## RESEARCH

**Open Access** 

# Prediction modeling of cigarette ventilation rate based on genetic algorithm backpropagation (GABP) neural network



Jiaxin Wei<sup>1</sup>, Zhengwei Wang<sup>1</sup>, Shufang Li<sup>1</sup>, Xiaoming Wang<sup>1</sup>, Huan Xu<sup>2\*</sup>, Xiushan Wang<sup>2</sup>, Sen Yao<sup>2</sup>, Weimin Song<sup>3</sup>, Youwei Wang<sup>2</sup> and Chao Mei<sup>2</sup>

\*Correspondence: xh990321@163.com

 <sup>1</sup> Xuchang Cigarette Factory of Henan Zhongyan Industry Co., Ltd, Xuchang 461001, China
 <sup>2</sup> College of Mechanical and Electrical Engineering, Henan Agricultural University, Zhengzhou 450002, China
 <sup>3</sup> Technology Center, China Tobacco Henan Industrial Co., Ltd., Zhengzhou 450000, Henan, China

## Abstract

The ventilation rate of cigarettes is an important indicator that affects the internal guality of cigarettes. When producing cigarettes, the unit may experience unstable ventilation rates, which can lead to a decrease in cigarette guality and pose certain risks to smokers. By establishing the ventilation rate prediction model, guide the design of unit parameters in advance, to achieve the goal of stabilizing unit ventilation rate, improve the stability of cigarette ventilation rate, and enhance the quality of cigarettes. This paper used multiple linear regression networks (MLR), backpropagation neural networks (BPNN), and genetic algorithm-optimized backpropagation (GABP) to construct a model for the prediction of cigarette ventilation rate. The model results indicated that the total ventilation rate was significantly positively correlated with weight (P < 0.01), circumference, hardness, filter air permeability, and open resistance. The results showed that the MLR models' (RMSE = 0.652,  $R^2$  = 0.841) and the BPNN models' (RMSE = 0.640,  $R^2$  = 0.847) prediction ability were limited. Optimization by genetic algorithm, GABP models were generated and exhibited a little better prediction performance (RMSE = 0.606,  $R^2$  = 0.873). The results indicated that the GABP model has the highest accuracy in the prediction of predicting ventilation rate and can accurately predict cigarette ventilation rate. This method can provide theoretical guidance and technical support for the stability study of the ventilation rate of the unit, improve the design and manufacturing capabilities and product quality of short cigarette products, and help to improve the quality of cigarettes.

**Keywords:** Tobacco, Multiple linear regression, BP neural network, Genetics algorithm, Ventilation rate

## **1** Introduction

The ventilation rate of cigarette is an important intrinsic indicator affecting the quality of cigarettes, and varies greatly when produced by different cigarette machines [1]. The prediction of open resistance and ventilation rate is realized based on the relationship between physical indicators and raw and auxiliary materials. Wang et al. [2] established a linear network model for cigarette airflow flow based on the series and parallel relationships between the various components of the cigarette, and calculated the total



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http:// creativecommons.org/licenses/by/4.0/.

ventilation rate of the cigarette. This study requires first dividing the structure of the produced cigarettes and then conducting modeling analysis, which has a certain lag in cigarette production.

Qu et al. [3] conducted a correlation analysis on the physical indicators of thin cigarettes and studied the key indicators that affect the open resistance of thin cigarettes, such as the ventilation rate of cigarette, cigarette weight, and cigarette hardness. Wang et al. [4] used correlation analysis and multiple regression to statistically analyze the relationship between cigarette open resistance of conventional brands and various physical indicators, such as process parameters, and cut tobacco structure. They found the mathematical model between cigarette open resistance and other indicators in order to explore the relationship between the physical indicators of cigarette and open resistance. Zhang [5] explored the impact of various physical indexes on the total ventilation rate of cigarettes and obtained key physical indicators that affect the total ventilation rate of cigarettes, such as weight, circumference, hardness, and open resistance. These studies indicate that physical indicators such as weight, circumference, open resistance, length, hardness, and filter air permeability have a significant impact on predicting cigarette ventilation rate.

Wang et al. [6] established a prior model-based method between the open resistance and ventilation of cigarettes. This model guides the production of various cigarettes, with good universality and applicability. However, the model is established under the assumption of uniform distribution of various parts of the cigarette, which is difficult to control in actual production and has certain limitations. Learning-based methods can automatically learn the implicit relationships of complex physical problems, establish models for simulation and prediction, but require sufficient data for training [7, 8]. This paper used machine learning and linear regression to predict the ventilation rate of cigarettes, namely multiple linear regression (MLR), backpropagation neural network (BPNN), and genetic algorithm-optimized backpropagation (GABP). Multiple linear regression analysis (MLR) [9] is a statistical model that uses the correlation of variables to predict the value of the dependent variables. It is the basic method for the prediction of ventilation rate. Backpropagation neural network (BPNN) systematically solves the learning problem of the connection weights of hidden unit layer in multilayer neuron networks using error backpropagation. It has intelligent information processing functions such as adaptability, self-organization, high parallelism, robustness, and fault tolerance and is especially suitable for nonlinear system modeling [10-12]. In this paper, the ventilation rate of cigarette and the physical indicators of cigarettes were used to model, and the weights and thresholds of the models were constantly adjusted to predict the ventilation rate of cigarette. Genetic algorithm (GA) [13, 14] is a random search and optimization method based on biological natural selection and genetic mechanism. Since BPNN tends to converge to a local minimum, GA is often used to find the best initial weight and threshold for BPNN optimization. GA is suitable for dealing with complex and nonlinear problems that are difficult to be solved by traditional search methods and for screening variables. It can reduce memory use and improve the predictive performance of BPNN.

This study focused on optimization and training MLR, BPNN, and GABP, and prediction of ventilation rate according to the physical indicators of cigarettes collected from Xuchang cigarette factory. The ventilation rate and parameter data sets of cigarette were divided into two groups for model training and verification. Model accuracy was evaluated by comparison of the predicted and measured values. The results indicated that the prediction model of ventilation rate suitable for multifactor influence was found, improved the processing technology and quality level of cigarettes, provided a rich theoretical basis for cigarette enterprises, and achieved the stability analysis of ventilation rate.

#### 2 Materials and methods

#### 2.1 Data sources

The cigarette samples were taken from cigarettes produced by a unit on the cigarette production line of Xuchang cigarette factory in Henan Province. The experimental data were as follows: weight, circumference, length, hardness, filter air permeability, open resistance, and total ventilation rate. A total of 900 samples were selected, of which 800 groups were training samples, and the remaining data were testing samples (Additional file 1).

#### 2.2 Data preprocessing

The normalization of the data means that the value of each feature is scaled to between 0 and 1 [15].

#### 2.3 Correlation analysis between parameters

Correlation analysis refers to the analysis of two or more correlated variable elements, in order to measure the degree of correlation between two variable factors. Statistical software SPSS was used to Pearson correlation analysis, Pearson' correlation coefficient formula was as follows: [16, 17].

$$\operatorname{COR}(X,Y) = \frac{\sum_{1}^{n} (X_{i} - \overline{X}) (Y_{i} - \overline{Y})}{\sqrt{\sum_{1}^{n} (X_{i} - \overline{X})^{2} \sum_{1}^{n} (Y_{i} - \overline{Y})^{2}}}$$
(1)

#### 2.4 Principal component analysis

PCA is a data reduction technique that creates principal components (PC), which are linear combinations of the original variables, and create new, uncorrelated variables [18]. In order to improve predictive performance and efficiency, PCA dimensionality reduced principal components were chosen for the experiment.

#### 3 Model building

#### 3.1 Multiple linear regression model

Multiple linear regression was mainly used to reflect the mathematical relationship between multiple independent variables and an independent variable, and can be used to predict the change of a variable. The multivariate linear model between the total ventilation rate of cigarettes and the circumference, weight, length, filter air permeability, open resistance, and hardness was established as follows:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 \tag{2}$$

In formula (2),  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ ,  $\beta_3$ ,  $\beta_4$ ,  $\beta_5$ , and  $\beta_6$  are the overall regression coefficients, *Y* is the total ventilation rate of the cigarette,  $x_1$  is the circumference,  $x_2$  is the weight, and  $x_3$  is the length,  $x_4$  is the filter air permeability,  $x_5$  is the open resistance,  $x_6$  is the hardness.

#### 3.2 BP neural network model design

BPNN is widely used in signal processing, pattern recognition, machine control, and many other fields [19]. In this study, BPNN was used as the classifier model. This study adopted a single hidden layer BPNN model, and its structure is shown in Fig. 1.

The main steps of the proposed BPNN classification method were as follows:

- (1) Enter the training sample, normalize the feature value of the input sample to the range [0,1];
- (2) Network initialization.

The six nodes of the input layer correspond to six eigenvectors (weight, circumference, length, hardness, filter air permeability, and open resistance). One output layer node is total ventilation rate of cigarette. The number of hidden layer nodes is related to the number of neurons in the input and output layers. The quantity shall be selected according to the design experience and experiment. Here, we used the following empirical formula for calculation [20]:



Fig. 1 Structure of BPNN

$$n1 = \sqrt{n+m} + a \tag{3}$$

where n1 is the number of neurons in the hidden layer, n is the number of neurons in the output layer, m is the number of neurons in the input layer, and a is a constant that ranges between 1 and 10. Nine neurons were selected in the hidden layer. The neural network model was transferred from the input layer to the hidden layer by tansig function and from the hidden layer to the output layer by purelin function [21]. This study used the combination of the above transfer functions, and the expressions of the two transfer functions were shown in equations (4) and (5). Then, the maximum number of iterations for model training was set to 2000, the inertia coefficient was set to 0.8, the maximum allowable error was set to  $1e^{-6}$ , and the learning efficiency was set to 0.01. The initial parameters of the backpropagation neural networks are shown in Table 1.

tansig: 
$$f(x) = \frac{1 - e^{-x}}{1 + e^{-x}}$$
 (4)

$$purelin: f(x) = x \tag{5}$$

(3)During the training process of BPNN, the weights and thresholds were constantly adjusted until the final result was obtained. Then, the trained classifier was used to recognize the test samples.

#### 3.3 Based on GABP neural network model

The classic BP neural network algorithm is based on the gradient descent method to achieve the purpose of optimization, which is prone to converge to the local minimum [22]. After the standard BP neural network was applied to the training set, it was difficult to distinguish the local extreme points from the global extreme points, and the error was large when correcting the weight and threshold. Therefore, the genetic algorithm was selected to optimize BP neural network. The GABP flowchart is shown in Fig. 2.

- (1) Enter the training sample and normalize the feature value of the input sample to the range [0,1];
- (2) Network initialization, based on BPNN, set the optimization parameters, adjusted the parameters continuously through practice, selected the population size as 200,

Parameters	Value
Neurons at the input layer	6
Neurons at the hidden layer	9
Neurons at the output layer	1
Learning rate	0.01
Total epochs	2000 (best = 3)
Activation	Tansig, purelin
Training function	Gradient descent

Table 1	Parameters of the BP model



Fig. 2 Flowchart of GABP

Parameters	Value
Neurons at the input layer	6
Neurons at the hidden layer	9
Neurons at the output layer	1
Learning rate	0.01
Total epochs	2000 (best=4)
Activation	Tansig, purelin
Training function	Gradient descent
Population size	200
Number of evolutions	500
Selection factor	0.09
Crossover factor	0.4
Mutation factor	0.001

set the population algebra as 500, coded the generated population, and performed population selection, genetic crossover, and mutation operations on it. The selection factor was set to 0.09, the crossover factor was set to 0.4, and the variation factor was set to 0.001. The initial parameters of the GABP are shown in Table 2.

(3) In the process of GABP training, the algorithm was continuously optimized to obtain the optimal parameters, and the model was trained.

#### 3.4 Assessment of model performance

Two statistical parameters were used to evaluate the training and prediction performance of the neural network model, which were root-mean-square error (RMSE) and determination coefficient ( $R^2$ ) [23]. Their definitions are shown in Formula (6), (7), respectively.

RMSE = 
$$\sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2}$$
 (6)

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (y_{i} - f(x_{i}))^{2}}{\sum_{i=1}^{n} (y_{i} - \bar{y})^{2}}$$
(7)

where *n* was the number of samples,  $y_i$  was the experimental measured value,  $f(x_i)$  was the prediction value calculated by the model, and  $\overline{y}$  was the average of the experimental measurements. The closer RMSE was to zero and the closer  $R^2$  was to 1, the better the model fitting effect was, and the more accurate the prediction was.

## 4 Results and discussion

#### 4.1 Correlation analysis

It can be seen from Table 3 that there were different degrees of correlation among the factors affecting the ventilation rate of cigarettes. The weight had an extremely significant positive correlation (p < 0.01) with the circumference, filter air permeability, hardness, and open

Index	Weight (g)	Circumference (mm)	Length (mm)	Hardness (mm)	Filter air permeability (%)	Open resistance (Pa)	Total ventilation rate (%)
Weight	1.000						
Circumfer- ence	0.417**	1.000					
Length	0.011	- 0.072*	1.000				
Hardness	0.498**	0.162**	0.073*	1.000			
Filter air perme- ability	0.268**	0.102**	0.041	0.184**	1.000		
Open resistance	0.519**	0.224**	0.073*	0.325**	- 0.048	1.000	
Total ventilation rate	0.565**	0.265**	0.073*	0.342**	0.838**	0.25**	1.000

Tab	0.3	Corro	lation	anal	veic
lap	ie s	Corre	iation	anar	VSIS

\* and \*\*, respectively, represent a significant correlation at the p < 0.05 level and an extremely significant correlation at the p < 0.01 level

resistance. The circumference had an extremely significant positive correlation with the hardness, filter air permeability, and open resistance and a significant negative correlation with the length. The length had a significant positive correlation (p < 0.05) with the hardness and open resistance, and the hardness had a very significant positive correlation with the filter air permeability, and open resistance; the total ventilation rate had an extremely significant positive correlation with weight, circumference, hardness, filter air permeability and open resistance, but the correlation with length was low. Therefore, in the process of cigarette production, using the correlation between these indicators can provide theoretical guidance and technical support for the stability study of the ventilation rate of the unit.

#### 4.2 Model analysis

### 4.2.1 Multiple linear regression

Using SPSS for comprehensive calculation, the basic formula of the obtained multiple linear regression model is:

$$Y = -26.336 + 1.208x_1 + 18.339x_2 + 0.965x_4 + 5.226x_5 + 0.014x_6 \tag{8}$$

Table 4 shows that the coefficient of determination  $R^2$  of the model was 0.841, which indicated that 84% of the total ventilation rate of the dependent variable can be explained in the selected variables; the Durbin Watson value was 1.899, close to 2, explained that the autocorrelation between the variables of the equation can be accepted; the results of variance analysis (ANOVA), the *p* value (significance) of the model was 0.000, representing *p* < 0.001, which proved that the model contained at least one independent variable that had a significant impact on the dependent variable *y*. There was a linear relationship between the dependent variable and the independent variable, so the model significance relationship of the model was tenable.

#### 4.2.2 Backpropagation neural networks

The BP neural network with the parameters in Table 1 is used to train, and the following results are obtained. Figure 3 plots the variation of MSE with the training period, and the best neural network performance is achieved when the period is 3. Figure 4 shows the regression diagram of training sets, verification sets, test sets, and overall data. The regression correlation coefficient, R, of all sets is above 0.88, and most of the predicted results are close to the training set.

#### 4.2.3 Genetic algorithm-optimized backpropagation

The GABP model with the parameters in Table 2 is used to train, and the following results are obtained. Figure 5 shows the changes in MSE with the training epochs. When the epoch is 4, the GABP performance reaches its peak. Figure 6 shows the regression diagram of training sets, verification sets, test sets, and overall data. The regression correlation coefficient, *R*, of all sets is above 0.91, indicating that the GABP has good predictive performance.

Model	R	R <sup>2</sup>	Adjusted R <sup>2</sup>	RMSE	Significance	Durbin Watson
MLR	0.918	0.842	0.841	0.652	0.000	1.899

nary





Fig. 4 BP linear regressions of training sets, validation sets, test sets, and the overall data

### 4.3 Predicted and actual values under different models

Figure 7 shows the dispersion of predicted and actual values for each model. Compared with MLR (Fig. 7a), BP (Fig. 7b), and GABP (Fig. 7c), the results are similar, but GABP is more stable than MLR and BP, indicating that all data points between predicted and actual values are widely distributed without overestimation or underestimation.





Fig. 6 GABP linear regressions of training sets, validation sets, test sets, and the overall data

## 4.4 The cumulative variance interpretation rate and R<sup>2</sup> corresponding to different numbers of principal components

As shown in Fig. 8, with the increase in the number of principal components, the cumulative variance interpretation rate will gradually increase,  $R^2$  will first increase, then tend to plateau, and finally slowly rise. In this paper, the number of principal components selected through experiments is 2, and  $R^2$  reaches 0.792, which is



Fig. 7 Comparison between predicted values and experimental values of different algorithms. a MLR, b BPNN, c GABP



**Fig. 8** The cumulative variance interpretation rate and  $R^2$  corresponding to different numbers of principal components

basically the same as when the principal components are 3 and 4. When the principal component is 5,  $R^2$  will increase to 0.863, which is similar to the results without PCA.

## 4.5 Model comparison

In this paper, multiple linear regression, BP neural network, and GABP neural network were used to simulate and predict the ventilation rate of cigarettes. Based on the results in Table 5, it can be concluded that:

(1) Compared with the multiple linear regression model, the RMSE of the BPNN decreased by 1.7%, and the  $R^2$  increased by 0.7%, which showed that the BPNN model had little improvement in the prediction of cigarette ventilation rate compared with the multiple linear regression model, probably due to the tendency of BPNN to fall into local minima.

Model	Evaluation index		
	RMSE	R <sup>2</sup>	
MLR	0.652	0.841	
BPNN	0.640	0.847	
GABP	0.606	0.873	
PCA-GABP	0.823	0.793	

 Table 5
 Model evaluation index

- (2) After the genetic optimization of the BPNN model, compared with the multiple linear regression model, RMSE decreased by 6.9%, and R<sup>2</sup> increased by 3.8%, which showed that the GABP model had a little improvement in the prediction of cigarette ventilation rate compared with the multiple linear regression model. Compared to BPNN, there is also a certain improvement, this implies that GA has been successfully used to find the optimal weights and thresholds to generate better performing GABP models.
- (3) After PCA dimensionality reduction, when the principal component was 2, the RMSE of GABP increased by 35.8% and the  $R^2$  decreased by 10.1%, which is not as good as GABP without principal component analysis in terms of predictive performance.

#### **5** Conclusion

- (1) In this paper, the data of cigarettes from Xuchang Cigarette Factory were used to establish a model and test the ventilation rate of cigarettes. The test results showed that there was a correlation between the indicators that affect the ventilation rate of cigarettes and the total ventilation rate: the total ventilation rate had a very significant positive correlation with the weight, circumference, hardness, filter air permeability, and open resistance, but had a low degree of correlation with the length.
- (2) The RMSE and  $R^2$  of multiple linear regression, BP neural network model, and GABP neural network model were compared. The results showed that there were differences among the three models in predicting cigarette ventilation rate: the effect of genetic optimization BP neural network model was a little better than the BP neural network and multiple linear regression, these improvements can make a certain contribution to the stability of cigarette ventilation rate, which is beneficial for improving the quality of cigarettes.
- (3) After using PCA dimensionality reduction technology, when the number of principal components was 2, the running time was reduced and the efficiency was indeed improved, but the predictive performance decreased. The number of features may be crucial, and using PCA dimensionality reduction will lose some important information, which is not in line with the expectation of improving predictive performance. Therefore, GABP without PCA dimensionality reduction was selected.

(4) In the future, the genetically optimized BP neural network can be optimized, or other relevant factors affecting the cigarette ventilation rate can be added to improve the prediction accuracy of the cigarette ventilation rate model.

### **Supplementary Information**

The online version contains supplementary material available at https://doi.org/10.1186/s13634-024-01119-1.

Additional file 1: 900 sets of cigarette data including weight, circumference, length, hardness, filter air permeability, open resistance, and total ventilation rate.

#### Acknowledgements

This research is supported by China Tobacco Henan Industrial Co., Ltd. (No. A2020040, No. 202102110120, No. AW2022025).

#### Author contributions

XH designed the work, analyzed and interpreted the data, and drafted the manuscript. WJX, WZW, and WXM participated in the design of the study, performed the experiments and analysis, and helped to draft the manuscript. WXS and YS contributed to literature investigation. SWM, WYW, and MC contributed to revise the manuscript. All the authors read and approved the final manuscript. LSF provided great assistance in the paper revision process.

#### Availability of data and materials

We tested the performance of the proposed method based on cigarettes produced by a unit on the cigarette production line of Xuchang cigarette factory in Henan Province.

#### Declarations

**Ethics approval and consent to participate** Not applicable.

**Consent for publication** Not applicable.

#### **Competing interests**

The authors declare no competing interests.

Received: 23 August 2023 Accepted: 30 January 2024 Published online: 22 February 2024

#### References

- H. Dong, Y. Zhang, J. Wang et al., Design and application of quality monitoring samples for main physical indicators of cigarettes. Tob. Sci. Technol. 55(07), 83–89+112 (2022)
- L. Wang, M. You, X. Cui et al., Prediction method of cigarette suction resistance and ventilation characteristics based on linear network model. Tob. Technol. 50(12), 85–89 (2017)
- X. Qu, J. Zhang, S. Jiang et al., Correlation analysis of physical indicators of thin cigarettes. J. Hunan Univ. Arts Sci. (Natl. Sci. Edit.) 32(04), 64–68 (2020)
- H. Wang, H. Zhao, C. Zhao et al., Mathematical model and application of cigarette smoking resistance based on multiple regression. Food Ind. 43(05), 60–64 (2022)
- Z. Zhang, Explore the impact of various physical indicators on the total ventilation rate of cigarettes. Light Ind. Sci. Technol. 35(09), 117–119 (2019)
- 6. L. Wang, G. Deng, Z. Wu et al., The relationship between bite-by-mouth dynamic absorption resistance and dynamic ventilation rate of cigarettes based on linear network model. Tob. Technol. **55**(5), 66–72 (2022)
- R.A.A. Ramadhan, Y.R.J. Heatubun, S.F. Tan et al., Comparison of physical and machine learning models for estimating solar irradiance and photovoltaic power. Renew. Energy 178, 1006–1019 (2021)
- 8. H. Park, D.Y. Park, Comparative analysis on predictability of natural ventilation rate based on machine learning algorithms. Build. Environ. **195**, 107744 (2021)
- 9. B. Lin, Multiple linear regression analysis and its application. China Sci. Technol. Inf. 09, 60–61 (2010)
- H. Fei, J. Zhou, R. Yang et al., Prediction model of potassium chloride ratio of flue-cured tobacco based on particle swarm optimization BP neural network. Tob. Technol. 47(6), 49–53 (2014)
- F. He, L. Zhang, Prediction model of end-point phosphorus content in BOF steelmaking process based on PCA and BP neural network. J. Process. Control. 66, 51–58 (2018)
- Y. Wu, A. Li, S. Lei et al., Prediction of pyrolysis product yield of medical waste based on BP neural network. Process. Saf. Environ. Prot. 176, 653–661 (2023)
- 13. Y. Xi, T. Chai, W. Yun, Overview of genetic algorithms. Control Theory Appl. 06, 697–708 (1996)

- K. Liu, T. Lin, T. Zhong et al., New methods based on a genetic algorithm back propagation (GABP) neural network and general regression neural network (GRNN) for predicting the occurrence of trihalomethanes in tap water. Sci. Total. Environ. 870, 161976 (2023)
- 15. R. Chen, J. Song, M. Xu et al., Prediction of the corrosion depth of oil well cement corroded by carbon dioxide using GA-BP neural network. Constr. Build. Mater. **394**, 132127 (2023)
- J. Zhang, Analysis of Influencing Factors of Ship Collision Damage Based on Logistic Regression Model (Dalian Maritime University, 2012)
- W. Huang, M. Li, Analysis of main factors affecting the development of tourism areas based on SPSS data analysis. Software 40(01), 144–149 (2019)
- R. Hayati, A.A. Munawar, E. Lukitaningsih et al., Combination of PCA with LDA and SVM classifiers: a model for determining the geographical origin of coconut in the coastal plantation, Aceh Province, Indonesia. Case Stud. Chem. Environ. Eng. 9, 100552 (2024)
- J. Zhu, A. Wu, X. Wang et al., Identification of grape diseases using image analysis and BP neural networks. Multimed. Tools Appl. 79, 14539–14551 (2020)
- L.L. Zhao, M. Zhang, H.X. Wang et al., Monitoring of free fatty acid content in mixed frying oils by means of LF NMR and NIR combined with BP-ANN. Food Control 133, 108599 (2022)
- 21. Z.Z. Hu, Y. Yuan, X. Li et al., Yield prediction of "Thermal-dissolution based carbon enrichment" treatment on biomass wastes through coupled model of artificial neural network and AdaBoost. Bioresour. Technol. **343**, 126083 (2022)
- 22. Y. Lei, Research on the Application of BP Neural Network Optimized by Genetic Algorithm in Multimode Integrated Forecasting (Nanjing University of Information Engineering, 2018)
- H.O. Kargbo, J. Zhang, A.N. Phan, Optimisation of two-stage biomass gasification for hydrogen production via artificial neural network. Appl. Energy 302, 117567 (2021)

#### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.