International Journal of Data Science

ISSN online: 2053-082X - ISSN print: 2053-0811
https://www.inderscience.com/ijds

The emergency network public opinion risk identification and early warning model from BP neural network

Shuling Chen, Shuai Yuan

DOI: 10.1504/IJDS.2025.10072258

Article History:

Received:	03 April 2025
Last revised:	15 May 2025
Accepted:	26 May 2025
Published online:	16 January 2026

The emergency network public opinion risk identification and early warning model from BP neural network

Shuling Chen*

School of Social and Public Administration,
Lingnan Normal University,
Zhanjiang, 524000, China
Email: 78579123@qq.com
*Corresponding author

Shuai Yuan

Zhanjiang Customs Anti_Smuggling Bureau,
Zhanjiang, 524000, China
Email: 38323351@qq.com

Abstract: This paper aims to solve the problem of emergency network public opinion (NPO) risk identification (RI) and early warning (EW). Firstly, the back propagation neural network (BPNN) optimised by Genetic Algorithm (GA) is used to process and model the data obtained on the network, identify the public opinion risk of emergencies, and realise the risk prediction and early warning. Secondly, through the analysis and mining of NPO data of emergencies, the factors affecting the risk of NPO, such as social media platforms, user characteristics, and text content, are explored. These factors are incorporated into the model to improve the predictive ability of the model. Finally, through the research, effective Risk Management (RM) and countermeasures of NPO in emergencies are proposed to provide feasible RM schemes for governments, enterprises, and the public to ensure social stability and security.

Keywords: BPNN; back propagation neural network; data mining; emergencies; NPO; network public opinion; risk identification and early warning.

Reference to this paper should be made as follows: Chen, S. and Yuan, S. (2025) 'The emergency network public opinion risk identification and early warning model from BP neural network', *Int. J. Data Science*, Vol. 10, No. 7, pp.93–113.

Biographical notes: Shuling Chen graduated with a Doctorate from Northeast Agricultural University. Currently, she is teaching at the School of Social and Public Administration of Lingnan Normal University. Her research interests lie in agricultural economic management and public administration.

Shuai Yuan graduated from Hebei University of Economics and Business with a Master's degree. Currently, he works at the Anti-Smuggling Bureau of Zhanjiang Customs. His research interests lie in the fields of law and social management.

1 Introduction

1.1 Research background and motivations

In today's society, emergencies often attract widespread public attention. The popularity of the internet and social media has made the dissemination of emergency information faster and wider in rapidly changing circumstances. This information is often complex and uncertain, and the proliferation and influence of network public opinion (NPO) can also lead to unpredictable consequences. Taking the early outbreak of COVID-19 in 2020 as an example, relevant information spread rapidly on social platforms such as Weibo and WeChat, which not only accelerated the popularisation of epidemic prevention and control knowledge, but also triggered a lot of rumours and panic, which posed a serious challenge to social order and public safety. Similarly, in recent years, natural disasters, public health incidents, major accidents and other emergencies have occurred frequently, and the complexity and uncertainty of online public opinion are increasing. The spread of a large number of false information and emotional remarks not only aggravated public panic, but also had a negative impact on government decision-making and social stability. Therefore, timely and accurate identification and prediction of NPO risks in emergencies has important practical significance and application value.

At present, machine learning (ML) and data mining (DM) techniques are mainly used to identify and predict the risk of NPO in emergencies. However, existing research often has the following problems. The model accuracy is not high, the real-time performance is not strong, and the adaptability is insufficient (Lyu et al., 2022). Meanwhile, back propagation neural network (BPNN), as a commonly used artificial neural network model, has strong nonlinear modelling ability and adaptability. It has a wide range of applications in public opinion analysis. BPNNs can cope with complex data analysis and pattern recognition and have potential applications in text classification, sentiment analysis, and early warning (EW) of emergencies (Sivamani et al., 2022). Therefore, the use of BPNN to construct an emergency NPO Risk Identification (RI) and EW model is expected to improve the accuracy and real-time prediction and provide effective risk management and response measures for governments, enterprises, and the public.

1.2 Research objectives

The research objectives are as follows.

Firstly, an NPO RI model for emergencies based on BPNN is designed. The model can obtain data from the network, process data, establish models, and accurately identify the NPO risk of emergencies to achieve risk prediction and EW. To improve the model's prediction accuracy and real-time performance, genetic algorithm (GA) was used to optimise the weights and thresholds of the BPNN. Unlike traditional BPNN models, GA can avoid local minima during training, ensuring global optimisation of network weights and thresholds. This enhances the network's convergence and prediction ability on diverse data. Additionally, this study incorporated multiple factors from emergency event data and used improved data analysis methods. This gives the model stronger real-time response capabilities, enabling it to quickly identify potential risks and take timely measures. Finally, the study deeply explored various factors affecting NPOs in emergencies and included them in the model analysis, further strengthening the model's predictive power. Compared to existing models, the innovation of this study lies in using GA to optimise the BPNN, overcoming the limitations of traditional methods in handling complex data and real-time warnings. Specifically, the model can efficiently and accurately identify and warn of risks in multi-source, dynamically changing data environments, providing stronger decision support for NPOs dealing with emergencies.

2 Literature review

In recent years, BPNN-based NPO public opinion monitoring methods have been widely used in public opinion analysis and emergency early warning. Li et al. proposed a BPNN-based NPO public opinion monitoring method, collecting and preprocessing public opinion data from platforms such as Weibo, news, and forums. Through BPNN analysis and modelling, features related to public opinion were extracted, and the monitoring and analysis of NPOs were realised (Li et al., 2023). This method improved the comprehensiveness and accuracy of monitoring through the integration of multi-platform data, but still had certain shortcomings in terms of data timeliness and multidimensional feature extraction. Xin et al. (2022) suggested that research on the identification and early warning of online public opinion risks in emergencies should start from the collection, processing, and application of public opinion data. Using data mining and natural language processing technologies, they constructed a BPNN-based NPO public opinion risk identification and early warning model. The model effectively identified the sources of public opinion risks through multidimensional analysis of public opinion data, but its processing efficiency for large-scale public opinion data needed to be improved. Carnegie and Gaikwad (2022) reviewed the theoretical basis, algorithm principles, and application characteristics of BPNN. They believed that BPNN had the advantages of a simple model structure, fast training speed, and high prediction accuracy, and pointed out that BPNN had broad application prospects in NPO analysis. However, existing BPNN applications still lack in-depth exploration of algorithm optimisation. Especially when facing complex public opinion data, there is still room for further optimisation. Sunhare et al. (2022) found that the BPNN-based emergency NPO public opinion risk identification and early warning model had high accuracy and timeliness, and had been verified in practical applications. However, it still faced problems of data uncertainty and insufficient samples, which were major challenges for improving the model's universality and reliability.

As research progresses, an increasing number of studies have proposed more diversified models and methods. Li et al. (2024) introduced a dual-layer network model that integrated social media public opinion and geographic information. This model could better identify public opinion risks in crises, emphasising the importance of geographic information in enhancing situational awareness and providing new perspectives for emergency response. This approach offers a new perspective for emergency response, particularly in the innovative application of geographic information. Bao et al. (2025) established an early warning index system for online public opinion, combining grey relational analysis and K-Means clustering algorithm for event grading. They developed a BPNN-based early warning model. Experiments had shown that BPNN had significant application potential in predicting online public opinion events. However, this method still needs further optimisation in dealing with the diversity and complexity of public opinion data. Cao et al. (2025) proposed a public opinion crisis prediction model that combined the Grey Wolf Optimiser (GWO) algorithm and Long Short-Term Memory (LSTM) network. The model is mainly used to analyse hot topics on Sina Weibo. The prediction accuracy of the model has been verified, showing its effectiveness in trend prediction and intervention. However, the model's scope of application is relatively limited, and it still lacks widespread applicability and universality.

In recent years, a large amount of research has combined intelligent optimisation algorithms with neural networks in model optimisation, achieving good results. Bai et al. (2024) developed a hybrid project portfolio risk prediction model that combined GAs and backpropagation neural networks, and introduced entropy numbers-trapezoidal fuzzy numbers to improve prediction accuracy. Chen et al. (2025) established a prediction model using GAs and backpropagation neural networks to predict friction behaviour in bolted connections. The GA-BP prediction model, validated through experiments, provided valuable insights for the design and optimisation of bolted connections. Wang (2025) optimised BPNN using GAs to address local optimisation and overfitting issues, developing an intelligent field irrigation early warning system based on an enhanced GA-BPNN model to improve the accuracy of agricultural irrigation flow prediction. In addition, Liu et al. (2024) integrated particle swarm optimisation algorithm with gated loop units to propose a dynamic asset volatility prediction model for financial markets. They emphasised the advantages of the model in processing time-series data and improving generalisation ability, demonstrating the strong potential of optimisation algorithms in dynamic systems. Dong and Asif (2024) integrated Siamese networks, temporal convolutional networks, and random forest techniques to form a composite neural network model. They established a classification and financial prediction model for green technology investment in enterprises, which showed high classification accuracy in enterprise decision support. Chao et al. (2024) combined visual transformers, graph neural networks, and generative adversarial networks for the analysis and prediction of advertising marketing effectiveness, significantly improving the effectiveness of advertising click through rate prediction. The above optimisation methods have shown excellent performance in fields such as engineering, finance, enterprise management, and digital marketing. However, most of these studies focus on scenarios with specific structures or relatively stable data characteristics, and lack deep adaptation and specialised optimisation for high-dimensional, dynamic, and variable NPO data processing.

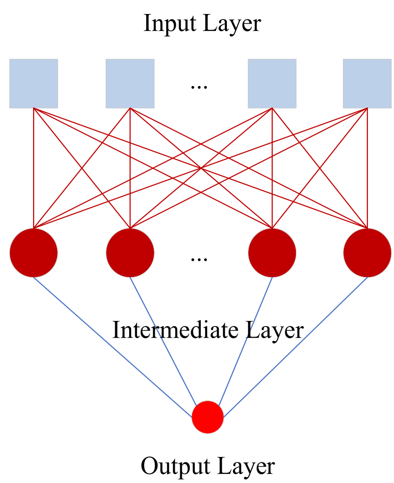
Although existing research has made remarkable progress in public opinion analysis and emergency early warning, there are still some limitations. Existing models often have deficiencies in aspects such as incomplete data, sample bias, and insufficient feature extraction. These deficiencies affect the universality and prediction accuracy of the models. Moreover, many studies are still not in-depth enough in exploring the optimisation algorithms of the BPNN, especially in the application of public opinion data processing and feature extraction. Therefore, in this study, by introducing the GA to optimise the BPNN model, targeted improvements have been made in data processing and feature extraction. This has enhanced the prediction accuracy and real-time response ability of the model, which has great innovation and application value.

3 Research model

3.1 BPNN

The BPNN is an artificial neural network commonly used in pattern recognition and prediction. Its basic structure consists of an input layer, a hidden layer, and an output layer (Chen and Chen, 2022; Liu et al., 2022; Qie et al., 2022). Each neuron is connected to the neurons in the previous layer and the next layer through connection weights. The core idea of BPNN is to continuously adjust the connection weights between neurons to reduce the error between the network output and the actual value to achieve the goal of training and optimising the model. Figure 1 shows the structure diagram of a simple BPNN.

Figure 1 Simple BPNN structure diagram (see online version for colours)



In the BPNN model, the input neuron and output neuron are the same, and the operation rules of the neuron in the middle hidden layer and the output layer are shown in equation (1).

$$Y_{kj} = f\left(\sum_{i=1}^{n} W_{(k-1)i,kj} Y_{(k-1)i}\right) \tag{1}$$

In equation (1), $Y_{(k-1)i}$ is the output of the *i*th neuron in the *k*–1 layer and the input of the *k*th layer neuron. $W_{(k-1)i,kj}$ is the link weight between the *i*th neuron of the *k*–1 layer and the *j*th neuron of the *k*th layer. Y_{kj} is the output of the *j*th neuron of the *k*th layer and the output of the neuron of the *k*+1 layer. *f* is the Sigmoid function and can be expressed as equation (2). *n* is the number of neurons in the *k*–1 layer.

$$F(x) = \frac{1}{1+e^{-x}} \tag{2}$$

The basic processing units of BPNNs have a nonlinear input-output relationship except for the input layer, allowing continuous adjustment of their input and output values (Dong et al., 2023). The essence of BPNNs is to learn and train samples according to the provided sample data, form a nonlinear mapping relationship between input and output, and extract its implicit feature relationship to form a network with classification ability (Zhao et al., 2022). Specifically, the basic idea of the BP algorithm is to propagate the error of each iteration backwards from the output layer, through the hidden layer to the input layer. The gap between the output value and the expected value is narrowed by adjusting the connection weights between the individual neurons. This process is repeated until the error reaches the set target and accuracy requirements. The BPNN algorithm effectively solves the nonlinear problem by adding the hidden layer in the middle and the corresponding learning rules (Chen and Cheng, 2022). Regarding the prediction of the online public opinion risk of a certain emergency, in the input layer of the BPNN, relevant information such as the discussion volume on social media about the event and the emotional tendency is inputted. After being processed by the hidden layer, the output layer gives the predicted public opinion risk value. Through backpropagation, the BPNN continuously adjusts each weight to optimise the prediction accuracy of the model. The advantage of the BPNN lies in its powerful nonlinear modelling ability and adaptability. Traditional linear models often cannot handle complex patterns and relationships. However, by introducing the hidden layer, the BPNN can establish a nonlinear relationship between the input and the output, enabling it to deal with more complex real-world problems, such as public opinion analysis and emergency prediction. Due to its excellent classification and prediction capabilities, the BPNN is widely applied in many fields, including speech recognition, image recognition, financial prediction, and so on.

3.2 Internet public opinion

NPO refers to the public's opinions, emotions, and attitudes towards an event or topic formed on the internet. It reflects the ideas, values, and ideologies of society (Chen et al., 2022; Lin et al., 2022). The characteristics of NPO are as follows: extensive, fast, diverse, and interactive (Liu et al., 2022; Zhang et al., 2022). The dissemination mechanism of online public opinion is complex and is usually influenced by multiple factors. These factors include the credibility of information sources, information dissemination channels, the structure of social networks, the bias of media reports, and changes in public sentiment, among others. The rapid spread of information on social media platforms is often accompanied by the polarisation and quick diffusion of emotions. This has led to the important roles played by 'opinion leaders' and 'information disseminators' in the formation of public opinion. The content of public opinion is continuously amplified through different dissemination channels. This may trigger fluctuations in the emotions of groups, and in turn, affect social stability and public decision-making.

The analysis of NPO is one of the issues that has attracted much attention from all walks of life. Its main task is to analyse the emotions of public opinion texts. This process requires mathematical calculations on public opinion data and the use of corresponding numerical values to judge the authenticity of public opinion. The dissemination of online public opinion is driven not only by the media and public sentiment but also influenced by deep-seated factors such as individual cognition, social structure, and cultural background. For instance, the algorithm recommendation mechanism of social platforms may push specific content based on users' historical preferences, thus enhancing the spread of certain emotions or viewpoints. The influencing factors of online public opinion also include the sensitivity of social events, the public's attention to these events, and the degree of intervention by the government and the media. These factors jointly determine the development direction of public opinion and its social influence.

In response to these influencing factors, there are various methods for public opinion analysis. Common network sentiment analysis methods include: text mining methods, ML methods, social network analysis methods, data visualisation methods, deep learning (DL) methods, and NLP methods (Zhang et al., 2022).

Among them, the DL methods can process massive amounts of data without excessive human intervention. The accuracy is high, and it has good advantages over other methods (Čábelková et al., 2022). Therefore, this paper will use DL algorithms and theories to construct an emergency NPO RI and EW model.

According to the Law of the People's Republic of China on Emergency Responses, emergencies can be divided into four levels according to the degree of urgency, scope of impact, development trend, and degree of damage caused by them (Liu, 2022; Wang et al., 2023). Through the model exercise of multiple case data in the early stage, it is found that four classification categories always appear in the simulation results. To consider the feasibility of actual operation, the public opinion of accident disaster is divided into four levels according to popularity, of which the first level is the highest. Based on the training results of training data and the experience of previous scholars, the criteria for each level are determined, as shown in Table 1.

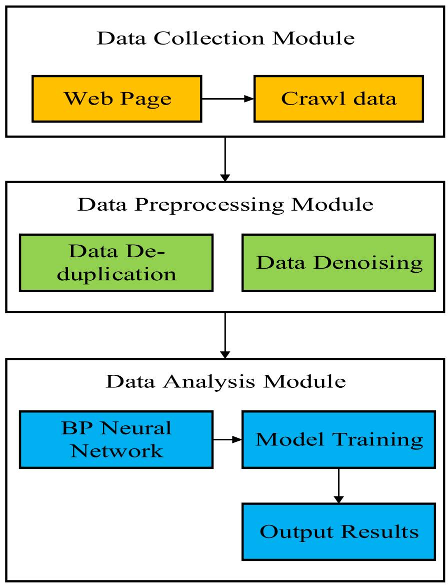
Table 1 Public opinion evaluation indicators for emergencies

Index	*Level 1*	*Level 2*	*Level 3*	*Level 4*
Number of deaths	>10	>5	>2	>1
Number of people injured	>10	>5	>2	>1
The inverse of economic loss	>4	>3	>2	>1
Duration of public opinion	>12d	>7d	>5d	>3d
Percentage of negative information	>60%	>55%	>50%	>45%
Number of comments	>1,000,000	>500,000	>100,000	>10,000
Number of Retweets	>1,000,000	>500,000	>100,000	>10,000
Number of Likes	>500,000	>200,000	>100,000	>50,000
Number of participating media	>1,500	>800	>200	>100
Total number of media releases	>500,000	>200,000	>50,000	>10,000
Peak data volume of public opinion	>200,000	>50,000	>20,000	>10,000
Timeliness	4	3	2	1
Degree of truthfulness	4	3	2	1
Accident handling speed	4	3	2	1
Satisfaction with accident results	4	3	2	1

3.3 Model building

The NPO RI and EW model constructed here can be divided into three modules: data collection module, preprocessing module, and analysis module. The data collection module uses a data crawler to collect data from specified webpages and store it in a database. The collected data also needs to be deduplicated and denoised in the preprocessing module before it can be used by the analysis module. Finally, the data analysis module uses the normalised data to train the BPNN model to determine the authenticity output of public opinion in emergencies and provide EW, as shown in Figure 2.

Figure 2 Emergency NPO RI and EW model (see online version for colours)



1 *Data collection module*

The data collection scripts used are based on the Scrapy framework. Data collection scripts written in Python can quickly traverse website data based on user needs (Foos and Bischof, 2022; Lyu et al., 2022). Unlike traditional data collection scripts, Scrapy data collection scripts can read a website's data interface, thereby increasing the speed of data collection (Liu et al., 2022; Velte, 2022). The data collection script structure of the Scrapy framework includes the script body, engine, scheduling plugin, download module, middleware, and pipeline (Kiran et al., 2022; Wang and Hu, 2022). The footer ontology is used for the management of uniform resource locators (URLs), ensuring that the content of specified websites can be accurately crawled. Once the data collection script obtains the content of the target URL address, it will input the content into the pipeline for storage (Azizi et al., 2022). The engine is responsible for transferring data content among various modules and coordinating the work of each link. The scheduling plugin is used to schedule the resource requests required by the engine, ensuring that the crawling tasks can be efficiently allocated and executed. The download module is controlled by the script. When it is necessary to download the content of a webpage, the download module will obtain the data of the target webpage by calling the downloader (Pascual-Ferrá et al., 2022; Qin et al., 2022). The middleware is used to handle the conversion of requests and responses, while the pipeline is used to clean the crawled data and store it in the database. To ensure the accuracy and integrity of the data, multiple inspection mechanisms have been added to the crawler script. Regular URL validity checks have been set up to ensure that the crawled webpage addresses are valid and the content is up-to-date. For the crawled data, a deduplication algorithm is adopted to prevent the repeated crawling of the same information. An exception capture and logging mechanism has also been introduced. During the data crawling process, if an error or interruption occurs, the system will automatically record it and retry, minimising the problems of data loss and incomplete crawling.

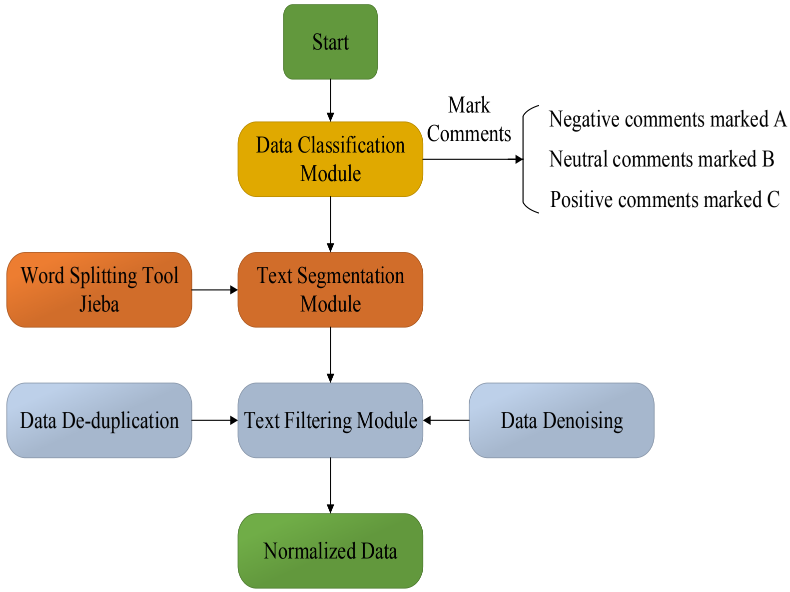
2 *Data preprocessing module*

The data preprocessing module includes three parts: a data classification module, a text word segmentation module, and a text filtering module. The data classification module is used to label the collected data, labelling negative comments as A, neutral comments as B, and positive comments as C (Batool et al., 2023; Yang et al., 2023). This categorical data will be used as a validation dataset. In the text word segmentation module, the third-party word segmentation tool Jieba based on the Python language is used. Given that there are no obvious word boundaries in Chinese texts, Jieba can effectively perform word segmentation on Chinese texts. It conducts segmentation through dictionary matching and word frequency-based statistical methods, and can accurately divide long texts into meaningful words. The text filtering module includes duplicate data deletion and noise removal operations. When removing duplicate data, a content-based deduplication algorithm is used. By performing hash matching on the text content, duplicate comment data is deleted. During the noise removal process, strategies such as removing stop words, low-frequency words, and non-Chinese characters are adopted to ensure that the remaining data is more representative and valid. To better handle noisy data and outliers, this text filtering module also uses the Z-Score-based method to detect outliers in the data, automatically eliminating those samples that obviously deviate from the normal data. Figure 3 presents the functions and connections of each module of the data preprocessing process.

3 *Data analysis module*

This module uses BPNNs for data analysis, as standard BPNNs have multiple minimums during training. It causes the training results to fall into local minimisation, which affects the convergence effect of the network (Santos-Pereira et al., 2022; Zhang et al., 2022). To solve this problem, the genetic algorithm (GA) is used to optimise the network parameters so that the optimised network can avoid limitations and perform global optimisation processing on the weights to obtain good predicted output values. The basic principle is to represent the initial weight and threshold with the individual in the GA, take the initialised individual error as the fitness value, and reduce the error through a series of genetic operator operations to find the optimal individual with strong fitness, which is the optimal initial weight or threshold (Franceschini et al., 2022; Muñoz-Rodríguez et al., 2023). The steps to optimise weights and thresholds are as follows.

Figure 3 Data preprocessing diagram (see online version for colours)



1 *Population initialisation*

The real-string encoding method is adopted to represent individuals. The real-string method can represent weights and thresholds more precisely and has a wider search range. The chromosome of each individual contains all the weights and thresholds of the network. It can be understood that the structure, weights, and thresholds of the network are all determined. According to experience, this encoding method has higher precision and broader search ability compared with other encoding methods. The encoding length of the chromosome is determined by the number of neurons in the input layer and the output layer, and the calculation equation is as follows:

$$S = rS_1 + rS_2 + S_1 + S_2 \tag{3}$$

In equation (3), r is a constant, S_1 is the number of input layers, and S_2 is the number of output layers.

2 *Fitness function*

According to the initial weight and threshold of the network obtained by chromosomes, the output of the BPNN model is calculated using Matrix Laboratory (MATLAB). The individual fitness calculation equation used is:

$$F = k\left(\sum_{i=1}^{n} abs\left(y_i - o_i\right)\right) \tag{4}$$

In equation (4), n is the number of output layers, k is the coefficient, y_i is the expected output, and o_i is the actual output.

3 *Selection, crossover, and mutation operations*

In the GA, the roulette wheel selection strategy is used to choose individuals with stronger fitness from the current population as parents. The parents generate a new generation of individuals through the crossover operation. During the crossover process, the single-point crossover method is adopted, that is, a crossover point is randomly selected to exchange the gene sequences of the parent individuals. To increase the diversity of the population, a random mutation operation is also introduced. By randomly adjusting the genes of individuals, it can avoid getting trapped in local optimal solutions. The roulette wheel selection method is used as the selection strategy, where individuals are selected based on their fitness ratios. The single-point crossover is used for the crossover operation, where the genes of the parent individuals are exchanged at a random position. The mutation operation is a random mutation, where certain positions of an individual's genes are randomly modified according to a certain probability.

4 Experimental design and performance evaluation

4.1 Datasets collection

The evaluation indicators in this paper include qualitative and quantitative indicators, which are used to evaluate the popularity of public opinion. Among them, the qualitative indicators are obtained by the expert scoring method. Three experts with the title of associate senior or above engaged in NPO research and five doctoral students engaged in public opinion research are selected to score the study. The three invited experts in the research of non-profit organisations all hold a deputy senior professional title or above, and they have at least five years of research experience in the field of non-profit organisations. They are familiar with the changing patterns of public opinion during emergencies at home and abroad. The five public opinion research experts are all doctoral supervisors or post-doctoral researchers, and their research directions cover areas such as online public opinion and public crisis management. They possess strong public opinion analysis capabilities and academic backgrounds. These experts have a high level of professional proficiency and research experience in their respective fields, and they can provide objective and authoritative evaluations for the research. During the scoring process, to reduce the subjective biases in the experts' scoring, a five-point scoring method is adopted to quantify the influence of public opinion. The specific scoring criteria are as follows: A score of 1 indicates that the influence of public opinion is almost ineffective or extremely weak. A score of 2 indicates that the influence of public opinion is relatively weak with a limited scope of influence. A score of 3 indicates that the influence of public opinion is average, having caused a certain degree of social response. A score of 4 indicates that the influence of public opinion is relatively strong with widespread social response. A score of 5 indicates that the influence of public opinion is extremely strong, attracting widespread attention from all sectors of society and having a significant impact. The scores of all qualitative indicators are quantified through the geometric mean method to reduce the scoring biases of individual experts and ensure the objectivity and consistency of the scoring results.

The data of quantitative indicators is mainly collected by writing Python programs to crawl data and use third-party public opinion monitoring platforms 'Baidu Index' and 'Qingbo Public Opinion'. To ensure the validity of the data, the Scrapy framework is used to crawl the hot news in the advanced search column of microblogs, and the keywords are 'name', 'forward', 'comment', and 'like'. In addition, the number of deaths, the number of injured, the proportion of negative information, the total amount of information released by the media, and the highest amount of public opinion data obtain from the official release of 'Qingbo Public Opinion'. To ensure the timeliness of public opinion information, the data of public opinion events from the beginning to the death are selected. After the processing of the data preprocessing module in the model proposed here, the raw data of the four cases in Table 2 are obtained.

Table 2 Raw data for individual cases

Index	*Chongqing bus crash*	*Thailand cruise ship capsizing incident*	*Zhangjiakou chemical plant explosion*	*Jiang an explosion and combustion incident*
Number of deaths	14	40	23	19
Number of Injured	2	13	22	12
Inverse of the degree of economic loss	2	3	4	2
Duration of the event	10d	13d	4d	3d
Percentage of negative information	59.71%	56.52%	85.49%	77.93%
Number of comments	1,890,810	542,560	57,021	30,549
Number of Retweets	1,172,606	2,303,444	107,703	87,188
Number of Likes	6,522,337	2,232,444	107,703	87,188
Number of participating media	1,619	791	231	96
Total number of media releases	502,704	253,357	36,440	10,667
Peak data volume of public opinion	173,703	52,155	19,074	5,872
Timeliness	5	5	2	1
Degree of truthfulness	5	4	2	1
Accident handling speed	4	3	2	1
Satisfaction with accident results	3	3	2	1

5 Experimental environment and parameter setting

This experiment is conducted on a 64-bit Windows 10 operating system. The hardware configuration includes an Intel(R) Core(TM) i7-7700 3.60GHz processor, 16GB of memory, and four NVIDIA GeForce GTX 1080 graphics cards for acceleration. All the experiments are completed using MATLAB R2021b. The Deep Learning Toolbox in MATLAB is employed for the training and simulation of the network model. To further enhance the development efficiency of the experiment, the network simulation and training processes also uses Python version 3.9 and the TensorFlow deep learning framework for auxiliary development. To ensure that the unit dimension of the data does not affect the accuracy of the model calculation results, the mapminmax function is used to normalise the data before simulation. Simulation is divided into two parts, parameter optimisation and network simulation training. In the first part, the chromosomes are first encoded. Then, the optimal individual is selected, and the operator operation is performed. In the second part, it is mainly divided into three steps. The first step is to generate a BP network using the 'newff' function in the MATLAB environment. The second step is to train the network using the 'train' function. The third step is to perform network simulation using the 'sim' function (Feng et al., 2022; Yang et al., 2022). The model is repeatedly trained, and the training results are revealed in Figure 4.

Figure 4 Neural network training results (see online version for colours)

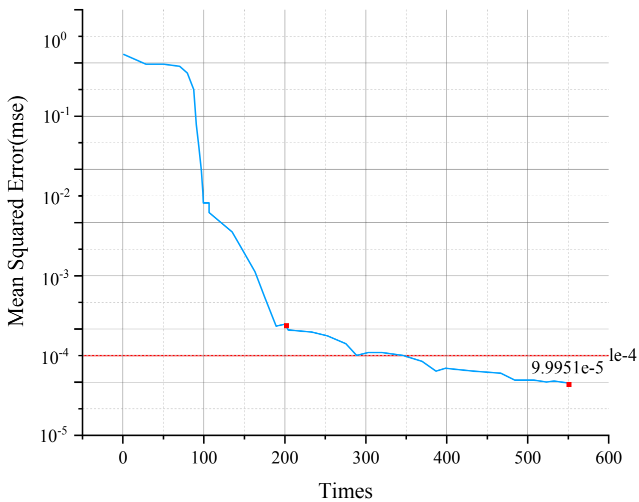


Figure 4 shows that the convergence effect has been significant when the number of training sessions is about 200 times. When the number of trainings is 574, the value of Mean Squared Error (MSE) is 9.9951e-5. The error is less than le-4, which already meets the accuracy requirements.

Table 3 shows the linear fit between the actual output values and the test values.

Table 3 The linear fit between the actual output value and the test value

Target	*Output*
–0.99654	–0.982376
–0.340328	–0.33566
0.336826	0.332145
0.996553	0.968266

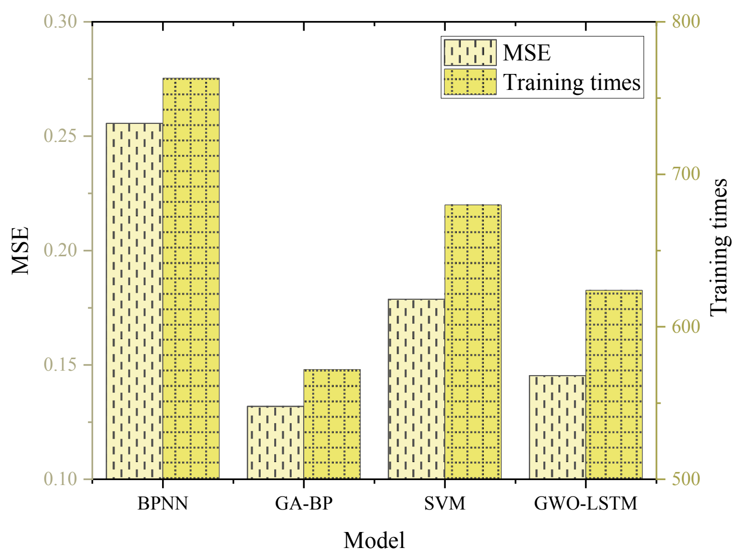
The regression coefficient R^2 is calculated to be 0.999, which is very close to 1. Therefore, the difference between the actual value and the test value is very small, and the two are highly consistent. This suggests that the simulation ability of the model is very strong, and the model is established.

The adaptive momentum gradient descent neural network algorithm is adopted. This algorithm is an improved gradient descent method by reducing oscillations, optimising the algorithm, and exponentially averaging gradients (Mayet et al., 2022). The S-type logarithmic function 'logsig' is selected as the hidden layer neuron transfer function, the S-type tangent function 'tansig' is selected as the output layer transfer function, and trained using the traingdx function suitable for pattern classification (Zheng et al., 2022). The maximum number of training times is set to 5000 times, the initial learning rate is 0.01, the training requirements are to reach the accuracy of le-4, and 50 training iterations are performed. For the GA in the model, the number of individuals is set to 100 to ensure that the algorithm can cover a wide search space. The crossover probability is determined based on experimental experience and optimisation objectives, and it is set at 70%. The mutation probability is set at 5% to maintain the diversity of the population and prevent premature convergence.

5.1 Performance evaluation

To verify the superiority of the constructed model, it is necessary to compare it with the standard BP model and other advanced models. The models involved in the comparison include BPNN, support vector machine (SVM), and GWO-LSTM. A cross-validation method is introduced to evaluate the generalisation ability and stability of the models. MATLAB is used to calculate the models, and training experiments are carried out on the public opinion heat of accident disasters. K-fold cross-validation ($k = 5$) is adopted to train and validate the BPNN, GA-BP model, SVM, and GWO-LSTM models. Each model undergoes five rounds of training on the same dataset, and its mean square error (MSE) and the number of training times are calculated. The results are summarised in Figure 5.

Figure 5 Comparison of training results of different models (see online version for colours)



In Figure 5, GA-BP model is superior to other models in MSE and training times, which shows its obvious advantages in public opinion heat prediction. The average number of training times for the BP model is 763 times, while the average training time for the GA-BP model is 572 times. This reveals that the GA-BP model has a great advantage in global optimisation and can reduce the number of trainings. The average MSE of the standard BP model is 0.2556, while the average MSE of the GA-BP model is only 0.1319. This shows that the GA-BP model has good stability and robustness. When compared with the SVM and GWO-LSTM models, the GA-BP model not only has a lower MSE, but also significantly reduces the number of training times, further demonstrating the advantages of optimisation by the GA. When dealing with complex data, the GA-BP model can effectively avoid the problem of local minima in traditional BP networks, thus improving the training efficiency and prediction accuracy of the model. Therefore, the GA-BP model has obvious advantages, performing excellently in terms of training efficiency, prediction accuracy, and generalisation ability. It provides powerful support for the public opinion prediction and early warning system.

The four cases in Table 2 are input into the model built here. The output results of the four cases are: 0.686, 1.744, 2.672, and 2.912. The output value range of the grade set here is $0 \sim 4$, and the correspondence between the grade and the standard value of the grade is provided in Table 4.

Table 4 The correspondence between the level of public opinion and the standard value of the rank

Grade	*1*	*2*	*3*	*4*
Grade standard value	(0,1]	(1,2]	(2,3]	(3,4]

From Table 4, the four emergencies in the original data in Table 2 correspond to the public opinion popularity levels of level 1, level 2, level 3, and level 3, respectively.

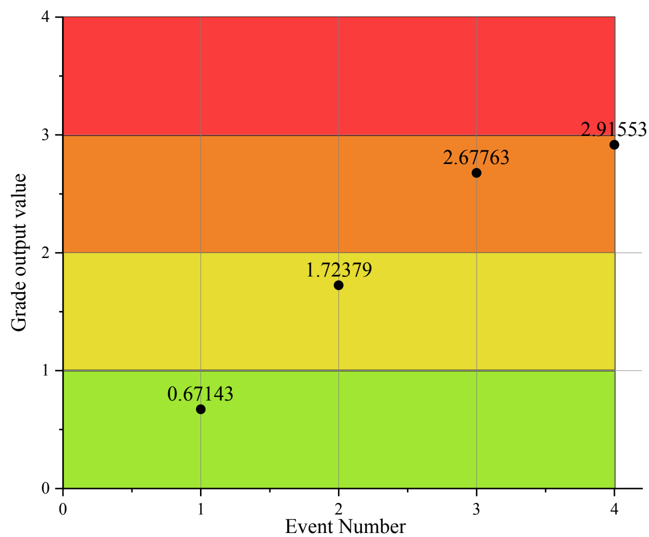
The strength of public opinion for the four emergencies in the original data is plotted in Figure 6.

From the output results, the public opinion popularity of emergency 1 is the lowest, 0.6861, while the popularity of emergency 4 is the highest, 2.9125. This indicates that the former has received more attention, while the latter has received relatively less. The results are consistent with real-world experience and subjective feelings. It is also in line with the attention and assistance of the government and society to these accidents. Therefore, it is very reasonable to use this model for public opinion popularity evaluation, and the output results have a high degree of credibility.

6 Discussion

In recent years, public opinion monitoring and early warning models have been continuously evolving. Zhang et al. (2023) proposed a pattern-based public opinion identification and early warning system. It uses an improved frequent pattern mining algorithm and word weight method to identify public opinion hotspots, and employs cosine similarity and hash table retrieval to distinguish events. Although this study performs excellently in keyword identification and retrieval efficiency, it relies on static pattern rules and is insufficient in dealing with the dynamic evolution of sudden public opinion. Wang et al. (2023) utilised blockchain technology to optimise the risk prediction and credibility detection of online public opinion and constructed a risk management system based on smart contracts. However, the system has a high complexity and its real-time performance still needs improvement. Chen and Zhang (2023) combined edge computing with a deep learning model to establish a sentiment recognition model based on Weibo texts and explored the evolution mechanism of online public opinion in public emergencies. Nevertheless, the predictive ability of this model for the response to sudden incidents is limited. Compared with the above-mentioned studies, the GA-BPNN model proposed in this paper has stronger adaptability and practicality. Its data collection and training mechanism based on time series can achieve real-time evaluation and trend prediction of the public opinion heat of non-profit organisations (NPOs) during emergencies, providing more accurate decision-making support for the government and relevant institutions in public crisis events. In addition, the GA optimisation mechanism introduced in this paper significantly improves the convergence speed and prediction accuracy of the neural network, effectively alleviating the local optimum problem that traditional BPNN may encounter when dealing with large-scale dynamic data.

Figure 6 The relationship between the rank value output and the rank for different cases (see online version for colours)



This paper is based on time series data collection. Therefore, it can not only evaluate the popularity of the entire NPO but also conduct real-time popularity evaluation of the NPO through a specific time point. The evaluation results can help the government understand the attention of public opinion, distinguish the importance and urgency of different events, and make corresponding decisions. In addition, the government can grasp the trend of public opinion changes and predict the future public opinion situation. This is very important for government regulatory departments, because it can prepare in advance, respond to negative public opinion timely, and guide hot topics to develop in a positive and rational direction to ensure the healthy development of public opinion. This makes it possible to detect and monitor the public opinion of emergencies in real time.

This provides public opinion information services for government departments, netizens, and the media. Before the emotions and attitudes of public opinion participants evolve into excessive or too intense, this monitoring and evaluation can provide reference for rumour-debunking and public opinion supervision departments to effectively avoid the secondary hidden harm of public opinion.

However, the model still faces several challenges in practical applications. Firstly, when confronted with large-scale, multi-source heterogeneous sudden public opinion events, the computational efficiency and response speed of the model may still become bottlenecks. Although the GA optimisation has improved the training efficiency, further compression of the calculation time is still required when deploying it into the actual public opinion monitoring system. Secondly, public opinion dissemination paths vary for different types of emergencies such as natural disasters, social conflicts, and political events. The stability and generalisation ability of the model when facing changes in situations still need to be further enhanced. To address the above issues, the following improvements can be made in the future. Firstly, introduce edge computing and distributed processing architectures to achieve local rapid data processing and analysis. Secondly, combine multi-model fusion strategies, such as integrating deep learning, graph neural networks, and sentiment analysis technologies, to improve the ability to capture complex semantic information. Thirdly, strengthen the model's transfer learning ability in different public opinion scenarios to enhance its cross-domain application performance. In conclusion, the constructed GA-BPNN model outperforms existing methods in terms of prediction accuracy, real-time performance, and model optimisation, providing a new technical path for the research on the identification of online public opinion risks and early warning. At the same time, it has important reference value for the government's public opinion management and social risk governance practices. In the future, this model can also be extended and applied to fields such as financial risk early warning and medical information public opinion monitoring, further enhancing its academic influence and application scope.

7 Conclusion

7.1 Research contribution

The following conclusions are drawn from the research. Firstly, the BPNN optimised by GA is used to establish an NPO recognition and EW model for emergencies, which can greatly reduce the influence of initial weights and thresholds on evaluation results. The average MSE of the standard BP model is 0.2556, while the average MSE of the GA-BP model is only 0.1319, which indicates that the GA-BP model has good stability and robustness. At the same time, compared with SVM and GWO-LSTM, GA-BP also shows lower prediction error. Secondly, relevant public opinion events are used, and data is extracted to establish evaluation standards. The NPO of the emergency is successfully scientifically evaluated by constructing the index system and model and using MATLAB for accurate calculation. The evaluation results show that this paper can truly and efficiently reflect the current situation of the selected public opinion cases and solve the problem of NPO identification and EW of emergencies.

7.2 Future works and research limitations

The identification and EW of NPO is a combination of many factors, subject to a series of constraints. Especially when the external conditions change, and the era conditions of public opinion change, it remains to be tested whether the evaluation indicators and models reported here can adapt. To avoid evaluation failures due to changes in new conditions, the next step will be to train models using newer and more comprehensive data to improve the universality of recognition and EW models.

Funding

This work was supported by “Research on the Operation Effect and Optimisation Path of Risk Sharing Model for Mortgage Financing of Agricultural Land Management Rights in Western Guangdong Province” (Grant No. GD24CYJ04).

Conflicts of interest

All Authors declare that they have no conflict of interest.

References

Azizi, H., Aaleagha, M.M., Azadbakht, B. and Samadyar, H. (2022) ‘Identification and Assessment of health, safety and environmental risk factors of Chemical Industry using Delphi and FMEA methods (a case study)’, *Anthropogenic Pollution*, Vol. 6, No. 2, pp.108–121.

Bai, L., Wei, L., Zhang, Y., Zheng, K. and Zhou, X. (2024) ‘GA-BP neural network modeling for project portfolio risk prediction’, *Journal of Enterprise Information Management*, Vol. 37, No. 3, pp.828–850.

Bao, Q., Chen, Y. and Ji, X. (2025) *Research on Evolution and Early Warning Model of Network Public Opinion Based on Online Latent Dirichlet Distribution Model and BP Neural Network*, arXiv preprint arXiv: 2503.03755.

Batool, S., Rashid, J., Nisar, M.W., Kim, J., Kwon, H.Y. and Hussain, A. (2023) ‘Educational data mining to predict students’ academic performance: a survey study’, *Education and Information Technologies*, Vol. 28, No. 1, pp.905–971.

Čábelková, I., Smutka, L. and Strielkowski, W. (2022) ‘Public support for sustainable development and environmental policy: a case of the Czech Republic’, *Sustainable Development*, Vol. 30, No. 1, pp.110–126.

Cao, C., Li, Z., Li, L. and Luo, F. (2025) ‘Research on the evolution of college online public opinion risk based on improved grey wolf optimizer combined with LSTM model’, *PloS One*, Vol. 20, No. 1, p.e0311749.

Carnegie, A. and Gaikwad, N. (2022) ‘Public opinion on geopolitics and trade: theory and evidence’, *World Politics*, Vol. 74, No. 2, pp.167–204.

Chao, Y., Zhu, H. and Zhou, Y. (2024) ‘Integrating visual transformer and graph neural network for visual analysis in digital marketing: exploring and predicting advertising effectiveness’, *Journal of Organizational and End User Computing (JOEUC)*, Vol. 36, No. 1, pp.1–28.

Chen, D. and Cheng, P. (2022) ‘Development of design system for product pattern design based on kansei engineering and BP neural network’, *International Journal of Clothing Science and Technology*, Vol. 34, No. 3, pp.335–346.

Chen, M. and Zhang, L. (2023) 'Application of edge computing combined with deep learning model in the dynamic evolution of network public opinion in emergencies', *The Journal of Supercomputing*, Vol. 79, No. 2, pp.1526–1543.

Chen, W., Li, Y., Liu, Z., Zhang, C., Zhao, Y. and Yan, X. (2025) 'Prediction model for bearing surface friction coefficient in bolted joints based on GA-BP neural network and experimental data', *Tribology International*, Vol. 201, p.110217.

Chen, Y.J. and Chen, Z.S. (2022) 'A prediction model of wall shear stress for ultra-high-pressure water-jet nozzle based on hybrid BP neural network', *Engineering Applications of Computational Fluid Mechanics*, Vol. 16, No. 1, pp.1902–1920.

Chen, Y., Chen, S., Wu, Z., Dai, B., Xv, L. and Wu, G. (2022) 'Optimization of genetic algorithm through use of back propagation neural network in forecasting smooth wall blasting parameters', *Mathematics*, Vol. 10, No. 8, p.1271.

Dong, Y., Li, X., Zhang, J., Li, Z. and Hou, D. (2023) 'Application of fractional theory in quantum back propagation neural network', *Mathematical Methods in the Applied Sciences*, Vol. 46, No. 3, pp.3080–3090.

Dong, Z. and Asif, M. (2024) 'Application of compound neural networks to classifying corporate green technology investments', *Journal of Organizational and End User Computing (JOEUC)*, Vol. 36, No. 1, pp.1–24.

Feng, L., Tian, X., El-Kassaby, Y.A., Qiu, J., Feng, Z., Sun, J. and Wang, T. (2022) 'Predicting suitable habitats of Melia Azedarach L. in China using data mining', *Scientific Reports*, Vol. 12, No. 1, p.12617.

Foos, F. and Bischof, D. (2022) 'Tabloid media campaigns and public opinion: quasi-experimental evidence on euroscepticism in England', *American Political Science Review*, Vol. 116, No. 1, pp.19–37.

Franceschini, R., Rosi, A., Catani, F. and Casagli, N. (2022) 'Exploring a landslide inventory created by automated web data mining: the case of Italy', *Landslides*, Vol. 19, No. 4, pp.841–853.

Kiran, B.R., Kopperi, H. and Venkata Mohan, S. (2022) 'Micro/nano-plastics occurrence, identification, risk analysis and mitigation: challenges and perspectives', *Reviews in Environmental Science and Bio/Technology*, Vol. 21, No. 1, pp.169–203.

Li, X., Wang, J. and Yang, C. (2023) 'Risk prediction in financial management of listed companies based on optimized BP neural network under digital economy', *Neural Computing and Applications*, Vol. 35, No. 3, pp.2045–2058.

Li, Y., Xu, X., Riaz, M. and Su, Y. (2024) 'Risk identification of public opinion on social media: a new approach based on cross-spatial network analysis', *The Electronic Library*, Vol. 42, No. 4, pp.576–597.

Lin, Y., Yu, K., Hao, L., Wang, J. and Bu, J. (2022) 'An indoor Wi-Fi localization algorithm using ranging model constructed with transformed RSSI and BP neural network', *IEEE Transactions on Communications*, Vol. 70, No. 3, pp.2163–2177.

Liu, C. (2022) 'Who supports expanding surveillance? exploring public opinion of Chinese social credit systems', *International Sociology*, Vol. 37, No. 3, pp.391–412.

Liu, C., Feng, Y. and Wang, Y. (2022) 'An innovative evaluation method for undergraduate education: an approach based on BP neural network and stress testing', *Studies in Higher Education*, Vol. 47, No. 1, pp.212–228.

Liu, D., Liu, C., Tang, Y. and Gong, C. (2022) 'A GA-BP neural network regression model for predicting soil moisture in slope ecological protection', *Sustainability*, Vol. 14, No. 3, p.1386.

Liu, J., Liu, L., Tu, Y., Li, S. and Li, Z. (2022) 'Multi-stage internet public opinion risk grading analysis of public health emergencies: an empirical study on microblog in COVID-19', *Information Processing and Management*, Vol. 59, No. 1, p.102796.

Liu, J., Xu, Z., Yang, Y., Zhou, K. and Kumar, M. (2024) 'Dynamic prediction model of financial asset volatility based on bidirectional recurrent neural networks', *Journal of Organizational and End User Computing (JOEUC)*, Vol. 36, No. 1, pp.1–23.

Lyu, H., Wang, J., Wu, W., Duong, V., Zhang, X., Dye, T.D. and Luo, J. (2022) 'Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination', *Intelligent Medicine*, Vol. 2, No. 01, pp.1–12.

Lyu, Z., Yu, Y., Samali, B., Rashidi, M., Mohammadi, M., Nguyen, T.N. and Nguyen, A. (2022) 'Back-propagation neural network optimized by K-fold cross-validation for prediction of torsional strength of reinforced concrete beam', *Materials*, Vol. 15, No. 4, p.1477.

Mayet, A.M., Salama, A.S., Alizadeh, S.M., Nesic, S., Guerrero, J.W.G., Eftekhari-Zadeh, E. and Iliyasu, A.M. (2022) 'Applying data mining and artificial intelligence techniques for high precision measuring of the two-phase flow's characteristics independent of the pipe's scale layer', *Electronics*, Vol. 11, No. 3, p.459.

Muñoz-Rodríguez, J.M., Alonso, C.P., Pessoa, T. and Martín-Lucas, J. (2023) 'Identity profile of young people experiencing a sense of risk on the internet: a data mining application of decision tree with CHAID algorithm', *Computers and Education*, Vol. 197, 104743.

Pascual-Ferrá, P., Alperstein, N. and Barnett, D.J. (2022) 'Social network analysis of COVID-19 public discourse on twitter: implications for risk communication', *Disaster Medicine and Public Health Preparedness*, Vol. 16, No. 2, pp.561–569.

Qie, X., Kang, C., Zong, G. and Chen, S. (2022) 'Trajectory planning and simulation study of redundant robotic arm for upper limb rehabilitation based on back propagation neural network and genetic algorithm', *Sensors*, Vol. 22, No. 11, p.4071.

Qin, Q., Yang, X., Zhang, R., Liu, M. and Ma, Y. (2022) 'An application of deep belief networks in early warning for cerebrovascular disease risk', *Journal of Organizational and End User Computing (JOEUC)*, Vol. 34, No. 4, pp.1–14.

Santos-Pereira, J., Gruenwald, L. and Bernardino, J. (2022) 'Top data mining tools for the healthcare industry', *Journal of King Saud University-Computer and Information Sciences*, Vol. 34, No. 8, pp.4968–4982.

Sivamani, S., Prasad, B.N., Nithya, K., Sivarajasekar, N. and Hosseini-Bandegharaei, A. (2022) 'Back-propagation neural network: box–Behnken design modelling for optimization of copper adsorption on orange zest biochar', *International Journal of Environmental Science and Technology*, Vol. 19, No. 5, pp.4321–4336.

Sunhare, P., Chowdhary, R.R. and Chattopadhyay, M.K. (2022) 'Internet of things and data mining: an application oriented survey', *Journal of King Saud University-Computer and Information Sciences*, Vol. 34, No. 6, pp.3569–3590.

Velte, P. (2022) 'Meta-analyses on corporate social responsibility (CSR):'*a Literature Review'*, *Management Review Quarterly*, Vol. 72, No. 3, pp.627–675.

Wang, E.S.T. and Hu, F.T. (2022) 'Influence of self-disclosure of internet celebrities on normative commitment: the mediating role of para-social interaction', *Journal of Research in Interactive Marketing*, Vol. 16, No. 2, pp.292–309.

Wang, M., Dong, Y., Cai, J., Liu, H., Liu, Z. and Zhang, Z. (2023) 'Method for reservoir permeability prediction and quality evaluation based on BP neural network', *Progress in Geophysics*, Vol. 38, No. 1, pp.321–327.

Wang, X. (2025) 'The artificial intelligence-based agricultural field irrigation warning system using GA-BP neural network under smart agriculture', *PloS One*, Vol. 20, No. 1, e0317277.

Wang, Z., Zhang, S., Zhao, Y., Chen, C. and Dong, X. (2023) 'Risk prediction and credibility detection of network public opinion using blockchain technology', *Technological Forecasting and Social Change*, Vol. 187, p.122177.

Xin, J., Chen, J., Li, C., Lu, R.K., Li, X., Wang, C. and He, R. (2022) 'Deformation characterization of oil and gas pipeline by ACM technique based on SSA-BP neural network model', *Measurement*, Vol. 189, p.110654.

Yang, H., Zhou, L., Cai, J., Shi, C., Yang, Y., Zhao, X. and Yin, X. (2023) 'Data mining techniques on astronomical spectra data–II classification analysis', *Monthly Notices of the Royal Astronomical Society*, Vol. 518, No. 4, pp.5904–5928.

Yang, Y., Yuan, Z. and Meng, R. (2022) 'Exploring traffic crash occurrence mechanism toward cross-area freeways via an improved data mining approach', *Journal of Transportation Engineering, Part A: Systems*, Vol. 148, No. 9, p.04022052.

Zhang, M., Fan, J., Sharma, A. and Kukkar, A. (2022) 'Data mining applications in university information management system development', *Journal of Intelligent Systems*, Vol. 31, No. 1, pp.207–220.

Zhang, X.R., Sun, X., Sun, W., Xu, T., Wang, P.P. and Jha, S.K. (2022) 'Deformation expression of soft tissue based on BP neural network', *Intelligent Automation and Soft Computing*, Vol. 32, No. 2, pp.1041–1053.

Zhang, Y., Chen, F. and Rohe, K. (2022) 'Social media public opinion as flocks in a murmuration: conceptualizing and measuring opinion expression on social media', *Journal of Computer-Mediated Communication*, Vol. 27, No. 1, p.zmab021.

Zhang, Z., Lin, X. and Shan, S. (2023) 'Big data-assisted urban governance: an intelligent real-time monitoring and early warning system for public opinion in government hotline', *Future Generation Computer Systems*, Vol. 144, pp.90–104.

Zhao, Y., Chen, W., Arashpour, M., Yang, Z., Shao, C. and Li, C. (2022) 'Predicting delays in prefabricated projects: SD-BP neural network to define effects of risk disruption', *Engineering, Construction and Architectural Management*, Vol. 29, No. 4, pp.1753–1776.

Zheng, Y., Lv, X., Qian, L. and Liu, X. (2022) 'An optimal BP neural network track prediction method based on a GA-ACO hybrid algorithm', *Journal of Marine Science and Engineering*, Vol. 10, No. 10, p.1399.