Retraction

Retracted: Abnormal Data Monitoring and Analysis Based on Data Mining and Neural Network

Journal of Sensors

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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

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

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

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

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

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

References

[1] Y. Chen, "Abnormal Data Monitoring and Analysis Based on Data Mining and Neural Network," *Journal of Sensors*, vol. 2022, Article ID 2635819, 7 pages, 2022.

Research Article

Abnormal Data Monitoring and Analysis Based on Data Mining and Neural Network

Yanyan Chen [ID]

Zhejiang Financial College, Hangzhou, Zhejiang 310018, China

Correspondence should be addressed to Yanyan Chen; 185101029@xzyz.edu.cn

In order to solve the problems of low efficiency, high consumption of human and time resources, and low degree of intelligence in the current financial abnormal data detection system in computerized accounting, this paper proposes a financial abnormal data monitoring and analysis algorithm based on data mining and neural network. The analysis algorithm uses the method of data mining to process the original financial data, remove invalid information, retain valuable information, and standardize the data to solve the problem of large labor and time consumption. Then, neural network-related algorithms are used to identify the anomalies of standardized data, so as to realize the intelligent early warning of financial abnormal data. Compared with the financial audit algorithm in traditional accounting computerization, this algorithm has the advantages of high efficiency, low energy consumption, and high intelligence. The test results show that the classification accuracy of the proposed algorithm for abnormal data can reach more than 90%. It is proved that the algorithm is effective and improves the efficiency at the same time. The classification error rate of the classifier designed in this paper is 22.5%, and the accuracy rate is 77.5%. Both estimated and actual values represent the number of times, and there is no physical unit. The experiment shows that the main reason for the error rate is the delay of the inspection results of financial abnormalities. Through the example analysis, it can be concluded that the proposed intelligent analysis method of financial abnormal data based on deep learning has good effectiveness and accuracy and has a certain practical value.

1. Introduction

The 129 financial data in the financial statements of Chinese listed companies and the financial ratios derived from these financial data can be called financial indicators [1]. If the financial indicators reflecting debt repayment, operation, profit, growth, and cash flow are abnormal, there is no doubt that the enterprise has problems in internal operation, external marketing, or finance. Through data mining of these abnormal financial indicators, find one or several abnormal transmission paths connecting them. If there are economically meaningful items on the abnormal transmission path, this abnormal transmission path is of great significance to enterprises [2, 3].

From a macro perspective, since the establishment of the Shanghai Stock Exchange and Shenzhen Stock Exchange in 1990, China's stock market has developed rapidly following the pace of national economic growth. It is very important

for securities regulators to formulate the supervision system for listed companies. At present, in the regulatory environment where China's regulatory mechanism is not perfect enough, the risk of fraud in the listed companies increases. Fraud is reflected in the financial statements, and a series of abnormal financial indicators are bound to appear. Microscopically speaking, for mature listed companies, due to the poor management of normal business activities, some financial indicators will be abnormal. For newly listed companies, due to the strict listing conditions (especially on the main board), the company may whitewash the financial statements before listing in order to achieve the purpose of listing [4]. Once the company is listed, some financial indicators are bound to decline or rise, resulting in abnormalities.

With the rapid development of the economy, the market is more and more prosperous, and the competition among enterprises is becoming increasingly fierce. The authenticity of financial data has attracted more and more attention.
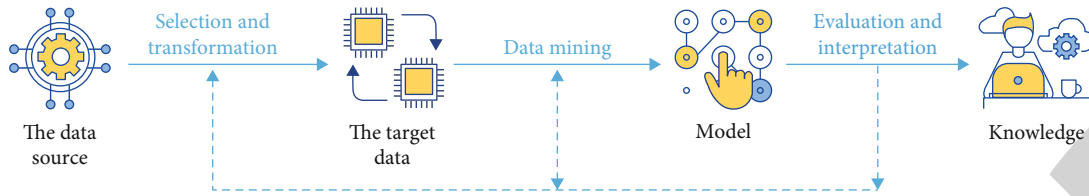
FIGURE 1: Workflow of data mining.

Academia has also taken a variety of measures and methods to analyze the abnormal situation of enterprise financial data and achieved corresponding results.

2. Literature Review

The research on the identification of financial anomaly characteristics of listed companies appeared earlier. Jiang and others believe that there is a certain relationship between company size and financial fraud, and the size of the company may directly affect the possibility of financial information fraud [5]. According to the research of Kim et al., there is a large correlation between financial information fraud and companies in financial distress. When the company is in financial difficulties, the management of the company is more likely to commit fraud in order to cover up its temporary financial difficulties [6]. Aydemir and others found that the company's abnormal financial information or even fraud has a great relationship with the company's industry. His experiments have proved that compared with other industries, the computer industry and manufacturing industry are more likely to have financial fraud, which needs special attention [7]. According to the research of Ali and others, according to the analysis of the financial report, we can find the signs of abnormal financial information, such as some unexplained changes in the financial report, some obviously different large transactions, and sudden increase of profits. In addition, frequent changes in management and faster growth of expenses than income may also be signs of financial fraud [8]. Asami and others conducted an empirical study on the relationship between insider trading and financial fraud of listed companies and established a cascaded logit regression model. The research points out that insider trading is a signal to reveal the potential possibility of financial fraud. Whether the company has financial fraud can be identified according to the analysis of insider trading variables and the specific financial characteristics of the company [9]. Moayedi and others proposed a new financial fraud identification model—fuzzy neural network (FNN) model. This model combines fuzzy logic, neural networks, and other methods to simulate the uncertainty in human rationality. The empirical results show that this model can effectively make up for the shortcomings of auditors or reduce their bias. Later research mainly uses artificial neural network (ANN) [10]. Xiang and others collected 164 fraudulent and nonfraudulent company data to verify the effectiveness of using machine learning technology to identify financial information fraud [11].

Most of these studies focus on the theoretical analysis of finance and accounting, only give the financial model analysis, but do not give the specific implementation scheme, resulting in poor realizability. Aiming at this problem, this paper focuses on the implementation scheme and puts forward the financial abnormal data monitoring and analysis algorithm based on data mining and neural network. For the problem of financial abnormal data, the specific identification model and implementation method are given, as shown in Figure 1.

Based on the current research, in order to solve the problems of low efficiency, high consumption of human and time resources, and low degree of intelligence in the current financial abnormal data detection system in computerized accounting, this paper proposes a financial abnormal data monitoring and analysis algorithm based on data mining and neural network. The analysis algorithm uses the method of data mining to process the original financial data, remove invalid information, retain valuable information, and standardize the data to solve the problem of large labor and time consumption.

3. Research Methods

3.1. System Framework. The overall frame structure of the system is shown in Figure 2. As can be seen from the figure, the system is mainly divided into four parts: identification object part, data processing part, identification process part, and result and early warning part. Among them, the identification object is the traditional financial system. This paper takes the data in the traditional financial system as the data set, takes it as the data acquisition channel, and carries out intelligent data monitoring [12]. Next is the data processing part, which processes the obtained original data through data mining technology. It is mainly divided into three modules: data acquisition, data preprocessing, and data coding. Among them, the data acquisition module is responsible for using the programming language to obtain the relevant original data required for abnormal data monitoring from the traditional financial database. The data preprocessing module mainly includes removing unique attributes, integrating related attributes, processing missing values, and data standardization so that the original data can become effective data. The data coding module is mainly responsible for coding the standardized data so that all data can be directly input into the intelligent algorithm to lay the foundation for the subsequent intelligent identification part. Next is the identification process part, which uses the processed financial data to train the intelligent algorithm through the artificial intelligence algorithm. The last part is the result
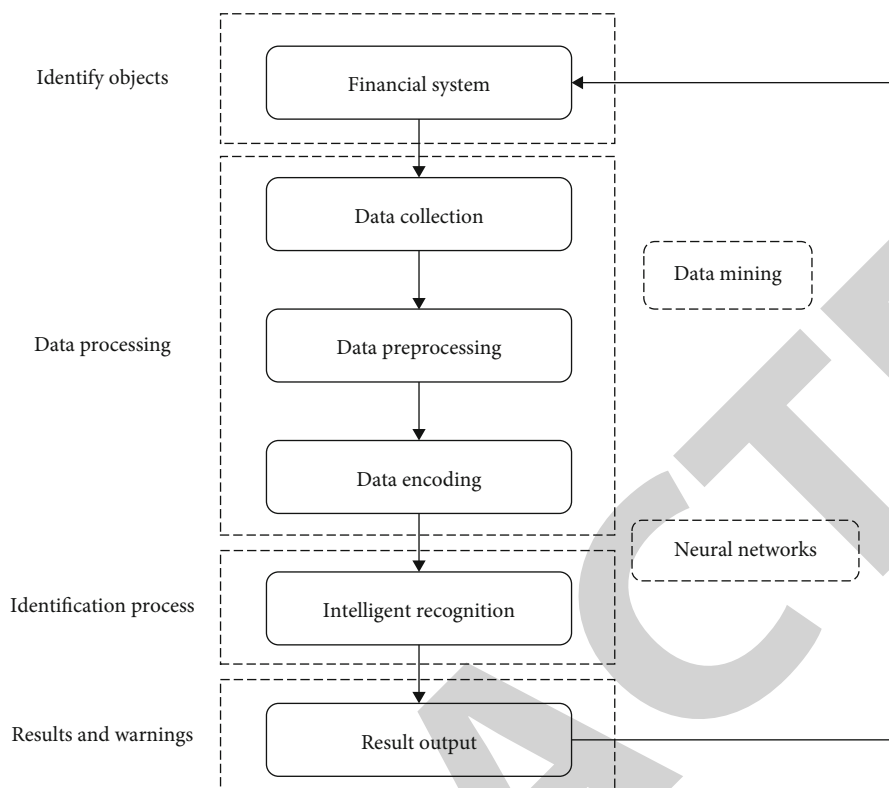
FIGURE 2: Overall framework of the system.

and early warning part, which uses the trained model to identify the financial data, mark and early warning the abnormal data, and feedback the early warning system to the financial system to form a closed loop.

Due to the infinite approximation ability of intelligent algorithm to nonlinear relationship, all kinds of abnormal identification methods of financial data are similar. This paper only takes the abnormal identification of employee travel invoice reimbursement as an example to illustrate the implementation process of the system.

3.2. Data Acquisition and Data Processing. In this system, the acquisition method of basic financial data is similar to that of traditional financial system data. It is mainly obtained through the statistics of financial department or personal report, and then, the original data is converted into standard data that can be directly input into intelligent control algorithm through corresponding data preprocessing methods [13]. The system is mainly divided into two parts: data acquisition and data processing.

3.2.1. Realization of Data Acquisition. The data required by this system usually consists of three types: personal information, including name, gender, age, department, length of service, professional title, position, basic salary, and other data. This kind of data can be obtained directly from the human resource department and the financial department through the internal network of the company. The second category is the bonus category, including the data of

quarterly bonus, year-end bonus, various honor bonuses, and subsidy bonuses of corresponding employees. This kind of data is usually distributed by their department according to their work performance, so it can be obtained directly through the department database. The third category is personal reimbursement data, mainly including travel location, travel duration, travel type, vehicle time, and various invoices and amounts. This kind of data is mainly reported by individuals and verified by department heads. With the development of network technology, the collection method of such data has also been greatly affected.

3.2.2. Realization of Data Processing. Data processing can be divided into two parts: data preprocessing and data coding. The implementation process of these two parts is described in the following two parts [14].

(1) Implementation of Data Preprocessing. The main process of data preprocessing is shown in Figure 3.

Figure 3 shows the problems to be solved in data preprocessing and the corresponding solutions. The following implementation scheme is given by taking Python language as an example:

(1) Unique attribute processing: the unique attribute refers to the name, job number, and other features that can uniquely identify the sample to be identified, but have no effect on the identification of abnormal data
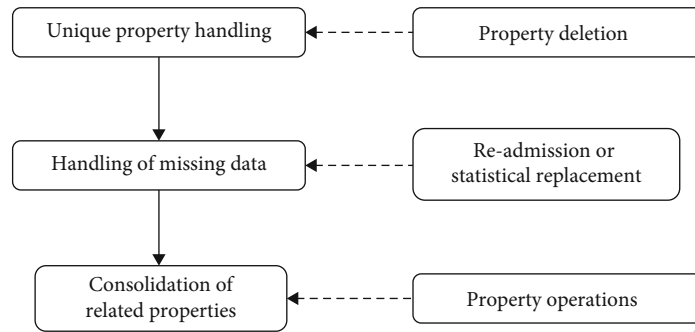
Figure 3: Data preprocessing.

and can be deleted directly. The instructions deleted in Python are as follows: use this command to delete the entire row or column data of the feature_name attribute, where axis = 1 means to delete the column and axis = 0 means to delete the row

(2) Processing of missing data: the processing steps of missing value are as follows: ① find the location of missing value; ② padding of missing values. Using Python language, the specific processing methods are as follows: the index information of the missing value of a characteristic data can be quickly obtained by using Pandas scientific calculation library. The function returns the location logical index of the missing value in the feature_name attribute. The missing value can be processed by using the location logical index

(3) Related attribute merging: for the characteristic data with obvious correlation, the operation of data can be used to merge the relevant data into one data and delete other data, so as to achieve the purpose of data merging and reducing the amount of calculation. The addition of two attribute columns feature 1 and feature 2 in data is realized, and the deletion operation after attribute merging is realized. For example, length of service and year of employment, such data can be consolidated

So far, the preprocessing of the original data has been completed, making the original data set into a data set without missing data and containing only effective features.

3.2.3. Realization of Data Coding. Data coding mainly realizes the transformation of feature data from semantic data convenient for human understanding to digital data convenient for machine processing. Data coding can be divided into discrete data and continuous data for processing, respectively. The specific processing methods are as follows:

(1) For discrete data: among the corresponding attributes, one or more discrete attributes can be classified into one class, and each class can be coded separately by One_Hot coding method. In Python language, Pandas scientific computing library can be used to realize this function conveniently

(2) For continuous data: for the coding of continuous data, in order to facilitate processing, the continuous data can be discretized in the form of segments and then encoded in sequence by One_Hot [15]. In the Python language, you can use custom operations to discretize continuous data. Next, take the binary classification as an example to illustrate the implementation method of continuous data discretization. It realizes the operation of separating and dispersing the feature data feature of continuous feature feature_name according to the threshold. Then, the discrete coding method can be used to encode the data

3.3. Realization of Identification Process and Result Output. The identification of abnormal data can be understood as the classification of data; that is, all data are divided into normal data and abnormal data or into normal data, doubtful data, and abnormal data. At present, the main classification algorithms are support vector machine (SVM). For example, SVM is used to build a prediction and classification system for the spread range of microblog rumors and the order volume of e-commerce. SVM is used to evaluate and classify the network system risk, personal credit, and daily stress state, and the application of gesture state classification and fault diagnosis is realized by the method of support vector machine. At present, SVM algorithm is widely used in classification applications, but it is still difficult to implement large-scale training samples.

Decision tree algorithm uses the decision tree algorithm to realize the classification of network loans, ECG signals, and financial crisis prediction models, but the decision tree algorithm has the problems of pruning and overfitting under large data samples [16]. K-means algorithm uses K-means to realize the classification of text and plants, but the algorithm has the problem that the K value has a great impact on the classification effect but cannot be obtained accurately [17].

Considering the application environment of financial abnormal data monitoring and analysis, the system adopts BP neural network with mature technology, ideal adaptability, and realization as the identification algorithm [18, 19]. The implementation method is as follows.
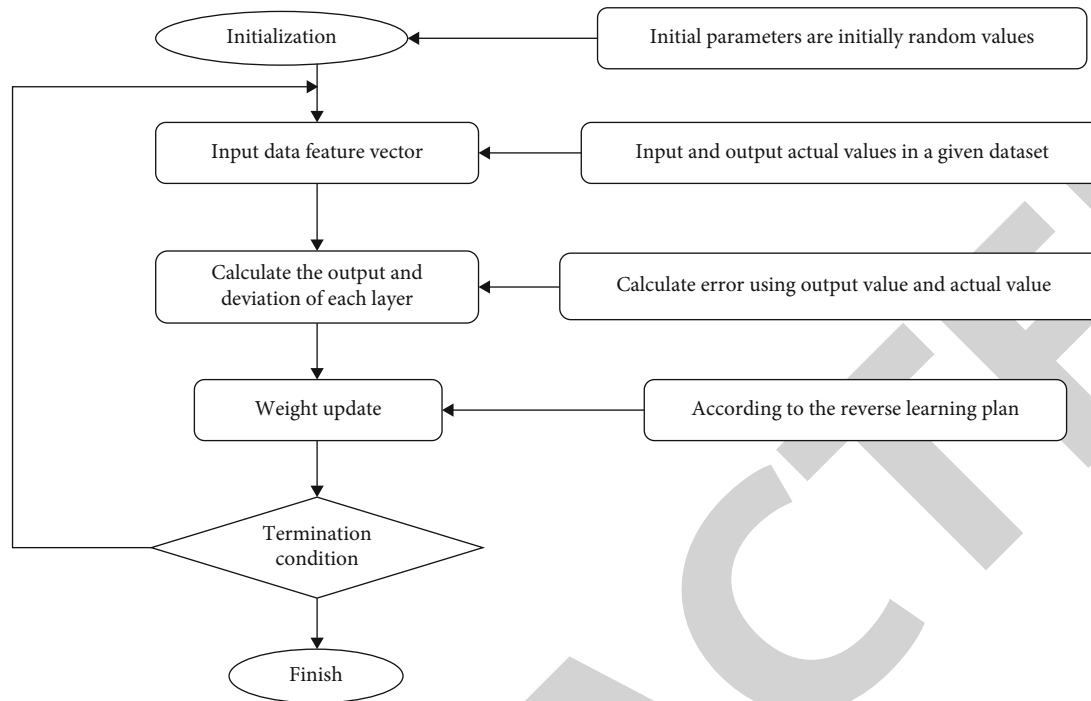
Figure 4: Training process of BP neural network.

(1) The number of input features and the number of coding bits of each feature determine the number of neurons in the input layer of the neural network

$$N_{\text{in}} = \sum_{1}^{N} n_i. \tag{1}$$

The calculation method is shown in formula (1), where N is the feature tree, n_i is the coding digit of the i-th feature, and N_{in} is the number of neurons in the input layer.

(2) Determination of intermediate layer: according to experience, the middle layer is usually 1 layer, and the number of neurons in the middle layer is $2 * N_{\text{in}}$, which can be adjusted according to the actual situation. The neuron correlation function is generally conventional.

(3) Output layer design: the number of neurons in the output layer is equal to the number of demand classification. One_Hot coding method is used to encode the demand state, and the output neuron function is selected as Logsig function.

The hidden layer in this paper is mainly used for feature learning and extraction through convolution and pooling. Different layers will extract different features. In order to ensure the optimal classification performance of the hidden layer, the training data and test data in the experimental data are divided according to the ratio of 4 : 1 and input into the convolutional neural network for solution optimization [20].

So far, the neural network has been designed and trained by using the processed data. The training method is shown in Figure 4 [21].

Finally, the monitoring and analysis of abnormal financial data can be realized by using the trained neural network. Give the identification results of normal and abnormal financial data, and feedback the abnormal financial data to the financial system.

4. Result Analysis

The BP neural network classifier in this paper is realized by MATLAB programming. In this section, the financial data of traditional indicators are processed by the timing construction method proposed above and then input into the classifier. After running, the classification effects of traditional models and various forms of models can be obtained. The operation results of each model are shown in Figure 5.

As can be seen from Figure 5, the recall rate of the time series index model in the ratio form and the first relative value form is the highest. In terms of precision, the time series index model in the form of difference is the best. It can also be seen from Figure 5 that the recall rate and precision rate show the law of one ebb and flow [22]. The comparison of classification accuracy under different hidden layer structures is shown in Table 1.

It can be seen from Table 1 that hidden levels 1, 2, 3, and 4 show good classification accuracy, all reaching more than 91%. After 400 iterations, the classification accuracy of the second level reaches 99.21%, which is the maximum of the classification accuracy. Therefore, it can be concluded that the convolutional neural network model with 4-layer hidden layer structure has good classification accuracy.

It can be seen from Table 2 that the classification and recognition error rate of the classifier designed in this paper is 22.5% and the accuracy rate is 77.5%. The estimated and
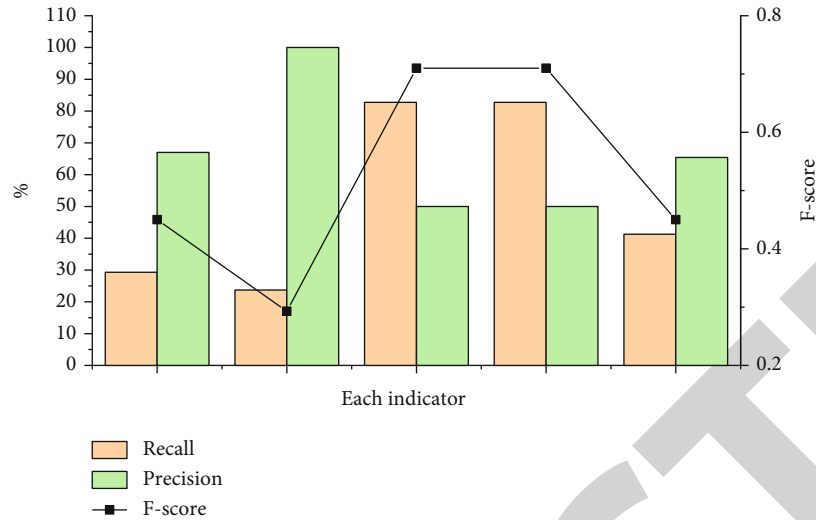
Figure 5: Comparison results of recognition effects of various forms of models.

Table 1: Comparison of classification accuracy under different hidden layer structures.

Number of iterations	Accuracy			
	Level 1	Level 2	Level 3	Level 4
100	0.9614	0.9637	0.9268	0.9177
200	0.9742	0.9763	0.9503	0.9315
300	0.9847	0.9873	0.9689	0.9655
400	0.9898	0.9921	0.9841	0.9678
500	0.9898	0.9922	0.9841	0.9678
600	0.9898	0.9921	0.9919	0.9738
700	0.9898	0.9892	0.9919	0.9794
800	0.9874	0.9897	0.9919	0.9851
900	0.9844	0.9897	0.9895	0.9875
1000	0.9819	0.9843	0.9895	0.9903

Table 2: Error rate of neural network classification.

	Abnormal	No abnormality
Estimated value	55	25
Actual value	64	16
Error rate	16.36%	36.00%
Total error rate		22.5%

actual values in Table 2 represent the number of times without physical units. The experiment shows that the main reason for the error rate is the delay of the inspection results of financial abnormalities [23]. Through the example analysis, it can be concluded that the proposed intelligent analysis method of financial abnormal data based on deep learning has good effectiveness and accuracy and has certain practical value [24].

5. Conclusion

This paper presents a detailed analysis of the current financial anomaly monitoring algorithm based on the combination of neural network and artificial intelligence and solves the problem of abnormal data mining. Taking the abnormal recognition of business trip invoice reimbursement as an example, this paper expounds the implementation process of data analysis and early warning and verifies the effectiveness of the system. The BP neural network classifier in this paper is realized by MATLAB programming. In this section, the financial data of traditional indicators are processed by the timing construction method proposed above and then input into the classifier. After running, the classification effects of traditional models and various forms of models can be obtained. At the same time, there are still some deficiencies in the system, which are mainly reflected in the long training time of neural network and the lack of guiding opinions on the selection of initial value of network. The optimization of network initial value and the adaptation of network parameters in the system will be the further research direction.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

[1] W. Silva, F. Nascimento, and M. Bourguignon, "A change-point model for the *r*-largest order statistics with applications to environmental and financial data," *Applied Mathematical Modelling*, vol. 82, pp. 666–679, 2020.

[2] F. De Arriba-Pérez, S. García-Méndez, J. Á. Regueiro-Janeiro, and F. J. González-Castaño, "Detection of financial opportunities in micro-blogging data with a stacked classification system," *IEEE Access*, vol. 8, pp. 215679–215690, 2020.

[3] R. Yang, L. Yu, Y. Zhao et al., "Big data analytics for financial market volatility forecast based on support vector machine," *International Journal of Information Management*, vol. 50, pp. 452–462, 2020.

[4] S. L. Li, "Data mining of corporate financial fraud based on neural network model," *Computer Optics*, vol. 44, no. 4, pp. 665–670, 2020.

[5] Q. Jiang, X. Zhang, Q. Lin, G. Chen, and S. Liu, "Internet penetration and regional financial development in China: empirical evidence based on Chinese provincial panel data," *Scientific Programming*, vol. 2020, no. 8, Article ID 8856944, 14 pages, 2020.

[6] H. Kim, J. Park, K. Min, and K. Huh, "Anomaly monitoring framework in lane detection with a generative adversarial network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, pp. 1–13, 2020.

[7] G. Aydemir and B. Acar, "Anomaly monitoring improves remaining useful life estimation of industrial machinery," *Journal of Manufacturing Systems*, vol. 56, pp. 463–469, 2020.

[8] B. Ali, R. A. Ahmed, J. A. Yaba, N. M. Hamawandy, and A. Ayoob, "The effects of computerized accounting system on auditing process: a case study from northern Iraq," *Solid State Technology*, vol. 63, no. 5, pp. 8564–8578, 2021.

[9] H. Asami, M. Golabi, and M. Albaji, "Simulation of the biochemical and chemical oxygen demand and total suspended solids in wastewater treatment plants: data-mining approach," *Journal of Cleaner Production*, vol. 296, no. 2-4, article 126533, 2021.

[10] H. Moayedi, M. M. Abdullahi, H. Nguyen, and A. Rashid, "Comparison of dragonfly algorithm and Harris hawks optimization evolutionary data mining techniques for the assessment of bearing capacity of footings over two-layer foundation soils," *Engineering with Computers*, vol. 37, no. 1, pp. 437–447, 2021.

[11] Y. Xiang and G. Yamamoto, "A data mining approach to investigate the carbon nanotubes mechanical properties via high-throughput molecular simulation," *Materials Science Forum*, vol. 1023, pp. 29–36, 2021.

[12] E. P. Booker and G. E. Jabbour, "Antiviral nanoparticle ligands identified with datamining and high-throughput virtual screening," *RSC Advances*, vol. 11, no. 37, pp. 23136–23143, 2021.

[13] Y. Li, R. K. Shyamasundar, and X. Wang, "Special issue on computational intelligence for social media data mining and knowledge discovery," *Computational Intelligence*, vol. 37, no. 2, pp. 658-659, 2021.

[14] Y. Liu, Z. Yu, and Y. Yang, "Diabetes risk data mining method based on electronic medical record analysis," *Journal of Healthcare Engineering*, vol. 2021, no. 6, Article ID 6678526, 11 pages, 2021.

[15] J. Fan, M. Zhang, A. Sharma, and A. Kukkar, "Data mining applications in university information management system development," *Journal of Intelligent Systems*, vol. 31, no. 1, pp. 207–220, 2022.

[16] M. Pargari, F. Marahel, and B. M. Goodajdar, "Kinetic spectrophotometric method and neural network model application for the quantitation of epinephrine by starch-capped agnps sensor in blood and urine," *Journal of Analytical Chemistry*, vol. 77, no. 4, pp. 484–494, 2022.

[17] B. S. N. Reddy, S. K. Pramada, and T. Roshni, "Selection of level and type of decomposition in predicting suspended sediment load using wavelet neural network," *Acta Geophysica*, vol. 70, no. 2, pp. 847–857, 2022.

[18] N. Zhao, H. Gao, X. Wen, and H. Li, "Combination of convolutional neural network and gated recurrent unit for aspect-based sentiment analysis," *IEEE Access*, vol. 9, pp. 15561–15569, 2021.

[19] J. Shen, J. K. Liu, and Y. Wang, "Dynamic spatiotemporal pattern recognition with recurrent spiking neural network," *Neural Computation*, vol. 33, no. 11, pp. 1–25, 2021.

[20] Y. Chen, W. Zhang, L. Dong, K. Cengiz, and A. Sharma, "Study on vibration and noise influence for optimization of garden mower," *Nonlinear Engineering*, vol. 10, no. 1, pp. 428–435, 2021.

[21] D. Kumar, A. Sharma, R. Kumar, and N. Sharma, "Restoration of the network for next generation (5G) optical communication network," in *2019 International Conference on Signal Processing and Communication (ICSC)*, NOIDA, India, 2019.

[22] J. Jayakumar, S. Chacko, and P. Ajay, "Conceptual implementation of artificial intelligent based E-mobility controller in smart city environment," *Wireless Communications and Mobile Computing.*, vol. 2021, pp. 1–8, 2021.

[23] L. Xin, L. Jianqi, C. Jiayao, and Z. Fangchuan, "Degradation of benzene, toluene, and xylene with high gaseous hourly space velocity by double dielectric barrier discharge combined with Mn_3O_4/activated carbon fibers," *Journal of Physics D: Applied Physics*, vol. 55, no. 12, article 125206, 2022.

[24] R. Huang, S. Zhang, W. Zhang, and X. Yang, "Progress of zinc oxide-based nanocomposites in the textile industry," *IET Collaborative Intelligent Manufacturing*, vol. 3, no. 3, pp. 281–289, 2021.