Bulletin of the New Zealand Society for Earthquake Engineering, Vol. 57, No. 4, December 2024

PREDICTION MODEL FOR EARTHQUAKE DEATH TOLL BASED ON PCA-BAS-ELM IN MAINLAND CHINA

Chenhui Wang[1], Xiaoshan Wang[2], Xiaotao Zhang[3*], Guojun Lv[4], Libing Wang[5] and Na Luo[6]

(Submitted *May 2024*; Reviewed *July 2024*; Accepted *August 2024*)

ABSTRACT

In recent years, China has experienced frequent catastrophic earthquakes, causing huge casualties. If the death toll can be quickly predicted after a disaster, then relief supplies can be delivered in a timely and reasonable manner, and the death toll and property losses can be minimized. Therefore, rapid and effective prediction of earthquake deaths plays a key role in guiding post-earthquake emergency rescue. However, there are many factors affecting the number of deaths in an earthquake. Aimed at this issue, a prediction model for earthquake deaths based on extreme learning machine (ELM) optimized by principal component analysis (PCA) and beetle antennae search (BAS) algorithm has been proposed in this study. Firstly, this study selected sample data of destructive earthquakes in mainland China in the past 50 years, then PCA was used to reduce the dimensionality of the factors affecting earthquake deaths, the principal components with lower contribution rates were removed, and the principal components with higher contribution rates were used as the input variables of ELM. Meanwhile, the earthquake deaths were used as the output variable, and the connection weights and thresholds of ELM was optimized using BAS. Finally, the prediction model for earthquake deaths based on PCA-BAS-ELM was established. The established model was used to predict the test samples. The results showed that the prediction results of PCA-BAS-ELM model had a higher fit with the actual values, and its mean square error, mean absolute percentage error and root mean square error were 2.433, 2.756% and 5.443, respectively, which suggested higher prediction accuracy.

https://doi.org/10.5459/bnzsee.1699

INTRODUCTION

China is one of the countries with the highest risk of earthquake disasters. The huge population base and the fragile engineering systems have caused China to suffer huge casualties and property losses in many earthquakes. Since it is usually difficult to understand the situation in the affected area immediately after an earthquake, the government needs to decide the level of response, activate emergency plans, deploy rescue forces, and carry out rescue work reasonably and efficiently based on preliminary predictions of the earthquake effect. Among them, the prediction of earthquake deaths plays an important role in earthquake disaster assessment, and the death situation is also a key basis for determining the earthquake emergency response level [1-2].

In the past 40 years, scholars have conducted a lot of research on earthquake death prediction methods. Earthquake death assessment models can be divided into two categories: empirical models based on seismic parameters and analytical models based on building vulnerability. The first type of model generally performs regression analysis based on historical earthquake damage data, and then obtaining empirical equations for magnitude or intensity, without considering the distribution of disaster-bearing bodies and their vulnerability [3-6]. The second type of model considers detailed building vulnerability data and estimates the number of earthquake deaths based on the vulnerability of different components and building structures in earthquakes [7-11]. In addition, with the gradual maturity of machine learning methods, machine learning has also been introduced into the research related to earthquake death assessment methods to improve the accuracy of the assessment models [12-14]. As a feedforward neural network, extreme learning machine (ELM) has a wide applicability, and it appears to be suitable for dealing with small sample and nonlinear problems such as earthquake death prediction. Therefore, this paper proposes an earthquake death prediction model based on ELM optimized by principal component analysis (PCA) and beetle antennae search (BAS). Firstly, we used PCA to reduce the dimension of influencing factors for earthquake death toll. Then using BAS to train ELM to obtain the optimal parameters, and the prediction model for earthquake death toll was established, which was used to predict the test samples with good results.

BASIC PRINCIPLE

Principal Component Analysis (PCA)

The PCA method constructs a linear combination of the original variables to obtain new variables with fewer dimensions and no correlation with each other. The solution process is as follows: Let ($X_1, X_2, \ldots Xn$) be n samples of the population X, each sample corresponds to an m-dimensional variable, and the corresponding matrix is:

$$X_{m\times n} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \quad (1)$$

Firstly, find the covariance matrix of the above matrix and convert the covariance matrix into correlation coefficient matrix. Then find the eigenvalue of the correlation coefficient

[1] *Intermediate Research Engineer, Hebei Earthquake Agency, China, Email: Caesar621@163.com*
[2] *Senior Research Engineer, Hebei Earthquake Agency, China, Email: .wangxsh2022@163.com*
[3] *Corresponding Author, Hebei Earthquake Agency, China, Email: 249595254@qq.com*
[4] *Senior Research Engineer, Hebei Earthquake Agency, China, Email: 328293676@qq.com*
[5] *Senior Research Engineer, Hebei Earthquake Agency, China, Email: 276877415@qq.com*
[6] *Senior Research Engineer, Hebei Earthquake Agency, China, Email: 93640348@qq.com*

matrix $\lambda_1 \geq \lambda_2 \geq ...\lambda_m \geq 0$. The contribution rate of the *i*-th principal component is λ_i / P , where *i*=1,2,3…m, $\mathrm{P} = \sum_{i=1}^{m} \lambda_i$, and the cumulative contribution rate of the first *q* principal components is $\sum_{i=1}^{q} \lambda_i / \mathrm{P}$. The newly generated first *q* principal component data form the new ELM input data matrix $Z_{k\times m}$. Usually, when the cumulative contribution rate reaches more than 85%, we can use the first *q* principal components to replace the original influencing factors, which meets the actual requirements [15-17].

Beetle Antennae Search (BAS) Algorithm

The beetle whisker algorithm is inspired by the beetle's foraging behaviour. The beetle finds the location of food by judging the odour concentration received by the left and right antennae. That is to say, the direction with high concentration is the direction that the beetle constantly adjusts until the food is found. Compared with traditional optimization algorithms, the advantage of BAS is that it can continuously approximate the true function expression without knowing any relevant information of the unknown function, which greatly simplifies the calculation process [18-19]. The specific ideas are as follows:

1) Set the initial value. The global dimension is *W* (*W* represents the number of newly generated principal components after dimensionality reduction of the original impact indicators.), the number of termination iterations is *T*, the step size is δ_0, the distance between the two antennae is *d*, the ratio of δ_0 to *d* is *c*, and the center particle position of beetle is x_0.
2) Establish the position function *f(x)*. Define the optimal position $x'_{\mathrm{best}} = x_0$, the maximum concentration value between the two antennae $f'_{\mathrm{best}} = f(x'_{\mathrm{best}})$. f'_{best} represents the highest concentration perceived by the beetle antennae, which corresponds to the optimal parameters found by the BAS algorithm for the ELM.
3) Set the direction vector of beetle whiskers and standardize it:

$$\boldsymbol{a} = \frac{\mathrm{rand}(W, 1)}{\|\mathrm{rand}(W, 1)\|} \tag{2}$$

where rand () is a custom function, and the left side of $\boldsymbol{a}$ is defined as positive.

4) Define the spatial function of beetle whiskers:

$$\begin{cases} x_{lt} = x_t + \mathrm{d}\boldsymbol{a} / 2 \\ x_{rt} = x_t - \mathrm{d}\boldsymbol{a} / 2 \end{cases} \tag{3}$$

where x_{lt} is the position of the left antennae when searching for the best solution *t* times, x_{rt} is the position of the right antennae when searching for the best solution *t* times, and x_t is the center of mass location for the beetle when searching for the best solution *t* times.

5) The position function of beetle at iteration *t*+1:

$$x_{t+1} = x_t - \delta_t \mathrm{sign}(f(x_{rt}) - f(x_{lt}))\boldsymbol{a} \tag{4}$$

where δ_t is the step length when optimizing *t* times, sign () is the carrier function; $f(x_{rt})$ and $f(x_{lt})$ are the concentrations received by the two whiskers.

6) Compare the concentration values and update the optimal coordinate position. When $x_{rt}=x_{lt}$, if $f(x_{t+1}) > f'_{\mathrm{best}}$, then $x'_{\mathrm{best}} = x_{t+1}$, $f'_{\mathrm{best}} = f(x_{t+1})$.

Extreme Learning Machine (ELM)

ELM is a single hidden layer feedforward neural network, unlike traditional neural networks, it only needs to determine the number of hidden layers in the neural structure and generate a unique optimal solution to avoid falling into a local optimal solution. Therefore, ELM has the characteristics of fast learning efficiency and good generalization performance [20-21]. The network structure of ELM is shown in Figure 1.

The principal component data generated by PCA is used as the input of ELM. Assuming that there are *P* and *Q* neurons in the hidden layer and output layer, the output result corresponding to sample *k* is:

$$\hat{\mathbf{y}}_k = \begin{bmatrix} \sum_{p=1}^{P} \beta_{p1} g(\boldsymbol{\omega}_p \boldsymbol{z}_k + b_p) \\ \sum_{p=1}^{P} \beta_{p2} g(\boldsymbol{\omega}_p \boldsymbol{z}_k + b_p) \\ \mathrm{M} \\ \sum_{p=1}^{P} \beta_{pQ} g(\boldsymbol{\omega}_p \boldsymbol{z}_k + b_p) \end{bmatrix} \tag{5}$$

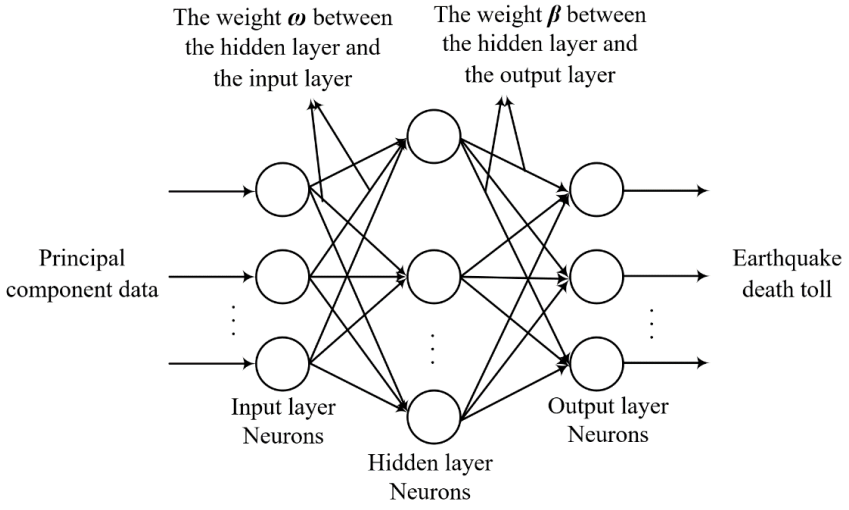

Figure 1: Network structure of ELM.

where: p = 1,2,L...,P; q = 1,2,L...,Q; k = 1,2,L...,K; ω_p is the coefficient between the p-th hidden layer and the input layer, b_p is the hidden layer threshold, β_{pQ} is the coefficient of the p-th hidden layer neuron and the q-th output neuron: z_k is the kth principal component data; $g(x)$ is the activation function.

This section focuses on the roles of the activation function $g(x)$ and the hidden layer threshold b_p. The activation function $g(x)$ converts the linear combination results of the hidden layer neurons into nonlinear outputs, enabling the neural network to approximate arbitrarily complex functions and achieve a high degree of fit between earthquake fatalities and principal component data. The hidden layer threshold b_p adjusts the input values of the activation function, allowing each hidden layer node to respond more flexibly to different principal component data features. This adjustment influences the output of the activation function, making the neurons more sensitive or robust to the principal component data, thereby increasing the model's flexibility and expressive capability.

The matrix form of equation (5) is:

$$\boldsymbol{H}\boldsymbol{\beta} = \hat{\boldsymbol{Y}}^{\mathrm{T}} \tag{6}$$

where $\boldsymbol{H}$ is the output matrix of hidden layer, in earthquake mortality prediction, it uses nonlinear transformation to map the input principal component data to a new feature space. This enables the model to better capture and represent the complex relationship between the input data and earthquake mortality, thereby improving prediction accuracy and robustness. $\boldsymbol{\beta}$ is the coefficient matrix between hidden layer and output layer, $\hat{\boldsymbol{Y}} = [\hat{\boldsymbol{y}}_1, \hat{\boldsymbol{y}}_2, \mathrm{L}\ , \hat{\boldsymbol{y}}_K]$ is the output matrix of the model and represents the predicted values of earthquake fatalities corresponding to different earthquake samples. When $\boldsymbol{H}$ is obtained, the coefficient between output layer and hidden layer will be achieved:

$$\boldsymbol{\beta} = H^{\mathrm{T}}(\frac{1}{C} + HH^{\mathrm{T}})^{-1}\hat{\boldsymbol{Y}}^{\mathrm{T}} \tag{7}$$

where C represents the penalty parameter, H^T and H are transposed. By adjusting the value of C, one can find the optimal balance between fitting the training data and preventing the model from becoming overly complex. This helps control the size of the output layer weights $\boldsymbol{\beta}$, preventing the model from overfitting the training data and enhancing its generalization ability, thereby avoiding an overly complex or simple model.

The ELM model construction process is as follows: First, determine the number of input neurons. Then, specify the activation function and the initial number of hidden layer neurons. Finally, solve for the hidden layer output matrix and the weights between the hidden and output layers, adjusting the number of hidden layer neurons to determine the final network structure.

MODEL BUILDING AND PREDICTION

Due to the number of ELM input nodes will affect the complexity of network structure, and the randomly generated input weights and hidden layer thresholds will cause dimensionality disaster and poor prediction accuracy. In order to solve this issue, this paper used PCA to reduce the dimension of input data, then introduced BAS to find the ELM parameters. The specific implementation steps are as follows: First, the raw data was normalized, as detailed in equation (11) in section 3.2; Second, PCA was used to reduce the dimension of impact indicators of earthquake deaths, and the newly generated principal component data was randomly divided into training samples and test samples. Then initialized the BAS parameters and determined the constraint function which was the mean square error between actual value and predicted value. The BAS algorithm was used to continuously optimize the ELM weight until the maximum number of iterations was reached. Finally, an earthquake deaths prediction model based on PCA-BAS-ELM was established. The prediction flow chart of PCA-BAS-ELM model is shown in Figure 2.

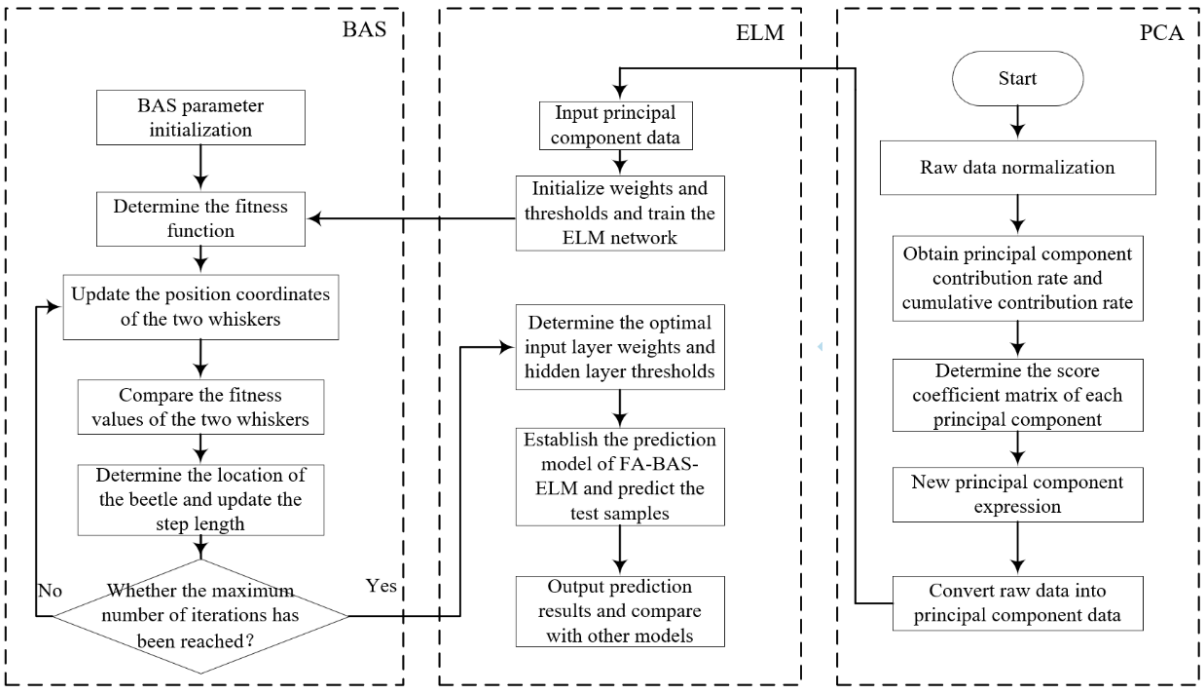

Figure 2: The flow chart of PCA-BAS-ELM model.

Model Evaluation Indicators

In order to verify the effectiveness and superiority of PCA-BAS-ELM model, the following three statistical parameters are used to evaluate the performance of the model, namely, mean absolute error (*MAE*), mean absolute percentage error (MAPE) and root mean square error (*RMSE*), which are calculated as follows:

$$MAE=\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i| \tag{8}$$

$$MAPE=100\times\frac{1}{N}\sum_{i=1}^{N}\frac{|y_i-\hat{y}_i|}{y_i} \tag{9}$$

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2} \tag{10}$$

where y_i is the actual value and $\hat{y}_i$ is the predicted value.

Impact Indicator Analysis and Data Sources

As is known to all, earthquake deaths are influenced by multiple impact indicators such as earthquake magnitude, earthquake time, focal depth, epicenter intensity, seismic fortification intensity, the difference between the epicentral intensity and the seismic fortification intensity, and the population density.

Magnitude: This measures the energy release of an earthquake. The larger the magnitude, the more energy is released, resulting in greater damage to the surface and buildings, and typically leading to more casualties.

Focal depth: This refers to the depth of the earthquake's focus from the earth's surface. Generally, the shallower the focus, the stronger the surface shaking and the greater the potential damage and casualties. Shallow earthquakes (focal depth less than 70 km) are usually more destructive than deep earthquakes (focal depth greater than 300 km) because the energy of the seismic waves diminishes less by the time they reach the surface.

Epicentral intensity: This indicates the degree of ground damage at the earthquake's epicenter. Higher epicentral intensity means more severe damage to the ground and buildings, leading to higher casualties. Areas with high seismic intensity often experience severe building collapse and infrastructure damage, resulting in increased direct casualties.

Seismic fortification intensity: This reflects the seismic strength considered during the design and construction of buildings and infrastructure according to national or regional seismic codes. Higher seismic fortification intensity implies that buildings are constructed to higher seismic standards, enhancing their ability to withstand earthquakes and thus reducing casualties. In areas with lower seismic fortification intensity, even moderate earthquakes can cause significant building collapse and casualties.

Difference between epicentral intensity and seismic fortification intensity: This gap indicates the difference between the actual earthquake intensity and the designed seismic resistance of buildings. If the epicentral intensity exceeds the seismic fortification intensity (a positive difference), buildings may not withstand the earthquake's strength, leading to severe damage and high mortality rates. Conversely, if the epicentral intensity is less than or equal to the seismic fortification intensity (zero or negative difference), buildings can better resist the earthquake, reducing casualties.

Population density: This is the number of people per unit area, typically measured in persons per square kilometer. Higher population density means more people are affected during an earthquake, potentially leading to more casualties. High population density areas are prone to secondary disasters like building collapse, fires, and traffic blockages, which can increase casualties.

Earthquake time: This is the specific time an earthquake occurs, usually recorded in hours and minutes. The timing significantly affects casualty numbers. For example, earthquakes occurring at night, when people are asleep, may result in higher casualties due to people being indoors with less time to escape and slower reaction times. Conversely, earthquakes during the day, when people are likely outdoors or at work, may result in fewer casualties due to easier evacuation.

Before conducting principal component analysis, we classified the influencing factors as follows: magnitude, focal depth, and epicentral intensity are considered seismic parameters; seismic fortification intensity and the difference between epicentral intensity and seismic fortification intensity are engineering parameters; and population density and earthquake timing are social activity parameters. These classifications helped ensure that our principal component analysis captured the most relevant factors affecting earthquake casualties.

Based on this, this paper selected 35 destructive earthquake sample data in the past 50 years, 25 samples were randomly selected as learning samples, and the remaining 10 samples were used as test samples. The partial raw data is shown in Table 1, where S1 is the earthquake time, the earthquake time 1 represents daytime (7:00-19:00), 0 represents nighttime (19:00-7:00), S2 is the magnitude, S3 is focal depth (km), S4 is seismic fortification intensity, S5 is seismic fortification intensity, S6 is the difference between the epicentral intensity and the seismic fortification intensity, and S7 is the population density (people·km^{-2}). In order to eliminate the difference between dimensions, to avoid the influence caused by different dimensions and to ensure the stability and effectiveness of the model training process, we normalized the original data on earthquake fatalities. This normalization ensures that each influencing factor's data falls within the [0,1] range. The normalization equation is as follows:

$$X_i^{*}=(X_i-X_{\min})/(X_{\max}-X_{\min}) \tag{11}$$

where: X_i and X_i^* are the i^{th} sample value and its corresponding normalized value of each influencing factor, respectively; X_{min} and X_{max} are the minimum and maximum value of each principal influencing factor, respectively.

PCA of the Data

Due to the large number of impact indicators, in order to improve the independence of impact factors and remove the correlation among them, SPSS 20 was used to perform PCA analysis on seismic sample data, then the eigenvalues, the contribution rate of each principal component and the cumulative contribution rate were shown in Table 2, and the score coefficients of each principal component were shown in the Table 3. Table 2 shows that the cumulative contribution rate of the first four principal components is 89.417%, covering the main information of the original data, thus the first four principal components can be used as new input variables for the prediction model. The expression of the first four principal components can be obtained as Equation (12) according to Table 3. Table 3 shows that the magnitude and epicenter intensity mainly belong to the principal component Y_1, the focal depth and seismic fortification in-tensity mainly belong to the principal component Y_2, the population density mainly belongs to the principal component Y_3, and the earthquake time mainly belongs to the principal component Y_4. As a result, the input dimension of the model was reduced from 7 dimensions to 4 dimensions, which improved the model operation efficiency to a certain extent.

Table 1: Seismic sample data.

Serial number	Date	Location	S1	S2	S3	S4	S5	S6	S7	Total deaths
1	1970.12.03	Xiji	1	5.5	12	7	8	-1	142	117
2	1976.04.06	Heilinger	0	6.3	18	7	7	0	55	28
3	1976.07.28	Tangshan	0	7.8	12	11	8	3	401	242769
4	1978.05.18	Yingkou	0	5.9	13	7	7	0	1038	11
5	1979.07.09	Liyang	0	6	12	8	7	1	512	41
6	1981.01.24	Daofu	1	6.9	12	8	8	0	7	123
7	1986.11.15	Hualian	0	7.3	34	9	9	0	619	12
8	1990.04.26	Gonghe	1	6.9	12	9	7	2	8	119
9	1995.07.12	Menglian	0	7.3	10	8	8	0	59	11
10	1995.07.22	Yongdeng	1	5.8	10	8	8	0	82	10
11	1995.10.24	Wuding	1	6.5	15	9	7	2	79	58
12	1996.02.03	Lijiang	0	7	10	9	8	1	59	309
13	1996.03.19	Jiashi	0	6.9	11	8	8	0	46	24
14	1996.05.06	Baotou	1	6.4	24	8	8	0	480	26
15	1997.01.21	Jiashi	1	6.9	11	8	8	0	46	14
16	1998.01.10	Zhangbei	1	6.2	10	8	7	1	88	49
17	2003.02.24	Jiashi	1	6.8	25	9	8	1	46	268
18	2003.07.21	Dayao	0	6.2	6	8	7	1	68	16
19	2003.10.25	Minle	0	6.1	18	8	7	1	81	10
20	2003.12.01	Zhaosu	1	6	18	8	8	0	14	10
21	2005.11.26	Jiujiang	1	5.7	10	7	6	1	751	13
22	2006.07.22	Yanjin	1	5.1	9	6	7	-1	171	22
23	2008.05.12	Wenchuan	1	8	14	11	8	3	27	69227
24	2008.08.30	Panzhihua	1	6.1	10	8	7	1	325	41
25	2008.10.06	Dangxiong	1	6.6	8	8	9	-1	3	10
26	2010.01.14	Yushu	1	7.1	14	9	7	2	9	2698
27	2011.03.10	Yingjiang	1	5.8	10	8	7	1	59	25
28	2012.09.07	Yiliang	1	5.7	14	8	7	1	177	81
29	2012.04.20	Ya’an	1	7	13	9	7	2	87	196
30	2013.07.22	Zhangxian	1	6.6	20	8	8	0	87	95
31	2014.11.22	Kangding	1	6.3	18	8	9	-1	11	10
32	2016.02.06	Gaoxiong	1	6.7	15	9	8	1	943	117
33	2017.08.08	Jiuzhaigou	0	7.0	20	9	8	-1	15	25
34	2022.09.05	Luding	1	6.8	16	9	8	-1	40	93
35	2023.12.18	Jishishan	0	6.2	10	8	7	-1	263	151

Table 2: Eigenvalues, contributions of principal components and cumulative contributions.

component	normalised eigenvalue	contributions	cumulative contributions
Y_1	2.511	35.875%	35.875%
Y_2	1.610	23.003%	58.878%
Y_3	1.165	16.639%	75.516%
Y_4	0.973	13.901%	89.417%
Y_5	0.494	7.052%	96.470%
Y_6	0.183	2.613%	99.083%
Y_7	0.064	0.917%	100.000%

Table 3: Score coefficient matrix of principal components.

Influencing Factors	Components						
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7
S1	0.134	-0.007	-0.202	0.813	0.398	-0.124	-0.817
S2	0.370	0.005	-0.007	-0.067	-0.336	-1.721	0.913
S3	0.138	0.547	0.200	-0.012	0.968	-0.132	-0.149
S4	0.366	-0.126	0.101	-0.187	-0.261	0.760	-2.653
S5	0.196	0.189	-0.625	0.101	-0.472	1.112	1.548
S6	0.213	-0.190	0.308	0.040	0.371	0.812	2.148
S7	-0.075	-0.077	0.140	0.562	-0.624	-0.033	-0.028

$$\begin{aligned} Y_1 &= 0.0134S1+0.37S2 + 0.138S3 + 0.366S4 + 0.196S5\text{-}0.213S6 - 0.075S7 \\ Y_2 &= -0.007S1 + 0.005S2+0.547S3 - 0.126S4+0.189S5 - 0.19S6 - 0.077S7 \\ Y_3 &= -0.202S1 - 0.007S2 + 0.2S3+0.101S4\text{-}0.625S5+0.308S6+0.14S7 \\ Y_4 &= 0.813S1 - 0.067S2\text{-}0.012S3\text{-}0.187S4 + 0.101S5 + 0.04S6+0.562S7 \end{aligned} \quad (12)$$

BAS Parameter Setting and Optimization

The four principal components were used as the input vector of ELM, and the output vector was the earthquake deaths. BAS parameter settings: ELM input nodes were 4, output node was 1, initial step length $\delta_0 = 9$, the ratio of step length to left and right whisker distance was $c = 3.5$, attenuation coefficient e=0.88, maximum number of iterations $T = 90$, fitness function was MSE=$\frac{1}{K}\sum_{k=1}^{K}(\eta_k - \hat{\eta}_k)^2$, through multiple tests, when the number of hidden layer nodes was 16, the model had the highest prediction accuracy. The optimization results of ELM input weights and hidden layer thresholds are shown in Table 4.

Prediction Results Analysis

The optimized parameters were substituted into ELM, and the established PCA-BAS-ELM model was used to predict the test samples. The prediction results were compared with that of the PCA-ELM model and PCA-BAS-BP model. The BP Neural Network is a multi-layer feedforward neural network trained using the backpropagation algorithm. It is highly flexible, with multiple hidden layers and varying numbers of neurons. Its structure is complex, computationally intensive, and the training process is relatively slow. Additionally, it is prone to getting trapped in local optima. The prediction results are shown in Figure 3, and the relative error comparison results are shown in Figure 4 and Table 5.

As shown in Figure 3, the predicted values of the PCA-BAS-ELM model were basically consistent with the actual values, which confirmed the effectiveness of BAS in terms of optimizing ELM and the superiority of ELM over BP. At the same time, as shown in Figure 4 and Table 5, the relative error of PCA-BAS-ELM model was lower than that of the other two models except for sample 5, which proved again that PCA-BAS-ELM model had higher prediction accuracy.

Finally, the three indicators of Equation (8), Equation (9) and Equation (10) were used to comprehensively analyze the prediction model results, and the comparison results are shown in Table 6. From Table 6, it can be seen that the mean square error (MAE), the mean absolute percentage error (MAPE) and the root mean square error (RMSE) of the PCA-BAS-ELM model were 2.433, 2.756% and 5.443, respectively, which were lower than that of the other two models, indicating that the PCA-BAS-ELM model has higher prediction accuracy. Therefore, PCA and BAS had effectively optimized the network structure of ELM, which obviously improved the prediction performance of the model.

Table 4: Optimization results of input weights and hidden layer thresholds of ELM.

Implicit Node	Input Node				Hidden layer threshold
	X_1	X_2	X_3	X_4	b_p
h_1	0.4583	-0.0304	-0.3489	-0.4946	-0.2459
h_2	0.4882	0.3256	-0.2772	-0.2059	0.4782
h_3	-0.0448	-0.4664	-0.4764	0.4442	0.0202
h_4	0.4564	-0.2423	0.4228	-0.2842	0.0393
h_5	0.4785	0.3746	0.2464	0.4464	0.4994
h_6	-0.4864	0.2028	0.0425	-0.3859	-0.4639
⋮	⋮	⋮	⋮	⋮	⋮
h_{16}	0.4843	-0.4798	-0.2935	0.4424	0.2365

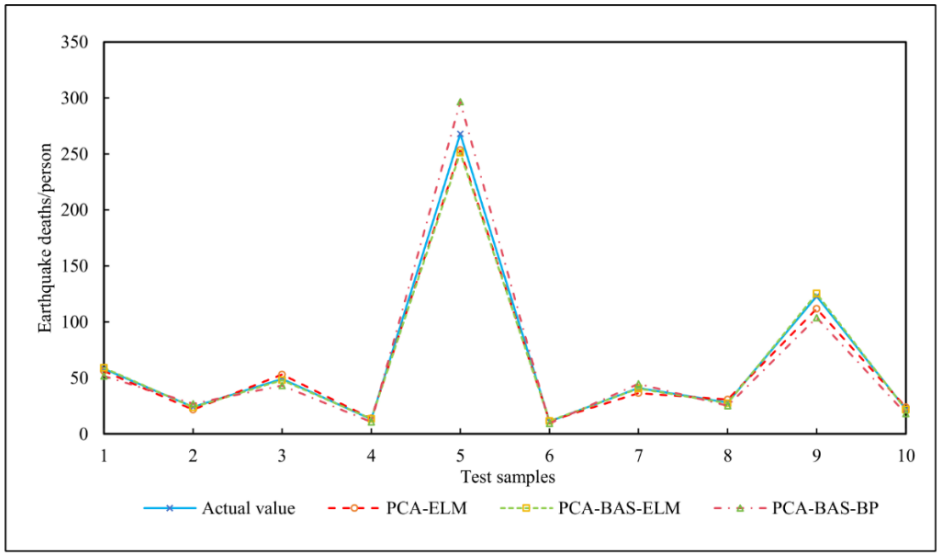

Figure 3: Comparison of prediction results.

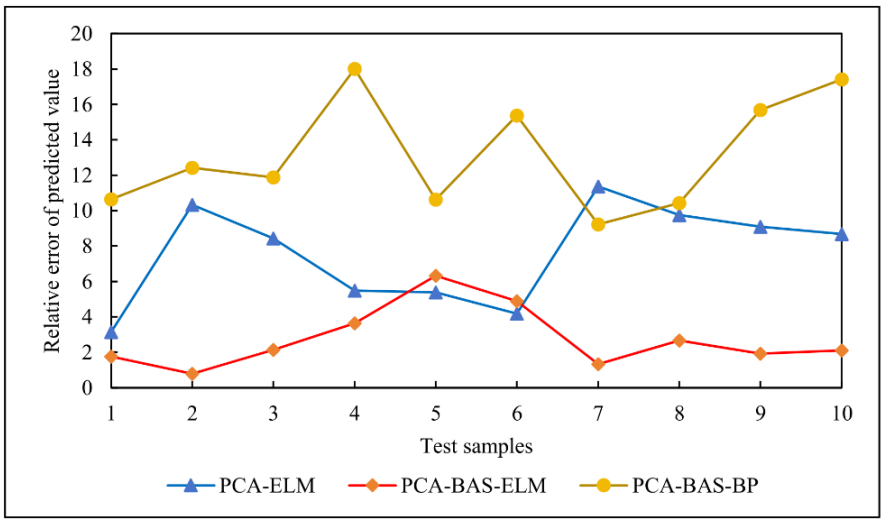

Figure 4: Comparison of relative errors.

Table 5: Comparative analysis of prediction results among three models.

Test samples	Actual value	PCA-ELM		PCA-BAS-ELM		PCA-BAS-BP	
		predicted value	Relative error/%	predicted value	Relative error/%	predicted value	Relative error/%
1	58	56.17	3.15	59.02	1.75	51.83	10.64
2	24	21.52	10.33	23.81	0.79	26.98	12.42
3	49	53.14	8.44	47.95	2.14	43.18	11.88
4	13	13.71	5.49	13.47	3.64	10.66	18.01
5	268	253.55	5.39	251.06	6.32	296.49	10.63
6	11	11.46	4.18	11.54	4.89	9.31	15.37
7	41	36.34	11.37	40.45	1.33	44.78	9.22
8	28	30.73	9.76	27.26	2.66	25.08	10.43
9	123	111.82	9.09	125.37	1.93	103.70	15.69
10	22	23.91	8.68	21.54	2.11	18.17	17.41

Table 6: Comprehensive analysis of prediction results.

Prediction model	MAE	MAPE/%	RMSE
PCA-ELM	4.455	7.588	6.276
PCA-BAS-ELM	2.433	2.756	5.443
PCA-BAS-BP	7.732	13.17	11.449

To prevent overfitting of the PCA-BAS-ELM model, the k-fold cross-validation technique was employed to calibrate and validate the model. Specifically, the 5-fold cross-validation was used, dividing the principal component dataset into five subsets. Each time, four subsets were used for training, and the remaining subset was used for testing. This process was repeated five times, each time using a different subset for testing, ensuring the model's performance across different data combinations. The MAPE of the predictions for each subset is shown in Figures 5 through 9.

Figures 5 to 9 show that when samples 1-7 were used as the test samples, the PCA-BAS-ELM model's average MAPE of the predictions was 2.98%. When samples 8-14 were used as the test samples, the average MAPE was 2.64%. For samples 15-21, the average MAPE was 3.15%; for samples 22-28, it was 2.85%; and for samples 29-35, the average MAPE was 3.42%. These results indicate that the prediction accuracy for each subset remains consistently high. This demonstrates that the PCA-BAS-ELM model, after fitting the training data, maintains good predictive performance and generalization ability on unseen data, making it effective for predicting post-earthquake deaths.

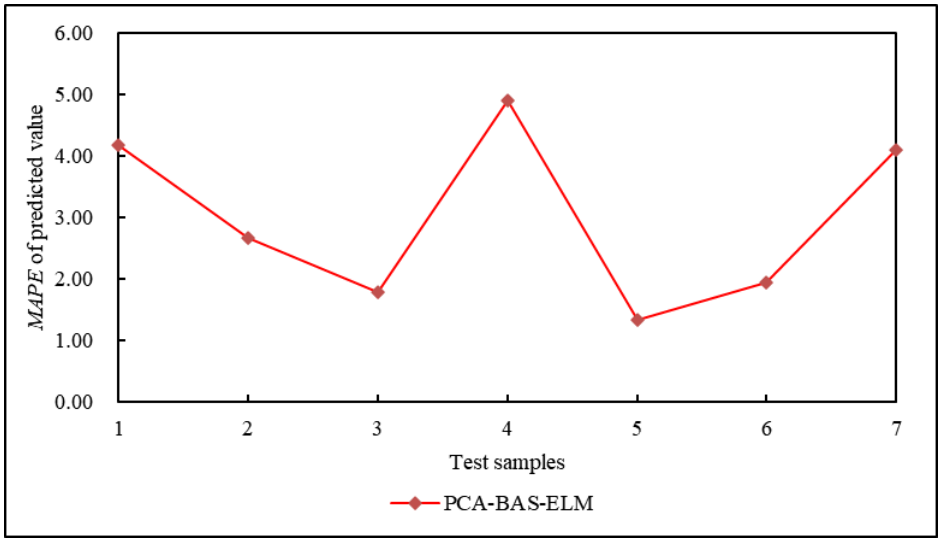

Figure 5: The MAPE value for prediction results of sample 1 to 7.

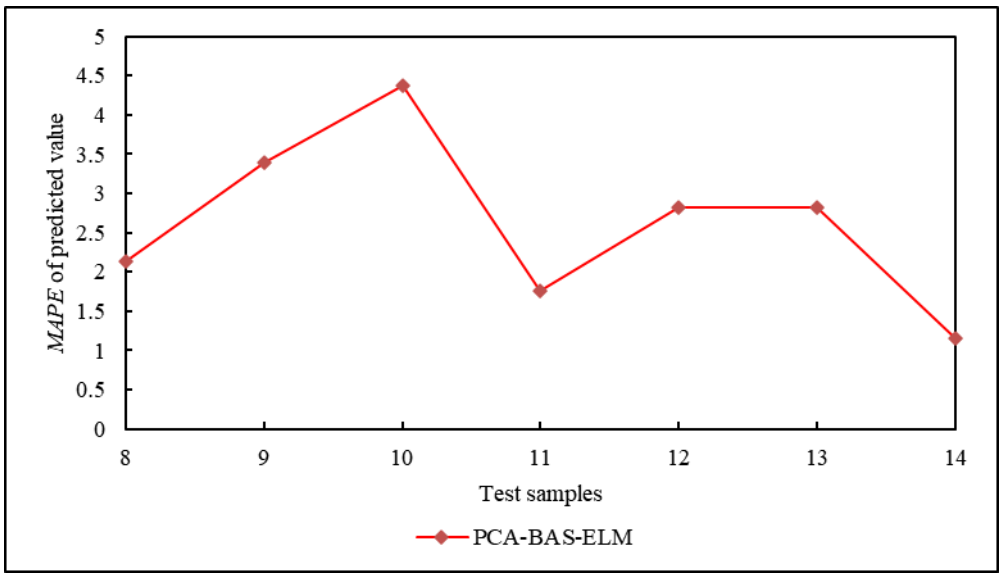

Figure 6: The MAPE value for prediction results of sample 8 to 14.

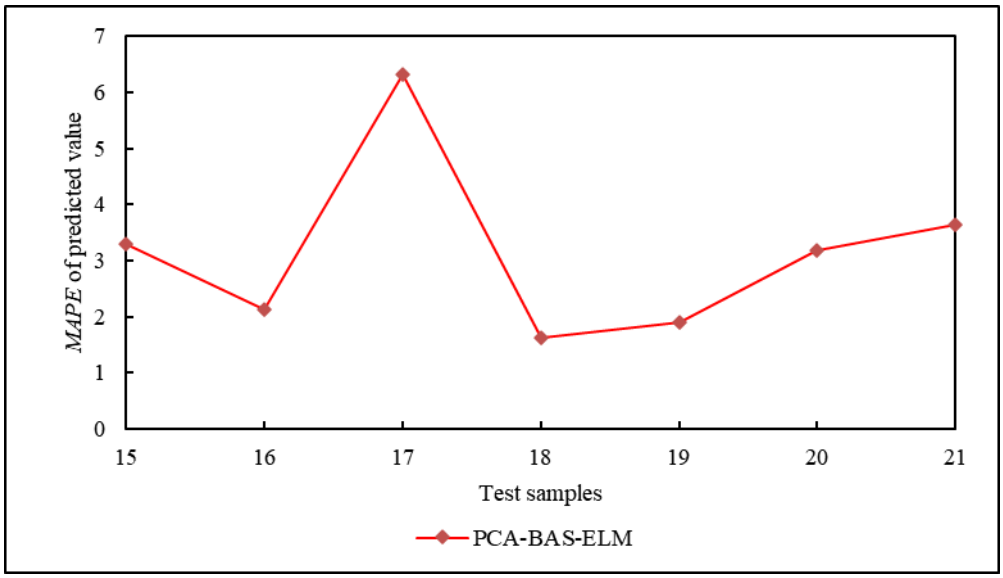

Figure 7: The MAPE value for prediction results of sample 15 to 21.

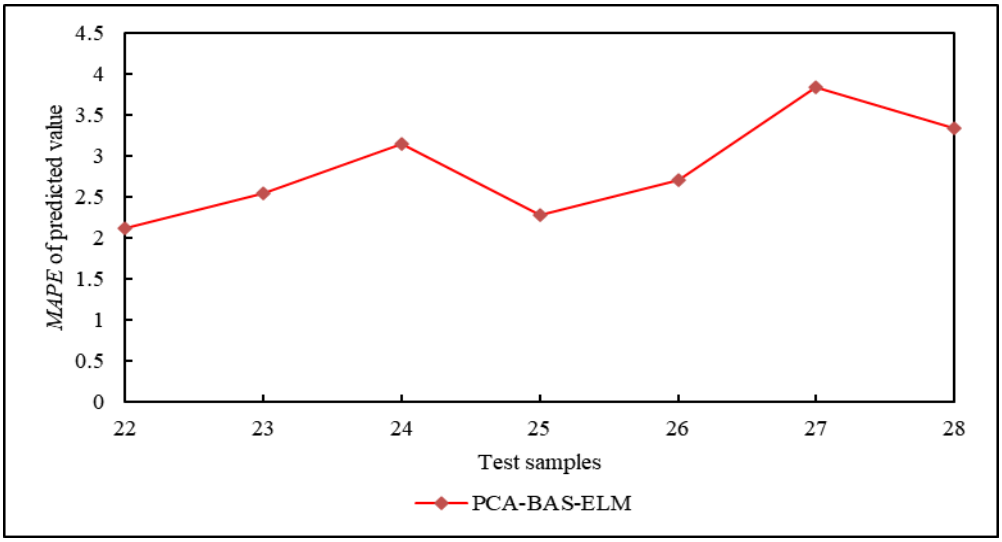

Figure 8: The MAPE value for prediction results of sample 22 to 28.

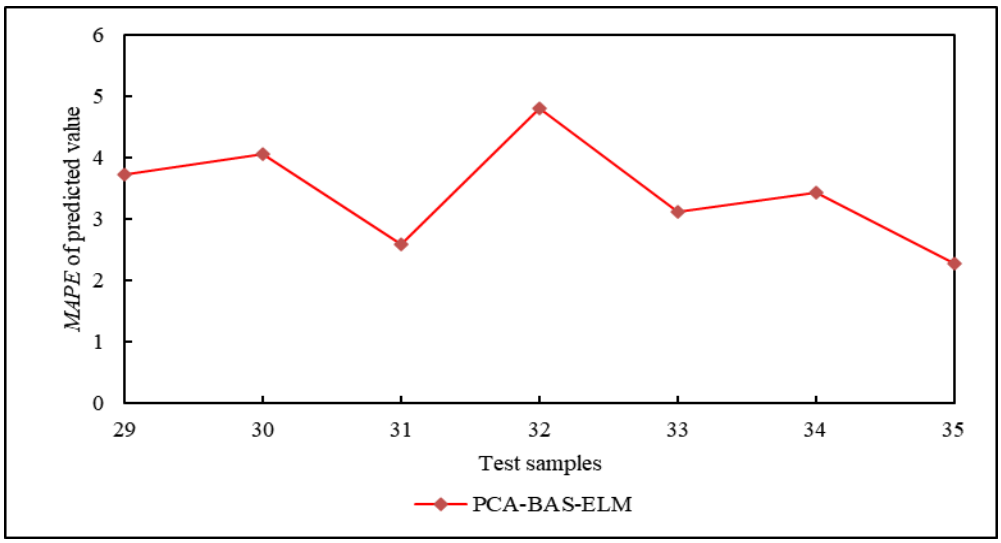

Figure 9: The MAPE value for prediction results of sample 29 to 35.

CONCLUSIONS

(1) The principal component analysis (PCA) was used to reduce the dimensionality of the original data, and the model input dimension was reduced from 7 dimensions to 4 dimensions, which improved the operating efficiency while ensuring the prediction accuracy.

(2) The beetle antennae search (BAS) algorithm was used to optimize the extreme learning machine (ELM) parameters, determine the optimal weights and thresholds of ELM, and improve the fault tolerance and optimization efficiency of the ELM.

(3) An earthquake death toll prediction model was established based on PCA-BAS-ELM. Compared with the PCA-ELM model and the PCA-BAS-BP model, the predicted value of PCA-BAS-ELM model was closer to the actual value, indicating that PCA-BAS-ELM model can effectively deal with the nonlinear problem between the earthquake deaths and the influencing factors.

(4) The mean square error (MAE), the mean absolute percentage error (MAPE) and the root mean square error (RMSE) of the PCA-BAS-ELM were 2.433, 2.756% and 5.443, respectively, which were lower than that of the other two models, indicating that the PCA-BAS-ELM model has higher prediction accuracy and can provide a basis for rapid assessment of earthquake deaths in mainland China.

(5) To further verify the reliability of the PCA-BAS-ELM model, 5-fold cross-validation method was employed for calibration and validation. The results show that the prediction accuracy for each subset remains consistently high, indicating that the PCA-BAS-ELM model, after fitting the training data, maintains good predictive performance and generalization ability on unseen data.

ACKNOWLEDGMENTS

This study was supported by Hebei Province Earthquake Science and Technology Spark Project (DZ2024052200003).

REFERENCES

1 Li Y et al. (2022). “Rapid estimation of earthquake fatalities in mainland China based on physical simulation and empirical statistics - A case study of the 2021 Yangbi earthquake”. *International Journal of Environmental Research and Public Health*, **19**(11): 6820. https://doi.org/10.5459/bnzsee.53.3.137-143

2 Fujimoto M and Nishiura H (2021). “Predicting the cumulative number of disaster deaths during the early stage of earthquakes”. *Annals Translational Medicine*, **9**(3): 241. 10.21037/atm-20-5784

3 Nichols JM and Beavers JE (2003). “Development and calibration of an earthquake fatality function”. *Earthquake Spectra*, **19**(3): 605-633. https://doi.org/10.1193/1.1596916

4 Samardjieva E and Badal J (2002). “Estimation of the expected number of casualties caused by strong earthquakes”. *Bulletin of the Seismological Society of America*, **92**(6): 2310-2322. https://doi.org/10.1785/0120010112

5 Liu JF et al. (2009). “Research on earthquake risk analysis model in mainland China (II): Life vulnerability model”. *Journal of Beijing Normal University (Natural Science Edition)*, **45**(4): 404-407. 10.3321/j.issn:0476-0301.2009.04.018

6 Liu JL and Lin JQ (2012). “Research on earthquake casualties assessment method based on epicenter intensity”. *Journal of Natural Disasters*, **21**(5): 113-119.

7 Kircher CA, Whitman RV and Holmes WT (2006). “HAZUS earthquake loss estimation methods”. *Natural Hazards Review*, **7**(2): 45-59. https://doi.org/10.1061/(ASCE)1527-6988(2006)7:2(45)

8 So E and Spence R (2013). “Estimating shaking-induced casualties and building damage for global earthquake events: a proposed modelling approach”. *Bulletin of Earthquake Engineering*, **11**(1): 347-363. https://doi.org/10.1007/s10518-012-9373-8

9 Saadat S, Camp CV and Pezeshk S (2014). “Seismic performance-based design optimization considering direct economic loss and direct social loss”. *Engineering Structures*, **76**: 193-201. https://doi.org/10.1016/j.engstruct.2014.07.008

10 Yin ZQ (1991). “Research on earthquake disaster loss prediction”. *Earthquake Engineering and Engineering Vibration*, **11**(4): 87-96.

11 Zhao ZD and Zheng XY (2006). “Numerical simulation and evaluation of casualties in the Tangshan earthquake”. *Earthquake Engineering and Engineering Vibration*, **26**(3): 28-30. 10.3969/j.issn.1000-1301.2006.03.006

12 Aghamohammadi H et al. (2013). "Seismic human loss estimation for an earthquake disaster using neural network". *International Journal of Environmental Science and Technology*, **10**(5): 931-939. https://doi.org/10.1007/s13762-013-0281-5

13 Qian FL and Cui J (2013). "Application of BP neural network model in emergency demand forecasting - taking earthquake casualties *forecasting* as an example". *Journal of China Safety Science*, **23**(4): 20-25. 10.16265/j.cnki.issn1003-3033.2013.04.017

14 Tian X and Zhu RR (2012). "Research on earthquake casualties prediction model based on principal component analysis and BP neural network analysis". *Northwestern Seismological Journal*, **34**(4): 365-368. 10.3969/j.issn 1000-0844.2012.04.0365

15 Ricciardi C et al. (2020). "Linear discriminant analysis and principal component analysis to predict coronary artery disease". *Health Informatics Journal*. **26**(3): 2181-2192. 10.1177/1460458219899210

16 Chahboun S and Maaroufi M (2021). "Principal component analysis and machine learning approaches for photovoltaic power prediction: A comparative study". *Applied Sciences*. **11**(17): 7943. https://doi.org/10.3390/app11177943

17 Yang K et al. (2021). "A novel principal component analysis integrating long short-term memory network and its application in productivity prediction of cutter suction dredgers". *Applied Sciences*. **11**(17): 8159. https://doi.org/10.3390/app11178159

18 Chen C et al. (2024). "A comprehensive survey of convergence analysis of beetle antennae search algorithm and its applications". *Artificial Intelligence Review*, **57**(6): 141. https://doi.org/10.1007/s10462-024-10789-0

19 Chen D, Li X and Li S (2023). "A novel convolutional neural network model based on beetle antennae search optimization algorithm for computerized tomography diagnosis". *IEEE Transactions on Neural Networks and Learning Systems*, **34**(3): 1418-1429. 10.1109/TNNLS.2021.3105384

20 Wang J, Lu S, Wang SH and Zhang YD (2022). "A review on extreme learning machine". *Multimedia Tools and Applications*, **81**: 41611–41660. https://doi.org/10.*1007*/s11042-021-11007-7

21 Wang G, Wong KW and Lu J (2021). "AUC-based extreme learning machines for supervised and semi-supervised imbalanced classification". *IEEE Transactions on Systems, Man and Cybernetics: Systems*, **51**(12): 7919-7930. 10.1109/TSMC.2020.2982226