Research Article

Forecast of Foreclosure Property Market Trends during the Epidemic Based on GA-BP Neural Network

YaoChing Fang (ORCID)

Institute of China and Asia-Pacific Studies, National Sun Yat-sen University, Kaohsiung 80424, Taiwan

Correspondence should be addressed to YaoChing Fang; d106070005@nsysu.edu.tw

Due to the epidemic, foreclosure, as a special transaction commodity, has also been affected to a certain extent. This paper uses the sample data of foreclosure auctions combined with the genetic algorithm multi-layer feedforward neural network to apply in the forecast of the foreclosure market trend. First of all, to select the indicators that affect the price fluctuation of foreclosed houses, use factor analysis technology to reduce the dimensionality of the original data to obtain two common factors; then combine the BP neural network and the genetic algorithm to propose a GA-BP neural network The prediction model of the foreclosure housing market takes the influencing factors as the input of the model and the housing price as the output. Finally, the prediction results are compared with the multiple linear regression and the extreme learning machine, which shows that the GA-BP neural network model is in It is a good choice for research, and it provides an effective method for predicting housing prices.

1. Introduction

As a circulating commodity, foreclosed houses are also in line with the laws of market operation. The supply of foreclosed houses and market demand determine their market position [1–5]. The two are related, but on another level, foreclosed houses are born with distinctive features that are different from other houses. The characteristic, that is, the national judicial coercive force and enforcement power granted by itself, to a certain extent, occupy a great advantage, that is, the convenience and uniqueness of the sales and transfer links [6, 7]. Therefore, the epidemic cannot have a subversive impact on the foreclosure. More viewpoints point out that in recent years, the domestic market has been sluggish and lacklustre, many companies and individuals are facing bankruptcy, and banks and financial institutions have increasingly defaulted. Affected by the epidemic at the beginning of the year, the economic situation forecast makes people unable to be optimistic [8–11]. It is believed that foreclosure properties will soon usher in a peak supply period. The supply increases and the market purchasing power has not followed up. Then the advantages

of foreclosure properties cannot be reflected. With the decrease in investment or rigid demand, prices will inevitably fall, and the market price of foreclosure housing will undoubtedly hit the bottom [1, 12–15].

How long will it take to transform from ordinary real estate to forensic auction, from litigation to the final execution of the auction? Two years as fast as possible, and three to five years as slow as possible [16, 17]. The duration of this cycle determines the stability of the supply of foreclosures. From this, it is not difficult to know that from the analysis of the supply source of foreclosure houses, the probability that the rate of foreclosure auctions will surge or decrease in a short period of time is extremely low. As we all know, since the foreclosure auction room has been converted from offline to online, the China Auction Platform, Taobao, JD.com and other online platforms have joined hands with the people's courts, and professional auction institutions have settled in to assist, and the national court system's online auction work has become more standardized and functioning well., Overall, the volume of foreclosure auctions has steadily increased year by year [5, 18–21].

With the various conveniences of online foreclosure transactions being widely accepted by home buyers, foreclosures are undoubtedly a beacon in the long night for the dream of buying a house blocked by the purchase restriction policy, bringing it to many buyers and investors Hope. For middle-class and above families, whether it is for the next generation to deploy in big cities in advance to help them settle down, or to make family assets safer and more appreciation and preservation, buying a house should be a relatively safe and feasible way. The foreclosure room just meets the needs of these families. The shining sign of "Unlimited Purchase" has attracted the attention of the whole society, and all foreclosures have a formal trading platform, and the foreclosures are endorsed by the people's court, the operation is standardized, and the bidding is transparent, which can interpret the auction to the greatest extent." The connotation of fairness, impartiality and openness; more importantly, professional auxiliary agencies provide a variety of thoughtful and thoughtful services to dispel doubts and worries. For those who really understand the foreclosure, it is undoubtedly a "leakage," although "Leak-picking" is a speculative behavior, but it can also be understood as a market behavior, and a large number of rigidly needed groups have always existed. The huge market demand and gap determine the rationality of the market. Affected by the epidemic, which affects all levels of society, foreclosure as a special transaction commodity should be listed among them. As for the extent and depth of its impact, there is still no conclusion [12]. Up to now, great progress has been made in housing price forecasting research methods. Hu Feng analyzed the historical data from 2006 to 2015 by establishing a regression model, and obtained the main factors affecting housing prices. Zhang Shuangni established a housing price prediction model based on stepwise regression. Liu Cong established a regression analysis model based on principal components when analyzing the influencing factors of housing prices, and established a time series model when predicting housing prices.

In order to further improve the accuracy of the foreclosure house price prediction model, this paper attempts to use the foreclosure information data combined with the genetic algorithm multi-layer feedforward neural network (genetic algorithm-back propagation, GA-BP) to be applied in the foreclosure market trend forecast. First, select the indicators that affect the fluctuation of the house price for foreclosure, and use factor analysis technology to reduce the dimension of the original data to obtain two common factors; then, use the GA-BP neural network to combine the BP neural network and the genetic algorithm., put forward a prediction model for the foreclosure housing market based on GA-BP neural network, which takes the influencing factors as the input of the model and the housing price as the output; finally, the prediction results are compared with multiple linear regression and extreme learning machine, which proves that the method is suitable for foreclosure auctions. Advantages in housing price forecasts.

2. Determination of the Influencing Factors of Foreclosure House Price Based on Factor Analysis

2.1. Factor Analysis Method Selection Index. The factor analysis method is a multivariate statistical method that extracts common factors from variable groups, thereby realizing dimension reduction of multiple variables. Factor analysis was first used in the field of psychology to study the fundamental factors affecting human intelligence, and now it has been widely used in various disciplines. This paper uses the number of auction houses and transactions in the court system of Yulin City (including districts and counties), counts the number of auction houses and transactions in the court system from 2018 to 2021 as a sample, and selects indicators that affect the fluctuation of foreclosure house prices. Using SPSS19 for factor analysis, the indicators are shown in Table 1.

(1) Applicability test. Enter the relevant variables of the foreclosure room into SPSS19, and the KMO value is 0.765, which is greater than 0.5, so there is a correlation between the variables. In the Bartlett sphere test, the P value is 0, which is less than 0.05, and it is considered that there is no significant difference between the variables. Therefore, common factors can be extracted, which is suitable for factor analysis.

(2) Factor rotation and interpretation. Two common factors are calculated: the first factor has a higher load on x1, x2, x3, x4, and x5, and it is named the foreclosure market impact factor; the second factor has on x6, x7, and x8 The higher the load, it is named the real estate market impact factor. According to the characteristic value greater than 1 and the cumulative variance contribution rate reached 92.689%, the original 8 variables were finally reduced in dimensionality.

(3) Calculate factor function. Get the factor function according to the score coefficient matrix, Among them, z_i ($i = 1, 2, 8$) is the standardized variable of x_i: $F_1 = 0.237z_1 + 0.003z_2 + 0.221z_3 + 0.176z_4 + 0.192z_5 + 0.271z_6 + 0.099z_7 + 0.166z_8$, $F_2 = -0.081z_1 + 0.266z_2 - 0.045z_3 + 0.036z_4 + 0.01z_5 - 0.177z_6 - 0.403z_7 - 0.459z_8$.

2.2. Data Preprocessing. The data used in the research comes from the foreclosure housing price data set of a provincial court system. Then the following data processing methods were adopted according to the data of the data set. Data normalization is an important technique. Generally speaking, before using the neural network model, it must be used to preprocess the data. It can standardize the range of each feature in the input data set. If the model is not preprocessed before training, the model may not be able to converge to the optimal solution. In the multi-index evaluation method, due to the different nature of each evaluation index, it usually has different dimensions and orders of magnitude. If the initial data is used directly, the data of different dimensions will have different degrees of influence on the network, which

TABLE 1: Impact indicators.

Serial number	Index
x_1	Participate in registration
x_2	Buyer market
x_3	Total price
x_4	Appraisal price
x_5	House area
x_6	GDP
x_7	Seller's taxes
x_8	Present value of house

will eventually make it difficult to learn the model. . Therefore, it is necessary to perform a series of standardized processing on the original data, and then input it into the neural network model, which can reduce the difficulty of subsequent model training and improve the performance of the network model. Data standardization is to scale the data to a smaller specific interval. It is usually used in the comparison or evaluation of the indicators to remove the unit limit of the original data, eliminate the magnitude of different data, and is suitable for comprehensive comparison and evaluation of indicators of different dimensions.

2.2.1. Duplicate value processing. When obtaining data, there are often some repeated data, and the repeated data will affect the statistical results and mislead the decision-makers. Use the duplicated function that has been written in python, which will compare all the columns. If the value of each column in two rows is the same, it will be marked as a duplicate value.

2.2.2. Outlier handling. The Z-SCORE standardization method is to standardize the data according to the mean and standard deviation of the original data. All data can be converted to an area close to zero. This method is suitable for situations where the maximum range and minimum range of attribute data are uncertain or exist. In the case of outliers, the processed data will satisfy a normal distribution with a mean of 0 and a standard deviation of 1.

$$Z = \frac{(X - \mu)}{\sigma}. \tag{1}$$

Among them: μ is the overall average, X-μ is the deviation from the mean, and σ is the standard deviation. The absolute value of z represents the distance between the original score within the standard deviation and the overall mean. Determine whether the Z-score score is greater than 2.2 (here 2.2 represents an empirical value), if it is, it is True, otherwise it is False. Directly delete the data row whose judgment result is true.

2.3. Model Evaluation Method. It is very important to evaluate the performance of a predictive model. It is generally measured from the accuracy of the model and training time, and an appropriate cost function is selected to measure the results of the model to evaluate the pros and cons of the model. If the loss is large, it means that the predicted value

deviates far from the actual value, and the predicted result is not ideal; the smaller the loss, the better the model effect; if the predicted value is equal to the true value, it means that there is no loss. There are different methods for selecting appropriate and effective error analysis for different models. The 4 evaluation methods of MSE, MAE, EV, and R^2 are used to evaluate the training effect of the model.

2.3.1. Mean square error (MSE)

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (f(x_i) - y_i)^2. \tag{2}$$

The range is $[0, +\infty)$. When the predicted value is exactly the same as the true value, it is called a complete model; otherwise, the greater the error, the greater the value.

2.3.2. Mean absolute error (MAE)

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |f(x_i) - y_i|. \tag{3}$$

The range is $[0, +\infty)$. When the predicted value is exactly the same as the true value, it is called a complete model; otherwise, the greater the error, the greater the value.

2.3.3. Root mean square error (RMSE)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (f(x_i) - y_i)^2}{n}}. \tag{4}$$

The formula, also known as the standard error, represents the expected value of the square of the error and can reflect the degree of dispersion of a data set. n is the number of model measurements.

2.3.4. Mean absolute percentage error (MAPE)

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{f(x_i) - y_i}{f(x_i)} \right| \times 100\%. \tag{5}$$

In the range $[0, +\infty)$, a map of 0% indicates a perfect model, and a MAPE greater than 100% indicates an inferior model. From the formula point of view, MAPE and MAE have similarities, the difference is mainly in the position of the denominator. When the true value has data equal to 0, there is a problem of dividing the denominator by 0, and this formula is not available.

2.3.5. Absolute coefficient (R^2)

$$R^2 = 1 - \frac{\sum \text{Squared\_Residuals}/N}{\text{Variance}_{Y_{\text{true}}}}. \tag{6}$$

The closer to 1, the better the effect. The meaning of R2 is from the perspective of least squares (that is, the second-order variance), indicating how much of the variance of the actual y value is explained by the predicted y value.

2.3.6. Difference Interpretation Score (EV)

$$EV = 1 - \frac{\text{Variance}\left(Y_{\text{true}} - Y_{\text{pred}}\right)}{\text{Variance}_{Y_{\text{true}}}}. \quad (7)$$

When the mean of the residual is 0, it is the same as R2.

3. 2GA-BP Neural Network Algorithm

Genetic algorithm is a kind of global search evolution algorithm that refers to biological evolution process. Analogous to the biological evolution process, genetic algorithms simulate biological evolution through selection, inferiority, and mutation to produce next generation solutions. By calculating and ranking the fitness of individuals, individuals with low fitness are eliminated, and the proportion of individuals with high fitness is increased. After several iterative calculations, the individual solution with the best fitness function value is found. The genetic algorithm's heuristic optimal solution search method combined with the BP neural network can effectively solve the problem that the BP neural network is easy to fall into the local optimal solution, and improve the accuracy of housing price prediction. The GA-BP neural network process is shown in Figure 1.

3.1. Genetic algorithm initialization. The genetic algorithm needs to decode the solution data of the solution space into genotype string structure data before using it. The commonly used encoding method is the binary encoding method, and different string structure data constitute different points. The N string structure data generated by encoding represent different individuals. Through the representation of different individuals, the total group of births is finally formed, also known as the initial population.

3.2. Fitness assessment. The fitness indicates the individual's ability to adapt, and also indicates the individual's strengths and weaknesses. The formula for calculating the individual's adaptive ability is

$$F = \text{mse}\,(Y - O) = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - o_i\right)^2,$$
$$F = \text{mse}\,(X - O) = \frac{1}{m}\sum_{i=1}^{m}\left(x_i - o_i\right)^2. \quad (8)$$

In the formula, m is the total number of samples; x_i is the output result of the genetic algorithm; O_i is the actual output result of the genetic algorithm, and mse represents the mean square error function.

3.3. Genetic manipulation. The principle of survival of the fittest is used in the group selection of genetic algorithm to select good individuals for the next generation of genetic reproduction. Individuals with strong adaptability can pass on excellent genes to their offspring through inheritance. In this paper, a proportional selection strategy is adopted, and
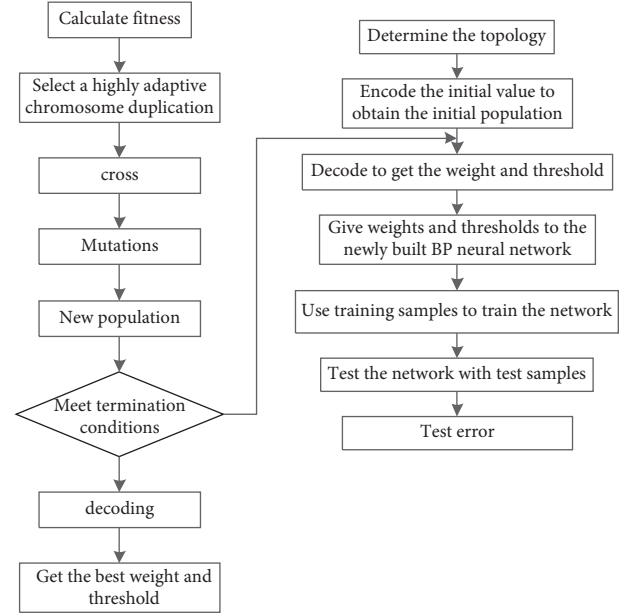

Figure 1: GA-BP network flow chart.

the population data amount is set as M, then the fitness formula of individual i is

$$P_i = \frac{k/F_i}{\sum_{i=1}^{N} k/F_i}. \quad (9)$$

In the formula, k is the genetic coefficient.

In the genetic algorithm, the next generation of individuals is reproduced through crossover operations, so as to obtain new individuals with new characteristics. The crossover operation is used to exchange genetic information, and the mutant individual randomly changes the value of a string in the string structure data with a certain probability, which has the characteristic of small value.

3.4. Combination of genetic algorithm and BP neural network. The optimal value search using genetic algorithm can reduce the calculation amount of BP neural network and optimize the calculation process of BP neural network.

BP neural network draws on the model structure of neural network, and builds a highly complex learning system through the interconnection of a large number of neurons. It has strong self-organization, self-adaptation and self-learning capabilities, arbitrarily complex pattern classification capabilities and excellent multi-dimensional function mapping capabilities. BP network is composed of input layer, output layer and hidden layer. The hidden layer calculates data by setting one or more layers of neurons, and each layer of neurons can have several nodes. The BP neural network common language pattern classification problem such as double hidden has the advantage of fast classification speed, so the BP neural network structure with double hidden layers is used in this article. The neural network is adjusted through the weight matrix and error feedback between each layer to achieve the expected output result of the load. Compared with the traditional artificial neural

network, the double hidden layer BP neural network has improved in terms of parallel processing of massive data and accuracy. BP neural network is shown in Figure 2.

In this paper, the index factors that affect housing prices summarized above are used as input factors, and the predicted housing prices are used as output.

The input vector of BP neural network is $K = (k_1, k_2, k_3, \ldots k_n)^T$, The weight matrix of the input layer and the hidden layer is $S_{ni}(1<n<N, 1<i<I)$, and the weight matrix between the first hidden layer and the second hidden layer is $W_{ij}(1<i<I, 1<j<J)$, the output vector of the first hidden layer is $B = (b_1, b_2, b_3, \ldots, b_n)^T$, The threshold on node i is θi, The output vector of the second hidden layer is $C = (c_1, c_2, c_3, \ldots, c_n)^T$. The threshold at node j is θj, The weight matrix of the second hidden layer and the output layer is Wjn, The output vector of the output layer is $D = (d_1, d_2, d_3, \ldots, d_n)^T$, The threshold at node n is θ_n, f(x) is the activation function. Input the n-dimensional vector A, then the output of the first hidden layer node i is

$$b_i = f\left(\sum_{i=1}^{m} w_{mi}a_m - \theta_i\right). \tag{10}$$

The output of the second hidden layer node j is

$$c_j = f\left(\sum_{j=1}^{I} w_{ij}b_i - \theta_j\right). \tag{11}$$

The output result of the nth node in the output layer is

$$d_n = f\left(\sum_{n=1}^{J} w_{jm}c_j - \theta_n\right). \tag{12}$$

In order to improve the convergence speed of the BP neural network, the input data is normalized to reduce the range of changes and increase the flexibility of interval selection. The formula is as follows:

$$t'_i = \frac{t_i - t_{\min}}{t_{\max} - t_{\min}}, \tag{13}$$

where: t_j is the input value of the neural network, t_i is the normalized value of the interval [0, 1], and $t_{\min}$ is the minimum and $t_{\max}$ maximum value of the sample, respectively.

The activation function is as follows:

$$f(x) = \frac{1}{1 + e^{-x}}. \tag{14}$$

In view of the above analysis, by selecting the factors that affect the foreclosure property, the GA-BP neural network-based foreclosure property market prediction model is established, which can effectively predict the price of the property.

4. Experiment Analysis

In this paper, Matlab platform is used to design and implement a foreclosure market prediction model based on GA-BP neural network. According to the results of factor
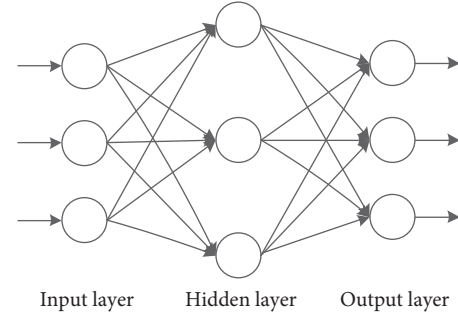

FIGURE 2: BP network topology.

analysis, the obtained F1 and F2 are used as the model input, and the housing price as the output, to establish a housing price prediction model. The experimental data comes from the number of auctions and transactions of the court system in a certain province from 2018 to 2021 as sample data. This paper divides the data into training set and test set, the numbers are 580 and 480 respectively. The GA algorithm parameters are set as follows: the number of iterations is 200, the selection probability is 0.8, the crossover probability is 0.6, and the mutation probability is 0.08.

When the model first iterated, the fitness value of the population individual was far from the optimal fitness value, and the individual fitness value increased significantly. At the later stage of the model iteration, due to the continuous convergence of the model, the population individual fitness value became more and more Close to optimal fitness. After completing the above work, the global optimal initial weight threshold searched by the genetic algorithm can be obtained, and the initial weight threshold can be brought into the network to train the BP network. Set the training parameters of the BP network: the number of neurons in the input layer is 6, the number of neurons in the hidden layer is 9, the number of neurons in the output layer is 1, and the unipolar sigmoid function is selected as the transfer Function, select the gradient descent BP training algorithm function training as the training function, the maximum number of training times is 200, the learning rate is 0.1, and the minimum mean square error of learning is 0.003. According to the selected 12 input vectors, the output data is normalized. The summary steps for establishing a housing price prediction model are as follows:

(1) Enter the original index data in SPSS19

(2) Use SPSS19 to do factor analysis to achieve dimensionality reduction of indicators, and to obtain new variable values

(3) Use the data from 2018 to 2020 as the training sample, and the data from 2017 to 2018 as the test sample

(4) Establish a GA-BP neural network prediction model based on training samples

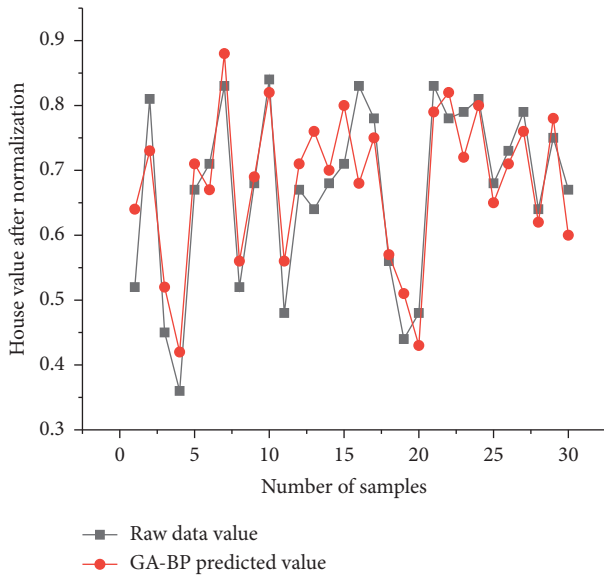(5) Based on test samples, verify the accuracy of the model

FIGURE 3: Forecast results.

TABLE 2: Model evaluation error analysis table.

Algorithm	MSE	MAE	EV	R^2
Multiple linear regression	0.0067	0.065	0.71	0.71
Extreme learning machine	0.0065	0.069	0.74	0.74
BP neural network	0.0051	0.073	0.68	0.67
GA-BP	0.0015	0.031	0.94	0.94

According to the constructed model, Matlab is used to establish a BP neural network model with two hidden layers, input the normalized data and iterate until the output results meet the prediction accuracy required by the experiment. The prediction results of training and testing samples obtained through GA-BP neural network algorithm analysis are shown in Figure 3, and samples are selected for display.

Figure 3 is a fitting diagram of the model and the training data. The dotted line is the original training set, and the solid line is the prediction result of the GA-BP model trained with the training set on each data in the training set. It can be seen that the prediction results can be better The trend of housing price changes is tracked, the prediction error is small, and the fitting effect is good.

It can be seen from Figure 3 that the GA-BP algorithm's prediction has a certain effect, but it is impossible to draw accurate conclusions only by observing the diagram. Therefore, a model evaluation algorithm needs to be used to evaluate the model. 4 evaluation methods such as MSE, MAE, EV, and R^2 are used to evaluate the training effect of the model. The error analysis details of the model evaluation are shown in Table 2.

It can be seen from Table 2 that the better model is the GA-BP model. When the model is used to fit the training data, the other three models are slightly less effective. When the GA-BP algorithm trains the model, it learns errors many times, so the prediction effect on the training data is better.

5. Conclusion

This article first selects the indicators that affect the price fluctuations of foreclosed houses, uses factor analysis technology to reduce the dimensionality of the original data, and obtains two common factors; then, combines the foreclosed house sample data with the method of GA-BP model to determine the The court system's foreclosure housing prices are predicted, and the predicted value is relatively close to the actual value. Moreover, by comparing the prediction results with multiple linear regression and extreme learning machine, it proves the superiority of this method in the accuracy of foreclosed house price prediction, which shows that the GA-BP neural network model is effective in the research of foreseeable house price prediction. A good choice, provides an effective method for the forecast of housing prices. [22].

Data Availability

The dataset can be accessed upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] P. Yue, R. Bai, and Y. Wang, "Review on second hand house price forecast," *Journal of Trend in Scientific Research and Development*, vol. 5, no. 1, 2020.

[2] K. Zhang and X. Zhu, "Analysis of Z city house price forecast based on application of multiple linear regression," in *Proceedings of the First International Symposium on Economics,Management,and Sustainable Development(EMSD 2019)*, pp. 74–77, Clausius Scientific Press, Hangzhou, China, 2019.

[3] Y. Liu and K. Li, "Research on house price forecast based on grey system GM(1, 1)," in *Proceedings of 2019 5th International Conference on Finance,investment,and Law(ICFIL 2019)*, pp. 208–214, Francis Academic Press, Hangzhou, China, 2019.

[4] J. Lai, L. Zhang, C. F. Duffield, and L. Aye, "Engineering reliability analysis in risk management framework: development and application in infrastructure project," *IAENG International Journal of Applied Mathematics*, vol. 43, no. 4, 2013.

[5] C. Pierdzioch, J. Christoph Rülke, and S. Georg, "House price forecasts, forecaster herding, and the recent crisis," *Italian Journal of Food Science*, vol. 1, no. 1, 2012.

[6] Y. Zhu, "Research on house price forecast based on multiple linear regression," *Basic and Clinical Pharmacology and Toxicology*, p. 128, 2021.

[7] J. Hu, Y. Wang, S. Wang, and A. Li, "Real estate warning based on BP neural network," in *Proceedings of the 2020 International Conference on Financial Economics and Investment Management*, Hangzhou, China, 2020.

[8] H. Oliver and R. Horst, "Euro area house prices and unconventional monetary policy surprises," *Economics Letters*, vol. 205, 2021.

[9] G. McGrevy Ryan and K. Sorensen, "A spatial model averaging approach to measuring house prices," *Journal of Spatial Econometrics*, vol. 2, no. 1, 2021.

[10] S. Fu, "Prediction and analysis of wuhan commercial housing prices based on grey prediction," *Modern Judaism*, vol. 10, no. 06, 2021.

[11] C. Zhou, "House price prediction using polynomial regression with Particle Swarm Optimization," *Journal of Physics: Conference Series*, vol. 1802, no. 3, 2021.

[12] Y. Sun, "Investigation on house price prediction with various gradient descent methods," *Journal of Physics: Conference Series*, vol. 1827, no. 1, 2021.

[13] A. R. Demong Nur, J. Lu, and K. Hussain Farookh, "An adaptive personalized property investment risk analysis method based on data-driven approach," *International Journal of Information Technology and Decision Making*, vol. 20, no. 02, 2021.

[14] N. H. Zulkifley, S. Abdul Rahman, N. H. Ubaidullah, and I. Ibrahim, "House price prediction using a machine learning model: a survey of literature," *International Journal of Modern Education and Computer Science*, vol. 12, no. 6, 2020.

[15] H. Zhang, K. Wang, M. Li, X. He, and R. Zhang, "House price prediction with an improved stack approach," *Journal of Physics: Conference Series*, vol. 1693, no. 1, 2020.

[16] Z. Wu, "Prediction of California house price based on multiple linear regression," *Academic Journal of Engineering and Technology Science*, vol. 3, no. 7.0, 2020.

[17] *UK House Price Index to return*, M2 Presswire, England, 2020.

[18] B. Nazemi and M. Rafiean, "Forecasting house prices in Iran using GMDH," *International Journal of Housing Markets and Analysis*, 2020.

[19] Y. Niu, "Linear regression model of house price in boston," *Science Discovery*, vol. 8, no. 3, 2020.

[20] R. Bala, K. Surya, T. Chandravas, and J Manikandan, "A machine learning based advanced house price prediction using logistic regression," *International Journal of Computer Application*, vol. 176, no. 28, 2020.

[21] M. Thamarai and S. P. Malarvizhi, "House price prediction modeling using machine learning[J]," *International Journal of Information Engineering and Electronic Business*, vol. 12, no. 2, 2020.

[22] A. N. Alfiyatin, R. E. Febrita, H. Taufiq, and W. F. Mahmudy, "Modeling house price prediction using regression analysis and particle swarm optimization case study: Malang, East Java, Indonesia," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 8, 2017.