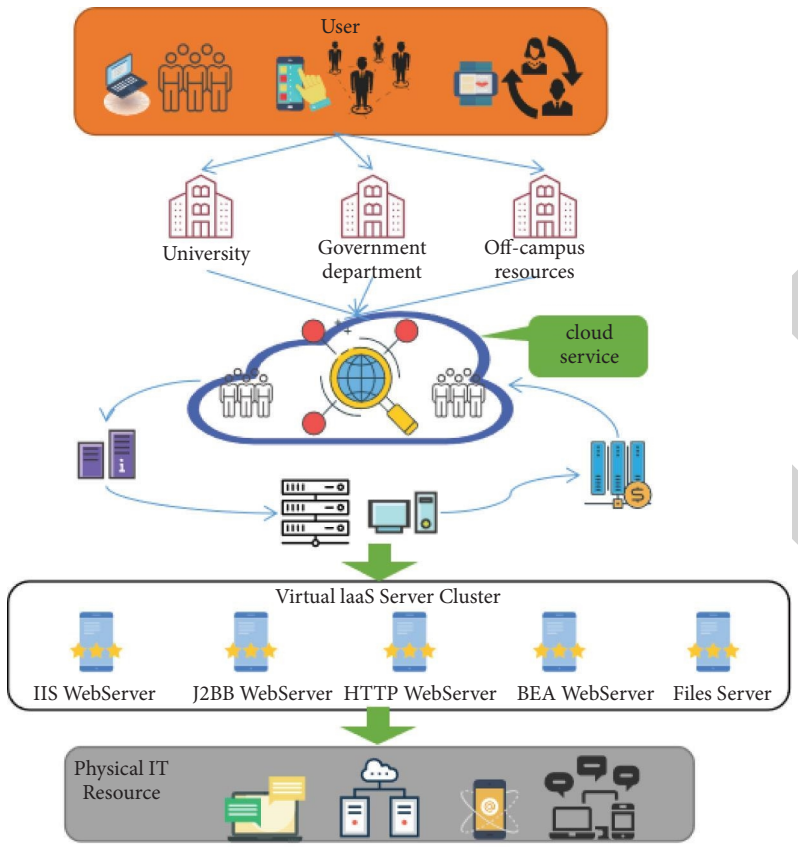


Figure 3: Architecture diagram of a multimedia-based cloud service platform.

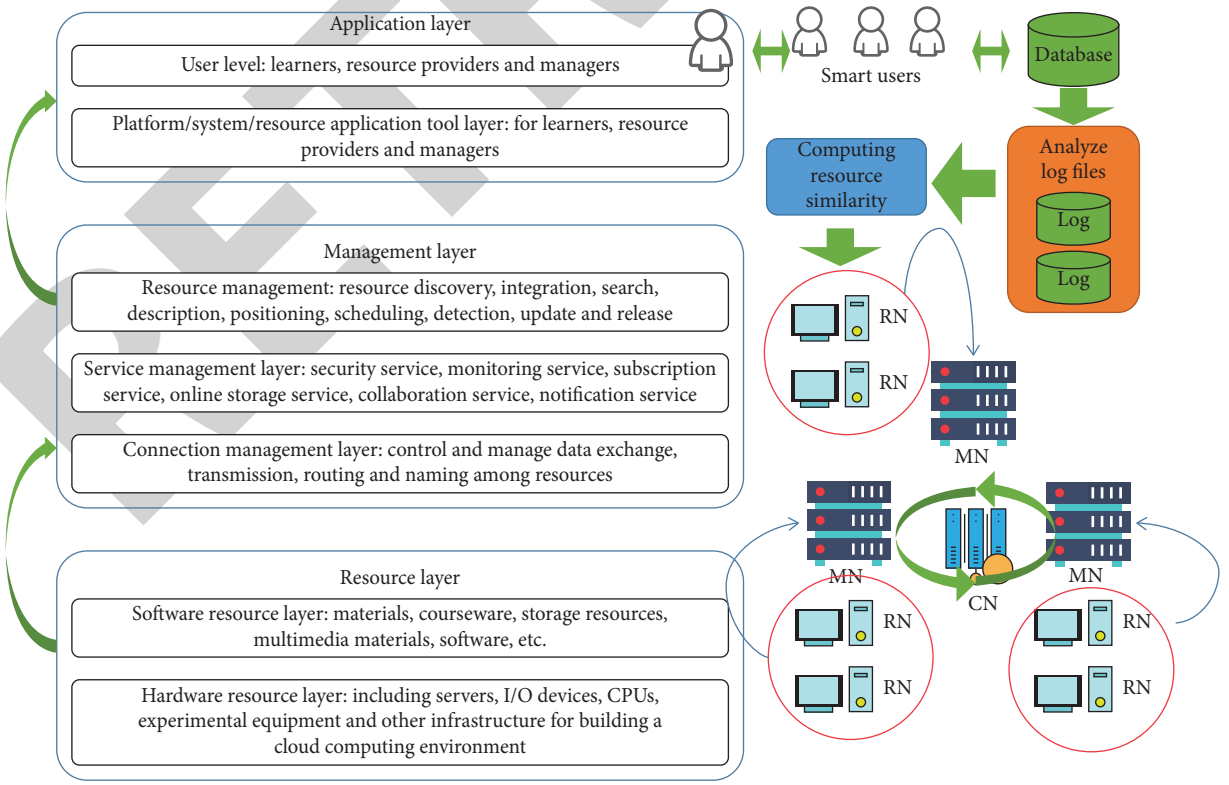


Figure 4: Online learning resource-sharing model based on smart education cloud service platform.

Through actual analysis and application, it is found that, in a local area, the user requests of the educational resource grid sometimes have the characteristics of dispersion. In this case, we introduced a new strategy to solve the problem.

The basic idea of this strategy is similar to that of ABU, which aggregates the same historical records layer by layer from bottom to top. When aggregating, the f of each record with the same parent node in the history record and requesting the same logical file name is aggregated. In addition, the education resource grid generally does not exceed 6 layers from the root node to the bottom node, which are the central node, provincial node, municipal node, county node, school management node, and bottom school resource node. According to the actual situation, within the same period of time, the similarity of resources requested by users in the jurisdiction of a node decreases as the level of the node increases. Therefore, each time it aggregates, it does not aggregate to the central node from the bottom up but only aggregates to the city-level nodes.

3.3. Dynamic Resource Scheduling Algorithm. In this paper, a dynamic resource scheduling algorithm is designed. Our approach is based on a multivirtual machine cloud computing resource sharing system. On this basis, a dynamic multiresource equitable distribution algorithm is proposed for users to submit multiple sets of limited time-changing tasks. Considering a cloud computing resource sharing system with multiple virtual machines, resource requirements for each task submitted by different users may be different. Therefore, it is necessary to select the most suitable virtual machine to run according to the resource requirements of the task submitted by the user to meet higher resource utilization.

We define the remaining resources of the m-th virtual machine in the system at time a as

$$R = (r_1, \ldots, r_m) = (n - m_1(a), n - m_2(a), \ldots, n - m_m(a)). \tag{6}$$

Among them, $ri = n{-}mi\ (a)$ is the remaining amount of the i-th resource in the m virtual machines at time a.

The amount of resources required by each task of user u at time a is

$$Ru_i(a) = (Ru_{i1}(a), Ru_{i2}(a), \ldots, Ru_{im}(a)). \tag{7}$$

Among them, Ru_i is the demand of user u for the i-th resource in each task at time a.

We define the following heuristic formula to measure the matching degree between user task resource requirements and virtual machines to complete the matching between user tasks and virtual machines:

$$Q(u, m) = \frac{Ru_i(a)}{\sum_i Ru_i(a)} - \frac{n - m_i(a)}{\sum_i n - m_i(a)}. \tag{8}$$

When the time a in $Q\ (u, m)$ is smaller, it means that the task demand of user u at time a is more matched with the remaining resource capacity of the m-th virtual machine. We select the virtual machine with the smallest matching degree to allocate task resources to user u.

The main idea of the algorithm has three aspects:

(1) When there are n users and m virtual machines in the system, according to the resource requirements Ru_i (a) of each task of any user $u \in U$ in the system at different times, select the best match in each round according to formula (8). The server m allocates the resources needed to perform a task to the user u. And update user u's cumulative global dominant resource share and global sharing coefficient p, sort from small to large, and select the user with the smallest sharing coefficient among users:

$$u = \min_i (1, n * p). \tag{9}$$

If $p < 1$, find the most matching virtual machine to allocate a task resource to the user u and update the system resource allocation status.

(2) Perform the above steps in sequence. If the minimum global sharing coefficient $p \geq 1$ for all users in the system, it means that all users in the system will no longer share defects. Select the user with the smallest cumulative share of dominant resources among all users:

$$\min\_u = \min (1, n * u), \tag{10}$$

and allocate the resources needed to perform a task for the user min_u and update the system resource allocation status.

(3) Perform Step 1 and Step 2. When a certain resource i, $1 \leq i \leq m$ in the system is allocated, or when all user tasks are executed, the resource allocation ends.

4. Results and Discussion

4.1. Configuration of Teaching Resource Platform and Experimental Settings. This paper uses the CPU and Memory resources required for running 12,951 tasks from the Alibaba cluster dataset [36] as the user resource requirements in this experiment. Each user in the dataset submits multiple sets of task resource requirements in the form of job. Each job contains multiple sets of task requirements. Each set of tasks consists of multiple instances with the same resource requirements. The experimental simulation server configuration in this paper is shown in Table 1, where the number of servers is the server resources of each user. This article assumes that the number of virtual machines is the same as the number of physical servers after users share resources (see Table 1).

4.2. The Impact of Different Copy Strategies on the Sharing Model. The average job execution time refers to the total execution time of jobs submitted by users divided by the number of jobs, which directly reflects the impact of replica creation strategy on grid system performance. Under the same workload, the shorter the average job execution time of the system, the better the performance. Based on this indicator, we, respectively, simulated, compared, and analysed

Table 1: Cluster server configuration.

User number	5	20	100
Per user own machines	1	1	1
CPU capacity	250	250	250
Memory capacity	40	40	40

the impact of factors such as the number of jobs, node storage capacity, and user access patterns on the performance of educational resource grid systems using different copy creation strategies.

4.2.1. Number of Jobs. The storage capacity of the underlying node is set to 100 G and 50 G, respectively, to conduct two sets of experiments to analyse the impact of the number of jobs on the performance of the educational resource grid system using different strategies. It can be seen from Figure 5 that, compared with no copy, various copy placement strategies can significantly reduce the average execution time of the job, which fully verifies that creating a copy can improve the performance of the grid system.

Figure 5(a) is the experimental result when the node storage capacity is 50 G. We observe that when the node storage capacity becomes 50 G, the performance of our strategy is significantly better than the other three strategies, and only when our strategy is adopted, time of the average job execution can be reduced as the number of jobs increases. This is because when the storage capacity of a node is insufficient, our strategy can select neighbour nodes to create a copy instead of frequent copy replacement, which saves a lot of remote file access time.

Figure 5(b) is the experimental result when the node storage capacity is 100 G. First of all, we notice that as the number of jobs increases, the average job execution time of the four copy creation strategies decreases, and the performance is not much different. This is because when the storage capacity of the underlying node is large, these strategies can create copies on the underlying node, reducing the number of remote file accesses. In addition, we can also observe that when the number of jobs is the same, the performance of our strategy is the best, and the performance of the copy creation strategy based on the economic model is the worst (see Figure 5).

4.2.2. Node Storage Capacity. In the two cases of 100 and 500 jobs, the impact of node storage capacity on the performance of the educational resource grid system using different strategies is analysed. It can be seen from Figure 6(a) and Figure 6(b) that the average job execution time of the three strategies of LRU, LFU, and the economic model is much longer than that of ours. When the storage capacity of the node is small, the difference is more obvious. This fully reflects the superiority of our strategy to select neighbour nodes to create copies when the storage capacity of the node is small.

We can also observe that as the storage capacity of the node decreases, the average job execution time of the three strategies of LRU, LFU, and the economic model gradually approaches the state of no copy. This fully shows that these three strategies are not suitable for educational resource grids with small node storage capacity. In addition, by comparing Figure 6(a) and Figure 6(b), we can also find that even when the storage capacity of the node is small, the average job execution time of our strategy can be reduced with the increase of the number of jobs.

4.2.3. Access Mode. The storage capacity of the underlying node is set to 50 G and the number of jobs to 500. Under the two modes of decentralized access and centralized access, the performance of the system when different strategies are adopted is analysed. It can be seen from Figure 7 that when the same job is submitted, the execution time of the distributed access job is less than the execution time of the centralized access job. This is because, in centralized access, all jobs are concentrated on a few specific nodes, making the storage capacity of the underlying nodes insufficient, resulting in remote access to files and frequent replacement of copies (see Figure 6).

4.2.4. Number of Copies Created. The number of copies created refers to the total number of copies created during the execution of the job. Since the creation of a copy requires the IO and network bandwidth of the storage node, when the same job is executed, the fewer the number of times the copy is created, the better while the job execution time is guaranteed. In our statistics, the number of jobs is 100 and 500. The number of replica creations of different strategies varies with the storage capacity of the node (see Figure 7).

It can be seen from Figure 8 that when the same job is executed, as the storage capacity of the node increases, the number of copies created by the three strategies of LRU, LFU, and the economic model continues to increase, while our strategy can always keep the number of copies created. This fully shows that in this paper, we analyse the three strategies of LRU, LFU, and the economic model. When the storage capacity of the node is small, it is easy to cause frequent creation and deletion of replicas. This conclusion is correct. In addition, from the comparison between Figure 8(a) and Figure 8(b), it can be seen that the number of copies of our strategy does not increase with the increase in the number of jobs. This shows the superiority of choosing neighbour nodes to establish replicas when the storage capacity of the underlying nodes is insufficient (see Figure 8).

4.2.5. The Impact of Our Strategy on System Performance. According to the experimental results, three factors including the number of jobs, the storage capacity of nodes, and the user access mode are integrated to analyse the impact of different strategies on the performance of the educational resource grid system.

Taking into account the actual situation of the educational resource grid, a distributed user access mode is adopted. When the storage capacity of the underlying node is 100 G, 50 G, and 10 G, 500 jobs are submitted, and the

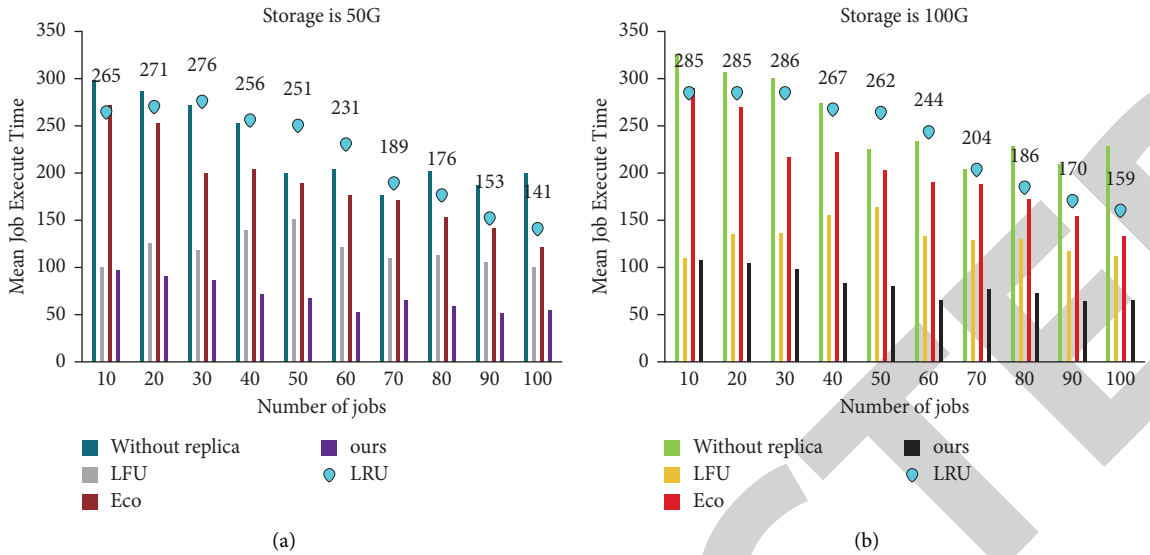


Figure 5: The relationship between the number of jobs and the average job execution time. (a) Storage is 50 G. (b) Storage is 100 G.

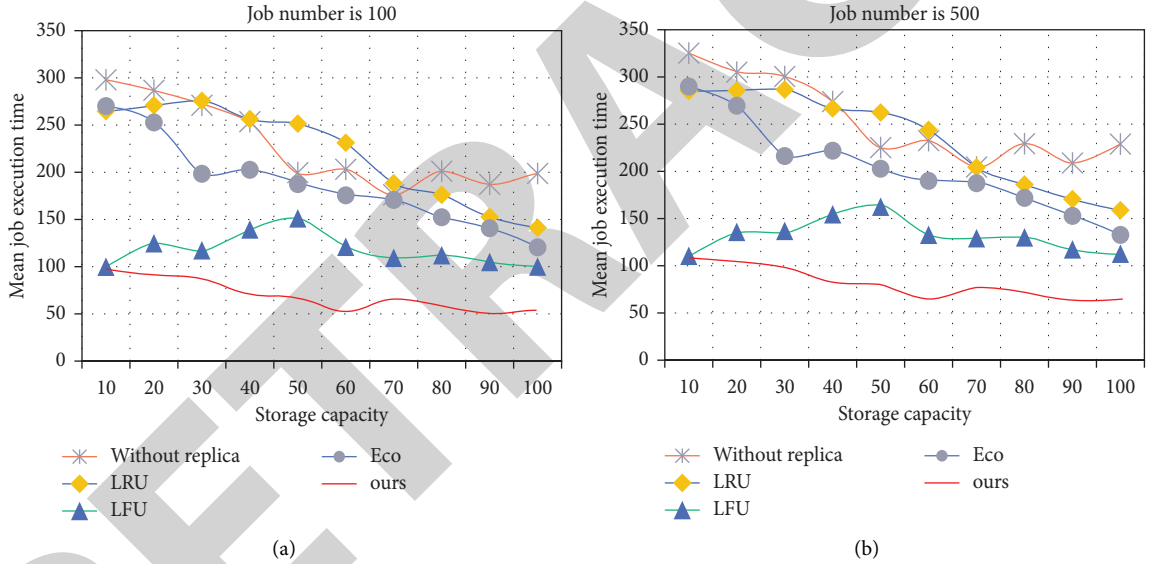


Figure 6: The relationship between storage capacity and average job execution time. (a) Job number is 100. (b) Job number is 500.

average job execution time and copy creation times are counted. The results are shown in Table 2.

When the node storage capacity is 100 G, that is, when the node capacity is large, there is little difference between different strategies. At this time, compared with LRU, the performance of our strategy is only improved by 0.5%. When the storage capacity of the node becomes 50 G, the average job execution time of LRU, LFU, and the economic model is not much different, but the number of copies of the economic model is significantly better than that of LRU and LFU. For our strategy, the average job execution time and the number of replica creation times are significantly better than the other three strategies. At this time, compared with the economic model, the two performance parameters of our strategy are increased by 78% and 89%, respectively. When the storage capacity of the node is further reduced to 10 G, the performance of the economic model begins to deteriorate, and our strategy improves the system performance more obviously. At this time, compared with the LFU strategy, the two performance parameters of our strategy are increased by 84% and 95%, respectively (see Table 2).

From the above analysis, it can be seen that, in various situations, the performance of the educational resource grid system using our strategy is better than that using the other three strategies. Especially, when the node storage capacity is small, our strategy can create fewer copies. Under the premise, a smaller average job execution time is obtained. It can be seen that it is appropriate to choose our strategy in the educational resource grid where the nodes are based on the general network environment.

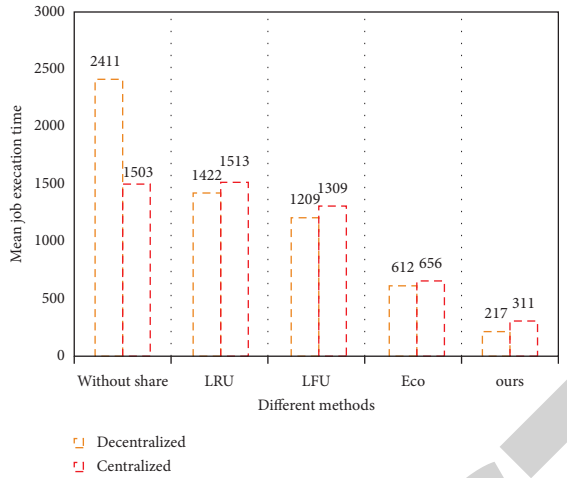


Figure 7: The relationship between access patterns and average job execution time.

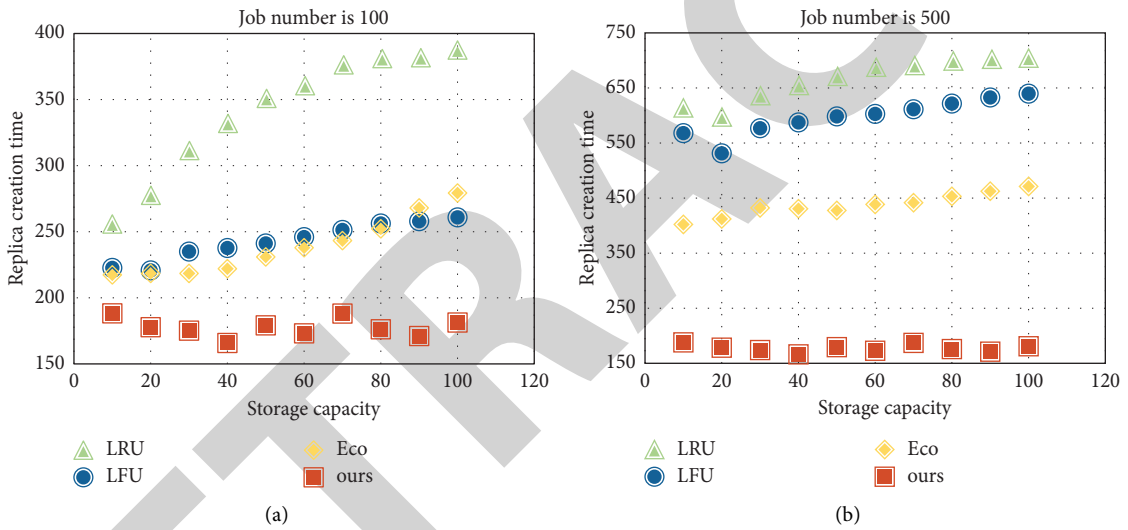


Figure 8: The relationship between storage capacity and the number of copies created. (a) Job number is 100. (b) Job number is 500.

Table 2: The impact of different strategies on grid system performance.

(G)		Without replica	LRU	LFU	Eco	Ours
10	Job execution time	2318.761	2276.512	2187.231	2213.876	521.876
	Replica creation time	0	3601	4672	2562	245
50	Job execution time	2211.876	1307.982	1588.657	1267.091	289.009
	Replica creation time	0	3287	3211	1678	365
100	Job execution time	1932.091	138.091	142.651	199.092	156.781
	Replica creation time	0	183	188	181	179

4.3. Performance of Dynamic Resource Scheduling Algorithm. This paper compares the algorithm in this paper with the user in the case of nonsharing of resources, and the user's smallest user accumulatively dominates the resource share and the resource utilization in the case of 20,100 users through experiments and executes 100 times to average to evaluate the performance of the algorithm in this paper.

4.3.1. Making the User's Accumulated Dominant Resource Share Meet the Lexicographical Optimal. It can be seen from Figure 9 that the goal of the algorithm in this paper is to make the user's accumulated dominant resource share meet the maximum and minimum. Therefore, compared with the situation where users do not share resources, the algorithm in this paper has the smallest cumulative resource share. As

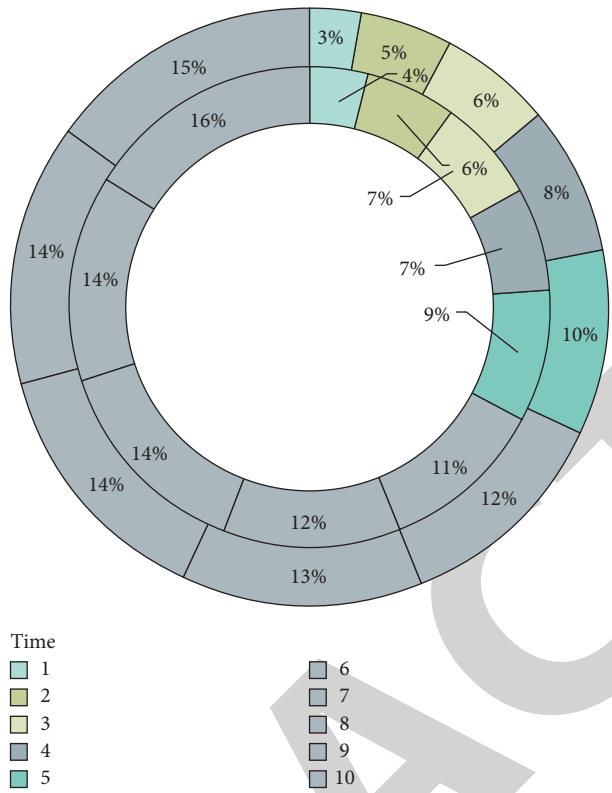


Figure 9: Cumulative share of resources.

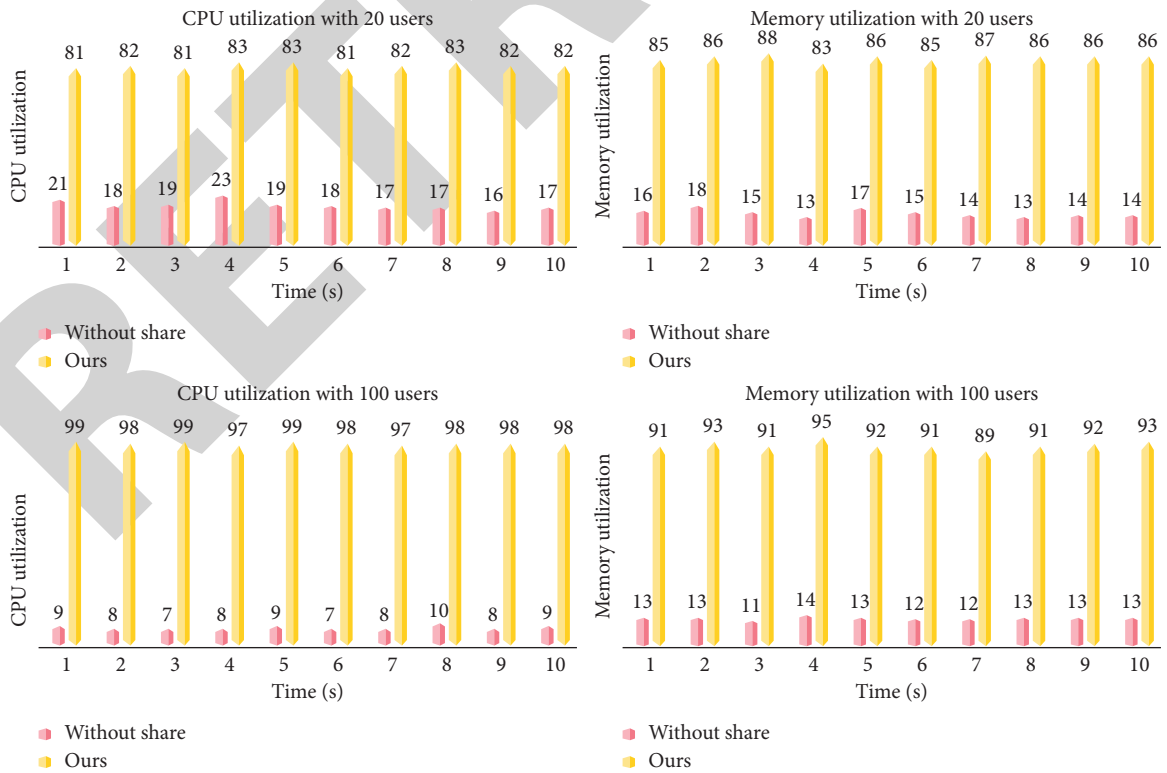


Figure 10: Utilization of CPU and Memory.

time goes by, the gap is getting bigger and bigger. This shows that the algorithm in this paper ensures that the user with the smallest accumulated dominant resource in the allocation process can allocate the most resources, that is, the user completes the largest number of tasks. Compared with the nonsharing algorithm, it ensures that the number of accumulated tasks completed by the user is as large as possible (see Figure 9).

4.3.2. High Resource Utilization. This paper compares the resource utilization rate under the algorithm of this paper with the resource rate of the system under the condition that users do not share resources. Figure 10 shows the resource utilization of the system when the number of users is 20 and the number of users is 100, respectively. It can be seen that the system resource utilization rate of the algorithm in this paper at any time is higher than the system resource utilization rate under the condition that users do not share resources.

In short, the algorithm in this paper satisfies fairness and enables all assigned users to complete more tasks while having high resource utilization (see Figure 10).

5. Conclusion

In the cloud computing environment, it is inevitable to realize the comprehensive sharing of network learning resources. It can not only make up for the shortcomings of traditional networks but also avoid the high cost and serious waste caused by the repeated construction of resources. On the other hand, it has promoted the development of education information, which has strong practical significance. Aiming at the problems existing in the sharing of learning resources in colleges and universities, this article first builds a smart education cloud service platform. Secondly, a hierarchical grid model of educational resource sharing is proposed. According to the characteristics of the educational resource grid, the key factors affecting the performance of the replica creation strategy are analysed, and a dynamic replica creation strategy is proposed. Considering the problem of multiuser resource allocation in the resource sharing system, a multiresource fair allocation algorithm based on the concept of resource sharing fairness is proposed. Experiments show that the grid sharing model proposed in this paper has better performance on the education cloud service platform. At the same time, when resource sharing users put forward multiple sets of time-varying resource requirements, the resource allocation algorithm proposed in this paper can ensure that users meet the optimal lexicographical order of accumulated dominant resource shares and have higher resource utilization.

Due to the need to judge user sharing defects and cumulative resource allocation in the process of algorithm execution, the algorithm does not have an advantage in allocating execution speed when there are a large number of users and virtual machines. Therefore, how to effectively optimize the allocation of tasks and virtual machines in order to improve the allocation speed will be the focus of the next research.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] L. Rajendran and R. Veilumuthu, "A cost effective cloud service for E-learning video on demand," *European Journal of Scientific Research*, vol. 55, no. 4, pp. 569–579, 2011.

[2] S. Gupta and B. B. Gupta, "XSS-secure as a service for the platforms of online social network-based multimedia web applications in cloud," *Multimedia Tools and Applications*, vol. 77, no. 4, pp. 4829–4861, 2018.

[3] J. H. Bing, "University Ideological and Political Multimedia Network Teaching Based on MOOC," in *Proceedings of the International Conference on E-Learning, E-Education, and Online Training*, pp. 167–180, Xinxiang, China, June 2021.

[4] S. Mayoof, H. Alaswad, S. Aljeshi, A. Tarafa, and W. Elmedany, "A hybrid circuits-cloud: development of a low-cost secure cloud-based collaborative platform for A/D circuits in virtual hardware E-lab," *Ain Shams Engineering Journal*, vol. 12, no. 2, pp. 1197–1209, 2021.

[5] J. J. Youn, S. J. Jeong, K. H. Lee, and K. Byung Man, "Development and Application of a Digital Curation System to Promote Total Creative Personality Based on Multimedia," *Multimedia Tools and Applications*, vol. 80, pp. 34369–34387, 2021.

[6] A. McDonald, H. McGowan, M. Dollinger, R. Naylor, and H. Khosravi, "Repositioning students as co-creators of curriculum for online learning resources," *Australasian Journal of Educational Technology*, vol. 37, no. 6, pp. 102–118, 2021.

[7] S. Namasudra, "Data access control in the cloud computing environment for bioinformatics," *International Journal of Applied Research in Bioinformatics*, vol. 11, no. 1, pp. 40–50, 2021.

[8] S. Khan, A. Al-Dmour, V. Bali, M. R. Rabbani, and K. Thirunavukkarasu, "Cloud computing based futuristic educational model for virtual learning," *Journal of Statistics & Management Systems*, vol. 24, no. 2, pp. 357–385, 2021.

[9] S. Yao, D. Li, A. Yohannes, and H. Song, "Exploration for network distance teaching and resource sharing system for higher education in epidemic situation of COVID-19," *Procedia Computer Science*, vol. 183, pp. 807–813, 2021.

[10] Z. Jiang, S. Yuan, J. Ma, and Q. Wang, "The evolution of production scheduling from Industry 3.0 through Industry 4.0," *International Journal of Production Research*, pp. 1–21, 2021.

[11] J. Zhou and W. Zou, "Construction and application of wenhua education cloud based on huawei Fusion Sphere," in *Proceedings of the 2021 IEEE international conference on artificial intelligence and computer applications (ICAICA)*, pp. 477–481, Dalian, China, June 2021.

[12] R. G. Barreto, L. Aversari, C. N. A. P. Gomes, and N. C. Q. Lino, "H-KaaS: a Knowledge-as-a-Service architecture for E-health," *Brazilian Journal of Biological Sciences*, vol. 5, no. 9, pp. 3–12, 2018.

[13] J. Martinez-Guijosa, A. López-Alonso, C. Gortazar, A. Pelayo, J. T. María, and V. Joaquín, "Shared use of mineral supplement in extensive farming and its potential for infection