

Research Article

Evaluation of Provincial Economic Resilience in China Based on the TOPSIS-XGBoost-SHAP Model

Zhan Wu 

School of Economics and Management, Shanghai Ocean University, Shanghai 201306, China

Correspondence should be addressed to Zhan Wu; m220751363@st.shou.edu.cn

Received 14 June 2023; Revised 27 September 2023; Accepted 4 October 2023; Published 12 October 2023

Academic Editor: Ding-Xuan Zhou

Copyright © 2023 Zhan Wu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The aim of this research is to propose a framework for measuring and analysing China's economic resilience based on the XGBoost machine learning algorithm, using Bayesian optimization (BO) algorithm, extreme gradient-boosting (XGBoost) algorithm, and TOPSIS method to measure China's economic resilience from 2007 to 2021. The nonlinear effects of its key drivers are also analysed in conjunction with the SHAP explainable model to explore the path of China's economic resilience enhancement. The results show that the level of China's economic resilience is improving, but the overall level is low; R&D expenditure and the number of patents granted are important factors affecting China's economic resilience with a significant positive relationship. The BO-XGBoost model outperforms the benchmark machine learning algorithm and can provide stable technical support and scientific decision-making basis for China's economic resilience measurement analysis and high-quality economic development.

1. Introduction

With the rapid advancement of globalisation and informatization, the domestic and international environments are becoming increasingly complex, and China's economic development is facing increasing external risks [1]. In recent years, economic resilience, as a new concept that emphasises the resistance, resilience, and evolution of the economic system in response to external shocks, is the most powerful support for risk prevention [2]. Therefore, analysing the nonlinear effects of key factors that affect China's economic resilience is of great significance to China to enhance the risk absorption capacity of the economic system and promote long-term stable and high-quality economic development.

At present, relevant studies focus mainly on two aspects of economic resilience measurement and analysis of influencing factors. In terms of economic resilience influencing factors, Wang and Wei [3] pointed out that factors such as human capital, trade openness, and entrepreneurship can promote economic resilience. Jiang et al. [4] population agglomeration can enhance the resilience of cities to economic crises, and is more conducive to improving the economic recovery and adjustment capacity of cities, and has

a positive spatial spillover effect on neighbouring cities. Wang et al. [5] found that innovation and entrepreneurship dynamics have a significant positive effect on economic resilience. In terms of economic measurement, many scholars comprehensively evaluate economic resilience from multidimensional and multiattribute indicators. Briguglio et al. [6] evaluated economic resilience based on four dimensions: market efficiency, economic stability, social development, and political system. Tan et al. [7] used principal component analysis to measure the level of economic resilience in 19 resource-based cities in northeast China. The study found that forests-based cities improved the most in terms of economic resilience. To overcome the shortcomings of subjective and objective weights in traditional models, Xun and Yuan [8] chose the evaluation model with good comprehensive performance of intuitionistic fuzzy set theory and TOPSIS method to evaluate the economic resilience of Dalian City, China.

However, the above-mentioned evaluation methods such as principal component analysis, hierarchical analysis, and TOPSIS model cannot meet the requirements of non-normal and nonlinear processing of high-dimensional data [9]. In recent years, with the continuous development of

artificial intelligence technology, more scholars have used BP neural networks, support vector machines (SVMs), and other machine learning methods for the evaluation of complex problems [10–12]. Compared to traditional machine learning models, extreme gradient boosting (XGBoost) is an integrated gradient-boosting learning algorithm that depicts the underlying mechanism between input characteristics and target outcomes and has the advantages of high prediction accuracy and less overfitting [13]. However, there is little literature on applying the XGBoost algorithm and the SHAP interpretable framework to economic resilience in China.

Therefore, this study attempts to introduce the XGBoost machine learning algorithm into the study of China's economic resilience measurement and fuse it with the Bayesian optimization (BO) algorithm to optimize the hyperparameters. The evaluation index system of China's economic resilience is constructed from three levels of reactivity, adaptive capacity, and recovery capacity, respectively, and the evaluated values of the TOPSIS model are used as the target values of the XGBoost regression algorithm for training and testing, and combined with the SHAP explanatory framework, to accurately excavate the key variables of economic resilience, and to analyze the non-linear effects of the factors affecting the resilience of China's economy. It can provide effective technical support and scientific decision-making basis for the analysis of China's economic toughness measurement and high-quality economic development.

2. Indicator System and Data Sources

2.1. Construction of the China's Economic Resilience Indicator System. The evaluation index system for economic resilience in existing studies has considered multiple dimensions, including environmental, economic, and social dimensions [14, 15], but so far a standard evaluation system has not been formed in academia. Based on the actual situation and existing studies [2, 7, 16], a resilience evaluation index system for China is constructed from three dimensions: resilience, recovery, and evolutionary capacity, as shown in Table 1.

Resistance indicates the ability of an economy to maintain stable operation, reduce losses, and avoid recession when it is subjected to external shocks in the development process. The stronger the resistance, the less likely an economy will be affected by a shock, and the strength of the resistance depends on the economy's own conditional endowment. Conditional endowments have automatic stabilisation mechanisms, and a well-endowed economy can spontaneously resist the impact of external shocks. The registered urban unemployment rate is an important indicator of urban employment and improvement of people's livelihood; the value added of the secondary industry reflects the innovation capacity and productivity of industrial enterprises. The natural growth rate of resident population affects the country's demographic structure, labour supply, social security, and resources and environment and is an important basis for formulating national development

strategies and policies. Education expenditure is a strategic investment to modernise education and enhance national competitiveness and innovation capacity, and the size of the economy is a strategic investment to achieve the modernisation of education. Education expenditure is a strategic investment to modernise education and improve national competitiveness and innovation capacity. Profits from industrial enterprises above the scale reflect the quality and efficiency of industrial development. The above indicators are a good reflection of China's economic condition endowment.

Resilience indicates that the economy can bounce back to its original state or close to it quickly, which means that the economy is able to adapt to external shocks and return to normal operation quickly after the shocks have passed. When the shock is over, government departments must flexibly adjust economic policies to mitigate the shock and promote economic recovery. Therefore, indicators are selected to fully reflect the country's economic size, growth, social stability, and innovation inputs. Among them, GDP per capita reflects the production capacity, consumption capacity, investment capacity, and international competitiveness of a country or region; expenditure on science and technology is an important guarantee to enhance national competitiveness and promote economic and social development. The total import and export is an important basis for studying balance of payments, economic growth, international competitiveness, and economic structure. The technology market turnover reflects the total supply and transformation efficiency of scientific and technological achievements. Total retail sales of consumer goods are an effective tool to measure the operation of national economy and the contribution of consumption to economic growth. Per capita consumption expenditure of urban residents reflects the level of urban economic development and the contribution of consumption to economic growth.

The evolutionary force indicates the ability of an economy to achieve economic structure optimization, development model transformation, and growth momentum conversion through its own adjustment and innovation after suffering from external shocks, so as to maintain healthy, sustainable, and high-quality economic development. The stronger the evolutionary force, the more it can take the initiative to update and adjust the original economic structure and choose new development methods and paths, thus enhancing economic resilience. The selected indicators need to reflect the improvement of people's livelihood, the ability to absorb foreign investment, infrastructure construction, and independent innovation. The accumulated balance of the basic pension insurance fund for urban workers reflects the financial status and sustainability of the basic pension insurance system for urban workers. The total investment in foreign-invested enterprises reflects the level of China's opening up to the outside world and its attractiveness. R&D expenditure measures the scale, structure, and efficiency of investment in R&D activities. The total investment in infrastructure construction is a reflection of the investment in transportation, energy, water conservancy, municipal, information, and others. The number of

TABLE 1: China's economic resilience evaluation index system.

Target layer	Dimensional layers	Indicator layers	Units	Properties
Resistance		Urban registered unemployment rate	%	-
		Secondary industry value added	Billion yuan	+
		Natural growth rate of resident population	%	+
		Education expenses	Billion yuan	+
China's economic resilience		Total profit of industrial enterprises above the scale	Billion yuan	+
		GDP per capita	Yuan/person	+
		Science and technology expenditures	Billion yuan	+
		Total import and export	Million dollars	+
		Technology market turnover	Million yuan	+
		Total retail sales of social consumer goods	Billion yuan	+
Evolutionary power		Per capita consumption expenditure of urban residents	Yuan/person	+
		The accumulated balance of the basic pension insurance fund for urban employees	Billion yuan	+
		Total investment in foreign-invested enterprises	Million dollars	+
		R&D expenditure	Billion yuan	+
		Total investment in infrastructure construction	Billion yuan	+
		Number of invention patent applications by domestic applicants	Individual	+

invention patents by domestic applicants reflects the level and competitiveness of domestic scientific and technological innovation.

2.2. Data Sources. All the data were collected from China's National Bureau of Statistics (2007–2021) and China Statistical Yearbook (2007–2021). Considering the availability of the data, the data sample in this paper involves 30 provinces excluding Tibet, Hong Kong, Macao, and Taiwan, and a complete and valid sample of 450 groups is obtained after data collation.

3. Construction and Rationale of China's Economic Resilience Assessment Model

3.1. Principle of the Entropy-Weighted TOPSIS Model. The TOPSIS model is called the distance method of superior and inferior solutions, which is applicable to multiple indicators and multiple solutions for comparison to make the best decision [17]. The advantage of the TOPSIS model is that it does not have strict requirements for the sample and is universally applicable, so it is often used in various evaluation works [18].

3.2. Principle of the XGBoost Algorithm. The extreme gradient boosting (XGBoost) algorithm is an integrated learning algorithm developed by Tianqi Chen et al. in recent years. It is an efficient implementation of gradient-boosting decision tree (GBDT). A strong classifier is constructed by integrating multiple weak classifiers, a second-order Taylor expansion of the loss function [19], while using a regular term to prevent overfitting of the model [20], and training with the objective function. The objective function of the algorithm is as follows [21]:

$$\text{Obj}(\theta) = \sum_{i=1}^M l(y_i, \hat{y}_i) + \sum_{j=1}^t \Omega(f_N). \quad (1)$$

In equation (1), Obj denotes the objective function, θ denotes the model parameters, M denotes the number of samples, $l(y_i, \hat{y}_i)$ denotes the loss function between the predicted and true values, t denotes the number of trees, f_N denotes N trees, and $\Omega(f_N)$ denotes the regularization term. XGBoost solves for the best model parameters by optimizing the objective function to obtain the best prediction results. The loss function represents the prediction effect of the model.

Regularization terms help to control the complexity of the model and avoid overfitting the model. The use of strongly convex regularization terms makes the model smoother and reduces the likelihood of overfitting. A common strongly convex regularization is L2 regularization (also known as weight decay), which is implemented by applying a squared penalty to the weights [22]. Another alternative is the entropy regularization term. This regularization term is based on the concept of information entropy, which bounds the model complexity by relating it to the information entropy. In this setting, the model

complexity is not only related to the size of the model weights but also to the distribution of the model weights [23]. In XGBoost, L1 and L2 regularization terms are two important methods used in XGBoost to control model complexity. L1 regularization generates sparse models for feature selection, while L2 regularization constrains model complexity to prevent overfitting.

3.3. SHAP Explanatory Model. Lundberg and Lee proposed the SHAP model to explain various machine learning algorithms in 2017 [24]. The SHAP value originated from game theory and is mainly used to quantify the contribution of each feature to the model prediction by calculating the marginal contribution of a feature when it is added to the model, and then by calculating the different marginal contributions of this feature in all the feature sequences, and then by finally calculating the SHAP value. The related calculation formula is shown as follows [25]:

$$y_i = y_{\text{base}} + f(x_{i1}) + f(x_{i2}) + \dots + f(x_{ip}), \quad (2)$$

where y_{base} is the mean value of the target variable over all samples and $f(x_{ij})$ is the SHAP value of x_{ij} . The advantage of the SHAP value is that it reflects the contribution of the features in each sample and indicates the positivity or negativity of the effect [26]. This allows for a better understanding of the prediction results for each sample and provides insight into the extent to which each feature influences the prediction results. In this paper, SHAP is utilized to provide explanations for the prediction model.

3.4. TOPSIS-BO-XGBoost Evaluation Model. Based on the evaluation results of the TOPSIS model, the comprehensive evaluation values are input into various types of machine learning models, as previous samples for training and testing. During the training process, the Bayesian optimization (BO) algorithm is introduced to optimize the hyperparameters. This combined TOPSIS and machine learning model is applied to the analysis of China's economic resilience measurement. In this paper, the machine learning algorithm environment is configured as follows: Alienware x17 R2 Windows 11 64-bit operating system; GPU model is NVIDIA A100 Tensor Core GPU 40G; Python configuration environment is TensorFlow 2.8.0, Pytorch 1.11.0, Python 3.9, and CUDA11.6; third-party libraries: sklearn machine learning library, seaborn library, and Bayesian-optimization library. In this paper, we use Python software to first call the Python API from the XGBoost library to build the XGBoost regression model, then we use the Bayesian optimizer from the sklearn library to optimize the hyperparameters of the XGBoost regression model, and finally we use the SHAP library to interpret the results of the machine learning regression predictions.

In this paper, the root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE), and the coefficient of determination R^2 are selected for the comprehensive evaluation of the XGBoost regression prediction model. Among the four performance measures, RMSE, MAE, and MAPE can reflect the optimization error

of the model, and the smaller the value means, the better the prediction error of the model. Among them, R^2 can reflect the fitting accuracy of the prediction results of the XGBoost regression prediction model with the real value, and the value of R^2 is between $[0, 1]$, and the bigger the value of the coefficient of determination, the higher the accuracy of the model is. The above formula is defined as follows:

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y - \hat{y}_i)^2}, \\ \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |y - \hat{y}_i|, \\ R^2 &= \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \\ \text{MAPE} &= \frac{1}{n} \sum_{i=1}^n \left| \frac{(y_i - \hat{y}_i)}{y_i} \right| \times 100\%. \end{aligned} \quad (3)$$

4. Empirical Results and Analysis

4.1. Changes in China's Economic Resilience. This paper applies the TOPSIS method to measure the economic resilience level of China's 30 provinces, as shown in Table 2, and divided China's economic resilience level into five grades: the economic resilience level in the range of 0 to 0.2 is grade 1, indicating a low level of economic resilience; the economic resilience level in the range of 0.2 to 0.4 is grade 2, indicating a relatively low level of economic resilience; the economic resilience level in the range of 0.4 to 0.6 range is level 3, indicating a medium level of economic resilience; the level of economic resilience in the range of 0.6 to 0.8 is level 4, indicating a relatively high level of economic resilience; the level of economic resilience in the range of 0.8 to 1, then the economic resilience is level 5, indicating a high level of economic resilience.

4.1.1. Time Series Analysis of China's Economic Resilience. The average level of economic resilience of 30 Chinese provinces increased from low resilience to lower resilience, with an overall upward trend, and the average value of their economic resilience index increased from 0.11 in 2007 to 0.25 in 2021. For space reasons, the economic resilience values of China in 2007, 2010, 2016, and 2021 are listed, as shown in Table 2. Based on the average economic resilience index between provinces, Guangdong has the highest economic resilience index of 0.43, which is at the general economic resilience level; Jiangsu, Shanghai, Zhejiang, and Beijing have economic resilience indexes in the range of 0.2 to 0.4, which is at the lower economic resilience level, and the rest of the provinces are at the lower economic resilience level. From the annual economic toughness index of each province, the percentage of provinces with general economic toughness increased from 0% in 2007 to 13.33% in 2021, and the percentage of provinces with lower economic toughness increased from 3.33% in 2007 to 33.33% in 2021.

4.1.2. Spatial Analysis of China's Economic Resilience. In terms of spatial distribution, the level of economic resilience in the eastern region is significantly higher than in the central and western regions. Provinces with lower economic resilience and above in the eastern region account for 75% and provinces with lower economic resilience in the central region account for 44.44%, while those in the western region account for only 22.22%. Guangdong in the eastern region is the first to enter the higher resilient provinces with its well-developed industry, well-established service sector institutions, and strong science and technology innovation capabilities. The economic resilience of the central region is slowly increasing, but the overall level is low, and there is a large gap with the east, which needs to further strengthen policy support and reform and innovation to improve the vitality and competitiveness of economic development in the central region. The economic resilience of the western region shows obvious geographical differences, among which the central region, represented by Sichuan and Shaanxi, shows a relatively strong level of economic resilience due to its strong resource endowment and industrial base, while other regions have relatively weaker economic resilience due to various constraints and limitations.

4.2. Influencing Factors and Feature Selection. Since the relationship between the economic resilience evaluation indicators selected in this article is unknown and there may be a problem of multicollinearity. Pearson's coefficients were used to divide all explanatory variables and identify high, moderate, weak, and irrelevant variables in the explanatory variables [27]. In this paper, indicator factors were used as explanatory variables and TOPSIS evaluation values were used as explanatory variables, and the Pearson correlation was detected using Python language to obtain the correlation results between the explanatory variables and the explanatory variables, and then the strength of the correlation between the explanatory variables and the explanatory variables was determined based on the magnitude of the Pearson coefficients, and the coefficient results are shown in Figure 1.

In order to avoid the influence of too many indicators, which increases the computation time of the machine learning model and the influence of irrelevant variables on the accuracy of the experiment. According to the correlation size, RTS, RDE, SIVA, EXED, NPA, BPIFE, STEX, TPIES, and TIE were selected as the explanatory variables with Pearson's coefficients greater than 0.8, representing the total retail sales of consumer goods, R&D expenditure, value added of the secondary industry, education expenditure, the number of domestic invention patents licensed, the cumulative balance of the urban workers' pension insurance fund, science and technology expenditure, the profit of the industrial enterprises above designated size, and the total import and export amount, respectively. The nine explanatory variables with Pearson's coefficients greater than 0.8 are total imports and exports.

4.3. Model Parameter Optimization. In this paper, the benchmark XGBoost model parameters are set as follows: the parameters of the XGBoost extreme gradient-boosting model are set to a base learner of 100, and the maximum

TABLE 2: Values of economic resilience of China in 2007, 2010, 2016, and 2021.

Provinces	2007		2010		2016		2021	
	Rating	Grade	Rating	Grade	Rating	Grade	Rating	Grade
Beijing	0.12	I	0.18	I	0.35	II	0.52	III
Tianjin	0.10	I	0.11	I	0.15	I	0.21	II
Hebei	0.11	I	0.13	I	0.16	I	0.19	I
Liaoning	0.11	I	0.12	I	0.14	I	0.18	I
Shanghai	0.15	I	0.19	I	0.28	II	0.39	II
Jiangsu	0.16	I	0.23	II	0.39	II	0.57	III
Zhejiang	0.12	I	0.16	I	0.27	II	0.44	III
Fujian	0.11	I	0.12	I	0.18	I	0.25	II
Shandong	0.13	I	0.19	I	0.29	II	0.42	III
Guangdong	0.21	II	0.33	II	0.45	III	0.73	IV
Hainan	0.10	I	0.09	I	0.09	I	0.30	II
Shanxi	0.09	I	0.10	I	0.11	I	0.13	I
Jilin	0.10	I	0.10	I	0.11	I	0.11	I
Heilongjiang	0.11	I	0.11	I	0.12	I	0.11	I
Anhui	0.11	I	0.11	I	0.17	I	0.26	II
Jiangxi	0.09	I	0.10	I	0.13	I	0.18	I
Henan	0.10	I	0.13	I	0.19	I	0.26	II
Hubei	0.11	I	0.13	I	0.18	I	0.27	II
Hunan	0.11	I	0.12	I	0.17	I	0.22	II
Inner Mongolia	0.10	I	0.11	I	0.13	I	0.15	I
Guangxi	0.10	I	0.11	I	0.12	I	0.15	I
Chongqing	0.10	I	0.11	I	0.14	I	0.18	I
Sichuan	0.11	I	0.12	I	0.18	I	0.28	II
Guizhou	0.10	I	0.10	I	0.11	I	0.16	I
Yunnan	0.11	I	0.12	I	0.12	I	0.16	I
Shaanxi	0.10	I	0.11	I	0.14	I	0.22	II
Gansu	0.09	I	0.09	I	0.09	I	0.12	I
Qinghai	0.10	I	0.11	I	0.10	I	0.09	I
Ningxia	0.12	I	0.12	I	0.12	I	0.13	I
Xinjiang	0.11	I	0.10	I	0.11	I	0.12	I

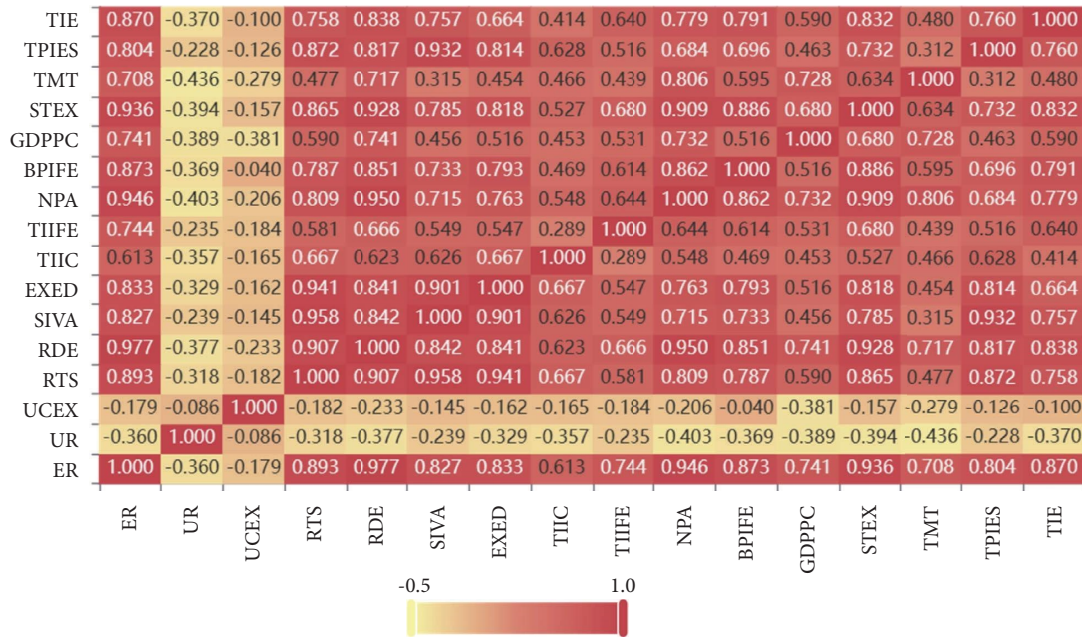


FIGURE 1: Pearson's correlation coefficient.

depth of the tree is 3. In order to avoid the overfitting problem of the XGBoost model, we need to optimize the model parameters.

The hyperparameters in the XGBoost regression prediction model directly affect the performance and prediction effect of the XGBoost model, and the main hyperparameters are `n_estimators`, `max_depth`, and `learning_rate`. `n_estimators` is the number of base learners, and the larger the number is, the better the model's learning ability is, but the more prone it is to overfitting. `max_depth` is the depth of the tree, and if the maximum depth of the tree is larger, the more prone it is to overfitting; if it is too small, it will lead to an oversimplified model. `min_child_weight` is the minimum weight of the leaf nodes. `learning_rate` is the speed of iterative decision making, and the `learning_rate` is the minimum weight of the leaf nodes. `learning_rate` is the step size of the iterative decision tree, which is also known as the learning rate, which controls the iteration speed of the algorithm and is usually used to prevent overfitting. The values of these four hyperparameters have a great impact on the model performance, so they are chosen as optimization hyperparameters.

Bayesian optimization is a global optimization algorithm that searches for the optimal solution that minimizes the objective function in a high-dimensional and nonconvex search space, proposed by Pelikan et al. [28, 29], which uses Bayes' theorem to construct an agent model, and determines the next step of optimization by continually updating this agent model with better convergence theory guarantees of the optimization model's hyperparametric approach [30, 31]. In XGBoost hyperparameter tuning, Bayesian optimization can automatically find the optimal hyperparameter configuration to minimize the model validation error. In addition to Bayesian optimization, there are many other hyperparameter computation methods, such as grid search and stochastic search. Grid search works by exhaustively enumerating a series of hyperparameter combinations and selecting the optimal one. Random search, on the other hand, randomly selects a small number of hyperparameter combinations to search, and then selects the optimal combination. The Bayesian optimization algorithm is able to select the next sampling point based on the existing data and the confidence level of the model construction, thus achieving efficient and accurate global optimization. The method shows superior performance in handling complex, nonconvex, nonlinear optimization problems, even when the problem size is very large.

In addition, Bayesian optimization has the advantage that it is computationally efficient and can find the global optimal solution quickly. This is because it employs a Gaussian process to take into account previous parameter information and constantly update the prior, and it is faster with relatively fewer iterations. In contrast, the lattice search algorithm does not take into account previous parameter information, and the search is slower, which may lead to dimensionality explosion when the number of parameters increases. In addition, Bayesian optimization remains robust for nonconvex problems, while grid search is prone to obtain locally optimal solutions for nonconvex problems. In order

to select the optimal hyperparameters more reasonably and accurately, this study uses Bayesian optimization to accurately optimize the hyperparameters of the XGBoost model.

This paper identifies the most important parameters and value ranges that may affect the effectiveness of the XGBoost model. These parameters and value ranges are replaced in the Bayesian optimization algorithm. The combinations of hyperparameters used for the XGBoost model prediction are as follows: the optimal range of `n_estimators` is [10, 500]; the optimal range of `max_depth` is [1, 15]; and the optimal range of `learning_rate` is [0.01, 1]. The optimization parameters of XGBoost are as follows: the base learning number is 373, the learning rate is 0.027, the minimum weight of the samples in the leaf nodes is 3.48, and the maximum depth of the tree is 9.19. The specific parameter adjustments are shown in Table 3. In order to reduce the time complexity of the simulation optimization step, 20% of the original data are used as the test set and 80% as the training set, and the fitting degree and error of the model are examined in the test set.

4.4. Model Performance Evaluation. K-fold cross-validation is a data splitting technique where the data are divided into n mutually exclusive subsets, and at each iteration, $n - 1$ subsets are taken as the training set and 1 subset as the test set, so that n sets of test sets and training sets can be obtained, and thus n trainings can be completed [32]. In this paper, we use the five-fold cross-validation method to verify the model performance, i.e., the data are divided into 5 parts, 4 of which are used as the training data, and the remaining 1 part is used as the test data, and the cycle is repeated 5 times in order to ensure that each set of data is tested 1 time. The final evaluation result is the average of 5 times. When applying the K-fold method, usually choose 5-fold or 10-fold cross-validation, based on the consideration of the amount of data, this paper chooses 5-fold cross-validation, and the results show that the XGBoost regression prediction model performs well, and there is no overfitting phenomenon. The corresponding algorithm evaluation index results are shown in Table 4.

In Table 5 and Figure 2, it is shown that the Bayesian optimization algorithm can effectively improve the performance of the algorithm.

The accuracy of the XGBoost prediction model optimized by the BO algorithm is improved by 0.62% compared to the baseline XGBoost model. The BO-XGBoost measure of China's economic resilience is shown in Table 6. In addition, we use random sampling with no restriction on scrambling the temporal and spatial order of the samples in the training and test sets, dividing the training set into 8:2 ratios to demonstrate the robustness of the different machine learning approaches. As can be seen in Figure 3, the BO-XGBoost model still has a high prediction accuracy, which is consistent with the conclusions obtained based on time series sampling.

In addition, we plot the scatter intersection of the real and predicted values, and Figure 3 shows the prediction results of China's economic toughness by the BO-XGBoost model. The horizontal axis is the real economic toughness

TABLE 3: Bayesian parameter optimization.

Model name	Parameter names	Parameter settings
BO-XGBoost	n_estimators	373.13
	learning_rate	0.027
	max_depth	9.19
	min_child_weight	3.48

TABLE 4: Predictive evaluation metrics results of the BO-XGBoost algorithm under 5-fold cross-validation.

Evaluation indicators	Train R^2	Test R^2	Train RMSE	Test RMSE
1	0.9999	0.9678	0.0008	0.1842
2	0.9999	0.9664	0.0008	0.1411
3	0.9998	0.9489	0.0009	0.0210
4	0.9999	0.9175	0.0009	0.2661
5	0.9999	0.9671	0.0008	0.019
Average value	0.9999	0.9535	0.0008	0.0198

TABLE 5: Model evaluation table.

Model names	MAPE	R^2	MAE	MAPE
XGBoost	0.01	0.968	0.008	0.059
B0-XGBoost	0.01	0.974	0.007	0.056

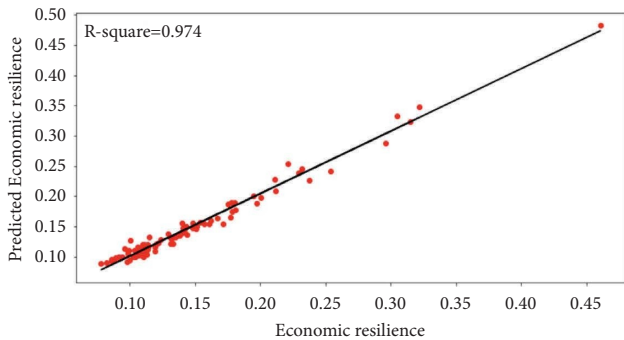


FIGURE 2: The forecasting result of the BO-XGBoost model.

and the vertical axis is the predicted economic toughness. $y = x$ is indicated by the black solid line. The training results based on the training set are shown by the red scatter in the figure. Ideally, the scatters should converge on the solid line $y = x$. The scatters of the BO-XGBoost model are almost always clustered around the $y = x$ line, which shows high prediction accuracy.

4.5. *Analysis of Drivers of Economic Resilience.* Figure 1 shows the SHAP global feature analysis of the XGBoost model. The higher the SHAP value of a feature, the more resilient China’s economy is. RTS, RDE, SIVA, EXED, NPA, BPIFE, STEX, TPIES, and TIE represent the nine variables of total retail sales of consumer goods, R&D expenditures, value added of the secondary industry, education expenditures, the number of domestic invention patents licensed,

TABLE 6: Measures of China’s economic resilience based on the BO-XGBoost model.

Provinces	2021		2021	
	TOPSIS	Grade	BO-XGBoost	Grade
Beijing	0.52	III	0.52	III
Tianjin	0.21	II	0.20	II
Hebei	0.19	I	0.19	I
Liaoning	0.18	I	0.18	I
Shanghai	0.39	II	0.39	II
Jiangsu	0.57	III	0.58	III
Zhejiang	0.44	III	0.44	III
Fujian	0.25	II	0.24	II
Shandong	0.42	III	0.42	III
Guangdong	0.73	IV	0.70	IV
Hainan	0.30	II	0.26	II
Shanxi	0.13	I	0.13	I
Jilin	0.11	I	0.12	I
Heilongjiang	0.11	I	0.12	I
Anhui	0.26	II	0.26	II
Jiangxi	0.18	I	0.17	I
Henan	0.26	II	0.25	II
Hubei	0.27	II	0.27	II
Hunan	0.22	II	0.22	II
Inner Mongolia	0.15	I	0.15	I
Guangxi	0.15	I	0.15	I
Chongqing	0.18	I	0.19	I
Sichuan	0.28	II	0.28	II
Guizhou	0.16	I	0.16	I
Yunnan	0.16	I	0.15	I
Shaanxi	0.22	II	0.22	II
Gansu	0.12	I	0.12	I
Qinghai	0.09	I	0.10	I
Ningxia	0.13	I	0.13	I
Xinjiang	0.12	I	0.12	I

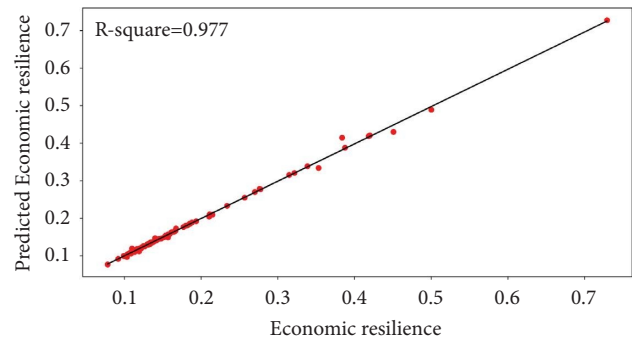


FIGURE 3: Prediction results of the BO-XGBoost model under dataset shuffling.

the cumulative balance of the urban workers’ pension fund, science and technology expenditures, the profits of industrial enterprises above the large-scale, and the total amount of imports and exports, respectively. As can be seen from Figure 4, R&D expenditures, invention patent authorizations, and S&T inputs are important factors affecting the resilience of the Chinese economy.

In order to explore more intuitively how the characteristics affect the output of the model and to extract valuable information to help relevant government departments take

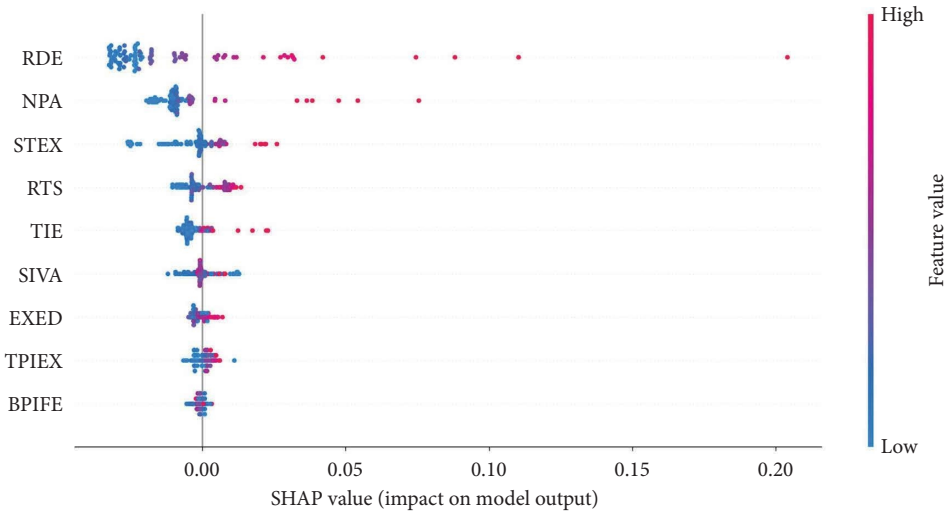


FIGURE 4: Summary chart of China's economic resilience SHAP.

targeted measures, this paper uses SHAP value mapping plots to show the nonlinear relationships between variables. Unlike the partial dependency graph, the vertical coordinate of the SHAP value mapping graph is the SHAP value rather than the output label value [33]. This leads to the threshold values of the key variables to improve the resilience of the Chinese economy. On the one hand, there is a clear segmentation between each variable and the increase in the level of China's economic resilience, which can well reflect the size of the marginal effect.

4.5.1. The Mapping Relationship between R&D Funding Investment and China's Economic Resilience. In Figure 5, when the investment in R&D funding is within the interval of [0, 500], it shows a smooth negative effect. When the investment in R&D funding exceeds 50 billion, the positive SHAP reflects that R&D funding investment has a significant positive effect on economic resilience. It can be seen that increasing the investment in R&D funding, deeply implementing major scientific and technological special projects and major projects, focusing on national strategic needs and frontier fields, breaking through a number of key core technologies, and forming a number of original, leading and supportive scientific and technological achievements will contribute to the improvement of China's economic resilience level.

4.5.2. Mapping the Number of Invention Patents to China's Economic Resilience. In Figure 6, when the number of patents is within the interval of [0, 12000], it shows a rapid upward trend and the SHAP value is less than 0, which has a negative effect on the level of economic resilience. When the number of invention patents exceeds 12,000, it has a significant positive effect on economic resilience. This indicates that promoting the transformation and application of scientific and technological achievements, improving the institutional mechanism of evaluation and protection of scientific and technological achievements, motivating

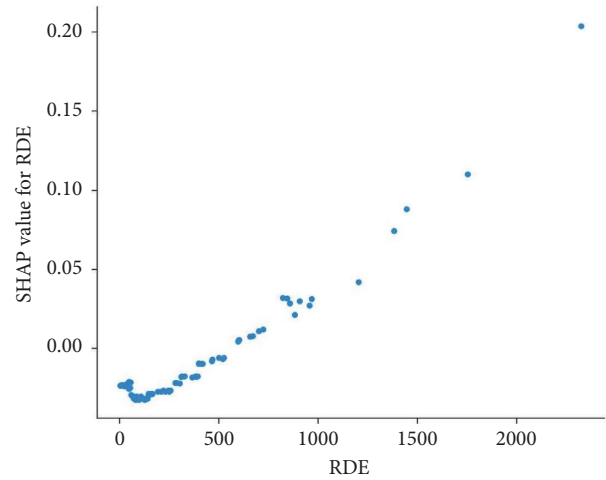


FIGURE 5: Mapping of R&D investment and China's economic resilience.

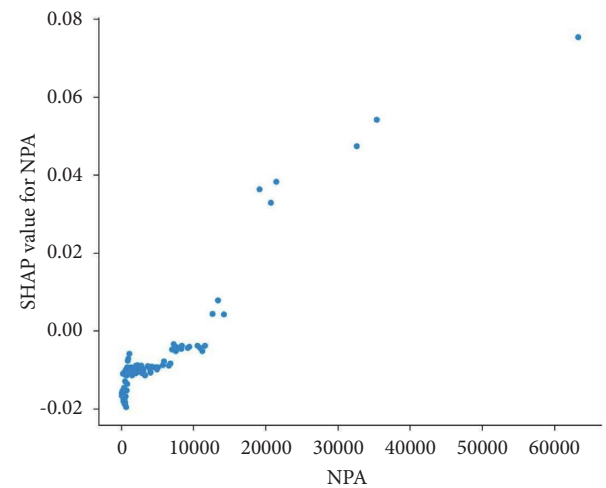


FIGURE 6: Mapping the number of invention patents to China's economic resilience.

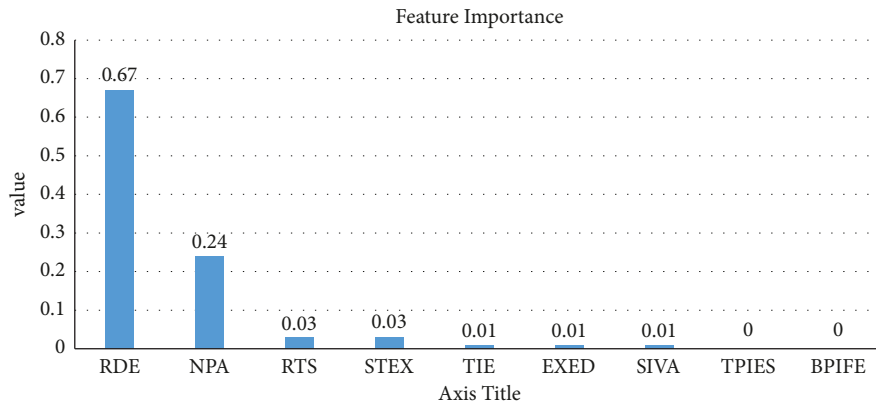


FIGURE 7: Feature importance.

researchers and enterprises to strengthen cooperation, and improving the marketability and social benefits of scientific and technological achievements will play a positive role in improving the resilience of China's economy.

4.6. Importance Analysis. In XGBoost, the `importance_type` parameter can be used to specify how feature importance is calculated. In this paper, the parameter is set to `importance_type="gain"` to calculate the gain of each feature during the training process, as a way to evaluate the importance of each feature in the model and to find out the most important features for the prediction results. Gain reflects the degree of contribution of each factor to the model prediction. The larger the value of the gain of a feature, the larger the reduction of that feature to the loss function at the split node, and the higher the importance of that feature in the model prediction.

Figure 7 shows the importance of each factor when using XGBoost-Bayes in training and testing. The importance can further explain which features contribute more to the prediction results under the XGBoost model. Figure 1 shows that RD expenditure and the number of invention patents granted to domestic applicants have the highest correlation with China's economic resilience, with feature importance of 67% and 24%, respectively. It indicates that these characteristics are the main factors to enhance China's economic resilience. Compared to the SHAP summary plot, the feature rankings are basically the same, which indicates the robustness of the feature importance rankings produced by the XGBoost model.

5. Conclusion

In this paper, the TOPSIS-BO-XGBoost-SHAP model is used to implement the fuzzy problem in the quantitative expression of economic resilience. The collected results show that the overall level of China's economic resilience is in steady increase from 2007 to 2021, and the average value of its level of economic resilience level increases from 0.11 in 2007 to 0.25 in 2021. Second, the XGBoost model optimized by Bayesian algorithm shows higher accuracy and stronger generalisation ability, which indicates that the XGBoost model has good applicability in China's economic resilience assessment. Finally, through visual

analysis of the results of SHAP interpretable tool, we found that the key factors affecting China's economic resilience are R&D expenditure, number of invention patents, and science and technology investment with positive effects.

In conclusion, this paper bridges the gap of machine learning algorithms in nonlinear causal analysis of economic resilience by using the TOPSIS-BO-XGBoost-SHAP model. The method is scientifically sound and reasonable in exploring and evaluating the resilience and enhancement mechanisms of economic systems. Future research can focus more on the interpretable analysis of machine learning, so that machine learning can be well applied to various fields of real life by gradually showing its intrinsic mechanisms while ensuring accuracy.

However, this study also has some limitations. First, from the perspective of index construction, this paper may be deficient in the selection of indicators, which may have an impact on the results of economic resilience assessment due to the limited availability of indicators and the long sequence length. In future research, adding suitable indicators can be considered to improve this problem. Second, in terms of methodology, a new time series generation framework, TimeGAN, can be used in the future to combine the flexibility of unsupervised learning and the strong control advantage of supervised training to generate synthetic time series data based on the joint training of GAN and self-encoder, and data augmentation [34] can be used to improve the prediction accuracy and robustness of the XGBoost regression prediction model. Using federated learning, distributed computing is performed to solve the problem of data availability and invisibility and data immobilization model movement and enhance privacy computation to construct a security tree model combining federated learning and XGBoost [35]. Finally, further research could be built on the existing foundation by adding variables related to government policy implementation to measure the impact of policy implementation on economic resilience. Therefore, future efforts must be made to fill this gap.

Data Availability

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

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] M. Li and X. Wang, "How regions react to economic crisis: regional economic resilience in a Chinese perspective," *Sage Open*, vol. 12, no. 4, Article ID 215824402211425, 2022.
- [2] L. Liu, Y. Lei, B. D. Fath et al., "The spatio-temporal dynamics of urban resilience in China's capital cities," *Journal of Cleaner Production*, vol. 379, Article ID 134400, 2022.
- [3] Z. Wang and W. Wei, "Regional economic resilience in China: measurement and determinants," *Regional Studies*, vol. 55, no. 7, pp. 1228–1239, 2021.
- [4] J. Jiang, X. Zhang, and C. Huang, "Influence of population agglomeration on urban economic resilience in China," *Sustainability*, vol. 14, no. 16, Article ID 10407, 2022.
- [5] W. Wang, J. Wang, S. Wulaer, B. Chen, and X. Yang, "The effect of innovative entrepreneurial vitality on economic resilience based on a spatial perspective: economic policy uncertainty as a moderating variable," *Sustainability*, vol. 13, no. 19, Article ID 10677, 2021.
- [6] L. Briguglio, G. Cordina, N. Farrugia, N. Farrugia, and S. Vella, "Economic vulnerability and resilience: concepts and measurements," *Oxford Development Studies*, vol. 37, no. 3, pp. 229–247, 2009.
- [7] J. Tan, P. Zhang, K. Lo et al., "Conceptualizing and measuring economic resilience of resource-based cities: case study of Northeast China," *Chinese Geographical Science*, vol. 27, no. 3, pp. 471–481, 2017.
- [8] X. Xun and Y. Yuan, "Research on the urban resilience evaluation with hybrid multiple attribute TOPSIS method: an example in China," *Natural Hazards*, vol. 103, no. 1, pp. 557–577, 2020.
- [9] W. Wang, C. Xia, C. Liu, and Z. Wang, "Study of double combination evaluation of urban comprehensive disaster risk," *Natural Hazards*, vol. 104, no. 2, pp. 1181–1209, 2020.
- [10] L. Jiao, L. Wang, H. Lu, Y. Fan, Y. Zhang, and Y. Wu, "An assessment model for urban resilience based on the pressure-state-response framework and BP-GA neural network," *Urban Climate*, vol. 49, Article ID 101543, 2023.
- [11] D. Liu, C. Wang, Y. Ji et al., "Measurement and analysis of regional flood disaster resilience based on a support vector regression model refined by the selfish herd optimizer with elite opposition-based learning," *Journal of Environmental Management*, vol. 300, Article ID 113764, 2021.
- [12] D. Liu, Z. Fan, Q. Fu et al., "Random forest regression evaluation model of regional flood disaster resilience based on the whale optimization algorithm," *Journal of Cleaner Production*, vol. 250, Article ID 119468, 2020.
- [13] B. Tan, Z. Gan, Y. Wu, Z. Gan, and Y. Wu, "The measurement and early warning of daily financial stability index based on XGBoost and SHAP: evidence from China," *Expert Systems with Applications*, vol. 227, Article ID 120375, 2023.
- [14] R. Li, M. Xu, and H. Zhou, "Impact of high-speed rail operation on urban economic resilience: evidence from local and spillover perspectives in China," *Cities*, vol. 141, Article ID 104498, 2023.
- [15] S. Man, X. Wu, Y. Yang, and Q. Meng, "An assessment approach to urban economic resilience of the rust belt in China," *Complexity*, vol. 2021, Article ID 1935557, 16 pages, 2021.
- [16] H. Wang and Q. Ge, "Spatial association network of economic resilience and its influencing factors: evidence from 31 Chinese provinces," *Humanities and Social Sciences Communications*, vol. 10, no. 1, p. 290, 2023.
- [17] B. Lv, C. Liu, T. Li et al., "Evaluation of the water resource carrying capacity in Heilongjiang, eastern China, based on the improved TOPSIS model," *Ecological Indicators*, vol. 150, Article ID 110208, 2023.
- [18] H. S. Shih, H. J. Shyur, E. S. Lee, H. J. Shyur, and E. S. Lee, "An extension of TOPSIS for group decision making," *Mathematical and Computer Modelling*, vol. 45, no. 7-8, pp. 801–813, 2007.
- [19] M. Ma, G. Zhao, B. He et al., "XGBoost-based method for flash flood risk assessment," *Journal of Hydrology*, vol. 598, Article ID 126382, 2021.
- [20] O. Sagi and L. Rokach, "Approximating XGBoost with an interpretable decision tree," *Information Sciences*, vol. 572, pp. 522–542, 2021.
- [21] T. Chen and C. Guestrin, "Xgboost: a scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, San Francisco, CA, USA, August, 2016.
- [22] S. Villa, S. Matet, B. C. Vü, and L. Rosasco, "Implicit regularization with strongly convex bias: stability and acceleration," *Analysis and Applications*, vol. 21, no. 01, pp. 165–191, 2023.
- [23] S. Huang, Y. Feng, and Q. Wu, "Learning theory of minimum error entropy under weak moment conditions," *Analysis and Applications*, vol. 20, no. 01, pp. 121–139, 2022.
- [24] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Proceedings of the Advances in Neural Information Processing Systems*, pp. 4765–4774, Long Beach, CA, USA, December, 2021.
- [25] T. Zhang, W. Zhu, Y. Wu, Z. Wu, C. Zhang, and X. Hu, "An explainable financial risk early warning model based on the DS-XGBoost model," *Finance Research Letters*, vol. 56, Article ID 104045, 2023.
- [26] D. Shi, F. Zhou, W. Mu et al., "Deep insights into the viscosity of deep eutectic solvents by an XGBoost-based model plus SHapley Additive exPlanation," *Physical Chemistry Chemical Physics*, vol. 24, no. 42, pp. 26029–26036, 2022.
- [27] I. Jebli, F. Z. Belouadha, M. I. Kabbaj, A. Tilioua, M. I. Kabbaj, and A. Tilioua, "Prediction of solar energy guided by pearson correlation using machine learning," *Energy*, vol. 224, Article ID 120109, 2021.
- [28] P. Martin, E. G. David, and C. P. Erick, "BOA: the Bayesian optimization algorithm," in *Proceedings of the genetic and evolutionary computation conference GECCO-99*, vol. 1, Orlando, FL, USA, July, 1999.
- [29] Y. Shen, Y. Zuo, L. Sun, X. Zhang, L. Sun, and X. Zhang, "Modified proximal symmetric ADMMs for multi-block separable convex optimization with linear constraints," *Analysis and Applications*, vol. 20, no. 03, pp. 401–428, 2022.
- [30] K. Y. Huang, J. B. K. Hsu, and T.-Y. Lee, "Characterization and identification of lysine succinylation sites based on deep learning method," *Scientific Reports*, vol. 9, no. 1, Article ID 16175, 2019.
- [31] J. Hoffer, S. Ranftl, and B. C. Geiger, "Robust bayesian target value optimization," *Computers and Industrial Engineering*, vol. 180, Article ID 109279, 2023.
- [32] T. T. Wong, "Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation," *Pattern Recognition*, vol. 48, no. 9, pp. 2839–2846, 2015.

- [33] K. K. Yun, S. W. Yoon, and D. Won, "Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process," *Expert Systems with Applications*, vol. 186, Article ID 115716, 2021.
- [34] J. Yoon, D. Jarrett, and M. van der Schaar, "Time-series generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 32, 2019.
- [35] L. Y. Wei, Z. Yu, and D. X. Zhou, "Federated learning for minimizing nonsmooth convex loss functions," *Mathematical Foundations of Computing*, vol. 6, no. 4, pp. 753–770, 2023.