Informatica

An International Journal of Computing and Informatics

Special Issue:

Recent Trends and Advances of Informatics in E-Commerce: Opportunities, Challenges and Solutions

Guest Editors:

Ruihang Huang, Amit Sharma, Ashutosh Sharma

1977

Editorial Boards

Informatica is a journal primarily covering intelligent systems in the European computer science, informatics and cognitive community; scientific and educational as well as technical, commercial and industrial. Its basic aim is to enhance communications between different European structures on the basis of equal rights and international refereeing. It publishes scientific papers accepted by at least two referees outside the author's country. In addition, it contains information about conferences, opinions, critical examinations of existing publications and news. Finally, major practical achievements and innovations in the computer and information industry are presented through commercial publications as well as through independentevaluations.

Editing and refereeing are distributed. Each editor from the Editorial Board can conduct the refereeing process by appointing two new referees or referees from the Board of Referees or Editorial Board. Referees should not be from the author's country. If new referees are appointed, their names will appear in the list of referees. Each paper bears the name of the editor who appointed the referees. Each editor can propose new members for the Editorial Board or referees. Editors and referees inactive for a longer period can be automatically replaced. Changes in the Editorial Board are confirmed by the Executive Editors.

The coordination necessary is made through the Executive Editors who examine the reviews, sort the accepted articles and maintain appropriate international distribution. The Executive Board is appointed by the Society Informatika. Informatica is partially supported by the Slovenian Ministry of Higher Education, Science and Technology.

Each author is guaranteed to receive the reviews of his article. When accepted, publication in Informatica is guaranteed in less than one year after the Executive Editors receive the corrected version of the article.

IJCAI-ECAI 2022: Can Europe Revive its Position in AI after Lagging Behind the US and China?
Subtitle: AI is dead, long live AI!

Editorial by Matjaž Gams

As the subtitle suggests, the old AI is dead, and a new AI is ascending the throne. Can IJCAI [1] provide us with answers about the new AI?

The joint IJCAI-ECAI 2022 conference with workshops was held at the Messe Center, with 55,000 m^2 and a capacity for 25,999 visitors in Vienna, Austria (see Figure 1), from the 23rd to the 29th of July. It was the 25th European Conference on Artificial Intelligence and the 31st International Joint Conference on Artificial Intelligence, making it the longest-running major conference series spanning all areas of artificial intelligence. It was the first in-person conference after the unfortunate COVID-19 period. This fact alone was enough to make it an exciting event, without even considering the advances highlighted in this editorial. The second central theme was the relative progress in AI made in China, the US and the EU.

Figure 1: In 2020, IJCAI-ECAI was held in Vienna, often described as the world's most livable city.

In recent years people have detected an ominous lag in European AI. For example, in 2021 the European Investment Bank [2] published a report on *Artificial intelligence, blockchain and the future of Europe* with the subtitle *"How disruptive technologies create opportunities for a green and digital economy"*. At that time Europe still had an upper hand in some categories, e.g., there were 43,064 AI researchers in Europe, plus 7998 in the UK, 28,539 in the US and 18,232 in China. However, while AI and blockchain technologies accounted for €25 billion in annual investments, 80% of that amount was covered by the US and China (€20 billion), and only €1.75 billion, or

7% of the investment, was from the 27 EU Member States. The report advises the EU to invest nearly €10 billion in blockchain and AI, to match the progress in AI in the other two superblocks. Similarly, scientific progress by the Chinese at IJCAI was observed [3].

These three blocks are well aware that AI is not only one of the most progressive scientific disciplines, it is also boosting the digital transformation across industries and societies at a global level. While the blocks are similarly concentrated on AI, their progress is very different. The US was – and still is – the leader in AI technologies; China has begun to catch up after a long period of delay; and the EU is a story on its own. In this period China has overtaken the US as the largest economy in terms of real GDP, i.e., PPP. Based on several metrics, the EU is currently positioned third, but with a clear potential to deliver on AI and catch up. The progress of the three blocks is also closely related to Brexit and the war in Ukraine, which has delivered a huge economic blow to progress in EU and a more modest one to the US.

Back in 2020 the EU recognized the importance of AI in Europe [4] and devoted reasonable funds to it. It also tried to forge its own way: towards trustworthy and human-centered AI. At IJCAI 2022 some researchers even claimed that the usual metrics are no longer relevant to the EU's AI since it is now differently oriented. On the other hand, some people are of the opinion that the EU diverted from the path of conventional research in the direction of social-sciences-oriented AI, which may on its own represent an additional obstacle to AI progress in traditional and technological ways. In [4], the overview concluded with "Europe needs to find a way to protect its research base, encourage governments to be early adopters, foster its startup ecosystem, expand international links, and develop AI technologies as well as leverage their use efficiently." Whatever the case, several reports about AI, similar to [4], conclude that "Disruptive technologies create opportunities for a green and digital economy".

Looking at search engines, we get an impression of the general relations. In this field, Google from the US and Baidu from China are not matched by an EU search engine. These companies not only use AI in every search, they also provide an intense top-class AI research. A decade and half ago the EU's approach to search engines resulted in a novel, distributed search-

engine concept based on genres [5], but as is typical with EU projects, after the research-project phase ended there were no funds to implement it in real life. In contrast, the Chinese (albeit with some issues related to democracy) promoted its own search engine, Baidu. By November 2013, Google's search market share in China had declined to 1.7% from its August 2009 level of 36.2%. Had the EU governments decided like the Chinese to actually implement ALVIS as its own search engine, it would be competing at the global level. Alternatively, the EU could buy a competing global search engine and adapt it to EU standards and needs. Unfortunately, and unlike the global fast-train initiative accepted recently, there is no EU initiative to setup a European search engine containing major AI elements.

In a report published in 2022 by the Joint Research Centre "AI Watch Index 2021" [6] the overall conclusion is that the US is the leading country in several categories, while Europe is in third place. For example, in terms of AI organizations (companies and institutions), the US has 14,000, China 11,000, and the EU 6,000. The report also observes an important reduction in AI activities in the EU due to Brexit. But while Europe is third, the gap is smaller than is often suggested. The European Commission is set to invest additional €1 billion per year in AI and bring overall EU spending up to €20 billion annually.

The report also contradicts the claims [2,3] that China is emerging as a world leader in AI. While China has experienced an explosion in the filing of patents, its innovative potential is kind of modest. Similarly, while in 2019 China accounted for 22.4% of the world's peer-reviewed AI publications, more than the EU (16.4%) and the US (14.6%), according to the Artificial Intelligence Index Report 2021 by Stanford University, and China overtook the US for the first time in AI journal citations, the major achievements still seem to remain related to the US. For example, 56% of China's top AI talents are employed in the United States. Nine out of ten Chinese students who studied in the field of AI in the US stayed on after graduation.

Back to IJCAI-ECAI. Would you expect one of the three top scientific journals to publish a paper about one of the year's AI achievements? It happened in the journal *Nature* in 2022 [7, 8], see Figure 2. Naturally, this achievement was presented and discussed at the conference during several events and subtasks, e.g., best lap, best overtaking, and similar. The catch is that the AI algorithm/method outcompeted human champions in the Gran Turismo racing game. In simple words – a program was driving better than the best human drivers. Another task where AI programs outperformed the best humans, but just consider how

much this task is different from the previous ones solved by AI.

It is more or less common knowledge that AI outperforms humans at chess and formal games and tasks. At IJCAI-ECAI 2022, the world chess competition was going on with Ginko coming out the winner (see Figure 3). However, the main attraction was the car racing. Consider again the major difference between the two tasks, i.e., chess and driving a car, the latter dealing with sliding, breaking, and overtaking on the limit. Would you expect it a year ago? Where is the limit for AI?

Figure 2: AI outcompeted humans in a car racing game.

Figure 3: The computer chess championship was held at the conference, resulting in several astonishing games.

There were lots of "normal" papers dealing with regular issues. The research described in an IJCAI paper [9], and also in Figure 4, first fed the agents, e.g., with anti-vaccination videos and observed how they became anti-vaccines oriented. However, after watching debunking videos, on average the agents turned somehow "normal", but to different degrees in the five areas analyzed: 9/11, chemtrails, anti-

vaccination, flat Earth, Moon landing. The agents did not have human cognitive properties, they mainly performed an extraction from the input into their "beliefs". It is fascinating that agents as well as humans seem to have a low-level of free will and a resistance to information tampering. The effect of commercials, web advertising, recommendation systems, conventional media and the information overflow seems to increasingly change humans into "mental zombies". The expectation of the web's visionaries that the vast amount of information at hand on the Web and the possibility to cross-check anything will create humans who are more knowledgeable and cautious is, on the whole, failing. Increasingly, people are becoming trapped in their information bubbles, leading to dispute and hate between different political and ideological groups.
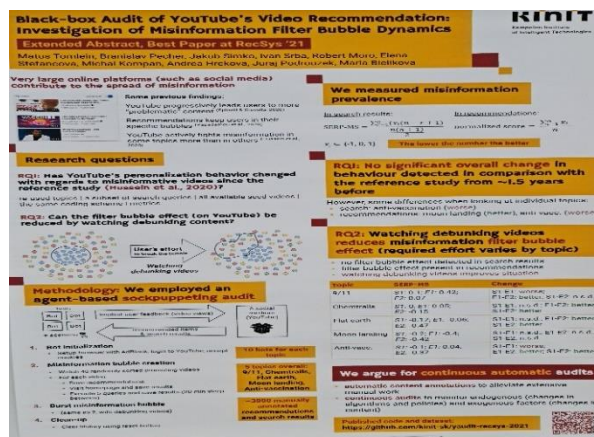

Figure 4: Agents demonstrate the power of YouTube information bubbles.

That paper also indicated how to deal with disinforming videos and other information sources: present quality debunking information that leaves no question. Years ago, scientists proposed Wikipedia as the main resource for human knowledge and truths, but unfortunately, even that top-quality source of knowledge in the form of an encyclopedia is becoming biased by radical ideologies. The main reason is that knowledge sources like Wikipedia or Quora started dealing with political issues, e.g., whether some action by President Donald Trump was legal or not. Such information has no place in quality scientific sources. Therefore, the current advice is to trust only factual data, and hold reservations and double check when dealing with political, ideological and subjective issues.

There was also a tutorial on opinion formation in social networks. Several models were explained, e.g., of De Groot, Friedkin-Jensen and similar. They enable a formal analysis of behavior, although it seems that some semantics is lacking to explain actual behavior.

But they enable a formalization, which is an important improvement in itself.

Another interesting area was automated story generation. From generation to generation, programs have improved their performance. There are several programs like GPT3, OPL, Lambda, Comet, etc. of which the public is probably aware of a couple. On average, they are not as good as humans, but the difference is shrinking fast.

It is worth pointing out that xGBoost and deep neural networks, which are now referred to as "neural networks" (since now they are all deep), compete for the best results in various domains. In one way, both methods are different, one relying on trees and the other on layers of neurons, but in another way they both exploit multiple/redundant knowledge, which is the source of their success.

Among the increasingly popular areas is federated learning, because it efficiently resolves anonymity problems. Among explanations, counterfactual reasoning provides the best ones – if only somebody could explain that to the bureaucrats.

The panel on career development concentrated on the differences between academia and industry. All over the world, salaries are larger in industry and risks higher, but academia is more open to new ideas.

Climate, oceans and environment deserved a special workshop at IJCAI.

Among the invited presentations, Gerhard Widmer, as usual, extracted the most passion from the audience, this time by introducing feelings into classical music. Luc Steels reminded us that AI is currently by far the most exciting field, and the one that will raise our society to the next level. Tim Miller analyzed explainable AI and showed that AI publications are slowly but surely moving from purely algorithmic/technical into the social and cognitive subfields. Pete Wurman explained how they won the world competition in the Gran Turismo racing game (see Figure 2). Jerome Lang presented an observation and vision of how AI is moving toward incorporating some social sciences using agent studies. Markus Hecher was the recipient of the EurAI dissertation award for an improvement in ASP by changing graph problems into trees. Sumit Gulvani from Microsoft Research explained his module for learning in Excel that is based on learning from a couple, one or even zero examples. Judea Pearl is no doubt one of the most famous scientists in probability and AI since he invented Bayesian networks. The key is in the causal inference. Unfortunately, time was too short to catch all his ideas. Michaela van der Schaar dealt with medical problems and emphasized the role

of time and explanation. SimpleEx is supposed to explain any black box in the form of an equation. Ana Pavia presented the engineering society and collaboration in AI systems. Bo Li introduced trustworthy ML. Michael Littman analyzed the decrease of complexity due to novel approaches. Stuart Russel presented an overview of AI development and potential future directions, and relations between AI and humans.

In summary, to attend IJCAI is to harvest the world's AI knowledge and to exchange ideas about future work. As such, IJCAI remains the premier AI conference in the world.

P.S. To demonstrate that we can and should do better in relation to the environment, a billboard promoting a grass field for insects in the center of Vienna is presented in Figure 5.

Figure 5: Vienna demonstrates that there is room for plants and insects in cities, symbolizing a new approach to the environment.

References

[1] IJCAI Proceedings, https://www.ijcai.org/past_proceedings

[2] Artificial intelligence, blockchain and the future of Europe, European Investment Bank, https://www.eib.org/en/publications/artificial-intelligence-blockchain-and-the-future-of-europe-report

[3] Gams, M., IJCAI 2018 - Chinese dominance established: editorial. *Informatica: an international journal of computing and informatics*, ISSN 0350-5596, 2018, vol. 42, no. 3, pp. 285-289.

[4] Brattberg, E., Csernatoni, R., Rugova, V., Europe and AI: Leading, Lagging Behind, or Carving Its Own Way? https://carnegieendowment.org/2020/07/09/europe-and-ai-leading-lagging-behind-or-carving-its-own-way-pub-82236

[5] Vedulin, V., Luštrek, M., Gams, M., Training a Genre Classifier for Automatic Classification of Web Pages. *CIT: journal of computing and information technology*, ISSN 1330-1136, 2007, vol. 15, no. 4, pp. 305-311.

[6] AI Watch Index, EU Report 2021, https://ai-watch.ec.europa.eu/publications/ai-watch-index-2021_en

[7] Wurman, P. R. et al. *Nature* **602**, 223–228 (2022).

[8] Gerdes J.C., Neural networks overtake humans in Gran Turismo racing game, February 2022, Nature News, https://www.nature.com/articles/d41586-022-00304-2

[9] Tomlein, M., Pecher, B., Simko, J., Srba, I., Moro, R., Stefancova, E., Kompan, M., Hrckova, A., Podrouzek, J., Bielikova, M., Black-box Audit of YouTube's Video Recommendation: Investigation of Misinformation Filter Bubble Dynamics, IJCAI Proceedings, 2022, https://www.ijcai.org/proceedings/2022/0749.pdf

Guest Editorial Preface

Recent Trends and Advances of Informatics in E-Commerce: Opportunities, Challenges and Solutions

The objective of this special issue is to concentrate on all aspects and future research directions related to this specific area of E-commerce toward online shopping, online food services, E-healthcare, E-care, E-solution, service oriented modeling, reliable and secure systems design and analysis. We have received more than 50 manuscripts in total for this special issue across the globe and after the rigorous review process, only 13 manuscripts have been accepted for publication. A short review about the commitments for this Special Issue is as underneath:

***Zhan Guo** et al.* contribute an article entitled *"Design and Study of Urban Rail Transit Security System Based on Face Recognition Technology"*. This paper studies an urban rail transit security system based on face recognition. The analysis of the main mode of face recognition is carried out utilizing the practical application design ideas. ***Jun Ding** et al.* contributes an article entitled *"Big Data Intelligent Collection and Network Failure Analysis Based on Artificial Intelligence"*. This paper presents intelligent data collection and network error analysis based on artificial to study smart data collection and network error analysis. ***Danna Su** et al.* contribute an article entitled *"Construction of lean control system of prefabricated mechanical building cost based on Hall multi-dimensional structure model"*. This paper studies the prefabricated mechanical building cost lean control system. The results shows that the original design components and the number of open models is 72, the optimized types of components and the number of open models is 51, which reduce 21 models machining. This results reduction in the models cost up to 25%. ***Yongqing Tian** et al.* contribute an article entitled *"Improved artificial electric field algorithm based on multi-strategy and its application"*. This article unveils that artificial electric field algorithm is a new swarm bionic optimization algorithm. In this paper, an artificial electric field algorithm based on opposition learning is proposed to improve the global exploration ability and local development ability of artificial electric field algorithms. The comparative results show that the IAEFA-SVM model has high prediction accuracy and provides an effective method for sand liquefaction identification when compared with the traditional methods. ***Haiyan Fan** et al.* contribute an article entitled *"Computer-aided architectural design optimization based on BIM Technology"*. This paper explores the architectural design process based on the BIM platform and puts forward the structural design method based on the BIM platform. The results obtained for experimentation show that the period ratio,

displacement ratio, and the first six modes calculated by the two methods in the modal analysis are consistent. ***Xiaoming Liu** et al.* contribute an article entitled *"Chaotic association feature extraction of big data clustering based on Internet of Things"*. This article addresses the stabilization of chaotic characteristics in abnormal data by proposing chaotic correlation feature extraction of big data clustering based on the Internet of things. The results show that when dealing with the same amount of data, the energy consumption of the proposed algorithm is significantly lower than that of the traditional algorithm. ***Hongwei Liang** et al.* contribute an article entitled *"Application and study of artificial intelligence in railway signal interlocking fault"*. This paper utilizes the deep learning algorithm of artificial intelligence for investigating the interlocking faults in the railway transportation. It is demonstrated that deep learning integration is an effective method to improve the classification performance of turnout fault diagnosis model. ***Ying Zhang** et al.* contribute an article entitled *"Design and Implementation of a New Intelligent Warehouse Management System Based on MySQL Database Technology"*. This article makes an overall design of the warehouse management system, builds a MySQL database, and realizes the design and application of a new intelligent warehouse management system. ***Rong Wang** et al.* contribute an article entitled *"Automatic classification of document resources based on naive Bayesian classification algorithm"*. This paper introduces the relevant theories of naive Bayes classification and the automatic document classification system. Experiments show that the naive Bayesian classification algorithm can effectively complete the automatic capture, processing and classification of massive academic documents, which can not only improve the classification accuracy, but also reduce the running time of automatic classification. ***Zheng Zheng** et al.* contribute an article entitled "Intelligent analysis and processing technology of big data based on clustering algorithm*"*. In this paper, an attribute category clustering method has been proposed to study the big data intelligent analysis and processing technology. The experimental results show that proposed the proposed method can effectively merge attributes, reduce the dimension after binary transformation and effectively reduce the amount of data under the condition of ensuring data information. ***Yujiao Liu** et al.* contribute an article entitled "The application of Internet of Things and Oracle database in the research of intelligent data management system*"*. This paper demonstrates an intelligent data management consisting resource allocation mechanism to provide timely and effective decision

for the resource allocation. The comparison results show that the same bitmap index only occupies about 1/30 of the original table, and the data size is reduced by more than 10 times. *Jing Feng et al.* contribute an article entitled "Intelligent engineering management of prefabricated building based on BIM Technology". This paper solves the problem of China's construction industry adopted by the traditional extensive construction mode for a long time. This paper puts forward a new mode of fine construction management based on BIM. It is demonstrated that BIM Technology has brought good economic and social effects to aid fine management. *Boyang Li et al.* contributes an article entitled "Application of interactive Genetic Algorithm in landscape planning and design". This article aims at improving the design effect of garden landscape space environment and optimizes the structure of garden landscape space environment. The proposed method achieves better

optimization of landscape spatial environment structure, and achieves good landscape spatial environment design effect.

I hope that the quality research work published in this special issue will be able to serve the concerned science, environment, and technology.

Guest Editors

Ashutosh Sharma
(sharmaashutosh1326@gmail.com), University of Petroleum and Energy Studies, Dehradun, India

Amit Sharma (amit.amitsharma90@gmail.com), Institute of Computer Technology and Information Security, Southern Federal University, Russia.

Ruihang Huang (1209125@mail.dhu.edu.cn), Donghua University, Shanghai, China

Improved Artificial Electric Field Algorithm Based on Multi-Strategy and its Application

Yongqing Tian[1], Libo Liu[1]*, Xiaolei Wang[1], Lin Dong[1], Rana Gill[2], Ravi Tomar[3]

[1]College of Civil Engineering, Hebei University of Engineering, Handan, Hebei, 056038, China
[2]Department of AIT-CSE, Chandigarh University, Mohali, Punjab-140413, India
[3]Persistent Systems, India

E-mail: yongqingtian2@126.com, liboliu7@163.com, xiaoleiwang8@126.com, lindong811@163.com, rana.cse@cumail.in, ravitomar7@gmail.com

Artificial electric field algorithm is a new swarm bionic optimization algorithm, which uses the interaction force of charged particles to create a mathematical model to solve the problem. To improve the global exploration ability and local development ability of artificial electric field algorithms, an artificial electric field algorithm based on opposition learning is proposed. The chaos strategy is used to strengthen the quality of the initial population, and the opposition learning strategy is used to increase the diversity of the population and the development ability of the algorithm. The excellent performance of the algorithm is proved by simulation experiments. The improved artificial electric field algorithm is combined with SVM to construct the sand liquefaction identification model by selecting seven measured indexes, including intensity, underground water level, overlying effective pressure, standard penetration hit number, average particle size, non-uniformity coefficient, and shear stress ratio. Compared with traditional methods such as the standard method and seed simplification method, the results show that the IAEFA-SVM model has high prediction accuracy and provides an effective method for sand liquefaction identification.

Povzetek: Predstavljen je izboljšan algoritem umetnega električnega polja na osnovi mnogoterih strategij.

1 Introduction

The artificial electric field algorithm (AEFA) is a new intelligent optimization algorithm proposed by Indian scholar Anita in 2019 [1]. Anita's intelligent optimization algorithm, which is inspired by Coulomb's law of static electricity, has the characteristics of fewer parameters, lower computational complexity, better scalability, exploitability, and many others. However, it is easy to get into the local optimum and lacks exploration.

To improve the performance of AEFA, Aysen [2] integrated the opposition-based learning strategy into the initialization and updating process of AEFA and proposed the oppositional learning-based AFEA (OBAEFA), which improved the exploring ability of AEFA. Anita [3-4] and others extend the AEFA algorithm for constrained optimization by introducing new velocity and location constraints. The existence of boundary allows particles to interact within the scope of the problem, and to learn from each other in the problem space. The introduction of the strategy makes a better balance effect on the exploration and development of the algorithm. In the following study, Anita extends the artificial electric field algorithm with combinatorial

higher-order graph matching problems and introduces the discrete artificial electric field algorithm. The framework combines redefinition of location, speed representation, use of addition and subtraction, updating rules for speed and location, and initialization of specific problems with heuristic information [5, 6]. The algorithm is proved to be superior to other existing algorithms in matching degree and accuracy [7].

To improve the exploratory ability of AEFA and solve the problem of easily falling into local optimal solution, the AFEA is improved in the following aspects:

i. The chaotic technique is introduced into the AEFA, and the initial population is generated in the search space by the randomness and universality of the chaotic motion, and the probability of finding the optimal solution is increased.

ii. The diversity of the population is maintained and the possibility of jumping out of the local optimum is improved by the opposite learning strategy.

iii. The greedy strategy is used to get the optimal value of the population quickly. Then, through the simulation of 9 test functions, the IAEFA algorithm is

compared with other improved algorithms to verify its effectiveness of the IAEFA algorithm.

Finally, the improved artificial electric field algorithm, in combination with the support vector machine (SVM) is applied for the identification of sand liquefaction and the results are compared with the traditional method of identification of sand liquefaction. This project is not limited to industrial applications but the overall growth of social life with the integration of the Internet of Things, AI, and robotics [8-11].

The rest of this article is organized as: Section 2 presents the principles of the algorithm. Section 3 consists of the information about artificial electric field algorithms based on chaotic learning and opposition-based learning strategy. The results and analysis part is covered in section 4. Section 5 describes several common assessment methods of sand liquefaction. At last, the concluding remarks are presented in Section 6.

2 Principles of the algorithms

2.1 Artificial Electric Field Algorithm (AEFA)

AEFA is inspired by Coulomb's Law of electrostatic force, which states that the force that occurs between charged particles and charged particle is proportional to the product of their charges. The force is also inversely proportional to the square of the distance between the charges, each individual in the population is considered to be a charged particle, their strength is measured by their charge, and the position of the charge corresponds to the solution to the problem, the charge is defined as the fitness value of the candidate solution and the fitness function of the population. In the AEFA algorithm, only the electrostatic gravitation is considered, so that the charged particle with the largest charge ("The best individual") attracts other lower charged particles and moves slowly in the search space. The AEFA shown in Figure 1 can be considered as an isolated system of charges, and the position of the optimal fitness value for any electron i at any time t is given by Equation 1.

$$p_i^d(t+1) = \begin{cases} p_i^d(t), f(p_i(t)) > f(x_i(t+1)) \\ x_i^d(t+1), f(p_i(t)) \leq f(x_i(t+1)) \end{cases} \quad (1)$$

The total number of charged particles are denoted by N and the total number of parameters by d. The position of the particle with the best fitness is represented by $p_{best} = x_{best}$ and the force exerted on the particle i at time t by the particle j as shown in Equation 2.

$$F_{ij}^d = k(t) \frac{Q_i(t) \cdot Q_j(t) \cdot (P_j^d(t) - X_i^d(t))}{R_{ij}(t) + \varepsilon} \quad (2)$$

$Q_i(t)$ and $Q_j(t)$ are the charges of the i particle and j particle at arbitrary time t. $k(t)$ is the Coulomb constant of the arbitrary time t. ε is a relatively small random number. $R_{ij}(t)$ is the Euclidean distance between the two particles, represented by the Equation 3.

$$R_{ij}(t) = \|x_i(t), x_d(t)\|_2 \quad (3)$$

$k(t)$ is the number of iterations and the maximum number of iterations, given by the following Equation 4.

$$k(t) = k_0 \cdot e^{\left(\frac{-\alpha \cdot (iter)}{maxiter}\right)} \quad (4)$$

α is the parameter and k_0 is the initial value, $iter$ is the current iteration, and maxiter is the maximum number of iterations. At the beginning of the algorithm, use constant k_0 in a large initial value can make a better exploration. Then it is reduced by iteration to control the search accuracy. The total electric force of the other particles at any time t on particle i is expressed in Equation 5.

$$F_i^d(t) = \sum_{j=1, j \neq i}^{N} rand \cdot F_{ij}^d(t) \quad (5)$$

F_i^d denotes the resultant force acting on the charged particle i in d dimensional at time t. And rand refers to the uniformly generated random number in the range of [0,1]. Random numbers can provide randomness. The electric field of the charged particle i in d dimension at time t is given in Equation 6.

$$E_i^d(t) = \frac{F_i^d(t)}{Q_i(t)} \quad (6)$$

By using Equation 6 and Newton's law, it can be deduced that the particle i has an acceleration at time t in d dimension and expressed in Equation 7.

$$a_i^d(t) = \frac{Q_i(t) \cdot E_i^d(t)}{M_i(t)} \quad (7)$$

$M_i(t)$ denotes the unit mass of a particle i at time t t, the velocity v and position x of the particle are represented by the following Equation 8 and Equation 9 respectively.

$$V_i^d(t+1) = rand \cdot V_i^d(t) + a_i^d(t) \quad (8)$$

$$X_i^d(t+1) = X_i^d(t) + V_i^d(t+1) \quad (9)$$

$$Q_i(t) = Q_j(t) = \frac{q_i(t)}{\sum_{i=1}^{N} q_i(t)} i, j = 1, 2, \cdots N \quad (10)$$

Rand denotes the uniformly generated random numbers in the range of [0,1]. The charge of the particle is calculated according to Equation 10 and it is supposed that each particle has an equal charge.

In Equation 10, $q_i(t)$ denotes the max normalized value ($Q_{best=1}$) of the best particle of the selected suitable charge function, calculated as Equation 11.

$$q_i(t) = e^{\left(\frac{fit_i(t)-worst(t)}{best(t)-worst(t)}\right)} \quad (11)$$

$fit_i(t)$ is the fitness value of particle i at time t. $best(t)$ is the fitness value of the best particle. $worst(t)$ is the fitness value of the worst particle. The minimization problem is defined as the following Equation 12.

$$best(t) = \min\left(fit_i(t)\right), i \in (1, 2, \cdots, N)$$

$$worst(t) = \max\left(fit_i(t)\right), i \in (1, 2, \cdots, N) \quad (12)$$

The flowchart of the AEFA algorithm is shown in Figure 1. From the flowchart, you can see that the algorithm starts with randomly initializing the particles. Then, for each iteration, the fitness of each particle is evaluated, and the fitness values for the best and worst particles are calculated. In the next iteration, the velocity and position of each particle are updated. This process is repeated until the maximum number of iterations is reached to obtain the optimal solution.
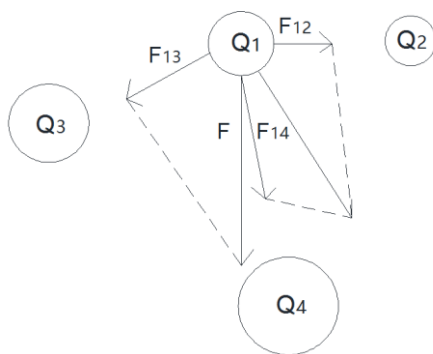

Figure 1: The interaction of particle

2.2 Basic ideas of opposition-based learning strategy

The opposition-based learning strategy was proposed by scholar Tizhoosh [12] in 2005. Compared with other algorithms, it takes time to get the efficiency of the new solution. Genetic algorithms, for example, require several generations or more of algebra to introduce new directions through genetic variation. In

recent years opposition-based learning OBL has been effectively applied to various swarm intelligence algorithms. When solving problems, it is considered that there may be a better solution on the opposite side of an ineffective solution. The quality of a population can be improved by introducing opposite solutions rather than two independent random solutions.

If there is a number X on [l, u], then the antithesis of X is defined as $\bar{x} = l + u - x$. Extending the definition of the opposite point to the n-dimensional space, supposing p as a point in the n-dimensional space, where $x_i \epsilon [l, u], i = 1, 2, \ldots, n$, the opposite point is $p' = (x'_1, x'_2, \cdots, x'_n)$. Among them, $x'_i = l_i + u_i - x_i$.

Suppose x as a random number on [l, u], $\tilde{x}$ as its reverse solution, $f(x)$ the objective function, $g(\cdot)$ the proper evaluation function. Calculating $f(x)$ and $f(\tilde{x})$ in each iteration, if $g(f(x))$ greater than $g(f(\tilde{x}))$, then retains the value of x and vice versa.

3 Artificial electric field algorithm based on chaotic learning and opposition-based learning strategy

This section includes the discussion of artificial electric field algorithms for chaotic learning and opposition-based learning strategy.

3.1 Basic ideas of the algorithm

For complex optimization problems, especially for multi-modal functions in high latitude, the basic artificial electric field algorithm is easy to get into the local optimal solution, and the ability of global exploration is insufficient. Based on chaos and oppositional learning, a hybrid artificial electric field algorithm (IAEFA) is proposed. In the basic artificial electric field algorithm, the population initialization of the chaotic map sequence and the opposition-based learning strategy is introduced. Below are three areas for improvement.

3.2 The main process of the IAEFA algorithm

3.2.1 Initialization of Chaos method

The process of the initialization of the standard artificial electric field algorithm takes random allocation and can not distribute the population uniformly in the solution domain. Especially when optimizing the multi-peak function of high latitude, the diversity of the population is reduced, causing precocious puberty. At present, the research shows that the variables generated by the logistics chaotic map [13] have strong universality, which can improve the shortage of initial population diversity generated by random allocation.

$$Z_{n+1} = \mu(1 - Z_n) \quad (13)$$

In Equation 13, μ for random numbers between [0, 4]; Z_n for the nth chaotic variable, the value range for [0, 1].

3.2.2 Greedy strategy

The matrix P of $M \times N$ can be obtained by the updating the position of particle x in the formula, and P_{ij} denotes the position of particle i in position j.

In the optimal problem, each column of matrix P has only one selected P_{ij}, and the selected P_{ij} is the smallest or smaller value of the column. So greedy strategy is introduced to make a quick selection and the specific steps are as follows:

i. Randomly select column $j\ (l, u)$ as the starting column, and select the minimum value of column j.

ii. From column j forward, select the minimum value of the column that meets the constraint conditions column by column.

iii. From column j backward, select the minimum value of the column that meets the constraint conditions column by column.

3.2.3 Opposition-based learning strategy

The opposition-based learning strategy can expand the searching range of the group, exploit the new searching area, and enhance the diversity of the group. Mixed with the artificial electric field algorithm, it can improve the global search ability of the algorithm and prevent the algorithm from falling into the local optimal solution. Therefore, after population updating, the strategy of oppositional learning is applied to the population.

When the position of particle swarm in n-dimensional space is updated as $x^k = (x_1^k, x_2^k, \cdots, x_n^k)$, the corresponding opposite is the elite opposite $\bar{x}^k = (\ \bar{x}_1^k,\ \bar{x}_2^k \cdots\ \bar{x}_n^k)$, where $\bar{x}_i^k = \gamma * (l_i - u_i) - x_i^k$, $\gamma \in$ [0,1] for the random number under the uniform distribution [14]. The sum of the population x^k and $\bar{x}_i^k$ is merged, and 2N particles are sorted according to the ascending order of fitness value, and the N particles before fitness value are selected as the new particle population.

The basic procedures are described below:

Step 1: Initializes the basic parameter and initial population of the algorithm, determines the particle dimension D, the number of charge Population N, and initializes the position x and velocity v of N particles by logistic chaotic map in a given range.

Step 2: Calculate the fitness value of each charge, calculate the Coulomb constant $k(t)$ of the charge, global optimum $best(t)$, and the worst value $worst(t)$.

Step 3: Calculate the Columbian force and acceleration of the charge. Update the velocity v and the position x.

Step 4: Adopt the opposition-based learning strategy to the renewed x population and select the first n individuals of the fitness.

Step 5: Use a greedy strategy to choose x.

Step 6: Judge whether the convergence condition of the algorithm is satisfied, if the termination condition is not satisfied, then return to Step 2; otherwise, output the optimal solution. End the loop.

4 Results and Analysis

This section illustrates the analysis of results obtained from the comparison of the IAEFA and AEFA algorithms and their comparison with other algorithms.

To verify the effectiveness of the improved basic artificial electric field algorithm, nine standard test functions are used to test its performance, and a comparison between the particle swarm optimization algorithm and the basic artificial electric field algorithm is made. The benchmark functions are shown in Table 1. In addition, comparisons between (IAEFA) with other intelligent algorithms are made. The experimental environment of the algorithm is based on the computer under Windows 7 system, MATLAB simulation platform, Inter Core i7-4720 processor, the main frequency of 2.6 GHz.

To verify the validity of the improved IAEFA, contrast experiments are made based on seven algorithms including the improved IAEFA and PSO, AEFA, literature [2] based on the opposite learning AEFA algorithm, Archimedes optimization algorithm (AOA) [15], Condor algorithm (BES) [16], SSA [17], to guarantee the fairness and validity of the experiment. In the simulation experiment, the initial population and iteration times of each algorithm are set to 30 and 1000; the remaining parameters are suggested in the corresponding reference [18], as shown in Table 2.

4.1 Comparison between IAEFA and AEFA on the performance

Table 3 is the experimental results of the two algorithms running 30 times independently on the 9 test function. The spatial dimension is 30. The evaluation results are from the optimal value, the worst value, the average value, the standard deviation, and the running time, and the optimum values are indicated in bold type [19].

In solving the problem of minimum or maximum, the average value can reflect the searching ability of the algorithm, the best value and the worst value can reflect the quality of the solution, and the standard deviation can reflect the robustness of the algorithm. From Table 3, it can be concluded that the overall optimization ability of IAEFA is better than that of AEFA. In the 9 algorithms, 7 of them searched the theoretical optimum and the quality of the solution is better than that of AEFA. It illustrates that in the global search stage, the chaos strategy is used to ensure the diversity of the population

and enhance the ability of global search [20]. From the average results, the unimodal functions F1, F3, F4, F6, and multimodal functions F7-F9, the average values of IAEFA are all 0, and F8 tends to improve compared with the basic algorithm AEFA. It shows that the accuracy of the algorithm in the late period is further improved by introducing an opposition-based learning strategy and a greedy strategy. From the standard deviation results, we can see that the results of IAEFA are better than that of AEFA. Excluding F2 and F5 test functions, the values of the remaining seven test functions are all 0. IAEFA maintains very good robustness; in terms of the running time of the algorithm, that of IAEFA is slightly longer than that of AEFA due to the addition of more policies, which, in combination with other aspects, is within acceptable limits.

4.2 Comparison between IAEFA and AEFA on the improved algorithm and other algorithms

Table 4 is the experimental results of 6 algorithms running 30 times on 9 test functions independently. The space dimension is set to 30, and the evaluation criteria are mean value and standard deviation. The "-" table does not provide the corresponding data in the references, and the optimal results are expressed in bold.

As can be seen from Table 4, the average values of IAEFA in the unimodal functions F1, F3, F4, F6, and

multimodal functions F7, and F9 are all 0. Compared with the other six algorithms, IAEFA is better in the quality of feasible solutions and search precision. There are many local extremum points in function F8, and it is difficult for the algorithm to jump out of the local extremum points in the process of solving. The precision of the improved IAEFA algorithm increased by 15 orders of magnitude compared with AEFA. In the same way, the results of the improved algorithm OBAEFA are relatively good, and the optimal values are found on F7 and F9. Based on the analysis of the standard deviation results, the improved IAEFA algorithms have a standard deviation of 0 in the 7 of the 9 test functions. The results show that the IAEFA algorithm has little fluctuation in the iterative process, and its stability is better than the other 6 algorithms.

The results show that the performance rank of the six algorithms is IAEFA, OBAEFA, BES, AOA and AEFA, PSO. Through the simulation experiment, the effectiveness of the improved algorithm is proved. Finally, it can be concluded that the improved algorithm IAEFA not only keeps the diversity of the population but also speeds up the convergence speed of the algorithm. To a certain extent, it avoids falling into the local optimal solution and further improves the optimization accuracy of the algorithm.

Types	Functions	Function Expressions	Region of search	Extreme value
Unimodal function	Sphere	$f_1(x) = \sum_{i=1}^{n} x_i^2$	[-100.100]	0
	Quartic	$f_2(x) = \sum_{i=1}^{n} i x_i^4 + rand[0,1]$	[-1.28.1.28]	0
	Schwefel2.21	$f_3(x) = max_i\{|x_i|, 1 \le i \le n\}$	[-100,100]	0
	Schwefel2.22	$f_4(x) = \sum_{i=1}^{n} |x_i| + \prod_{i=1}^{n} |x_i|$	[-10,10]	0
	Rosenbrock	$f_5(x) = \sum_{i=1}^{n-1} \left[100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2\right]$	[-30,30]	0
	Rotator hyper-ellipsoid	$f_6(x) = \sum_{i=1}^{n} ([x_i + 0.5])^2$	[-100,100]	0
Multimodal function	Griewank	$f_7(x) = \frac{1}{4000}\sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos\left(\frac{x_i}{\sqrt{i}}\right) + 1$	[-600,600]	0
	Ackley	$f_8(x) = -20 \exp\left(-0.2\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}\right) - \exp(\frac{1}{N}\sum_{i=1}^{N} cos(2\pi x_i))+20+e$	[-32,32]	0

Rastrigin $f_{10}(x) = \sum_{i=1}^{n} \left[x_i^2 - 10\cos^2(2\pi x_i + 10) \right]$ [-5.12,5.12]   0

Table 1:   Benchmark test functions

Algorithm	Parameter
AEFA	Alfa=30;K0=150;
OB-AEFA	Alfa=30;K0=150;
IAEFA	Alfa=30;K0=150;
PSO	W=0.9;c1c2=2.03;wmin=0.4;
SSA	R1,R2,R3=0-1;
BES	A=10;r=1.5;
AOA	C1=2;c2=6;c3=2;c4=0.5;u=0.9;l =0.1;

Table 2: Specific parameters set by each algorithm

Function	Algorithm	Optimal value	The worst value	Mean value	Standard Deviation	Run time
F1	AEFA	1.24E-23	2.36E+00	2.75E-01	5.63E-01	2.6723
	IEAEFA	**0.00E+00**	**0.00E+00**	**0.00+00**	**0.00E+00**	2.1768
F2	AEFA	4.80E-02	3.42E-01	1.9E-01	8.06E-02	1.8059
	IEAEFA	**4.66E-07**	**6.85E-05**	**2.28E-05**	**2.60E-05**	2.3204
F3	AEFA	2.53E+00	8.57E+00	6.07E+00	1.67E+00	1.7088
	IEAEFA	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	2.0955
F4	AEFA	1.71E-03	1.82E+01	4.81E+00	4.88E+00	1.784
	IEAEFA	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	2.107
F5	AEFA	1.53E+02	1.92E+02	1.72E+02	2.73E+01	1.3438
	IEAEFA	**2.85E+01**	**2.86E+01**	**2.86E+01**	**6.38E-02**	1.646
F6	AEFA	5.98E+02	1.94E+03	1.23E+03	3.82E+02	2.052
	IEAEFA	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	2.4963
F7	AEFA	1.07E+01	3.22E+01	2.18E+01	6.86E+00	1.7691
	IEAEFA	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	1.8644
F8	AEFA	1.26E-09	1.77E+00	3.71E-01	5.38E-01	1.6274
	IEAEFA	**8.88E-16**	**8.88E-16**	**8.88E-16**	**0.00E+00**	1.6879
F9	AEFA	1.29E+01	4.87E+00	3.11E+01	8.76E+00	1.8159
	IEAEFA	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	1.8537

Table 3: Experimental results of IAEFA and AEFA

4.3 Comparison between IAEFA, AEFA, OB-AEFA, AOA, BES, and PSO on Convergent curve

According to the average fitness curve, in the unimodal functions F1 and F6, the IAEFA found the theoretical optimum values at about 350 iterations, and for F3, and F4, at about 700 iterations. The other five algorithms are all above IAEFA, the fitness fluctuation value is small, and the theoretical optimum value cannot be found after 1000 iterations. For the function, F8 has many local minimum values and it is easy to get into the local optimum. IAEFA has obtained the theoretical optimal value of about 30 iterations and keeps the state of continuous exploration. Figure 2 to Figure 10 depicts the average fitness curve of function F1, F2, F3, F4, F5, F6, F7, F8, F9 respectively.

For F7, F9, and IAEFA, the convergence speed and the precision are better than those of AEFA, OBAEFA, AOA, BES, and PSO. The effect of the algorithm is remarkable, the convergence curve is always at the bottom, and the theoretical optimal value is found in about 10 iterations. The results show that the algorithm can get the optimal population more quickly in the global search stage and avoid falling into the local optimal solution because of the guidance of the optimal individual in the local search stage. And the convergence

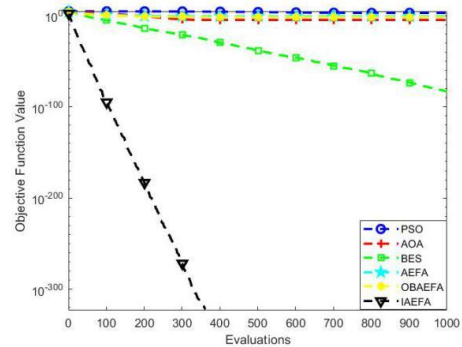speed and accuracy of the algorithm are improved to a great extent.
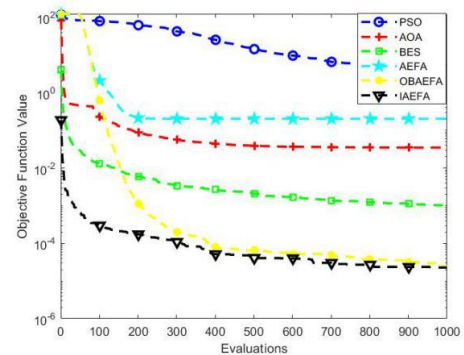

Figure 2: Average fitness curve of function F1

Figure 3: Average fitness curve of function F2

Function		PSO	BES	AOA	SSA	AEFA	OBAEFA	IAEFA
F1	Mean	3.01E-06	3.98E-84	2.11E-05	1.74E-06	2.67E-01	4.59E-02	**0.00E+00**
	Std.dev.	9.15E-06	1.39E-83	1.93E-05	1.05E-06	5.96E-01	2.00E-01	**0.00E+00**
F2	Mean	5.38E+00	1.02E-03	3.48E-02	1.52E-02	2.03E-01	2.39E-05	**2.28E-05**
	Std.dev.	6.66E+00	3.23E-03	1.86E-02	1.13E-02	8.53E-01	5.46E-05	**2.60E-05**
F3	Mean	3.01E+01	6.05E-02	2.79E+00	2.24E-05	6.07E+00	8.66E-01	**0.00E+00**
	Std.dev.	6.42E+00	1.39E-01	1.30E+00	7.80E-06	1.67E+00	4.33E-01	**0.00E+00**
F4	Mean	5.93E+01	8.07E-52	3.33E-04	5.30E-02	4.81E+00	1.61E-14	**0.00E+00**
	Std.dev.	1.89E+01	3.25E-51	2.36E-04	2.33E-01	4.88E+00	1.45E-14	**0.00E+00**
F5	Mean	4.65E+04	1.56E+01	2.80E+01	-	1.72E+02	2.85E+01	**2.82E+01**
	Std.dev.	6.14E+04	2.15E+00	2.13E+00	-	2.73E+01	2.59E-02	**1.38E-02**
F6	Mean	3.40E+04	3.52E-10	5.83E+01	1.40E+02	1.23E+03	6.51E-28	**0.00E+00**
	Std.dev.	1.19E+04	1.93E-09	5.51E+01	1.42E+02	3.82E+02	1.39E-27	**0.00E+00**
F7	Mean	3.01E+01	0.00E+00	1.91E-03	1.58E-02	2.18E+01	0.00E+00	**0.00E+00**
	Std.dev.	4.32E+01	0.00E+00	1.66E-02	1.11E-02	6.86E+00	0.00E+00	**0.00E+00**
F8	Mean	1.59E+01	2.34E-02	1.81E-01	2.16E+00	3.71E-01	4.67E-15	**8.88E-16**
	Std.dev.	7.06E+00	9.99E-02	5.56E-01	6.33E-01	5.38E-01	4.05E-15	**0.00E+00**
F9	Mean	1.58E+02	2.74E+01	2.15E+01	5.21E+01	3.11E+01	0.00E+00	**0.00E+00**
	Std.dev.	2.99E+01	4.91E+01	5.94E+00	1.64E+01	8.76E+00	0.00E+00	**0.00E+00**

Table   4:   Performance   comparison   of   IAEFA   with   modified   AEFA   and   other   algorithms
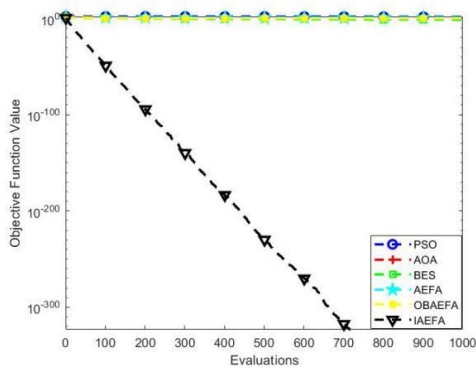

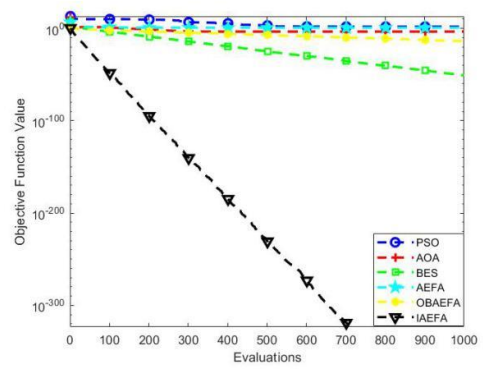
Figure 4: Average fitness curve of function F3

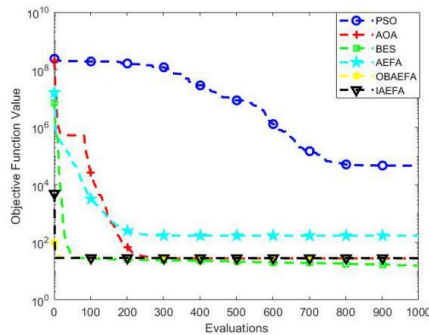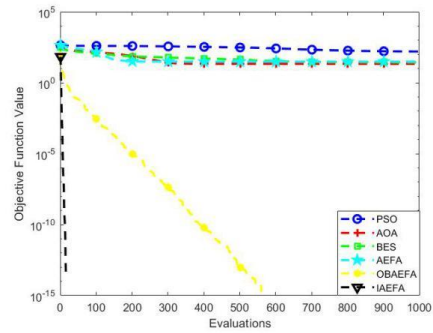Figure 5: Average fitness curve of function F4

Figure 6: Average fitness curve of function F5

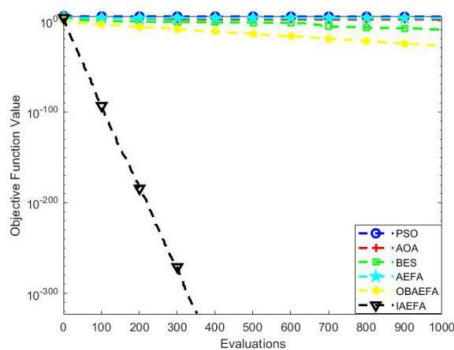Figure 7: Average fitness curve of function F6

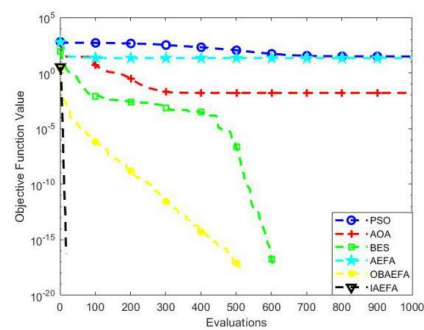Figure 8: Average fitness curve of function F7

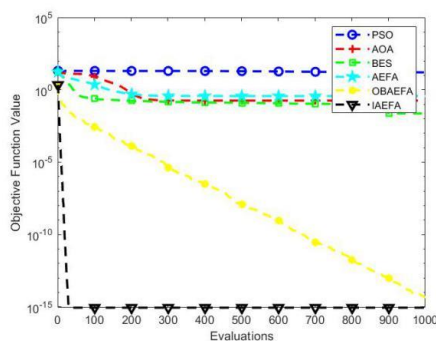Figure 9: Average fitness curve of function F8

Figure 10: Average fitness curve of function F9

5 Several common assessment methods of sand liquefaction

The influence factors of sand liquefaction can be summed up into three categories [22]. Dynamic load: seismic intensity, duration, seismic wave characteristics, etc.; burial conditions: geological factors, soil depth, groundwater level, etc.; Soil conditions: soil type, particle composition, density, etc. In addition, the site shape, geomorphology, and historical earthquake background also have an impact on the foundation soil liquefaction. A description of the factors is given in Table 5.

According to the analysis method of other scholars [23, 24], seven independent variables are selected among numerous influencing factors according to the seismic liquefaction data set provided by reference [25, 26]. Based on seven characteristic indexes, including intensity $I'(X_1)$, groundwater level $d_w(X_2)$, effective overburden pressure $\sigma_0'(X_3)$, blow counts of SPT $N_{63.5}(X_4)$, average grain size $d_{50}(X_5)$, non-uniformity coefficient $C_u(X_6)$ and shear-to-stress ratio $\tau_d/\sigma_0'(X_7)$, the liquefaction of sandy soil is divided into three grades according to the field conditions. The category set is {non-liquefaction (1), critical liquefaction (2) , obvious liquefaction (3)} . The discriminant results of the IAEFA-SVM model and Code for Seismic Design of Buildings (GB5011- 2010) [27] (hereinafter referred to as "Code") and that of the seed simplification method [28] are compared and analyzed. Raw data are shown in Table 6.

5.1 Critical blow counts of SPT for evaluating liquefaction

In the Code for Seismic Design of Buildings 2010 [28], clause 4.3.4 of the code puts forward the formula for evaluating sand liquefaction, within a depth of 20m below the ground, the critical blow counts of SPT of evaluating liquefaction can be calculated as follows in equation 14.

$$N_{cr} = N_0\beta[\ln(0.6d_s + 1.5) - 0.1d_w]\sqrt{3/\rho_c} \qquad (14)$$

In the formula: N_{cr} is the critical value of the blow counts of SPT for evaluating liquefaction; N_0 is the

reference value of the blow counts of SPT for evaluating liquefaction, which can be taken as follows in Table 7. d_s is the depth of penetration point for saturated soil m; d_w is the groundwater level, m; ρ_c is the clay content, when less than 3 or sand is used 3; β is the adjustment factor, the first group takes 0.80, the second group 0.95, and the third group takes 1.05.

5.2   Seed's "simplified procedure"

Seed's "simplified procedure" is the first method proposed abroad to evaluate the liquefaction of saturated sand in a horizontal site [29]. The essence is to compare the Cyclic Resistance Ratio CRR generated by vibration with the Cyclic Stress Ratio CSR to evaluate the liquefaction. The safety factor FS=CRR /CSR, if FS>=1, is judged not to be liquefied, otherwise, it is judged to be liquefied [30].

5.2.1   Cyclic Stress Ratio CSR

The Seed's "simplified procedure" is modified several times, and then converts the cyclic stress ratio into the equivalent CSR7.5 under the magnitude $M_s = 7.5$ after several corrections.

$$CSR_{7.5} = \frac{\tau_d}{\sigma_0'} = 0.65 \times \frac{\alpha_{max}}{g} \times \frac{\sigma_0}{\sigma_0'} \times \gamma_d \tag{15}$$

In the formula, $CSR_{7.5}$ for the earthquake cyclic stress ratio, kPa; τ_d for the average shear stress, kPa; α_{max} for the peak acceleration, m/s^2; g for the gravitational acceleration, m/s^2; σ_0 for the calculated depth of the soil divided by the total vertical stress, kPa; γ_d for the stress reduction factor.

$$\gamma_d = 1.000 - 0.00765z, z \leq 9.15m \tag{16}$$

$$\gamma_d = 1.174 - 0.0267z, 9.15m \leq z \leq 23m \tag{17}$$

z is the depth of the calculated point.

Influencing factor	Description of influencing factors
Dynamic load	When an earthquake is less than magnitude 5, that is, when the epicentral intensity is less than 6, liquefaction will not occur generally [31]. The higher the earthquake intensity, the more serious the sand liquefaction.
Burial conditions	Deeper the sand layer is buried, greater the effective overburden pressure is, and the less easy the sand is to liquefy. The shallower the groundwater is, the smaller the effective pressure is, and the smaller the shear stress is, the easier the sand is to liquefy. The geological factors mainly refer to the geological age and geomorphologic unit. The older the geological age, the better the degree of consolidation, compactness, and structure, and the stronger the anti-liquefaction ability [32-34].
Soil conditions	The average grain size is the main basis for classifying sandy soil, which can reflect the gradation of soil particles. The size of soil particles is related to drainage conditions. The larger the particle size, the less likely it is to liquefy. The non-uniformity coefficient is an index to reflect the uniformity of the composted soil, and it can reflect the gradation of the soil. The well-graded soil has a relatively stable structure, so the well-graded sand is not easy to liquefy [35-37].

Table 5: Factors affecting liquefaction and their description

Serial number	$I(X_1)$	$d_w(X_2)$	$\sigma_0'(X_3)$	$N_{63.5}(X_4)$	$d_{50}(X_5)$	$C_u(X_6)$	$\tau_d/\sigma_0'(X_7)$	Categorization vector
1	7	1.09	50.3	5.0	0.41	2.9	0.1	2
2	7	1.2	34.6	8.0	0.187	4.0	0.09	2
3	7	0.8	20.3	6.0	0.111	2.0	0.08	2
4	7	0.5	21.1	3.0	0.166	1.7	0.1	2
5	7	1.1	42.1	7.0	0.17	1.7	0.1	2
6	7	1.1	71.5	9.0	0.14	2.8	1.11	2
7	7	1.4	55.5	9.0	0.14	1.6	0.1	2

⋮				⋮				⋮
88	8	0.65	57.7	1.1	0.080	1.74	0.234	3
89	9	1.5	76	16.0	0.160	1.80	0.4150	3
90	9	1.45	65.8	5.0	0.055	5.60	0.4070	3

Table 6: Model-training samples

The basic design earthquake acceleration (g)	0.1	0.15	0.2	0.3	0.4
The reference value of the blow counts of SPT for evaluating liquefaction	7	10	12	16	19

Table 7: Reference value of the blow counts of SPT for evaluating liquefaction N_0

5.2.2 Cyclic Resistance Ratio CRR

The cyclic resistance ratio CRR can be calculated from SPT values obtained from standard penetration tests, using the following formula (18):

$$CRR_{7.5} = \frac{1}{34 - (N_1)_{60CS}} + \frac{(N_1)_{60CS}}{135} + \frac{50}{[10(N_1)_{60CS} + 45]^2} - \frac{1}{200} \quad (18)$$

$$(N_1)_{60CS} = \alpha + \beta(N_1)_{60} \quad (19)$$

Among them:

When $FC \leq 5$, $\alpha = 0$, $\beta = 1.0$; when $5 \leq FC \leq 35$, $\alpha = \exp\left[1.76 - \left(\frac{190}{FC^2}\right)\right]$, $\beta = \left[0.99 - \left(\frac{FC^2}{1000}\right)\right]$; and when $FC \geq 35$, $\alpha = 0.5$, $\beta = 1.2$

$CRR_{7.5}$ for the cyclic resistance ratio, $(N_1)_{60CS}$ for the corrected blow counts of SPT, FC for the fines content, $(N_1)_{60}$ for the modified blow counts of SPT when the overburden load is 100kpa and the energy transfer efficiency is 60%.

$$(N_1)_{60} = C_N \cdot N \quad (20)$$

$$C_N = \sqrt{100/\sigma_0'} \quad (21)$$

In the formula, N is the actual blow count; C_N is the adjusted factor of overburden pressure, when C_N is less than 0,4, it takes 0.4, when it is more than 2, takes 2. σ_0' is the effective overburden pressure.

5.3 IAEFA-SVM Model

Support Vector Machine [38-40] is a machine learning approach proposed by Vapnik that has been widely used to analyze and identify patterns. Optimal Separate Hyperplane (Optimum Separate Hyperplane,

OSH) is obtained by using the training set to split the data into two categories to obtain the data categories. As shown in Figure 11 below.
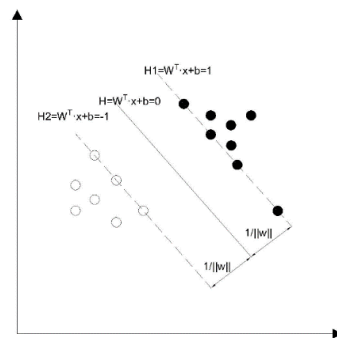

Figure 11: Support vector machine and Optimal Separating Hyperplane

The problem of solving in a linear Support vector machine can be translated into the following problem solving:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \xi_i \quad (22)$$

$$s.t. \, y_i(w \cdot x_i + b) \geq 1 - \xi_i, i = 1,2,\cdots,N \quad (23)$$

$$\xi_i \geq 0, i = 1,2,\cdots,N \quad (24)$$

w is the normal vector of the hyperplane, b is the classification threshold, $\xi_i \geq 0$ is the introduced slack variable, and C is the penalty factor. The size of C indicates the size of the misclassification penalty. The optimal decision function is obtained by the Lagrange multiplier:

$$f(x) = sign[y_i a_i(x \cdot x_i + b)] \quad (25)$$

The nonlinear problem is transformed into a linear problem by being transformed into a high-dimensional space to solve the problem of surface classification. Finally, the optimal decision function becomes:

$$f(x) = sign[\sum_{i=1}^{N} y_i a_i k(x \cdot x_i) + b] \qquad (26)$$

Where $k(x \cdot x_i)$ is the kernel function.

$$k(x \cdot x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \qquad (27)$$

SVM is suitable for solving the problem of small sample size, nonlinearity, high latitude, and local minimum. In the SVM model using Radial Basis Function (RBF) as kernel Function, penalty factor C and kernel function g both affect the performance of SVM. The parameters C and G are optimized by using the algorithm. The flow chart of seismic sand liquefaction evaluation based on IAEFA-SVM is shown in Figure 12.

Step 1: Through the 6:4, 7:3, and 8:2 comparison of seismic data, select the 9:1 ratio in the training set and test set and improve the performance of the model. The input variables are the seven parameters shown above.

Step 2: Set the range of values for C and g and the specific parameters for IAEFA.

Step 3: Calculate the fitness value of IAEFA-SVM.

Step 4: According to Formula (8) ~ (13), update the position of the particle, calculate the fitness value of the current position, and compare it with the previous fitness value, choose a better one.

Step 5: Select the max of iterations as the end indicator, the optimal values of the IAEFA output are the C and g parameters in the SVM model.

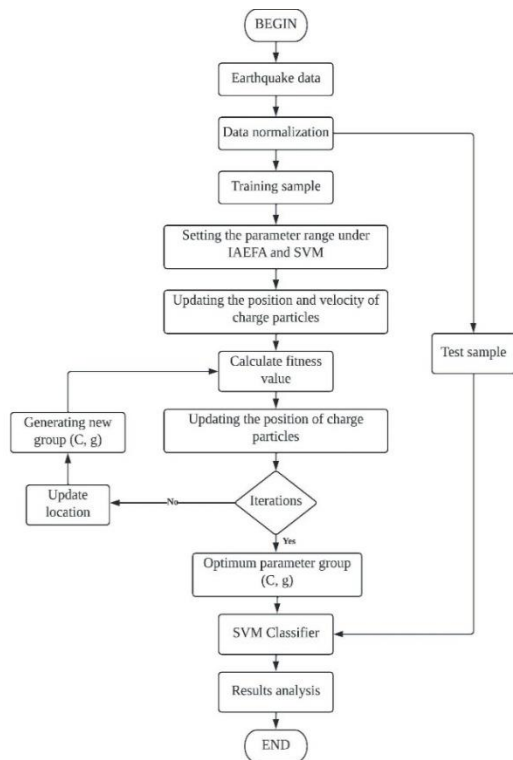Step 6: Take the obtained C and g parameters into the prediction model for testing, and analyze the results.

Figure 12: The flow chart of seismic sand liquefaction evaluation based on IAEFA-SVM

Serial number	I	d_w	d_s	σ'_v	$N_{63.5}$	d_{50}	C_u	τ_d/τ'_v	Measured value	Norm	Seed's	IAEFA-SVM
1	7	0.5	1.7	66.0	3	0.16	1.65	0.10	0	0	0	0
2	7	1.1	6.3	100.0	9	0.14	2.80	0.11	0	0	1	0
3	7	0.7	2.3	17	1	0.07	4.00	0.10	0	0	0	0
4	7	1.4	2.3	82.4	2	0.19	1.90	0.80	0	0	0	0
5	8	3.2	7.2	98.9	8	0.13	2.23	0.172	1	0	0	1
6	8	3.1	9.3	78.3	51	0.32	2.46	0.184	1	1	1	1
7	8	2.3	12.3	140.0	13	0.30	2.43	0.203	1	1	1	1
8	8	1.1	9.22	23.4	12	0.11	2.00	0.225	0	0	0	0
9	8	3	5.1	84.2	9	0.20	2.38	0.159	0	0	0	0
10	8	2	3.46	48.6	8	0.31	2.42	0.163	0	1	0	0
11	9	5	13.52	176.7	64	0.13	2.00	0.34	1	1	1	1
12	9	3.5	8.35	78.5	31	0.21	3.15	0.347	1	1	1	1

Table 8: Evaluation results of sand liquefaction by three methods

5.4 Comparison between Seed's simplification method and IAEFA-SVM model and norm

To prove the accuracy of the model, 78 groups of sample data were trained and 12 groups of sample data were evaluated. And they are also compared with the criterion and the results of Seed's simplification method as shown in Table 8.
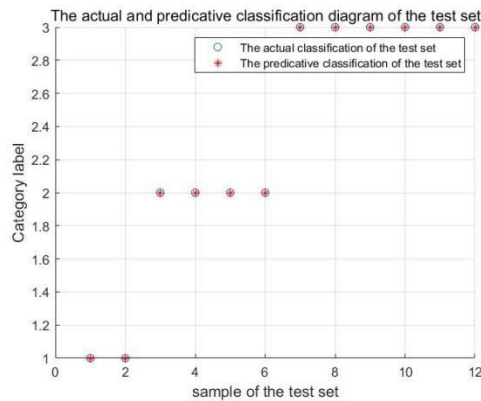

Figure 13: IAEFA-SVM identification diagram

From the comparison results in Table 8, it can be seen that two samples were misjudged by the normalization method and two samples were misjudged by the seed simplification method, the classification accuracy of the IAEFA-SVM model is illustrated. The reason for the error of the standard method is that the method does not take into account some key factors that affect the liquefaction of sand. The reason for the error of the seed simplification method is that it is the empirical discriminant of statistics, and it will have some deviation. It is affected seriously by human factors and has certain limitations.

From the identification diagram as depicted in Figure 13, it can be seen that the accuracy of identifying the degree of sand liquefaction by using the IAEFA-SVM model is 100%. Although there are some differences in the process of (C, g) parameter optimization with IAEFA, it is caused by the randomness of IAEFA in the process of optimization and it does not affect the accuracy of the model. It is proved that the classification effect of IAEFA-SVM is good and it can effectively solve the problem of earthquake liquefaction prediction of sand soil.

6 Conclusion

Based on the analysis of the iterative optimization process of the artificial electric field algorithm, the chaotic strategy is proposed to improve the initial population quality, and the opposite learning strategy and greedy strategy are used to enhance the ability of the algorithm to prevent the local optimal solution.

In the process of benchmark function quota optimization, the results prove the effectiveness of the improved strategy. Based on the analysis of standard deviation results, the IAEFA algorithm can find the theoretical optimal value in 7 out of 9 test functions, the standard deviation of 7 out of 9 test functions is zero, which shows that IAEFA keeps good robustness and has little fluctuation in the iterative process. According to the analysis of the average results, all the six test functions of IAEFA are zero, which shows that the quality of the feasible solution and the search precision of IAEFA can be improved obviously by introducing the opposition-based learning strategy.

Based on the measured data of the earthquake, the seven measured characteristic indexes include intensity, effective overlying pressure, groundwater level, blow counts of SPT, average grain diameter, asymmetrical coefficient, and the shear-to-stress ratio. These characteristics are used as the discriminant indexes of the IAEFA-SVM model. The standard method, seed simplification method, and IAEFA-SVM model were used to distinguish sand liquefaction. In 12 groups of samples, both the standard method and seed simplification method made two misjudges. The accuracy of IAEFA-SVM to identify sand liquefaction reached 100%, providing a new method for the identification of sand liquefaction.

References

[1] Yadav, A. (2019). AEFA: Artificial electric field algorithm for global optimization. *Swarm and Evolutionary Computation*, *48*, 93-108. *https://doi.org/10.1016/j.swevo.2019.03.013*

[2] Demirören, A., Ekinci, S., Hekimoğlu, B., & Izci, D. (2021). Opposition-based artificial electric field algorithm and its application to FOPID controller design for unstable magnetic ball suspension system. *Engineering Science and Technology, an International Journal*, *24*(2), 469-479. *https://doi.org/10.1016/j.jestch.2020.08.001*

[3] Yadav, A. (2020). Discrete artificial electric field algorithm for high-order graph matching. *Applied Soft Computing*, *92*, 106260. *https://doi.org/10.1016/j.asoc.2020.106260*

[4] Yadav, A., & Kumar, N. (2020). Artificial electric field algorithm for engineering optimization problems. *Expert Systems with Applications*, *149*, 113308. *https://doi.org/10.1016/j.eswa.2020.113308*

[5] Hassan, M. H., Kamel, S., El-Dabah, M. A., Khurshaid, T., & Domínguez-García, J. L. (2021). Optimal reactive power dispatch with time-varying demand and renewable energy uncertainty using Rao-3 algorithm. *IEEE Access*, *9*, 23264-23283. *https://ieeexplore.ieee.org/document/9344706*

[6] Sheikh, K. H., Ahmed, S., Mukhopadhyay, K., Singh, P. K., Yoon, J. H., Geem, Z. W., & Sarkar, R. (2020). EHHM: Electrical harmony based hybrid meta-heuristic for feature selection. *IEEE Access*, *8*,

158125-158141.
https://ieeexplore.ieee.org/document/9178740

[7] Xu, H., Zhai, X., Wang, Z., Cui, Z., Fu, Z., & Lu, Y. (2019). An epitaxial synaptic device made by a band-offset BaTiO3/Sr2IrO4 bilayer with high endurance and long retention. *Applied Physics Letters*, *114*(10), 102904.
https://doi.org/10.1063/1.5085126

[8] Singh, P. K., & Sharma, A. (2022). An intelligent WSN-UAV-based IoT framework for precision agriculture application. *Computers and Electrical Engineering*, *100*, 107912.
https://doi.org/10.1016/j.compeleceng.2022.107912

[9] Zeng, H., Dhiman, G., Sharma, A., Sharma, A., & Tselykh, A. (2021). An IoT and Blockchain-based approach for the smart water management system in agriculture. *Expert Systems*, e12892.
https://doi.org/10.1111/exsy.12892

[10] Sharma, A., & Singh, P. K. (2021). UAV-based framework for effective data analysis of forest fire detection using 5G networks: An effective approach towards smart cities solutions. *International Journal of Communication Systems*, e4826.
https://doi.org/10.1002/dac.4826

[11] Sharma, A., Singh, P. K., & Kumar, Y. (2020). An integrated fire detection system using IoT and image processing technique for smart cities. *Sustainable Cities and Society*, *61*, 102332.
https://doi.org/10.1016/j.scs.2020.102332

[12] Tizhoosh, H. R. (2005, November). Opposition-based learning: a new scheme for machine intelligence. In *International conference on computational intelligence for modelling, control and automation and international conference on intelligent agents, web technologies and internet commerce (CIMCA-IAWTIC'06)* (Vol. 1, pp. 695-701). IEEE.
https://ieeexplore.ieee.org/document/1631345/

[13] Karimi, F., Attarpour, A., Amirfattahi, R., & Nezhad, A. Z. (2019). Computational analysis of non-invasive deep brain stimulation based on interfering electric fields. *Physics in Medicine & Biology*, *64*(23), 235010.
10.1088/1361-6560/ab5229

[14] Hashim, F. A., Hussain, K., Houssein, E. H., Mabrouk, M. S., & Al-Atabany, W. (2021). Archimedes optimization algorithm: a new metaheuristic algorithm for solving optimization problems. *Applied Intelligence*, *51*(3), 1531-1551.
https://doi.org/10.1007/s10489-020-01893-z

[15] Alsattar, H. A., Zaidan, A. A., & Zaidan, B. B. (2020). Novel meta-heuristic bald eagle search optimisation algorithm. *Artificial Intelligence Review*, *53*(3), 2237-2264.
https://doi.org/10.1007/s10462-019-09732-5

[16] Sayed, G. I., Khoriba, G., & Haggag, M. H. (2018). A novel chaotic salp swarm algorithm for global optimization and feature selection. *Applied Intelligence*, *48*(10), 3462-3481.

https://doi.org/10.1007/s10489-018-1158-6

[17] Ge, Q., Li, A., Li, S., Du, H., Huang, X., & Niu, C. (2021). Improved Bidirectional RRT Path Planning Method for Smart Vehicle. *Mathematical Problems in Engineering*, *2021*.
https://doi.org/10.1155/2021/6669728

[18] Jeong, W., Jeong, S. M., Lim, T., Han, C. Y., Yang, H., Lee, B. W., & Ju, S. (2019). Self-emitting artificial cilia produced by field effect spinning. *ACS applied materials & interfaces*, *11*(38), 35286-35293.
https://doi.org/10.1021/acsami.9b09571

[19] Petwal, H., & Rani, R. (2020). An improved artificial electric field algorithm for multi-objective optimization. *Processes*, *8*(5), 584.
https://doi.org/10.3390/pr8050584

[20] Selem, S. I., El-Fergany, A. A., & Hasanien, H. M. (2021). Artificial electric field algorithm to extract nine parameters of triple-diode photovoltaic model. *International Journal of Energy Research*, *45*(1), 590-604.
https://doi.org/10.1002/er.5756

[21] Naderipour, A., Abdul-Malek, Z., Mustafa, M. W. B., & Guerrero, J. M. (2021). A multi-objective artificial electric field optimization algorithm for allocation of wind turbines in distribution systems. *Applied Soft Computing*, *105*, 107278.
https://doi.org/10.1016/j.asoc.2021.107278

[22] Yadav, A. (2021). An intelligent model for the detection of white blood cells using artificial intelligence. *Computer methods and programs in biomedicine*, *199*, 105893.
https://doi.org/10.1016/j.cmpb.2020.105893

[23] Sharma, A., & Jain, S. K. (2021). Day-ahead optimal reactive power ancillary service procurement under dynamic multi-objective framework in wind integrated deregulated power system. *Energy*, *223*, 120028.
https://doi.org/10.1016/j.energy.2021.120028

[24] Wang, H., Sharma, A., & Shabaz, M. (2022). Research on digital media animation control technology based on recurrent neural network using speech technology. International Journal of System Assurance Engineering and Management, 13(1), 564-575.
https://doi.org/10.1007/s13198-021-01540-x

[25] Sharma, P., Mishra, A., Saxena, A., & Shankar, R. (2021). A novel hybridized fuzzy PI-LADRC based improved frequency regulation for restructured power system integrating renewable energy and electric vehicles. *IEEE Access*, *9*, 7597-7617.
https://ieeexplore.ieee.org/document/9312597

[26] Alihodzic, A., Mujezinovic, A., & Turajlic, E. (2021). Electric and Magnetic Field Estimation Under Overhead Transmission Lines Using Artificial Neural Networks. *IEEE Access*, *9*, 105876-105891.
10.1109/ACCESS.2021.3099760

[27] Chen, M., Sharma, A., Bhola, J., Nguyen, T. V., &

Truong, C. V. (2022). Multi-agent task planning and resource apportionment in a smart grid. International Journal of System Assurance Engineering and Management, 13(1), 444-455. *https://doi.org/10.1007/s13198-021-01467-3*

[28] Kharrich, M., Kamel, S., Abdeen, M., Mohammed, O. H., Akherraz, M., Khurshaid, T., & Rhee, S. B. (2021). Developed approach based on equilibrium optimizer for optimal design of hybrid PV/Wind/Diesel/Battery microgrid in Dakhla, Morocco. *IEEE Access*, *9*, 13655-13670. *10.1109/ACCESS.2021.3051573*

[29] Chen, Y., Zhang, W., Dong, L., Cengiz, K., & Sharma, A. (2021). Study on vibration and noise influence for optimization of garden mower. Nonlinear Engineering, 10(1), 428-435. *https://doi.org/10.1515/nleng-2021-0034*

[30] Youd, T. L., & Idriss, I. M. (2001). Liquefaction resistance of soils: summary report from the 1996 NCEER and 1998 NCEER/NSF workshops on evaluation of liquefaction resistance of soils. *Journal of geotechnical and geoenvironmental engineering*, *127*(4), 297-313. *https://doi.org/10.1061/(ASCE)1090-0241(2001)127:4(297)*

[31] Chopra, S., Dhiman, G., Sharma, A., Shabaz, M., Shukla, P., & Arora, M. (2021). Taxonomy of adaptive neuro-fuzzy inference system in modern engineering sciences. *Computational Intelligence and Neuroscience*, *2021*. *https://doi.org/10.1155/2021/6455592*

[32] Zhan, X., Mu, Z. H., Kumar, R., & Shabaz, M. (2021). Research on speed sensor fusion of urban rail transit train speed ranging based on deep learning. Nonlinear Engineering, 10(1), 363-373. *https://doi.org/10.1515/nleng-2021-0028*

[33] Han, Z., Chen, M., Shao, S., & Wu, Q. (2022). Improved artificial bee colony algorithm-based path planning of unmanned autonomous helicopter using multi-strategy evolutionary learning. *Aerospace Science and Technology*, *122*, 107374.

https://doi.org/10.1016/j.ast.2022.107374

[34] Liu, C., Lin, M., Rauf, H. L., & Shareef, S. S. (2021). Parameter simulation of multidimensional urban landscape design based on nonlinear theory. Nonlinear Engineering, 10(1), 583-591. *https://doi.org/10.1515/nleng-2021-0049*

[35] Sharma, A., Singh, P. K., Hong, W. C., Dhiman, G., & Slowik, A. (2021). Introduction to the Special Issue on Artificial Intelligence for Smart Cities and Industries. Scalable Computing: Practice and Experience, 22(2), 89-91. *https://doi.org/10.12694/scpe.v22i2.1939*

[36] Wang, H., Wu, Z., Rahnamayan, S., Sun, H., Liu, Y., & Pan, J. S. (2014). Multi-strategy ensemble artificial bee colony algorithm. *Information Sciences*, *279*, 587-603. *https://doi.org/10.1016/j.ins.2014.04.013*

[37] Lu, H., Sun, S., Cheng, S., & Shi, Y. (2021). An adaptive niching method based on multi-strategy fusion for multimodal optimization. *Memetic Computing*, *13*(3), 341-357. *https://doi.org/10.1007/s12293-021-00338-5*

[38] Zhang, X., Rane, K. P., Kakaravada, I., & Shabaz, M. (2021). Research on vibration monitoring and fault diagnosis of rotating machinery based on internet of things technology. Nonlinear Engineering, 10(1), 245-254. *https://doi.org/10.1515/nleng-2021-0019*

[39] Sharma, A., Georgi, M., Tregubenko, M., Tselykh, A., & Tselykh, A. (2022). Enabling Smart Agriculture by Implementing Artificial Intelligence and Embedded Sensing. Computers & Industrial Engineering, 107936. *https://doi.org/10.1016/j.cie.2022.107936*

[40] Zhuang, D. Y., Ma, K., Tang, C. A., Liang, Z. Z., Wang, K. K., & Wang, Z. W. (2019). Mechanical parameter inversion in tunnel engineering using support vector regression optimized by multi-strategy artificial fish swarm algorithm. *Tunnelling and underground space technology*, *83*, 425-436. *https://doi.org/10.1016/j.tust.2018.09.027*

Computer-Aided Architectural Design Optimization Based on BIM Technology

Haiyan Fan [1*], Bhawna Goyal[2], Kayhan Zrar Ghafoor[3,4]
[1]Shandong Polytechnic, Ji Nan, Shandong, 250104, China
[2]Department of ECE, University Centre for Research & Development, Chandigarh University, Mohali, Punjab-140413, India
[3]Department of Computer Science, Knowledge University, Erbil 44001, Iraq
[4]Department of Software & Informatics Engineering, Salahaddin University-Erbil, Erbil, Iraq
E-mail: haiyanfan7@163.com, bhawna.e9242@cumail.in, kayhan.zrar@knu.edu.iq

This article addresses the problem of the non-circulation of information in each stage of architectural design. This paper explores the architectural design process based on the BIM platform and puts forward the structural design method based on the BIM platform. It carries out the seismic analysis of a high-rise building with a transfer floor structure and compares the analysis results with the structural analysis software commonly used by the current design institute. The results obtained for experimentation show that the period ratio, displacement ratio, and the first six modes calculated by the two methods in the modal analysis are consistent. The error between calculation results and PKPM calculation results is within a reasonable range. In the analysis of the mode decomposition response spectrum method, the seismic forces in X and Y directions, floor shear, overturning moment, floor average displacement, and displacement angle obtained by the two models are compared respectively. The analysis results of the two methods accord with the mechanical characteristics of the transfer floor structure, and the calculation error is within the allowable range. The structural design based on the BIM platform has the advantages of high visualization, parameter-driven component size, and high model accuracy, improving design drawing efficiency.

Povzetek: S platformo BIM so izboljšali arhitekturo snovanja na praktičnem primeru seizmične analize.

1 Introduction

In recent years, with the rapid growth of the social economy and the acceleration of urbanization, more and more complex residential buildings and transfinite high commercial complex buildings have sprung up [1-2]. The intervention of CAD has changed the design method and production mode of manual drawing with a drawing board. This not only liberates the engineering designers from the traditional design calculation and repeated manual drawing modification design mode, but also promotes the professionals involved in the project to focus more on the solution of professional problems and the optimization of the design scheme, improves the design quality and improves the modification efficiency of design drawings. However, with the continuous changes of the types of building structures and the structural forms of building components, the relatively simple two-dimensional expression has more and more limitations in the expression of architectural and structural design. BIM, as an extension of the production and application technology of the mechanical industry in the construction industry, provides a new technical idea for the information management and exchange of construction projects [3].

BIM Technology not only provides a solution to improve the quality of architectural design drawings, but also makes the building model and design information better transmitted in the process of building life cycle, and fundamentally solves the problem of non-circulation of information in each stage of design, construction, operation and maintenance [4]. It reproduces the real situation of buildings through computer simulation. It is the third technological revolution in the construction industry. The six characteristics of this technology are visualization, synergy, interoperability, simulation, relevance, and parameterization. BIM Technology has brought unprecedented changes to the traditional working mode and provided a better solution for the needs of fine design. The involvement of computer software has rapidly improved the work efficiency and design quality of the majority of design institutes. However, Auto CAD software presents its design information in the form of point, line, and surface based on the plane, which is basically consistent with the information carried by traditional manually drawn drawings, and does not have much impact on the design method. With more and more special-shaped buildings and more complex building functions, the architectural design method based on CAD software has increasingly exposed many deficiencies [5-6]. Figure 1 shows a

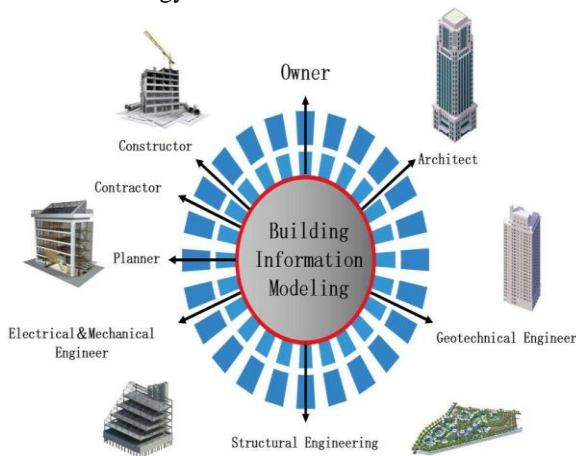design diagram of a computer-aided architecture based on BIM technology.

Figure 1: Optimization of computer-aided building design

The BIM software technology has become a boom these days as the design specifications are entered into the BIM software the 3D model plan is elevated along with the detailed design. The designers utilize the drawing for extracting basic design information in order to deal with the management limitations of the project.

This article addresses the collision problem of design and construction drawings caused by delayed communication among disciplines in the design stage. The collision problem in the process of collaborative design is solved in this article through practical engineering cases and puts forward a solution based on the BIM platform. This article basically analyses the BIM design software in the design stage. This work explores the architectural design process based on the BIM platform. A solution based on the BIM tool is proposed for structural seismic analysis. Aiming at the structural design method based on the BIM platform proposed in this paper, the seismic analysis of a high-rise building with a transfer floor structure is carried out. The analysis results are compared with the analysis results of the current structural analysis software used by the design institute.

The rest of this article is systematized as literature is presented in section 2 followed by research methods in section 3. Section 4 depicts the results and the conclusion is presented in section 5.

2 Related work

In this section various state-of-the-art work in the field of optimization design based on Computer-Aided architecture is presented.

With regard to the application of computer-aided technology in architectural design, Kamel and Memari [7] uses the BIM model established by the calculation software to directly convert the two-dimensional electronic diagram and generate collision reports automatically, in batches, or according to conditions. Nan Fangying and others use the ruling principle to automatically generate multiple design ranges that comply with laws and regulations, and then select the most ideal results of energy consumption simulation to assist in decision-making building volume design. Sayary and Omar [8] and others proposed a method to transform DFS rules into a computer language recognized by Revit, so as to automatically review the design and effectively identify construction safety risks. Du *et al.* [9] and others made a preliminary exploration of the inspection method and process of BIM model quality mainly with the help of the rule inspection software solibri model checker (SMC) v8.0 of solibri company in Finland. Hattab and Hamzeh [10] and others analyzed and summarized the technical advantages of rule checking, expounded the application methods of different types of rules, and further explained the application prospect of rule checking technology from the perspective of solving practical problems and improving work efficiency.

For the application of BIM Technology, Ning *et al.* [11] and others proposed a BIM 3D solid modeling based on a CAD graphics engine based on IFC Standard, which can be transformed into the surface model to meet the application requirements of BIM geometric data for different stages of construction engineering. This method improves the reusability and universality of avoiding data. Heaton *et al.* [12] studied how to combine the BIM technical concept with the current plane representation method of structural construction drawings in China, and analyzed the feasibility of the plane representation method of structural construction drawings based on the BIM platform. A plane representation method of structural construction drawing is proposed, which realizes the correlation of parameters through sharing parameters and label family, realizes the transformation from FIC standard to Revit structural software, and is verified by an example. Lin *et al.* [13] and others analyzed the value and application process of using BIM Technology in prefabricated buildings, studied how to apply BIM Technology to prefabricated houses, and analyzed their adaptability based on actual project cases, providing a reference for the further application of BIM Technology in prefabricated buildings. Mattern and Konig [14] studied the building information model based on Revit software to extract structural information and provide reliable information data for structural analysis, and gave the model conversion method between Revit software and international general structural analysis software SAP2000.

With the growth in the worldwide economy and improvement in technology, the design schemes of domestic engineered architecture have been improving daily, thus, combining the CAD architectural designing with BIM technology [15, 16]. Further with the development in construction technology, architectural designing is also changing from hand-made drawings to CAD-based architects [17, 18]. The CAD architects are using BIM technology which promotes the architectural design to be more scientific and stabilized, thereby improving the efficiency of design in architectural

construction [19]. The improvement in the construction industry is noticed with the involvement of BIM technology with the CAD architecture and has created a high value to the construction industry [20, 21]. This work is also considered for the industrial applications and contributing towards social life with the integration of the Internet of Things, AI, and robotics [22-25].

This article basically introduces the principle of BIM affecting the architectural design using CAD software and thus compared this novel strategy with the other CAD optimization approaches which apply BIM for their technological applicability.

3   Research methods

This section includes the project design process, structural seismic analysis and detailed modeling steps of proposed architecture.

3.1   BIM based construction project design process

When compared with the traditional architectural design method, the architectural design method based on BIM is characterized in that the professional engineers involved in the project do not need to imagine and build a three-dimensional drawing in front of a pile of simple and numerous two-dimensional plans [26]. The BIM-based construction project design process repeatedly compares and calculates the architectural design information, but arranges components and designs architectural information in the virtual three-dimensional space through computer software technology. Based on the understanding of the current BIM platform software, this paper attempts to establish the BIM building structure design process in the design stage. In the process of structural design, the main components of the structure should always be built around the building model, which does not affect the artistic effect and use of the function of the building. Based on the visualization characteristics of BIM core modeling software Revit, the CAD files of the building model and scheme design can be loaded into the new structure template by importing or linking. After completing the structural model in the BIM core modeling software, it is necessary to reasonably select the structural finite element calculation software for trial calculation [27, 28]. Based on the characteristics of BIM platform data sharing, the selection of structural finite element analysis software in the BIM platform shall be based on the following points:

i.   It has a data exchange interface corresponding to BIM core modeling software. The geometric dimensions, load cases, and boundary constraints in the structural model can be directly or indirectly transformed into the structural finite element software as analysis data, which can avoid repeated modeling in the structural analysis software. This data transfer method can improve the efficiency of structural analysis in the process of structural design.

ii.   The structural finite element analysis software

can feed back the model after calculation, analysis, and adjustment to the corresponding BIM core modeling software, so as to update or modify the original model.

The main task in the construction drawing stage is to reflect the final model of each discipline in the preliminary design of the two-dimensional drawing. Before sorting out the construction drawings, we should integrate the needs of architecture, structure, plumbing, and electricity, and further deepen the model of architecture and structure. Rigid structures and prefabricated buildings can simulate the construction of complex hoisting links. The final outcome document of the construction drawing level is to complete the trap drawings of various disciplines of architecture, structure, and equipment that meet the requirements of equipment and material procurement, non-standard equipment manufacturing, and construction [29].

BIM core modeling software Revit architecture software and Revit structure software are modeling software based on parametric design. When the building or structural model is completed, it can be converted into a construction drawing through the plan view of each level. And when the later design changes, whether the construction drawing of the Revit project browser is modified directly or in the 3D model, the components at the corresponding positions of other views will be modified, that is, if one change occurs, the corresponding parts of other drawings will also be changed [30, 31]. Through the project browser of Revit series software, you can efficiently manage design drawings, construction drawings, design descriptions, and other drawing files.

3.2   Structural seismic analysis based on BIM

The structural analysis model based on Revit software is formed while creating the structural geometric model. While creating the geometric model, the analytical model is automatically connected to the nodes. The creation process of the geometric model is carried out in the order of floor-by-floor construction from low to high in the actual construction project. The project consists of a podium and main building [32]. Therefore, when dividing the project for modeling, it can be divided into the main building and podium according to the primary and secondary structure of the project, so as to improve the modeling efficiency. The creation of the BIM structural 3D model is to build an information model with structural component properties through different component families, classes, and elements. This project belongs to frame supported shear wall structure. BIM model is created by taking basic structural components, beams, structural columns, and structural plates as basic elements [33]. The modeling steps of the proposed architecture are depicted in Figure 2.
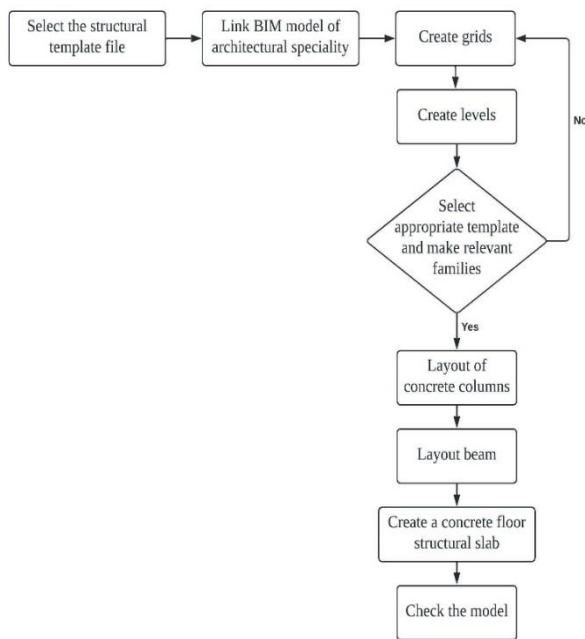
Figure 2: Modeling steps of the proposed architecture

The basic modeling steps are as follows: Select the structure template file; Link BIM model of architecture specialty; Create grids; Create levels; Select the appropriate family template file and make relevant component families; Layout of concrete columns; Layout beam; Create a concrete floor structural slab; Check the model.

After the BIM pattern is adjusted, the BIM pattern will be displayed.

Modal analysis is to analyze the properties of the structure itself. It is the most commonly used and effective analysis method in the seismic response analysis of uncoupled linear structures or decoupled linear structures [34, 35]. At the same time, structural modal analysis is also the analysis basis of response spectrum analysis and time history analysis.

According to D'Alembert's principle, the dynamic balance equation of structural system under earthquake action:

$$F_I(t) + F_D(t) + F_S(t) = F(t) \qquad (1)$$

Where: $F_I(t)$ is inertial force vector acting on node mass; $F_D(t)$ is viscous damping force vector or energy dispersive force vector; $F_S(t)$ is internal force vector borne by structure; $F(t)$ is the load vector imposed on the structure by the outside world.

For seismic action, when the external load F(t) in Equation (1) is equal to zero and the structure is undamped, it can be expressed as a second-order differential equation (2).

$$MX''(t) + KX(t) = 0 \qquad (2)$$

Where M and K are the mass matrix and stiffness matrix of the structural system respectively; X"(t) and X(t) are structural acceleration and displacement vector.

Assuming that each particle vibrates with the same frequency ω, the same phase angle ωt + φ and different amplitude X:

$$X(t) = \{X\}\sin(\omega t + \varphi) \qquad (3)$$

Substitute (3) into the natural vibration equation (4)

$$[K]\{X\}\sin(\omega t + \varphi) - \omega^2[M]\{X\}\sin(\omega t + \varphi) = 0 \qquad (4)$$

The above formula holds for any time, so there is a characteristic equation

$$([K] - \omega^2[M])\{X\} = 0 \qquad (5)$$

It is impossible to obtain $\{X\}$ by the vibration coefficient of each node in the determinant, so Equation (5) must be equal to zero

$$|[K] - \omega^2[M]| = 0 \qquad (6)$$

Through the finite element software SATWE and YJK, the two structural models are calculated respectively, and the 18th order vibration mode is selected for analysis. Read the first 6 vibration modes from the calculation result file, and the structural natural vibration period of each vibration mode is shown in Table 1. In structural design, in order to make the structure have good torsional resistance, the overall torsional deformation resistance of the structure is usually indirectly reflected by the period ratio, that is, the ratio of the first natural vibration period Tt with torsion to the first natural vibration period T1 dominated by translation.

Vibration mode	SATWE		YJK	
	Cycle	Torsion coefficient	Cycle	Torsion coefficient
1	3.0524	0.00	3.0058	0.00
2	2.9234	0.09	2.9043	0.07
3	2.5312	0.91	2.3595	0.93
4	0.9639	0.02	0.8981	0.03
5	0.7879	0.00	0.7505	0.00
6	0.6676	0.98	0.6101	0.96

Table 1: Natural vibration period and vibration mode characteristics of structure

The number of vibration modes calculated by yingjianke software is also 18, and the effective mass coefficient in X direction is 92.12%, and the effective mass coefficient in Y direction is 94.32%, both of which

are greater than 90%, which also meet the specification requirements. It shows that the calculation results of BIM structure model imported into YJK software are basically consistent with those of conventional calculation methods [36].

4   Results and Analysis

This section illustrates the analysis of results obtained by comparing the seismic forces calculated by two programs and finally presents its discussion and summary.

In Figure 3 (a, b), the horizontal seismic forces of each layer under X-direction seismic action and Y-direction seismic action of the two computer calculation methods are compared respectively. The calculation results of the two methods show that the overall variation trend of the horizontal seismic force along the structural height of the two calculation models is basically the same, whether in the X direction or in the Y direction. The seismic force of the fifth floor (i.e., the fourth floor of the building ground) of the two models in the figure is significantly greater than its adjacent upper and lower floors. This is because there are transfer beams with large section and transfer floor slab with thick section in the transfer floor, which makes the transfer floor have large mass and stiffness and will produce large inertial force under the action of earthquake. At the same time, due to the existence of transfer floor, the vertical stiffness of the structure changes suddenly at this floor, resulting in the rapid increase of horizontal seismic force at this floor. In addition, the podium floor at the lower part of the transfer floor has large structural stiffness, resulting in the increase of horizontal seismic forces on it compared with the upper layer. The results show that the high-rise building structure with transfer floor needs to strengthen the seismic design at the transfer floor. The figure shows that the seismic force on the structure in the Y direction is greater than that in the X direction, which also shows that the structural stiffness in the Y direction of the calculation model is greater than that in the X direction. The horizontal seismic force of several floors on the top of the structure has an obvious increasing trend, which shows that the top of the structure is vulnerable to the influence of high-order vibration modes.
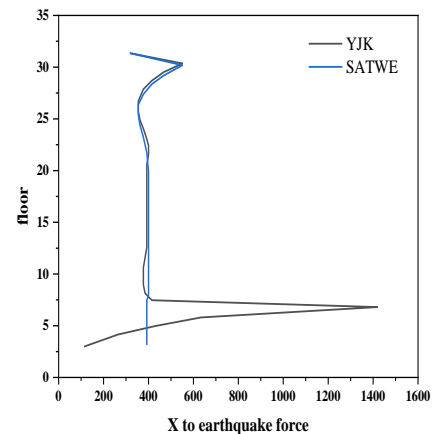
(a)
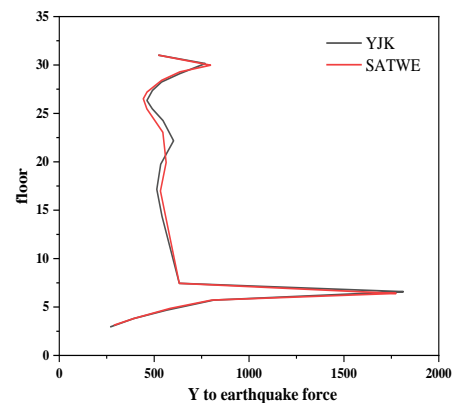
(b)

Figure 3 (a, b): Comparison of seismic forces calculated by two programs

The floor shear force calculated by the two programs under the action of x-direction and Y-direction earthquake is compared in Figure 3. It can be seen from the figure that the floor shear of the two models gradually increases from the top of the structure to the bottom of the structure. The increasing trend of model shear force calculated by YJK (Yingianke software) is basically consistent with that calculated by SATWE. The base shear in Y direction of the two calculation results is larger than that in Y direction. This also validates the analysis results of the above Fig. In the transfer floor and the lower floors of the transfer floor, the increasing trend of floor shear is more obvious than that of the floors above the transfer floor. The reason is analyzed: the transfer floor and the podium floor below have large overall stiffness and bear more seismic shear under horizontal earthquake.

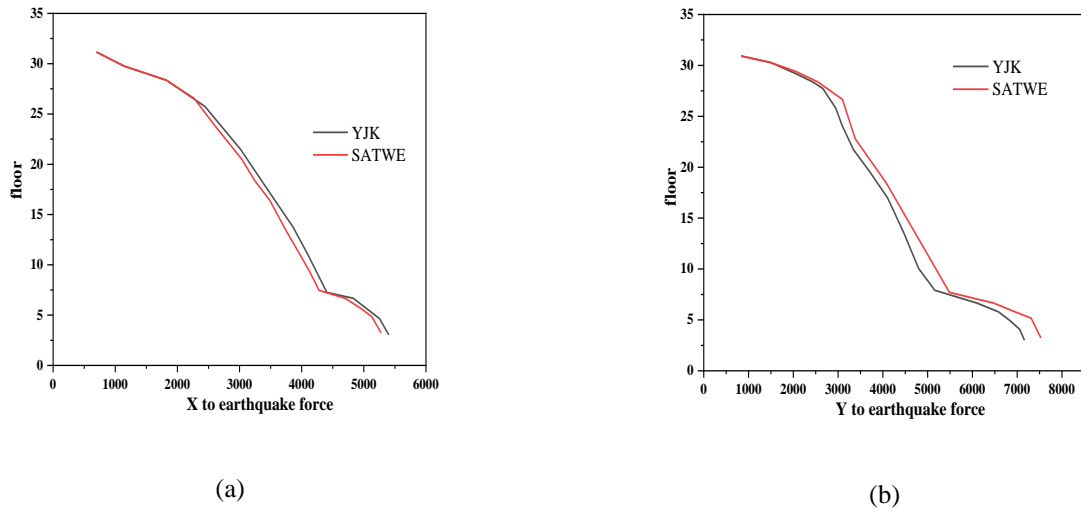(a)   (b)

Figure 4 (a, b): Calculation of floor shear force by two programs

The overturning bending moments in X direction and Y direction of the calculation model are compared through Figure 4 (a, b). It can be seen from the figure that the values of floor overturning bending moment calculated by the two programs are similar and the change trend is similar, which gradually decreases from the top to the bottom of the structure.

	X-direction seismic action			**Y-direction earthquake action**		
Seismic direction	Top floor displacement Δ/mm	Total displacement angle Δ/H	Maximum interlayer displacemen δ/h	Top floor displacement Δ/mm	Total displacement angle Δ/H	Maximum interlayer displacement δ/h
SATWE	57.82	1/1695	1/1036	62.39	1/1571	1/1195
YJK	55.32	1/1772	1/1173	65.75	1/1190	1/1095

Table 2: Top floor displacement and inter floor displacement angle of structural model

(a)   (b)
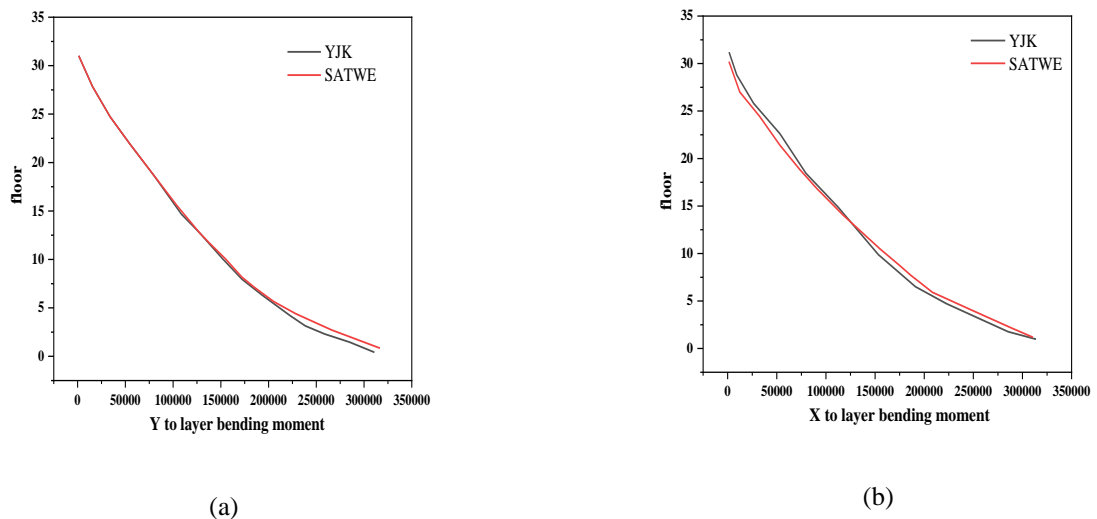
Figure 5 (a, b): Calculation of overturning moment by two programs

The bottom overturning moment calculated by YJK program is slightly smaller than that calculated by SATWE. The bottom overturning moment in Y direction is greater than that in X direction, which is consistent with the analysis results in the previous two figures, indicating that the structural stiffness in Y direction is greater than that in X direction. In the transfer floor and its lower floors, the seismic overturning moment of the floor increases significantly compared with that above the transfer floor, which also verifies the analysis results in Figure 5 (a, b). The outcomes indicate that the overall stiffness of the transfer floor and its lower floors is larger than that of the upper floors of the transfer floor and absorbs more energy from seismic action. According to the theoretical knowledge of seismic design, the floor seismic shear force and floor overturning moment are essentially determined by the magnitude of seismic action. From the above analysis, it can be seen that the transfer floor and its lower floors with large floor stiffness are also subject to large horizontal seismic action. Through the comparison of the above three figures, it can be seen that the calculated values of the two methods are basically similar, and the three data in each method can also be mutually verified.
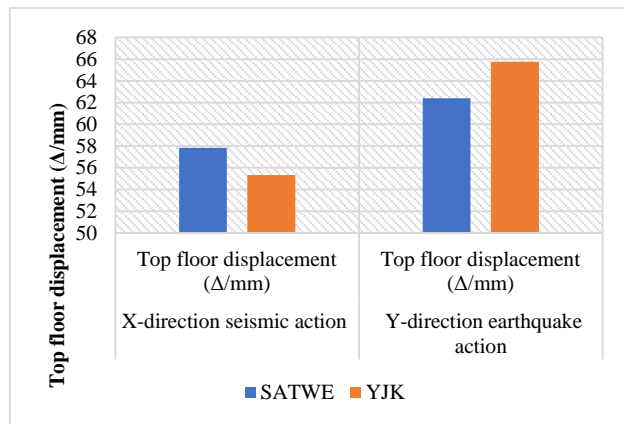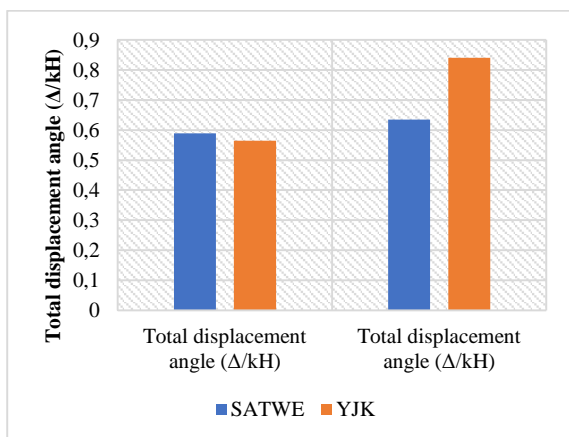
Table 2 lists the top floor displacement, maximum displacement angle and total displacement angle of displacement angle of the two calculation models under the earthquake action in X direction and Y direction. The graphical representation of structural model displacement length, angle and interlayer displacement is depicted in Figure 6, Figure 7 and Figure 8.
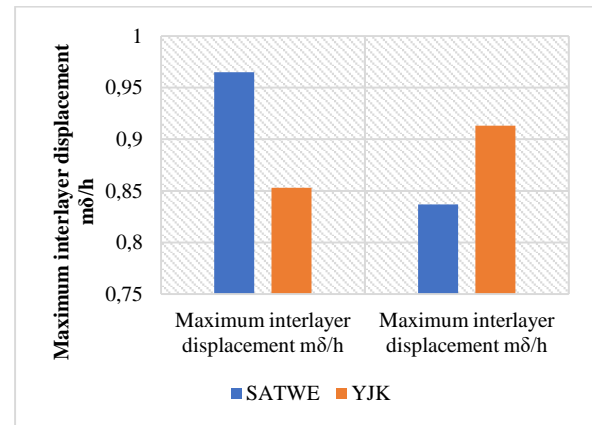

Figure 8: Structural model interlayer displacement

It can be seen from the table that the calculated value of top floor displacement of SATWE is slightly larger than that of YJK. The top displacement of the structure under X-direction seismic action calculated by SATWE is 57.82mm; The displacement of the top floor under Y-direction earthquake is 62.39mm. It is found that the difference range of inter story displacement angles corresponding to the corresponding floors of the two calculation models is within 5%, which meets the allowable error range.

5   Conclusions

The proposed WADO based retinal image transmission technology and structured numerical report in DICOM-SR can better solve the invulnerability problem of retinal image in different systems. The analysis done in this work for the investigation of invulnerable retinal imaging information can be used for quantitative analysis of morphological change of retinal vascular network. This work is mainly focused on the medical digital image transmission protocol Digital Imaging and Communications in Medicine (DICOM) version 3.0 and the retinal image Picture Archiving and Communication System (PACS) was constructed in the laboratory. The retinal image PACS system constructed in B/S mode can effectively store and transmit DICOM images when combined with the application program. This project will integrate quantitative features of retina in future research, providing more meaningful research data for data mining based on chronic disease management system. In addition, a study will be conducted on the conversion of retinal images and reports based on DICOM 3.0 standard and HL7 CDA documents. Therefore, in order to provide a technical basis for the integration of retinal images and existing

Figure 6: Structural model displacement length

Figure 7: Structural model displacement angle

resident health records with HL7 interfaces. The quantitative analysis of retinal morphology data and the original database system text information mining association rules can find more meaningful clinical information. The feasibility of the recognition rate and other evaluation parameters is justified by obtaining the 98.51% accuracy rate with comparatively better values of sensitivity, specificity and precision.

References

[1] Schlueter, A., & Geyer, P. (2018). Linking BIM and Design of Experiments to balance architectural and technical design factors for energy performance. *Automation in Construction*, *86*, 33-43.
https://doi.org/10.1016/j.autcon.2017.10.021

[2] Veloso, P., Celani, G., & Scheeren, R. (2018). From the generation of layouts to the production of construction documents: An application in the customization of apartment plans. *Automation in Construction*, *96*, 224-235.
https://doi.org/10.1016/j.autcon.2018.09.013

[3] Ohki, C., Okamoto, T., Ohga, H., & Yoshizawa, N. (2018). Energy performance evaluation of outside sun shadings using radiance and newhasp. *Journal of Environmental Engineering (Japan)*, *83*(753), 861-870. *https://doi.org/10.3130/aije.83.861*

[4] Ansah, M. K., Chen, X., Yang, H., Lu, L., & Lam, P. T. (2019). A review and outlook for integrated BIM application in green building assessment. *Sustainable Cities and Society*, *48*, 101576.
https://doi.org/10.1016/j.scs.2019.101576

[5] Ren, X., Li, C., Ma, X., Chen, F., Wang, H., Sharma, A., ... & Masud, M. (2021). Design of multi-information fusion based intelligent electrical fire detection system for green buildings. Sustainability, 13(6), 3405.
https://doi.org/10.3390/su13063405

[6] Yuan, Z., Sun, C., & Wang, Y. (2018). Design for Manufacture and Assembly-oriented parametric design of prefabricated buildings. Automation in Construction, 88, 13-22.
https://doi.org/10.1016/j.autcon.2017.12.021

[7] Kamel, E., & Memari, A. M. (2019). Review of BIM's application in energy simulation: Tools, issues, and solutions. *Automation in construction*, *97*, 164-180.
https://doi.org/10.1016/j.autcon.2018.11.008

[8] El Sayary, S., & Omar, O. (2021). Designing a BIM energy-consumption template to calculate and achieve a net-zero-energy house. *Solar Energy*, *216*, 315-320.
https://doi.org/10.1016/j.solener.2021.01.003

[9] Du, J., Zou, Z., Shi, Y., & Zhao, D. (2018). Zero latency: Real-time synchronization of BIM data in virtual reality for collaborative decision-making. *Automation in Construction*, *85*, 51-64.
https://doi.org/10.1016/j.autcon.2017.10.009

[10] Al Hattab, M., & Hamzeh, F. (2018). Simulating the dynamics of social agents and information flows in BIM-based design. *Automation in Construction*, *92*, 1-22.
https://doi.org/10.1016/j.autcon.2018.03.024

[11] Ning, G., Kan, H., Zhifeng, Q., Weihua, G., & Geert, D. (2018). e-BIM: a BIM-centric design and analysis software for Building Integrated Photovoltaics. *Automation in Construction*, *87*, 127-137.
https://doi.org/10.1016/j.autcon.2017.10.020

[12] Heaton, J., Parlikad, A. K., & Schooling, J. (2019). Design and development of BIM models to support operations and maintenance. *Computers in industry*, *111*, 172-186.
https://doi.org/10.1016/j.compind.2019.08.001

[13] Lin, Y. C., Chen, Y. P., Yien, H. W., Huang, C. Y., & Su, Y. C. (2018). Integrated BIM, game engine and VR technologies for healthcare design: A case study in cancer hospital. *Advanced Engineering Informatics*, *36*, 130-145.
https://doi.org/10.1016/j.aei.2018.03.005

[14] Mattern, H., & König, M. (2018). BIM-based modeling and management of design options at early planning phases. *Advanced Engineering Informatics*, *38*, 316-329.
https://doi.org/10.1016/j.aei.2018.08.007

[15] Mesároš, P., Mandičák, T., & Behúnová, A. (2020). Use of BIM technology and impact on productivity in construction project management. *Wireless networks*, 1-8.
https://doi.org/10.1007/s11276-020-02302-6

[16] Chen, R., & Sharma, A. (2021). Construction of complex environmental art design system based on 3D virtual simulation technology. *International Journal of System Assurance Engineering and Management*, 1-8.
https://doi.org/10.1007/s13198-021-01104-z

[17] Jia, M., Komeily, A., Wang, Y., & Srinivasan, R. S. (2019). Adopting Internet of Things for the development of smart buildings: A review of enabling technologies and applications. *Automation in Construction*, *101*, 111-126.
https://doi.org/10.1016/j.autcon.2019.01.023

[18] Sun, H., Fan, M., & Sharma, A. (2021). Design and implementation of construction prediction and management platform based on building information modelling and three-dimensional simulation technology in industry 4.0. *IET collaborative intelligent manufacturing*, *3*(3), 224-232.
https://doi.org/10.1049/ cim2.12019

[19] Wang, P., Sulaimani, H. J., & Kim, S. H. (2021). Digital model creation and image meticulous processing based on variational partial differential equation. *Applied Mathematics and Nonlinear Sciences*.
https://doi.org/10.2478/amns.2021.1.00047

[20] Sun, Y., & Sharma, A. (2022). Research and Design

of High Efficiency Superfine Crusher using CAD Technology. *Computer-Aided Design and Applications*, *19*(S2), 26-38. *https://doi.org/10.14733/cadaps.2022.S2.26-38*

[21] Gong, P., & Li, J. (2022). Application of Computer 3D Modeling Technology in Modern Garden Ecological Landscape Simulation Design. *Security and Communication Networks*, *2022*. *https://doi.org/10.1155/2022/7646452*

[22] Singh, P. K., & Sharma, A. (2022). An intelligent WSN-UAV-based IoT framework for precision agriculture application. *Computers and Electrical Engineering*, *100*, 107912. *https://doi.org/10.1016/j.compeleceng.2022.107912*

[23] Zeng, H., Dhiman, G., Sharma, A., Sharma, A., & Tselykh, A. (2021). An IoT and Blockchain-based approach for the smart water management system in agriculture. *Expert Systems*, e12892. *https://doi.org/10.1111/exsy.12892*

[24] Sharma, A., & Singh, P. K. (2021). UAV-based framework for effective data analysis of forest fire detection using 5G networks: An effective approach towards smart cities solutions. *International Journal of Communication Systems*, e4826. *https://doi.org/10.1002/dac.4826*

[25] Sharma, A., Singh, P. K., & Kumar, Y. (2020). An integrated fire detection system using IoT and image processing technique for smart cities. *Sustainable Cities and Society*, *61*, 102332. *https://doi.org/10.1016/j.scs.2020.102332*

[26] Hossain, M. A., Abbott, E. L., Chua, D. K., Nguyen, T. Q., & Goh, Y. M. (2018). Design-for-safety knowledge library for BIM-integrated safety risk reviews. *Automation in Construction*, *94*, 290-302. *https://doi.org/10.1016/j.autcon.2018.07.010*

[27] Simeone, D., Cursi, S., & Acierno, M. (2019). BIM semantic-enrichment for built heritage representation. *Automation in Construction*, *97*, 122-137. *https://doi.org/10.1016/j.autcon.2018.11.004*

[28] Cavalliere, C., Habert, G., Dell'Osso, G. R., & Hollberg, A. (2019). Continuous BIM-based assessment of embodied environmental impacts throughout the design process. *Journal of Cleaner Production*, *211*, 941-952. *https://doi.org/10.1016/j.jclepro.2018.11.247*

[29] Röck, M., Hollberg, A., Habert, G., & Passer, A. (2018). LCA and BIM: Integrated assessment and visualization of building elements' embodied impacts for design guidance in early stages. *Procedia CIRP*, *69*, 218-223. *https://doi.org/10.1016/j.procir.2017.11.087*

[30] Hilal, M., Maqsood, T., & Abdekhodaee, A. (2018). A scientometric analysis of BIM studies in facilities management. *International Journal of Building Pathology and Adaptation*. *https://doi.org/10.1108/IJBPA-04-2018-0035*

[31] Schlueter, A., & Geyer, P. (2018). Linking BIM and Design of Experiments to balance architectural and technical design factors for energy performance. *Automation in Construction*, *86*, 33-43. *https://doi.org/10.1016/j.autcon.2017.10.021*

[32] Yuan, Z., Sun, C., & Wang, Y. (2018). Design for Manufacture and Assembly-oriented parametric design of prefabricated buildings. *Automation in Construction*, *88*, 13-22. *https://doi.org/10.1016/j.autcon.2017.12.021*

[33] Veloso, P., Celani, G., & Scheeren, R. (2018). From the generation of layouts to the production of construction documents: An application in the customization of apartment plans. *Automation in Construction*, *96*, 224-235. *https://doi.org/10.1016/j.autcon.2018.09.013*

[34] Madan, J., Mani, M., Lee, J. H., & Lyons, K. W. (2015). Energy performance evaluation and improvement of unit-manufacturing processes: injection molding case study. *Journal of Cleaner Production*, *105*, 157-170. *https://doi.org/10.1016/j.jclepro.2014.09.060*

[35] Soust-Verdaguer, B., Llatas, C., & García-Martínez, A. (2017). Critical review of bim-based LCA method to buildings. *Energy and Buildings*, *136*, 110-120. *https://doi.org/10.1016/j.enbuild.2016.12.009*

[36] Ahsan, S., Guo, Z., Miao, Z., Sotiriou, I., Koutsoupidou, M., Kallos, E., & Kosmas, P. (2018). Design and experimental validation of a multiple-frequency microwave tomography system employing the DBIM-TwIST algorithm. *Sensors*, *18*(10), 3491. *https://doi.org/10.3390/s18103491*

Chaotic Association Feature Extraction of Big Data Clustering Based on the Internet of Things

Xiaoming Liu[1*], Thipendra Pal Singh[2], Rajeev Kumar Gupta[3], Edeh Michael Onyema[4]
[1]JingZhou Vocational College of Technology, Software Engineering Institute, Jingzhou, Hubei, 434000, China
[2]School of Computer Science, University of Petroleum and Energy Studies, Dehradun, India
[3]Pandit Deendayal Energy University, Gandhinagar, India
[4]Department of Mathematics and Computer Science, Coal City University, Enugu, Nigeria
Emails: xiaomingliu7@126.com, thipendra@gmail.com, rajeev.gupta@sot.pdpu.ac.in, michael.edeh@ccu.edu.ng

This article addresses the stabilization of chaotic characteristics in abnormal data by proposing chaotic correlation feature extraction of big data clustering based on the Internet of things. The chaotic features in big data usually show complex folding and distortion without obvious rules and order and non-synchronization. In this article, the dimension of extracted correlation is utilized as the chaotic feature for the clustering of big data. The one-dimensional time series that can be extended in multi-dimensional space is analysed based on phase space reconstruction, to extract the chaotic correlation dimension (CCD) features. After the relevant experimental analysis, this paper mainly compares the energy consumption and processing time of the two respective algorithms. In the simulation parameter design, the time interval of big data packet generation is 0.1s, and the data is generated from the simulation time of 300s. The results obtained show that when dealing with the same amount of data, the energy consumption of this algorithm is significantly lower than that of the traditional algorithm. When dealing with the same amount of data, the time required by this algorithm is significantly lower than that of the traditional algorithm. This is because this algorithm is easy to implement and has good clustering efficiency for data, so the clustering time is short. With the gradual increase in the amount of data, the correlation dimension of this algorithm tends to be stable. While the correlation dimension of the traditional algorithm fluctuates greatly, it is revealed that the proposed approach has high data clustering efficiency and verifies the effectiveness of this algorithm.

Povzetek: Za internet stvari je analizirana možnost stabilizacije nenavadnih podatkov znotraj velikih podatkov.

1   Introduction

With the rapid expansion of network technology, the network crime activities in the big data environment are gradually increasing, increasing the amount of abnormal data in the environment of huge data [1]. Therefore, seeking effective big data mining methods is of great consequence to ensure the security of related systems in a big data environment [2]. Most of the current big data mining methods carry out big data mining according to the known abnormal characteristics, which reduces the reliability and efficiency of big data mining, increases the overhead of processing big data, and reduces the overall availability and performance of big data. As revealed in Figure 1, the framework of big data mining and analysis platform [3]. Therefore, how to analyse the failure rate, probability analysis, and adjustment scheme of big data in different regions without interfering with the performance of huge data has an emphasis on the analysis of data mining [4]. In large-scale data mining, massive data brings great difficulties to the existing abnormal data mining

efficiency [5]. How to design sub-region mining algorithms for massive data has gained attention and becomes a research hotspot. Due to the huge amount of data, to reduce the pressure of hardware, when the data scale exceeds the upper limit, it is necessary to partition big data [6]. In the distributed cluster environment without fault tolerance, the efficiency of big data partitioning is inversely proportional to the hardware involved in mining [7]. Therefore, anomaly data mining of massive data is a challenging task. The traditional partition mining algorithm based on mean clustering is affected by data similarity. This kind of partition mining algorithm will produce a high communication load in the parallel process, which is difficult to achieve a high degree of parallelism [8].

There are certain research gaps in the traditional work like the problem of stabilization of chaotic characteristics in abnormal data by proposing the chaotic correlation feature extraction of huge data clustering based on the Internet of things. Also, the chaotic features in big data usually show complex folding and distortion

without obvious rules and order and non-synchronization. The chaotic features are very complex, which are described by the correlation dimension.

Thus, this article contributes to the extraction of the correlation dimension as the chaotic feature of huge data clustering. Based on the reconstruction of phase space, the 1D (one dimensional) time series can be extended in multi-dimensional space, to extract the chaotic correlation dimension features. Cluster analysis of big data is carried out according to the extracted chaotic correlation dimension (CCD). Relevant experimental analysis is carried out in this article and the traditional neural network algorithm is compared in terms of the energy consumption and processing time of the two algorithms. In the simulation parameter design, the time interval of big data packet generation is and the data is

generated at varying simulation times. In the experiment, the amount of data varied from 100MB to 1GB. The correlation dimension of this algorithm is observed to be stable, while the correlation dimension of the traditional algorithm fluctuates greatly, verifying the effectiveness of the proposed algorithm for high data clustering efficiency.

The structure of this paper is arranged as: the literature review is provided in section 2 and the huge data clustering process based on chaotic correlation dimension (CCD) feature extraction is depicted in section 3. The experimental outcomes are presented in section 4 while the conclusion is presented in section 5 of this article.
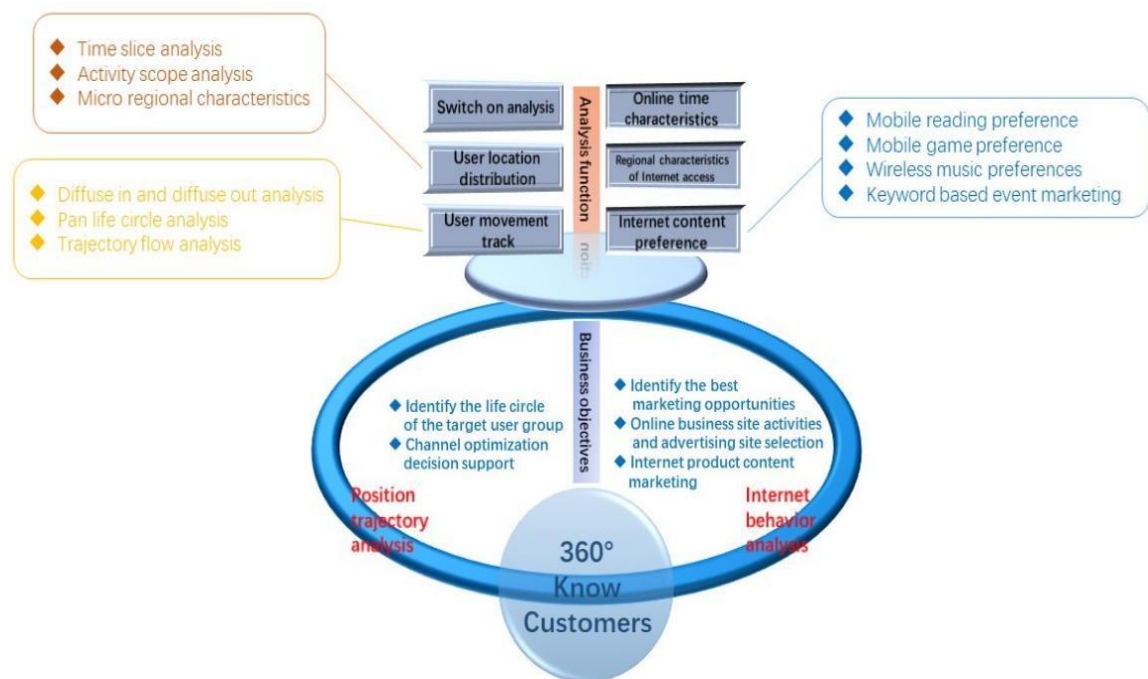

Figure 1: Big data mining and analysis platform

2 Related work

In this section, various state-of-the-art works in the field of feature extraction of big data clustering based on the Internet of Things are discussed.

For this research problem, there are many research methods related to big data clustering of the Internet of things. For example, the cluster analysis method of big data of the Internet of things proposed by Liu *et al.* [9]. Boushaki *et al.* proposed a multi-view fuzzy clustering algorithm based on the condensed information bottleneck audio event clustering method and representing point consistency constraints [10]. Single pass Bayesian fuzzy clustering algorithm and dynamic optimization cellular genetic fuzzy clustering method proposed by Yang *et al.* [11]. RNA SEQ data clustering method proposed by Park

and Lee [12]. Grid coupled data stream clustering method proposed by Cui [13].

Roy *et al.* proposed an uncertain data clustering algorithm based on Voronoi diagram in obstacle space

[14]. Fast density clustering algorithm for location big data proposed by Li *et al.* [15]. Mdfuzzyk modes clustering algorithm based on classification matrix object data proposed by Yan *et al.* [16]. Fast adaptive clustering algorithm based on representative comment scoring strategy and geographic spatiotemporal big data clustering method proposed by Chen *et al.* [17]. The clustering method of Internet of things data in the cloud proposed by song, t, and others has the ability to classify the event big data with chaotic correlation characteristics into their respective clustering centres, and can obtain

satisfactory clustering results. However, from the actual clustering effect, the above traditional methods have some key problems to be solved, such as large time consumption, slow speed, low agility, low data access load, slow convergence, large error, low efficiency of load balanced collaborative filtering. Research on more effective Internet of things big data clustering algorithm based on cloud mode event chaotic correlation feature extraction is rare [18].

Based on the current research, this paper presents the chaotic correlation feature extraction of huge data clustering based on the Internet of things. The chaotic features in big data usually show complex folding and distortion without obvious rules and order and non-synchronization. The chaotic features are very complex, which are described through the correlation dimension. In this article, the dimension of extracted correlation is used as the chaotic characteristic of huge data clustering. Based on the reconstruction of phase space, time series of one-dimensional space can be extended in multi-dimensional space, so as to extract the chaotic features of correlation dimension. Cluster analysis of big data is presented according to the extracted chaotic correlation dimension. The relevant experimental analysis depicts some simulation outcomes which show that the proposed method can accurately mine abnormal data for different large data sets, and has high feasibility and efficiency.

3 Huge data clustering algorithm depending on CCD feature extraction

This section includes the description of clustering algorithm based on chaotic correlation dimension along with the big data clustering implementation.

3.1 Feature extraction and analysis of CCD

The chaotic characteristics in big data are usually complex folding and distortion without obvious rules and order and non-synchronization. The chaotic characteristics are very complex and need to be described by correlation dimension [19].

A. Reconstruction of phase space
The data sequence belongs to nonlinear time series to a great extent, and the key of nonlinear time series is phase space reconstruction. Phase space reconstruction can keep many geometric features in the original system unchanged, establish a bridge between the original time series and multi-dimensional space analysis, and effectively extract the chaotic correlation dimension (CCD) features of data in multi bit phase space. The phase space reconstruction method is as follows: assuming that the time series is $\{x_1, x_2, …, x_N\}$, the phase space reconstruction result can be described as:

$$X = \left[X_1, X_2, …, X_K\right]$$
$$= \begin{bmatrix} x_1 x_2 … x_K \\ x_{1+\tau}, x_{2+\tau}, …, x_K \\ x_{1+(m-1)\tau}, x_{1+(m-1)\tau} … x_{M+(m-1)\tau} \end{bmatrix} \quad (1)$$

Wherein, $K = N - (m-1)\tau$, τ is used to describe the time delay; M is used to describe the embedding dimension. If $m \geq 2d + 1$ the geometric structure of the dynamic system will be completely opened, and d is used to describe the dimension of the chaotic attractor of the system. The selection of embedding dimension m and time delay is the key to phase space reconstruction. Only by selecting reasonable m and τ can we accurately reconstruct the phase space reflecting the characteristics of the original system. The detailed selection method is given below. For the selection of time delay τ. This study considers time delay τ denoted by the abscissa when the mutual data of delay time takes the first minimum value as the finest time delay for recreating phase space [20]. In the interval of data distribution, the probability distribution curve of data is established. p_i is used to describe the probability that $x(t)$ appears in the interval I of the data distribution curve; $p_{ij}(\tau)$ is used to describe the joint probability that $x(t)$ appears in i and delay $x(t + \tau)$ after a certain amount of delay τ appears in region j. Then the delay time mutual information can be described as:

$$I(\tau) = -\sum_{ij} p_{ij}(\tau) \ln \frac{p_{ij}(\tau)}{p_i p_j} \quad (2)$$

If $I(\tau) = 0$, $x(t + \tau)$ cannot be predicted, that is, $x(t)$ and $x(t + \tau)$ are independent of each other, and the smaller $I(\tau)$ is more independent $x(t)$ and $x(t + \tau)$. Therefore, when $I(\tau)$ reaches the minimum, the time delay τ corresponding to the abscissa can be utilized as the finest time delay for recreating the phase space. For the selection of embedding dimension m, this paper uses the virtual nearest neighbor algorithm for the estimation [21]. According to Takens theorem, the $m - dimensional$ vector formed in the $m - dimensional$ phase space can be described as:

$$X(n) = \{x(n), x(n+\tau), …, x(n+(m-1)\tau)\} \quad (3)$$

Obtaining the minimum embedding dimension of phase space reconstruction needs to meet the conditions described in equation (4). If yes, $X_{\eta(n)}$ is called the false nearest neighbor of X_n.

$$\frac{\left|x_{\eta(n)+m\tau} - x_{n+m\tau}\right|}{\left\|X_{\eta(n)} - X_n\right\|_2^{(m+1)}} \geq R_{tol} \quad (4)$$

Where R_{tol} is used to describe the threshold, usually R_{tol} takes 15. At this time, the proportion curve of false nearest neighbour points is required. If the proportion of false nearest neighbour points is less than 5%, it is considered that the obtained m is the minimum embedding dimension of phase space reconstruction [22].

B. Feature extraction of chaotic correlation dimension

In this paper, the extracted CCD is utilized as the chaotic element of huge data clustering. Based on phase space rebuilding, 1-D time series can be stretched out in multi-layered space to separate chaotic element aspect highlights [23]. As per the procedure analyzed in earlier section, the recreated time series can be acquired:

$$X_i = \left(x_i, x_{i+\tau}, \ldots, x_{i+(m-1)\tau}\right)^T \qquad (5)$$

In the $m - dimensional$ phase space recreated by the above-mentioned procedure, the focuses whose separation from phase point x_j to x_i additional x_i itself doesn't surpass r can be portrayed as:

$$Q = \sum_{j \neq i} H\left(r - \|x_i - x_j\|\right) \qquad (6)$$

Where H (*) is utilized to portray the Heaviside work. The idea of connection work is given here. All focuses that might be more modest than the given distance r are comparative with one another The extent of the complete point logarithm is known as the connection work, and the equation is portrayed as follows:

$$C_N(r) = \frac{2}{Q(Q-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} H\left(r - \|x_i - x_j\|\right) \qquad (7)$$

In the equation, the numerator is 2 to wipe out continued counting. The distance between two-stage focuses can be acquired by depicting the distance between two-stage focuses with standard, or at least, the greatest contrast among two vectors:

$$\|x_i - x_j\| = \max_{1 \leq k \leq m} |x_{i-(k-1)\tau} - x_{j-(k-1)\tau}| \qquad (8)$$

For a vector whose distance doesn't surpass r, it tends to be called a cooperative vector [24]. Expecting that there is n $1D$ estimated succession information, the quantity of vector focuses in stage space remaking is $N = m - (m-1)\tau$. Compute the extent of the stage point logarithm with connection in all conceivable $N(N-1)/2$ sets, which is known as the relationship aspect. The recipe is depicted as follows:

$$C_m(r) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N} H\left(r - \|x_i - x_j\|\right) \qquad (9)$$

Then the relationship aspect got above is the tumultuous trademark amount of large information grouping, and the bunching of huge information is acknowledged by the connection aspect.

3.2 Big data clustering implementation

The cluster analysis is to divide different samples into several categories, and make the samples of an aggregate class more similar than those of different aggregate classes [25]. In this paper, huge data is clustered and analysed as per the extracted CCD [26-28]. The flowchart of big data clustering implementation in this article is depicted in Figure 2 and the detailed implementation is provided in this section.

A. Input samples and parameters

Enter n data samples $\{x_1, x_2, \ldots, x_n\}$, According to the characteristics of chaotic correlation dimension, n cluster centers are selected from the above samples and described by $\{Z_1, Z_2, \ldots, Z_n\}$.

B. Divide n samples into the nearest cluster according to the following principles ω_j

$$\|x - z_j\| = \min\left(\|x - z_j\|\right) \qquad (10)$$

Where $\|x - Z_j\|$ is used to describe the distance between x and Z_j. At the same time, it is assumed that there are N_i samples in ω_j.
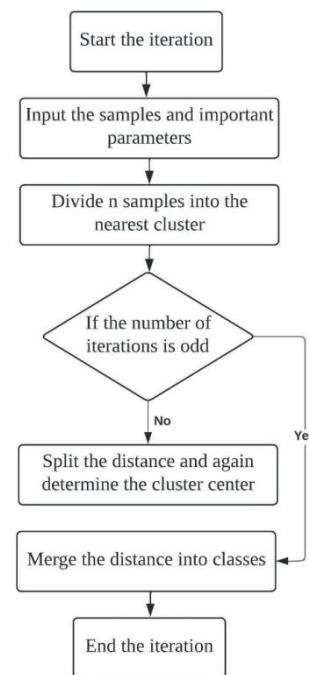

Figure 2: Flowchart of huge data clustering implementation

C. The cluster centre value is obtained by the following formula

$$z_j = \frac{1}{N_j} \sum_{x \in \omega_j} x C_m(r) \qquad (11)$$

If the number of iterations is odd, proceed directly to step (5); Otherwise, follow next step.

D. Split

Assuming $L = \max(x - Z_i), x \in \omega_j$, d_1 is used to describe the splitting distance. If $L > d_1$, ω_j is divided into two categories. At this time, the cluster center can be described as:

$$\begin{cases} z_{i1} = z_i + \lambda L \\ z_{i2} = z_i - \lambda L \end{cases} \qquad (12)$$

Where λ is used to describe a constant greater than 0. If $L < d_1$ and the last merge operation was not performed, proceed to step (6).

E. Merge

Assuming $l = \|Z_I - Z_J\| = \|Z_i - Z_j\|$, use d_2 to describe the merge distance. If $l < d_2$, then, ω_I, ω_J are merged into one class, and the merging center can be described as:

$$z_{IJ} = \frac{1}{N_I + N_J} [N_I z_I + N_J z_J] \qquad (13)$$

If $l < d_2$, and not classified last time, proceed to step (6), otherwise proceed to step (3).

F. End iteration

In this paper, the data with the same chaotic correlation characteristics are divided into one class through the above clustering analysis process, so as to realize the effective clustering of big data [29-31]. This work is also considered for the industrial applications and contributing towards social life with the integration of the Internet of Things, AI, and robotics [32-35].

4 Results and Analysis

This section presents the result analysis obtained for from the proposed big data clustering algorithm and finally presents its discussion and summary in conclusion section.

In order to validate the efficiency of the huge data clustering algorithm based on chaotic correlation feature extraction proposed in this paper, relevant experimental analysis is needed [36-38]. Taking the traditional neural network algorithm as a comparison, the energy consumption and processing time of the two algorithms are mainly compared [39-42]. In this paper, the algorithm is verified by simulation data. All the experimental programs are written in C++, which is in Ubuntu 12.04 operating system. The experimental hardware platform is

LenovoM4390 (i3-2100 CPU, 4UB memory, 2TB disk), processor Intel (R) core (TM) 2duocpu2 94GHz, memory: 8.00GB. In the simulation parameter design, the time interval of big data packet generation is 0.1s, and the data is generated from the simulation time of 300s. In the experiment, the amount of data is from 100MB to 1GB, with 100MB as the unit, the data increases nonlinearly, discrete scheduling and interval boundary approximation are carried out for big data, the time interval of big data feature acquisition is 0.1s, and the parameter configuration is listed in Table 1.

Parameter	Value (Mbps)
Data quantity	1000
Number of big data distribution Characteristics	5
Load per data access system	16
Data complexity size (GB)	2
Data execution time delay (MS)	2400
Maximum queue size	2200

Table 9: Parameter configuration

The algorithm in this paper and the traditional algorithm are used to cluster different amounts of data, and the clustering efficiency of the two algorithms is counted. The outcomes are listed in Table 2 and graphical represented is provided in Figure 3.

Data volume	Time required for the proposed algorithm (s)	Time required for traditional algorithm (s)
200	1925	5998
400	4433	12769
800	8343	29151
1024	10151	35832

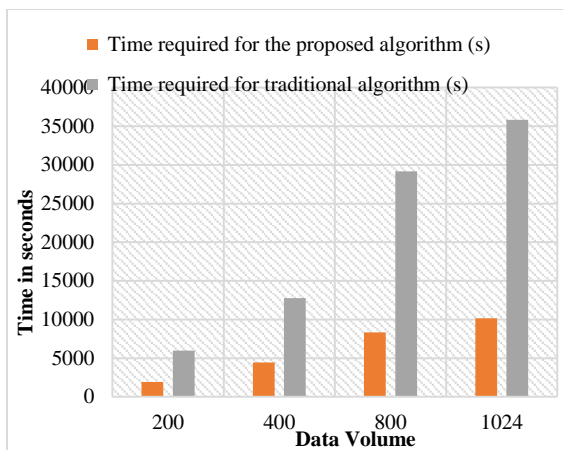Table 10: Comparison results of clustering efficiency of two algorithms

Figure 3: Graphical comparison of clustering efficiency of two algorithms

It can be observed from the analysis of Table 1 and Figure 3 that with the gradual increase of the amount of data, the time for data clustering of this algorithm and the traditional algorithm gradually increases. While this improvement occurs, the processing time required by this algorithm has been potentially lower than that of the traditional algorithm, which shows that this algorithm has high data clustering efficiency and verifies the effectiveness of this algorithm.

In order to further validate the effectiveness of this algorithm, this paper compares the energy consumed by the two algorithms to process the same amount of data. The results are shown in Figure 4.

By analysing Figure 4, it can be seen that when processing the same amount of data, the energy consumption of this algorithm is significantly lower than that of the traditional algorithm. This is because this algorithm is easy to implement and has high clustering efficiency for data, so it consumes less energy, which verifies the effectiveness of this algorithm.
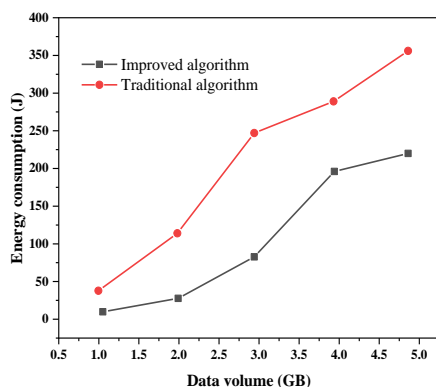

Figure 4: Comparison results of energy consumption of two algorithms

By analysing Figures 5, 6 and 7, it can be seen that when processing the same amount of data, the time required by the algorithm in this paper is significantly lower than that of the traditional algorithm. This is

because the algorithm in this paper is easy to implement and has good clustering efficiency for data, so the clustering time is short, which further verifies the effectiveness of the algorithm in this paper.
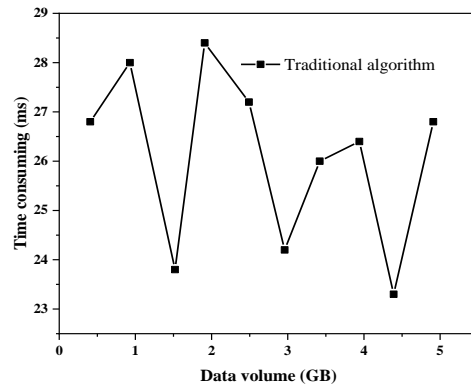

Figure 5: Time consuming of traditional algorithm
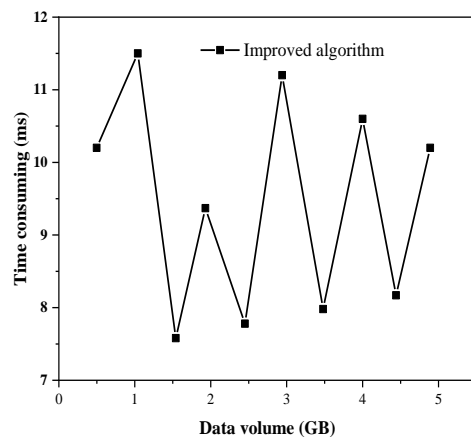

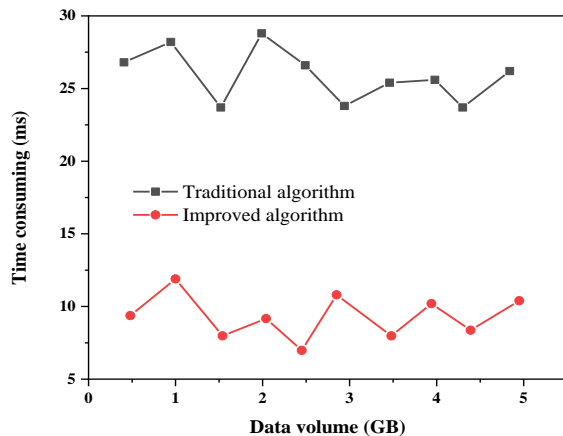Figure 6: Time consuming of improved algorithm

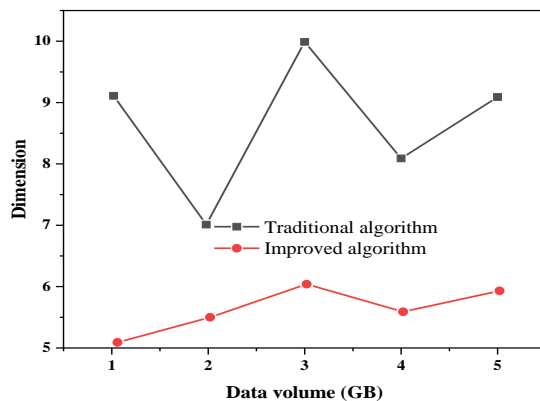Figure 7: Time consuming comparison results of two algorithms

Figure 8: Comparison results of correlation dimensions of two algorithms

It can be seen from the analysis of Figure 8 that with the gradual increase of the amount of data, the correlation dimension of the algorithm in this paper tends to be stable, while the correlation dimension of the traditional algorithm fluctuates greatly. This fluctuation shows that the algorithm in this paper has high data clustering efficiency and verifies the effectiveness of the algorithm in this paper [43-44].

5　Conclusions

This article presents the CCD feature extraction of huge data clustering based on the Internet of things is proposed. By reconstructing the phase space, a multi-dimensional state space vector and chaotic trajectory are established. It was revealed that many geometric features in the creative scheme remain unchanged, which provides an effective basis for analysing the chaotic characteristics of the original system. The false adjacent neighbour procedure is used to select the finest embedding dimension. The extracted CD is used as the chaotic feature of huge data clustering, and the big data is clustered according to the extracted chaotic correlation dimension. Simulation outcomes show that the proposed method can accurately mine abnormal data for different large data sets, and has high feasibility and efficiency. At present, the composition structure, operation mechanism and relevant standards of the Internet of things in cloud mode have not been completely unified. This can act as the future research scope of this article and therefore, the research on big data clustering of the Internet of things needs to be further discussed in many aspects in the future part of this research work.

References

[1] Bu, F. (2018). An efficient fuzzy c-means approach based on canonical polyadic decomposition for clustering big data in IoT. *Future Generation Computer Systems*, *88*, 675-682. *https://doi.org/10.1016/j.future.2018.04.045*

[2] Bu, F., Hu, C., Zhang, Q., Bai, C., Yang, L. T., & Baker, T. (2020). A Cloud-Edge-aided Incremental High-order Possibilistic c-Means Algorithm for Medical Data Clustering. *IEEE Transactions on Fuzzy Systems*, *29*(1), 148-155. *10.1109/TFUZZ.2020.3022080*

[3] Liu, Y., Zhang, J., & Zhan, J. (2021). Privacy protection for fog computing and the internet of things data based on blockchain. *Cluster Computing*, *24*(2), 1331-1345. *https://doi.org/10.1007/s10586-020-03190-3*

[4] Lye, G. X., Cheng, W. K., Tan, T. B., Hung, C. W., & Chen, Y. L. (2020). Creating personalized recommendations in a smart community by performing user trajectory analysis through social internet of things deployment. *Sensors*, *20*(7), 2098. *https://doi.org/10.3390/s20072098*

[5] Cai, G., Fang, Y., Chen, P., Han, G., Cai, G., & Song, Y. (2020). Design of an MISO-SWIPT-aided code-index modulated multi-carrier M-DCSK system for e-health IoT. *IEEE Journal on Selected Areas in Communications*, *39*(2), 311-324. *10.1109/JSAC.2020.3020603*

[6] Xia, H., Huang, W., Li, N., Zhou, J., & Zhang, D. (2019). PARSUC: A parallel subsampling-based method for clustering remote sensing big data. *Sensors*, *19*(15), 3438. *https://doi.org/10.3390/s19153438*

[7] Arora, S., Sharma, M., & Anand, P. (2020). A novel chaotic interior search algorithm for global optimization and feature selection. *Applied Artificial Intelligence*, *34*(4), 292-328. *https://doi.org/10.1080/08839514.2020.1712788*

[8] Maddumala, V. R. (2020). Big Data-Driven Feature Extraction and Clustering Based on Statistical Methods. *Traitement du Signal*, *37*(3). *10.18280/ts.370305*

[9] Liu, W., Wang, X., & Peng, W. (2019). Secure remote multi-factor authentication scheme based on chaotic map zero-knowledge proof for crowdsourcing internet of things. *IEEE Access*, *8*, 8754-8767.

10.1109/ACCESS.2019.2962912

[10] Boushaki, S. I., Kamel, N., & Bendjeghaba, O. (2018). A new quantum chaotic cuckoo search algorithm for data clustering. *Expert Systems with Applications*, *96*, 358-372. *https://doi.org/10.1016/j.eswa.2017.12.001*

[11] Yang, Q., Ruan, J., Zhuang, Z., & Huang, D. (2019). Chaotic analysis and feature extraction of vibration signals from power circuit breakers. *IEEE Transactions on Power Delivery*, *35*(3), 1124-1135. *10.1109/TPWRD.2019.2934123*

[12] Park, S. W., & Lee, I. Y. (2019). Enhanced signature RTD transaction scheme based on Chebyshev polynomial for mobile payments service in IoT device environment. *The Journal of Supercomputing*, *75*(8), 4617-4637. *https://doi.org/10.1007/s11227-018-2546-8*

[13] Cui, Y. (2018). Application of the improved chaotic self-adapting monkey algorithm into radar systems of internet of things. *IEEE Access*, *6*, 54270-54281. *10.1109/ACCESS.2018.2869632*

[14] Roy, S., Chatterjee, S., Das, A. K., Chattopadhyay, S., Kumari, S., & Jo, M. (2017). Chaotic map-based anonymous user authentication scheme with user biometrics and fuzzy extractor for crowdsourcing Internet of Things. *IEEE Internet of Things Journal*, *5*(4), 2884-2895. *10.1109/JIOT.2017.2714179*

[15] Li, L., Wen, G., Wang, Z., & Yang, Y. (2019). Efficient and secure image communication system based on compressed sensing for IoT monitoring applications. *IEEE Transactions on Multimedia*, *22*(1), 82-95. *10.1109/TMM.2019.2923111*

[16] Yan, Z., Liu, J., Vasilakos, A. V., & Yang, L. T. (2015). Trustworthy data fusion and mining in Internet of Things. *Future Generation Computer Systems*, *49*(C), 45-46. *https://doi.org/10.1016/j.future.2015.04.001*

[17] Chen, F., Li, Q., Li, M., Huang, F., Zhang, H., Kang, J., & Wang, P. (2021). Unclonable fluorescence behaviors of perovskite quantum dots/chaotic metasurfaces hybrid nanostructures for versatile security primitive. *Chemical Engineering Journal*, *411*, 128350. *https://doi.org/10.1016/j.cej.2020.128350*

[18] Song, T., Li, R., Mei, B., Yu, J., Xing, X., & Cheng, X. (2017). A privacy preserving communication protocol for IoT applications in smart homes. *IEEE Internet of Things Journal*, *4*(6), 1844-1852. *10.1109/JIOT.2017.2707489*

[19] Alarifi, A., Sankar, S., Altameem, T., Jithin, K. C., Amoon, M., & El-Shafai, W. (2020). A novel hybrid cryptosystem for secure streaming of high efficiency H. 265 compressed videos in IoT multimedia applications. *IEEE Access*, *8*, 128548-128573. *10.1109/ACCESS.2020.3008644*

[20] Niu, Z., Zheng, M., Zhang, Y., & Wang, T. (2019).

A new asymmetrical encryption algorithm based on semitensor compressed sensing in WBANs. *IEEE Internet of Things Journal*, *7*(1), 734-750. *10.1109/JIOT.2019.2953519*

[21] Li, L., Liu, L., Peng, H., Yang, Y., & Cheng, S. (2018). Flexible and secure data transmission system based on semitensor compressive sensing in wireless body area networks. *IEEE Internet of Things Journal*, *6*(2), 3212-3227. *10.1109/JIOT.2018.2881129*

[22] Yung, C., Chen, C. C., Yuan, Y. L., & Li, C. (2019). A Systematic Model of Big Data Analytics for Clustering Browsing Records into Sessions Based on Web Log Data. *J. Comput.*, *14*(2), 125-133. *10.17706/jcp.14.2.125-133*

[23] Jang, S. W., & Kim, G. Y. (2017). A monitoring method of semiconductor manufacturing processes using Internet of Things–based big data analysis. *International Journal of Distributed Sensor Networks*, *13*(7), 1550147717721810. *https://doi.org/10.1177/1550147717721810*

[24] Gong, X., Liu, L., Fong, S., Xu, Q., Wen, T., & Liu, Z. (2019). Comparative research of swarm intelligence clustering algorithms for analyzing medical data. *IEEE Access*, *7*, 137560-137569. *10.1109/ACCESS.2018.2881020*

[25] Lee, Y. C., Huang, S. C., Huang, C. H., & Wu, H. H. (2016). A new approach to identify high burnout medical staffs by kernel k-means cluster analysis in a regional teaching hospital in Taiwan. *Inquiry: The Journal of Health Care Organization, Provision, and Financing*, *53*, 0046958016679306. *https://doi.org/10.1177/0046958016679306*

[26] Shabaz, M., Sharma, A., Al Ajrawi, S., & Estrela, V. V. (2022). Multimedia-based emerging technologies and data analytics for Neuroscience as a Service (NaaS). *Neuroscience Informatics*, *2*(3), 100067. *https://doi.org/10.1016/j.neuri.2022.100067*

[27] Poongodi, M., Hamdi, M., Malviya, M., Sharma, A., Dhiman, G., & Vimal, S. (2022). Diagnosis and combating COVID-19 using wearable Oura smart ring with deep learning methods. *Personal and ubiquitous computing*, *26*(1), 25-35. *https://doi.org/10.1007/s00779-021-01541-4*

[28] Kumbinarasaiah, S., & Raghunatha, K. R. (2021). A novel approach on micropolar fluid flow in a porous channel with high mass transfer via wavelet frames. *Nonlinear Engineering*, *10*(1), 39-45. *https://doi.org/10.1515/nleng-2021-0004*

[29] Wang, H., Sharma, A., & Shabaz, M. (2022). Research on digital media animation control technology based on recurrent neural network using speech technology. *International Journal of System Assurance Engineering and Management*, *13*(1), 564-575. *https://doi.org/10.1007/s13198-021-01540-x*

[30] Ting, L., Khan, M., Sharma, A., & Ansari, M. D. (2022). A secure framework for IoT-based smart climate agriculture system: Toward blockchain and

edge computing. *Journal of Intelligent Systems*, *31*(1), 221-236.
https://doi.org/10.1515/jisys-2022-0012

[31] Sharma, D., Kaur, R., Sandhir, M., & Sharma, H. (2021). Finite element method for stress and strain analysis of FGM hollow cylinder under effect of temperature profiles and inhomogeneity parameter. *Nonlinear Engineering*, *10*(1), 477-487.
https://doi.org/10.1515/nleng-2021-0039

[32] Ren, Y., Rubaiee, S., Ahmed, A., Othman, A. M., & Arora, S. K. (2022). Multi-objective optimization design of steel structure building energy consumption simulation based on genetic algorithm. *Nonlinear Engineering*, 11(1), 20-28.
https://doi.org/10.1515/nleng-2022-0012

[33] Singh, P. K., & Sharma, A. (2022). An intelligent WSN-UAV-based IoT framework for precision agriculture application. *Computers and Electrical Engineering*, *100*, 107912.
https://doi.org/10.1016/j.compeleceng.2022.107912

[34] Zeng, H., Dhiman, G., Sharma, A., Sharma, A., & Tselykh, A. (2021). An IoT and Blockchain-based approach for the smart water management system in agriculture. *Expert Systems*, e12892.
https://doi.org/10.1111/exsy.12892

[35] Sharma, A., & Singh, P. K. (2021). UAV-based framework for effective data analysis of forest fire detection using 5G networks: An effective approach towards smart cities solutions. *International Journal of Communication Systems*, e4826.
https://doi.org/10.1002/dac.4826

[36] Sharma, A., Singh, P. K., & Kumar, Y. (2020). An integrated fire detection system using IoT and image processing technique for smart cities. *Sustainable Cities and Society*, *61*, 102332.
https://doi.org/10.1016/j.scs.2020.102332

[37] Gunupudi, R. K., Nimmala, M., Gugulothu, N., & Gali, S. R. (2017). CLAPP: A self constructing feature clustering approach for anomaly detection. *Future Generation Computer Systems*, *74*, 417-429.
https://doi.org/10.1016/j.future.2016.12.040

[38] Sharma, A. (2021). Integrity and Multimedia Data Management using Emerging Technologies in the Healthcare Applications-Part II. Recent Advances in Electrical & Electronic Engineering (Formerly Recent Patents on Electrical & Electronic Engineering), 14(7), 698-699.
https://doi.org/10.2174/235209651407211104091930

[39] Wang, Y., Chen, Q., Kang, C., & Xia, Q. (2016). Clustering of electricity consumption behavior dynamics toward big data applications. *IEEE transactions on smart grid*, *7*(5), 2437-2447.
https://doi.org/10.1109/TSG.2016.2548565

[40] Guo, Z., & Xiao, Z. (2021). Research on online calibration of lidar and camera for intelligent connected vehicles based on depth-edge matching. *Nonlinear Engineering*, 10(1), 469-476.
https://doi.org/10.1515/nleng-2021-0038

[41] Deng, Z., Hu, Y., Zhu, M., Huang, X., & Du, B. (2015). A scalable and fast OPTICS for clustering trajectory big data. *Cluster Computing*, *18*(2), 549-562.
https://doi.org/10.1007/s10586-014-0413-9

[42] Chen, Y., Zhang, W., Dong, L., Cengiz, K., & Sharma, A. (2021). Study on vibration and noise influence for optimization of garden mower. *Nonlinear Engineering*, 10(1), 428-435.
https://doi.org/10.1515/nleng-2021-0034

[43] Sharma, A., Singh, P. K., Hong, W. C., Dhiman, G., & Slowik, A. (2021). Introduction to the Special Issue on Artificial Intelligence for Smart Cities and Industries. Scalable Computing: Practice and Experience, 22(2), 89-91.
https://doi.org/10.12694/scpe.v22i2.1939

[44] Luna-Romera, J. M., García-Gutiérrez, J., Martínez-Ballesteros, M., & Riquelme Santos, J. C. (2018). An approach to validity indices for clustering techniques in big data. *Progress in Artificial Intelligence*, *7*(2), 81-94.
https://doi.org/10.1007/s13748-017-0135-3

Application and Study of Artificial Intelligence in Railway Signal Interlocking Fault

Hongwei Liang[1]*, Xiuxuan Wang[1], Anjali Sharma[2], Mohd Asif Shah[3]
[1]Zhengzhou Railway Vocational & Technical College, Zhengzhou, Henan ,451460, China
[2]School of Biological and Environmental Sciences, Shoolini University of Biotechnology and Management Sciences, Solan 173229 (H.P.), India
[3]Bakhtar University, Kabul, Afganistan
Emails: hongweiliang7@163.com,   xiuxuanwang9@126.com,   anjalisharmaas8749347@gmail.com, ohaasif@bakhtar.edu.af

The rapid development of railway transportation towards high speed, high density and heavy load has led to even higher requirements for the safety of railway signal equipment. The safety of railway signal equipment is an important part of ensuring railway traffic safety, thus, it is very necessary to study a system that can diagnose the fault of railway signal equipment according to the actual situation. This article utilizes the deep learning algorithm of artificial intelligence for investigating the interlocking faults in the railway transportation. This paper uses ADASYN data synthesis method to synthesize few category samples, uses TF-IDF to extract features and transform vectors, and proposes a deep learning integration method based on combined weight. The results show that BiGRU has better overall classification performance when evaluated on the index of primary and secondary fault classification accuracy. The classification accuracy improvement of 5% is achieved for primary fault classification and the comprehensive evaluation index of secondary fault classification is improved by about 9%. It was revealed that when compared with ADASYN + BiLSTM neural network, the comprehensive evaluation index of primary fault classification accuracy is improved by about 6%, and the comprehensive evaluation index of secondary fault classification is improved by about 10%. It is demonstrated that deep learning integration is an effective method to improve the classification performance of turnout fault diagnosis model.

Povzetek: Za železniški sistem je bila uporabljena metodologija globokih nevronskih mrež za iskanje napak v signalih.

1   Introduction

With the gradual increase of railway traffic density and operation speed in China, it is difficult to avoid various faults of railway signal equipment. If the faults cannot be handled in a short time, they will have a great impact on traffic safety, and even lead to the hidden dangers of major accidents, so as to reduce the efficiency and safety of railway operation. At the same time, it also brings new challenges to railway signal equipment maintenance personnel to check and maintain signal equipment timely and accurately.

High speed railway signal equipment is an important infrastructure to ensure high-speed train operation. The maintenance quality of signal equipment directly affects the traffic safety and transportation efficiency of high-speed railway. Signal equipment fault is diagnosed and handled according to the experience and knowledge of on-site maintenance personnel, which is easy to cause maintenance judgment error and maintenance time delay, and in serious cases, it will lead to equipment fault driving accident. The fault data of high-speed railway signal equipment records the fault phenomenon when the fault occurs in the form of text.

The fault phenomenon is analyzed based on text data mining technology. Combined with the diagnosis results of experts on the fault phenomenon, the fault diagnosis model of signal equipment is studied to assist maintenance personnel to quickly locate the fault location and cause according to the fault phenomenon. It will be of great significance to further improve the safety guarantee level of high-speed railway. The basic activity diagram of train fault detection method is shown in Figure 1.
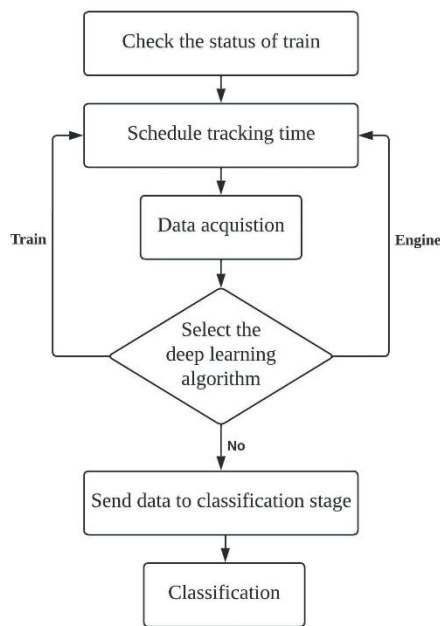
Figure 1: Activity diagram of Railway fault detection method

This limitation of imbalanced faults of different signal equipment is addressed in this article. In order to study the signal equipment fault diagnosis method based on unbalanced samples based on text mining technology, two problems need to be solved: one is the processing of unbalanced samples, and the other is the construction of fault diagnosis and classification model.

This article contributed in mainly using two methods to solve the sample imbalance problem: one is to synthesize the sample data by using data enhancement, under sampling or oversampling, and data generation methods such as SMOTE (Synthetic Minority Oversampling Technology) and ADASYN (Adaptive Synthetic Sampling). The other is to adjust the parameters of different categories for the classification learning algorithm. The sample synthesis algorithm can appropriately synthesize a few categories of samples according to the distribution of the overall samples, and can ensure that the sample data is not repeated. There are several articles which uses SVM-SMOTE method to automatically synthesize the few category samples of signal equipment fault, so as to solve the problem of signal equipment fault sample imbalance. This article utilizes the deep learning algorithm of artificial intelligence for investigating the interlocking faults in the railway transportation. This paper uses ADASYN data synthesis method to synthesize few category samples, uses TF-IDF to extract features and transform vectors, and proposes a deep learning integration method based on combined weight. The outcomes obtained for the proposed method reveals that BiGRU has better overall classification performance when evaluated on the index of primary and secondary fault classification accuracy.

The rest of this article is structured as: review of literature is provided in section 2 followed by research methodology involved in analysis of fault diagnosis of railway unlocking system in section 3. Section 4 provides the experimental results and discussion along with concluding remarks in section 5.

2 Related work

In this section various state-of-the-art work in the field of railway signal interlocking fault based on artificial intelligence and other technologies is presented.

With the advent of the intelligent era, artificial intelligence has become the mainstream technology in the world, and artificial intelligence technology has laid a solid research foundation [1]. Paek and Kim explores the future direction of education by examining the current impact of artificial intelligence and predicting the future impact [2]. Interlocking is a railway system, which can automatically control safety management route change and avoid train collision and derailment. Dobias and Kubatova analyzes the latest technologies used in several commercial interlocking equipment, and proposed the design and implementation of an interlocking system architecture based on FPGA technology [3]. In order to solve the problem of channel estimation based on demodulated reference signal (DMRs) in railway tunnel scene, Skiribou et al. proposed a deterministic model to accurately generate time-varying channel response [4]. Kiedrowski and Saganowski introduced a scheme of applying PLC technology to railway light signs. This paper introduces the structure of the network and a group of equipment to realize this specific type of wired sensor network, which is used to monitor the railway led sign network and maintenance parameters [5]. Yang et al. analyzed the requirements of clock synchronization of signal ground equipment in combination with the application status of clock synchronization of ground equipment in high-speed railway signal system. By analyzing the advantages and disadvantages of the world's mainstream satellite navigation system and the requirements of China's railway signal system, Beidou time service technology is selected as the clock synchronization technology of the ground equipment of high-speed railway signal system, and the overall scheme based on Beidou time service technology is constructed [6]. In order to evaluate the network access performance of railway signal equipment machine communication (MTC) in the next generation intelligent transportation system, Lin et al. divided the railway signal equipment machine communication traffic prediction model into station indoor model, station outdoor model and station outdoor model, and calculated the traffic and signaling overhead of the three models respectively. Based on Poisson distribution and Markov renewal process, an improved Markov modulated poisson process (immpp) for source traffic model is designed [7]. Wang et al. combined with the new technical characteristics of high-speed railway, analyzed the current situation of lightning protection technology and lightning faults of foreign railway signal equipment. At the same time, the functions of intelligent technologies such as lightning activity location and lightning fault diagnosis are

introduced, and the development direction of railway lightning protection in the future is prospected according to the characteristics of this technology [8]. In order to realize the real-time acquisition, monitoring and management of the technical status of railway signal equipment and meet the multi-dimensional business needs of railway signal system information sharing, data mining, analysis and display, Sahal *et al.* put forward the national technical big data platform of railway signal equipment on the basis of analyzing the current situation of railway signal system and the significance of signal big data platform construction [9]. Based on the common signal system equipment of rail transit stations at home and abroad, Cao *et al.* analyzed the common faults and their settings of the system, studied the common faults analysis, design and construction of the signal system, and developed the railway signal fault setting training system based on the core concept of fault safety design [10]. In order to solve the problem of railway transportation safety, Dong *et al.* carried out detection experiments on simulated images and real videos of railway signal lights based on machine vision. The image features of railway signal lights in different color spaces and their influence on railway signal light recognition are discussed [11].

Railway signal equipment safety is an important part of ensuring railway traffic safety, thus, it is very necessary to study a system that can diagnose the fault of railway signal equipment according to the actual situation. The literature suggests that there are many studies on using data synthesis method to solve the sample imbalance based on the deep learning of artificial intelligence approach [12-15]. This paper diagnoses the fault of high-speed railway signal equipment, improves the performance of equipment fault diagnosis, so as to improve the safety of railway.

3 Research methods

This section includes the description of small category sample generation based on ADASYN. The fault text features of high-speed railway signal are also represented in this section and fault diagnosis model is presented.

High speed railway signal fault diagnosis forms a turnout fault diagnosis model with deliverable evaluation indexes through the training and optimization of the fault diagnosis model based on deep learning integration [16]. The turnout fault phenomenon of high-speed railway is input into the fault diagnosis model, and the model automatically outputs the type and cause of the fault, so as to realize the intelligent diagnosis of turnout equipment fault [17-19]. The architecture of this research work is depicted in Figure 2.
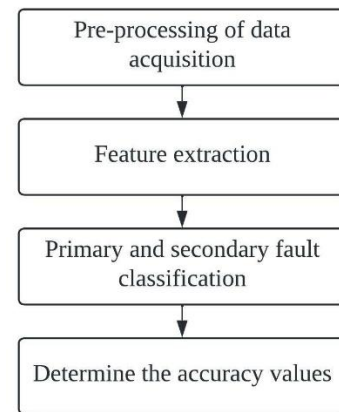

Figure 2: Architecture of research work

The basic structure of this research work includes pre-processing of data acquired from various sources. Further, the feature set is extracted followed by the classification of primary and secondary faults [20, 21]. At the final stage, accuracy values are determined for the proposed architecture. The development of artificial intelligence and Internet of Things is considered for several industrial applications and contributing towards social life [22-25].

3.1 Small category sample generation based on ADASYN

ADASYN adaptive synthesis oversampling method is to adaptively synthesize a small number of samples according to the distribution of a small number of samples, and synthesize fewer samples where it is easy to classify and more samples where it is difficult to classify. The key of the synthesis algorithm is to find a probability distribution r_i. Put r_i is the criterion for determining how many samples should be synthesized for each small category sample.

The proportion of the number of secondary categories included in each primary fault category of high-speed railway signal turnout fault is 12:17:8: 11:7:1:7. Therefore, ADASYN is used to synthesize fewer secondary fault category samples, and the imbalance of primary fault categories can be solved at the same time. The process of using ADASYN to adaptively generate turnout secondary few category samples is as follows:

Step 1: Calculate the unbalance degree of few categories, $d = m_s/m_l$, m_s and m_l represent the number of samples with few categories and multiple categories respectively, $d \in (0, 1]$.

Step 2: Calculate the total number of small category samples to be synthesized, $G = (m_l - m_s) \times \beta$, $\beta \in (0, 1]$, indicating the expected imbalance degree of the whole sample after adding the synthetic sample, $\beta= 1$ means that the sample category is completely balanced after adding the synthetic sample.

Step 3: For each sample of a few categories x_i. Find their K-nearest neighbors in *n*-dimensional space and

calculate the ratio $r_i = \Delta_i/K (i = 1,2,...,m)$, m is the total number of samples, Δ_i is the number of multiclass samples in the k-nearest neighbor of x_i, so $r_i \in (0,1]$。

Step 4: According to $\hat{r}_i = r_i / \sum_{i=1}^{m_s} r_i$, regularize r_i. So r_i is the probability distribution, and $\sum \hat{r}_i = 1$.

Step 5: Calculate the number of samples $g_i = \hat{r}_i \times G$ to be synthesized for each small category sample x_i, G is the total number of synthetic samples.

Step 6: According to the above steps, calculate the number of samples g_i synthesized by each small category sample x_i.

3.2 Fault text feature representation of high-speed railway signal equipment

TF-IDF is a text feature representation method based on weighting idea. Its core idea is that if a word appears frequently in one document and low in other documents, it indicates that the word has high recognition in the document and assigns its high weight. The feature extraction of signal equipment fault text first needs to realize Chinese word segmentation [26-29]. Because the high-speed railway signal equipment fault text data contains professional words such as switch machine, red light band and sealer, this paper constructs railway signal professional thesaurus and loads the thesaurus into Jieba word segmentation tool to realize the accurate word segmentation of fault text.

Text frequency (TF) in TF-IDF refers to the frequency of a given word in the document. For a given word t_i. In a document d_j, the degree of importance can be expressed as:

$$TF_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \tag{1}$$

Where: $n_{i,j}$ is the number of occurrences of the i-th word in document d_j. $\sum_k n_{k,j}$ is the total number of occurrences of each word in document d_j.

The inverse file frequency IDF is a measure of the general importance of a word. Its calculation formula is as follows. The larger the IDF, the better the ability to distinguish categories.

$$IDF_i = \log_2 \frac{|D|}{1 + |j: t_i \in d_j|} \tag{2}$$

Where: D is the total number of sample files, $|j: t_i \in d_j|$ contains the number of documents in the

word. If the word is not in the sample, it will cause the denominator to be zero. Therefore, adding 1 to the denominator is to avoid the situation that the denominator is 0.

$W_{i,j} = TF_{i,j} \times IDF_i$. Weight $\omega_{i,j}$ of words is obtained by multiplying the word frequency in the document by the low file frequency of the word in the whole document set.

According to the TF-IDF feature weight calculation method, the characteristics of turnout fault samples based on text are calculated. The characteristics of a turnout fault sample are expressed as $d_i = [\omega_i^1 \ \omega_i^2 \ ... \ \omega_i^m]$, m is the length of the sample, and the primary fault category and secondary fault category are expressed as matrix $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$ by one hot coding vectorization, $y_i = [0 \ 1 \ 0 \ ... \ c-1]$, c is the total number of categories, and the fault level I category feature is expressed as $D_{L1} = [d_i \ y_1]$, ($i=1,2,..., n$), n represents the total number of samples. The label of fault level I is also input into the feature vector by Fault secondary feature as a feature, $D_{L2} = [[d_i \ y_1] \ y_2]$.

3.3 Deep learning integrated fault diagnosis model

Integrated learning is to combine multiple weak supervised learning models to get a better and more comprehensive supervised learning model. The high-speed railway turnout fault diagnosis model adopts BiGRU and BiLSTM neural networks as the weak supervised learning model, inputs the feature vectors extracted from the features into the embedded layer of BiGRU and BiLSTM neural networks respectively, and the two neural networks output the classification and prediction probability of the feature vectors in the Softmax layer through learning. The prediction results of the two neural networks are integrated and calculated by the combined weighted integration method, and finally the classification results of the input data by the deep learning integration model are output [30].

GRU and LSTM are variants of RNN neural network. Gating units are designed in neurons to effectively calculate and control the input and output of information, as shown in Figure 3. The design of this gating unit solves the problem of text sequence length dependence. Since the output of sigmoid function is 0 ~ 1, 1 can mean that the information is retained, and 0 means that the information is discarded, GRU and LSTM process the input information through sigmoid function, and tanh function processes the output information.
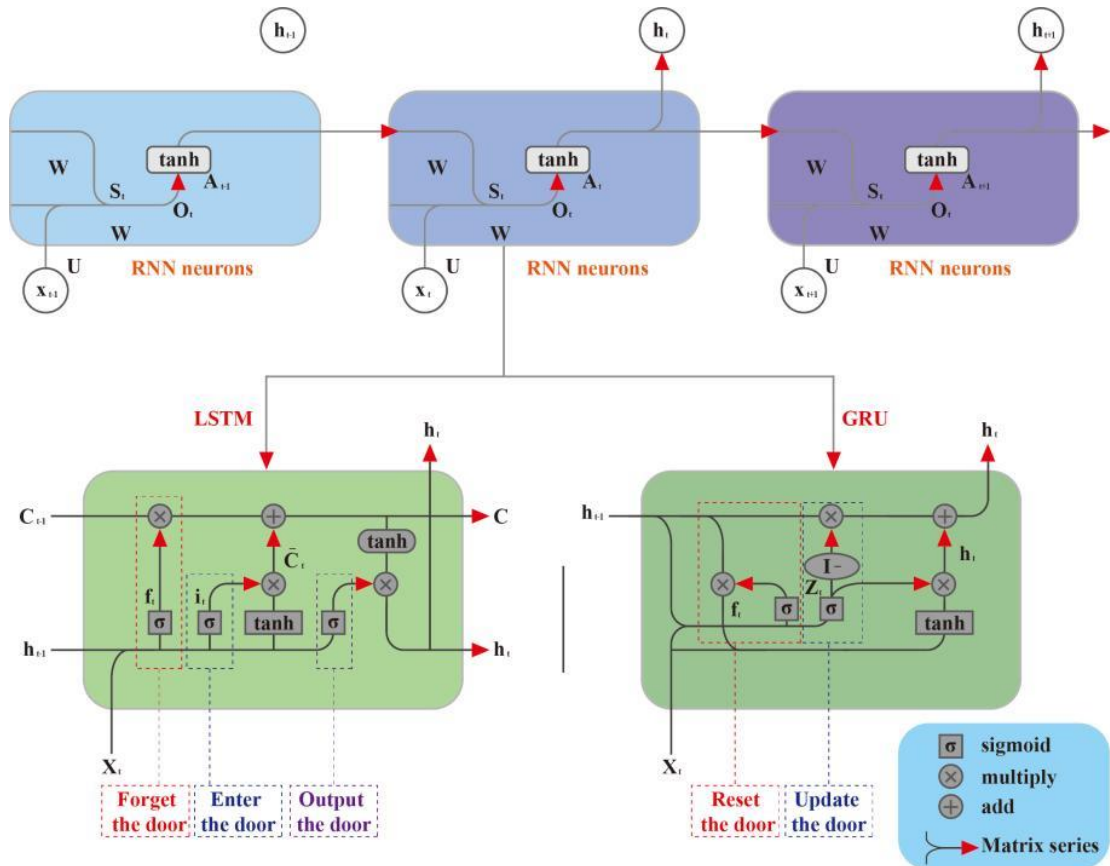
Figure 3: Structural units of RNN and its variant neurons

LSTM neural unit is composed of three gates, namely forgetting gate, input gate and output gate, as shown in Figure 3. LSTM first determines which information needs to be discarded through the forgetting gate, and calculates h_{t-1} x_i and output a vector between 0 and 1, the vector represents what information neuron C_{t-1} retains or discards. Then, the input gate is used to determine which information needs to be added in the neuron, and the candidate neuron $\widetilde{C}_t$ is obtained by tanh's calculation using h_{t-1} and x_i, which can be updated into the neuron. Finally, the output information is controlled by the output gate, and the LSTM neuron output is finally obtained by multiplying the 0 ~ 1 vector obtained by the output layer o_t and the neuron through the tanh layer.

$$f_t = \sigma(W_f \cdot [h_{t-1} \ x_i] + b_f) \quad (3)$$

$$i_t = \sigma(W_i \cdot [h_{t-1} \ x_i] + b_i) \quad (4)$$

$$\widetilde{C}_t = tanh(W_c \cdot [h_{t-1} \ x_i] + b_c) \quad (5)$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t \quad (6)$$

$$o_t = \sigma(W_o \cdot [h_{t-1} \ x_i] + b_o) \quad (7)$$

$$h_t = o_t * \tanh(C_t) \quad (8)$$

where: * is Hadamard product operator, which means multiplication of elements at the same position of the matrix.

GRU is a variant of LSTM, as shown in Figure 3. It combines the forgetting gate and input gate into an update gate z_t. z_t controls how much information needs to be forgotten from the previous hidden layer h_{t-1}, how much information needs to be added to the current hidden layer $\widetilde{h}_t$, and then obtains h_t. Reset gate r_t controls how much previous information needs to be retained. When r_t is 0, $\widetilde{h}_t$ only contains the information of the current word.

$$z_t = \sigma(W_z \cdot [h_{t-1} \ x_t]) \quad (9)$$

$$r_t = \sigma(W_r \cdot [h_{t-1} \ x_t]) \quad (10)$$

$$r_t = \tanh(W \cdot [r_t * h_{t-1} \ x_t]) \quad (11)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \widetilde{h}_t \quad (12)$$

The combination weighted integration method of LSTM and GRU combines the overall classification performance of a single neural network with the classification performance of each category by assigning weights. The combination weighted integration method includes overall weight and category weight. The higher

the overall classification performance of a single neural network, the higher the overall weight will be allocated. According to formula (13) and formula (14), the lower the error proportion of neural network in category classification, the better classification performance it has in this category, the higher the category weight will be allocated. Then add the overall weight of the neural network and the category weight according to equation (15) to recalculate the predicted value of the neural network in each category. This combined weighted integration method can avoid the influence of few values and extreme values in the integration method.

$$\epsilon_{ij} = \frac{error\ Num_{ij}}{text\ Num_{ij}} \tag{13}$$

$$\alpha_{ij} = \begin{cases} \ln\left(\frac{1-\epsilon_{ij}}{\epsilon_{ij}}\right) & \epsilon_{ij}<0.5 \\ 0 & \epsilon_{ij}\geq0.5 \end{cases} \tag{14}$$

$$P_i = \sum_{j=1}^{n} (\omega_j + \alpha_{ij}) \cdot P_{ij} \tag{15}$$

Where: ϵ_{ij} is the classification error ratio of neural network j in category i. $text\ Num_{ij}$ is the total number of samples of category i; $error\ Num_{ij}$ is the number of classification error samples of neural network j in category i. α_{ij} is the category weight of neural network j in category i., ω_j is the overall weight of neural network j, and $\sum_{j=1}^{n} \omega_j$.

In order to improve the generalization ability of deep learning integration model, K-fold cross validation training model is adopted. K-fold cross validation is to randomly divide the whole training sample into K parts, one of which is used as the validation set and the other K-1 is used as the training set, and cycle K times until all data are selected once.

4   Results and Analysis

This section illustrates the result and analysis of overall weight distribution, weight calculation and the classification of deep learning integration model.

4.1   Overall weight distribution of BiGRU and BiLSTM

BiGRU and BiLSTM have the same network parameters, in which the embedded layer dimension is 100, the hidden layer dimension is 512, K-fold cross validation K = 4, the number of iterations is 50, and the batch size is 256. After TF-IDF feature extraction and vector representation, the training set and verification set synthesized by ADASYN are input into BiGRU and BiLSTM networks for training. The change of *loss* function value in the training process of the two neural networks is shown in Figure 4. It can be seen from Figure 4 that with the increase of iteration times, the *loss* value of BiGRU is lower than that of BiLSTM, indicating that its overall classification performance is

better. In the two neural networks, the *loss* function value of the primary classification is lower than that of the secondary classification, indicating that the evaluation index of the primary classification of the neural network is higher than that of the secondary classification. Both neural networks are between 40 ~ 50 iteration rounds, and the *loss* function value tends to be stable, indicating that the number of iteration rounds of 50 can make the neural network training reach the best state.
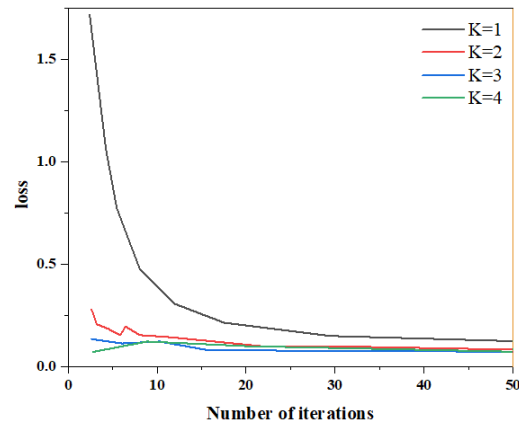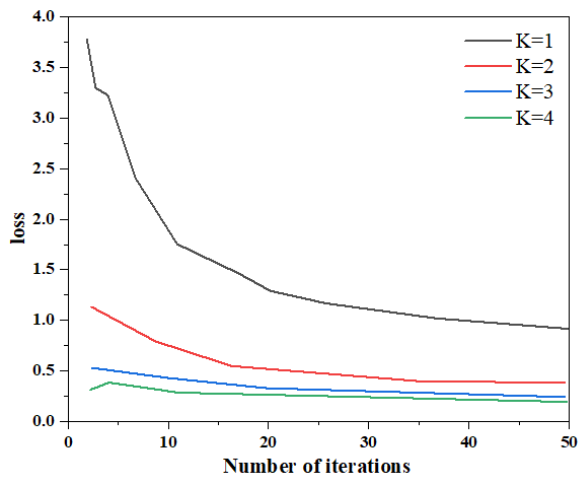

(a): BiGRU primary classification training process
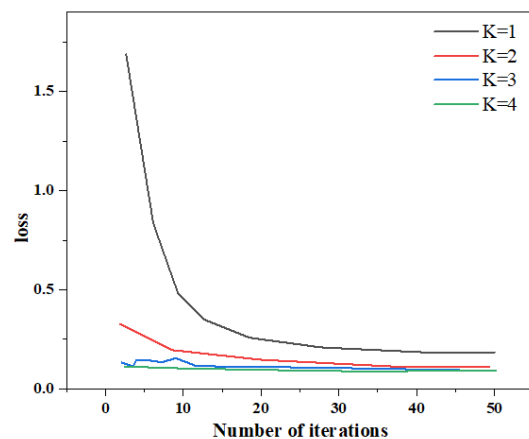
(b): BiGRU primary classification training process
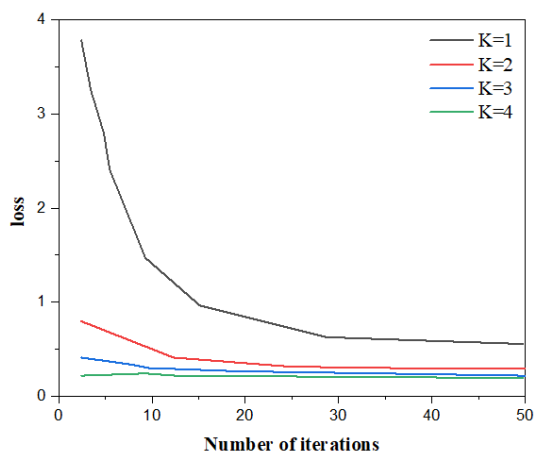
(c): BiGRU secondary classification training process

(d): BiGRU primary classification training process

Figure 4 (a, b, c, d): Variation of loss value in K-cross training of BiGRU and BiLSTM neural networks

After K = 4 training, 30% real samples are used to evaluate BiGRU and BiLSTM training models. The evaluation results are shown in Table 1 and is graphically presented in Figure 5.

Method	Level	Accuracy rate	Recall rate	F1 value
ADASYN ＋BiGRU	Primary fault classification	0.8742	0.8814	0.8779
	Secondary fault classification	0.7828	0.7421	0.7619
ADASYN ＋BiLSTM	Primary fault classification	0.8613	0.8765	0.8688
	Secondary fault classification	0.7601	0.7581	0.7591
BiGRU	Primary fault classification	0.7317	0.7098	0.7206
	Secondary fault classification	0.7081	0.6712	0.6891
BiLSTM	Primary fault classification	0.6912	0.7129	0.7019
	Secondary fault classification	0.6371	0.6214	0.6292

Table 1: Test results of K-fold cross validation + BiGRU and BiLSTM neural network
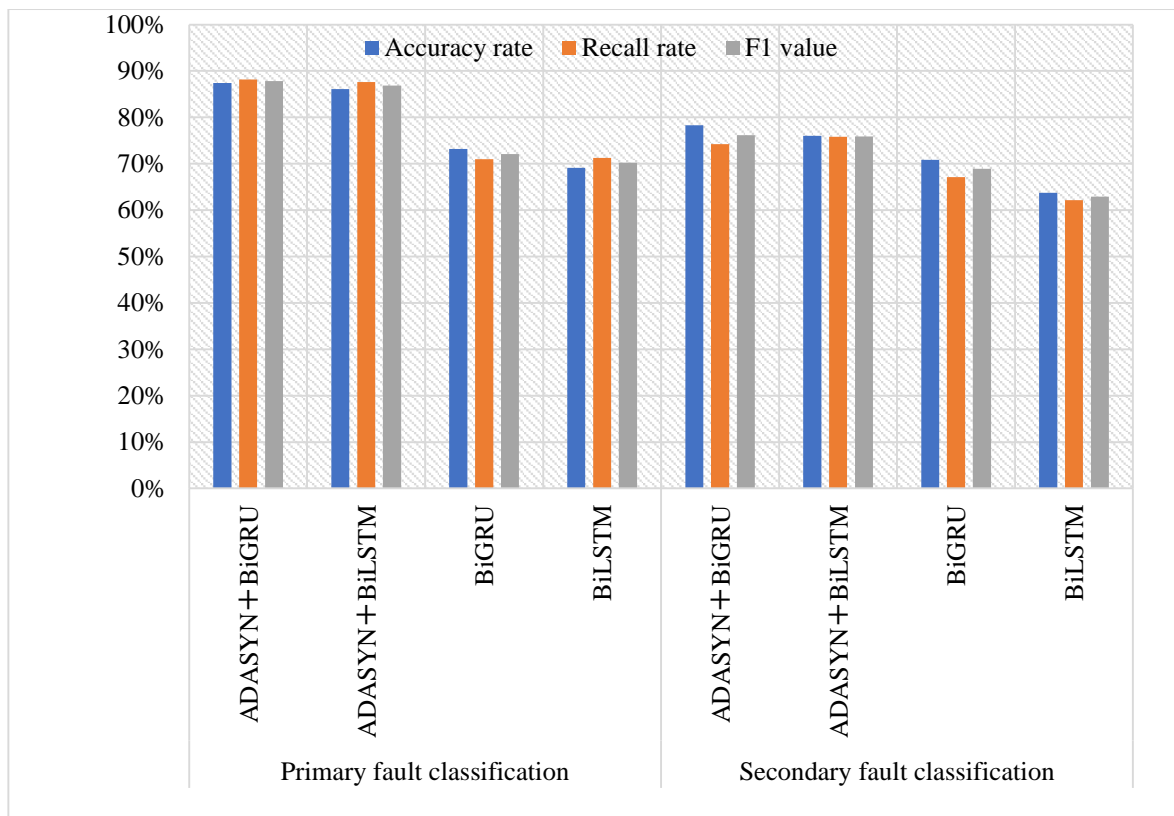
Figure 5: Graphical results of K-fold cross validation + BiGRU and BiLSTM neural network

It can be seen from Table 1 that after using ADASYN less category synthesis method, the evaluation indexes of BiGRU network are higher than BiLSTM network under the same parameters, so BiGRU network should be assigned a higher overall weight. The original samples are trained with the same network structure and parameters. The test results are shown in Table 1. It can be seen that after ADASYN synthesizes a small number of samples, the classification indexes of the two neural networks are significantly improved, the first-class rating indexes of BiGRU network with good performance are increased by nearly 15%, and all evaluation indexes of BiGRU network are higher than those of BiLSTM network. It is further concluded that the performance of BiGRU is better than BiLSTM, and a higher overall weight can be assigned to BiGRU network.

4.2 Weight calculation of BiGRU and BiLSTM

In order to more comprehensively obtain the performance of neural network in each category classification, a few category samples synthesized by ADASYN and all real samples are used. A total of 6327 samples are input into the trained ADASYN + BiGRU and ADASYN + BiLSTM neural networks. The category weight calculation results of the two neural networks in the primary classification are shown in Table 2.

It can be seen from Table 2 that although BiGRU has higher overall evaluation index and higher overall weight than BiLSTM, the performance of the two neural

networks are different in each category. BiLSTM has a larger category weight in the categories of security inspector, public works equipment and unknown reason, indicating that BiLSTM network has decision-making power in these three categories. Due to the large number of secondary classification categories of signal turnout equipment faults, considering the length, this paper only lists the weight calculation results of primary classification categories.

4.3 Deep learning integration model and classification

The various weights of the neural network are obtained through the above tests, and BiGRU should have higher overall weight than BiLSTM. Different overall weights are given to BiGRU and BiLSTM. The two deep learning neural networks are integrated through combined weighting, and the common classification prediction results are obtained through recalculation of the outputs of the two networks.

Under different overall weight distribution, see Figure 6 for the evaluation indexes of level 1 fault classification and level 2 fault classification of the deep learning integration model (where G represents BiGRU and L represents BiLSTM). It can be seen from Figure 6 that when the overall weight of BiGRU is 0.54 and the overall weight of BiLSTM is 0.46, the evaluation index of the deep learning integration model is the highest. The

final classification results of the deep learning integration
model are shown in Table 3 and Figure 7.

Classification	Classification method	Number of classification errors / total number of categories	Recall rate	Category weight
Switch machine	ADASYN＋BiGRU	266/2053	0.1295	1.9048
	ADASYN＋BiLSTM	288/2053	0.1403	1.8129
External locking and installation device	ADASYN＋BiGRU	163/1251	0.1303	1.8983
	ADASYN＋BiLSTM	192/1251	0.1534	1.7076
Paste checker	ADASYN＋BiGRU	81/567	0.1428	1.7918
	ADASYN＋BiLSTM	70/567	0.1235	1.9601
Turnout control circuit equipment	ADASYN＋BiGRU	167/1280	0.1305	1.8968
	ADASYN＋BiLSTM	189/1280	0.1477	1.7531
Permanent way equipment	ADASYN＋BiGRU	62/440	0.1409	1.8077
	ADASYN＋BiLSTM	55/440	0.1250	1.9459
Supporting equipment	ADASYN＋BiGRU	86/614	0.1401	1.8147
	ADASYN＋BiLSTM	80/614	0.1303	1.8984
Unknown reason	ADASYN＋BiGRU	21/124	0.1694	1.5902
	ADASYN＋BiLSTM	14/124	0.1129	2.0614

Table 2: Calculation results of class I classification weight of signal turnout equipment fault
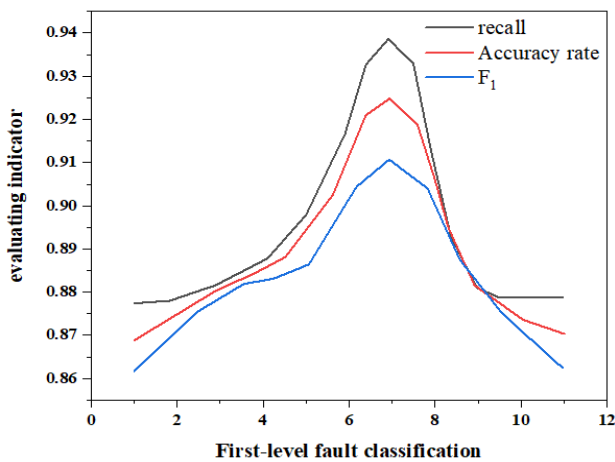
(a): First-level fault classification
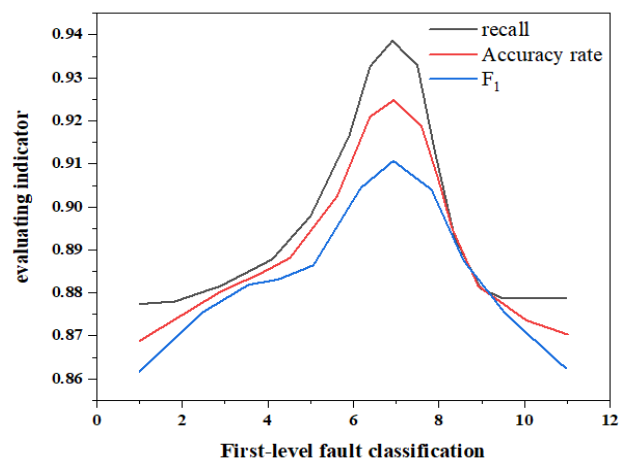
(b): BiLSTM secondary classification training process

Figure 6: Evaluation index values of deep learning integration model under different overall weight distribution

Method	Level	Accuracy rate	Recall rate	F1 value
Deep learning integration model	Primary fault classification	0.9106	0.9389	0.9245
	Secondary fault classification	0.8564	0.8612	0.8588

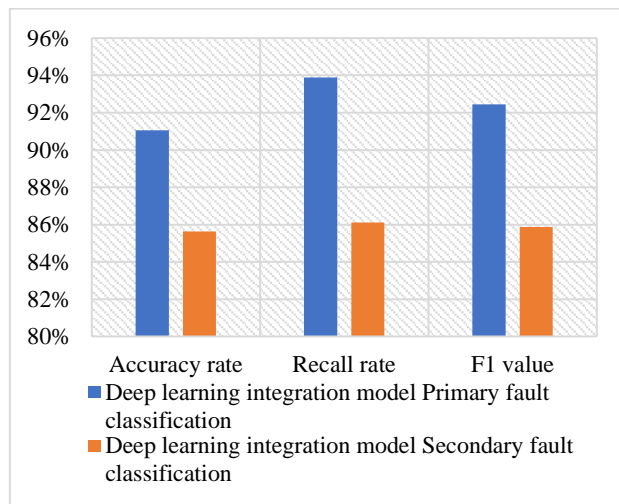Table 3: Classification test results of deep learning integration model

Figure 7: Graphical representation of classification test results of deep learning integration model

It can be seen from Table 3 and Figure 7 that compared with ADASYN + BiGRU neural network, the comprehensive evaluation index of primary fault classification is improved by about 5%, and the comprehensive evaluation index of secondary fault classification is improved by about 9%. Compared with ADASYN + BiLSTM neural network, the comprehensive evaluation index of primary fault classification is improved by about 6%, and the comprehensive evaluation index of secondary fault classification is improved by about 10%.

5 Conclusions

This paper studies the fault diagnosis model of signal turnout fault text data, uses ADASYN data synthesis method to synthesize few category samples. This article also uses TF-IDF to extract features and transform vectors, and puts forward a deep learning integration method based on combination weight. The sample synthesis algorithm can appropriately synthesize a few categories of samples according to the distribution of the overall samples. There are several articles which uses SVM-SMOTE method to automatically synthesize the few category samples of signal equipment fault, and solve the problem of signal equipment fault sample imbalance. Through experimental analysis, it is proved that deep learning integration is a method that can effectively improve the classification performance of turnout fault diagnosis model. At the same time, this method can also provide a new idea for railway text classification and fault diagnosis. This article utilizes the deep learning algorithm of artificial intelligence for investigating the interlocking faults in the railway transportation. This paper uses ADASYN data synthesis method to synthesize few category samples, uses TF-IDF to extract features and transform vectors, and proposes a deep learning integration method based on combined weight. The outcomes obtained for the proposed method reveals that BiGRU has better overall classification performance when evaluated on the index of primary and secondary fault classification accuracy.

References

[1] Kong, J. (2020, December). Application and research of artificial intelligence in digital library. In *International conference on Big Data Analytics for Cyber-Physical-Systems* (pp. 318-325). Springer, Singapore.
https://doi.org/10.1007/978-981-33-4572-0_47

[2] Paek, S., & Kim, N. (2021). Analysis of worldwide research trends on the impact of artificial intelligence in education. *Sustainability*, *13*(14), 7941.
https://doi.org/10.3390/su13147941

[3] Dobias, R., & Kubatova, H. (2004, August). FPGA based design of the railway's interlocking equipments. In *Euromicro Symposium on Digital System Design, 2004. DSD 2004.* (pp. 467-473). IEEE.
10.1109/DSD.2004.1333312

[4] Skiribou, C., Elbahhar, F., & Elassali, R. (2021). DMRS-based channel estimation for railway communications in tunnel environments. *Vehicular Communications*, *29*, 100340.
https://doi.org/10.1016/j.vehcom.2021.100340

[5] Kiedrowski, P., & Saganowski, Ł. (2021). Method of Assessing the Efficiency of Electrical Power Circuit Separation with the Power Line Communication for Railway Signs Monitoring. *Transport and Telecommunication*, *22*(4), 407-416.
10.2478/ttj-2021-0031

[6] Yang, J., Bai, X., Zhang, Z., Yang, M., Pan, P., Liu, T., & Tao, T. (2021, May). Research on the application of BDS/GIS/RS technology in the high speed railway infrastructure maintenance. In *IOP Conference Series: Earth and Environmental Science* (Vol. 783, No. 1, p. 012168). IOP Publishing.
10.1088/1755-1315/783/1/012168

[7] Lin, J., Hu, X., Dang, J., & Wu, Z. (2019). Traffic model of machine-type communication for railway signal equipment based on MMPP. *IET Microwaves, Antennas & Propagation*, *13*(8), 1072-1079.
https://doi.org/10.1049/iet-map.2018.6004

[8] Wang, X., Guo, J., Jiang, L., Fu, J., & Li, B. (2016, August). Intelligent fault diagnosis and prediction technologies for condition based maintenance of track circuit. In *2016 IEEE International Conference on Intelligent Rail Transportation (ICIRT)* (pp. 276-283). IEEE.
10.1109/ICIRT.2016.7588745

[9] Sahal, R., Breslin, J. G., & Ali, M. I. (2020). Big data and stream processing platforms for Industry 4.0 requirements mapping for a predictive maintenance use case. *Journal of manufacturing systems*, *54*, 138-151.
https://doi.org/10.1016/j.jmsy.2019.11.004

[10] Cao, Y., Li, P., & Zhang, Y. (2018). Parallel processing algorithm for railway signal fault diagnosis data based on cloud computing. *Future Generation Computer Systems*, *88*, 279-283.
https://doi.org/10.1016/j.future.2018.05.038

[11] Dong, C. Z., Ye, X. W., & Jin, T. (2018). Identification of structural dynamic characteristics based on machine vision technology. *Measurement*, *126*, 405-416.
https://doi.org/10.1016/j.measurement.2017.09.043

[12] Jia, Z., & Sharma, A. (2021). Review on engine vibration fault analysis based on data mining. *Journal of Vibroengineering*, *23*(6), 1433-1445.
https://doi.org/10.21595/jve.2021.21928

[13] Yin, M., Li, K., & Cheng, X. (2020). A review on artificial intelligence in high-speed rail. *Transportation Safety and Environment*, *2*(4), 247-259.
https://doi.org/10.1093/tse/tdaa022

[14] Ren, X., Li, C., Ma, X., Chen, F., Wang, H., Sharma, A., & Masud, M. (2021). Design of multi-information fusion based intelligent electrical fire detection system for green buildings. *Sustainability*, *13*(6), 3405.
https://doi.org/10.3390/su13063405

[15] Sharma, D., Kaur, R., Sandhir, M., & Sharma, H. (2021). Finite element method for stress and strain analysis of FGM hollow cylinder under effect of temperature profiles and inhomogeneity parameter. *Nonlinear Engineering*, *10*(1), 477-487.
https://doi.org/10.1515/nleng-2021-0039

[16] Afandizadeh, S., & Rad, H. B. (2021). Developing a model to determine the number of vehicles lane changing on freeways by Brownian motion method. Nonlinear Engineering, 10(1), 450-460.
https://doi.org/10.1515/nleng-2021-0036

[17] Shabaz, M., Sharma, A., Al Ajrawi, S., & Estrela, V. V. (2022). Multimedia-based emerging technologies and data analytics for Neuroscience as a Service (NaaS). Neuroscience Informatics, 2(3), 100067.
https://doi.org/10.1016/j.neuri.2022.100067

[18] Meher, M., & Rostamy, D. (2021). Hybrid of differential quadrature and sub-gradients methods for solving the system of Eikonal equations. Nonlinear Engineering, 10(1), 436-449.
https://doi.org/10.1515/nleng-2021-0035

[19] Mi, Z., Wang, T., Sun, Z., & Kumar, R. (2021). Vibration signal diagnosis and analysis of rotating machine by utilizing cloud computing. Nonlinear Engineering, 10(1), 404-413.
https://doi.org/10.1515/nleng-2021-0032

[20] Wang, H., Sharma, A., & Shabaz, M. (2022). Research on digital media animation control technology based on recurrent neural network using speech technology. International Journal of System Assurance Engineering and Management, 13(1), 564-575.
https://doi.org/10.1007/s13198-021-01540-x

[21] Yousaf, B., Qaisrani, M. A., Khan, M. I., Sahar, M. S. U., & Tahir, W. (2021). Numerical and

experimental analysis of the cavitation and study of flow characteristics in ball valve. Nonlinear Engineering, 10(1), 535-545. *https://doi.org/10.1515/nleng-2021-0044*

[22] Singh, P. K., & Sharma, A. (2022). An intelligent WSN-UAV-based IoT framework for precision agriculture application. *Computers and Electrical Engineering*, *100*, 107912. *https://doi.org/10.1016/j.compeleceng.2022.107912*

[23] Zeng, H., Dhiman, G., Sharma, A., Sharma, A., & Tselykh, A. (2021). An IoT and Blockchain-based approach for the smart water management system in agriculture. *Expert Systems*, e12892. *https://doi.org/10.1111/exsy.12892*

[24] Sharma, A., & Singh, P. K. (2021). UAV-based framework for effective data analysis of forest fire detection using 5G networks: An effective approach towards smart cities solutions. *International Journal of Communication Systems*, e4826. *https://doi.org/10.1002/dac.4826*

[25] Sharma, A., Singh, P. K., & Kumar, Y. (2020). An integrated fire detection system using IoT and image processing technique for smart cities. *Sustainable Cities and Society*, *61*, 102332. *https://doi.org/10.1016/j.scs.2020.102332*

[26] Zang, Y., Shangguan, W., Cai, B., Wang, H., & Pecht, M. G. (2019). Methods for fault diagnosis of high-speed railways: A review. Proceedings of the institution of mechanical engineers, part O: journal of risk and reliability, 233(5), 908-922. *https://doi.org/10.1177/1748006X18823932*

[27] Ting, L., Khan, M., Sharma, A., & Ansari, M. D. (2022). A secure framework for IoT-based smart climate agriculture system: Toward blockchain and edge computing. Journal of Intelligent Systems, 31(1), 221-236. *https://doi.org/10.1515/jisys-2022-0012*

[28] Minea, M., Dumitrescu, C. M., & Dima, M. (2021). Robotic Railway Multi-Sensing and Profiling Unit Based on Artificial Intelligence and Data Fusion. Sensors, 21(20), 6876. *https://doi.org/10.3390/s21206876*

[29] Luo, J., Wu, M., Gopukumar, D., & Zhao, Y. (2016). Big data application in biomedical research and health care: a literature review. *Biomedical informatics insights*, *8*, BII-S31559. *https://doi.org/10.4137/BII.S31559*

[30] Chen, F., Deng, P., Wan, J., Zhang, D., Vasilakos, A. V., & Rong, X. (2015). Data mining for the internet of things: literature review and challenges. *International Journal of Distributed Sensor Networks*, *11*(8), 431047. *https://doi.org/10.1155/2015/431047*

Design and Implementation of a New Intelligent Warehouse Management System Based on MySQL Database Technology

Ying Zhang *[1], Feng Pan [2]
YingZhang78@126.com, FengPan278@163.com
*Corresponding Author

[1]Changchun Polytechnic, Career Foundation Department, Jilin-Changchun, 130000, China
[2]Continental Automotive Corporation Changchun Co., LTD. Moonlake Branch, IT Department, Jilin-Changchun,130000, China

The handling industry of materials/goods is fundamental for companies for ensuring the warehouses smooth running. Efficiency within every aspect of the business is essential to gain a competitive advantage. In order to improve the material management level of enterprises based on MySQL database technology; this paper makes an overall design of the warehouse management system, builds a MySQL database, and realizes the design and application of a new intelligent warehouse management system. Through the operation and test of the system, the results show that the system mainly realizes the five necessary functional modules of warehouse management: basic information management, system management, procurement management, warehousing management and inventory management. In the test, the system runs normally, the unit test and integration test can meet the expected requirements, realize the functions required by the user, and get the desired results within the user's acceptable response time (within 3S). Whether the system is running on the local machine or on the real server in the network, it must use the appropriate hardware and software conditions. It can provide automatic and comprehensive records for the whole process of material management of the enterprise, and provide real-time and correct information for all warehouse activities, resources and inventory levels

Povzetek: Za MySQL je bil razvit nov inteligentni upravljalski sistem.

1 Introduction

Unnecessary labor costs and the incorrect use of storage systems and racking arrangements result are caused by the disorganized warehouse spaces in many companies and find their warehouse shelves full, with no space to receive new inventory. When inventory location is not organized and easily available, pickers will take longer to find items that need to be shipped. With the progress of the times and the continuous updating of technology, society has entered the era of big data with rapid development and informatization. High-end technologies and concepts such as big data, Internet of things and cloud storage have been applied to real life and work. Warehouse management system combines management science, computer science and other sciences [1]. With the progress of the times and the continuous renewal of technology, warehouse management system plays a very important role in the development of enterprises. It can help enterprise managers make correct decisions and predict the development direction of enterprises. The content of warehouse management is very rich. For example, it includes the layout and design of warehouse system, high-quality inventory management and efficient warehouse operation. The above-mentioned contents

complement each other. The production capacity and level of most Chinese enterprises lag behind the same type of foreign enterprises. In addition to the advanced technology and excellent talents of foreign enterprises, the information integration degree of domestic enterprises is not high and the operation efficiency is generally low, resulting in low profitability and even lower ability to resist market risks than foreign enterprises. In particular, China's warehousing management level is inefficient, the utilization rate of warehousing resources is not high, the operation conditions are poor, and it lacks its own development ability [2]. Like other management, enterprises need to develop towards specialization, specialization, functionalization and personalization. Most foreign enterprises have a good level of warehouse information management, including account processing and settlement processing, and providing real-time query; Location management, making documents and reports, stock control, etc. The efficient warehouse management of foreign enterprises is based on the effective control and organization of materials. Foreign enterprises have focused on the establishment of effective information networks for warehouses, manufacturers, material managers, material demanders, material descriptions and other contents, so as to realize the sharing of warehouse

information, and realize the networked and intelligent management of warehouse information through information network control [3].

This paper mainly introduces the technical research of warehouse management system. Firstly, the business process of warehouse management is studied and designed for analysis and refinement, which involves administrator login, purchase warehousing, standby transfer and scrap warehousing, outbound and inbound statistics. The specific implementation process of the functional modules such as purchase warehousing, material warehousing, material processing and query in the system is carried out. Finally, the SQL database background and the system use eclipse are realized. The test and analysis of the warehouse management system is mainly the specific analysis and description of the function test of each system module. At the same time, according to the test results, this paper deeply analyzes and studies the functional performance of the warehouse management system, and makes improvement suggestions.

(1) Plan the functional modules of the warehouse material management system

First of all, understand the relevant work tasks of each department involved in material management in the enterprise, and plan several modules required by the system, such as purchase warehousing, material warehousing, material processing, query statistics, basic material information, system management, etc. (as shown in Figure 1).
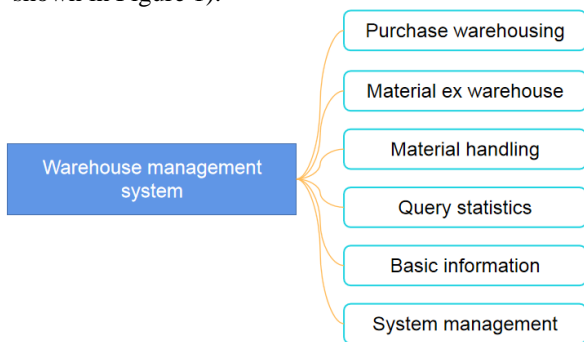

Figure 1: Schematic diagram of each module of warehouse management system

(2) Sort out the specific business and overall workflow of each module of the warehouse material management system

Based on the warehouse material management system, each module of the system has its own different business, and the business of each module also has contact and certain order. Determine the specific business of each module and the relationship between each module. Based on B / S architecture, with my SQL as the background database platform and my eclipse as the development tool, the functional design and implementation of warehouse management system are completed based on struts 2, hibernate and spring

framework. At the same time, boost is used to beautify the front-end page.

Contribution: This paper makes an overall design of the warehouse management system, builds a MySQL database, and realizes the design and application of a new intelligent warehouse management system in order to improve the material management level of enterprises, based on MySQL database technology.

The organization of the paper is as follows. Section 2 provides an overview of the exhaustive literature survey followed by a methodology adopted in section 3. A detailed discussion of obtained results is in section 4. Finally, Section 5 concludes the paper.

2   Literature review

With regard to the development and application of the Internet of things, J Liang studied the construction and key points of the storage system architecture based on the Internet of things environment, and conducted simulation research [4]. Zhao, J. studied the development and application of intelligent storage information system based on Internet of things technology, expounded the technical difficulties and doubts of system development, Zhong Yuangen studied the construction and smart design of mobile electronic vending public service platform based on Internet of things technology, and simulated the construction of simulated public service platform [5]. Zhao, K. studied the construction process of digital warehouse software architecture based on Internet of things technology, highlighting the characteristics of digitization [6]. Viloria, A. studied the development process of dangerous goods intelligent logistics system based on Internet of things, so that the transportation of dangerous goods can be monitored and handled in real time [7]. Zhang, Y. studied the design process of automatic cold storage management system based on Internet of things technology [8].

With regard to the development and application of intelligent warehousing, kermani, M. studied the application of intelligent warehousing based on WLAN and RFID, and proposed a combined system using wireless RF technology and wireless LAN technology [9]. Yu, S. studied the design of intelligent storage node based on ZigBee wireless sensor network, which solved the shortcomings of strong manual dependence and low automation level of traditional warehouse management [10]. Somasundaram, M. studied RHD middleware for the h-party intelligent warehousing, expounded RFID middleware and related specifications, described the application status and problems of RFID Middleware for the h-party warehousing, and explored solutions [11]. Nastasi, G. designed the intelligent warehouse management system, expounded the key technologies of swms system, and formulated the design scheme [12]. Kumar, R. S. studied the application of intelligent warehousing in modern logistics, expounded the current

situation of warehousing management, and put forward suggestions and methods for constructing intelligent warehousing system [13]. SHARIFI, H. studied the debugging problem of intelligent storage system, fully expounded the existing reasons and gave the corresponding solutions [14].

In the development of the Internet of things in intelligent storage, Xu, Z. studied the design of RFID based storage management information system, expounded the relevant theories of storage management information system, analyzed the requirements of RFID storage management information system, carried out the overall design and detailed design of RFID storage management information system, and finally carried out the simulation implementation [15]. Mo, Z. studied the upgrading of RFID Middleware in warehouse management Internet of things system, and discussed the concept, characteristics, infrastructure and application functions of RFID Middleware in Internet of things system [16]. Ad, A. studied the warehouse management system based on RFID technology, explored and improved the RFID anti-collision algorithm, and carried out RFID optimized inventory management. On this basis, he developed and implemented the warehouse management information system [17]. Many researchers have worked in this field in the previous years, some of the relevant articles are tabulated in Table 1.

Authors	Presented Work	Key points	Benefits	References
Someah Alangari et al., 2021	This paper present and analyze system that will be intelligent enough to help the organization users to manage their inventory that will be helpful enough for providing information as well as providing various amazing heuristic methods that will be helpful enough for the system content management.	The system prediction power is useful for many inventories and the power provides the notifications in advance to manage the system's components.	Users are able to access or request a particular object from the inventory. Manager handles all the entries inside the system.	[18]
R A Darajatun et al., 2017	Paper presents the design and development of Kanban of inventory storage and delivery system.	The author uses Java programming language for the application development used for building Java Web applications, while the database used is MySQL.	Goods are monitored and warehouse is divided into many locations.	[19]
Walaa hamdy et al.,	This paper proposed a framework for implementing the technology in a warehouse.	The presented work help in achieving more monitoring on the operations in the warehouse in real time.	Increased speed and efficiency. It prevents counterfeiting and inventory shortage.	[20]

| Reza Pulungan et al., 2013 | The state-of-the-art results are integrated in the field of intelligent systems—neural network, bee colony optimization, fuzzy control, and decision support system—together with the latest echnologies—RFID and Android-based handheld devices—in every part of business processes in WMS. | Discussions on the practical implementation of AI in the main WMS processes are provided. | Highly effective. | [21] |
| Jia Mao et al., 2018 | Effective scheduling method is presented and initially realizes the intelligent warehouse management system based on cloud model. | To integrate the resources effectively, a variety of automation, intelligence and information technology are utilized and discussed | Better scheduling solution and the certain robustness | [22] |

Table 1: Some relevant state of the art work in previous years

3   Research methods

3.1   Overall design of warehouse management system

The structure of the system adopts B / S architecture. All business processing logic is executed on the server. The client has only a browser (fire fox / Chrome / 360 / Sogou, etc.), and all interface presentation / operations send data to the server through the browser, which is processed by the corresponding module of the server, as shown in Figure 2.
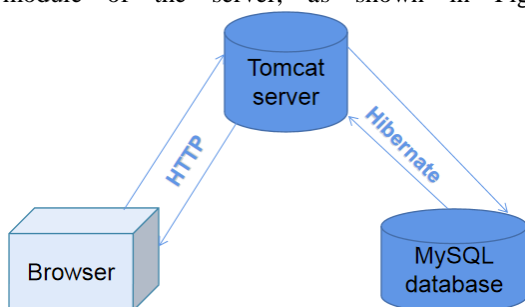

Figure 2: System B / S architecture

The system adopts a three-tier model to realize the client / server mode. The three-tier structure model of the system takes accessing the web database as the center, HTTP as the transmission protocol, and the client accesses the web server and its connected background database through the browser. The composition of its three-tier structure is shown in Figure 3.
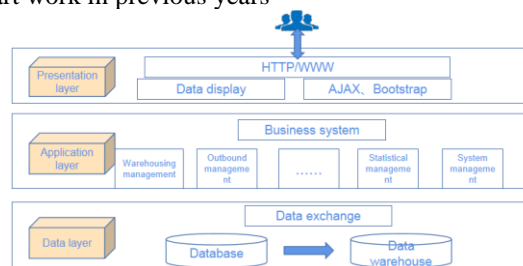

Figure 3: Three layers structure model

The first layer is the user interface layer, which is mainly responsible for user interaction processing and the interaction between the client and the background. When the user clicks a button in the page to trigger an event, the client sends a request to the background. This process may be synchronous or asynchronous through Ajax. The second layer is the business logic processing layer. When the client sends a request through a pre-defined interface, it parses the request according to the rules of the interface agreement, then processes the corresponding request, and finally returns to the client. The third layer is the data support layer, where the information records sent by the client are saved in the database through MySQL database, such as warehousing records, outbound records, etc.

3.2   MySQL database construction

My SQL is a small relational database management system. At present, my SQL is widely used in small and medium-sized websites on the Internet. Due to its small size, fast speed and low total cost of ownership, especially open source, many small and medium-sized

websites choose my SQL as the website database in order to reduce the total cost of ownership [23]. My SQL has the following features:

1. It is written in C and C + +, and tested with a variety of compilers to ensure the portability of the source code.

2. Support AIX, free BSD, HP-UX, Linux, Mac OS, Novell Netware, open BSD, OS / 2 wrap, Solaris, windows and other operating systems.

3. Provides API for multiple programming languages. These programming languages include C, C + +, python, Java, Perl, PHP, Eiffel, ruby and TCL.

4. Support multithreading and make full use of CPU resources.

5. The optimized SQL query algorithm can effectively improve the query speed.

6. It provides TCP / IP, ODBC, JDBC and other database connection channels.

7. Provides management tools for managing, checking, and optimizing database operations.

8. It can handle large databases with tens of millions of records.

First, download the appropriate version of the installation package on the official website of my SQL (we choose mysql-5.6.23-winx64). After installation, in order to use it directly under the console, you need to add the bin directory to the environment variable path. Finally, you need to add the file my.ini under the installation path and set the values of basedir and dataDir as the values corresponding to your installation directory.

This system belongs to Java website development and needs to run on a server that can run Java programs. Tomcat server is selected for the operation of this system. First, go to the official Tomcat website to download the installation package of the corresponding version. After downloading, unzip it to a path and then it can be used normally. The Tomcat version we selected is apache-tomcat-7.0.70. Go to bin / and click Startup bat to start the server directly. Click shutdown.bat to shut down the running server. The system needs to be debugged frequently during eclipse development. In order to facilitate future development and debugging, you need to configure eclipse to directly start the installed Tomcat server.

In Eclipse, click window - > preference - > server - > runtime environment, and then click Add to add. After adding, click Edit to modify the Tomcat path and the Tomcat server path.

3.3 Implementation of warehouse management system

The main purpose of the basic information management sub module is to realize the relatively static basic information management and maintenance of the logistics management system. The basic information mainly includes the information of logistics company staff, cooperative units, commodities and warehouses.

The mechanism of staff information management is shown in Figure 4.
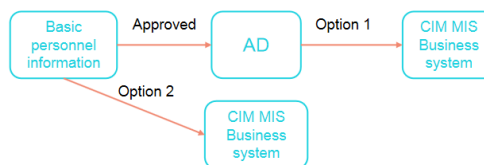

Figure 4: Mechanism of staff information management

As a warehouse management system, it is necessary to manage the necessary basic information to ensure that the subsequent operations can be met. The system can add, modify and delete materials, departments, construction groups and reservoir areas. For materials, you need to automatically generate IDs to meet your needs. Purchase order No., receipt Order No. and issue order No. in subsequent systems need to be generated automatically. The system generates ID through the database storage process. Taking the material table as an example, the database code of the storage process generate_WZID for generating material number is as follows:

```
BEGIN
  #Use WZ + year + 4-digit serial number as material number
  DECLARE current Date var CHAR (4); #current date
  #The last 5 digits of the serial number of the nearest qualified material number
  DECLARE max No INT DEFAULT 0;
  DECLARE newid VARCHAR (25);  #New item number
  SELECT DATE_FORMAT(NOW(),'%Y') INTO current Date; #4-digit year
  #Get the maximum ID number from the material list
  SELECT IFNULL(id,'') INTO old Order No FROM material
  WHERE SUBSTRING(id , 3 , 4) = current Date AND SUBSTRING(id,1,2) =
  'WZ' AND LENGTH(id) = 10 ORDER BY id DESC LIMIT 1;
  IF old Order No != '' THEN
    SET max No = CONVERT(SUBSTRING(old Order No,-4),DECIMAL);
  END IF;
  #Splice the new ID number into the newid
  SELECT              CONCAT('WZ',current Date,LPAD((max No+1),4,'0')) INTO newid;
  SELECT newid;
END
```

In Hibernate, the latest ID number can be generated by calling the stored procedure with the following code:

```
SQLQuery query = get Session().create SQLQuery("{call generate_WZID()}");
String id = (String) query.unique Result();
```

4   Results and discussion

4.1   System operation

Whether the system is running on the local machine or on the real server in the network, it must use the appropriate hardware and software conditions. Only the appropriate operating environment can ensure the normal operation of the system. Otherwise, problems such as bugs or poor system operation will occur due to configuration problems during testing or actual operation [24-26]. Table 2 shows the hardware and software configurations currently used by the system.

Server side		
Hardware configuration	**CPU**	**Intel i5 4750**
	Memory	4G
	Hard disk capacity	500G mechanical hard disk
	Network card	Gigabit Ethernet
Software configuration	Web server	Tomcat v9.0
	Database	My SQLv5.6.30
	JDK version	Java v8.0.7
Client		
Hardware configuration	Traditional PC, smooth Internet access	
Software configuration	Mainstream browsers, such as chrome, Firefox, Sogou browser, etc	

Table 2: System operation hardware configuration

After the development of the system, in order to be truly put into the production environment for users to use, the system must be published to a real server in the network. Users can directly enter the address of the server in the browser to log in to the system. To deploy the project to a real server, you need to package the project into a war package and put it into the server. Right click the project name, click export - > war file, select parameters and click Finish to generate war package in corresponding directory.

The project can be run through the above two methods. Enter "http: / / server address / warehouse /" in the browser to enter the login interface, as shown in Figure 5.
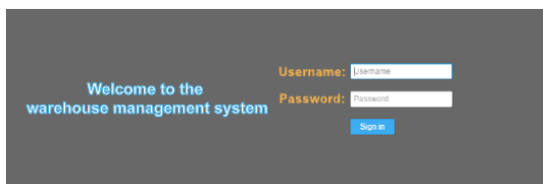

Figure 5: System login interface

4.2   System test

Unit testing is the smallest test method. This method tests a method or code block to find out whether the method or code block can complete the correct task [27, 28]. Because unit testing must fully understand the details of internal code design, it is most common for system developers rather than testers to complete this test. The system needs to conduct the corresponding unit test after the coding of a method or code segment to find problems. Due to the single similarity of the method of unit test, an example will be given below to introduce how the system uses unit test to complete the test work.

Because the system uses SSH framework and spring dependency injection to manage the creation of class objects in the system, it is difficult to use ordinary unit testing. Unit testing needs to be implemented through the unit test package provided by spring, and JUnit's jar package needs to be introduced for testing. Right click - > run as - > JUnit test to execute the test. After verification, the test is successful.

Through the unit test of this test method, the correctness of the results returned by the lower layer methods can be guaranteed to the greatest extent between the controller layer and the service layer methods [29-31]. It ensures the smooth and fast development of the system. After the unit test of the whole project, each function point has obtained the correct results. In this way, we can enter the integration test phase of the system.

Integration testing is based on unit testing to test whether each part of the work meets or realizes the corresponding technical indicators and requirements in the process of assembling all software units into modules, subsystems or systems according to the requirements of design specifications [32-34]. In other words, before integration testing, unit testing should have been completed, and the objects used in integration testing should be software units that have passed unit testing. This is very important because without unit testing, the effect of integration testing will be greatly affected, and the cost of software unit code error correction will be greatly increased.

After the design of each functional module, the system needs to test the correctness and complexity of the actual use of the module, the response speed of the website, the concurrent use of the system by multiple users, and the security of the system in actual use. Prevent the system from not working normally and poor user experience due to a large number of concurrency. Table 2 shows the description of each function point of the system and the description of the test cases of the

corresponding function points. Finally, the test results of the function points are obtained through the integration test. Table 3 shows the system function test results.

Serial number	Function description	Test case description	Test result
1	System user login	Enter the user name and password to log in to the system main interface	Realize
2	Basic information management	Add, delete, modify and query basic information such as materials and departments	Realize
3	Purchase materials	The purchaser adds a purchase order	Realize
4	Po approval	The reviewer reviews the purchase order	Can query and approve by Doc No
5	Purchase warehousing	The warehouse keeper queries the purchase order and receives it	Can query and stock in by Doc No
6	Reserve transfer / scrap receipt	The warehouse keeper adds a reserve transfer / scrap doc	Realize
7	Material ex warehouse	The warehouse keeper adds an issue document	Realize
8	Issue approval	The approver approves the issue doc	Can query and approve by Doc No
9	Inventory management	Query and statistics of receipt, issue and inventory information	The results can be queried and displayed within 3S
10	System log management	The system administrator queries the user operation log	Realize

Table 3: System function test results

In the process of system design, we need to test and improve constantly, find out the loopholes in the system through testing, and modify and improve them in time. Due to the small number of users of the system, the performance requirements are not too high, so the general requirements of the system are to realize the functions required by the user and obtain the desired results within the user's acceptable response time (the response time specified by the system is within 3S).

5 Conclusion

The system designed in this paper mainly realizes the five necessary functional modules of warehouse management: basic information management, system management, procurement management, warehousing management and inventory management. Provide automatic and comprehensive records for the whole process of material management of the enterprise, and provide real-time and correct information for all warehouse activities, resources and inventory levels. After the system running test, the system runs normally, the unit test and integration test can meet the expected requirements, realize the functions required by the user, and obtain the desired results within the user's acceptable response time (within 3S). Warehouse management system is the core of material management and an

indispensable part of an enterprise. The information content it provides is very important for enterprise decision-makers and managers. The development of this system not only improves the efficiency of its own material management, but also improves the material management level of its own enterprise. The effectiveness and the efficiency of the design can be increase by adopting the approach of artificial intelligence and this work can be extended in this direction in the future.

References

[1] Velicka, J. , Pies, M. , & Hajovsky, R. . (2018). Wireless measurement of carbon dioxide by use of iqrf technology. IFAC-PapersOnLine, 51(6), 78-83, DOI: https://doi.org/10.1016/j.ifacol.2018.07.133

[2] Raju, T. , & Kim, W. S. . (2019). Mobile guidance system for evacuation based on wi-fi system and node architecture. Journal of Information Technology Applications & Management, 26, https://doi.org/10.21219/jitam.2019.26.5.041

[3] Liu, Y. X. , & Li, X. Y. . (2020). Design and implementation of a business platform system based on java. Procedia Computer ence, 166, 150-153, DOI: https://doi.org/10.1016/j.procs.2020.02.038.

[4] J Liang, Wu, Z. , Zhu, C. , & Zhang, Z. H. .

(2020). An estimation distribution algorithm for wave-picking warehouse management. Journal of Intelligent Manufacturing(1), 1-14, DOI: 10.1007/s10845-020-01688-6

[5] Zhao, J. , Xue, F. , & Li, D. A. . (2019). Intelligent management of chemical warehouses with rfid systems. Sensors, 20(1), 123, DOI: 10.3390/s20010123

[6] Zhao, K. , Zhu, M. , Xiao, B. , Yang, X. , & Wu, J. . (2020). Joint rfid and uwb technologies in intelligent warehousing management system. IEEE Internet of Things Journal, PP(99), 1-1, DOI: 10.1109/JIOT.2020.2998484

[7] Viloria, A. , Rodado, D. N. , & Lezama, O. . (2019). Recovery of scientific data using intelligent distributed data warehouse. Procedia Computer Science, 151, 1249-1254

[8] Zhang, Y. , Yao, J. , & Guan, H. . (2018). Intelligent cloud resource management with deep reinforcement learning. IEEE Cloud Computing, 4(6), 60-69, DOI: 10.1109/MCC.2018.1081063

[9] Kermani, M. , Adelmanesh, B. , Shirdare, E. , Sima, C. A. , & Martirano, L. . (2021). Intelligent energy management based on scada system in a real microgrid for smart building applications. Renewable Energy, 171.

[10] Yu, S. , X Chen, Zhou, Z. , X Gong, & Wu, D. . (2020). When deep reinforcement learning meets federated learning: intelligent multi-timescale resource management for multi-access edge computing in 5g ultra dense network. IEEE Internet of Things Journal, PP(99), 1-1, DOI: 10.1109/JIOT.2020.3026589

[11] Somasundaram, M. , Junaid, K. , & Mangadu, S. . (2020). Artificial intelligence (ai) enabled intelligent quality management system (iqms) for personalized learning path. Procedia Computer Science, 172, 438-442, https://*doi*.org/10.1016/j.procs.2020.05.

[12] Nastasi, G. , Colla, V. , Cateni, S. , & Campigli, S. . (2018). Implementation and comparison of algorithms for multi-objective optimization based on genetic algorithms applied to the management of an automated warehouse. Journal of Intelligent Manufacturing, 29, 1545-1557, DOI: 10.1007/s10845-016-1198-x.

[13] Kumar, R. S. , Raghav, L. P. , Raju, D. K. , & Singh, A. R. . (2021). Intelligent demand side management for optimal energy scheduling of grid connected microgrids. Applied Energy, 285(march), 1-14, https://doi.org/10.1016/j.apenergy.2021.116435

[14] Sharifi, H. , Roozbahani, A. , & Shahdany, S. . (2021). Evaluating the performance of agricultural water distribution systems using fis, ann and anfis intelligent models. Water Resources Management, 35(6), 1797-1816, https://doi.org/10.1007/s11269-021-02810-w.

[15] Xu, Z. , Zhang, J. , Song, Z. , Liu, Y. , Li, J. , &

Zhou, J. . (2021). A scheme for intelligent blockchain-based manufacturing industry supply chain management. Computing, 103(8), 1771-1790, https://doi.org/10.1007/s00607-020-00880-z

[16] Mo, Z. , & Zhao, C. . (2021). Dynamic cost evaluation method of intelligent manufacturing enterprises based on dea model. International Journal of Manufacturing Technology and Management, 35, DOI: 10.1504/IJMTM.2021.118802.

[17] Ad, A. , Sk, A. , Pk, A. , & P B. (2020). Adaptive power management prototype employing intelligent scheduling of time and solar tracker - sciencedirect. Procedia Computer Science, 167, 1749-1760, https://doi.org/10.1016/j.procs.2020.03.385.

[18] Alangari, S., & Khan, N. A. (2021). Artificially Intelligent Warehouse Management System. *Asian Journal of Basic Science & Research*, *3*(3), 16-24, https://doi.org/10.1016/j.jbusres.2020.09.009

[19] Darajatun, R. A. (2017, December). C. In *IOP Conference Series: Materials Science and Engineering* (Vol. 277, No. 1, p. 012002). IOP Publishing.

[20] Mostafa, N., Hamdy, W., & Elawady, H. (2020). An intelligent warehouse management system using the internet of things. *The Egyptian International Journal of Engineering Sciences and Technology*, *32*(Mechanical Engineering), 59-65.

[21] Reza Pulungan. Et al., (2013). Design of an intelligent warehouse management system. *Information Systems International Conference*, 263-268, 10.1109/ISCID.2013.117

[22] Mao, J., Xing, H., & Zhang, X. (2018). Design of intelligent warehouse management system. *Wireless Personal Communications*, *102*(2), 1355-1367, https://dl.acm.org/doi/10.1007/s11277-017-5199-7

[23] Chen, M. , Wang, T. , Ota, K. , Dong, M. , & Liu, A. . (2020). Intelligent resource allocation management for vehicles network: an a3c learning approach. Computer Communications, 151, 485-494, https://doi.org/10.1016/j.comcom.2019.12.054

[24] Sharma, A., & Kumar, R. (2019). Risk-energy aware service level agreement assessment for computing quickest path in computer networks. *International Journal of Reliability and Safety*, *13*(1-2), 96-124, https://www.inderscienceonline.com/doi/abs/10.1504/IJRS.2019.097019

[25] Utracki, J. , & Boryczka, M. . (2020). A multi-agent approach to the optimization of intelligent buildings energy management. Procedia Computer Science, 176, 2665-2674, https://doi.org/10.1016/j.procs.2020.09.296

[26] RG Rodríguez, Mares, J. J. , & Christian, G. . (2020). Computational intelligent approaches for non-technical losses management of electricity. Energies, 13, https://doi.org/10.3390/en13092393

[27] Sharma, A., Cholda, P., Kumar, R., & Dhiman, G. (2021). Risk-aware optimized quickest path

computing technique for critical routing services. *Computers & Electrical Engineering*, *95*, 107436, https://doi.org/10.1016/j.compeleceng.2021.107436.

[28] M Manbachi, & M Ordonez. (2019). Intelligent agent-based energy management system for islanded ac/dc microgrids. IEEE Transactions on Industrial Informatics, PP(99), 1-1, DOI: 10.1109/TII.2019.2945371

[29] Sharma, A., & Kumar, R. (2019). Computation of the reliable and quickest data path for healthcare services by using service-level agreements and energy constraints. *Arabian Journal for Science and Engineering*, *44*(11), 9087-9104, DOI: 10.1007/s13369-019-03836-4

[30] Sharma, A., & Kumar, R. (2019). Service-level agreement—energy cooperative quickest ambulance routing for critical healthcare services. *Arabian Journal for Science and Engineering*, *44*(4), 3831-3848. DOI: 10.1007/s13369-018-3687-z

[31] Liu, X. , Shu, S. , Yang, K. , Wang, T. , & Geng, B. . (2019). Intelligent management of secondary water supply systems in downtown shanghai. Procedia Computer Science, 154, 206-209, https://doi.org/10.1016/j.procs.2019.06.031

[32] Cui, Q. , Wang, Y. , Chen, K. C. , Wei, N. , & Ping, Z. . (2018). Big data analytics and network calculus enabling intelligent management of autonomous vehicles in a smart city. IEEE Internet of Things Journal, PP(99), 1-1, https://doi.org/10.1016/j.procs.2019.06.031

[33] Zhang, D. , Wang, L. Y. , Jiang, J. , & Zhang, W. . (2018). Optimal power management in dc microgrids with applications to dual-source trolleybus systems. IEEE Transactions on Intelligent Transportation Systems, 1-10, DOI: 10.1109/TITS.2017.2717699

[34] Chergui, H. , & Verikoukis, C. . (2020). Big data for 5g intelligent network slicing management. IEEE Network, 34(4), 56-61, DOI: 10.1109/MNET.011.1900437.

Application of Interactive Genetic Algorithm in Landscape Planning and Design

Boyang Li[1*], Ashutosh Sharma[2]
[1]Jiangxi University of Technology, College of Art and Design, Nanchang, Jiangxi 330003, China
[2]School of Computer Science, University of Petroleum and Energy Studies, Dehradun, 248171, India
Emails: boyangli7@126.com, ashutosh.sharma@ddn.upes.ac.in

This article aims at improving the design effect of garden landscape space environment and optimizes the structure of garden landscape space environment. An optimization design method of garden landscape space environment based on interactive genetic algorithm is proposed in this article by designing a landscape space environment design with image visual feature space distributed monitoring model and fuzzy pixel area feature fusion reconstruction model. The landscape space environment design of the image by multistage decomposition and pixel gray characteristics has been established landscape space environment design image visual feature reconstruction model. The method proposed in this article combines the block area template matching method with landscape space environment design of the image features visual reconstruction. The visual space distributed detection is done with information fusion using the similarity model reconstruction for landscape space environment design of image visual perception process of information fusion. In order to extract fuzzy characteristic landscape space environment design of the image, interactive genetic algorithm is used to realize the quality assessment of landscape art information fusion perception and visual reconstruction. The simulation results show that compared with the traditional method, the visual reconstruction quality of landscape spatial environment design image processed by this method is better along with higher image recognition accuracy, and the output signal-to-noise ratio is improved by 14.6%. The experimental results prove that the introduction of interactive genetic algorithm in landscape planning and design can effectively solve the problems of multi-level feature decomposition and pixel feature separation in the process of landscape design. The proposed method achieves better optimization of landscape spatial environment structure, and achieves good landscape spatial environment design effect.

Povzetek: Za snovanje prostorskih krajinskih načrtov, npr. vrtov, je uporabljen interaktivni genetski algoritem.

1 Introduction

The development of computer vision information processing technology, using visual image processing method to analyze and extract the features of garden landscape space environment, and establishing the visual reconstruction model of garden landscape space environment design image can effectively improve the artistic feature identification and fusion reconstruction ability of garden landscape space environment. In recent years, the feature reconstruction technique of landscape spatial environment art has attracted extensive attention from scholars at home and abroad. Compared to traditional landscape design method of space environment, the introduction of artificial intelligence technology, such as interactive genetic algorithm (depicted in Figure 1) can effectively extract the geometrical characteristics of the landscape space environment design image analysis model.
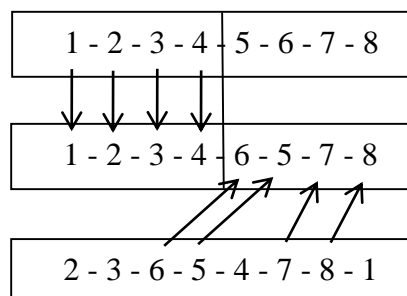

Figure 1: Interactive genetic algorithm

By machine vision characteristics such as 3-d reconstruction method, realization of the landscape space environment design and visual image reconstruction, the ability of geometric characteristic of visual recognition can be improved, thus, obtaining high-quality structural models [1].

1.1   Literature review

In view of this research problem, Shen proposed that the feature reconstruction of garden landscape spatial environment art is based on the visual reconstruction of garden landscape spatial environment design image, and the distributed fusion of visual image and binary recognition method can be used for the visual reconstruction processing of garden landscape spatial environment design image [2]. Based on the existing classification, Tamene *et al.* introduced the related concepts, development and role of ARTIFICIAL intelligence in landscape architecture research, and pointed out the specific application and existing problems of various artificial intelligence methods in landscape architecture analysis, design and evaluation [3]. Murgante *et al.* divided the mainstream ARTIFICIAL intelligence methods in landscape architecture research into artificial life, intelligent random optimization and machine learning, discussed the principle, development and characteristics of typical algorithms in each category, and then discussed the necessity of establishing a hybrid intelligence system and its future development prospect, as shown in Table 1 [4]. According to the attributes and applications of ARTIFICIAL intelligence, Ohman analyzed its limitations in the research of landscape architecture and pointed out the development trend of intelligent design of landscape architecture [5]. Smith proposed a CA model for landscape design based on inertial weight particle swarm optimization, introduced swarm intelligence into landscape design modeling, reduced the uncertainty of simulation, and established an efficient CA model to simulate landscape dynamics [6]. Anagnostopoulos and Mamanis used the improved Logics-CA mathematical model to simulate and predict the characteristics of landscape evolution space process under three conditions (historical extrapolation, endogenous development and exogenous development) in Tianjin Coastal area from 2011 to 2020, so as to further obtain the elements affecting landscape design and master the development process of landscape design [7]. Cho *et al.* proposed the use of interactive genetic algorithm to express the cycle iteration relationship of design activities in view of the complex iterative cycle in garden engineering design, so as to clearly reflect the data mutual extraction between design activities [8]. Eikelboom *et al.* proposed a DSM-GA algorithm compiled based on the critical path Method (CPM), and applied the genetic algorithm to the design activity matrix to find a better order arrangement of design activities, so as to optimize the design iteration

and shorten the design period [9]. Venema and Calamai extracted the geometric feature analysis model of garden landscape spatial environment design image, and realized garden landscape spatial environment design and image visual reconstruction through the 3d reconstruction method of computer vision features [10]. Banyai used pixel tracking and fusion technology to construct key feature quantity of landscape spatial environment design image, and realized landscape spatial environment design and optimization recognition [11].

1.2   Contribution

On the basis of the current research, this paper proposes an optimization design method of landscape space environment based on interactive genetic algorithm. A landscape space environment design is built in this article for image visual feature space distributed detection and fuzzy pixel area feature fusion reconstruction. The classification of artificial intelligence in landscape architecture research is depicted in Figure 2.

The proposed model achieves a similarity information fusion model is adopted to improve the landscape space environment design in the process of visual image reconstruction garden art landscape perception and block template matching area information fusion. This article aims at improving the design effect of garden landscape space environment and optimize the structure of garden landscape space environment. An optimization design method of garden landscape space environment based on interactive genetic algorithm is proposed in this article by designing a landscape space environment design with image visual feature space distributed monitoring model and fuzzy pixel area feature fusion reconstruction model. The method proposed in this article combines the block area template matching method with landscape space environment design of the image features visual reconstruction. The visual space distributed detection is done with information fusion using the similarity model reconstruction for landscape space environment design of image visual perception process of information fusion. The proposed method achieves better optimization of landscape spatial environment structure, and achieve good landscape spatial environment design effect.

The rest of this article is arranged as: section 2 presents the research methods depicting the visual sampling and fusion of landscape spatial environment design. Section 3 presents the research results and discussion followed by conclusion in section 4.
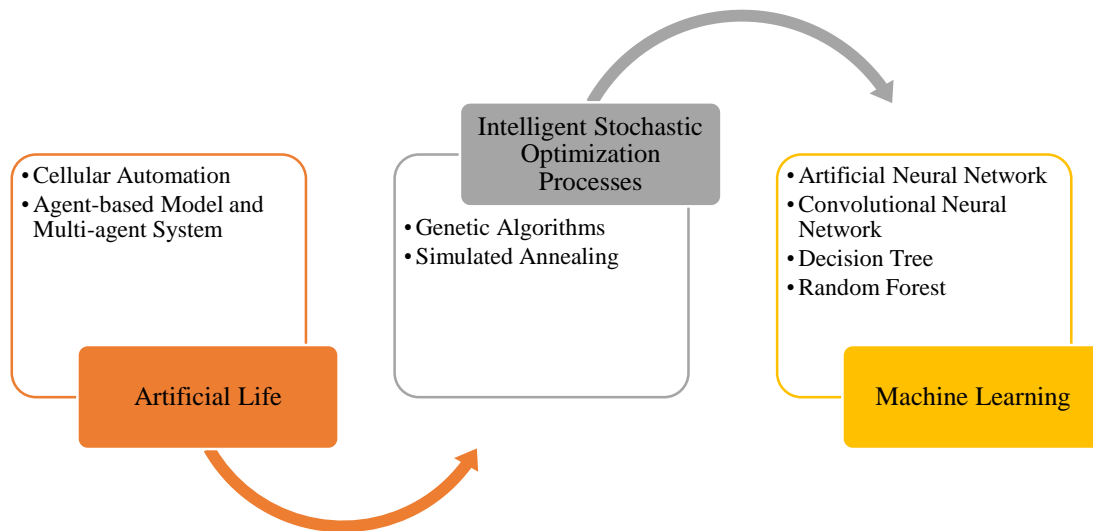
Figure 2: Classification of artificial intelligence in landscape architecture research

2 Research methods

2.1 Visual sampling and fusion of landscape spatial environment design

The flowchart of visual sampling and fusion of landscape spatial environment design is presented in Figure 3.
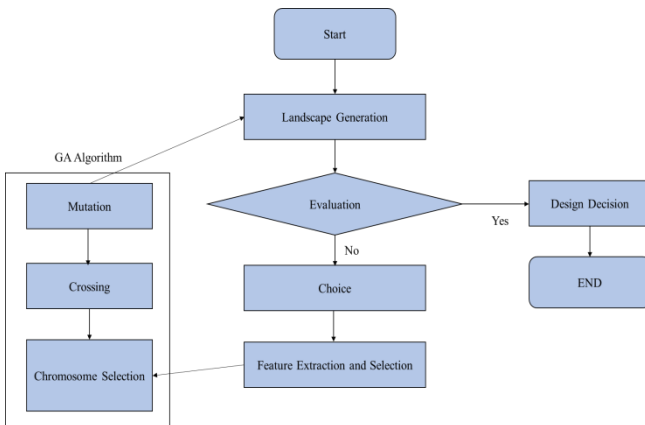

Figure 3: Flowchart of landscape spatial environment design

The system provides the genetic operation applications in urban landscapes for evolving the input variabilities in position of walls, their heights and building structure. The weighted estimate of feature matching is done using spare features for image reconstruction which is detailed in further subsection.

A. *Landscape space environment design image collection*

In order to achieve the landscape space environment design based on interactive genetic algorithm (GA) and visual image reconstruction, to build landscape space environment design of the image pixel space fusion model. The feature matching method is adopted to improve the landscape space environment design and

image feature detection [12], for the sparse feature of the landscape space environment design image reconstruction. The Atanassov extension method was used to match feature points of landscape spatial environment design images, and the template matching model for visual reconstruction of landscape spatial environment design images was constructed, as shown in Figure 4.
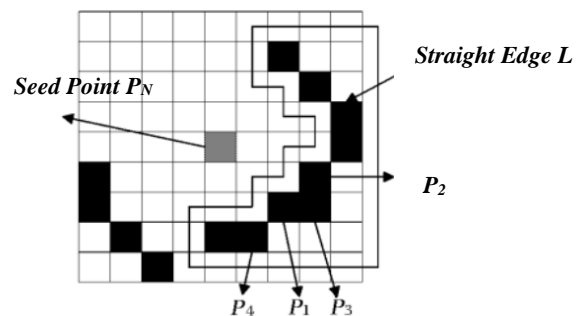

Figure 4: Template matching model of landscape space environment design

Assume that landscape space environment design of the image gray of pixel sets (I, j), as a pixel center, sharpening template block combination method is used to construct landscape space environment design of the image characteristics of visual reconstruction model. For the first k is the belt in the acquisition of landscape space environment design of the image grey value I_{swk}, in gray pixel distribution characteristics of space, The gradient characteristic components of the corresponding landscape spatial environment design image are as follows:

$$P_{rk} = (\frac{\sum_{j=1}^{c} I_{swk}(1,j)}{c}, \frac{\sum_{j=1}^{c} I_{swk}(2,j)}{c}, ..., \frac{\sum_{j=1}^{c} I_{swk}(1,j)}{c}, ..., \frac{\sum_{j=1}^{c} I_{swk}(r,j)}{c}) \quad (1)$$

$$P_{ck} = (\frac{\sum_{i=1}^{r} I_{swk}(i,1)}{r}, \frac{\sum_{i=1}^{r} I_{swk}(i,2)}{r}, ..., \frac{\sum_{i=1}^{r} I_{swk}(i,j)}{r}, ..., \frac{\sum_{i=1}^{r} I_{swk}(i,c)}{r}) \quad (2)$$

Where C is the column number of LGB vector quantization matrix of landscape spatial environment design image, and R is motion fuzzy feature quantity. Based on the fusion reconstruction method of fuzzy pixel regional features, the pixel set of artistic feature distribution of landscape spatial environment was obtained, and the information reconstruction and three-dimensional perception of landscape spatial environment design image were carried out to improve the ability of environment design.

B. Image feature fusion and reconstruction model

The spatial distributed detection model of visual feature of landscape spatial environment design image was constructed. Multi-stage feature decomposition and grey pixel feature separation of garden landscape spatial environment design image are carried out [13-15], and the visual feature reconstruction model of garden landscape spatial environment design image is established. The visual feature distribution of garden landscape spatial environment design image is as follows:

$$\vec{G(x)} = \sum_{j=1}^{p} \vec{G_j(x)} \quad (3)$$

Adaptive fusion method was used to reconstruct the image vision of landscape design, and the edge vision reconstruction model of landscape spatial environment design was constructed [16, 17]. The fuzzy proximity function of landscape spatial environment image was obtained as follows:

$$fitness(\vec{x}) = f(\vec{x}) + (Ct)^{\alpha} \sum_{j=1}^{p} G_j^{\beta}(\vec{x}) \quad (4)$$

It is assumed that the coordinate of garden landscape spatial environment design P_N is (X_{PN}, Y_{PN}), then the coordinate of all garden landscape spatial environment design edge points (X_k, y_k) on L is compared with P_N: when $X_K > X_{PN}$, $i_L = i_L + 1$; When $x_K < X_{PN}$, $i_L = i_L-1$; When $X_k = X_{PN}$, $i_L = i_L + 0$. The perception fusion model of landscape art information fusion is constructed, and the fitness function of landscape art information fusion is as follows:In this section various state-of-the-art work in the field of optimization design based on Computer-Aided architecture is presented.

$$fitness(\vec{x}) = \begin{cases} f(\vec{x}), \\ 1 + r\vec{G(x)}, \end{cases} \quad (5)$$

Considering the gray level f of the garden landscape spatial environment design, the resolution model of the garden landscape spatial environment vision is constructed by using the gray level invariant moment feature decomposition method [18-20], and the visual feature reconstruction model of the garden landscape spatial environment is as follows:

$$W_u(a,b) = e^{j2\pi k \ln \alpha} \times \frac{K}{\sqrt{\alpha}} \left\{ \begin{bmatrix} \dfrac{\alpha e^{\frac{j2\pi f_{min}}{\alpha}(b-b_a)}}{f_{min}} - \dfrac{e^{\frac{j2\pi f_{max}}{\alpha}(b-b_a)}}{f_{max}} \end{bmatrix} + j2\pi(b-b_a)[Ei(j2\pi f_{max}(b-b_a))] \\ -Ei(\dfrac{j2\pi f_{min}}{\alpha}(b-b_a))] \end{bmatrix} \right\} \quad (6)$$

Where $b_a = (1-a)\left(\frac{1}{af_{max}} - \frac{T}{2}\right)$, $Ei(.)$ represents the recombination output of visual information features of garden landscape spatial environment. Combined with model recognition method, garden landscape design is carried out.

2.2 Optimization of landscape design

A. Landscape space environment vision

After extracting all edge points on L, δ_1^2 is the local variance of landscape spatial environment design image, δ_n^2 is the optimization coefficient of landscape spatial environment design image. $\beta = \max[\frac{\delta_1^2-\delta_n^2}{\delta_1^2}, 0]$, using gradient descent method for visual landscape space environment of the regional block visual refactoring, make landscape space environment design of the image sparse eigenvalue meet $C \in S$, according to the sparse prior as a result, the environmental design of landscape space, image Fm the first m frames (x, y) (x, y) in the optimal visual reconstruction threshold. Based on the approximate sparse representation method, template matching of landscape spatial environment design image was carried out, and the matching coefficient was obtained as follows:

$$g_i^* = \begin{cases} Rs_j, z \le i \le x-y \\ g_i, 否则 \end{cases} \quad (7)$$

In the formula, R is a standard constant. In combination with block area template matching method, landscape spatial environment design and distributed detection are carried out, and contour point matching model is used to extract edge features of landscape spatial environment design [21-23]. The maximum gray value of image analysis department of landscape spatial environment design is:

$$n_{pg} = \frac{\mu_{pb}}{(\mu_{00})^{\gamma}} \quad (8)$$

Sparse representation and super-resolution reconstruction methods were used for visual reorganization of landscape spatial environment design,

and interactive genetic algorithm was used to realize landscape art information fusion perception. The information reconstruction model of landscape spatial environment design was expressed as follows:

$$g(x, y) = f(x, y) + \varepsilon(x, y) \tag{9}$$

$f(x,y), g(x,y)$ and $\varepsilon(x,y)$ respectively represent the original landscape spatial environment image, reconstructed image and gray scale image. In summary, interactive genetic optimization design of landscape design can be carried out.

B. Interactive genetic optimization

This paper proposes a visual reconstruction algorithm for landscape spatial environment design based on interactive genetic algorithm [24,25]. Template matching method combining block areas landscape space environment design of the image characteristics of visual reconstruction, based on local feature adaptive feature matching method constructs a model of information fusion visual landscape space environment, the construction of landscape space environment art characteristic expression model, under the genetic evolutionary optimization, get the garden art landscape information fusion expression is:

$$g = k \otimes f + n \tag{10}$$

Where $\otimes$ represents convolution operator, carries out vector set fusion processing on the collected design images of landscape spatial environment, constructs the visual feature decomposition model of landscape spatial environment, and obtains the best discerning feature value of landscape spatial environment vision:
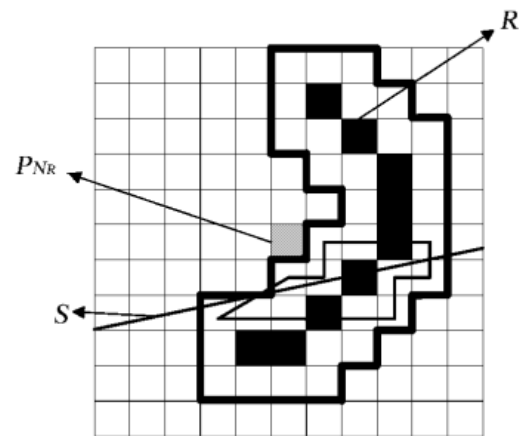
$$s_{PPM}(t) = \sum_{i=-\infty}^{\infty} \sum_{j=0}^{N_p-1} p(t - iT_s - jT_p - c_j T_c - a_i \varepsilon) \tag{11}$$

$$s_{PPM}(t) = \sum_{j=-\infty}^{\infty} d_j p(t - jT_s) \tag{12}$$

Where T_s is the optimization iteration width of genetic evolution. Under interactive inheritance, the block model of garden landscape spatial environment design is obtained, as shown in Figure 5.
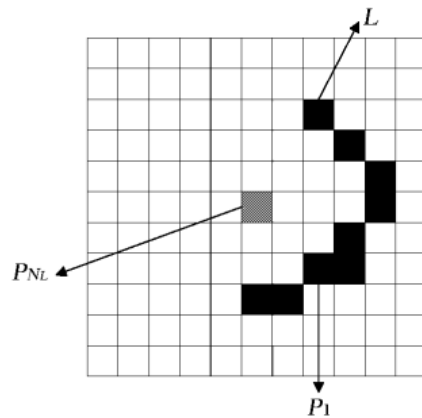
(a): Block matching of left seed points

(b): Block matching of right seed points

Figure 5(a, b): Landscape space environment design

Gray correlation constraint is added to determine the final matching point, and the image pixel decomposition model is expressed as:

$$x(t) = \sum_{m=1}^{M} \sum_{k=1}^{K(m)} \omega_{nk} s(t - T_m - \tau_{mk}) + v(t) \tag{13}$$

In the formula, ω_{nk} is the fuzzy feature component of landscape spatial environment vision. Under the genetic interactive evolution, the output of landscape optimization design is:

$$\begin{cases} x = R \sin \eta \cos \varphi, 0 \le \varphi \le 2\pi \\ y = R \sin \eta \sin \varphi, 0 \le \eta \le \pi \\ z = R \cos \eta, R = D / 2 \end{cases} \tag{14}$$

Among them, η represents the landscape spatial visual reconstruction function, φ represents the Angle function of landscape spatial environment image visual reconstruction, and R represents the template matching coefficient.

3 Research results and discussion

In order to verify the application performance of the proposed method in realizing the spatial environment design of landscape, simulation experiment analysis was conducted. It was assumed that the number of seed points in landscape design was 40, the coefficient of feature matching was 0.36, and the block size of pixel points was 200 × 200. The gray scale of landscape design was shown in Figure 6.
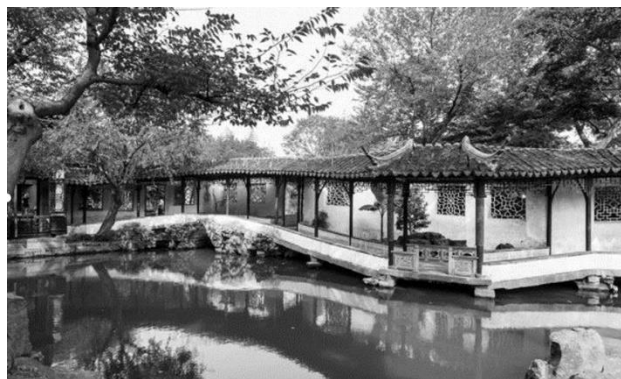
Design Sample 1

Design Sample 2

Figure 6: Gray scale of landscape design

Taking the image in Figure 6 as the research object, the similarity information fusion model is used to carry out the perception of landscape art information fusion and block region template matching in the process of visual reconstruction of landscape spatial environment design image to realize the design optimization, and the optimized design results are shown in Figure 7. The analysis of Figure 7 shows that the proposed method can effectively realize the optimal design of landscape space environment, with higher image recognition accuracy and improved design effect. The output SNR of the proposed method is 14.6% higher than that of the traditional method.

Design Sample 1

Design Sample 2

Figure 7: Optimization of landscape design

The proposed genetic algorithm-based approach is compared with the tradition method in terms of Signal to noise ratio as well as accuracy. The outcomes obtained are depicted in Figure 8.
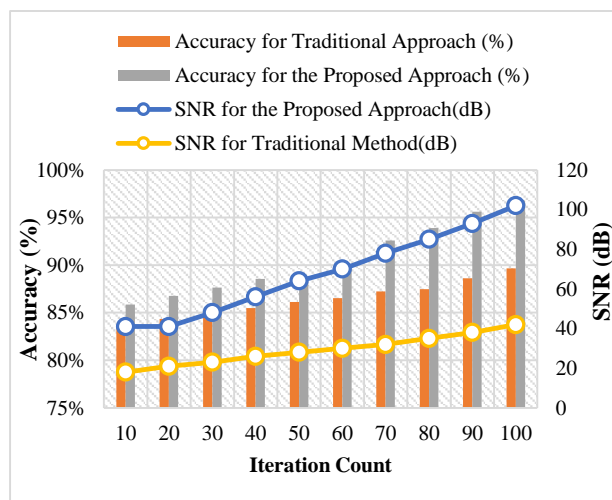

Figure 8: Comparative analysis of proposed and traditional method

The output signal-to-noise ratio is improved by 14.6% for the proposed approach as compared to the tradition approach. The accuracy value obtained using

the proposed method is improved by 7.9% comparative to the traditional approach. The outcomes obtained depicts the viability of the proposed approach.

4　Conclusion

This article presents an optimization design method of landscape space environment based on interactive genetic algorithm. Firstly, the multi-stage feature decomposition and grey pixel feature separation of landscape spatial environment design image are carried out by constructing distributed detection of visual feature space and reconstruction model of fuzzy pixel region feature fusion. On this basis, the visual feature reconstruction model of landscape spatial environment design image was established, the fuzzy feature quantity of landscape spatial environment design image was extracted, and the interactive genetic algorithm method was used to realize the landscape information fusion perception and visual reconstruction quality evaluation of landscape art. The sparse representation and super-resolution reconstruction methods were used for visual reorganization of landscape spatial environment design, and interactive genetic algorithm was used to realize landscape art information fusion perception and landscape design optimization. The simulation results show that compared with the traditional method, the visual reconstruction quality of landscape spatial environment design image processed by this method is better along with higher image recognition accuracy of 96.78%, and the output signal-to-noise ratio is improved by 14.6%. The experimental results prove that the introduction of interactive genetic algorithm in landscape planning and design can effectively solve the problems of multi-level feature decomposition and pixel feature separation in the process of landscape design. The method proposed achieves better optimization of landscape spatial environment structure and has a good design effect of landscape spatial environment.

References

[1]　Li, G. (2020, June). Urban Landscape Design Optimization Based on Interactive Genetic Algorithm. In *International Conference on Applications and Techniques in Cyber Security and Intelligence* (pp. 1097-1102). Springer, Cham.

[2]　Shen, C. (2021). Rural road ecological landscape planning system based on interactive genetic algorithm. In *E3S Web of Conferences* (Vol. 267, p. 01013). EDP Sciences.
https://doi.org/10.1051/e3sconf/202126701013

[3]　Tamene, L., Le, Q. B., & Vlek, P. L. (2014). A landscape planning and management tool for land and water resources management: an example application in northern Ethiopia. *Water resources management*, *28*(2), 407-424.
https://doi.org/10.1007/s11269-013-0490-1
10.1007/978-3-030-53980-1_166

[4]　Murgante, B., Borruso, G., & Lapucci, A. (2009). Geocomputation and urban planning. In *Geocomputation and urban planning* (pp. 1-17). Springer, Berlin, Heidelberg.
https://doi.org/10.1007/978-3-540-89930-3_1

[5]　Öhman, K. (2001). *Forest planning with consideration to spatial relationships* (Vol. 198, No. 198).
https://www.cabdirect.org/cabdirect/abstract/20033035907

[6]　Smith, K. A. (1991). Ecology of Arable Land: Organisms, Carbon and Nitrogen Cycling.(Ecological Bulletins 40). Edited by O. Andrén, T. Lindberg, K. Paustian and T. Rosswall. Copenhagen: Munksgaard (1990), pp. 221, no price stated. *Experimental Agriculture*, *27*(1), 89-89.
https://doi.org/10.1017/S0014479700019293

[7]　Anagnostopoulos, K. P., & Mamanis, G. (2011). Multiobjective evolutionary algorithms for complex portfolio optimization problems. *Computational Management Science*, *8*(3), 259-279.
https://doi.org/10.1007/s10287-009-0113-8

[8]　Cho, Y. J., Wang, Y., & Hsu, L. L. I. (2016). Constructing Taiwan's low-carbon tourism development suitability evaluation indicators. *Asia Pacific Journal of Tourism Research*, *21*(6), 658-677.
https://doi.org/10.1080/10941665.2015.1068193

[9]　Eikelboom, T., Janssen, R., & Stewart, T. J. (2015). A spatial optimization algorithm for geodesign. *Landscape and Urban Planning*, *144*(December), 10-21.
http://dx.doi.org/10.1016/j.landurbplan.2015.08.011

[10]　Venema, H. D., & Calamai, P. H. (2003). Bioenergy systems planning using location–allocation and landscape ecology design principles. *Annals of Operations Research*, *123*(1), 241-264.
https://doi.org/10.1023/A:1026135632158

[11]　Bányai, T. (2009, April). Modelling and genetic algorithm based optimisation of inverse supply chain. In *EGU General Assembly Conference Abstracts* (p. 3794).
https://meetingorganizer.copernicus.org/EGU2009/EGU2009-3794.pdf

[12]　Gao, Y., Wu, Z., Lou, Q., Huang, H., Cheng, J., & Chen, Z. (2012). Landscape ecological security assessment based on projection pursuit in Pearl River Delta. *Environmental monitoring and assessment*, *184*(4), 2307-2319.
https://doi.org/10.1007/s10661-011-2119-2

[13]　Liu, Y., Yuan, M., He, J., & Liu, Y. (2015). Regional land-use allocation with a spatially explicit genetic algorithm. *Landscape and Ecological Engineering*, *11*(1), 209-219.
https://doi.org/10.1007/s11355-014-0267-6

[14]　Ting, L., Khan, M., Sharma, A., & Ansari, M. D. (2022). A secure framework for IoT-based smart climate agriculture system: Toward blockchain and edge computing. *Journal of Intelligent Systems*, *31*(1), 221-236.
https://doi.org/10.1515/jisys-2022-0012

[15]　Sharma, D., Kaur, R., Sandhir, M., & Sharma, H. (2021). Finite element method for stress and strain analysis of FGM hollow cylinder under effect of temperature profiles and inhomogeneity parameter. *Nonlinear Engineering*, *10*(1), 477-487.
https://doi.org/10.1515/nleng-2021-0039

[16]　Tong, Z., & Ding, W. (2011, April). A genetic algorithm

approach to planning of buildings in urban green space. In *2011 International Conference on Electric Technology and Civil Engineering (ICETCE)* (pp. 2856-2859). IEEE. *10.1109/ICETCE.2011.5775697*

[17] Kaabar, M. K., Kalvandi, V., Eghbali, N., Samei, M. E., Siri, Z., & Martínez, F. (2021). A Generalized ML-Hyers-Ulam Stability of Quadratic Fractional Integral Equation. *Nonlinear Engineering*, *10*(1), 414-427. *https://doi.org/10.1515/nleng-2021-0033*

[18] Feng, J., Li, Y., Zhao, K., Xu, Z., Xia, T., Zhang, J., & Jin, D. (2020). DeepMM: deep learning based map matching with data augmentation. *IEEE Transactions on Mobile Computing*. *10.1109/TMC.2020.3043500*

[19] Chen, Y., Zhang, W., Dong, L., Cengiz, K., & Sharma, A. (2021). Study on vibration and noise influence for optimization of garden mower. *Nonlinear Engineering*, *10*(1), 428-435. *https://doi.org/10.1515/nleng-2021-0034*

[20] Sharma, R., Raju, C. S., Animasaun, I. L., Santhosh, H. B., & Mishra, M. K. (2021). Insight into the significance of Joule dissipation, thermal jump and partial slip: Dynamics of unsteady ethelene glycol conveying graphene nanoparticles through porous medium. *Nonlinear Engineering*, *10*(1), 16-27. *https://doi.org/10.1515/nleng-2021-0002*

[21] Cantelli, A., D'orta, F., Cattini, A., Sebastianelli, F., & Cedola, L. (2015). Application of genetic algorithm for the simultaneous identification of atmospheric pollution sources. *Atmospheric Environment*, *115*, 36-46. *https://doi.org/10.1016/j.atmosenv.2015.05.030*

[22] Shabaz, M., Sharma, A., Al Ajrawi, S., & Estrela, V. V. (2022). Multimedia-based emerging technologies and data analytics for Neuroscience as a Service (NaaS). *Neuroscience Informatics*, *2*(3), 100067. *https://doi.org/10.1016/j.neuri.2022.100067*

[23] Wagner, M. P., & Oppelt, N. (2020). Extracting agricultural fields from remote sensing imagery using graph-based growing contours. *Remote sensing*, *12*(7), 1205. *https://doi.org/10.3390/rs12071205*

[24] Kumbinarasaiah, S., & Raghunatha, K. R. (2021). A novel approach on micropolar fluid flow in a porous channel with high mass transfer via wavelet frames. *Nonlinear Engineering*, *10*(1), 39-45. *https://doi.org/10.1515/nleng-2021-0004*

[25] Wang, H., Sharma, A., & Shabaz, M. (2022). Research on digital media animation control technology based on recurrent neural network using speech technology. *International Journal of System Assurance Engineering and Management*, *13*(1), 564-575. *https://doi.org/10.1007/s13198-021-01540-x*

Automatic Classification of Document Resources Based on Naive Bayesian Classification Algorithm

Rong Wang[1*]
Email: rongwang9@126.com

World Wide Web has become big as the amount of documents collection is increasing rapidly. The automatic classification of document resources based on Naive Bayesian classification algorithm is detailed in this paper. Firstly, this paper introduces the relevant theories of naive Bayes classification and the automatic document classification system. Then, a massive network academic document automatic classification system is designed and implemented. The system uses modular design, including academic document automatic capture module, academic document word document matrix processing module, ontology integration module and semantic driven classification module. Finally, based on the Naive Bayesian classification algorithm, the training set of 12 categories preset is utilized in the professional classification directory of the Ministry of education.. Experiments show that the naive Bayesian classification algorithm can effectively complete the automatic capture, processing and classification of massive academic documents, which can not only improve the classification accuracy, but also reduce the running time of automatic classification. It solves the problems of the integration of two heterogeneous ontology libraries and also the problem that the traditional word vector space cannot meet people's needs for semantic classification.

Povzetek: Za avtomatsko klasifikacijo dokumentov s spleta je implementiran naivni Bayesov algoritem.

1   Introduction

The Internet is an information resource of text, images, audio and video. There is rapid increase in the amount of information available on the World Wide Web (WWW) at an exponential rate. This rich textual information is contained in the Web documents but the growth of the internet has made it difficult for users for location of relevant information quickly on the Web. At present, network academic resources show an upward trend in both breadth and depth, and have attracted more and more attention from the academic community. The massive network academic literature has a huge scale and fast update speed. Fully mining it has important academic value. However, these characteristics have also become a stumbling block for scientific researchers to make use of it. How to acquire and process a large amount of academic literature is a severe test for computer processing and throughput. Whether in terms of processing speed, storage space, fault tolerance or access speed, it is difficult for single computer platform architecture and processing capacity to successfully complete this task. Due to the huge number of network academic documents, it is difficult to make effective use of them, so it is of practical significance to classify them automatically based on disciplines. Automatic document classification is widely used in the fields of information retrieval, data mining, spam filtering, digital library and so on. There are two common classification methods: one is rule-based, which usually requires a large number

of domain experts to extract the rules of the text, which is time-consuming and laborious, and the classification effect is poor; Another kind of method is machine learning method based on statistics, including nearest neighbor method, support vector machine, naive Bayes, decision tree, neural network, etc. this kind of method usually uses feature vector space to train document classification model. However, word feature vectors ignore the semantic relationship between words and cannot reflect synonyms, polysemy and the upper and lower relationship between words, resulting in too high vector space dimension. When automatically classifying massive documents, there will be problems such as insufficient memory, slow classification speed and low classification performance, Automatic document classification technology and method cannot be more widely applied to the practice of specific fields [2]. In order to solve the problems existing in the traditional automatic document classification based on word vector space, a series of semantic driven automatic document classification methods are proposed, such as latent semantic analysis method, ontology semantic mapping method, concept lattice construction method, standardized concept analysis method and so on. Although the semantic driven automatic text classification method can greatly reduce the dimension of document vector space, it also has many defects, such as high requirements for semantic reasoning ability, high computational complexity, and unable to classify web documents quickly and effectively.

Bayesian classification (Figure 1) is proposed on the solid theoretical basis of Bayesian theorem. For a given sample, the posterior probability of belonging to each category is calculated according to the distribution of each category sample in the training set, and then the category of the sample is judged as the category corresponding to the maximum posterior probability. The principle of this method is simple, but when the number of attributes is large, training and learning a classification model completely according to Bayesian theorem will have a huge computational overhead and will be greatly limited in practical application [3]. Therefore, scholars simplified a hypothesis of attribute conditional independence, and proposed a practical naive Bayesian classification algorithm, which greatly reduced the computational overhead in the process of model training. At the same time, the research also shows that naive Bayesian classification method still has good performance in many practical applications.
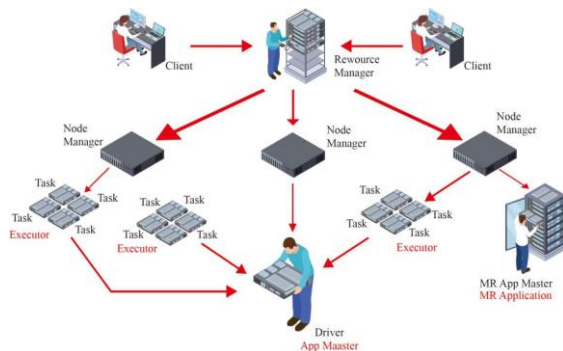
Figure 1: Bayesian classification

Contribution: This paper introduces the relevant theories of naive Bayes classification and the automatic document classification system. Then, a massive network academic document automatic classification system is designed and implemented.

The organization of the paper is as follows. Section 2 provides an overview of the exhaustive literature survey followed by the Automatic classification of massive network academic documents adopted in section 3. The experimental analysis is in section 4. Finally, Section 5 concludes the paper.

2   Literature review

Du, J. h. and others also proposed a network extended naive Bayesian classification model (BAN). This method extends the structure of naive Bayesian classifier to a greater extent. It is the same as the improved model of TAN. Its fundamental starting point is to weaken the assumption that attributes are independent to a greater extent. The ban model is the same as the TAN model in many aspects. The BAN model also stipulates that the class node is the root. At the same time, all other attribute nodes take its parent node and the BAN classifier uses Bayesian network as

the expression structure, which is the only difference [4]. Y Kumar's Bayesian augmented naive Bayesian classifier GBAN is based on genetic algorithm. GBAN model can meet the limitations of the network extended naive Bayesian classification model on the network structure, that is, any attribute node has at most M parent nodes (generally m < 4), but the category variables are not included [5]. The hybrid tree augmented naive Bayesian classification model proposed by DIAS, K. L. is based on rough set theory.

The composition process of augmented naive Bayesian classification model is as follows: Based on the attribute reduction theory of rough set, under the condition of keeping the classification ability unchanged, it is divided into two categories according to the impact of attribute variables on the classification results. It is assumed that the attribute variables that have no or little impact on the classification results are independent of each other, and these nodes can only have one parent node, The attribute variables that affect the classification results are not independent of each other, and these nodes can have two parent nodes [6]. Tajanpur proposed a hybrid model (nbtree) combining decision tree and naive Bayes. The process of learning nbtree by the algorithm is similar to that of decision tree (C4.5), but it is different in the selection of attribute splitting evaluation score function [7]. Gaber, A. and others proposed an average naive Bayesian tree model [8]. Lopes, F. and others proposed an improved naive Bayes model (LBR) combining lazy technology and naive Bayes, which can obtain high classification accuracy, but the classification efficiency of this method is not very high [9]. In terms of automatic document classification, the classification method based on coverage coefficient by an, Y. and others is a classification method based on the inherent attributes of document set. This method borrows mathematical tools to derive a classification step with rigorous reasoning.

The premise is that (under certain general assumptions) the class and number of classes of each document in the document set have been determined by the inherent attributes of the document set itself [10]. Rueda and others proposed an automatic acquisition and parallel processing model of massive network academic documents. The rules specified by the heritrix platform are used to capture the data of the seed site. For the captured file resources, they are judged according to the set academic literature feature rules, and then some of them are selected to invite domain experts for category indexing, train the machine learning classification algorithm, and finally realize the classification of all documents [11]. In previous years, many researchers have worked on this particular field, some of the relevant articles are tabulated in Table 1.

Authors	Presented Work	Key points	Benefits	References
Mohamed EL KOURDI et al.,	"Naive Bayes (NB) is a statistical machine learning algorithm utilized for the classification of non-vocalized Arabic web documents which is presented in this paper."	"The data set utilized during the experiments consists of 300 web documents per category."	High classification accuracy	[12]
Huaixin Chen et al., 2018	""Improved Naïve Bayes classifiers are presented utilizing multinomial model."	"The proposed method is able to improve the accuracy of Naïve Bayes classifiers dramatically."	Good scalability	[13]
Yong Wang et al., 2003	"An automatic document classification system, WebDoc, which classifies Web documents according to the Library of congress is presented."	"Performance of each method in terms of recall, precision, and F-measures is reported."	Highly effective and efficient.	[14]
A. B. Adetunji et al., 2018	"A University web site is used as a case study and a machine learning workbench called WEKA is discussed."	"General-purpose environment for automatic classification, clustering and feature selection are provided."	Naïve Bayes algorithm ability is to accurately classify the web document vast amount.	[15]
Yugang Dai a et al., 2014	"Naïve bayesian classification algorithm is presented by the author which is further combining with the rough set theory."	"This algorithm is implemented on a cloud platform utilizing map-reduce programming mode."	High recall rate	[16]

Table 1: Some existing and relevant articles in previous years

3 Automatic classifications of massive network academic documents

With the goal of automatically acquiring massive documents and automatically classifying documents, its framework is shown in Figure 2:
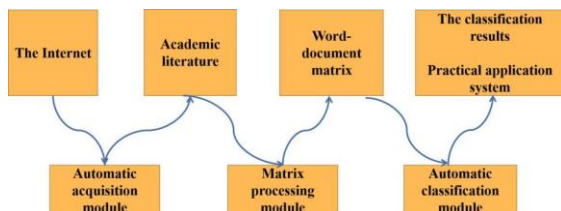

Figure 2: Framework of automatic classification system for massive network academic documents

The automatic document acquisition module first captures and determines academic documents from the Internet according to predetermined rules and conditions, so as to filter irrelevant documents; Then, through the matrix processing module, the academic literature is transformed into a word document matrix for subsequent processing; Finally, the word document matrix is imported into the automatic classification module after training and ontology integration to obtain the classification results [17, 18].

(1) Automatic acquisition of massive network academic documents

In the automatic classification system of massive network academic documents, it is necessary to obtain massive academic documents. First, use heritrix to grab all PDF files under the domain name from a specific website, read all PDF files with checkpdf, and identify academic literature through rule-based judgment method, as shown in Figure 3:
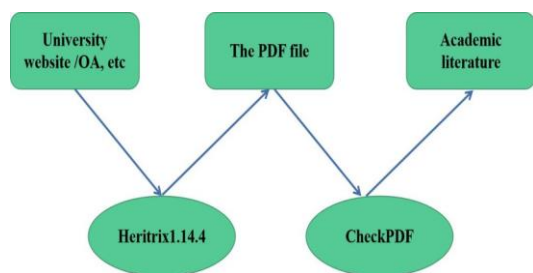
Figure 3: Automatic acquisition of massive network academic documents

In the selection of capture tools, the author studies and analyzes the network resource capture platforms such as nutch, heritrix, jspider and web harvest from the aspects of capture efficiency and scalability, and finally selects heritrix as the capture platform. Heritrix has high scalability [19, 20], can retain the original file structure and directory, and has a web user interface. It runs on Linux system and can ensure high capture speed. In terms of file format, considering the convenience of subsequent processing and the proportion of various file

types, PDF is selected as the main capture file type. After the PDF file is captured, it needs to be screened to retain the academic literature. The rule-based decision method is used, that is, the decision is made through keywords. By analyzing a large number of academic documents, it is found that its unique characteristic words include abstract, keywords, introduction, discussion, conclusion and recognition. Different documents may contain several words respectively. Therefore, a threshold can be set to judge according to the number of the above words [21-22].

(2) Massive network academic literature words -- document matrix processing

In view of the large number of documents to be processed, the word frequency matrix is generated by distributed processing. This part is implemented using Hadoop, including Hadoop namenode and Hadoop datanode. Namenode is responsible for the scheduling of parallel processing, and datanode is responsible for the actual parallel processing. Academic documents are first read into the Hadoop platform, and an index of all documents is saved on the namenode. The actual documents are saved on at least two datanodes in the form of redundancy, and finally passed Namenode calls the parallel processing program to generate the word document matrix of academic literature [23-25], as shown in Figure 4:
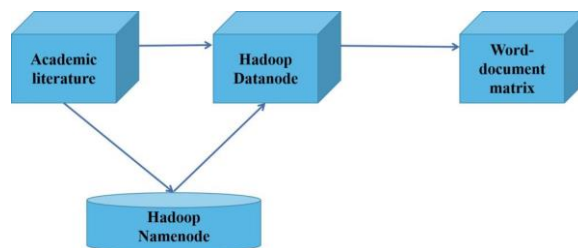

Figure 4: Massive network academic literature words - document matrix processing

In the map phase of Hadoop, stringtokenizer is used to extract the words in the literature in turn and generate a key \ value pair < word, document ID >. In the reduce phase of Hadoop, a reducer is used to process the same word, create an array with the length of documents, save the word frequency of the current word in the corresponding documents, and then accept the key \ value pair in turn and update the array. Output the matrix after all reducer work is completed. Since this matrix is sparse, you can delete 0 bits and output sparse matrix to reduce storage space [26, 27].

(3) Ontology integration

In order to understand natural language, the common method is to use ontology library to annotate and integrate text. This part mainly uses prompt. Prompt first reads the ontology, then analyzes the relationship between concepts, maps the same concepts, retains the special concepts in an ontology library, and finally generates an integrated integrated ontology, as shown in Figure 5.
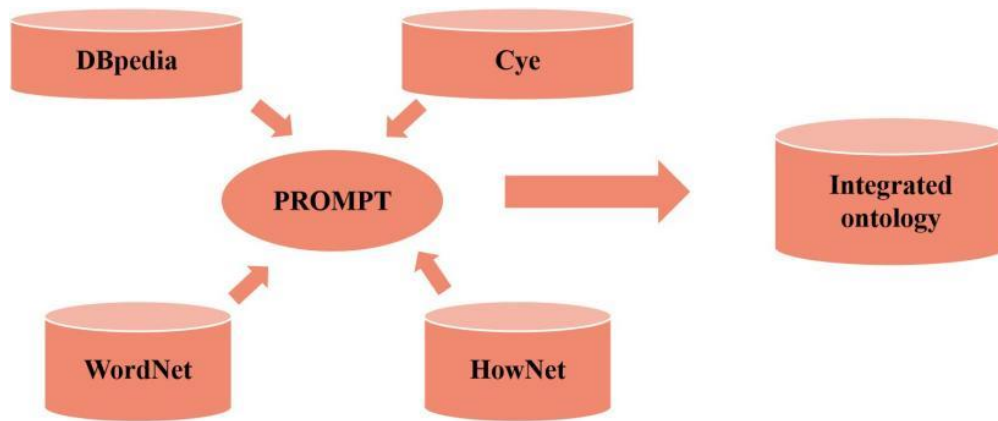
Figure 5: Ontology integration

3.1   Naive Bayesian algorithm

Before describing the naive Bayesian classification algorithm, the classification problem is formalized from the perspective of statistics. Let X represent the attribute set of the system data set, $X = (A_1, A_2, \cdots A_m)$, Y represent the class label set of the system data set, and $Y = (C_1, C_2, \cdots C_t)$. Because the relationship between class variables and attributes is uncertain, X and Y can be regarded as random variables, and $P(Y'|X')$ can be used to capture the relationship between them in a probabilistic manner. $P(Y'|X')$ is also called a posteriori probability of class Y. Correspondingly, $P(Y)$ is called a priori probability of Class Y [28, 29].

In the training stage of naive Bayesian classification algorithm, firstly, the information statistics of the training data set is carried out, and the a posteriori probability $P(Y|X)$ of each combination of attribute sets X and Y is calculated. After calculating these probabilities, the test sample X 'can be classified by finding the class Y ' that maximizes the delay probability $P(Y'|X')$. However, it is very difficult to accurately estimate the a posteriori probability of each possible combination of Class Y and attribute values, because even if the number of attributes is not many, a large training set is still required. At this time, the Bayesian theorem plays an important role, because posterior probability can be expressed by the prior probability $P(Y)$, the class conditional probability $P(X|Y)$ and the evidence $P(X)$ through the Bayesian theorem. The formula for calculating the posterior probability $P(Y|X)$ by the Bayesian theorem is formula (1).

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad （1）$$

When comparing the posterior probabilities of different Y values, the denominator $P(X)$ is always constant and can be ignored. The prior probability $P(Y)$ can be easily estimated by calculating the proportion of training samples belonging to each class in the total training samples in the training data set. However, for the training data with m attributes [30, 31], the calculation of class conditional probability $P(X|Y)$ is time-consuming. In order to improve the efficiency of calculating $P(X|Y)$, naive Bayesian classification algorithm assumes that the attributes are conditionally independent when estimating the conditional probability of classes. The assumption of attribute conditional independence can be expressed by formula (2):

$$P(X|Y = y) = \prod_{i=1}^{m} P(X_i|Y = y) \quad （2）$$

Through the conditional independence assumption, it is not necessary to calculate the class conditional probability of each value group sum of X, but to mark Y for a given class and calculate the conditional probability of each X_i. In contrast, the latter method is more practical. Because through the assumption of conditional independence, better probability statistics can be obtained without a large training data set [32-34].

In the classification test stage, naive Bayesian classification algorithm calculates a posteriori probability for each X, as shown in formula (3):

$$P(Y|X) = \frac{P(Y)\prod_{i=1}^{m} P(X_i)}{P(X)} \quad （3）$$

Because $P(Y)$ and $P(X)$ are fixed for fixed training data sets and determined test data. Therefore, it is sufficient to find the class that maximizes the molecular $P(Y)\prod_{i=1}^{m}P(X_i)$. For naive Bayesian classification algorithm, the biggest disadvantage is that naive Bayesian classification algorithm can only deal with discrete attributes [35, 36].

4 Experimental Analysis

The experimental classification standard selects 12 categories preset in the professional classification catalogue of the Ministry of education of the people's Republic of China, namely philosophy, economics, law, pedagogy, literature, history, science, engineering, agronomy, medicine, management and military science. The literature data sets used in the experiment include isolet, covtype and census_. The specific description of the data set is shown in Table 2.

Experiment No	Number of documents	Number of matrix rows	Number of matrix columns	Matrix size	Computing time
1	72	72	29876	5.7	3
2	728	728	175897	332.6	14
3	7159	7159	746239	17.8	27 minutes and 100 seconds
4	108026	108026	903452	198.6	7 hours 34 minutes 20 seconds

Table 2: Description of algorithm experimental data

In terms of data sources, after analyzing different target sources, it is found that famous university websites, some discipline portals and OA warehouses contain a large number of publicly published academic documents, which can be captured without restrictions. Therefore, it is determined to take university websites, OA warehouses and discipline portals as target sources. In order to make the results more representatives, the conference website and the researcher's home page were also added. The target sites selected in this experiment are shown in Table 3.

No.	Site	Brief Introduction	Type
1	https: //www. stanford.edu	Stanford University website	University website
2	https://www.omicsonline.org	Omnics group website	OA warehousing
3	https://www.acm.org	American Computer Society website	Subject Portal
4	https://webis.de	International Conference pan website	Conference website

Table 3: Document capture target sites

It can be seen from the experimental results that the classification accuracy of naive Bayes has been slightly improved after discretization. The reason is that after discretization, the continuous attributes are mapped into discrete classification attributes, which makes the system more complete, and avoids a potential problem in estimating a posteriori probability from training data to a certain extent: the class conditional probability of attributes is equal to zero, The extreme case that the class posterior probability of the whole class is equal to zero, resulting in classification error or inability to classify. The experimental results show that the classification accuracy of the algorithm can be greatly improved by discretizing the continuous data through the parallel attribute discretization algorithm based on direct.

In the aspect of algorithm execution efficiency, the running time of the two algorithms to deal with data classification tasks of different scales under the

environment of different number of nodes is recorded respectively. The specific running time is shown in Figure 6.
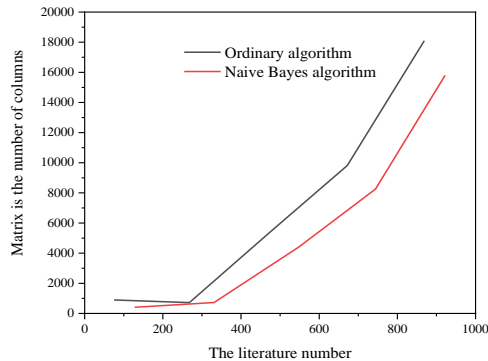

Figure 6: Comparison of algorithm running time

As can be seen from Figure 6, the classification results of all academic documents can be viewed through this system. The year data is not mined at the text level, but directly uses the PDF file metadata (creation date). On the document display page, you can view the title, category, original URL and excerpt of the text of the document. Each interface is equipped with faceted search function to facilitate users' secondary retrieval. The efficiency of these algorithms in terms if run time is calculated and shown in Figure 7.
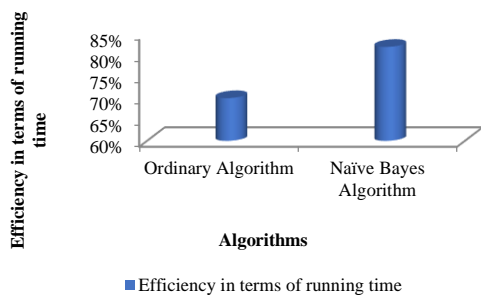
Figure 7: Comparative analysis of the algorithms in terms of efficiency

The Naïve algorithm is much more effective and efficient in terms of complexity. The Naïve Bayes algorithm is 82% efficient and the ordinary algorithm efficiency is 70%.

4   Conclusion

The successful design and implementation of naive Bayesian classification algorithm can not only solve the problems of large memory consumption, slow processing speed and high feature vector dimension in the process of massive document processing, but also enable scientific researchers to effectively obtain and use the documents. At the same time, it also solves the problems of the integration of two heterogeneous ontology libraries and how to apply them in specific fields. The problem is that

the traditional word vector space cannot meet people's needs for semantic classification, semantic navigation and semantic retrieval of massive network information resources due to high dimension and lack of semantics. Therefore, it has academic value and practical significance. The design idea and framework of the system can be directly applied to e-government system, portal website, vertical search engine, digital library website and so on. The main strength of the approach lies in its ability to classify the web documents into the right categories correctly and in zero seconds. The future work of this work can be on combining two classification techniques to increase the accuracy of a web page classification.

References

[1] Li, W. Q. , Li, Y. , Chen, J. , & Hou, C. Y. . (2017). Product functional information based automatic patent classification: method and experimental studies. Information Systems, 67(JUL.), 71-82. https://doi.org/10.1016/j.is.2017.03.007

[2] Agnihotri, D. , Verma, K. , & Tripathi, P. . (2018). An automatic classification of text documents based on correlative association of words. Journal of Intelligent Information Systems, 50(3), 549-572.

[3] Pajic, M. S. , Veinovic, M. , Peric, M. , & Orlic, V. D. . (2020). Modulation order reduction method for improving the performance of amc algorithm based on sixth – order cumulants. IEEE Access, PP(99), 1-1. DOI: 10.1109/ACCESS.2020.3000358

[4] Du, J. H. . (2017). Automatic text classification algorithm based on gauss improved convolutional neural network. Journal of Computational Science, 21(jul.), 195-200.

[5] Y Kumar, Sheoran, M. , Jajoo, G. , & Yadav, S. K. . (2020). Automatic modulation classification based on constellation density using deep learning. IEEE Communications Letters, PP(99), 1-1, DOI: 10.1109/LCOMM.2020.2980840

[6] Dias, K. L. , Pongelupe, M. A. , Caminhas, W. M. , & Errico, L. D. . (2019). An innovative approach for real-time network traffic classification. Computer networks, 158(JUL.20), 143-157, https://doi.org/10.1016/j.comnet.2019.04.004

[7] TajanpureRupalirupalidixit@gmail.comMuddanaAkk alakshmiamuddana@gitam.eduGITAM University,Hyderabad,Telangana,India. (2021). Circular convolution-based feature extraction algorithm for classification of high-dimensional datasets. Journal of Intelligent Systems, 30(1), 1026-1039, https://doi.org/10.1515/jisys-2020-0064

[8] Gaber, A. , Hamdy, A. , Abdelaal, H. M. , Elkattan, A. , & Youness, H. A. . (2021). Automatic classification algorithm for diffused liver diseases based on ultrasound images. IEEE Access, PP(99), 1-1, DOI: 10.1109/ACCESS.2021.3049341.

[9] Lopes, F. , Agnelo, J. , Teixeira, C. A. , Laranjeiro, N. , & Bernardino, J. . (2020). Automating orthogonal defect classification using machine

learning algorithms. Future generation computer systems, 102(Jan.), 932-947, DOI: 10.1109/ACCESS.2021.3049341

[10] An, Y. , Xu, M. , & Shen, C. . (2019). Classification method of teaching resources based on improved knn algorithm. International Journal of Emerging Technologies in Learning (iJET), 14(4), 73-88, https://doi.org/10.3991/ijet.v14i04.10131

[11] Rueda, C. A. , & Ryan, J. P. . (2020). Humpback whale song analysis based on automatic classification performance. The Journal of the Acoustical Society of America, 148(4), 2597-2597, https://doi.org/10.1121/1.5147215

[12] El Kourdi, M., Bensaid, A., & Rachidi, T. E. (2004). Automatic Arabic document categorization based on the Naïve Bayes algorithm. In *proceedings of the Workshop on Computational Approaches to Arabic Script-based Languages* (pp. 51-58), https://dl.acm.org/doi/10.5555/1621804.1621819

[13] Chen, H., & Fu, D. (2018, March). An improved Naive Bayes classifier for large scale text. In *2018 2nd International Conference on Artificial Intelligence: Technologies and Applications (ICAITA 2018)* (pp. 33-36). Atlantis Press, https://doi.org/10.2991/icaita-18.2018.9

[14] Wang, Y., Hodges, J., & Tang, B. (2003, November). Classification of web documents using a naive bayes method. In *Proceedings. 15th IEEE International Conference on Tools with Artificial Intelligence* (pp. 560-564). IEEE, DOI: 10.1109/TAI.2003.1250241

[15] Adetunji, A. B., Oguntoye, J. P., Fenwa, O. D., & Akande, N. O. (2018). Web Document Classification Using Naïve Bayes. *Journal of Advances in Mathematics and Computer Science*, 29(6), 1-11, DOI: https://doi.org/10.48550/arXiv.2006.01715

[16] Dai, Y., & Sun, H. (2014). The naive Bayes text classification algorithm based on rough set in the cloud platform. *Journal of Chemical and Pharmaceutical Research*, 6(7), 1636-1643, https://doi.org/10.1007/s00500-020-05410-9

[17] Koopman, B. , Zuccon, G. , Nguyen, A. , Bergheim, A. , & Grayson, N. . (2015). Automatic icd-10 classification of cancers from free-text death certificates. International journal of medical informatics, 84(11), 956-965, DOI: 10.1016/j.ijmedinf.2015.08.004

[18] Li, K. , & Sidorovskaia, N. . (2019). Detection and classification beaked whale vocalization calls based on unsupervised machine learning algorithm. The Journal of the Acoustical Society of America, 145(3), 1855-1856.

[19] Sharma, A., & Kumar, R. (2019). Computation of the reliable and quickest data path for healthcare services by using service-level agreements and energy constraints. *Arabian Journal for Science and Engineering*, 44(11), 9087-9104, 10.1007/s13369-019-03836

[20] Harakawa, R. , Ogawa, T. , Haseyama, M. , & Akamatsu, T. . (2018). Automatic detection of fish sounds based on multi-stage classification including logistic regression via adaptive feature weighting. The Journal of the Acoustical Society of America, 144(5), 2709-2718, DOI: 10.1121/1.5067373

[21] Hartvigsen, L. , Kongsted, A. , Vach, W. , Salmi, L. R. , & Hestbaek, L. . (2018). Does a diagnostic classification algorithm help to predict the course of low back pain? a study of danish chiropractic patients with one-year follow up. Journal of Orthopaedic and Sports Physical Therapy, 48(11), 1-35, DOI: 10.2519/jospt.2018.8083

[22] Sharma, A., & Kumar, R. (2019). Risk-energy aware service level agreement assessment for computing quickest path in computer networks. *International Journal of Reliability and Safety*, 13(1-2), 96-124.

[23] M Foroutan, & JR Zimbelman. (2017). Semi-automatic mapping of linear-trending bedforms using 'self-organizing maps' algorithm. Geomorphology, 293(PT.A), 156-166.

Heidari, M. , Lakshmivarahan, S. , Mirniaharikandehei, S. , Danala, G. , & Zheng, B. . (2021). Applying a random projection algorithm to optimize machine learning model for breast lesion classification. IEEE Transactions on Biomedical Engineering, PP(99), 1-1, https://doi.org/10.1109/TBME.2021.3054248

[24] Nardini, A. , & Brierley, G. . (2020). Automatic river planform identification by a logical-heuristic algorithm. Geomorphology, 375(1–2), 107558, https://doi.org/10.1016/j.geomorph.2020.107558

[25] Yan, J. , Lin, S. , Kang, S. B. , & Tang, X. . (2015). Change-based image cropping with exclusion and compositional features. International Journal of Computer Vision, 114(1), 74-87, DOI: https://doi.org/10.1007/s11263-015-0801-5

[26] Elsanadily, S. , Mahran, A. , & Elghandour, O. . (2018). Classification-based algorithm for bit-flipping decoding of gldpc codes over awgn channels. IEEE Communications Letters, PP(99), 1-1, DOI: 10.1109/LCOMM.2018.2840146

[27] Sharma, A., Kumar, R., & Bajaj, R. K. (2021). On Energy-constrained quickest path problem in green communication using intuitionistic trapezoidal fuzzy numbers. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, 14(1), 192-200, DOI: https://doi.org/10.2174/2213275911666181025125224

[28] Bahadure, N. B. , Ray, A. K. , & Thethi, H. P. . (2018). Comparative approach of mri-based brain tumor segmentation and classification using genetic algorithm. Journal of Digital Imaging, 31(1), 1-13, DOI: DOI: 10.1007/s10278-018-0050-6

[29] Redzic, M. , Laoudias, C. , & Kyriakides, I. . (2019). Image and wlan bimodal integration for indoor user localization. IEEE Transactions on Mobile Computing, 19(99), 1109-1122, DOI: DOI: 10.1109/TMC.2019.2903044

[30] Zhao, D. , Liu, S. , Yang, X. , Ma, Y. , & Chu, W. . (2021). Research on camouflage recognition in simulated operational environment based on hyperspectral imaging technology. Journal of Spectroscopy, 2021(2), 1-9, DOI: https://doi.org/10.1155/2021/6629661

[31] Poongodi, M., Sharma, A., Vijayakumar, V., Bhardwaj, V., Sharma, A. P., Iqbal, R., & Kumar, R. (2020). Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system. *Computers & Electrical Engineering*, *81*, 106527, DOI: https://doi.org/10.1016/j.compeleceng.2019.106527

[32] Ahmed, I. , Ali, R. , D Guan, Lee, Y. K. , Lee, S. , & Chung, T. C. . (2015). Semi-supervised learning using frequent itemset and ensemble learning for sms classification. Expert Systems with Applications, 42(3), 1065-1073, DOI: https://doi.org/10.1016/j.eswa.2014.08.054

[33] Kumar, C., Singh, A. K., Kumar, P., Singh, R., & Singh, S. (2020). SPIHT-based multiple image watermarking in NSCT domain. *Concurrency and Computation: Practice and Experience*, *32*(1), e4912, DOI: https://doi.org/10.1002/cpe.4912

[34] Dadaneh, B. Z. , Markid, H. Y. , & Zakerolhosseini, A. . (2016). Unsupervised probabilistic feature selection using ant colony optimization. Expert Systems with Applications, 53(Jul.), 27-42, https://doi.org/10.1016/j.eswa.2016.01.021

Big Data Intelligent Collection and Network Failure Analysis Based on Artificial Intelligence

Jun Ding[1], Roobaea Alroobaea[2], Abdullah M. Baqasah[3], Anas Althobaiti[4], Rajan Miglani[5*], Harsimranjit Singh Gill[6]

[1]Anhui Technical College of Industry and Economy, Hefei, Anhui, 230051, China
[2]Department Computer Science, College of Computers and Information Technology, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia
[3]Department of Information Technology, College of Computers and Information Technology, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia
[4]College of Environment and Technology, Bristol University, Bristol Bs1, United Kingdom
[5]Lovely Professional University, Punjab, India
[6]Guru Nanak Dev Engineering College, Ludhiana, India
Emails: [1]dingjun82@163.com, [2]r.robai@tu.edu.sa, [3]a.baqasah@tu.edu.sa, [4]anasuk.tu@gmail.com, [5]rajan.16957@lpu.co.in *corresponding author

To study smart data collection and network error analysis, this paper proposes intelligent data collection and network error analysis based on artificial intelligence. It examines the establishment of an enterprise-level information security situation awareness system and proposes specific information security models, architectures, and implementation methods. By designing and deploying the system, businesses can effectively detect information security threats, receive threats, filter risks, control threats, and comprehensively improve businesses' ability to detect security threats and security attacks. Test results: Through this platform, it is possible to manually intervene in the unknown threat of large data analysis in the system, and professionals can perform a detailed analysis to determine the means, goals and objectives of the attack and restore the complete picture. Intruder through artificial intelligence combined with big data knowledge and intrusion. Dimensional human characteristics. Including similar Trojans and malicious servers with different application forms, encodings, and attack principles, they "track" intruders by their general characteristics, constantly detect unknown threats, and ultimately ensure the accuracy of unknown threat detection, creating a local threat intelligence analytics platform. Practice has shown that the intelligent acquisition of large data by artificial intelligence can effectively analyze network failures.

Povzetek: S pomočjo umetne inteligence je narejena analiza napak v omrežjih in zbiranje podatkov.

1 Introduction

Artificial intelligence belongs to a key branch of computer science. Relying on the essence of intelligence, in order to generate intelligent machines similar to human intelligence, its key research objects are the application systems and technologies that simulate, expand and expand human intelligence. Artificial intelligence technology highly simulates many thinking processes and intelligent behaviors of human beings, providing great convenience for people's daily life, so it has received high attention from all fields of society [1-2]. As shown in Figure 1, the system will investigation and analysis of the whole security device logs and network traffic after discovering offensive behavior, so as to determine the specific degree of behavior, and solve these problems as far as possible. In the enterprise

passive cycle of information security defense system, the vast majority of enterprises will put more energy on the defense process, but ignore the determination and analysis of the attack cause. There is relatively little investment and research in system repair, which is usually passive repair based on patches from the original manufacturer's products. At the same time, enterprises constantly optimize and improve defense policies to improve system defense capabilities and effectively resist external attacks. The defense means of information security has been effectively optimized. Many enterprises have established a network anti-virus, terminal management, security audit, access restrictions, and integrated security systems, such as the leak was found to be able to ensure the safety of the enterprise business reliably, reduce the information security risk to the enterprise, achieve the unity of the enterprise early

be able to ensure the safety of the enterprise business reliably, reduce the information security risk to the enterprise, achieve the unity of the enterprise early warning, unified management and traceability, reduce information risk's influence on the enterprise normal business activities [4].

Intelligent fault diagnosis technology includes fuzzy technology, grey theory, pattern recognition, fault tree analysis, diagnosis expert system and so on. The first four technologies only use logical reasoning knowledge to some extent and partly solve the problems such as fuzzy information, incomplete information, fault classification and location in the diagnosis process, while the diagnosis expert system can take itself as a platform and integrate other diagnosis technologies to form a hybrid intelligent fault diagnosis system. The narrow sense of intelligent diagnosis technology generally refers to expert system. Due to its inherent super adaptability

and learning ability, artificial intelligence has been widely used in many fields and solved many problems that are difficult to be solved by traditional methods. The unique nonlinear adaptive information processing ability of neural network overcomes the defects of traditional artificial intelligence methods in intuition, such as pattern recognition, speech recognition and unstructured information processing, and makes it successfully applied in neural expert system, pattern recognition, intelligent control, combinational optimization, prediction and other fields. The combination of neural network and other traditional methods will promote the continuous development of artificial intelligence and information processing technology.
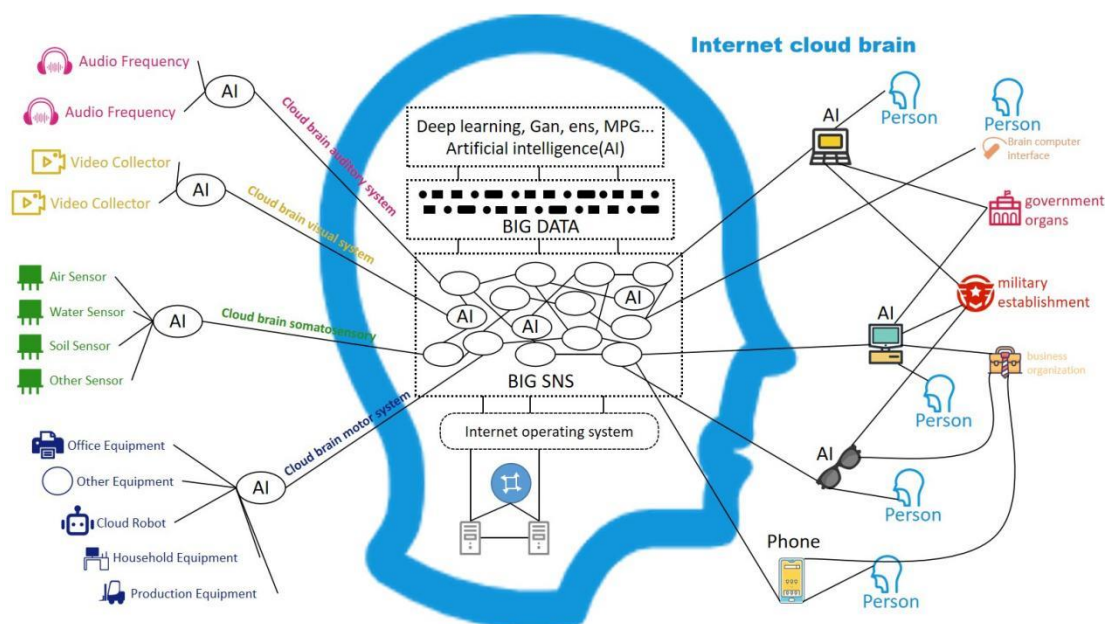

Figure 1 Intelligent collection of big data with artificial intelligence

2 Literature review

Fault diagnosis technology has developed greatly from methods to means, and the emergence, development and penetration of a large number of relevant scientific and technological achievements have also promoted its development. At the same time, because of the rapid development of computer technology, the fault diagnosis technology has unprecedented application value and popularization. However, the penetration of various disciplines only changes its methods and means, and its fundamental purpose is still to obtain and interpret the information of equipment operation state, so as to ensure

the normal operation of equipment and maintenance according to the situation, reduce or eliminate accidents. In all walks of life, due to the application of fault diagnosis technology, not only effectively prevent the occurrence of many serious accidents, but also achieve great economic and social benefits. Cheng, L. proposed a cloud computing based on big data, information preprocessing optimized clustering algorithm and Chinese NLP (natural language processing) sentiment tendency analysis algorithm artificial intelligence network public opinion analysis platform. Speed up the

screening speed of effective information and the speed of public sentiment-oriented analysis; ensure that under the environment of massive network data, timeliness and effectiveness of public opinion monitoring. Finally through the experiment, compared with the traditional statistical big data information analysis system, this method has fast information convergence speed, information analysis is efficient and reliable, especially after doing a good job of classification training in key areas of focus, as the amount of collected data grows, the results of public opinion-oriented analysis are also more accurate [5]. Wang, J. restored the data in combination with the technology of the system, made precise analysis of the data, and then published the specific data so that other enterprises could store and use the data through the enterprise platform [6]. Zhan, J. believes that with the continuous application of new technologies, the means and methods of attack are increasingly hidden and difficult to detect. Covert attack means that can bypass various traditional security detection and defense measures and achieve targeted attacks through careful camouflage, long-term latency and continuous penetration [7]. Mengyuan, H. Constructed artificial intelligence detection technology of malicious code through artificial search engine, based on a large number of samples of malicious software and normal software, searched for information data features existing in different samples, and constructed an effective machine learning model for security scanning of unknown programs [8]. Zhang, Z. Intelligent operation and maintenance mode of transmission network based on artificial intelligence and big data analysis, can save labor costs, save equipment investment and improve network performance, effectively support the company's Internet-based operation transformation, effectively support the market, improve customer perception, it has good application promotion value. With the emergence and development of new technologies such as artificial intelligence, big data, cloud computing and SDN/NFV, traditional operation and maintenance technology of communication operators based on manual methods, it has been unable to meet the needs of cost and efficiency, automation and intelligent operation and maintenance technology has become an inevitable choice [9]. Karim, A. H. Research results in data analysis and visualization technology, it is possible to build an epidemic prevention and control platform based on big data and artificial intelligence technology, the platform can provide timely and accurate epidemic information for government agencies at all levels, and decision-making support for epidemic prevention and control, provide technical

support for the implementation of the major policy of "highlighting key points, overall planning, classification guidance, and district implementation of policies" [10]. Hussein, H. A. T. outlines the basic meaning of network information retrieval, from FTP (File Transfer Protocol) search tools, menu-based search tools, three aspects of keyword-based search tools, analyze the classification of network information retrieval tools, and use this as a basis, put forward the application countermeasures of artificial intelligence in network information retrieval in the era of big data [11]. Raisan, A. Through barcode technology, radio frequency technology, internet of things, global positioning system technology, geographic information system technology, ERP, CRM, wide application of technology such as industrial control system, can quickly collect, process, and analyze data, promote industrial enterprises to realize the interconnection of all links in the production process. Regarding the current status of big data acquisition methods, the main problems in data acquisition methods, analyze the changes and strategies of future acquisition methods, and expounds the trend of change in the way of big data acquisition [12]. Lei, Y. will first introduce big data and artificial intelligence, after analyzing the application of artificial intelligence in computer networks in the current era of big data, in this way, it can be used as a reference for relevant people to communicate [13]. Xia, M. Can use the effective application of artificial intelligence virus detection and killing technology in enterprise information security situation awareness system can realize the effective identification and timely detection and killing of virus, and reduce the damage caused by virus to computer system [14-18].Because these studies have large loopholes, or the detection is not comprehensive enough, this paper proposes a method based on artificial intelligence on the basis of existing studies [19-27]. The design and deployment of the system, effective detection, threat perception, determination and threat risk tracking of information security threats can realize the comprehensive improvement of enterprises and their ability to detect security threats and security attacks [28-35].

3   Introduction to theory and computer network failures

The fault diagnosis of computer network is studied. Trained neural networks can store knowledge about processes and learn directly from historical fault information.

The main work is as follows:

①In the computer network management for computer network failure to select the appropriate data.

②Self-organizing feature mapping (SOM) neural network is used to cluster computer network faults.

③Set appropriate weights for the clustering results and add the sample data to establish the BP neural network model.

④ Computer numerical simulation, and the simulation results are compared.

3.1 Learning algorithm of BP neural network

The jth neuron in the k layer of the BP neural network has the following input and output relationship:

$$y_j^{(k)} = f_j^k \left(\sum_{i-1}^{N_{k-1}} W_{ij}^{(k-1)} y_i^{(k-1)} - \theta_j^{(k)} \right) (k = 1, 2, \cdots M; j = 1, 2, \cdots, N_k)$$

（1）

Among them, the connection weight of the ith node the $k-1$-th layer to this node is $W_{ij}^{(k-1)}$; $\theta_j^{(k)}$ is the function of the node; N_k is the number of nodes in layer k; M is the total number of layers. $f_j^{(k)}$ is taken as the Sigmoid function $f(x) = \dfrac{1}{(1 + e^{-x})}$.

BP neural network uses error back propagation algorithm for learning, and the adjustment of weights is as follows:

$$W_{ij}^{(k-1)}(t+1) = W_{ij}^{(k-1)}(t) + \eta \sum_{h-1}^{I} \delta_{hj}^{(k)} y_{hj}^{(k-1)} \quad (2)$$

Among them, I is the total number of samples, $0 < \eta < 1$ is the learning step size, and $\delta_{hj}^{(k)}$ is the error transmission term.

For the output layer:
$$\delta_{hj}^{(M)} = \left(\hat{y}_{hj}^{(M)} - y_{hj}^{(M)} \right) f_j \left(y_{hj}^{(M)} \right) \quad (3)$$

For other layers:
$$\delta_{hj}^{(k)} = f_j \left(y_{hj}^{(k)} \right) \sum_{i=1}^{N_{k+1}} \delta_{hj}^{(k+1)} W_{hj}^{(k)}(t) \quad (4)$$

The output error ε_1 of the network is calculated as follows:

$$\varepsilon_1 = \sum_{h-1}^{I} \sum_{j-1}^{N_M} \left(\hat{y}_{hj}^{(M)} - y_{hj}^{(M)} \right)^2 \quad (5)$$

3.3 LM and fuzzy theory

If $\varepsilon_1 > \varepsilon$ (ε is the error of the preselection setting), then continue to the next round of learning to adjust the weights, otherwise, the network stops learning. The network formed by the weight of W_{ij} after learning can achieve the desired output within the error range set by ε [36-38].

3.2 SOM network implementation process

The algorithm process of SOM network learning is

①Initialization. For N input neuron bands, the connection weights are small. Select the set S_j of output neurons j "adjacent neurons". Among them, $S_j(0)$ represents the "adjacent neuron" set of neuron j at time t=0, $S_j(t)$ represents the set of "adjacent neurons" at time t. Area $S_j(t)$ keeps shrinking with time [16].

②Provide a new input mode X.

③j is distance from the input and the output:

$$d_j = \| X - W_j \| = \sqrt{\sum_{i=1}^{N} \left[x_i(t) - w_{ij}(t) \right]^2} \quad (6)$$

And find a neuron j^* with the smallest distance, that is, a certain unit k is determined, so that for any j, there is $d_k = \min(d_j)$.

④Give a surrounding neighborhood $S_k(t)$.

⑤Correct the weight of output neuron j^* "adjacent neuron":

$$w_{ij}(r+1) = w_{ij}(t) + \eta(t) \left[x_i(t) - w_{ij}(t) \right] \quad (7)$$

η Gain item, gradually reduced to 0

$$\eta(t) = \frac{1}{t} \, or \, \eta(t) = 0.2 \left(1 - \frac{t}{10000} \right) \quad (8)$$

⑥Calculation output o_k:

$$o_k = f \left(\min_j \| X - W_j \| \right) \quad (9)$$

Among them, $f(.)$ is generally a 0-1 function or other non-linear function.

There are also some different strategies, such as BP algorithm combined with other techniques such as fuzzy theory or genetic algorithm [39-43].

Let x_k be the approximate value of the k-th iteration, and F be the objective function

$$H(s) = \sum_{t-1}^{M} w_i^2(r) = w^R(r)v(s) \quad (10)$$

$v(x) = (v_1(x), v_2(x), ..., v_N(x))^T$ in the above formula. Then the LM algorithm is:

$$\Delta s_k = s_{k+1} - s_k = -[H(s_k) + \mu_k I]^{-1} J^T(s_k)v(s_k)$$
（11）

J is the Jacobi matrix of F:

$$J(x) = \begin{bmatrix} \dfrac{\partial v_1(x)}{\partial x_1} & \dfrac{\partial v_1(x)}{\partial x_2} & \cdots & \dfrac{\partial v_1(x)}{\partial x_n} \\ \dfrac{\partial v_2(x)}{\partial x_1} & \dfrac{\partial v_2(x)}{\partial x_2} & \cdots & \dfrac{\partial v_2(x)}{\partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \dfrac{\partial v_N(x)}{\partial x_1} & \dfrac{\partial v_N(x)}{\partial x_2} & \cdots & \dfrac{\partial v_N(x)}{\partial x_n} \end{bmatrix} \quad (12)$$

H is the approximate matrix of the Hesse matrix of F, which is taken as:

$$H(x_k) = J^T(x_k)J(x_k) \quad (13)$$

μ_k is greater than 0, it will be gradually used in LM algorithm. When equal to 0, it approaches Gaussian Newton algorithm. At maximum, LM drops linearly[44-51].

4   Results and Discussion

Figure 2 shows the BP neural network training process without any improvement. After adjusting the sample weight, the sampling training process using the improved LM algorithm is shown in Figure 3. From Figure 4, when the unimproved neural network is trained 100 times, there is still a big gap from the error of 10^{-2}, and the neural network combined by the SOM method and the LM method, convergence is reached after 20 trainings [52-58]. Using the above combination of SOM method and LM method, the training process is shown in Figures 4, 5 and Figures 6, 7. It can be seen from Figures 4, 5 and Figures 6, 7 and the continuous era development and the rapid progress of computer technology, in recent years, the infection types of computer virus emerge in

endlessly, which seriously threatens the normal work of computer system. Through the application of artificial intelligence virus detection technology in the big data technology can improve the perception ability of enterprises to the virus, using a variety of virus location methods can improve the efficiency of the existing virus

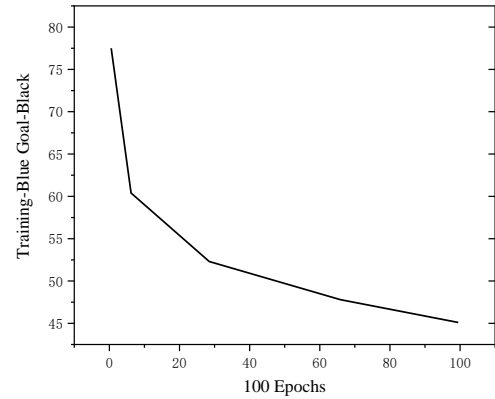search, improve the accuracy and scientific nature of virus detection.

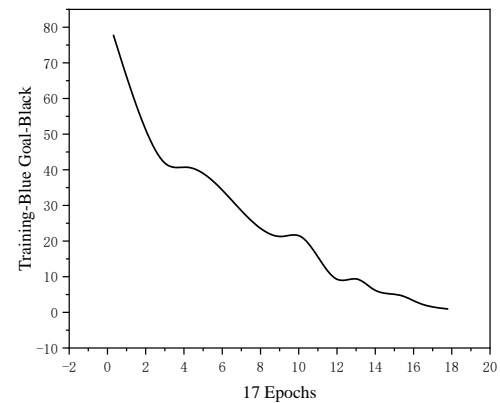Figure 2 Training process of the combined algorithm
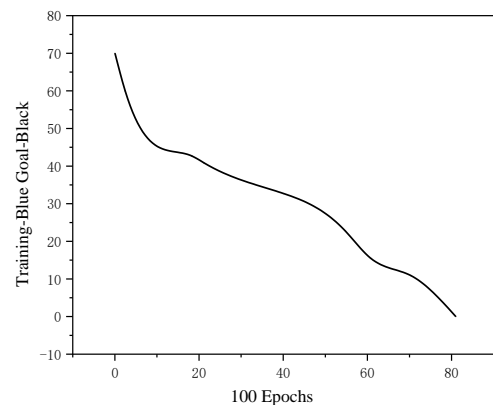

Figure 3 Training of the original algorithm

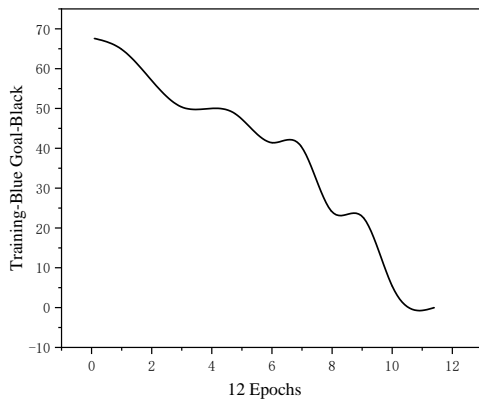Figure 4 Combined algorithm training process

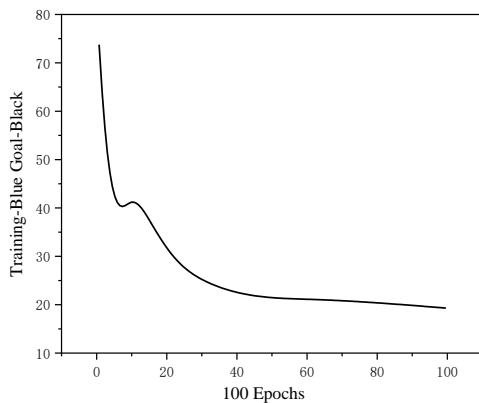Figure 5 Training process of the original algorithm
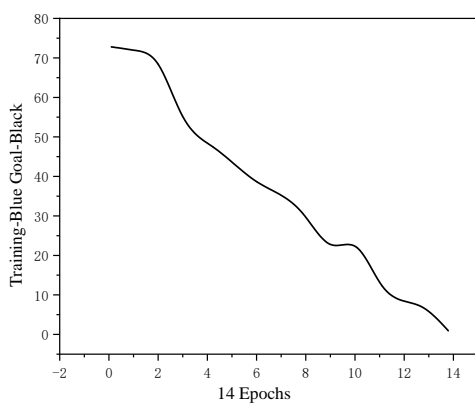
Figure 6 Combine the training process of the algorithm

Figure 7 Training process of the original algorithm

5 Conclusions

Although neural network has achieved good results in the field of fault diagnosis, the structure and training times of neural network have great influence on the fault

diagnosis ability of neural network. An insufficiently designed neural network may have poor performance in fault diagnosis. Based on large data and information security situational awareness of artificial intelligent technology for enterprise digital information security and the normal operation of the enterprise has the vital role, because the quantity of a security threat facing the enterprises growth trends, so enterprises must adopt more effective measures and means to show complete these threats, and take corresponding measures to solve it. The effective application of big data and artificial intelligence technology can improve the accuracy and accuracy of information processing, comprehensively assess the security risk status of information system, and realize the safe and orderly operation of enterprises. The main research direction in the future probably has two aspects: (1) better, fast design of the optimal neural network structure, in the fault diagnosis to achieve the best effect. If the neural network can be designed more scientifically and objectively, and can be carved scientifically with mathematical language, the neural network model can be established better and faster, laying a foundation for further research. (2) Further application of artificial intelligence. With the further development of computer science, it is believed that there will be more and better artificial intelligence models. If these intelligent algorithms are applied to the field, it is believed that in the near future there will be better results applied in the field of fault diagnosis, and further improve the scientific and intelligent fault diagnosis.

Acknowledgement

References:

[1] Guo, A. , & Yuan, C. . (2021). Network intelligent control and traffic optimization based on sdn and artificial intelligence. Electronics, 10(6), 700. https://doi.org/10.3390/electronics10060700

[2] Hu, H. , Bo, T. , Gong, X. , Wei, W. , & Wang, H. . (2017). Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks. IEEE Transactions on Industrial Informatics, 13(4), 2106-2116. https://doi.org/10.1109/tii.2017.2683528

[3] Hu, J. , Zhang, L. , Cai, Z. , & Wang, Y. . (2015). An intelligent fault diagnosis system for process plant

using a functional hazop and dbn integrated methodology. Engineering Applications of Artificial Intelligence, 45(OCT.), 119-135. https://doi.org/10.1016/j.engappai.2015.06.010

[4] Juying, Dai, Jian, Tang, Shuzhan, & Huang, et al. (2019). Signal-based intelligent hydraulic fault diagnosis methods: review and prospects. Chinese Journal of Mechanical Engineering, v.32(05), 11-32. https://doi.org/10.1186/s10033-019-0388-9

[5] Cheng, L. , & Yu, T. . (2018). Dissolved gas analysis principle-based intelligent approaches to fault diagnosis and decision making for large oil-immersed power transformers: a survey. Energies, 11(4), 913. https://doi.org/10.3390/en11040913

[6] Wang, J. , Wang, D. , Wang, S. , Li, W. , & Song, K. . (2021). Fault diagnosis of bearings based on multi-sensor information fusion and 2d convolutional neural network. IEEE Access, PP(99), 1-1. https://doi.org/10.1109/access.2021.3056767

[7] Zhan, J. , Wang, R. , Yi, L. , Wang, Y. , & Xie, Z. . (2019). Health assessment methods for wind turbines based on power prediction and mahalanobis distance. International Journal of Pattern Recognition & Artificial Intelligence, 33(2), 1951001.1-1951001.17. https://doi.org/10.1142/s0218001419510017

[8] Mengyuan, H. , D Qiaolin, Shutao, Z. , & Yao, W. . (2017). Research of circuit breaker intelligent fault diagnosis method based on double clustering. Ieice Electronics Express, 14(17), 20170463-20170463. https://doi.org/10.1587/elex.14.20170463

[9] Zhang, Z. . (2020). Big data analysis with artificial intelligence technology based on machine learning algorithm. Journal of Intelligent and Fuzzy Systems, 39(5), 1-8. https://doi.org/10.3233/jifs-191265

[10] Karim, A. H. , Hassaan, G. A. , & Hegazy, A. . (2021). Artificial neural network based intelligent fault identification of rotating machinery. International Journal of Web Engineering and Technology, 2(No 6,), 26-39.

[11] Hussein, H. A. T. , Ammar, M. E. , & Hassan, M. A. M. . (2017). Three phase induction motors stator turns fault analysis based on artificial intelligence. International Journal of System Dynamics Applications, 6(3), 1-19. https://doi.org/10.4018/ijsda.2017070101

[12] Raisan, A. , Yaacob, M. M. , & Alsaedi, M. A. . (2015). Faults diagnosis and assessment of transformer

insulation oil quality: intelligent methods based on dissolved gas analysis a-review. International Journal of Engineering & Technology, 4(1), 54. https://doi.org/10.14419/ijet.v4i1.3941

[13] Lei, Y. , Jia, F. , Lin, J. , Xing, S. , & Ding, S. X. . (2016). An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. IEEE Transactions on Industrial Electronics, 63(5), 3137-3147. https://doi.org/10.1109/tie.2016.2519325

[14] Xia, M. , Li, T. , Liu, L. , Xu, L. , & Silva, C. . (2017). An intelligent fault diagnosis approach with unsupervised feature learning by stacked denoising autoencoder. IET Science Measurement ? Technology, 11(6), 687-695. https://doi.org/10.1049/iet-smt.2016.0423

[15] E. -M. Amhoud, M. Chafii, A. Nimr and G. Fettweis, "OFDM with Index Modulation in Orbital Angular Momentum Multiplexed Free Space Optical Links," 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), 2021, pp. 1-5, doi: 10.1109/VTC2021-Spring51267.2021.9448928. https://doi.org/10.1109/vtc2021-spring51267.2021.9448928

[16] Gill, H. S., Singh, T., Kaur, B., Gaba, G. S., Masud, M., & Baz, M. (2021). A Metaheuristic Approach to Secure Multimedia Big Data for IoT-Based Smart City Applications. *Wireless Communications and Mobile Computing*, *2021*. https://doi.org/10.1155/2021/7147940

[17] Kumar, A., Sehgal, V. K., Dhiman, G., Vimal, S., Sharma, A., & Park, S. (2021). Mobile networks-on-chip mapping algorithms for optimization of latency and energy consumption. Mobile Networks and Applications, 1-15. https://doi.org/10.1007/s11036-021-01827-0

[18] Boguszewicz, C., Boguszewicz, M., Iqbal, Z., Khan, S., Gaba, G., Suresh, A., & Pervaiz, B. The Fourth Industrial Revolution-Cyberspace Mental Wellbeing: Harnessing Science & Technology for Humanity.

[19] E. Amhoud, G. R. Othman and Y. Jaouën, "Concatenation of Space-Time Coding and FEC for Few-Mode Fiber Systems," in IEEE Photonics Technology Letters, vol. 29, no. 7, pp. 603-606, 1 April1, 2017, doi: 10.1109/LPT.2017.2675919. https://doi.org/10.1109/lpt.2017.2675919

[20] E. -M. Amhoud et al., "Experimental Demonstration of Space-Time Coding for MDL Mitigation in Few-Mode Fiber Transmission Systems," 2017 European Conference on Optical Communication (ECOC), 2017, pp. 1-3, doi: 10.1109/ECOC.2017.8345841.
https://doi.org/10.1109/ecoc.2017.8345841

[21] Gaba, G.S. (2021). *Privacy-Preserving Authentication and Key Exchange Mechanisms in Internet of Things Applications* (Doctoral Dissertation, Lovely Professional University Punjab).

[22] Choudhary, K., & Gaba, G. S. (2021). Artificial intelligence and machine learning aided blockchain systems to address security vulnerabilities and threats in the industrial Internet of things. *Intelligent Wireless Communications*, 329.
https://doi.org/10.1049/pbte094e_ch13

[23] Zerhouni, K., Amhoud, E. M., & Chafii, M. (2021). Filtered Multicarrier Waveforms Classification: A Deep Learning-Based Approach. IEEE Access, 9, 69426-69438.
https://doi.org/10.1109/access.2021.3078252

[24] Gaba, G. S., Kumar, G., Monga, H., Kim, T. H., Liyanage, M., & Kumar, P. (2020). Robust and lightweight key exchange (LKE) protocol for industry 4.0. *IEEE Access*, *8*, 132808-132824.
https://doi.org/10.1109/access.2020.3010302

[25] Sharma, A., & Kumar, N. (2021). Third eye: an intelligent and secure route planning scheme for critical services provisions in internet of vehicles environment. IEEE Systems Journal.
https://doi.org/10.1109/jsyst.2021.3052072

[26] Kumar, P., & Gaba, G. S. (2020). Biometric-based robust access control model for industrial internet of things applications. *IoT Security: Advances in Authentication*, 133-142.
https://doi.org/10.1002/9781119527978.ch7

[27] M. Hedabou. Cryptography for addressing Cloud Computing Security, Privacy and Trust Issues. Book on Computer and Cyber Security: Principles, Algorithm, Applications and Perspective. CRC Press, Francis and Taylor Publisher. USA, 2018.
https://doi.org/10.1201/9780429424878-11

[28] Z. Iggaramen, M. Hedabou. FADETPM: Novel approach of file assured deletion based on trusted platform module. In Lecture Notes in Networks and Systems, vol. 49, pp. 49-59. Springer Verlag, 2017.

https://doi.org/10.1007/978-3-319-97719-5_4

[29] Azougaghe, M. Hedabou, M. Belkasmi. An Electronic Voting System Based On Homomorphic Encryption and Prime Numbers. In International Conference On Information Assurance and Security. Marrakech 2015.
https://doi.org/10.1109/isias.2015.7492759

[30] Bentajer, M. Hedabou. AN IBE-Based Design For Assured Deletion In Cloud Storage. In Journal of Cryptologia vol 141, pp. 559-564. Springer-Verlag, 20119.
https://doi.org/10.1080/01611194.2018.1549123

[31] Gaba, G. S., Kumar, G., Monga, H., Kim, T. H., & Kumar, P. (2020). Robust and lightweight mutual authentication scheme in distributed smart environments. *IEEE Access*, *8*, 69722-69733. https://doi.org/10.1109/access.2020.2986480

[32] M. Hedabou. Some Ways to secure elliptic curves cryptosystems. In Journal of Advances in Cliford Algebras, Vol 18, pp 677-688, 2008.
https://doi.org/10.1007/s00006-008-0093-8

[33] Gaba, G. S., Kumar, G., Kim, T. H., Monga, H., & Kumar, P. (2021). Secure device-to-device communications for 5g enabled internet of things applications. *Computer Communications*, *169*, 114-128.
https://doi.org/10.1016/j.comcom.2021.01.010

[34] Sharma, A., Podoplelova, E., Shapovalov, G., Tselykh, A., & Tselykh, A. (2021). Sustainable Smart Cities: Convergence of Artificial Intelligence and Blockchain. Sustainability, 13(23), 13076.
https://doi.org/10.3390/su132313076

[35] Bentajer, M. Hedabou,K. Abouelmehdi, S. ELFEZAZI. CS-IBE : A Data Confidentiality System in Public Cloud Storage System. In Procedia Computer Science vol 141, pp. 559-564. Elsevier, 2018.
https://doi.org/10.1016/j.procs.2018.10.126

[36] Azougaghe, M. Hedabou, O. Oualhaj, M Belkasmi, A. Kobbane. Many-to -One matching game towards secure virtual machine migrating in cloud computing. International Conference on Advanced Communication System and Information Security. Marrakech, 2016.
https://doi.org/10.1109/acosis.2016.7843922

[37] Masud, M., Gaba, G. S., Choudhary, K., Hossain, M. S., Alhamid, M. F., & Muhammad, G. (2021). Lightweight and anonymity-preserving user authentication scheme for IoT-based healthcare. *IEEE Internet of Things Journal*. https://doi.org/10.1109/jiot.2021.3080461

[38] Sharma, A., Singh, P. K., Sharma, A., & Kumar, R. (2019). An efficient architecture for the accurate detection and monitoring of an event through the sky. Computer Communications, 148, 115-128. https://doi.org/10.1016/j.comcom.2019.09.009

[39] Masud, M., Gaba, G. S., Choudhary, K., Alroobaea, R., & Hossain, M. S. (2021). A robust and lightweight secure access scheme for cloud based E-healthcare services. *Peer-to-peer Networking and Applications*, *14*(5), 3043-3057. https://doi.org/10.1007/s12083-021-01162-x

[40] M. Hedabou. A Frobenius Map Approach for an Efficient and Secure Multiplication on Koblitz curves. International Journal of Network Security, Vol. 3, N. 2, PP.233-237. 2006.

[41] Sharma, A., Georgi, M., Tregubenko, M., Tselykh, A., & Tselykh, A. (2022). Enabling Smart Agriculture by Implementing Artificial Intelligence and Embedded Sensing. Computers & Industrial Engineering, 107936. https://doi.org/10.1016/j.cie.2022.107936

[42] H. Boukhriss, M. Hedabou, A. Azougaghe New Technique of Localization a Targeted Virtual. In Proceedings of the 5th International Workshop on Codes, Cryptography and Communication Systems, El Jadida November 27-28, 2014. https://doi.org/10.1109/wcccs.2014.7107907

[43] Liu, Z. , Dai, C. , Hu, K. , & He, S. . (2016). A new search algorithm of mbd based on spider web and its application in power distribution network fault diagnosis. International Journal of Artificial Intelligence Tools, 25(02), 1650002. https://doi.org/10.1142/s0218213016500020

[44] Wang, X. , Su, Y. , Li, Q. , & Han, F. . (2021). Research on intelligent operation and maintenance management method of enterprise it. Journal of Physics: Conference Series, 1732(1), 012059 (7pp). https://doi.org/10.1088/1742-6596/1732/1/012059

[45] Yang, Z. , & Yin, R. . (2017). Design and research of electronic circuit fault diagnosis based on artificial intelligence. Revista de la Facultad de Ingenieria, 32(12), 766-772.

[46] Qiang, L. I. , Wenbin, W. , & Xue, L. . (2015). Intelligent recognition of axis orbits with fish-based algorithms and neural networks with mentors. Shuili Fadian Xuebao/Journal of Hydroelectric Engineering, 34(6), 191-196.

[47] Stéfano Frizzo Stefenon, Silva, M. C. , Bertol, D. W. , Meyer, L. H. , & Nied, A. . (2019). Fault diagnosis of insulators from ultrasound detection using neural networks. Journal of Intelligent and Fuzzy Systems, 37(5), 6655-6664. https://doi.org/10.3233/jifs-190013

[48] Chen, L. , Lan, S. , & Jiang, S. . (2019). Elevators fault diagnosis based on artificial intelligence. Journal of Physics: Conference Series, 1345(4), 042024 (10pp). https://doi.org/10.1088/1742-6596/1345/4/042024

[49] Bode, G. , Thul, S. , Baranski, M. , & D Müller. (2020). Real-world application of machine-learning-based fault detection trained with experimental data. Energy, 198(May1), 117323.1-117323.8. https://doi.org/10.1016/j.energy.2020.117323

[50] Noureldeen, O. , Hamdan, I. , & Hassanin, B. . (2019). Design of advanced artificial intelligence protection technique based on low voltage ride-through grid code for large-scale wind farm generators: a case study in egypt. SN Applied Sciences, 1(6), 515-. https://doi.org/10.1007/s42452-019-0538-9

[51] Liu, Q. , & Huang, Z. . (2020). Research on intelligent prevention and control of covid-19 in china's urban rail transit based on artificial intelligence and big data. Journal of Intelligent and Fuzzy Systems, 39(21), 1-6. https://doi.org/10.3233/jifs-189307

[52] Xu, K. , Li, S. , Li, R. , Lu, J. , & Zeng, M. . (2021). Deep domain adversarial method with central moment discrepancy for intelligent transfer fault diagnosis. Measurement Science and Technology, 32(12), 124005 (16pp). https://doi.org/10.1088/1361-6501/ac20f1

[53] Samara, S. , & Natsheh, E. . (2020). Intelligent pv panels fault diagnosis method based on narx network and linguistic fuzzy rule-based systems. Sustainability, 12(5), 2011. https://doi.org/10.3390/su12052011

[54] Lee, K. , Han, S. , Pham, V. H. , Cho, S. , & Lee, S. W. . (2021). Multi-objective instance weighting-based

deep transfer learning network for intelligent fault diagnosis. Applied Sciences, 11(5), 2370.
https://doi.org/10.3390/app11052370

[55] Li, S. , & Zhou, D. . (2016). Study on a new fault diagnosis method based on combining intelligent technologies. International Journal of Multimedia and Ubiquitous Engineering, 11(6), 61-72.
https://doi.org/10.14257/ijmue.2016.11.6.06

[56] P. Ahmad et al., "MH UNet: A Multi-Scale Hierarchical Based Architecture for Medical Image Segmentation," in IEEE Access, vol. 9, pp. 148384-148408, 2021, doi: 10.1109/ACCESS.2021.3122543.

https://doi.org/10.1109/access.2021.3122543

[57] M. A. Razzaq et al., "The 3-Axis Scalable Service-Cloud Resource Modeling for Burst Prediction Under Smart Campus Scenario," in IEEE Access, vol. 9, pp. 116927-116941, 2021,
https://doi.org/10.1109/access.2021.3105539

[58] Abbas, A., Alroobaea, R., Krichen, M. *et al.* Blockchain-assisted secured data management framework for health information analysis based on Internet of Medical Things. *Pers Ubiquit Comput* (2021).
https://doi.org/10.1007/s00779-021-01583-8

Intelligent Analysis and Processing Technology of Big Data Based on Clustering Algorithm

Zheng Zheng[1], Fukai Cao[1*], Song Gao[2], Amit Sharma[3]
[1]Jitang College, North China University of Science and Technology, Tangshan,063210, China
[2]Tangshan Power Supply Company, State Grid Jibei Electric Power Co., Ltd, Tangshan,063000, China
[3]Southern Federal University, Russia
Emails: zhengzheng873@163.com, fukaicao@126.com, songgao56@163.com, amit.amitsharma90@gmail.com

An attribute category clustering method based on hierarchical clustering is proposed in order to study the big data intelligent analysis and processing technology. The proposed model combines the attribute categories with similar fault type distribution, reduces the data dimension, and binarizes it. To address the problem of more missing values of continuous data, a data completion method based on attribute distribution function is adopted. Through the perspective of selection and estimation of project unit price in construction enterprises, this paper summarizes the data mining process facing the characteristics of project cost data, and puts forward the method of analyzing and processing project cost data based on clustering algorithm. Finally, the processed data sets are subjected to bottom-up hierarchical clustering analysis, and finally the ideal analysis results can be obtained. The experimental results show that the preprocessing method based on attribute clustering proposed in this paper can effectively merge attributes, reduce the dimension after binary transformation and effectively reduce the amount of data under the condition of ensuring data information.

Povzetek: S hierarhičnim gručenjem je narejena inteligentna analiza velikih podatkov.

1 Introduction

The hidden value of big data promotes the derivation of big data mining technology and methods. Big data mining is to mine valuable knowledge for data processing through massive multiple data sources. Therefore, how to quickly and accurately mine valuable knowledge through big data has attracted much attention. In fact, data mining is also a decision support process. Its common methods mainly include classification, clustering, prediction, regression analysis, association rules and so on. Clustering is the most key technology. Big data is unstructured data, which is difficult and large in processing and analysis, making the structural analysis mode too complicated, and the traditional data analysis cannot effectively process, mine and analyze as shown in Figure 1 [1]. The classical methods of cluster analysis can be summarized as: partition method, hierarchical method, density-based method, grid-based method, model-based method, neural network method based on computational intelligence, evolutionary computing method, fuzzy method and so on, as well as the semi supervised clustering method which has attracted much attention at present. Recently, the new cluster integration method has rapidly become a new research hotspot of cluster analysis. The purpose of clustering integration is to fuse the results from multiple clustering algorithms to obtain higher quality and robust clustering results. The method based on graph theory is one of the fast-

developing methods recently. It is a method to realize clustering by using the principles of graph theory and graphics. Compared with traditional algorithms, this algorithm can deal with more complex cluster structures, such as nonconvex structures, and can converge to the global optimal solution [2].

In recent years, with the rapid development of network information technology, the era of big data has come and penetrated into many fields. There are more and more big data application research for specific professional fields. However, for the field of project cost, this aspect has always been a blank. Every day, with the help of the Internet and various project cost systems, a large number of project cost data are generated, but there is no scientific and accurate processing method to process it, so that it is lost in vain. The acquisition and transmission of project cost information still rely on the traditional way, and the timeliness and accuracy cannot meet the needs of today's project management field [3]. To process and mine these huge project cost information data and provide basis and reference for the decision-making of project management process, it is not enough to rely on manual processing technology. We should innovate and apply data mining technology to make full use of the value of massive project cost data, so as to promote the rapid and healthy development of the industry.
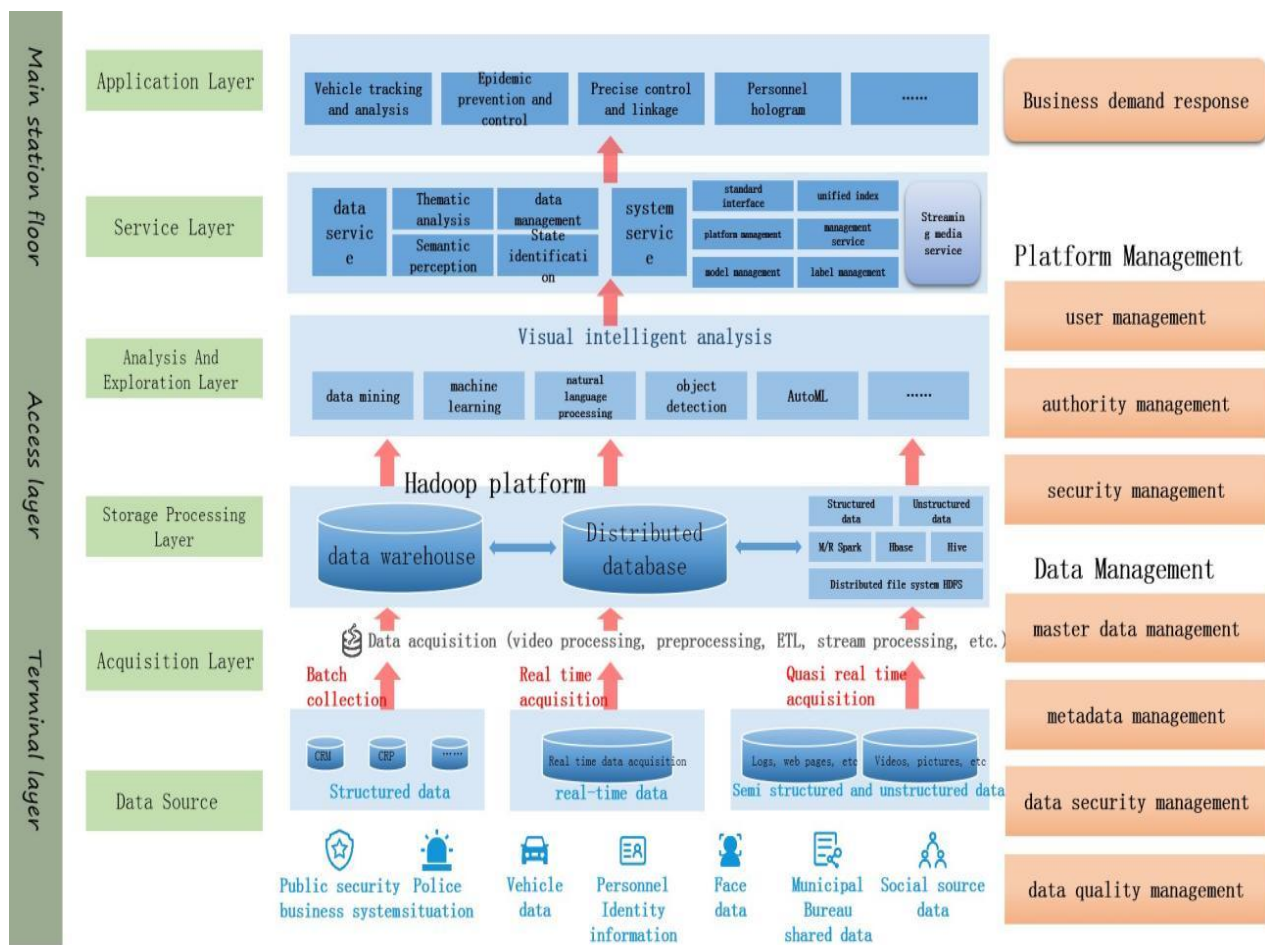
Figure 1: Big data intelligent analysis and processing technology

The rest of manuscript is organized as the most recent work done is discussed in Section 2. The research methodology, optimization of clustering algorithm, complexity and project acquisition is presented in Section 3. Results and analysis of the proposed model is discussed in Section 4 which is followed by the conclusion in Section 5.

2 Related work

In this section various state-of-the-art work in the field of big data processing based on clustering algorithm is presented.

Zhu *et al.* proposed an initial clustering center selection method based on point density, and processed outliers specially [4]. Ser *et al.* proposed an improved algorithm to determine the optimal cluster number k by calculating the contour coefficient of each object in the cluster under different K values, and determine the initial cluster center by hierarchical aggregation method [5]. Wu proposed a clustering method based on patent technology efficacy matrix. This method uses K-means to cluster by calculating the similarity of technology, and achieves good results. K-medoids and PAM algorithms are very effective for small data sets, but they do not have good scalability for large data sets [6]. Duan and

Wang proposed a new heuristic search algorithm clarans algorithm based on PAM [7]. The algorithm finds the center point of the representative cluster by random search of the graph. Clarans algorithm is the first clustering algorithm successfully applied in the field of spatial data mining. It overcomes the shortcomings that other classical clustering algorithms cannot deal with large-scale data sets, but it still fails to solve the problem of low execution efficiency. Its time complexity is 2O (KN). In order to speed up the execution speed of the algorithm, the parallel clarans algorithm based on PVM mechanism proposed by Xing and Li effectively improves the speed of the algorithm [8]. In the artificial neural network, Cai applied the classical hierarchical clustering algorithm and partition algorithm to cluster SOM, which aims to reduce the computational complexity of the classical clustering method [9]. In addition, in terms of network application: Xu *et al.* proposed a three-dimensional facial expression clustering method based on network, which overcomes the shortcomings of limited information contained in data and sharp decline in recognition performance in the case of two-dimensional facial expression recognition [10]. In terms of project cost, Li *et al.* others established the power grid cost management method system and the construction framework of cost analysis information platform under the big data environment [11]. Shi and

Zhu designed the cost management system of mine engineering construction project based on cost data [12]. Wendong *et al.* put forward the statistics and analysis method of project cost information data under the background of big data, and constructed the statistical calculation model of project cost information data [13]. The evolution of artificial intelligence and Internet of Things is considered for several industrial applications and contributing towards social life [14-17].

3 Research methods

This section includes the project design process, structural seismic analysis and detailed modeling steps of proposed design.

As unstructured data, big data is difficult to be characterized by two-dimensional logic table of database. The multi-dimensional de clustering analysis algorithm shows the hidden structure of observation variables through the Bayesian network model structure, and constructs the logical correlation between leaf nodes and other nodes. In this model, multiple hidden variables are allowed to exist, corresponding to the corresponding data clustering methods. Based on the probability dependence between random variables, the multi-dimensional de reunion class analysis algorithm analyzes unstructured data, and quantitatively describes the reasonable distribution with the conditional concept as the carrier. The specific flow of data processing is as follows:

Data preprocessing, that is, data cleaning, avoiding noise and solving the problem of data loss. During data processing, discrete the continuous values in the attribute and convert the data. The data result set and test training set are studied, and the data set is divided into two parts: data result set and test training set. The classifier is constructed by classification algorithm. Through the test set, the accuracy evaluation mode is selected to evaluate the classifier. The classifier that meets the accuracy standard is applied in practice, otherwise it will be modified.
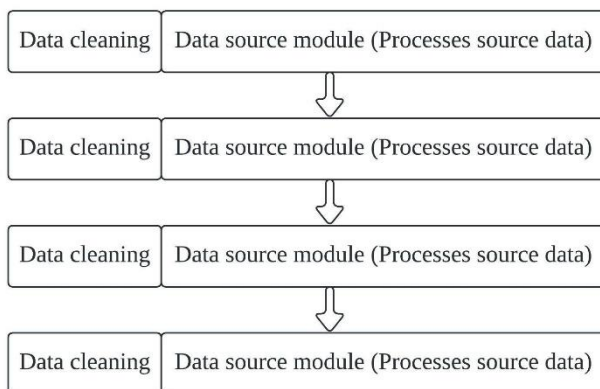

Figure 2: Data processing flow of cluster analysis model

Word segmentation and document vectorization processing, reorganize the continuous word sequence according to the established norms to form the word sequence. In order to transform the document after word

segmentation into a pattern that can be recognized and processed by computer, it is necessary to quantify the word features as the feature vector, which is currently processed by vector space model. Feature selection and multi-dimensional cluster analysis, word features will lead to a certain sparsity and high dimensionality in the document vector feature space, so an effective feature selection method is selected to reduce the dimension of the feature space and further improve the classification efficiency and accuracy [18]. The detailed data processing steps of the analysis model are shown in Figure 2.

Clearly build a functional model for the classification process of big data and unstructured data. The problem can be described as a given data set and category set which is evaluated using Equation 1 and 2.

$$F = \left\{ F_1, F_2, F_3, F_4, \cdots\cdots, F_m \right\} \tag{1}$$

$$G = \left\{ G_1, G_2, G_3, \cdots\cdots, G_m \right\} \tag{2}$$

The classification problem is to clarify the function mapping to make the data items of the data set map to the corresponding categories. Given the big data variable set, the variable takes the parent node set as the carrier, the carrier correlation between nodes can be characterized by a directed graph, that is, for each variable, it can be characterized as a node, and each node guides a directed edge from each directed node of the parent node set to enter the variable. Suppose that the variables of Bayesian network are a and b respectively, and X is the node set without a and b. once z separates a and b, the conditions remain independent based on a given z. The so-called isolation and conditional independence show the close relationship between the graph theory side and probability theory side of Bayesian network. Set to classify objects based on the evidence provided by the feature vector, then:

$$e\left(\frac{vj}{x}\right) > e\left(\frac{vi}{x}\right) i \neq j \tag{3}$$

$$e\left(\frac{x}{vj}\right) e(vj) > e\left(\frac{x}{vi}\right) e(vi) \tag{4}$$

Decision rules are likelihood test rules which are evaluated using Equation 3 and 4. Bayesian network reasoning, through probability decomposition, reduces the reasoning complexity to localize the operation. Through the edge processing and analysis of the elimination process, the decision rules can be tested by likelihood rate for all given large data sets to obtain the minimum error probability calculation samples [19].

3.1   Optimization of clustering algorithm

Based on the function model, an optimized clustering algorithm is constructed to divide the overall big data into multiple data intervals, which are stored through multiple files, and each file represents the corresponding interval. After scanning and comparing all the data, divide them into multiple sections, and sort and remove multiple files. The data quantity of each file is $1M$ and $2M$ respectively. After the data is de duplicated, cluster analysis is carried out and Bayesian formula is used to calculate, that is calculated in Equation 5.

$$BIC\left(\frac{N}{F}\right) = Max\,\alpha\log e\left(\frac{N}{F},\alpha\right) - f(n)\log\frac{M}{2} \qquad (5)$$

$Max\,\alpha\log e\left(\frac{N}{F},\alpha\right)$ represents the effect of data and model integration; $f(n)log$ represents that when the data is closely integrated with the model, it should be taken as the negative amount of difference, while when combined with sparse, it should be taken as the compensation amount. Based on the specific specification of Bayesian formula and the organic combination of model and data, on the basis of meeting the clustering characteristics, it is necessary to calculate and analyze the model through multi-dimensional clustering algorithm. The input of this algorithm contains m objects. The objects in the same cluster have high similarity, on the contrary, the similarity is small. The algorithm description process is shown in Figure 3.
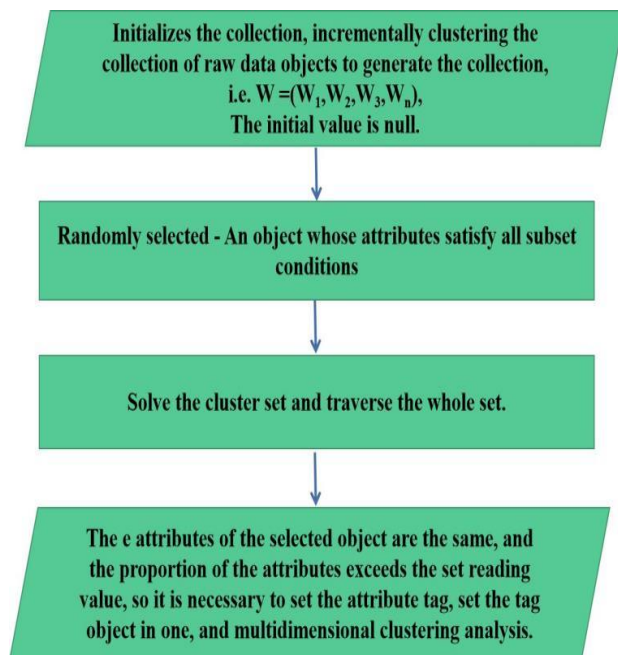
Figure 3: Algorithm description process

3.2   Complexity

The space cost generated by the new algorithm needs to fully consider the characteristic samples of big data. If hierarchical clustering is used to optimize the clustering algorithm, all clusters to be clustered need to be reasonably set according to the serial mode, the total clustering time (R) and the cost (n). Then the space complexity (W) is expressed in Equation 6.

$$W = n + R\left(m^2\log m\right) \qquad (6)$$

In terms of optimization rules, when the model and data fusion are sparse, set x and y as the dimension of the data set. When dividing attributes, only scan the data set at one time, in which z identifies clustering data, and the results will not be affected by factors such as multidimensional space and input order [20, 21]. Then the multi-dimensional spatial clustering can be found in time through the evaluation of weight and threshold, and the amount of calculation can be simplified. The total clustering time $u \times n$ is divided by the linear arrangement of the consumption time (n) and the de duplication time (m); Total weight removal time $u \times n$; Time complexity $u \times m\ 2\log m$, then the total time complexity of the algorithm is calculated by Equation 7.

$$R(m) = u \times m + u \times m + u \times m^2\log m \qquad (7)$$

3.3   Acquisition of project cost data

There are two ways to obtain project cost data based on big data.

i.   There are generally two methods of internal collection in the platform. First method is to build a unified project cost information data collection template, collect and import the relevant data in the platform according to the user-defined unified specifications, so as to directly convert the target cost data information and store it in the local database for backup. The second method is to set up fields conforming to certain specifications on the relevant cost information platform, collect the information of the same field and store it in the local database [22].

ii.   The specific methods and principles are as follows: create a unified data exchange format through the corresponding platform interface, and realize the information exchange of relevant businesses inside and outside the platform.

According to the collection method and the form of price change trend, we generally use the box method to process the project cost data studied in this paper. Before processing, we must first solve the problem of detection. For the detection of noise data, the change of cost data is mainly based on the overall change of market economy [23]. From the perspective of time series, it changes continuously, and is largely affected by the overall economic development. Generally, there will be no major

fluctuations and changes. We set the percentage of the annual change threshold range of cost data to 19%. Within the sampling range, the data points exceeding 20% of the average value are regarded as noise, the regression curve is calculated, and its value is re solved and corrected [24-26]. Handling method of inconsistent data format: to deal with the problem of inconsistent data format, the common method is to establish a general data acquisition template and collect according to the general data template to ensure the consistency of data acquisition format. According to the requirements and characteristics of data analysis in this paper, the data acquisition template is established, as shown in Table 1 and Table 2.

Number	Region	Unit	Time	Source
1	Jiangsu	yuan	January	Data survey
2	Shanghai	yuan	February	Data survey
3	Beijing	yuan	March	Data survey

Table 1: Data collection template - labor unit price expense template

Listing	Type	Accuracy	Format	Explain
Region	text	--	--	Area code
Number	double	1	XXX	Sample number
Company	text	--	--	Collection unit
Unit Price	double	0.02	XX.XX	Unit Price
Single time	Date	s	--	Acquisition time
Source	Date	…	--	Data sources

Table 2: Template description - labor unit price expense template

As the material cost accounts for a large proportion of the project cost, usually about 0% ~ 70%, the material price has a great impact on the specific final settlement results and decisions [27]. Therefore, this paper selects the material price as the research object, and focuses on the specific application of material price data in the fields of relevant project cost index prediction, project price

information analysis and investment estimation. Due to the dynamic, massive, multi-source and heterogeneous characteristics of project cost big data, we choose K-means clustering algorithm for specific solution [28].

4 Results and Analysis

This section illustrates the analysis of results obtained by comparing the seismic forces and finally presents its discussion and summary. In this proposed model, cluster analyze the quotations of 20 local suppliers for composite Portland cement.

Number	Region	Specifications	Unit Price	Source
1	SSX	PC32.1	452	Merchant A
2	SSX	PC32.1	326	Merchant A
3	SSX	PC32.1	419	Merchant A
4	SSX	PC32.1	385	Merchant A
5	SSX	PC32.1	453	Merchant A
6	SSX	PC32.1	376	Merchant A
7	SSX	PC32.1	413	Merchant A
8	SSX	PC32.1	306	Merchant A
9	SSX	PC32.1	378	Merchant A
10	SSX	PC32.1	403	Merchant A
11	SSX	PC32.1	487	Merchant A
…	SSX	PC32.1	…	Merchant A
20	SSX	PC32.1	346	Merchant A

Table 3: Data acquisition results

The 20 data listed in Table 3 are combined according to the price and serial number to obtain the initial data set A, A is $\{x_1, x_2, x_3, …, x_{20}\}$. Before calculation, it should be noted that K-Means algorithm must give K value before solution, which directly

determines the accuracy and efficiency of the algorithm. This paper determines the K value according to the following methods: firstly, compare the distance between each sample in the sample data set, select the point furthest from other points as the initial center point of the calculation according to the calculation results, and then determine the value of K through the newly generated classification [29, 30].

   i.   Select the two data with the smallest distance in the data sequence. In this example, the distance between the two points x_9 and x_{12} is the largest. Take these two points as the center of the cluster for cluster calculation to obtain two cluster sets.

They are: $S_{21} = \{x_9, x_2, x_4, x_8, x_{10}, x_{13}, x_{14}, x_{18}\}$ and $S_{22} = \{x_{12}, x_3, x_5, x_7, x_{11}, x_{12}, x_{16}, x_{17}\}$.

   ii. Combined with the above clustering calculation results, for the two clustering sets, first solve the first type of data and cluster center x_9 respectively, for example, to obtain the farthest distance of 83, the second type of data and cluster center x_{12} respectively, with the maximum distance of 85, and then select the point x_{11} with the maximum distance as the third cluster point.

   iii. Recalculate, select x_9, x_{12} and x_{11} as three cluster centers, and calculate three cluster sets as follows:

$S_{31} = \{x_9, x_2, x_{10}, x_{20}\}$, $S_{32} = \{x_{12}, x_1, x_5\}$ and $S_{33} = \{x_{11}, x_3, x_4, x_6, x_7, x_8, x_{10}, x_{13}, x_{14}, x_{15}\}$

   iv. Calculate the distance between the data elements in the three set classes and each cluster center, continue the cluster analysis, and then obtain four cluster sets [31].

   v.   Based on the above calculation results, the cluster numbers of different cluster centers are listed, as shown in Table 4.

Serial number	Center point	Numerical value	Number of clusters
1	x_9	315	5
2	x_{11}	406	4
3	x_{12}	475	4
4	x_{18}	413	9

Table 4: Cluster analysis results

According to the analysis of the results of the clustering algorithm in Table 4 and Figure 4, point x_{18} is

the center with the largest number of clustering samples in all clustering centers, so it can better reflect the real price of the market compared with other centers [32]. Taking this as an example, in the practical application of project cost budget and final accounts, we can analyze the market price of materials through the data mining algorithm proposed in this paper. By analyzing the solution results, we can assist relevant personnel to accurately grasp the market price information and help auditors judge the authenticity of price information in time.
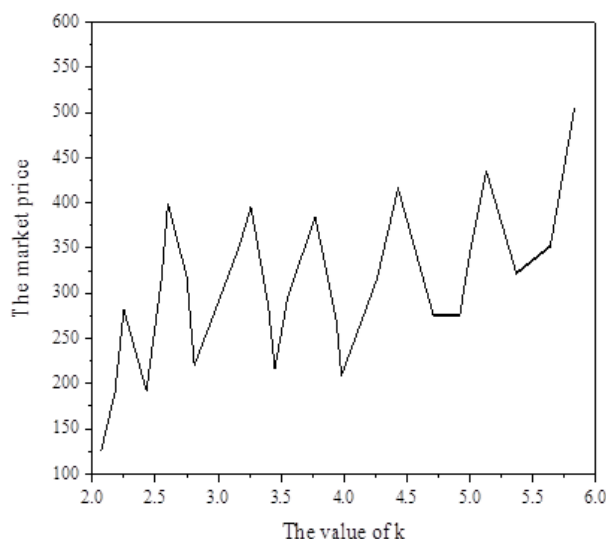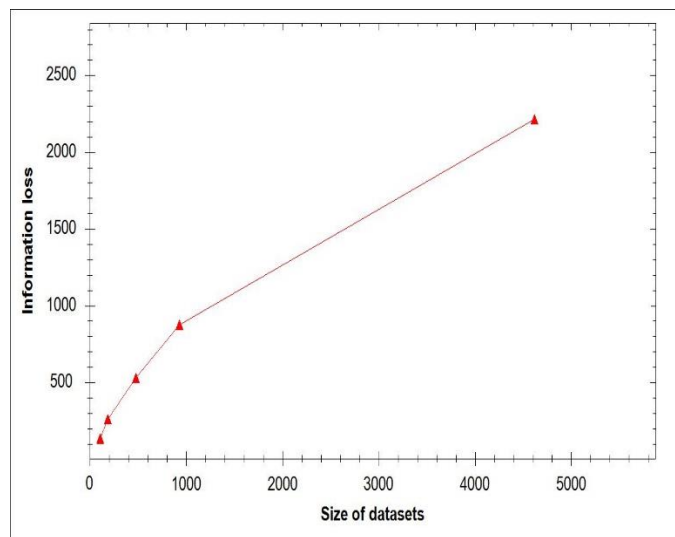

Figure 4: Results of clustering algorithm

Figure 5(a): Result for different size of datasets for information loss
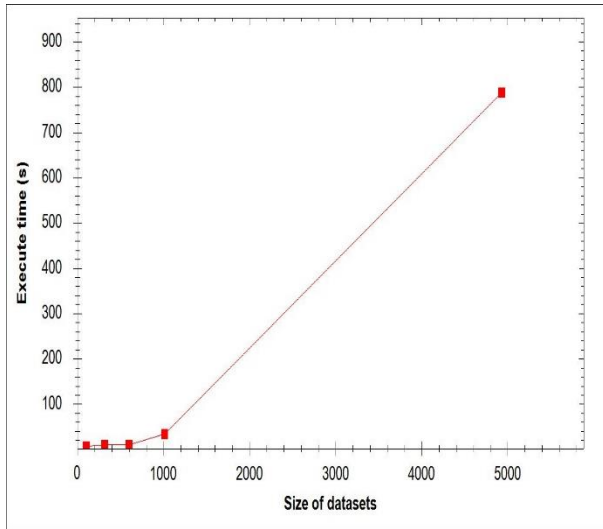
Figure 5(b): Result for different size of datasets for execute time

Different number of records are separated from the grown-up dataset and assess the exhibition of further developed anonymity model on various size of datasets, as depicted in Figure 5 (a and b). As shown in this figure, execute time increment and information loss with the increasing size of datasets. Execute time rises quickly, yet the incensement of information loss reportedly slows progressively. Clearly, the rising size of datasets genuinely affects execution time on the grounds that the grouping system of finding proportionality classes is perplexing and time consuming.
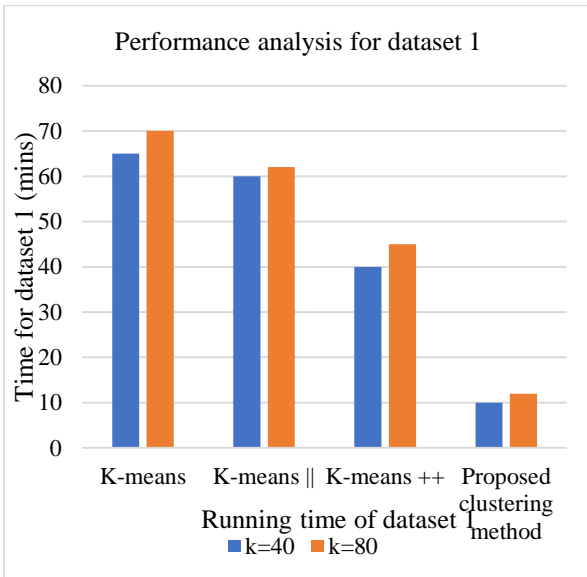
Figure 6: Performance comparison of time measured for dataset 1
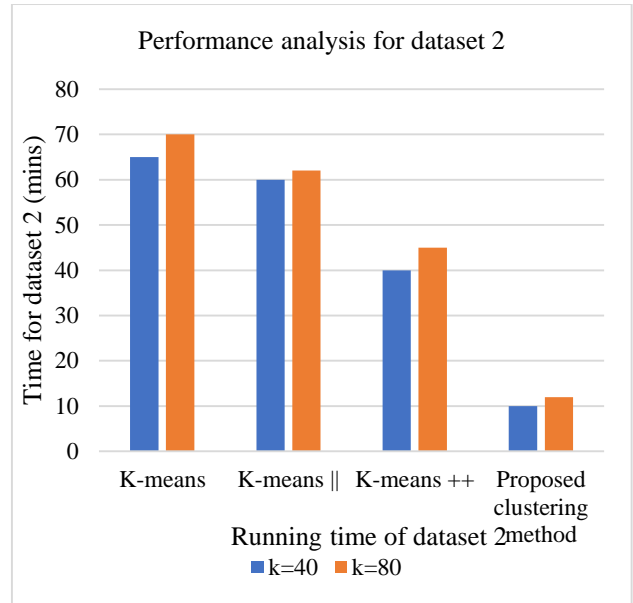
Figure 7: Performance comparison of time measured for dataset 2

The performance of the proposed clustering scheme is measured on two different datasets, dataset 1 i.e., BoW (Bag of words) dataset and dataset 2 i.e., HOUSE (household electric power consumption) dataset. To analyze the performance of clustering cost of proposed algorithm we have compared it with existing baseline models. The value of k is considered as 40 and 80 for BoW and HOUSE datasets. Figure 6 and 7 illustrates the experimental analysis of HOUSE and Bag of words (BoW) datasets and the total running time of proposed model is observed. It is observed from the experimentation that the proposed model achieves higher performance in comparison with K-means ++, K-means and K-means || when implanted to execute in parallel.

5 Conclusions

Different data analysis and mining methods are required for different purposes of project cost data mining under the background of big data. From the perspective of the selection and estimation of engineering unit price in construction enterprises, this paper summarizes the data mining process facing the characteristics of engineering cost data, and puts forward the method of analyzing and processing engineering cost data based on clustering algorithm. The proposed model provides a meaningful exploration for the research of massive engineering cost data mining. From the experimentation it is analyzed that the proposed clustering model achieves better time measurement when compared with existing baseline models. The clustering models based on computational intelligence are proposed. However, these intelligent technologies are not organically integrated. Machine learning and data mining technology have made great breakthroughs in today's academic and industrial circles. Therefore, how to integrate various intelligent technologies to give full play

to the functional characteristics of this kind of algorithm applied to cluster analysis is also one of the future research directions.

References

[1]   Li, W., & Huang, Q. (2017). Research on intelligent avoidance method of shipwreck based on bigdata analysis. *Polish Maritime Research*.
10.1515/pomr-2017-0125

[2]   Li, L., Wang, J., & Li, X. (2020). Efficiency analysis of machine learning intelligent investment based on K-means algorithm. *Ieee Access*, *8*, 147463-147470.
10.1109/ACCESS.2020.3011366

[3]   Dong-rui, L. (2017). Cluster analysis algorithm based on key data integration for cloud computing. *International Journal of Reasoning-based Intelligent Systems*, *9*(3-4), 123-129.
10.1504/IJRIS.2017.090041

[4]   Zhu, K., Joshi, S., Wang, Q. G., & Hsi, J. F. Y. (2019). Guest editorial special section on big data analytics in intelligent manufacturing. *IEEE Transactions on Industrial Informatics*, *15*(4), 2382-2385.
10.1109/TII.2019.2900726

[5]   Del Ser, J., Sanchez-Medina, J. J., & Vlahogianni, E. I. (2019). Introduction to the special issue on online learning for big-data driven transportation and mobility. *IEEE Transactions on Intelligent Transportation Systems*, *20*(12), 4621-4623.
10.1109/TITS.2019.2955548

[6]   Wu, C. (2019, June). Research on Clustering Algorithm Based on Big Data Background. In *Journal of Physics: Conference Series* (Vol. 1237, No. 2, p. 022131). IOP Publishing.
10.1088/1742-6596/1237/2/022131

[7]   Duan, S., & Wang, Z. (2021). Research on the service mode of the university library based on data mining. *Scientific Programming*, *2021*.
https://doi.org/10.1155/2021/5564326

[8]   Xing, Z., & Li, G. (2019). Intelligent classification method of remote sensing image based on big data in spark environment. *International Journal of Wireless Information Networks*, *26*(3), 183-192.
https://doi.org/10.1007/s10776-019-00440-z

[9]   Cai, Z. M. (2020). Network community partition based on intelligent clustering algorithm. *Компьютерная оптика*, *44*(6), 985-989.
10.18287/2412-6179-CO-724

[10] Xu, Z., Shi, D., & Tu, Z. (2021). Research on diagnostic information of smart medical care based on big data. *Journal of Healthcare Engineering*, *2021*.
https://doi.org/10.1155/2021/9977358

[11] Li, W., Luo, Y., Tang, C., Zhang, K., & Ma, X. (2021). Boosted Fuzzy Granular Regression Trees. *Mathematical Problems in Engineering*, *2021*.
https://doi.org/10.1155/2021/9958427

[12] Shi, F., & Zhu, L. (2019). Analysis of trip generation rates in residential commuting based on mobile phone signaling data. *Journal of Transport and Land Use*, *12*(1), 201-220.
http://dx.doi.org/10.5198/jtlu.2019.1431

[13] Wendong, X., Yuanfeng, L., & Deli, C. (2017). Algorithm of key data ensemble clustering and approximate analysis in cloud computing. *International Journal of Reasoning-based Intelligent Systems*, *9*(3-4), 177-184.
10.1504/IJRIS.2017.090038

[14] Singh, P. K., & Sharma, A. (2022). An intelligent WSN-UAV-based IoT framework for precision agriculture application. *Computers and Electrical Engineering*, *100*, 107912.
https://doi.org/10.1016/j.compeleceng.2022.107912

[15] Zeng, H., Dhiman, G., Sharma, A., Sharma, A., & Tselykh, A. (2021). An IoT and Blockchain-based approach for the smart water management system in agriculture. *Expert Systems*, e12892.
https://doi.org/10.1111/exsy.12892

[16] Sharma, A., & Singh, P. K. (2021). UAV-based framework for effective data analysis of forest fire detection using 5G networks: An effective approach towards smart cities solutions. *International Journal of Communication Systems*, e4826.
https://doi.org/10.1002/dac.4826

[17] Sharma, A., Singh, P. K., & Kumar, Y. (2020). An integrated fire detection system using IoT and image processing technique for smart cities. *Sustainable Cities and Society*, *61*, 102332.
https://doi.org/10.1016/j.scs.2020.102332

[18] Tseng, F. H., Cho, H. H., & Wu, H. T. (2019). Applying big data for intelligent agriculture-based crop selection analysis. *IEEE Access*, *7*, 116965-116974.
10.1109/ACCESS.2019.2935564

[19] Zhao, Y., Ding, F., Li, J., Guo, L., & Qi, W. (2019). The intelligent obstacle sensing and recognizing method based on D–S evidence theory for UGV. *Future Generation Computer Systems*, *97*, 21-29.
https://doi.org/10.1016/j.future.2019.02.003

[20] Yuan, W., Deng, P., Taleb, T., Wan, J., & Bi, C. (2015). An unlicensed taxi identification model based on big data analysis. *IEEE Transactions on Intelligent Transportation Systems*, *17*(6), 1703-1713.
10.1109/TITS.2015.2498180

[21] Wang, L. (2021, December). Intelligent analysis of accounting information processing under the background of big data. In *2021 2nd International Conference on Big Data Economy and Information Management (BDEIM)* (pp. 461-464). IEEE. *10.1109/BDEIM55082.2021.00100*

[22] Ma, X., Wang, Z., Zhou, S., Wen, H., & Zhang, Y. (2018, June). Intelligent healthcare systems assisted by data analytics and mobile computing. In *2018 14th International Wireless Communications & Mobile Computing Conference (IWCMC)* (pp. 1317-1322). IEEE. *10.1109/IWCMC.2018.8450377*

[23] Hu, H., Tang, B., Gong, X., Wei, W., & Wang, H. (2017). Intelligent fault diagnosis of the high-speed train with big data based on deep neural networks. *IEEE Transactions on Industrial Informatics*, *13*(4), 2106-2116. *10.1109/TII.2017.2683528*

[24] Vedavathi, N., Dharmaiah, Ghuram, Venkatadri, Kothuru and Gaffar, Shaik Abdul. Numerical study of radiative non-Darcy nanofluid flow over a stretching sheet with a convective Nield conditions and energy activation. *Nonlinear Engineering*, *10*(1), 159-176, 2021. *https://doi.org/10.1515/nleng-2021-0012*

[25] Hayat, Tasawar, Ullah, Inayat, Muhammad, Khursheed and Alsaedi, Ahmed. Gyrotactic microorganism and bio-convection during flow of Prandtl-Eyring nanomaterial. *Nonlinear Engineering*, *10*(1), 201-212, 2021. *https://doi.org/10.1515/nleng-2021-0015*

[26] Li, Zhenfang, Gao, Dong, Wu, Chuanji, Lv, Guoqing, Liu, Xin, Zhai, Haoran and Huang, Zhanfang. Mechanical performance of aerated concrete and its bonding performance with glass fiber grille. *Nonlinear Engineering*, *10*(1), 240-244, 2021. *https://doi.org/10.1515/nleng-2021-0018*

[27] Liang, H., Yun, C., Kan, M. J., & Gao, J. (2019). Research and application of element logging intelligent identification model based on data mining. *IEEE Access*, *7*, 94415-94423. *10.1109/ACCESS.2019.2928001*

[28] He, Z., He, Y., Liu, F., & Zhao, Y. (2019). Big data-oriented product infant failure intelligent root cause identification using associated tree and fuzzy DEA. *IEEE Access*, *7*, 34687-34698. *10.1109/ACCESS.2019.2904759*

[29] He, X., Wang, K., Lu, H., Xu, W., & Guo, S. (2020). Edge qoe: Intelligent big data caching via deep reinforcement learning. *IEEE Network*, *34*(4), 8-13. *10.1109/MNET.011.1900393*

[30] Lei, Y., Jia, F., Lin, J., Xing, S., & Ding, S. X. (2016). An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Transactions on Industrial Electronics*, *63*(5), 3137-3147. *10.1109/TIE.2016.2519325*

[31] Srivani, B., Sandhya, N., & Padmaja Rani, B. (2020). Literature review and analysis on big data stream classification techniques. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, *24*(3), 205-215. *10.3233/KES-200042*

[32] Liu, X., Sun, Q., Lu, W., Wu, C., & Ding, H. (2020). Big-data-based intelligent spectrum sensing for heterogeneous spectrum communications in 5G. *IEEE Wireless Communications*, *27*(5), 67-73. *10.1109/MWC.001.1900493*

The Application of Internet of Things and Oracle Database in the Research of Intelligent Data Management System

Yujiao Liu[1]*, Rajeev Kumar[2], Ashutosh Tripathi[3], Anil Sharma[4], Muskaan Rana[5]

[1]Chongqing Open University, Chongqing Business Vocational College, Chongqing, 400000, China
[2]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India
[3]Department of ECE, University Institute of Engineering, Chandigarh University, Chandigarh, India
[4]Department of Computer Science, Faculty of Technology, Debre Tabor University, Ethiopia
[5]Department of Computer Science and Engineering Chandigarh University, Mohali, India
Emails: yujiaoliu7@126.com, Rajeev.kumar@chitkara.edu.in, ashu20034@gmail.com, anilsharma@dtu.edu.et, muskaan.e11410@cumail.in

The most critical issue in manufacturing is known as resource allocation. This article demonstrates an intelligent data management consisting resource allocation mechanism. The aim of the proposed system is to provide timely and effective decision for the resource allocation. Aiming at the needs of general large-scale monitoring systems, this paper designs an intelligent data management system that can provide fast data query and relieve sudden data congestion through in-depth research on Oracle database and data division. To the data access request from the front desk, the system can respond quickly through the real-time data monitoring module and the online analysis software OLAP mode database, which has far-reaching significance for the development of the Internet of Things and related systems. The experimental results show that, compared with the traditional system, the same bitmap index only occupies about 1/30 of the original table, and the data size is reduced by more than 10 times. The proposed model is compared with other state of art classifiers for evaluating percentage efficiency and F score. The experimental data verifies the characteristics of the system in this paper to strengthen the background data receiving and processing capabilities, and alleviate the problems such as the reduction of the system running rate and even the system paralysis caused by the sudden mass data.

Povzetek: Mehanizem dodeljevanja virov je implementiran s pomočjo inteligentnega sistema in baze Oracle.

1 Introduction

As a new generation of monitoring system development, IP-based network digital monitoring system has gradually become the main monitoring method in the contemporary era. At present, most of the monitoring is mainly used for indoor video monitoring with a small empty range. However, with the vigorous development of the Internet of Things technology, the information transmission technology with the object state as the basic data has broadened the development of the monitoring system [1]. More and more monitoring systems are gradually developed to rely on the Internet of Things technology to conduct unmanned monitoring in large outdoor spaces, such as intelligent bridge health detection, intelligent fire protection systems, environmental monitoring, etc. These systems play an important role in people's lives. A typical database - the modular design of Oracle database in intelligent data management system is shown in Figure 1 [2].
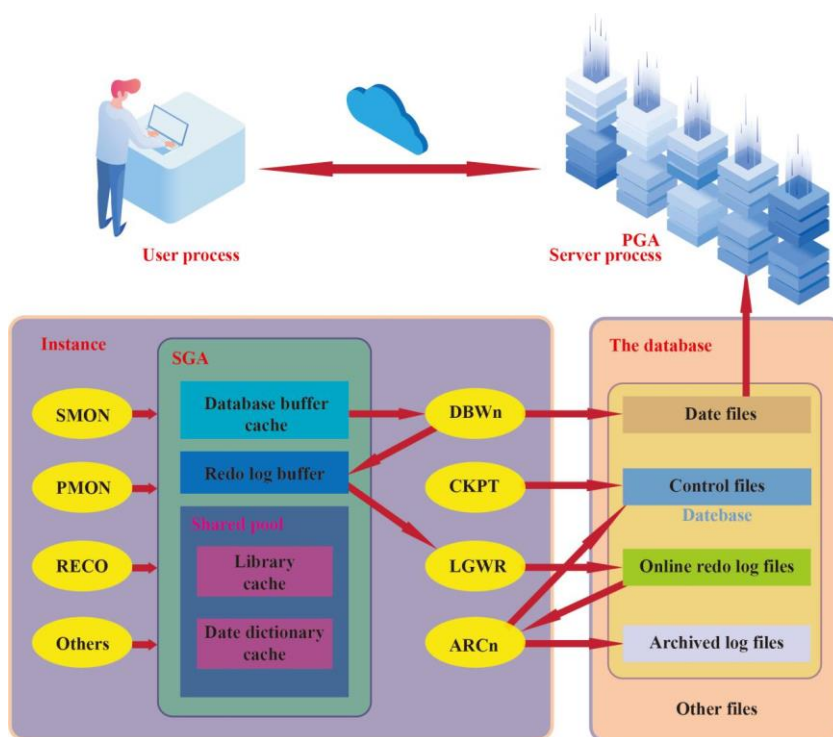
Figure 1: Modular design of Oracle database in intelligent data management system

In order to solve this problem in the monitoring system of the Internet of Things, this topic gives the detailed optimization design and the specific realization of the system from various aspects. The main research content of this topic is to develop the data layer of a new generation monitoring system based on the Internet of Things technology. In view of the impact of the sudden massive data generated by high-frequency collection on the background server, it can effectively solve the problem of excessive data loading and data volume. Too large and other problems [3]. Through in-depth research, there is still no universally applicable solution in the field of Internet of Things monitoring in China. The research on the undergraduate topic can effectively fill this gap, and the problem of reading and writing sudden massive data has been effectively discussed. It provides theoretical help and experimental data reference for relevant researchers in the future. The final output of this project is a data model that can solve the sudden massive data loading and reading and writing. This model can provide good support for the background data layer of the Internet of Things monitoring system and avoid the above problems caused by data [4, 5].

The rest of this article is organized as: Section 2 presents the most recent work carried out in the field of intelligent data management system. Section 3 consists the information about research methodology including system overflow and the implementation of business logic layer. The results and analysis part of the proposed scheme is covered in section 4. Section 5 describes the concluding remarks along with the future scope.

2   Related work

In this section the most recent work in intelligent data management system is discussed. One of the biggest characteristics of traditional monitoring systems is that there is less human-computer interaction, and the monitoring content is mostly image information. The amount of system data is usually maintained at a scale that increases linearly, and there is often a lack of analysis of the overall fluctuation trend of data over time [6]. The biggest feature of the new monitoring system based on the Internet of Things technology is to use the network to complete state-based monitoring, and to use the change trend of the monitoring object's own state as the standard for monitoring and analysis, so as to obtain comprehensive state information of the monitoring object. This system greatly increases the number of people According to the content of computer interaction, the monitor can change the monitoring mode according to the needs, which makes the originally stable growth of data volume more unpredictable [7]. Usually, due to the special needs of monitoring, high-density status collection of monitoring objects in a specific range will be performed, resulting in inevitable information peaks, which will bring greater data processing pressure to the background server processing. When the amount of data is overloaded, it may cause server congestion, slow message response, or even server crash. In order to avoid the problem of reading and writing caused by sudden massive data, it is necessary to provide a data layer structure suitable for the monitoring system mode, so as to perform data buffering and fast reading and writing of these data [8].

At present, there are few researches on the sudden mass database in the Internet of Things. Especially for massive data processing, basically most of the research papers are mainly analyzed in a database environment, such as the discussion of database partitioning technology, and the domestic research situation such as data table index design scheme as follows. Meng introduced a study from the National University of Defense Technology realized a real-time loading technology for TB-level massive data, and proposed a real-time loading system based on this technology. It mainly uses the SQL*Loader mechanism in the Oracle database to quickly process data storage, while using Database-specific swap partition method to quickly complete data loading [9]. Guo *et al.* from the School of Computer and Electronic Information of Guangxi University proposed a method to process massive data on a server, which avoids a series of huge initial hardware investment caused by the use of minicomputers with strong data storage capabilities at the hardware level question [10]. Zhen *et al.* from the Department of Ordnance Science and Technology of the Naval Aviation Engineering College proposed a realization method of multi-threading and double-buffer theory in the field of real-time data reception and storage [11]. Chen *et al.* gave a design and implementation of a massive burst signal acquisition system, and proposed an effective solution to the problems involved in high-speed acquisition frequency [12].

Foreign countries are much more in-depth than domestic research in massive data research. For example, the PI real-time database system developed by OSI software company in the United States is one of the most popular real-time databases today. It uses revolving door compression technology and secondary filtering technology to compress the massive data loaded into the database extremely efficiently, saving a lot of money. Hard disk space [13]. Research such as MARS [14] developed by Southern Methodist University, and System [15] in Princeton University's "Mass Storage Machine" project designed the "master version" of the database into the memory environment to make the system as a whole Architecture with greatly improved performance. Mitzutani *et al.* [16] presented a parallel processing structure using dual CPUs on the recovery architecture. Sidlauskiene [17] proposed a method based on the combination of log and shadow to solve the problem of occupying more memory space and needing to maintain a large number of page pointers.

Research on loading massive amounts of data in database clusters is still in its infancy. The American Supercomputer Application Center and the Department of Astronomy of Illinois State University jointly conducted research on the storage and query system for massive astronomical data [18]. The SDSS project in the United States has studied how to use SQL server clusters to quickly store data [19]. The evolution of artificial intelligence and Internet of Things is considered for several industrial applications and contributing towards social life [20-23].

In general, the current massive data processing technology is still a hot research topic, especially in today's booming Internet of Things, the stored data not only far exceeds the data generated by previous applications, but also has higher storage requirements. more stringent requirements. As a problem often faced in the development and application of current and future actual systems, the sudden mass data processing technology is the focus of research and solution.

3   Research methodology

This section includes the research design and methods about the system overflow. The implementation of business logic layer is also presented in this section.

3.1   System workflow

In order to meet the cross-platform characteristics and facilitate the general application in the Internet of Things, the background communication method adopts Web Service connection port, which enables unimpeded communication between Java EE architecture and .Net architecture [24]. In business, data is buffered by means of double buffering technology and file writing, and Memcached technology is also used to process, buffer and store data in memory, and then use multi-threading and batch processing to load data in the background [25]. Secondly, in terms of database, Oracle database is used to save data, which is mainly divided into two parts: real-time database and historical database. The sudden massive data is mainly stored in the real-time database, so this paper will elaborate on the design and implementation of the real-time database [26]. The background part mainly solves the problem of suddenness, mainly using buffering technology and caching technology to solve the problem of reading and writing, so the database is implemented by a single database. The main monitoring function process of the system is shown in Figure 2.
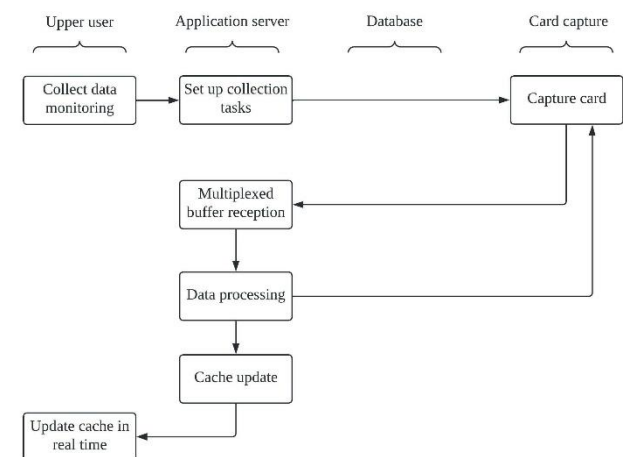

Figure 2: Flow chart of burst massive data collection function

The above process is the main execution steps when data is collected at high frequency. When the data is collected by other methods such as low frequency, the basic steps are the same as above. When performing data query, it mainly performs database access operations.

3.2   Implementation of business logic layer

The business logic layer mainly processes and responds to various requests of users on the server side, maintains the timed task queue, and continuously schedules the acquisition card for data collection according to the acquisition task. Using in-memory databases and caches to improve data processing efficiency [27].

Basic operations mainly include basic transaction operations such as adding, deleting, updating, and querying [28] performed by the user in the foreground. This module is responsible for receiving and parsing user requests from the foreground, and calling the corresponding module to access the database to obtain information, or modify the database content. After that, the obtained data is packaged according to the rules and returned to the front-end user for display.

The operation of the server on data interacts with the database through the ORM framework. The main monitoring objects are encapsulated by the decorator pattern in the design pattern idea. Taking the bridge object in bridge monitoring as an example, the class diagram is depicted in Figure 3.
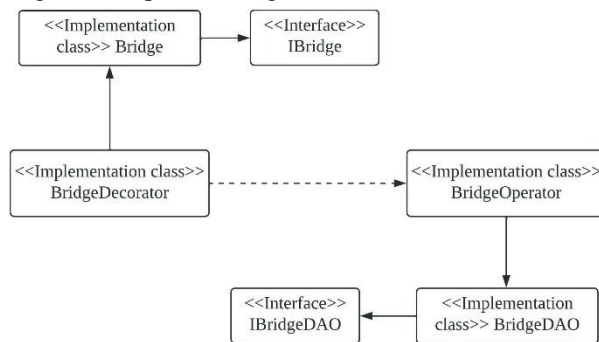
Figure 3: Basic module class diagram

IBridge completes the basic function definition of bridges, such as bridge number query, bridge parameter setting, bridge health status, etc. The Bridge class implements the IBridge interface, and rewrites each method as a concrete implementation. The BridgeDecorator class is used as the decoration class of the Bridge, and the Bridge object is called as a basic property. At the same time, the BridgeOperator inherits from BridgeDAO, which will extend the function of the database access class, and also exists as a base property for BridgeDecorator. In this way, new methods about bridges can be dynamically added to the BridgeDecorator. This interface-oriented design method meets the characteristics of JavaEE programming and is easy to maintain and upgrade.

The construction of other basic modules is similar to the above-mentioned modules. For example, the modules such as Line, Section and other modules complete the operation of business logic and database access by constructing their respective Decorator and Operator modules.

4   Results and Analysis

This section presents the description of result analysis of proposed scheme and the performance comparison of various indexes is also discussed in this section.

4.1   Data calculation transfer and batch loading method

The test data is mainly divided into 100,000 data loading and 1 million data loading. The data loading request is sent to the server through the simulation acquisition module to observe the processing efficiency of the system.

When performing a single insert operation of 100,000 pieces of data on the voltage acquisition table with calculation operation in the Oracle database, the average time for multiple sets of data is 2 minutes and 27 seconds; it is modified to a single insert operation of 1 million pieces of data. The data shows an average time of 23 minutes and 31 seconds. Instead, use 10 data as a group for batch data loading, and perform the insertion operation of 100,000 data in the voltage acquisition table with calculation operation. Multiple sets of data show that the average time is 2 minutes and 17 seconds; modified to 1 million data. For the insertion operation, multiple sets of data show that the average time is 23 minutes and 5 seconds. Using this batch method for insertion operations, the performance improvement is not obvious. The data is inserted directly into the database without calculation. In a single insert operation of 100,000 pieces of data, the average time of multiple sets of data is 19 seconds; if it is modified to a single insert operation of 1 million pieces of data, multiple sets of data show that the average time was 4 minutes and 44 seconds. Instead, 10 pieces of data are used as a group for batch data loading, and 100,000 pieces of data are inserted without calculation operations. Multiple sets of data show that the average time is 14 seconds; Multiple sets of data show that the average time is 3 minutes and 20 seconds.
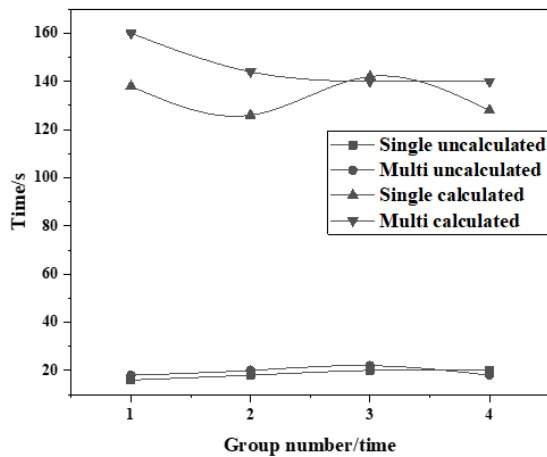
Figure 4: 100,000 data insertion tests

To sum up, it can be seen that when the server is used for batch data loading, as the amount of inserted data increases, the data insertion efficiency also increases gradually, but the efficiency does not increase linearly, and the efficiency increase is limited. If the calculation processing of the data is performed on the database, the consumption time is about 6 to 8 times compared with the simple insertion operation. 100,000 pieces of data are shown in Figure 4, and 1 million pieces of data are shown in Figure 5.
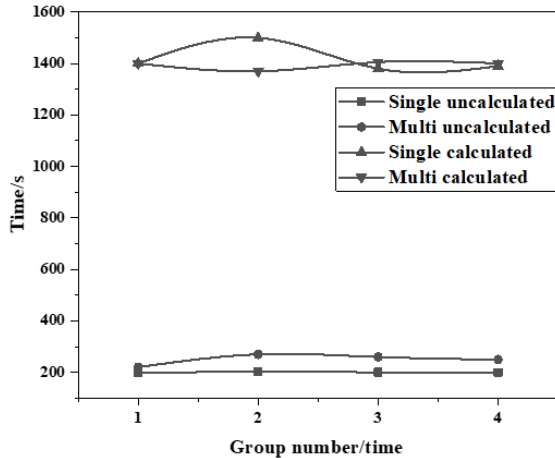
Figure 5: 1 million data insertion tests

When the amount of data is large, it is not advisable to temporarily buffer the data with memory, so the data needs to be cached in another way. Considering that Oracle database has a file bulk loading mechanism, a large amount of data can be buffered into files [29].

When the burst data is all stored in the form of files, the database tool SQL Loader is used to import the massive data into the database in parallel in the form of files. Taking 100,000 pieces of data as the test unit, it takes 0.23 seconds for the data to be buffered to the file, 2 seconds for batch importing into the database, and the comprehensive time is less than 2.2 seconds; with 1 million pieces of data as the test unit, the time for data buffering to the file is 6.8 seconds Second, the batch

import database takes 110 seconds, and the comprehensive time is less than 117 seconds.

Although using SQL Loader to load data is not as efficient as direct bulk loading, this method does not have the risk of memory overflow and is more reliable. Through the analysis of the experimental data, it is known that when the data loading rate is higher than 100MB/S~150MB/S, the system is very likely to have the risk of memory overflow, so the method of loading the data into a file should be used for processing [30].

4.2 Data query

First, the B-tree query efficiency is compared. Data query experiments are carried out in four data tables. The data volumes in the data tables are 10,000, 100,000, 1 million, and 10 million, respectively. By building a B-tree index on it and querying it, the number of consistent reads per table is 3 data blocks, 3 data blocks, 4 data blocks, and 4 data blocks, respectively. It can be seen that even when the amount of data is very different, when conditional queries are performed on fields with unique constraints, the resources they consume, that is, the SQL execution efficiency, are almost indistinguishable. However, if the index is not built, the data consistency read will be greatly increased. Through experiments, it is found that there are 750, 8823, 102391, and 894721 data reads and writes respectively. This is extremely inefficient for massive data query. The index performance comparison chart is depicted in Figure 6.
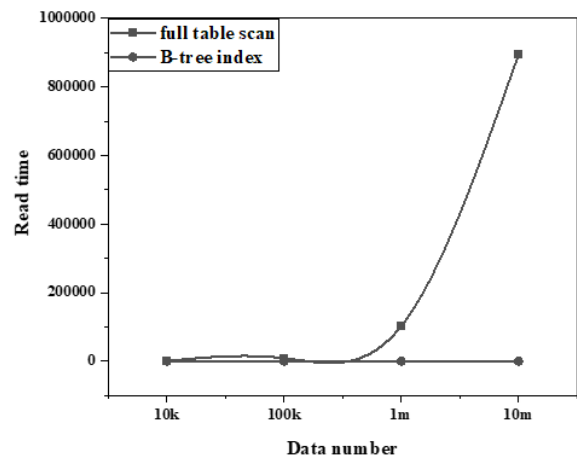

Figure 6: Performance comparison between B-tree index and full table scan
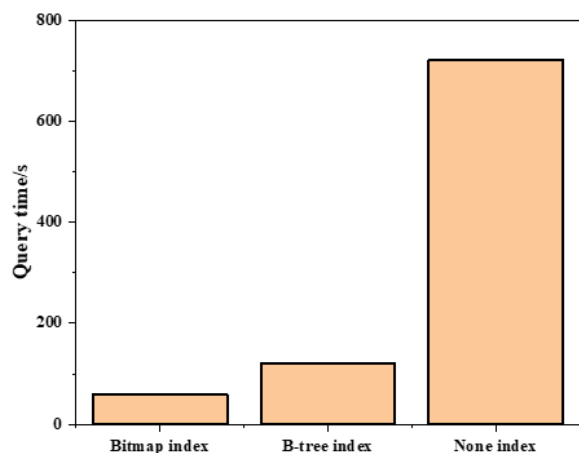
Figure 7: Performance comparison of various indexes

Figure 8: Efficiency and F score analysis of proposed scheme

Secondly, for the parameter table, in some cases, the use of bitmap index can have a better improvement. When the data dispersion is very low, the use of B-tree index is often not a good choice [31]. As shown in Figure 8 below, for a parameter table with a data volume of 10,000, by constructing bitmap index and B-tree index to count and query parameters, the query efficiency under the bitmap index is significantly higher than the other two.

In addition to greatly improving the execution efficiency of specific queries, bitmap indexes can also greatly reduce the disk space occupied by the index. For a table with a data volume of 100,000, the space occupied by the B-tree index is basically more than half of the original table, while the same bitmap index only occupies about 1/30 of the original table, and the data size is larger than that of the B-tree index. More than 10 times smaller.

Finally, by adopting the partition strategy, the acceleration effect can also be mentioned for the query. For data tables, large-scale query operations are not suitable for indexes, and the database optimizer often uses partitions to query data directly. For example, in a large-scale data query in a data table with a data size of 50,000, due to the use of partition pruning, the efficiency of range partitioning will be higher than that of hash partitioning and other partitioning strategies. When performing a specific data query, that is, using "=" to determine, the data query efficiency of hash partitioning will be higher than that of range partitioning and other partitioning strategies. From a purely performance point of view, hash partitioning has high performance when the field repetition rate is low and the operation result set is small. For range partitioning and list partitioning, if the data is the same, the execution plan of the two is roughly the same, that is, there is no big difference in performance, but list partitioning can solve some specific data distribution problems, which is beneficial to the data according to certain way to manage [32].
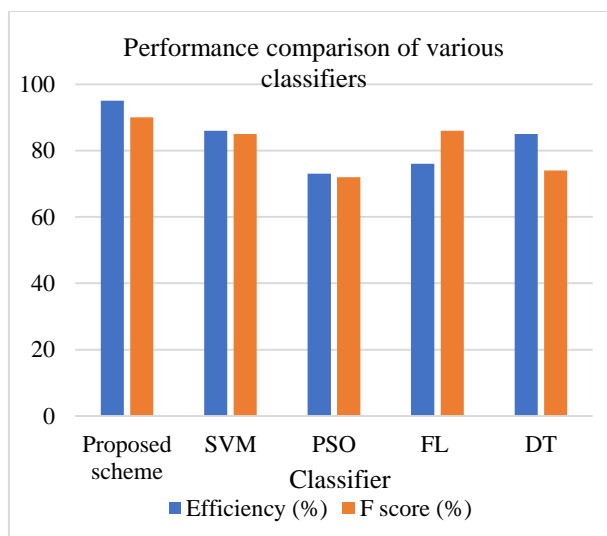
The performance analysis of proposed scheme in terms of efficiency and F score is compared with existing state of art classifiers. This performance comparison is presented in Figure 8. The simulation results in terms of efficiency and F score of the proposed system and existing state of art systems are analyzed and the result is depicted in the above figure. The proposed scheme has the highest efficiency (95%) and F score (90%). The simulation analysis is done in less than 10 minutes and the overall performance is analyzed. It is observed from the analysis that the proposed scheme has superior performance in comparison with existing systems such as support vector machine (SVM), particle swarm optimization (PSO), fuzzy logic (FL), and decision tree (DT).

5   Conclusion

Through the research of the subject, the design and implementation of the monitoring system to deal with the sudden mass data have been basically completed, and the data test has been carried out for the part that has been realized, and good results have been obtained. The experiments have proved that by using double-buffering multi-threading and loading data, good processing efficiency can be obtained in a single-database environment. Although there are many problems in the design and implementation of sudden massive data, such as memory cache and multi-threaded data processing, etc., but in the end, they have been well solved by consulting the data, not only making them aware of the sudden massive data. With a deeper understanding of handling, the abilities are also exercised to solve scientific research problems. At present, there are still many aspects of the data layer that need to be studied in depth, such as distributed databases and other issues. From the experimentation, it is analyzed that the proposed model achieves better efficiency and F score when compared with existing state of art classifier such as SVM, PSO, FL and DT.

However, it is believed that with the continuous development of monitoring system technology and wider research on IoT-related applications on the basis of current popular cloud computing and cloud storage technology, there will be a lot of application space. The Internet of Things monitoring system with the help of new technology will have a better solution for the processing of sudden massive data. At that time, data will no longer be the bottleneck of system operation, and the Internet of Things will be widely used in society, bringing greater benefits to the society.

References

[1] Zheng, F., & Zheng, B. (2021, August). Research on the Optimization and Application of Intelligent Data Acquisition and Alarm System Based on Internet of Things. In *Journal of Physics: Conference Series* (Vol. 1992, No. 2, p. 022083). IOP Publishing. *1088/1742-6596/1992/2/022083*

[2] Hemanth, D. J., Shakya, S., & Baig, Z. (Eds.). (2019). *Intelligent Data Communication Technologies and Internet of Things: ICICI 2019* (Vol. 38). Springer Nature. *https://doi.org/10.1007/978-3-030-34080-3*

[3] Dan, J., Zheng, Y., & Hu, J. (2022). Research on sports training model based on intelligent data aggregation processing in internet of things. *Cluster Computing*, *25*(1), 727-734. *https://doi.org/10.1007/s10586-021-03469-z*

[4] Meher, M., & Rostamy, D. (2021). Hybrid of differential quadrature and sub-gradients methods for solving the system of Eikonal equations. Nonlinear Engineering, 10(1), 436-449. *https://doi.org/10.1515/nleng-2021-0035*

[5] Mi, Z., Wang, T., Sun, Z., & Kumar, R. (2021). Vibration signal diagnosis and analysis of rotating machine by utilizing cloud computing. Nonlinear Engineering, 10(1), 404-413. *https://doi.org/10.1515/nleng-2021-0032*

[6] Xue, B. (2021, January). Information Fusion and Intelligent Management of Industrial Internet of Things under the Background of Big Data. In *2021 13th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)* (pp. 68-71). IEEE. *10.1109/ICMTMA52658.2021.00025*

[7] Wang, Z., & Sharma, A. (2021). Research on transformer vibration monitoring and diagnosis based on Internet of things. *Journal of Intelligent Systems*, *30*(1), 677-688. *https://doi.org/10.1515/jisys-2020-0111*

[8] Sittrop, D., & Crosthwaite, C. (2021). Minimising Risk—The Application of Kotter's Change Management Model on Customer Relationship Management Systems: A Case Study. *Journal of Risk and Financial Management*, *14*(10), 496. *https://doi.org/10.3390/jrfm14100496*

[9] Meng, Q. (2019, August). A Study on the Urban Emergency Management System Based on the Internet of Things. In *International Conference on Management Science and Engineering Management* (pp. 645-655). Springer, Cham. *https://doi.org/10.1007/978-3-030-21248-3_47*

[10] Gao, L., Yang, Q., Zou, B., Liu, Q., & Wang, C. (2021, March). Research on Data Asset Management System of Graph Database Based on Internet of Things. In *Journal of Physics: Conference Series* (Vol. 1802, No. 3, p. 032134). IOP Publishing. *10.1088/1742-6596/1802/3/032134*

[11] Zhen, H., Kumar, P. M., & Samuel, R. D. J. (2021). Internet of Things Framework in Athletics Physical Teaching System and Health Monitoring. *International Journal on Artificial Intelligence Tools*, *30*(06n08), 2140016. *https://doi.org/10.1142/S0218213021400169*

[12] Chen, G., Xiao, X., Zhao, X., Tat, T., Bick, M., & Chen, J. (2021). Electronic textiles for wearable point-of-care systems. *Chemical Reviews*. *https://doi.org/10.1021/acs.chemrev.1c00502*

[13] Jiang, Z. G., & Shi, X. T. (2021). Application Research of Key Frames Extraction Technology Combined with Optimized Faster R-CNN Algorithm in Traffic Video Analysis. *Complexity*, *2021*. *https://doi.org/10.1155/2021/6620425*

[14] Wen, C., Yang, J., Gan, L., & Pan, Y. (2021). Big data driven Internet of Things for credit evaluation and early warning in finance. *Future Generation Computer Systems*, *124*, 295-307. *https://doi.org/10.1016/j.future.2021.06.003*

[15] Aneja, S., En, M. A. X., & Aneja, N. (2022, January). Collaborative adversary nodes learning on the logs of IoT devices in an IoT network. In *2022 14th International Conference on COMmunication Systems & NETworkS (COMSNETS)* (pp. 231-235). IEEE. *10.1109/COMSNETS53615.2022.9668602*

[16] Mitzutani, I., Ramanathan, G., & Mayer, S. (2021, November). Semantic data integration with DevOps to support engineering process of intelligent building automation systems. In *Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 294-297). *https://doi.org/10.1145/3486611.3492413*

[17] Sidlauskiene, J. (2021). What Drives Consumers' Decisions to Use Intelligent Agent Technologies? A Systematic Review. *Journal of Internet Commerce*, 1-38. *https://doi.org/10.1080/15332861.2021.1961192*

[18] Sun, H., Tan, Y. A., Zhu, L., Zhang, Q., Li, Y., & Wu, S. (2022). A fine-grained and traceable multidomain secure data-sharing model for

intelligent terminals in edge-cloud collaboration scenarios. *International Journal of Intelligent Systems*, *37*(3), 2543-2566. *https://doi.org/10.1002/int.22784*

[19] Khalil, R. A., Saeed, N., Masood, M., Fard, Y. M., Alouini, M. S., & Al-Naffouri, T. Y. (2021). Deep learning in the industrial internet of things: Potentials, challenges, and emerging applications. *IEEE Internet of Things Journal*, *8*(14), 11016-11040. *10.1109/JIOT.2021.3051414*

[20] Singh, P. K., & Sharma, A. (2022). An intelligent WSN-UAV-based IoT framework for precision agriculture application. *Computers and Electrical Engineering*, *100*, 107912. *https://doi.org/10.1016/j.compeleceng.2022.107912*

[21] Zeng, H., Dhiman, G., Sharma, A., Sharma, A., & Tselykh, A. (2021). An IoT and Blockchain-based approach for the smart water management system in agriculture. *Expert Systems*, e12892. *https://doi.org/10.1111/exsy.12892*

[22] Sharma, A., & Singh, P. K. (2021). UAV-based framework for effective data analysis of forest fire detection using 5G networks: An effective approach towards smart cities solutions. *International Journal of Communication Systems*, e4826. *https://doi.org/10.1002/dac.4826*

[23] Sharma, A., Singh, P. K., & Kumar, Y. (2020). An integrated fire detection system using IoT and image processing technique for smart cities. *Sustainable Cities and Society*, *61*, 102332. *https://doi.org/10.1016/j.scs.2020.102332*

[24] Khan, I. K., Wei, Q., Chapman, S., Kc, D. B., & Kihara, D. (2015). The PFP and ESG protein function prediction methods in 2014: effect of database updates and ensemble approaches. *GigaScience*, *4*(1), s13742-015. *https://doi.org/10.1186/s13742-015-0083-4*

[25] Chen, C., Huang, T. S., Huang, J. C., Shih, C. H., & Du, Y. (2021). Music Intelligent Push Play and Data Analysis System Based on 5G Internet of Things. *Mathematical Problems in Engineering*, *2021*. *https://doi.org/10.1155/2021/6670534*

[26] Wang, X. (2022). Application of 3D-HEVC fast coding by Internet of Things data in intelligent decision. *The Journal of Supercomputing*, *78*(5), 7489-7508. *https://doi.org/10.1007/s11227-021-04137-0*

[27] Gao, X., Li, Q., & Liu, F. (2021, April). Research on the New normal Technology and Application of artificial Intelligence in the Internet of things. In *Journal of Physics: Conference Series* (Vol. 1865, No. 4, p. 042062). IOP Publishing. *10.1088/1742-6596/1865/4/042062*

[28] Qi, Y., & Wu, H. (2021, April). Fusion Application of Big Data and Cloud Computing In the Internet of Things. In *Journal of Physics: Conference Series* (Vol. 1881, No. 3, p. 032013). IOP Publishing. *10.1088/1742-6596/1881/3/032013*

[29] Zhou, G. (2021, March). The Application of Computer in Enterprise Economic Management Under the Background of Internet of Things. In *The International Conference on Cyber Security Intelligence and Analytics* (pp. 281-289). Springer, Cham. *https://doi.org/10.1007/978-3-030-70042-3_41*

[30] Caiqian, Z., & Xincheng, Z. (2021). Multimedia system and database simulation based on internet of things and cloud service platform. *Journal of Intelligent & Fuzzy Systems*, *40*(2), 2613-2624. *10.3233/JIFS-189253*

[31] Yue, S., Du, Y., & Zhang, X. (2021). Research and application of agricultural internet of things technology in intelligent agriculture. In *Journal of Physics: Conference Series* (Vol. 1769, No. 1, p. 012020). IOP Publishing. *10.1088/1742-6596/1769/1/012020*

[32] Zhou, B., Liu, Y., Xie, Y., Wang, J., Hao, Z., & Meng, J. (2021, May). Research and Application of Intelligent Street Lamp Platform Based on Ubiquitous Internet of Things. In *Journal of Physics: Conference Series* (Vol. 1920, No. 1, p. 012068). IOP Publishing. *10.1088/1742-6596/1920/1/012068*

Intelligent Engineering Management of Prefabricated Building Based on BIM Technology

Jing Feng1, Zhiying Zhang[1*], Yuequn Xu[1], Aiju Zhang[1]
[1]Architecture Department, Shijiazhuang Institute of Railway Technology, Shijiazhuang, Hebei Province, 050041, China
Emails: jingfeng59@126.com, zhangzhiying67@163.com, yuequnxu7@163.com, aijuzhang@126.com

This article solves the problem of China's construction industry adopted by the traditional extensive construction mode for a long time. The traditional methods was falling behind as they have the largest number of accidents among various types of safety accidents in the construction industry. This paper puts forward a new mode of fine construction management based on BIM. This article depicts the experimental analysis considering 277 accidents of falling from height, accounting for 54% of the total. There were 72 collapse accidents, ranking second among all types of safety accidents in the construction industry. It further discusses the application measures and benefits of BIM Technology in fine management from four aspects of quality management, schedule control, cost management and safety management. It is demonstrated from experimentation that BIM Technology has brought good economic and social effects to aid fine management.

Povzetek: S strojnim učenjem in pomočjo BIM tehnologije so bile analizirane nesreče v kitajskem gradbeništvu.

1   Introduction

China's construction industry has adopted the more traditional extensive construction mode from a very long time. With the improvement of the construction market, prefabricated buildings have gradually attracted extensive attention [1]. As the main difference between prefabricated buildings and traditional buildings, the construction site of prefabricated buildings is not only in the assembly site, but also in the manufacturing plant. It is precisely because of the extension of the construction site that the construction management becomes more and more complex, from the traditional single management of the construction site to the current project management of both manufacturing plant and construction site. In addition, the construction mode is changed from wet operation to dry operation, and from cast-in-situ to assembly, which also changes the whole construction management system [2].

At the same time, due to the development of the construction technology of prefabricated buildings, the traditional problems existing in prefabricated buildings, such as the connection quality of components and fittings and the production and manufacturing of large components and fittings, have been solved, thus promoting the rapid popularization of prefabricated buildings to a certain extent. However, in the process of popularization, with the extension of the management chain and the increase of the management process, new problems continue to emerge. In order to coordinate the construction process of prefabricated buildings at the management level, improve the construction management efficiency of prefabricated buildings. It further promotes the development of prefabricated buildings from the management level has become a major problem to be solved [3]. The major applications of BIM technology in building designing are depicted in Figure 1.
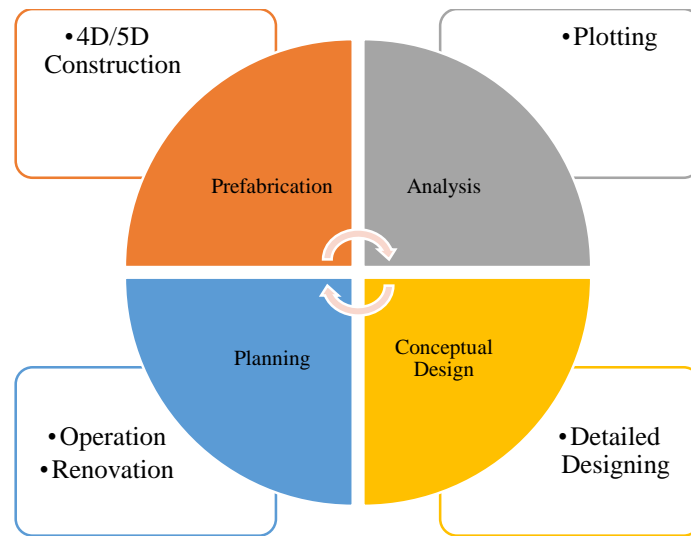
Figure 1: The major applications of BIM technology in building designing

The research gap lies in the view of the complex management problems of prefabricated buildings. This paper contributes in introducing BIM Technology into the construction process of prefabricated buildings, in order to find an appropriate construction application management mode of prefabricated buildings based on BM technology. With the help of the built BM model platform, it can effectively coordinate the management of manufacturing plant and assembly site. The proposed methodology can eliminate the information island effect in the management process, and integrate a series of management processes such as production and manufacturing, logistics and transportation. It further addresses the temporary storage and on-site assembly of components and parts, so as to provide some reference for the application of BIM Technology in the management of assembly building construction.

Further, this article is organized as: section 2 presents the literature review followed by discussion of methods in section 3. Research results of experimentation are depicted in section 4 followed by conclusion in section 5.

2   Related work

With the rapid development of society and the advancement of urbanization, the requirements for the construction industry are higher and higher. Fabricated building components are made in factories, with fast construction speed, good precision and quality. They can meet the green building design and construction requirements of "four sections and one environmental protection" to the greatest extent. They are in line with the development of modern construction industry and have received strong support from China [4]. The development of BIM Technology in China is relatively late. At present, it is mainly concentrated in the design stage, and its application in prefabricated building construction project management is relatively small. Guo and Wei combined with the characteristics of prefabricated buildings and BIM Technology, analyzed

the application value of BIM Technology in the whole life cycle of prefabricated buildings, and established a collaborative platform of prefabricated buildings based on BIM Technology [5]. Li *et al.* used Revit API and c# high-level programming language technology to establish the data statistical analysis process of light assembly construction process [6]. Szelag discussed the application of BIM Technology in the design and construction of prefabricated buildings from four aspects: model creation, collision detection, progress simulation and real-time roaming [7]. Zhang constructed the ISM model of restrictive factors and believed that the fundamental reason restricting the development of prefabricated buildings in China is the lack of professionals [8]. Wesz *et al.* constructed the assembly building integration system based on BIM platform, which promoted the application of BIM Technology in assembly building [9]. Qianqian put forward the assembled building management mode with BIM Technology as the information means and lean construction as the guiding ideology [10].

Abey and Anand established the maturity evaluation model of BIM Technology in the construction stage of prefabricated buildings [11]. Ngo *et al.* constructed a BIM application capability evaluation model of prefabricated buildings based on grey clustering, and proposed a new construction management and quality control method of prefabricated system based on BIM Technology and laser scanning [12]. Wang and Srinivasan established a quality management system for assembly component production by combining the core values of BIM Technology and RFID technology [13]. Serrano analyzed the role of BIM Technology in pre-construction planning, component management and control, construction schedule management, site dynamic layout and cost management of prefabricated buildings [14]. However, the construction process of prefabricated buildings is different from ordinary cast-in-situ buildings. Its construction site is not only a construction site, but also a factory.
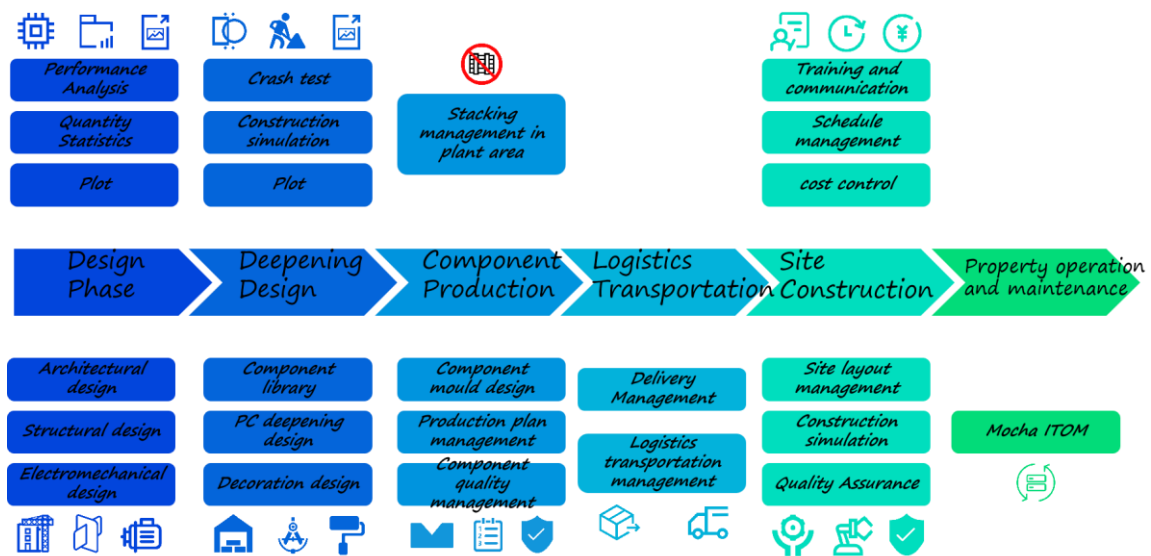
Figure 2: Development trend of prefabricated building construction

The cost management, quality management, safety management and schedule management in these two aspects can be better and faster realized under the coordination of BIM Technology, as shown in Figure 2. Therefore, it is of great theoretical and practical significance to analyze the role of fine management in the process of prefabricated building construction from BIM Technology.

3 Research method

Fine construction management is to control the details of the construction process accurately and standard, so as to save resources and reduce costs to the greatest extent [15]. BIM Technology is the integration and circulation of various information of buildings, which can provide complete and accurate information for fine construction and improve the efficiency of fine construction management. In previous engineering projects, BIM and fine construction management were not used at the same time, but from the perspective of theoretical research, it is feasible to apply BIM concept and fine construction management to engineering projects.

3.1 BIM Technology and fine construction management have common goals

Fine construction management is to formulate a specific and clear responsibility system from the perspective of management, implement the responsibility requirements of each participant, minimize the resources consumed in the construction process, achieve accurate control of the construction process, reduce material waste and reduce construction cost [16]. BIM Technology is to accurately divide the tasks by stages through information, visualization and other means, simulate the construction process, find a weak link in the process, and correct the construction scheme in time, so as to reduce

the construction cost and improve the project benefit. It can be seen that the objectives of the two are the same, and ultimately to reduce the construction cost.

3.2 Both refined construction management and BIM require the participation of all units

BIM is an information sharing technology in the whole life cycle of buildings, which involves many participants and each stage of construction, and requires the cooperation and exchange of each participant such as the owner, the design unit and the government [17]. Refined construction management is a comprehensive management method, which penetrates into every link of the work. Each activity participant needs to form refined ideas and earnestly implement the refined system. Form a corporate culture with the fine concept as the core, which is an important guarantee for fine construction management. Both have a common mass base and are consistent in terms of participants.

3.3 Fine construction management can make up for the deficiency of BIM from the management level

At present, most of the research and application of BIM in China are focused on the technical aspects of drawing deepening design, site dynamic layout, construction progress simulation, construction process simulation, BIM calculation, pipeline comprehensive optimization and so on, lacking the research on the management mode based on BIM Technology. To really use BIM well, we not only need advanced software and single node technology application, but also need advanced management scheme to match it, so as to give full play to the role of BIM, grasp the construction

objectives as a whole, reduce resource waste and ensure the completion of

construction objectives. Through fine construction management, the deficiencies in BIM management are made up, and the obstacles to the development of BIM Technology are eliminated from the root [18].

3.4 BIM Technology in turn promotes the development of fine construction management

BIM Technology injects "information" elements into fine construction management, which in turn promotes the development of fine construction management [19]. The core of BIM is to realize the transmission and sharing of information. BIM model stores all kinds of building information. This building information model can be used as the basis of the project, provide accurate and real data for the construction of various disciplines, optimize the construction scheme, and reasonably allocate the use of personnel and materials, so as to promote the development of fine construction management.

3.5 The refined construction management integrated with BIM Technology is more operable

The application of BIM in fine construction management provides accurate and real data support for fine management, so that the work is refined and the assessment quantification is based on, rather than based on experience. So that the fine management is no longer an empty rules and regulations, but carried out with good reasons, which enhances the operability of the fine management.

The construction fine management mode based on BIM includes fine management objectives, management contents, management elements and management system. BIM based construction fine management mode is based on fine management and BIM Technology as the core. It decomposes and refines the construction process accurately and in detail, implements the responsibilities of each step, and clarifies, concretizes and quantifies the responsibilities, with the main goal of minimizing the resources occupied by management and reducing management costs. Figure 3 depicts the construction fine management mode based on BIM [20].

Under the traditional mode, the quality control of engineering projects is mainly in the design and construction stages, and mainly carried out by the construction unit, the construction unit and the supervision unit. Generally, the quality of the project is inspected and accepted by relevant units or personnel organized by the supervising engineer (or the project leader of the construction unit) on the basis of the self-inspection and evaluation of the construction unit according to the qualified quality standard [21-23]. The whole process of project quality control under the traditional mode is shown in Figure 4.

Under the traditional mode, the construction unit generally enters the project from the bidding stage of the project, and rarely or basically does not participate in the design of the project. Therefore, the quality control in the design stage of the project is mainly responsible by the design unit and the construction unit. The work content of quality control in the design stage mainly includes two aspects: the control of quality standards adopted by the project and the control of design work quality itself [24,25].

In recent years, with the maturity of the construction market, Chinese construction enterprises have gradually established a quality-oriented business philosophy, which has steadily improved the quality level of construction projects. However, the extensive construction management mode cannot be completely changed in the short term. The construction quality control system under the old mode is still adopted by most projects, and there are still many problems in construction quality:

i. In the traditional working mode, there are a large number of CAD drawings in the construction stage, and the drawings of various disciplines are independent of each other, resulting in the disharmony between a large number of drawings, which brings hidden dangers to the construction. At the same time, for buildings with strange shapes and complex structures, two-dimensional drawings are difficult to express and workers are difficult to understand, making technical disclosure difficult, which may cause construction quality problems.

ii. The project lacks construction quality control scheme. At this stage, the project quality control mainly depends on supervision, self inspection and spot check. It is too late to check the construction problems, and the hidden dangers cannot be eliminated before they occur.

iii. The key parts of quality control are not included in the construction scheme. The quality inspectors are not clear about the location, testing time and requirements of quality inspection objects, resulting in non-standard inspection in the construction process, untimely quality inspection and evaluation, and the project management personnel do not understand the quality of the construction process. The construction quality will not be evaluated as a whole until the project is completed.

iv. The construction of engineering project is a systematic and complex process, which requires mutual coordination and cooperation between different disciplines and types of work. However, in engineering practice, due to different majors or different affiliated units, it is difficult to coordinate and communicate among various types of work in advance. This leads to the poor coordination of various professional types of work in the actual construction, resulting in the discontinuous progress of the project, or the need for frequent rework, as well as the collision, even mutual

destruction and interference between various types of work, which seriously affects the quality of the project. For example, the work sequence arrangement of other professional teams such as water and electricity and the main construction team is unreasonable, resulting in the arbitrary gouging and opening of bearing walls, plates, columns and beams during the construction of hydropower, which destroys the main structure and affects the quality problem of structural safety.

v. China has strict regulations and division on the quality of building materials, and individual enterprises also have their own quality standards for the use of materials. However, in the actual construction process, the management of building material quality is often not paid enough attention. In order to pursue additional benefits, individual construction units will intentionally or unintentionally use some non-standard engineering materials in the construction process of engineering projects, resulting in problems in the final quality of engineering projects. In the traditional two-dimensional design, the disciplines and drawings are independent and not related to each other, so it is inevitable that there will be some problems of disharmony between the drawings. In BIM model, each individual building component is represented only once, such as shape, attribute and position in the model. All drawings, reports and analysis information sets obtained in the same version of BIM model are interrelated, and they are changed and updated everywhere. This function can solve the problem of disharmony among drawings. And in the process of

establishing the three-dimensional model, we can have an intuitive and comprehensive understanding of the project, so as to find the errors and defects in the design before the project construction, improve the engineering design quality and eliminate the engineering quality problems from the source. The speed and accuracy of establishing BIM model are very key. The speed and accuracy of modeling directly affect the effect of later engineering application. Autodesk Autodesk Revit 2015 software is selected for the initial modeling of the project. Revit has the powerful functions of architectural design, structural design and electromechanical design modeling, and can accurately and flexibly represent the geometric and physical characteristics of components. In the Revit model, all drawings, plan views, 3D views and schedules are established in the same database of the building information model. There is a close correlation between the 3D model and the drawings, so one modification will be automatically modified everywhere else, saving a lot of manpower and time to adjust the drawings and ensure the coordination between the drawings. At the same time, you can accumulate and create your own parametric family library, and create the current model by adjusting the parameters of the original component family when creating the model, which can greatly improve the modeling speed [26-28]. Complete the establishment of BIM model within the specified working days, record the errors found in the drawings during the creation process, and submit them to the designer in writing for modification opinions. See drawing joint review record Table 1 for some parts.
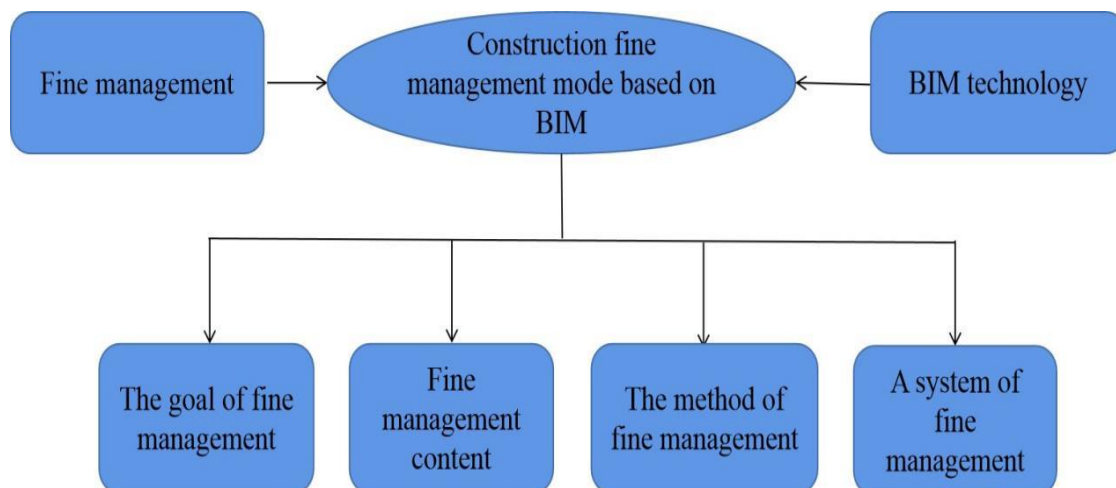

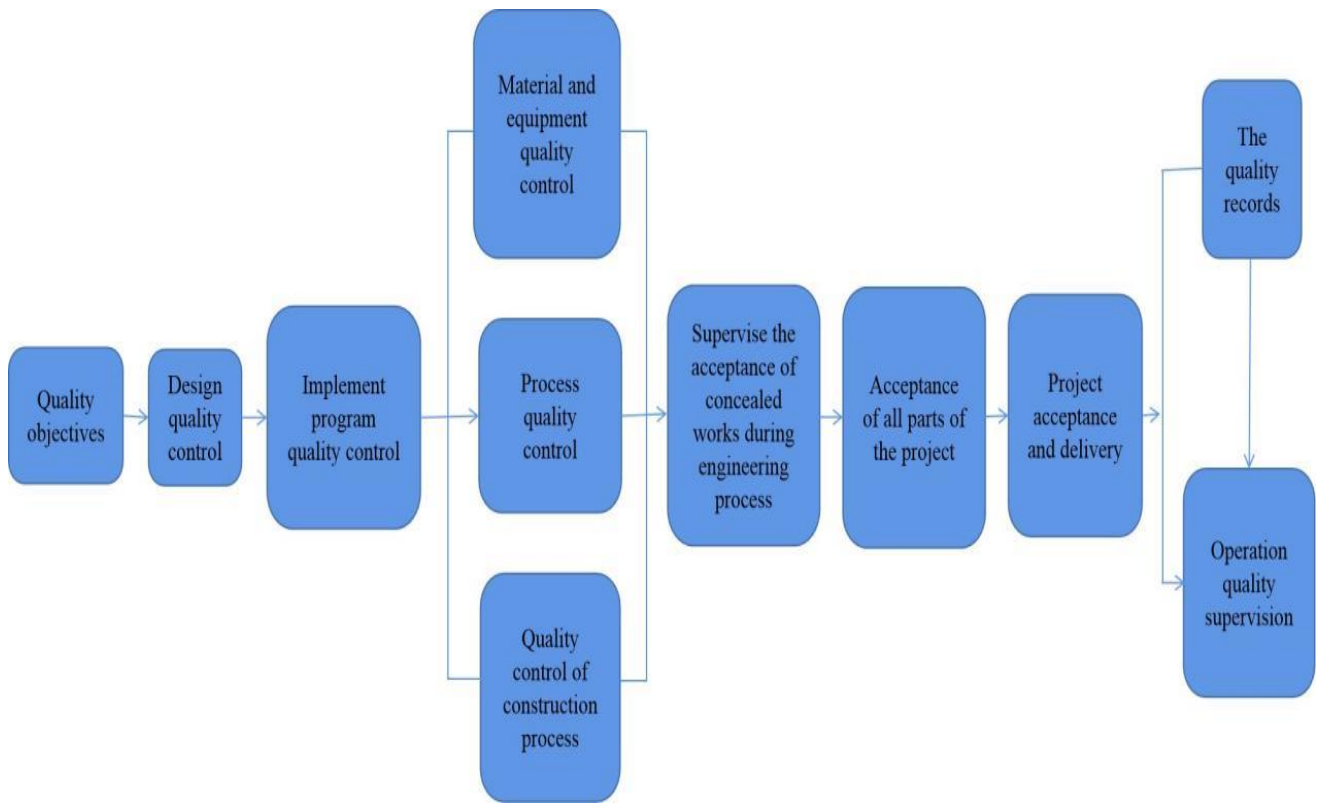Figure 3: Construction fine management mode based on BIM

Figure 4: Quality control process of engineering project under traditional mode

Project name	Phase II (C10 plot) project of resettlement housing in a large residential community		Major	Civil engineering
Joint examination place	Project meeting room		Date	
Serial number	Drawing No	Questions about drawings	Reply comments	
1	Building construction-01 and construction-03	The distance between axis m and axis L between axes 15-17 is inconsistent with the two drawings	Subject to building construction-01	
2	Structural construction-04	On the 11m floor of axis L and axis A, the beam position is inconsistent with the elevation. In the architectural drawing, the roof beam is aligned with the lower part.	Change the beam position to align with the lower part	
3	Structural construction-01	In the first point of 3.10, in the strength grade of concrete components, the rest floors of the frame beam of the main building are (soil above 0.000) C30, and the rest floors of the main floor slab are (soil above 0.000) C25. The	The strength grade of slab concrete shall be changed to that of frame beam of the same floor	

		concrete strength of the beam slab is different. Can the concrete strength grade of the beam slab be changed to the same strength grade? Is it feasible?	
4	Structural construction-05	There is no dimension for the opening beam at the junction of axis D and axis 6 of the second floor beam	According to the size of opening beam on the first floor
5	Structural construction-08	The beam between axis F and axis G on axis 7 is not marked	The beam is kl16
6	Structural construction-08	There is no kz-15 method in the column table	To the top of the third floor, the reinforcement shall be kzi5 on the first floor

Table 1: Drawing joint review record

4 Results and Analysis

Due to the characteristics of the construction industry and low safety investment, the labor environment and safety situation of construction workers are not optimistic. According to the accident statistics of the project quality and safety supervision department of the Ministry of housing and urban rural development. The statistics of safety accidents in China's construction industry are shown in Figure 5, and the statistics of accident deaths are shown in Figure 5. It can be seen that the number of safety accidents and deaths maintain a downward trend, and the safety production situation tends to be stable on the whole, but it is still at a high level, and the decline is not obvious.

The complete comparison of accident number with corresponding accident fatalities is depicted in Figure 7. The types of production safety accidents in China's construction industry are shown in Figure 8. The types of construction safety accidents in China mainly include falling accidents, collapse accidents, object strike accidents, etc. It can be seen that falling from height has the largest number of accidents among various types of safety accidents in the construction industry. It is revealed that 277 accidents, accounting for 54% of the total while collapse accidents ranked second among all types of safety accidents in the construction industry, with 72, accounting for 15%. The third type of safety accidents in the construction industry is object strike accidents, 66, accounting for 15%. Lifting injury accidents, machine injury accidents and electric shock accidents account for 10%, 5% and 2% of the total safety accidents respectively, ranking 4th-6th respectively.
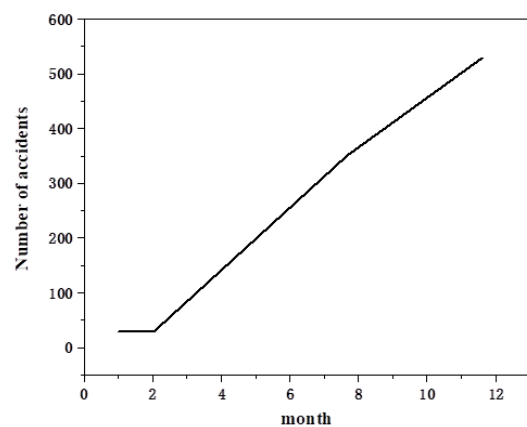
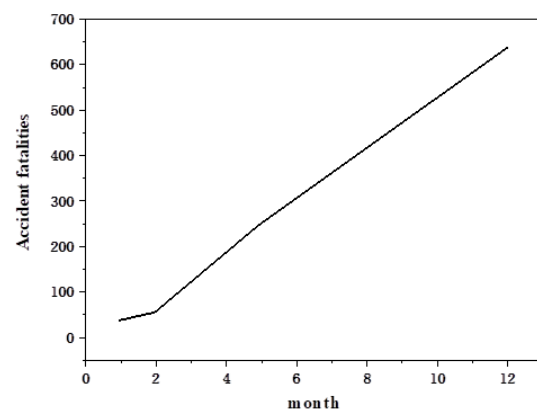Figure 5: Number of accidents
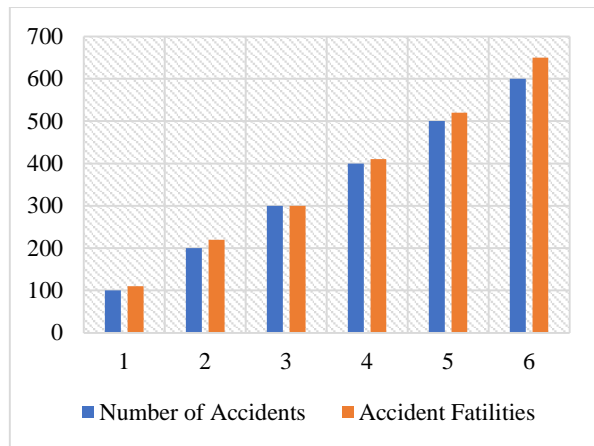
Figure 6: Accident fatalities

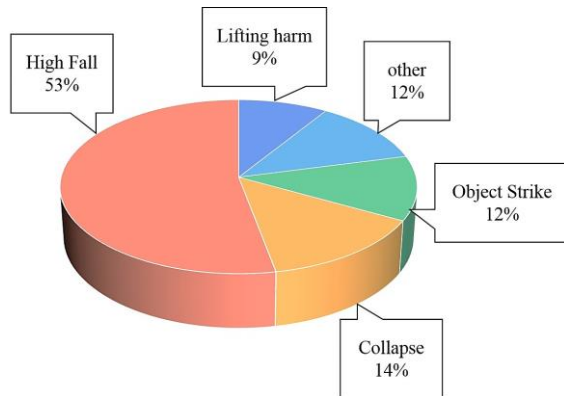Figure 7: Comparison of accident number with corresponding accident fatalities

Figure 8: Accident types in the year 2020

Based on the above safety statistics of China's construction industry, it can be seen that the current safety production situation of China's construction industry is still not optimistic, safety accidents are still not effectively controlled, causing huge economic losses, casualties and unnecessary losses every year. Falling from height, collapse and object strike are the most frequent types of accidents. The cumulative number of the above three types of safety accidents accounted for about 80% of the total number of accidents in 2020, and the mortality rate is also the highest among all kinds of accidents.

The main causes of safety accidents include non-standard construction market behavior, imperfect safety management system, ineffective preventive measures, incomplete elimination and treatment of hidden dangers in safety production, backward safety management level and technology, weak safety awareness of construction workers, taking chances and not strictly abiding by professional norms. The understanding of the construction process is not thorough enough, and there are potential safety hazards in the construction process or site layout. The difficulty of fall prevention management of large-scale construction projects is that it is difficult to find all edges and openings that need protection [29,30]. The traditional management method is mainly based on the two-dimensional drawings and the environmental inspection and supervision management of the

construction site to find the four openings that need protection: staircase, elevator, entrance and exit and reserved opening.

Five temporary edges: the periphery of balustraded balcony, the periphery of roof without external frame protection, the periphery of frame engineering floor, both sides of stair ramp and the outer side of unloading platform. With heavy workload and low efficiency, it is difficult to find all potential falling safety hazards of the project and formulate corresponding safety protection measures in time. For the safety measures of openings, the safety protection measures taken for openings of different sizes are different, as shown in Table 2.

Opening size	Safety measures
Greater than 150 cm	Guardrail protection and safety flat net protection shall be added around the opening
50-150 cm	A layer of grid grid formed by fastening steel pipes must be set and covered with scaffold board
25-50 cm	For openings during the installation of prefabricated components and openings formed temporarily due to lack of components, bamboo and wood can be used as cover plates to cover the openings
2.5-25 cm	Use solid cover plate to cover the opening for protection
Less than 2.5cm	Considering the size of the hole and the reduced possibility of falling objects, it is ignored

Table 2: Safety protection measures for portal

Using BIM modeling, 4D virtual construction technology and visualization characteristics, we can find out the potential falling safety hazards in different construction stages and parts in the process of 3D model and 4D virtual construction. Then the fall protection model is established and imported into the structural model for detection to ensure that there are no security vulnerabilities in the fall protection system.

The state-of-the-art comparison of the proposed method with the other techniques depicted in the literature is presented in Figure 9. This figure reveals the state-of-the-art comparison of accident prediction accuracy for various methods reported in the literature survey.
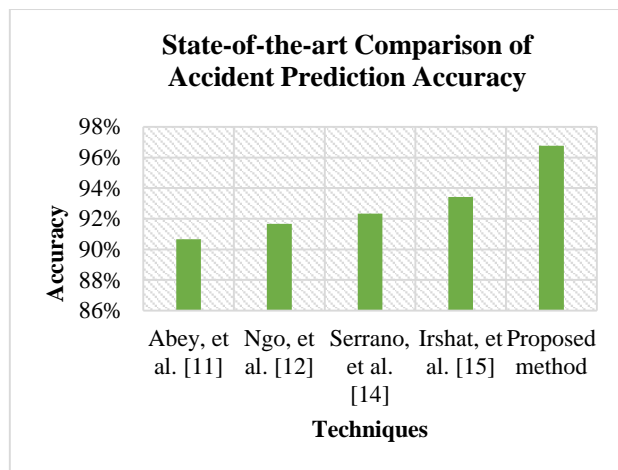
Figure 9: State-of-the-art Comparison Accident Prediction Accuracy

The study of this model reveals that it is easy to find out all the edges and openings with potential fall safety hazards in the whole project. Then place the built edge and opening fall protection model in the structural model to form a fall protection system, provide a visual management platform for managers, and strengthen the communication effect of safety plan. Before the actual construction, the simulated construction environment can be observed to identify and analyze the hazard sources. Optimize the construction scheme and site layout, or formulate emergency measures to control safety risks and avoid safety accidents. In large and complex projects, many workers often carry out construction in different parts, but it is difficult for us to grasp the overall situation on site. In the virtual construction model, we can clearly see the potential risk factors in different parts.

5   Conclusions

China accounted for 80% of the total number of accidents in 2020, and the mortality rate was also the highest among all kinds of accidents. The main causes of safety accidents include irregular behavior in the construction market. In order to solve these problems, this paper puts forward the specific application of BIM Technology in engineering construction safety management, and discusses the advantages and application effects of BIM technology in construction safety management. BIM based safety management can enable project managers to discover in advance, the risks that may affect the project construction progress or lead to safety accidents during project implementation. This article further formulates the corresponding control measures, strengthen the communication of safety plan and emergency plan between project management personnel and construction personnel. It maintains the integration and sharing of information and reduces the occurrence of accidents. Thus, facilitating the implementation of safety plan and the control of safety risks, further promoting the refinement and digitization of construction safety management.

References

[1]   Wang, R., Dong, X., Wang, K., Sun, X., Fan, Z., & Duan, W. (2019). Two-step approach to improving the quality of laser micro-hole drilling on thermal barrier coated nickel base alloys. *Optics and Lasers in Engineering*, *121*, 406-415. https://doi.org/10.1016/j.optlaseng.2019.05.002

[2]   Wang, Q., Zhang, B., Yu, S., Xiong, J., Yao, Z., Hu, B., & Yan, J. (2020). Waste-printed circuit board recycling: Focusing on preparing polymer composites and geopolymers. *ACS omega*, *5*(29), 17850-17856. https://doi.org/10.1021/acsomega.0c01884

[3]   Hardin, M. (2019). Design-Build for Discovery: Applied Research on the Construction Site. *Building Technology Educator's Society*, *2019*(1), 9. https://doi.org/10.7275/642c-vp30

[4]   Akinade, O. O., Oyedele, L. O., Ajayi, S. O., Bilal, M., Alaka, H. A., Owolabi, H. A., & Arawomo, O. O. (2018). Designing out construction waste using BIM technology: Stakeholders' expectations for industry deployment. *Journal of cleaner production*, *180*, 375-385. https://doi.org/10.1016/j.jclepro.2018.01.022

[5]   Guo, S. J., & Wei, T. (2016). Cost-effective energy saving measures based on BIM technology: Case study at National Taiwan University. *Energy and Buildings*, *127*, 433-441. https://doi.org/10.1016/j.enbuild.2016.06.015

[6]   Li, X., Xu, J., & Zhang, Q. (2017). Research on construction schedule management based on BIM technology. *Procedia engineering*, *174*, 657-667. https://doi.org/10.1016/j.proeng.2017.01.214

[7]   Szeląg, R. (2017). The use of BIM technology in the process of analyzing the increased effort of structural elements. *Procedia Engineering*, *172*, 1073-1076. https://doi.org/10.1016/j.proeng.2017.02.165

[8]   Zhang, Y. (2022). Application of BIM Technology in Project Construction Schedule Management. In *2021 International Conference on Big Data Analytics for Cyber-Physical System in Smart City* (pp. 77-85). Springer, Singapore. https://doi.org/10.1007/978-981-16-7469-3_8

[9]   Wesz, J. G. B., Formoso, C. T., & Tzortzopoulos, P. (2018). Planning and controlling design in engineered-to-order prefabricated building systems. *Engineering, Construction and Architectural Management*. https://doi.org/10.1108/ECAM-02-2016-0045

[10]   Qianqian, S. U. N. (2018). Design of Prefabricated Old-age Building Based on Modularization: A Case Study of Institutional Elderly Houses Design in Taigou Village, Xi'an City. *Journal of Landscape*

Research, *10*(5).
10.16785/j.issn 1943-989x.2018.5.016

[11] Abey, S. T., & Anand, K. B. (2019). Embodied energy comparison of prefabricated and conventional building construction. *Journal of The Institution of Engineers (India): Series A*, *100*(4), 777-790.
https://doi.org/10.1007/s40030-019-00394-8

[12] Ngo, T. D., Nguyen, Q. T., & Tran, P. (2016, October). Heat release and flame propagation in prefabricated modular unit with GFRP composite facades. In *Building Simulation* (Vol. 9, No. 5, pp. 607-616). Springer Berlin Heidelberg.
https://doi.org/10.1007/s12273-016-0294-3

[13] Wang, Z., & Srinivasan, R. S. (2017). A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. *Renewable and Sustainable Energy Reviews*, *75*, 796-808.
https://doi.org/10.1016/j.rser.2016.10.079

[14] Serrano, W. (2019, September). iBuilding: artificial intelligence in intelligent buildings. In *UK Workshop on Computational Intelligence* (pp. 395-408). Springer, Cham.
https://doi.org/10.1007/978-3-030-29933-0_33

[15] Irshat, K., Petr, R., & Irina, R. (2018, October). The selecting of artificial intelligence technology for control of mobile robots. In *2018 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon)* (pp. 1-4). IEEE.
10.1109/FarEastCon.2018.8602796

[16] Brooks, R. A. (1991). Intelligence without representation. *Artificial intelligence*, *47*(1-3), 139-159.
https://doi.org/10.1016/0004-3702(91)90053-M

[17] Winfield, A. F., & Jirotka, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *376*(2133), 20180085.
https://doi.org/10.1098/rsta.2018.0085

[18] Vaidyanathan, P. K., Sidduraj, M., & Woods, G. (2016). *U.S. Patent No. 9,405,531*. Washington, DC: U.S. Patent and Trademark Office.

[19] Tian, B. (2018). Building Artificial Intelligence for Dermatological Practice. *Open Access Library Journal*, *5*(04), 1.
10.4236/oalib.1104541

[20] Chiba, D., Akiyama, M., Yagi, T., Hato, K., Mori, T., & Goto, S. (2018). DomainChroma: Building actionable threat intelligence from malicious

domain names. *computers & security*, *77*, 138-161.
https://doi.org/10.1016/j.cose.2018.03.013

[21] Gadakari, T., Hadjri, K., & Mushatat, S. (2016, August). Relationship between building intelligence and sustainability. In *Proceedings of the Institution of Civil Engineers-Engineering Sustainability* (Vol. 170, No. 6, pp. 294-307). Thomas Telford Ltd. ,
https://doi.org/10.1680/jensu.16.00028

[22] Ko, C.H., & Li, S. C. (2014). Enhancing submittal review and construction inspection in public projects. Automation in construction, 44, 33-46.

[23] Liu, J., & Shi, G. (2017). Quality control of a complex lean construction project based on KanBIM technology. EURASIA Journal of mathematics, science and technology education, 13(8), 5905-5919.

[24] Sharma, D., Kaur, R., Sandhir, M., & Sharma, H. (2021). Finite element method for stress and strain analysis of FGM hollow cylinder under effect of temperature profiles and inhomogeneity parameter. *Nonlinear Engineering*, *10*(1), 477-487.
https://doi.org/10.1515/nleng-2021-0039

[25] Ting, L., Khan, M., Sharma, A., & Ansari, M. D. (2022). A secure framework for IoT-based smart climate agriculture system: Toward blockchain and edge computing. *Journal of Intelligent Systems*, *31*(1), 221-236.
https://doi.org/10.1515/jisys-2022-0012

[26] Froufe, M. M., Chinelli, C. K., Guedes, A. L. A., Haddad, A. N., Hammad, A. W., & Soares, C. A. P. (2020). Smart buildings: Systems and drivers. *Buildings*, *10*(9), 153.
https://doi.org/10.3390/buildings10090153

[27] Kumbinarasaiah, S., & Raghunatha, K. R. (2021). A novel approach on micropolar fluid flow in a porous channel with high mass transfer via wavelet frames. *Nonlinear Engineering*, *10*(1), 39-45.
https://doi.org/10.1515/nleng-2021-0004

[28] Gruszczak, A. (2016). *Intelligence security in the European Union: Building a strategic intelligence community*. Springer.
https://doi.org/10.1057/978-1-137-45512-3

[29] Horowitz, M. C., Kahn, L., & Mahoney, C. (2020). The Future of Military Applications of Artificial Intelligence: A Role for Confidence-Building Measures?. *Orbis*, *64*(4), 528-543.
https://doi.org/10.1016/j.orbis.2020.08.003

[30] Huang, Z., Lin, K. J., Tsai, B. L., Yan, S., & Shih, C. S. (2018). Building edge intelligence for online activity recognition in service-oriented IoT systems. *Future Generation Computer Systems*, *87*, 557-567.
https://doi.org/10.1016/j.future.2018.03.003

Construction of Lean Control System of Prefabricated Mechanical Building Cost Based on Hall Multi-dimensional Structure Model

Danna Su[1], Miao Fan[1], Ashutosh Sharma[2]

[1]College of Railway Engineering, Zheng Zhou Railway Vocational & Technical College, Zhengzhou, Henan, 450000, China

[2]School of Computer Science, University of Petroleum and Energy Studies, Dehradun, 248171, India

Emails: dannasu9@126.com, miaofan121@163.com, ashutosh.sharma@ddn.upes.ac.in

Based on the systematic idea of Hall's multidimensional structure and the theory and practice of prefabricated building cost lean management, the prefabricated mechanical building cost lean control system based on Hall's multidimensional structure model is proposed and constructed. The application of the lean management of the hall multidimensional structure model from the perspective of the time dimension, logic, and knowledge dimension. The example analysis results show that the original design components and the number of open modes is 72, the optimized types of components and the number of open modes is 51, reduce 21 mold machining, mold costs were reduced by about 25%. The number of original design components and the lifting times of components is 129 kinds, the number of components and lifting of components is 103, the number of components per layer was decreased by 26, lifting time is shortened by about 20%, the comprehensive construction period is shortened by more than 40 days, improve the management efficiency, lean cost control of the project plays a positive role. It provides a reference for the lean control system management of the hall multidimensional structural model

Povzetek: Razvit je vitki nadzorni sistem za gradnje na osnovi Hallovega multi-dimenzionalnega modela.

1 Introduction

In recent years, with the orderly progress of the transformation and upgrading of the construction industry, prefabricated buildings have become the direction of sustainable development of the construction industry due to the advantages of energy conservation, environmental protection, green, and efficiency [1]. However, the development of prefabricated buildings has also brought a series of quality management problems, such as the transformation of the extended construction mode of the construction industry chain [2]. BIM technology as the information development of construction industry technology, applied to the project construction design, construction, management, can be integrated into the construction process of prefabricated building, realize the different links of construction industry chain information exchange, coordination, and simulated in the virtual environment, control the real project construction. With the maturity of the development system of prefabricated buildings and the improvement of the development scale, the cost is lower than that of traditional cast-in-place buildings, among which the cost of prefabricated buildings in the United States is only half the cost of traditional cast-in-place buildings [3]. In economic economics of prefabricated buildings Tezel, A. through the mail to 100 construction units, design company, prefabricated component manufacturers and workers questionnaire, detailed analysis of the use of prefabricated building system, most

of the contractors think in the use of prefabricated building system, if the staff has a high degree of specialization and information communication between the participating units smooth, prefabricated building system will reflect more economical [4]. Xing *et al.* outlined the development situation of prefabricated housing and prefabricated and traditional cast-in-place building in cost and construction process difference, with the help of cost software and fixed cost difference of cost analysis, and the key factors related to cost sensitivity analysis, and then affect the key factors of cost put forward suggestions on the control [5]. In the prefabricated building design phase. Goger and Bisenberger on the basis of fully considering the prefabricated building cost control, the design method is optimized. In the production stage of prefabricated components, the economic advantages of prefabricated buildings are analyzed from the production stage of prefabricated components, which shows the broad development prospects of prefabricated buildings and the huge economic benefits brought by [6]. The main reason for the high cost of prefabricated buildings than the traditional cast-in-place buildings is the high fixed asset investment and industrial worker training cost of prefabricated component factories. In the logistics stage of prefabricated components, for prefabricated building prefabricated components in the process of logistics due to not smooth information transmission caused by distribution delay, repeated detection, stacking error phenomenon to increase labor cost and mechanical costs,

RFID technology combined with GPS technology to the logistics transportation of prefabricated components, can quickly and simple positioning of prefabricated building prefabricated components, the site management of prefabricated components. Based on the basis of the present research, this paper to construct a lean control system based on the Hall multidimensional structure model, from the perspective of a time dimension, logic, and knowledge dimension application, through example analysis results show that combining the participants of various stages, recycling seven logical steps, and constantly refine cost management objectives and operation to achieve the goal of overall cost control.

The rest of this article is organized as: Section 2 presents the related works in various domains. Section 3 consists of methods comprising the concept and flowcharts of the proposed 3D structural model. Results and analysis are discussed in section 4 followed by concluding remarks in section 5.

2 Related work

The construction manufacturing business is considered the major reason for the degradation of the environment [7]. The construction businesses consume an excessive number of natural resources and are responsible for the wastage of C&D (construction and demolition) [8]. In the year 2018, approximately 600 million tons of waste is reported in the United States, even though this waste can be recycled and reused. In one study it is discussed that approximately 50% of C&D waste is recycled and reused and transferred to energy facilities [9]. It is estimated that approximately 40% of C&D waste after the recycled and reused treatment is transferred to the landfills without any further direction and use [10]. It is noticed from the observation that the adverse environmental impacts of C&D can be reduced by maximizing the recycling and reuse process [11]. Economic waste management activities can also help in reducing C&D waste [12]. Instead of giving attention to the issues of C&D waste, the low recycled and reused measures of C&D are considered to be major limitations. In the United States, the recycling of concrete material is estimated at approximately 55% [13]. The design of the construction waste management system is very essential for the recycling and reusing of industrial waste and to divert the industrial waste from landfills to reusability [14]. An efficient system for industrial and construction waste management systems incorporates the estimation of recycling and reusing quantities and the methods for storing and reducing construction waste [15]. This project is not limited to industrial applications but the overall growth of social life with the integration of the Internet of Things, AI, and robotics [16-19].

Moreover, such a system can also provide information about the stakeholders who are responsible for waste disposal. The benefits of recognition of such a system also present their implementation challenges in terms of delay and productivity [20]. In order to meet all such requirements, efficient planning is the foremost requirement to address issues such as budget, safety, and schedule [21]. BIM (building information modeling) is recognized as the main expansion for Construction, architectural, and engineering industries [22]. Over the last 10 years, BIM technology has gained attention and the majority of BIM applications are considered for construction waste management systems [23]. The planning of construction waste management can be improved by several capabilities of BIM such as simulation, visualization, and parametric modeling. However, one study on the requirement of BIM for construction waste management presents that the advanced computer-aided tools have the capability for enhancing the performance of construction waste management throughout the several phases of development [24]. An exhaustive review presents the application of BIM toward construction waste management, highlights that there is less evidence of such systems that can discretize the generation of construction waste for recycling and reusing without depending on some external issues, and addressing precise actions in the schedule of construction and hence admitting reuse of construction waste [25]. The authors have presented a four-dimensional BIM model for enhancing the recycling and reusing of construction waste and addressing the previous limitations. Their work considers on-site reusing and off-site recycling of construction waste and specific actions are indicated for admitting the reuse of construction waste [26]. With the integration of the temporal dimension to BIM, the generation of construction waste can be imagined as the activities of construction, therefore enabling the construction waste planning for on-site reusing and off-site recycling [27].

The four-dimensional BIM application in the planning of recycling and reusing is demonstrated for non-residential case studies in the streams of drywall and concrete [28]. These waste streams are nominated as they are the largest construction waste streams that are produced in the US. Concrete possesses a high potential for both recycling and reusing, whereas drywall possesses a good potential for recycling only. The maximum resource recovery can be achieved by the efficient planning of the construction waste recycling and reusing process, and thereby reduction can be observed in landfills of construction waste [29]. The prime objective of this study is to highlight the planning of construction waste for recycling and reusing for projects by designing a model based on a temporal and visual approach by using the available data of construction projects. The proposed model is also considered to be applicable for several projects that are independent of their locations. The major contribution of this study is to provide an approach for the identification of on-site reusing activities of construction waste.

3   Method

In this section, the concept of the multidimensional model and the flowchart of the proposed 3D structural model is described.

3.1   Cost structure of the prefabricated building, Hall 3D structure theory, and lean cost management thought

C.   *Cost composition of the prefabricated building*

Prefabricated building cost refers to all the costs involved in the life cycle of the prefabricated building project. It can be divided into the following four categories: planning and design cost, construction and production cost, warehousing and logistics cost, and construction and installation cost. Therefore, lean management of these processes is a key to reducing the cost of prefabricated buildings [30].
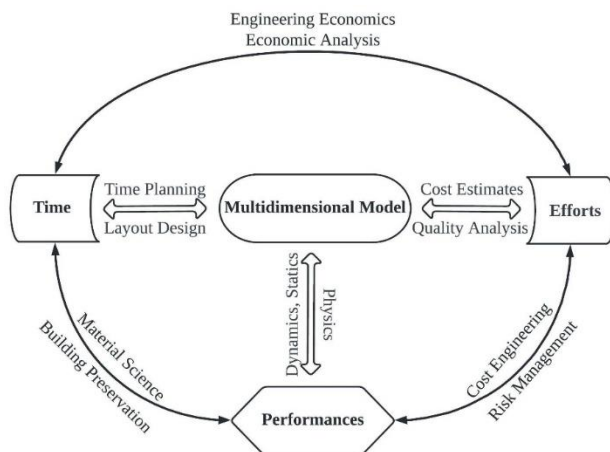
Figure 1: Concept of a multidimensional model

The concept of the multidimensional model is depicted in figure 1. It consists of a cycle of time, effort, and performance. In the proposed model, the time and effort cycles combine the process of planning, layout design for the estimation of cost parameters, and quality analysis considering economics. Effective planning and constant efforts lead to quality products and improved performance is achieved through dynamics, physics, and statics.

D.   *Hall 3-dimensional structure theory*

The theoretical method of 3-dimensional spatial structure solves the management problems of planning, organization, and coordination of some large and complex projects. Hall's three-dimensional structure theory divides the objects of system engineering research into knowledge dimension, time dimension, and logic dimension according to different stages, knowledge, and logic methods used. Using relevant expertise provides effective analysis tools for solving large and complex projects.

E.   *Lean cost management thought*

Compared with the traditional cost management method of construction projects, lean management pays more attention to the cost management of the whole process of the project, so using the idea of lean management for cost management is more comprehensive. Lean cost management is studied in the bidding, design, construction, logistics and other aspects of construction projects analyze the factors affecting cost at each stage, and then puts forward targeted cost management methods, so as to achieve the purpose of improving efficiency and reducing cost [31].

3.2   Construction and analysis of the Hall 3 D structural model

3.2.1   Time dimension

In this section, the working of the proposed design is discussed. The proposed design is divided into four categories as depicted in figure 2.
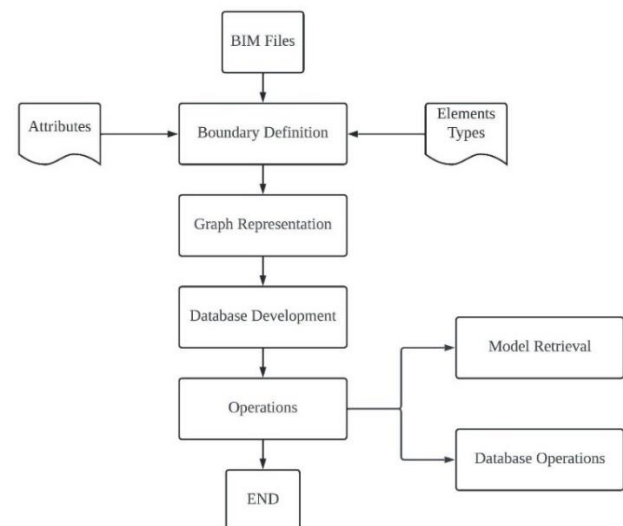

Figure 2: Proposed design of Hall 3D structural model

In the first step, system boundaries are determined through the inputs of selection attributes and building elements. In the next step, the information from the selected attribute or element is represented graphically and the same process is repeated for each module. In the third step, the graphical information is imported to a graph database. In the next step, graph-based operations are performed for module retrieval and performing other graphical applications.

A.   *Cost management in the planning and design stage*

The beginning stage of the cost composition of prefabricated construction projects is the planning and design stage. Usually, at this stage, the preliminary work should be strengthened, and the planning and design should be made according to relevant knowledge and

regulations, so as to obtain the minimum investment and obtain the maximum income. According to the lean cost management idea, the following two methods are put forward in deepening the design: method-is to implement the parallel design. In the prefabricated building planning and design stage, the relevant subjects of each stage can send technicians to participate. Methods Second, fine management based on BIM, collaborative operation of each major; using BIM technology to find omissions and collision inspection is conducive to reducing the cost generated by design change; the information platform built by BIM technology, establish a standard component library, realize the standardized design and reduce between later design cost [32].

B. Cost management in the construction and production stage

The cost of the production stage is the largest part of the life cycle cost of prefabricated buildings. Prefabricated components use the following lean cost management methods in the production stage: Method first, is to implement standardized production, which refers to collecting product information, ensuring the supply of raw materials, and conducting standardized mass production according to the information in the component information database. Methods second, to conduct lean supply chain management and establish BIM raw material supply information sharing platform.

C. Cost management in the warehousing and transportation stage

The cost management in the warehousing and transportation stage is mainly realized through the implementation of nine on-time productions and strengthening the protection of component transportation. The implementation of on-time production mainly refers to the reasonable planning of the production and completion time of prefabricated components and controlling the one-time production, so as to effectively use the storage space and reduce the inventory cost. Strengthening the transportation protection of components refers to the prefabricated components that are transported to the construction site after the production of the factory, and conduct strict quality checks on the loading stage and transportation stage to reduce the cost of secondary repair.

D. Cost management in the construction and assembly stage

The prefabricated components should be assembled after being transported to the construction site. At this stage, the I site should be managed in an orderly manner, and various construction information should be organized and coordinated to ensure normal operation. The following methods are proposed for the cost management of the construction and assembly stage based on the lean management method: 5s site construction management, which is very efficient for the site management of prefabricated construction projects

and is synchronously controlled through the integration of various aspects of information. Formulate a reasonable prefabricated hoisting plan, and the prefabricated components to be transported to the construction site shall be assembled in time, otherwise, the site will be occupied, and the storage cost on the site will be increased. The application of s site management method in the construction and assembly stage of prefabricated construction projects is shown in Table 1.

Designation	Concrete operations
Arrange	Organize and distinguish the relevant items on the site, and remove the irrelevant items on the site
Rectify	Place items in a reasonable location for easy search
Clear	Clean up the dust and garbage on the site to ensure that the site is clean and tidy
Cleaning	Continue to thoroughly implement the three links of sorting out, rectification, and cleaning
Accomplishment	Cultivate the comprehensive quality of relevant personnel on-site and improve the mental outlook of employees

Table 1: 5S practices for managing prefabricated building sites

3.2.2 Logical dimension

The logical dimension refers to the thinking procedure that the work content should be followed in each stage of the time dimension, that is, refers to the thinking process of each stage of cost management of lean management thought. When using system engineering ideas to solve engineering problems, logic dimensions can be divided into the following steps, as shown in Table 2.

Step	Concrete operations
Make clear the problem	The main purpose is to set the completion goals of various stages, schedule, consider possible problems, and prepare measures to respond
Set goals	After determining the overall goal, the objectives need to refine the goal and develop phased goals at each stage
Comprehensive plan	According to the characteristics of the target, the scientific scheme comparison method is used to finally determine the optimal scheme
Systems analysis	Considering the advantages and disadvantages of different schemes, then deeply analyze the unique advantages of each scheme, and comprehensively judge the efficiency and ease of completion of each scheme according to the corresponding indicators and rank

Scheme comparison	The optimal scheme is chosen according to the different objectives and the constraints existing in the actual process
Make policy	After systematic analysis and comparison of numerous schemes, the optimal implementation of the research problem is determined
Put into effect	Use the final scheme as the implementation scheme of the cost management site of the prefabricated construction project

Table 2: Steps for logical dimension

3.2.3 Knowledge dimension

The knowledge dimension of lean management of prefabricated buildings mainly includes project knowledge, financial knowledge legal knowledge, and management knowledge. Project knowledge refers to the process to be familiar with the design, production, transportation, and assembly stages of matching construction projects. Financial knowledge refers to discussing the cost composition of all stages of the prefabricated building project, analyzing the main factors affecting the cost, analyzing the content of cost management from the micro and macro perspective, and coordinating the interests of the participating subjects. Legal knowledge refers to the life cycle of prefabricated construction projects, from the bidding stage to the project completion stage, various legal risks should be avoided. Management knowledge refers to the flexible use of lean management theory, including lean value management theory and lean management characteristics [33].

4 Results and analysis

This section includes the result and analysis of the proposed model consisting of example analysis and risk assessment.

4.1 Lean cost management model of the prefabricated building based on Hall 3D structure

A. Prefabricated construction

Treated project cost management hall 3D structure model activity matrix hall 3D structure model by combining time-dimensionality, logical dimensionality, the effective combination of intellectual dimension, it may clearly understand a certain state in space and have targeted research cost management mode and method. The three-dimensional structural model can also choose two dimensions to simplify the two-dimensional planar structure, which can more intuitively understand the connection between two dimensions. According to the matrix theory, select the two dimensions of time and

logic dimensions, cross the two dimensions across the plane $m * n$ matrix of each element [34].

Using system engineering theory knowledge, the four phases of the time-dimensional time dimension in assembly buildings and seven steps of logical dimensions constitute 28 elements of the two-stage building profit cost management activity matrix, such as Table 3 Show, a_{ij} indicates the specific activity of lean cost management at all stages.

Time dimension	Logical dimension						
	Make clear the problem	Set goals	Comprehens-ive plan	Systems analysis	Scheme comparison	Make policy	Put into effect
Planning programming	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}
Construction and production	a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	a_{26}	a_{27}
Storage and transportation	a_{31}	a_{32}	a_{33}	a_{34}	a_{35}	a_{36}	a_{37}
Construction assembly	a_{41}	a_{42}	a_{43}	a_{44}	a_{45}	a_{46}	a_{47}

Table 3: Lean cost management activity matrix for prefabricated buildings

B. Assembly building cost management model based on Hall 3 D structure

Based on the above cost management ideas of prefabricated building projects of Hall's three-dimensional structure from three dimensions of the time dimension, logic dimension, and knowledge dimension, all elements are organically combined to build the lean cost management model of corresponding prefabricated building projects. Prefabricated building projects involve a large number of participants and complex uncertainties. With the advancement of all stages in the assembly time life cycle, the subject and object of cost management work have changed accordingly, so richer and extensive knowledge support is needed. Lean cost management thought closely connects the scattered stages through the knowledge dimension and time dimension. The logical dimension runs through all stages of prefabricated construction projects. For the cost management objects of different time dimensions, combined with the participants in each stage, seven logical steps are recycled, continuously decompose, and refine the cost management objectives and operations, in order to achieve the target of the overall cost control [35].

C. Example analysis

Take a single building as an example to analyze the benefits obtained in cost control. As shown in figure 3, the original design components and opening types are 72,51,51,21 molds, 125%, 129 components, 103 components, 26 components, lifting time by about 20%,

and 40 days, improving management efficiency and promoting lean control of the project cost.
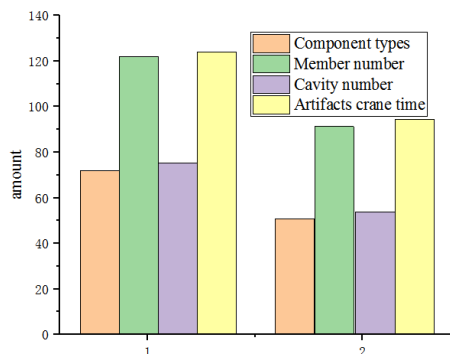

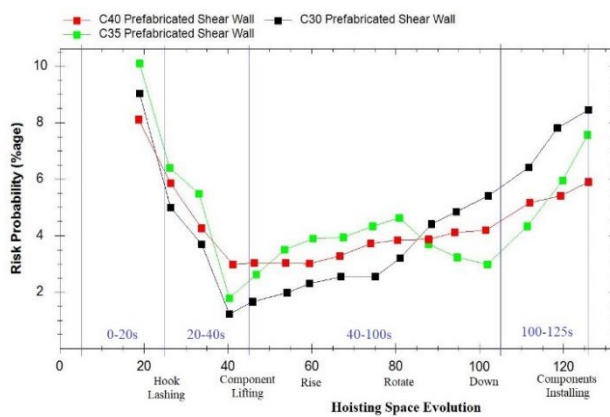Figure 3: Comparison between original design and optimization

Figure 4: Safety risk assessment

In the very first step, the BIM model is handled. This study changes over the BIM model into IFC design records, then the arrangement documents are parsed and handled in JavaScript. The BIM model was carried on the website page involving WebGL as depicted in figure 4 after the lightweight of the BIM model. This structure gives an effective information connection strategy to plan multidimensional data on location to virtual model in time. Moreover, the system can rapidly send feedback and estimation results of virtual space to directors and administrators, and work fair and square of well-being the executives.

5 Conclusions

This paper is based on systematic ideas of hall 3-dimensional structure, a Study on the construction of a lean control system of prefabricated machinery construction cost, through an analysis of the hall 3-D structure model, and the construction of a prefabricated building cost management model based on the hall 3D structure, benefit analysis of one building. The results show that the original design components and the number of open modes is 72, the optimized types of components and the number of open modes is 51, reduce 21 mold machining, mold costs were reduced by about 25%. The number of original design components and the

lifting times of components is 129 kinds, the number of components and lifting of components is 103, the number of components per layer was decreased by 26, lifting time is shortened by about 20%, the comprehensive construction period is shortened by more than 40 days, improve the management efficiency, lean control of the cost of the project plays a positive role. Combined with the participants in each stage, seven logical steps are recycled to continuously decompose and refine the cost management objectives and operations, so as to achieve the goal of the overall cost control. Due to the limited time and level, the research in this paper still has some shortcomings. In the future, BIM technology can be combined with wireless RF identification (REID) technology, the internet of things, a global positioning system (GPS), and other information technologies, to form the whole construction process of a prefabricated buildings-a system that can identify, locate and monitor prefabricated components automatically and in real-time, and more effectively control the cost of prefabricated buildings.

References

[1] Lin, J., & Lu, Y. (2018). Construction of mesoscale two-dimensional honeycomb structures: a route from self-assembly building blocks to highly-organized superstructures. *Science China Chemistry*, *61*(7), 759-760. *https://doi.org/10.1007/s11426-018-9256-x*

[2] Gu, L., Xie, M. Y., Jin, Y., He, M., Xing, X. Y., Yu, Y., & Wu, Q. Y. (2019). Construction of antifouling membrane surfaces through layer-by-layer self-assembly of lignosulfonate and polyethyleneimine. *Polymers*, *11*(11), 1782. *https://doi.org/10.3390/polym11111782*

[3] Zou, H., Sun, H., Wang, L., Zhao, L., Li, J., Dong, Z., ... & Liu, J. (2016). Construction of a smart temperature-responsive GPx mimic based on the self-assembly of supra-amphiphiles. *Soft matter*, *12*(4), 1192-1199. *https://doi.org/10.1039/C5SM02074C*

[4] Tezel, A., Koskela, L., & Aziz, Z. (2018). Current condition and future directions for lean construction in highways projects: A small and medium-sized enterprises (SMEs) perspective. *International Journal of project management*, *36*(2), 267-286. *https://doi.org/10.1016/j.ijproman.2017.10.004*

[5] Xing, W., Hao, J. L., Qian, L., Tam, V. W., & Sikora, K. S. (2021). Implementing lean construction techniques and management methods in Chinese projects: A case study in Suzhou, China. *Journal of cleaner production*, *286*, 124944. *https://doi.org/10.1016/j.jclepro.2020.124944*

[6] Goger, G., & Bisenberger, T. (2018). Tunnelling 4.0–Construction-related future trends: Tunnelbau 4.0–Baubetriebliche Zukunftstrends. *Geomechanics*

and Tunnelling, *11*(6), 710-721.
https://doi.org/10.1002/geot.201800058

[7]   Kabirifar, K., Mojtahedi, M., Wang, C., & Tam, V. W. (2020). Construction and demolition waste management contributing factors coupled with reduce, reuse, and recycle strategies for effective waste management: A review. *Journal of Cleaner Production*, *263*, 121265.
https://doi.org/10.1016/j.jclepro.2020.121265

[8]   Aslam, M. S., Huang, B., & Cui, L. (2020). Review of construction and demolition waste management in China and USA. *Journal of Environmental Management*, *264*, 110445.
https://doi.org/10.1016/j.jenvman.2020.110445

[9]   Akhtar, A., & Sarmah, A. K. (2018). Construction and demolition waste generation and properties of recycled aggregate concrete: A global perspective. *Journal of Cleaner Production*, *186*, 262-281.
https://doi.org/10.1016/j.jclepro.2018.03.085

[10] Tam, V. W., Soomro, M., & Evangelista, A. C. J. (2018). A review of recycled aggregate in concrete applications (2000–2017). *Construction and Building materials*, *172*, 272-292.
https://doi.org/10.1016/j.conbuildmat.2018.03.240

[11] Huang, B., Wang, X., Kua, H., Geng, Y., Bleischwitz, R., & Ren, J. (2018). Construction and demolition waste management in China through the 3R principle. *Resources, Conservation and Recycling*, *129*, 36-44.
https://doi.org/10.1016/j.resconrec.2017.09.029

[12] Yuan, H. (2013). Key indicators for assessing the effectiveness of waste management in construction projects. *Ecological Indicators*, *24*, 476-484.
https://doi.org/10.1016/j.ecolind.2012.07.022

[13] Aslam, M. S., Huang, B., & Cui, L. (2020). Review of construction and demolition waste management in China and USA. *Journal of Environmental Management*, *264*, 110445.
https://doi.org/10.1016/j.jenvman.2020.110445

[14] Wahi, N., Joseph, C., Tawie, R., & Ikau, R. (2016). Critical review on construction waste control practices: legislative and waste management perspective. *Procedia-Social and Behavioral Sciences*, *224*, 276-283.
https://doi.org/10.1016/j.sbspro.2016.05.460

[15] Liu, C., Lin, M., Rauf, H. L., & Shareef, S. S. (2021). Parameter simulation of multidimensional urban landscape design based on nonlinear theory. *Nonlinear Engineering*, 10(1), 583-591.
https://doi.org/10.1515/nleng-2021-0049

[16] Singh, P. K., & Sharma, A. (2022). An intelligent WSN-UAV-based IoT framework for precision agriculture application. *Computers and Electrical Engineering, 100, 107912*.

https://doi.org/10.1016/j.compeleceng.2022.107912

[17] Zeng, H., Dhiman, G., Sharma, A., Sharma, A., & Tselykh, A. (2021). An IoT and Blockchain-based approach for the smart water management system in agriculture. *Expert Systems*, e12892.
https://doi.org/10.1111/exsy.12892

[18] Sharma, A., & Singh, P. K. (2021). UAV-based framework for effective data analysis of forest fire detection using 5G networks: An effective approach towards smart cities solutions. *International Journal of Communication Systems*, e4826.
https://doi.org/10.1002/dac.4826

[19] Sharma, A., Singh, P. K., & Kumar, Y. (2020). An integrated fire detection system using IoT and image processing technique for smart cities. *Sustainable Cities and Society*, *61*, 102332.
https://doi.org/10.1016/j.scs.2020.102332

[20] Yang, W. S., Park, J. K., Park, S. W., & Seo, Y. C. (2015). Past, present and future of waste management in Korea. *Journal of Material Cycles and Waste Management*, *17*(2), 207-217.
https://doi.org/10.1007/s10163-014-0301-7

[21] Akinade, O. O., Oyedele, L. O., Munir, K., Bilal, M., Ajayi, S. O., Owolabi, H. A., & Bello, S. A. (2016). Evaluation criteria for construction waste management tools: towards a holistic BIM framework. *International Journal of Sustainable Building Technology and Urban Development*, *7*(1), 3-21.
https://doi.org/10.1080/2093761X.2016.1152203

[22] Umar, U. A., Shafiq, N., Malakahmad, A., Nuruddin, M. F., & Khamidi, M. F. (2017). A review on adoption of novel techniques in construction waste management and policy. *Journal of Material Cycles and Waste Management*, *19*(4), 1361-1373.
https://doi.org/10.1007/s10163-016-0534-8

[23] Ge, X. J., Livesey, P., Wang, J., Huang, S., He, X., & Zhang, C. (2017). Deconstruction waste management through 3d reconstruction and bim: a case study. *Visualization in engineering*, *5*(1), 1-15.
https://doi.org/10.1186/s40327-017-0050-5

[24] Lu, W., Webster, C., Chen, K., Zhang, X., & Chen, X. (2017). Computational Building Information Modelling for construction waste management: Moving from rhetoric to reality. *Renewable and Sustainable Energy Reviews*, *68*, 587-595.
https://doi.org/10.1016/j.rser.2016.10.029

[25] Won, J., & Cheng, J. C. (2017). Identifying potential opportunities of building information modeling for construction and demolition waste management and minimization. *Automation in Construction*, *79*, 3-18.
https://doi.org/10.1016/j.autcon.2017.02.002

[26] Li, C. Z., Zhao, Y., Xiao, B., Yu, B., Tam, V. W., Chen, Z., & Ya, Y. (2020). Research trend of the application of information technologies in construction and demolition waste management. *Journal of Cleaner Production*, *263*, 121458.
https://doi.org/10.1016/j.jclepro.2020.121458

[27] Akinade, O. O., Oyedele, L. O., Ajayi, S. O., Bilal, M., Alaka, H. A., Owolabi, H. A., & Arawomo, O. O. (2018). Designing out construction waste using BIM technology: Stakeholders' expectations for industry deployment. *Journal of cleaner production*, *180*, 375-385.
https://doi.org/10.1016/j.jclepro.2018.01.022

[28] Jupp, J. (2017). 4D BIM for environmental planning and management. *Procedia engineering*, *180*, 190-201.
https://doi.org/10.1016/j.proeng.2017.04.178

[29] Martins, S. S., Evangelista, A. C. J., Hammad, A. W., Tam, V. W., & Haddad, A. (2020). Evaluation of 4D BIM tools applicability in construction planning efficiency. *International Journal of Construction Management*, 1-14.
https://doi.org/10.1080/15623599.2020.1837718

[30] Ren, Y., Rubaiee, S., Ahmed, A., Othman, A. M., & Arora, S. K. (2022). Multi-objective optimization design of steel structure building energy consumption simulation based on genetic algorithm. *Nonlinear Engineering*, 11(1), 20-28.
https://doi.org/10.1515/nleng-2022-0012

[31] Jahromi, A. E., & Miller, F. K. (2016). Construction and experimental validation of a simple, compact, resealable, and reliable Vycor® superleak assembly for use at low temperatures. *Review of Scientific Instruments*, *87*(4), 045112.
https://doi.org/10.1063/1.4947232

[32] Endo, N., Shimoda, E., Goshome, K., Yamane, T., Nozu, T., & Maeda, T. (2019). Construction and operation of hydrogen energy utilization system for a zero emission building. *International Journal of Hydrogen Energy*, *44*(29), 14596-14604.
https://doi.org/10.1016/j.ijhydene.2019.04.107

[33] Wang, T., Gao, S., Li, X., & Ning, X. (2018). A meta-network-based risk evaluation and control method for industrialized building construction projects. *Journal of Cleaner Production*, *205*, 552-564.
https://doi.org/10.1016/j.jclepro.2018.09.127

[34] Hamdaoui, S., Mahdaoui, M., Allouhi, A., El Alaiji, R., Kousksou, T., & El Bouardi, A. (2018). Energy demand and environmental impact of various construction scenarios of an office building in Morocco. *Journal of cleaner production*, *188*, 113-124.
https://doi.org/10.1016/j.jclepro.2018.03.298

[35] Liu, Q., Zhang, W., Bhatt, M. W., & Kumar, A. (2021). Seismic nonlinear vibration control algorithm for high-rise buildings. *Nonlinear Engineering*, 10(1), 574-582.
https://doi.org/10.1515/nleng-2021-0048

Design and Study of Urban Rail Transit Security System Based on Face Recognition Technology

[1]Zhan Guo, [1]Zuming Xiao, [2]Roobaea Alroobaea, [3]Abdullah M. Baqasah, [4]Anas Althobaiti, [5*]Harsimranjit Singh Gill

[1]Mechanical and Electronic Department, Jingdezhen University, Jingdezhen Jiangxi, 333400, China
[2]Department of Computer Science, College of Computers and Information Technology,
Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia
[3]Department of Information Technology, College of Computers and Information Technology, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia
[4]College of Environment and Technology, Bristol University, Bristol Bs1, United Kingdom
[5]Guru Nanak Dev Engineering College, Ludhiana, India
Emails: guozhan91@163.com,   xiaozuming7@126.com,   r.robai@tu.edu.sa,   a.baqasah@tu.edu.sa, anasuk.tu@gmail.com, harsimran18@gmail.com

In the modern world, it is difficult to prevent terrorism due to the relatively closed environment, dense personnel, large passenger flow, long line and wide coverage of urban rail transit. Identity recognition is a core element of security. The design and study of an urban rail transit security system based on face recognition technology are proposed in this paper. Through the study on the face recognition algorithm of intelligent security systems in urban rail transit, the related introduction of face recognition technology is done. The analysis of the main mode of face recognition is carried out utilizing the practical application design ideas. The results by experimental analysis show that if FAR is set to a very low range (such as 0.1% or even 0.01%) meanwhile FRR can reach a very low level (such as less than 1%). Such a system has practical value and otherwise, it may face a large number of passenger affairs and complaints to be handled. When FAR is set to 0.1% and N is 1.6 million, FRR can reach 2.1%. However, according to the test, when the picture quality deteriorates (during image captured by a webcam), the FRR will increase by 2 to 3 times. If a Webcam is used for recognition in Mugshot, the lowest FRR of the three top algorithms is only 5.21%.

Povzetek: Tehnologija prepoznavanja obrazov je uporabljena za nadzor osumljencev - teroristov na vlakih.

1   Introduction

At present, each urban rail transit is equipped with a video surveillance system. A large number of surveillance cameras are installed within the station area for subway operation and public security monitoring. Such a video monitoring system has almost become a tool to provide post-evidence recording and has lost the ability to prevent or stop criminal activities from occurring [1]. Urban rail transit has small space and large passenger flow, so safe operation is always the most concerning thing for government departments, operating units, and public security organs. How to identify the dangerous elements hidden in the crowd timely and accurately when entering the scope of rail transit is an urgent problem to be solved by the operation unit and the Ministry of public security. With the progress of society, some high technologies have been continuously used in industrial or civilian production and life. As the most advanced biometric technology and image processing technology in the world, face recognition technology is developing and improving day by day and is constantly applied to various fields in society. The technical development of the urban rail transit integrated monitoring system is shown in Figure 1 [2]. As a highly intelligent security monitoring means, face recognition technology is gradually applied in various fields of society. The introduction of a face recognition system in urban rail transit will certainly help to reduce the work pressure of public security personnel, provide good technical support for the safe operation of urban rail transit and criminal investigation and investigation, contribute to the personal and property safety of passengers, and maintain social stability. But due to the particularity of urban rail transit, the application of the technology in urban rail transit also needs to constantly improved. Many problems hinder face recognition technology in the further application of urban rail transit, so need continuous research to improve the technology.

Based on the current research, this paper proposes the design and research of an urban rail transit security system based on face recognition technology, through the study of face recognition algorithm and face recognition technology and the analysis of the main mode of face recognition. Conduct the practical application of the design ideas. The results show that by setting False

Acceptance Rate (FAR) to a very low range, such as 0.1% or even 0.01%, while False Rejection Rate (FRR) can reach a very low level, such as below 1%, the system has utility, otherwise, the application may face a large number of passenger affairs and complaints to be handled. When the FAR is set at 0.1%, and N = 1.6 million, the FRR can reach 2.1%. However, according to the test, when the picture quality deteriorates, such as

Webcam (images collected by web camera), the FRR will increase by 2 to 3 times. If Webcam is identified in Mugshot, the lowest FRR of the three top algorithms is only 5.21%. The application of face recognition in the rail transit Automated Fare collection (AFC) system puts forward higher requirements for the integration, processing, and analysis ability of data.
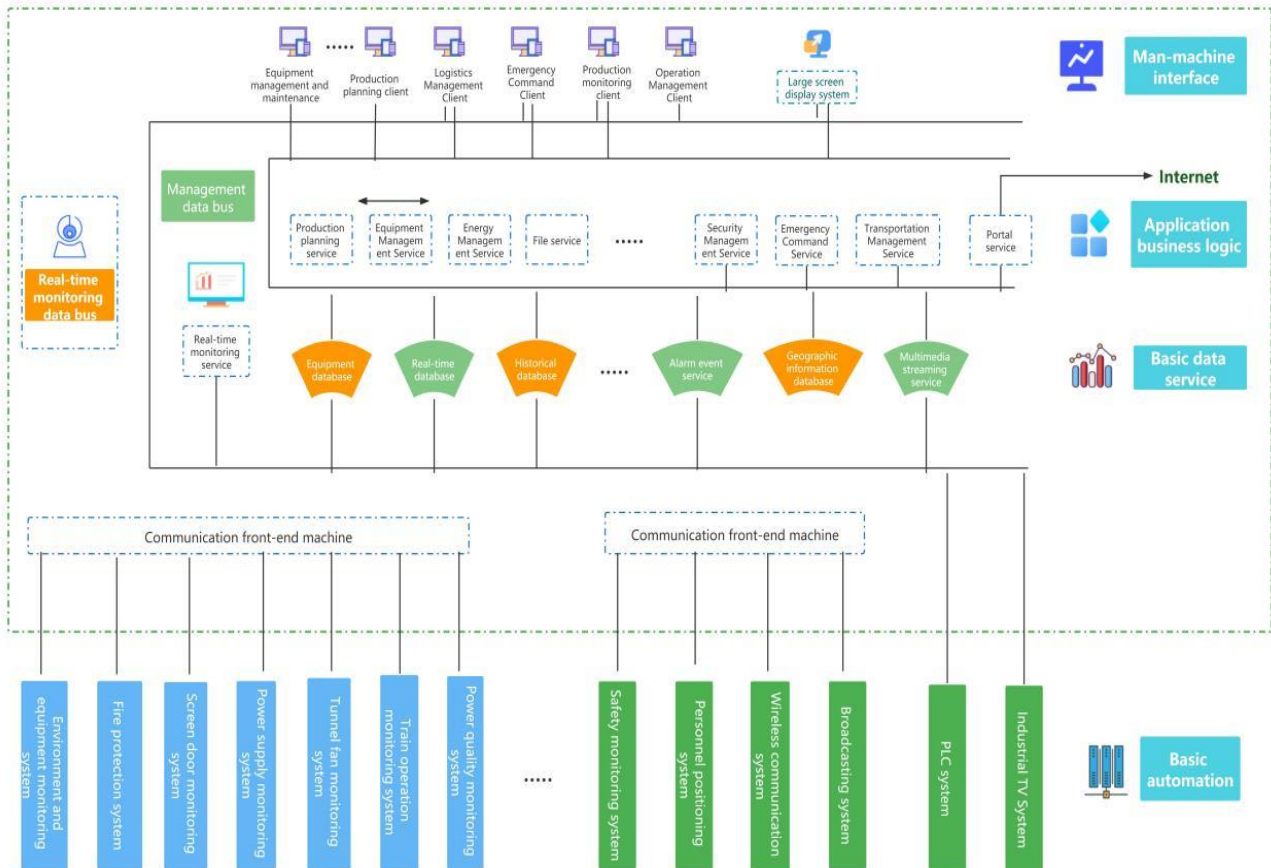
Figure 1: Technical development process of urban rail transit integrated monitoring system

It also provides a good platform for the development and application of big data, cloud computing, Internet of Things, artificial intelligence and other technologies.

To precisely acquire the concentration level of the travelers, it is necessary to precisely get the traveler stream of each key area. This paper plans identification gear for assessing traveler stream, which has incorporated the data acquisition equipment and algorithm process. From the assortment of image recognition, the most recent examination aftereffects of profound learning are utilized to recognize the traveler stream, and the blunder of the video traveler stream is rectified with the WIFI test hardware. To screen the early and late peak time of the megacities and the traveler stream in the vital place of the tram station, the abrupt occasions in the station and the activity span are decreased, while diminishing the unexpected occasions.

The rest of this article is organized as: Section 2 presents the related works in various domains. Section 3

consists of methods comprising the concept. Results and analysis are discussed in Section 4 followed by concluding remarks in section 5.

2 Literature review

In the literature, Yanpeng *et al.* analyzed the main technical indicators for the application of face recognition in urban rail transit AFC system. They investigated the technical level of the current top face recognition algorithms in the world, and proposed a design idea for the large-scale application of 1: N face recognition system in urban rail transit AFC system [3]. Feng *et al.* believe that the recognition server outputs the comparison results and alarms to the station's local monitoring terminal and remote monitoring center according to the comparison results [4]. Chen *et al.* for rail transit industry in recent years rapid development needs of public security and intelligence operations proposes a face recognition system based on intelligent

subway design. The face recognition function design, front-end acquisition unit selection and deployment location, overall system architecture design is analyzed and discussed, and for other domestic embarks on a facial recognition system of the construction of the rail transit to provide reference [5]. Li *et al.* believe that the extraction of facial features is a key step in face recognition, which directly affects the accuracy of recognition. It is completed by the video analysis server set up in the station. The face recognition process is shown in Figure 2 [6]. Gao *et al.* present an article that puts forward the design and implementation method of

the technology so that the school can complete the technical design and management of the student attendance system through the application of the equipment, and give full play to the supervision effect of the equipment [7]. Liu *et al.* present their view of the inability to obtain real-time learning status of learners in online learning. Their paper uses face recognition technology to monitor and analyze learners' learning status in front of the camera.
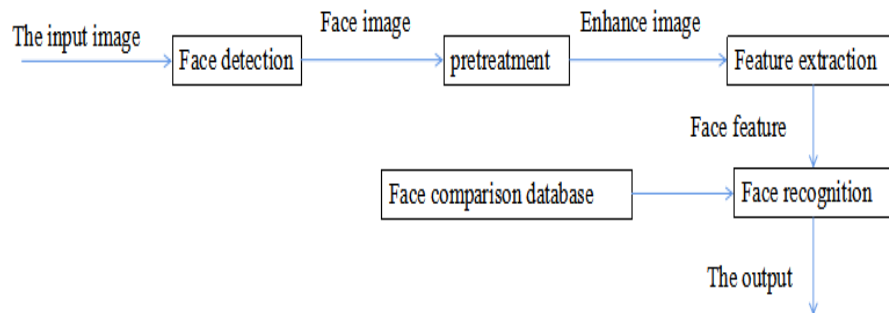
Figure 2: Face recognition process

The authors have designed and developed an intelligent supervising assistant system based on face recognition. The system collects students' images in real-time through the camera, establishes a mathematical model according to the main features of the face, extracts the relevant characteristic values, obtains the data such as face plane, three-dimensional rotation Angle, and eye closure state, and judges the learning state of learners. The experiment proves that the system can better assist managers to monitor the learning state of learners and improve learners' concentration during online learning [8]. Zhu *et al.* display information, such as the similarity ratio of face recognition, the location of the face, etc., through a graphical user interface, and switch the relevant video image to the monitoring terminal for easy tracking and monitoring [9]. Lu *et al.* study and analyze large-angle deflection face mainly by three-dimensional facial feature extraction. Three-dimensional facial feature extraction is a method of extracting facial feature points and head posture information from a given area [10]. Shen *et al.* believe that the location of facial features needs to adapt to the changes in various aspects of the face in different positions to the greatest extent, which can further improve the accuracy of the algorithm [11]. Shi and Lei believe that image preprocessing techniques include geometric ruler normalization, glasses extraction, and image gray-scale attribute correction. When normalizing the geometric size, the eyes and jaw points are automatically positioned, the two eyes are aligned by scaling and rotation, and the distance between the two centers of jaw points is a predefined constant, and then the image is cropped to a fixed size [12].

Although face recognition technology is in continuous development and improvement [13-20], the recognition rate, anti-counterfeiting, fingerprint, retina, and other has a large gap [21-25]. The large passenger flow and complex environmental characteristics of urban

rail transit also affect the application of face recognition technology in urban rail transit. In this regard, the application of a face recognition system in urban rail transit should also pay attention to the following aspects: the uncertainty in the process of video image acquisition [26-29]. Due to the complex environment of urban rail

transit, such as lighting, installation location, occlusion, and human posture, the acquired video and image quality are different, and the face acquisition is unclear or lacking [30-32], which affects the recognition rate. Therefore, in the installation and selection of the camera should pay attention to the installation position, light, lens exposure angle, and the selection of wide dynamic function, to improve the quality of video images [33-37].

Diversity of face patterns and uncertainty about-face plastic deformation. Because the same face has a diversity of faces such as beard, glasses, hairstyles, and the shape and deformation of different expressions, it affects the precise extraction of facial characteristics. The use of more advanced image processing technology and stable and accurate face expression method is the basis of the wide application of face recognition technology in urban rail transit. Urban rail transit has the characteristics of large passenger flow [38]. There are often many faces in a video image, and the task of video analysis and comparison is large. To realize real-time investigation, it is necessary to configure the number of video analysis

servers and comparison servers reasonably [39], and the face detection and comparison technology are constantly optimized and improved. The face recognition system based on the original video monitoring system must establish a linkage with the original video monitoring system, make full use of the advantages of large-scale monitoring coverage, and realize the tracking of personnel through linkage control, to improve the case handling efficiency of police officers [40].

3 Research method

3.1 Study on the Face Recognition Algorithm of Intelligent Security System in Urban Rail Transit

The Adaboost face detection algorithm mainly uses the gray distribution characteristics of the face area to construct the classifier. The first step is to extract the Haar feature of the face gray distribution and use the integral diagram to quickly calculate the feature value. The second step is to use the weighted voting method to achieve the construction of the Adaboost strong classifier. The third step is to obtain a stronger joint classifier. The Haar feature represents a simple rectangular feature. Haar features can reflect a variety of image features, including horizontal, vertical, edge, center, linear and diagonal features [41].

Each Haar feature corresponds to a weak classifier, and the definition formula is:

$$h_i(x) = \begin{cases} 1, if, P_i f_i(x) \leq P_i \theta_i \\ 0, other \end{cases} \quad (1)$$

From equation 1: $f_i(x)$ represents the rectangular feature value of the ith rectangle, $i(x)$ represents the classification result of the rectangular feature i on x. When the value is 0, it means that it is a non-human sample. If its value is 1, it means that it is a human face sample. P_i is used to determine the direction of the inequality, and θ_i represents the optimal threshold of the rectangular feature i [42].

3.2 Overview of face recognition technology

Face recognition technology is a technology that combines digital image processing, computer graphics, pattern recognition, visualization technology, human physiology, cognitive science, psychology, and other research fields to analyze the collected face graphics, determine the position, size, and posture of the face, and extract effective recognition information for face feature comparison, to realize identity recognition [43].

3.3 Advantages of face recognition over other biometric technologies

Biometrics is a technology that uses the inherent physiological or behavioral characteristics of the human body to perform identification. The physiological or behavioral characteristics require universality (covering a wide range of people) and differences (there should be identifiable differences between different individuals), stability (will not change within a certain period), and vitality (cannot be simulated by simulation) [44]. The comparison of their technical characteristics is shown in Tables 1 and 2 below.

Project	Face recognition	Palm vein recognition
Deployment cost	Non-contact type	High
Data acquisition	No feeling, no need to cooperate	Non-contact type
Accessibility	Fast	Palm extension fit
Recognition speed	High	Common
Safety	Light and dress blocking; Age change	Very high
Possible interference	Face recognition	Age, physiological changes

Table 1: Technical features of face recognition and palm vein recognition

Project	Fingerprint recognition	Iris recognition
Deployment cost	Contact type	medium
Data acquisition	Finger extension fit	Non-contact type
Accessibility	Fast	High fit required
Recognition speed	Common	Slow
Safety	Dirt and skin wear	Very high
Possible interference	Fingerprint recognition	Contact lenses

Table 2: Technical characteristics of fingerprint recognition and iris recognition

It can be seen from Tables 1 and 2 that although palm vein recognition has advantages in terms of safety and stability, it is not easy to promote due to the cost of equipment and recognition speed. Fingerprint recognition is a contact check, which has problems of insanity and susceptibility to interference. Iris recognition is difficult to operate and slow to recognize [45-49].

3.4 Face recognition training process

Face tracking function was tested for different face conditions, including single face, multiple faces, and face in and out or interleaved [50]. The system automatically initializes a tracking window before the face tracking starts, and assigns a Camshift tracker to each face in the

video surveillance to realize multiple face recognition. Through many system function tests, it is found that the Camshift algorithm can have a good tracking effect and robustness in practical application, even for large Angle deflection or position change of face [51-54]. Even if there is a complex situation of people in and out of the area or face interleave, the system can normally start the face detection program and constantly correct the face tracking window, the system can quickly obtain all the face areas in each frame of the image, and is more rapid and stable than the face detection frame by frame.

4 Results and Analysis

The following key indicators are used to replace traditional physical tickets with facial features to realize face-to-pay rides in urban rail transit. First, the face feature database N: the number of registered person

images in the system's face database [55]. Second, the false acceptance rate (FAR): refers to the probability of identifying an unregistered user as a registered user. Suppose that in a face recognition test, the threshold is set to T. If the comparison value of a test object is greater than T, it is considered that the object should be recognized as true, as presented in Equation 2.

One 1: N face recognition can be regarded as one face recognition performed N times, as presented in Equation 3. When FAR(1, T) is very small, FAR(N, T)=N(FAR(1, I), it can be seen that FAR will increase linearly as N increases [56-60]. Third, the false rejection rate (FRR): refers to the probability that registered users are rejected and is presented in Equation 4.

$$FAR(N,T) = \frac{\text{Comparison value of non-registered personnel} > \text{The number of T}}{\text{Total number of comparisons in the library by non-registered personnel}} \quad (2)$$

$$FAR(N,T) = 1 - \left(1 - FAR(1,t)\right)^N \quad (3)$$

$$FAR(N,T) = \frac{\text{Comparison value of non-registered personnel} > \text{The number of T}}{\text{Total number of comparisons in the library by non-registered personnel}} \quad (4)$$

Since a registrant is recognized as truly independent of other comparison data, it is generally considered that FRR (N, T) = FRR (1, T). But in actual tests, since the registration database is not a linear data structure, it may be an index or tree structure, causing FRR (N, T) to slowly increase as N increases. Fourth, the recognition time t: the time from the extraction of features from the face image to the completion of the feature comparison.

In the AFC system, FAR determines the safety of the system, FRR determines the accessibility of the system, and t determines the passing speed of the gate. For most face recognition algorithms, the threshold T used by the computer for face recognition is different, and FAR and FRR are also different. FAR increases with the decrease of T (relaxation conditions), and FRR decreases with the decrease of T. Therefore, FAR and FRR are almost contradictory indicators. Different applications have different requirements for FAR and FRR indexes, as shown in Figure 3. High-security applications have a low tolerance for FAR. For public security agencies to find personnel, it is suitable to use a lower FN.

Figure 3: Applications for different indicators

For urban rail transit, due to the huge passenger flow and the money transaction involved, it should be classified as a high-security application. Therefore, the FAR needs to be set to a very low range, such as 0.1% or even 0.01%, and the FRR can reach a very low level, such as less than 1%. Such a system has practical value. Otherwise, once applied, it may face a large number of passenger affairs and complaints to be handled.

Figures 4, 5, and 6 show the performance curves of the three top international algorithms in the latest Face Recognition Vendor Test (FRVT) evaluation. The processed images are all Mugshot (face photos). When FAR is set to 0.1% and N is 1.6 million, FRR can reach 2.1%. However, according to the test, when the picture

quality deteriorates, such as when using a Webcam (image captured by a web camera), the FRR will increase by 2 to 3 times. If a Webcam is used for recognition in Mugshot, the lowest FRR of the three top algorithms is only 5.21%.
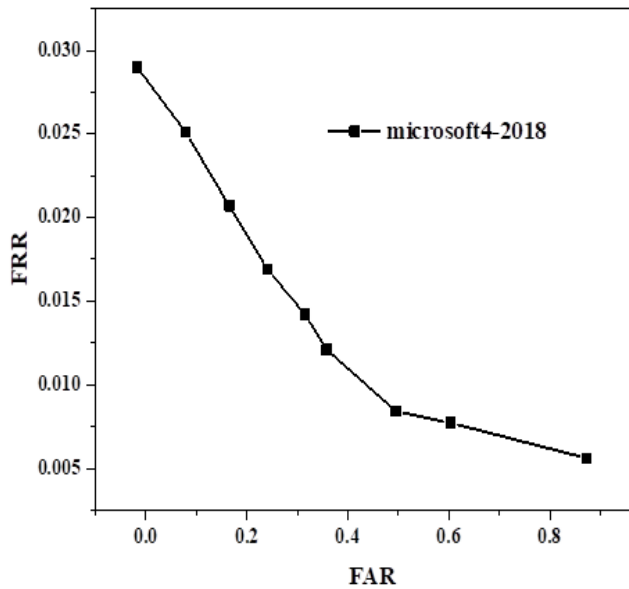

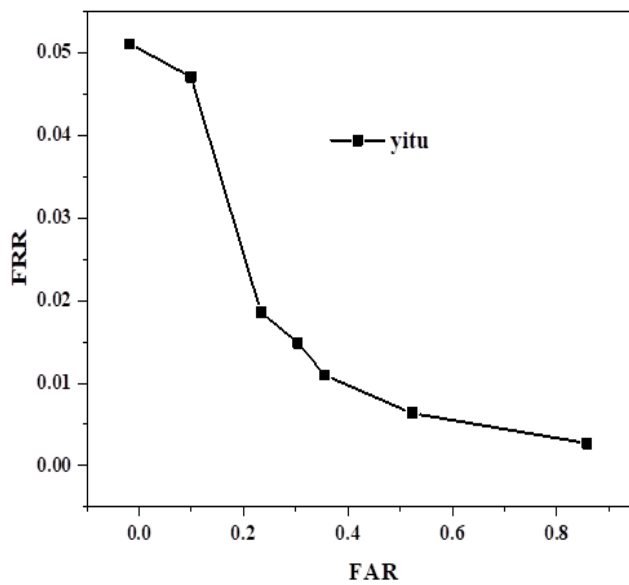Figure 4: Performance curve of microsoft4-2018 top algorithms
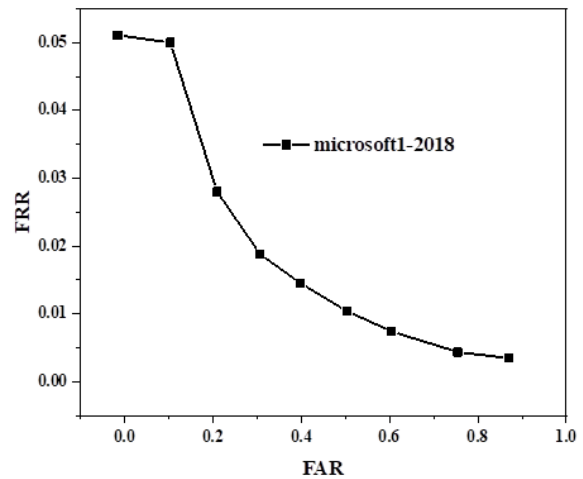

Figure 5: Performance curve of yitu top algorithm

Figure 6: Performance curve of microsoft1-2018 top algorithms

Therefore, the actual application effect of face recognition is often different from the evaluation results. At present, domestic manufacturers focusing on intelligent face recognition, such as Yitu, Me gvii, Sense time, and other companies, have launched 1: N face recognition products, but at present, they can only meet the above requirements of FAR and FRR and achieve the face feature library N to the order of tens of thousands of yuan, and the recognition time t is controlled within 1s. It can be seen that with the current computer processing ability and face recognition algorithm performance, it cannot be directly applied on a large scale, and only small-scale pilot ideas can be carried out using specific personnel schemes, single-line commuter passenger schemes, and post-payment based on the third-party payment platform.

5   Conclusion

The urban rail transit should be classified as high-security application due to the huge passenger flow and the money transaction involved. This article presents the design idea of large application 1: N face recognition system in urban rail transit AFC system. The designed system can realize various functions, and has higher detection efficiency, lower error rate, and better application value. Therefore, the FAR needs to be set to a very low range, such as 0.1% or even 0.01%, and the FRR can reach a very low level, such as less than 1%. However, according to the test, when the picture quality deterior, such as Webcam (images collected by web camera), the FRR will increase by 2 to 3 times. If Webcam is identified in Mugshot, the lowest FRR of the three top algorithms is only 5.21%. The application of face recognition in the rail transit AFC system puts forward higher requirements for the integration, processing, and analysis ability of data. However, it also provides a good platform for the development and application of big data, cloud computing, Internet of Things, artificial intelligence and other technologies. It is believed that the future implications of this research

work lie in further exploration of face recognition technology which will certainly shine in the rail transit industry.

Acknowledgments:

References

[1] Li, S., Wu, S., Xiang, S., Zhang, Y., Guerrero, J. M., & Vasquez, J. C. (2020). Research on synchronverter-based regenerative braking energy feedback system of urban rail transit. *Energies*, *13*(17), 4418.

[2] Nie, X. L., & Wei, Q. C. (2015). Research on the location of emergency rescue stations for urban rail transit based on the PSO [J]. *Journal of Railway Engineering Society*, *32*(07), 100-105.

[3] Yanpeng, Z., Jianwu, D., Xiaojuan, L., & Fan, Y. (2014). Optimized Handover Algorithm Based on Stackelberg Games in CBTC Systems for Urban Rail Transit. *Journal of Engineering Science & Technology Review*, *7*(3).

[4] Feng, X., Zhang, H., Gan, T., Sun, Q., Ma, F., & Sun, X. (2016). Random coefficient modeling research on short-term forecast of passenger flow into an urban rail transit station. *Transport*, *31*(1), 94-99.

[5] Chen, H., Wang, B., He, W., & Zheng, J. (2020). Research on passenger flow early warning of urban rail transit station based on system dynamics. In *MATEC Web of Conferences* (Vol. 308, p. 01003). EDP Sciences.

[6] Li, J., Wang, J., Xu, N., Hu, Y., & Cui, C. (2018). Importance degree research of safety risk management processes of urban rail transit based on text mining method. *Information*, *9*(2), 26.

[7] Gao, H., Liu, S., Cao, G., Zhao, P., Zhang, J., & Zhang, P. (2020). Big data analysis of beijing urban rail transit fares based on passenger flow. *IEEE Access*, *8*, 80049-80062.

[8] Liu, H., Xie, Y., Liu, Y., Nie, R., & Li, X. (2019). Mapping the knowledge structure and research evolution of urban rail transit safety studies. *IEEE Access*, *7*, 186437-186455.

[9] Zhu, K., Xun, P., Li, W., Li, Z., & Zhou, R. (2019). Prediction of passenger flow in urban rail transit based on big data analysis and deep learning. *IEEE Access*, *7*, 142272-142279.

[10] Lu, Q. C., Zhang, L., Xu, P. C., Cui, X., & Li, J. (2022). Modeling network vulnerability of urban rail transit under cascading failures: A Coupled Map Lattices approach. *Reliability Engineering & System Safety*, *221*, 108320.

[11] Shen, X., Wei, H., & Lie, T. T. (2020). Management and utilization of urban rail transit regenerative braking energy based on the bypass DC loop. *IEEE Transactions on Transportation Electrification*, *7*(3), 1699-1711.

[12] Shi, Z., Zhang, N., & Zhu, L. (2019). Understanding the Propagation and Control Strategies of Congestion in Urban Rail Transit Based on Epidemiological Dynamics Model. *Information*, *10*(8), 258.

[13] Amhoud, E. M., Chafii, M., Nimr, A., & Fettweis, G. (2021, April). OFDM with index modulation in orbital angular momentum multiplexed free space optical links. In *2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring)* (pp. 1-5). IEEE.

[14] Gill, H. S., Singh, T., Kaur, B., Gaba, G. S., Masud, M., & Baz, M. (2021). A metaheuristic approach to secure multimedia big data for IoT-based smart city applications. *Wireless Communications and Mobile Computing*, *2021*.

[15] Kumar, A., Sehgal, V. K., Dhiman, G., Vimal, S., Sharma, A., & Park, S. (2022). Mobile networks-on-chip mapping algorithms for optimization of latency and energy consumption. *Mobile Networks and Applications*, *27*(2), 637-651.

[16] Boguszewicz, C., Boguszewicz, M., Iqbal, Z., Khan, S., Gaba, G. S., Suresh, A., & Pervaiz, B. (2021). The fourth industrial revolution-cyberspace mental wellbeing: Harnessing science & technology for humanity. *Global foundation for cyber studies and research*.

[17] Amhoud, E. M., Othman, G. R. B., & Jaouën, Y. (2017). Concatenation of space-time coding and FEC for few-mode fiber systems. *IEEE Photonics Technology Letters*, *29*(7), 603-606.

[18] Amhoud, E. M., Othman, G. R. B., Bigot, L., Song, M., Andresen, E. R., Labroille, G., & Jaouën, Y. (2017, September). Experimental demonstration of space-time coding for MDL mitigation in few-mode fiber transmission systems. In *2017 European Conference on Optical Communication (ECOC)* (pp. 1-3). IEEE.

[19] Singh, G. (2021). *Privacy-preserving authentication and key exchange mechanisms in internet of things applications* (Doctoral dissertation, Lovely Professional University Punjab).

[20] Choudhary, K., Gaba, G. S., Miglani, R., Kansal, L., & Kumar, P. (2021). Artificial intelligence and machine learning aided blockchain systems to address security vulnerabilities and threats in the industrial Internet of things. Institution of Engineering and Technology.

[21] Zerhouni, K., Amhoud, E. M., & Chafii, M. (2021). Filtered multicarrier waveforms classification: a deep learning-based approach. *IEEE Access*, *9*, 69426-69438.

[22] Gaba, G. S., Kumar, G., Monga, H., Kim, T. H., Liyanage, M., & Kumar, P. (2020). Robust and lightweight key exchange (LKE) protocol for industry 4.0. *IEEE Access*, *8*, 132808-132824.

[23] Sharma, A., & Kumar, N. (2021). Third eye: an intelligent and secure route planning scheme for

critical services provisions in internet of vehicles environment. *IEEE Systems Journal*, *16*(1), 1217-1227.

[24] Kumar, P., & Gaba, G. S. (2020). Biometric-based robust access control model for industrial internet of things applications. *IoT Security: Advances in Authentication*, 133-142.

[25] Hedabou, M. (2018). Cryptography for Addressing Cloud Computing Security, Privacy, and Trust Issues. In *Computer and Cyber Security* (pp. 281-304). Auerbach Publications.

[26] Igarramen, Z., & Hedabou, M. (2017, October). FADETPM: Novel approach of file assured deletion based on trusted platform module. In *International Conference of Cloud Computing Technologies and Applications* (pp. 49-59). Springer, Cham.

[27] Azougaghe, A., Hedabou, M., & Belkasmi, M. (2015, December). An electronic voting system based on homomorphic encryption and prime numbers. In *2015 11th International Conference on Information Assurance and Security (IAS)* (pp. 140-145). IEEE.

[28] Bentajer, A., Hedabou, M., Abouelmehdi, K., Igarramen, Z., & El Fezazi, S. (2019). An IBE-based design for assured deletion in cloud storage. *Cryptologia*, *43*(3), 254-265.

[29] Gaba, G. S., Kumar, G., Monga, H., Kim, T. H., & Kumar, P. (2020). Robust and lightweight mutual authentication scheme in distributed smart environments. *IEEE Access*, *8*, 69722-69733.

[30] Hedabou, M., Bénéteau, L., & Pinel, P. (2008). Some ways to secure elliptic curve cryptosystems. *Advances in applied Clifford algebras*, *18*(3), 677-688.

[31] Gaba, G. S., Kumar, G., Kim, T. H., Monga, H., & Kumar, P. (2021). Secure device-to-device communications for 5g enabled internet of things applications. *Computer Communications*, *169*, 114-128.

[32] Sharma, A., Podoplelova, E., Shapovalov, G., Tselykh, A., & Tselykh, A. (2021). Sustainable smart cities: convergence of artificial intelligence and blockchain. *Sustainability*, *13*(23), 13076.

[33] Bentajer, A., Hedabou, M., Abouelmehdi, K., & Elfezazi, S. (2018). CS-IBE: a data confidentiality system in public cloud storage system. *Procedia computer science*, *141*, 559-564.

[34] Azougaghe, A., Oualhaj, O. A., Hedabou, M., Belkasmi, M., & Kobbane, A. (2016, October). Many-to-one matching game towards secure virtual machines migration in cloud computing. In *2016 International Conference on Advanced Communication Systems and Information Security (ACOSIS)* (pp. 1-7). IEEE.

[35] Masud, M., Gaba, G. S., Choudhary, K., Hossain, M. S., Alhamid, M. F., & Muhammad, G. (2021). Lightweight and anonymity-preserving user authentication scheme for IoT-based healthcare. *IEEE Internet of Things Journal*, *9*(4),

2649-2656.

[36] Sharma, A., Singh, P. K., Sharma, A., & Kumar, R. (2019). An efficient architecture for the accurate detection and monitoring of an event through the sky. *Computer Communications*, *148*, 115-128.

[37] Masud, M., Gaba, G. S., Choudhary, K., Alroobaea, R., & Hossain, M. S. (2021). A robust and lightweight secure access scheme for cloud based E-healthcare services. *Peer-to-peer Networking and Applications*, *14*(5), 3043-3057.

[38] Hedabou, M. (2006). A frobenius map approach for an efficient and secure multiplication on Koblitz curves. *International Journal of Network Security*, *3*(3), 239-243.

[39] Sharma, A., Georgi, M., Tregubenko, M., Tselykh, A., & Tselykh, A. (2022). Enabling smart agriculture by implementing artificial intelligence and embedded sensing. *Computers & Industrial Engineering*, *165*, 107936.

[40] Boukhriss, H., Azougaghe, A., & Hedabou, M. (2014, November). New technique of localization a targeted virtual machine in a Cloud Platform. In *2014 5th Workshop on Codes, Cryptography and Communication Systems (WCCCS)* (pp. 124-127). IEEE.

[41] Zhang, H., & You, J. (2018). An empirical study of transport efficiency of urban rail transit based on data envelopment analysis and tobit model. *Journal of Tongji University (Natural Science)*, *46*(09), 1306-1311.

[42] Hu, H. (2017). Design and Research of English Self-study System Based on Computer Network. *Revista de la Facultad de Ingenieria*, *32*(4), 534-540.

[43] Ning, F. A. N. G. (2017). Design and research of multi axis motion control system based on plc. *Academic Journal of Manufacturing Engineering*, *15*(1).

[44] Lin, W. (2017). Research on the Urban Rail Transit Design between Subcenter and Center of City. *Journal of railway engineering society*, *12*, 73-76.

[45] Jin, H., Zhou, X., Sun, X., & Li, Z. (2021). Decay rate of rail with egg fastening system using tuned rail damper. *Applied Acoustics*, *172*, 107622.

[46] Wang, Y., Wang, P., Li, Z., Chen, Z., & He, Q. (2020). Forecasting Urban Rail Transit Vehicle Interior Noise and Its Applications in Railway Alignment Design. *Journal of Advanced Transportation*, *2020*.

[47] Mehmood, R., Katib, S. S. I., & Chlamtac, I. (2020). *Smart infrastructure and applications*. Springer International Publishing.

[48] Mo, Z., Zeng, M., & Guan, J. (2021, October). Analysis of the Dynamic Performance Between Pantograph and Rigid Conductor Rail Considering Construction Error of the Contact Wire Height. In *International Conference on Electrical and Information Technologies for Rail*

Transportation (pp. 502-514). Springer, Singapore.

[49] Wang, X., Guo, Y., Bai, C., Liu, S., Liu, S., & Han, J. (2020). The effects of weather on passenger flow of urban rail transit. *Civil Engineering Journal*, *6*(1), 11-20.

[50] Gill, H. S., Singh, T., Kaur, B., Gaba, G. S., Masud, M., & Baz, M. (2021). A metaheuristic approach to secure multimedia big data for IoT-based smart city applications. *Wireless Communications and Mobile Computing*, *2021*.

[51] Alghamdi, A., Al-Badi, A., Alroobaea, R., & Mayhew, P. (2013). A comparative study of synchronous and asynchronous remote usability testing methods. *International Review of Basic and Applied Sciences*, *1*(3), 61-97.

[52] Zhu, Z., Zhao, F., Yuan, D., Wang, H., Fei, X., & Author, J. K. (2018). Research and development of high-power and high-speed dc circuit breaker for urban rail transit. *High Voltage Engineering*, *44*(2), 417-423.

[53] Pan, D., Zhao, L., Luo, Q., Zhang, C., & Chen, Z. (2018). Study on the performance improvement of urban rail transit system. *Energy*, *161*, 1154-1171.

[54] Alhakami, W., ALharbi, A., Bourouis, S., Alroobaea, R., & Bouguila, N. (2019). Network anomaly intrusion detection using a nonparametric Bayesian approach and feature selection. *IEEE Access*, *7*, 52181-52190.

[55] Kaur, G., Singh, K., & Gill, H. S. (2021). Chaos-based joint speech encryption scheme using SHA-1. *Multimedia tools and applications*, *80*(7), 10927-10947.

[56] Krichen, M., Lahami, M., Cheikhrouhou, O., Alroobaea, R., & Maâlej, A. J. (2020). Security testing of internet of things for smart city applications: A formal approach. In *Smart infrastructure and applications* (pp. 629-653). Springer, Cham.

[57] Xiaofeng, Y., Hao, X., & Zheng, T. Q. (2019). Stray current and rail potential dynamic simulation system based on bidirectional variable resistance module. *Transactions of China Electrotechnical Society*, *34*(13), 2793-2805.

[58] Anh, A., Phuong, V., Van Lien, N., & Hai, N. (2018). Braking energy recuperation for electric traction drive in urban rail transit network based on control super-capacitor energy storage system. *Journal of Electrical Systems*, *14*(3).

[59] Lawes, D., Ran, L., & Xu, Z. (2014, September). Design of a solid-state DC circuit breaker for light rail transit power supply network. In *2014 IEEE Energy Conversion Congress and Exposition (ECCE)* (pp. 350-357). IEEE.

[60] Li, P. (2017). Discussion on the construction cost control measures for urban rail transit. *Journal of Railway Engineering Society*, *34*(8), 89-92.

JOŽEF STEFAN INSTITUTE

Jožef Stefan (1835-1893) was one of the most prominent physicists of the 19th century. Born to Slovene parents, he obtained his Ph.D. at Vienna University, where he was later Director of the Physics Institute, Vice-President of the Vienna Academy of Sciences and a member of several sci- entific institutions in Europe. Stefan explored many areas in hydrodynamics, optics, acoustics, electricity, magnetism and the kinetic theory of gases. Among other things, he originated the law that the total radiation from a black body is proportional to the 4th power of its absolute tem- perature, known as the Stefan–Boltzmann law.

The Jožef Stefan Institute (JSI) is the leading independent scientific research institution in Slovenia, covering a broad spectrum of fundamental and applied research in the fields of physics, chemistry and biochemistry, electronics and information science, nuclear science technology, en- ergy research and environmental science.

The Jožef Stefan Institute (JSI) is a research organisation for pure and applied research in the natural sciences and technology. Both are closely interconnected in research de- partments composed of different task teams. Emphasis in basic research is given to the development and education of young scientists, while applied research and development serve for the transfer of advanced knowledge, contributing to the development of the national economy and society in general.

At present the Institute, with a total of about 900 staff, has 700 researchers, about 250 of whom are postgraduates, around 500 of whom have doctorates (Ph.D.), and around 200 of whom have permanent professorships or temporary teaching assignments at the Universities.

In view of its activities and status, the JSI plays the role of a national institute, complementing the role of the uni- versities and bridging the gap between basic science and applications.

Research at the JSI includes the following major fields: physics; chemistry; electronics, informatics and computer sciences; biochemistry; ecology; reactor technology; ap- plied mathematics. Most of the activities are more or   less closely connected to information sciences, in particu- lar computer sciences, artificial intelligence, language and speech technologies, computer-aided design, computer architectures, biocybernetics and robotics, computer automa- tion and control, professional electronics, digital communi- cations and networks, and applied mathematics.

The Institute is located in Ljubljana, the capital of the in dependent state of **Slove**nia (or S♡nia). The capital today isconsidered a crossroad between East, West and Mediter-ranean Europe, offering excellent productive capabilities and solid business opportunities, with strong international connections. Ljubljana is connected to important centers such as Prague, Budapest, Vienna, Zagreb, Milan, Rome, Monaco, Nice, Bern and Munich, all within a radius of 600 km.

From the Jožef Stefan Institute, the Technology park "Ljubljana" has been proposed as part of the national strat- egy for technological development to foster synergies be- tween research and industry, to promote joint ventures be- tween university bodies, research institutes and innovative industry, to act as an incubator for high-tech initiatives and to accelerate the development cycle of innovative products.

Part of the Institute was reorganized into several high-tech units supported by and connected within the Technol-ogy park at the Jožef Stefan Institute, established as the beginning of a regional Technology park "Ljubljana". The project was developed at a particularly historical moment, characterized by the process of state reorganisation, privati-sation and private initiative. The national Technology Park is a shareholding company hosting an independent venture-capital institution.

The promoters and operational entities of the project are the Republic of Slovenia, Ministry of Higher Education, Science and Technology and the Jožef Stefan Institute. The framework of the operation also includes the University of Ljubljana, the National Institute of Chemistry, the Institute for Electronics and Vacuum Technology and the Institute for Materials and Construction Research among others. In addition, the project is supported by the Ministry of the Economy, the National Chamber of Economy and the City of Ljubljana.

Jožef Stefan Institute
Jamova 39, 1000 Ljubljana, Slovenia
Tel.:+386 1 4773 900, Fax.:+386 1 251 93 85
WWW: http://www.ijs.si
E-mail: matjaz.gams@ijs.si
Public relations: Polona Strnad

INFORMATICA

AN INTERNATIONAL JOURNAL OF COMPUTING AND INFORMATICS

INVITATION, COOPERATION

Submissions and Refereeing

Please register as an author and submit a manuscript at: http://www.informatica.si. At least two referees outside the author's country will examine it, and they are invited to make as many remarks as possible from typing errors to global philosoph- ical disagreements. The chosen editor will send the author the obtained reviews. If the paper is accepted, the editor will also send an email to the managing editor. The executive board will inform the author that the paper has been accepted, and the author will send the paper to the managing editor. The paper will be pub- lished within one year of receipt of email with the text in Infor- matica MS Word format or Informatica LATEX format and figures in .eps format. Style and examples of papers can be obtained from http://www.informatica.si. Opinions, news, calls for conferences, calls for papers, etc. should be sent directly to the managing edi- tor.

SUBSCRIPTION

Please, complete the order form and send it to Dr. Drago Torkar, Informatica, Institut Jožef Stefan, Jamova 39, 1000 Ljubljana, Slovenia. E-mail: drago.torkar@ijs.si

Since 1977, Informatica has been a major Slovenian scientific journal of computing and informatics, including telecommuni- cations, automation and other related areas. In its 16th year (more than twentyeight years ago) it became truly international, although it still remains connected to Central Europe. The ba- sic aim of Informatica is to impose intellectual values (science, engineering) in a distributed organisation.

Informatica is a journal primarily covering intelligent systems in the European computer science, informatics and cognitive com- munity; scientific and educational as well as technical, commer- cial and industrial. Its basic aim is to enhance communications between different European structures on the basis of equal rights and international refereeing. It publishes scientific papers ac- cepted by at least two referees outside the author's country. In ad- dition, it contains information about conferences, opinions, criti- cal examinations of existing publications and news. Finally, major practical achievements and innovations in the computer and infor- mation industry are presented through commercial publications as well as through independent evaluations.

Editing and refereeing are distributed. Each editor can conduct the refereeing process by appointing two new referees or referees from the Board of Referees or Editorial Board. Referees should not be from the author's country. If new referees are appointed, their names will appear in the Refereeing Board.

Informatica web edition is free of charge and accessible at http://www.informatica.si.

Informatica print edition is free of charge for major scientific, ed- ucational and governmental institutions. Others should subscribe.

Informatica

An International Journal of Computing and Informatics

Web edition of Informatica may be accessed at: http://www.informatica.si.

Informatica

An International Journal of Computing and Informatics