

Article

Intelligent Forecast Model for Project Cost in Guangdong Province Based on GA-BP Neural Network

Changqing Li ^{1,2}, Yang Xiao ¹, Xiaofu Xu ¹, Zhuoyu Chen ¹, Haofeng Zheng ¹ and Huling Zhang ^{1,*}¹ College of Ocean Engineering and Energy, Guangdong Ocean University, Zhanjiang 524005, China² Guangdong Provincial Key Laboratory of Intelligent Equipment for South China Sea Marine Ranching, Guangdong Ocean University, Zhanjiang 524088, China

* Correspondence: educauniversity@126.com

Abstract: Project cost forecasting is a complex and critical process, and it is of paramount importance for the successful implementation of engineering projects. Accurately forecasting project costs can help project managers and relevant decision-makers make informed decisions, thereby avoiding unnecessary cost overruns and time delays. Furthermore, accurately forecasting project costs can make important contributions to better controlling engineering costs, optimizing resource allocation, and reducing project risks. To establish a high-precision cost forecasting model for construction projects in Guangdong Province, based on case data of construction projects in Guangdong Province, this paper first uses the Analytic Hierarchy Process (AHP) to obtain the characteristic parameters that affect project costs. Then, a neural network training and testing dataset is constructed, and a genetic algorithm (GA) is used to optimize the initial weights and biases of the neural network. The GA-BP neural network is used to establish a cost forecasting model for construction projects in Guangdong Province. Finally, by using parameter sensitivity analysis theory, the importance of the characteristic values that affect the project cost is ranked, and the optimal direction for controlling the project cost is obtained. The results showed: (1) The determination coefficient between the forecasting and actual values of the project cost forecasting model based on the BP neural network testing set is 0.87. After GA optimization, the determination coefficient between the forecasting and actual values of the GA-BP neural network testing set is 0.94. The accuracy of the intelligent forecast model for construction project cost in Guangdong Province has been significantly improved after optimization through GA. (2) Based on sensitivity analysis of neural network parameters, the most significant factor affecting the cost of construction projects in Guangdong Province is the number of above-ground floors, followed by the main structure type, foundation structure, above-ground building area, total building area, underground building area, fortification intensity, and building height. The results of parameter sensitivity analysis indicate the direction for cost control in construction projects. The research results of this paper provide theoretical guidance for cost control in construction projects.



Citation: Li, C.; Xiao, Y.; Xu, X.; Chen, Z.; Zheng, H.; Zhang, H. Intelligent Forecast Model for Project Cost in Guangdong Province Based on GA-BP Neural Network. *Buildings* **2024**, *14*, 3668. <https://doi.org/10.3390/buildings14113668>

Academic Editor: Gerardo Maria Mauro

Received: 12 October 2024

Revised: 13 November 2024

Accepted: 15 November 2024

Published: 18 November 2024

Keywords: project cost; bp neural network; genetic algorithm (GA); forecast model; parameter sensitivity analysis



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The accurate estimation of project cost is of great significance. In the project proposal and feasibility study stage, accurate estimation of project cost can provide a key basis for investment decision-making of the project. In the preliminary design and construction drawing design stages, accurate estimation of project cost is an important basis for the preparation of preliminary design estimates and construction drawing budgets. In the project construction stages, accurate estimation of project cost can better control project costs, optimize resource allocation, and reduce project risks. However, the current project cost estimation methods are no longer able to meet the accuracy requirements of project cost estimation, and the accuracy of project cost estimation has become a key factor limiting the successful completion of engineering.

At present, establishing project cost estimation models is a research hotspot in the field of construction, mainly involving various types of projects such as building construction [1–3], water conservancy and hydropower engineering [4,5], electric power engineering [6,7], railway engineering [8,9], highway engineering [10,11], bridge engineering [12,13], and tunnel engineering [14–16]. The commonly used research methods currently include time series analysis [17], support vector machine regression [18,19], random forest [20], Bayesian linear regression [21], neural networks [22–25], etc.

Ottaviani and Marco [26] developed a linear model to improve the accuracy of standard EAC and minimize the variance of errors. This study is conducted on the EVM dataset, which includes 29 real-life projects with a total of 805 observations. The results indicate that their proposed model has higher accuracy and lower variance compared to the standard formula. Inan et al. (2022) proposed a machine learning model based on long short-term memory to predict project costs. This model uses seven-dimensional feature vectors, including schedule and cost performance factors and their moving averages, as predictors. Based on the cost variation pattern during the training phase, they conducted 300 experiments during the testing phase to validate the model. The results indicate that their proposed model produces more accurate cost estimates compared to traditional models based on earned value management indices [27]. Altuncan and Vanhoucke proposed a new hybrid prediction model that utilizes input parameters from project scheduling and risk analysis literature to predict the final time and cost of a project. This hybrid method combines two well-known risk models. Among them, structural equation modeling is used to construct and validate a theoretical risk model to represent the known relationship between project indicators and project performance. Bayesian networks are used to train theoretical models using artificial project data from literature. Then, the model is validated through 33 empirical projects, and the results showed that their proposed risk model demonstrated significant advantages in predicting time and cost [28]. Marco et al. proposed a framework for estimating ongoing project costs based on trend and seasonal analysis of project cost performance using the Holt–Winters method. Their proposed framework has also been used for estimating project completion time. The results indicate that their proposed framework has high accuracy [29].

At present, there are mainly the following problems in the research on project cost forecasting models: (1) The current project cost forecasting models are difficult to effectively characterize the nonlinear relationship between the characteristic values of projects and project costs, resulting in low accuracy of project cost forecasting models; (2) Due to the difficulty in building a database that includes project cases from various regions around the world, the current project cost forecasting models have strong regional applicability, making it difficult to promote forecasting models. Guangdong is an important economic province in China with a large number of construction projects. All parties involved in the projects urgently need to establish an accurate, fast, and convenient project cost forecasting model for production guidance. So, it is urgent to establish a project cost forecasting model in Guangdong Province. Based on case data from construction projects in Guangdong Province, the Analytic Hierarchy Process (AHP) is used to obtain the characteristic parameters that affect project cost in this paper. Then, a neural network training and testing dataset is constructed, and a project cost forecast model is established using the BP neural network algorithm. Due to the susceptibility of BP neural networks to getting trapped in local minima during the computation process, which can result in reduced model accuracy, this paper employs the genetic algorithm (GA) to optimize the initial weights and biases of the neural network. This optimization enhances the accuracy and robustness of the project cost forecast model. Consequently, the project cost forecast model for Guangdong Province is established using a GA-BP neural network. Finally, by using parameter sensitivity analysis theory, the importance of the characteristic values that affect the project cost is ranked, and the optimal direction for controlling the project cost is obtained.

2. Methodology

2.1. AHP

AHP is a quantitative analysis method used for decision-making with multiple objectives, proposed by Thomas L. Saaty in 1970. The basic principle is to decompose complex decision-making problems into a hierarchical structure of goals, criteria, and solutions (factors). Then, through constructing a judgment matrix, it compares feature parameters, determines the importance of feature parameters, and calculates weights. Finally, conduct a comprehensive evaluation for each feature parameter. During this process, consistency checks need to be conducted to ensure the rationality and reliability of these judgments. The AHP can help decision-makers systematically analyze and compare various options, accurately determine the importance of each factor, and ultimately make reasonable decisions. The objective layer is the highest level of the AHP, typically consisting of only one factor that represents the overall goal or ultimate pursuit of the decision-making problem. In the AHP, the objective layer provides clear direction and guidance for the entire decision-making process. The criterion layer is located below the target layer and serves as a bridge connecting the target layer and the factor layer. The criteria layer includes various criteria or standards that need to be followed to achieve the overall goal. These criteria can be qualitative or quantitative, and together they form the basis for evaluating the factors under the target layer. The factor layer is the bottom layer of the AHP, which includes various specific factors that need to be considered to achieve the overall goal. In the AHP, by comparing each factor in the factor layer pairwise, a judgment matrix can be constructed to determine the weight of each factor on the criteria or objectives of the previous layer. The AHP structure is shown in Figure 1.

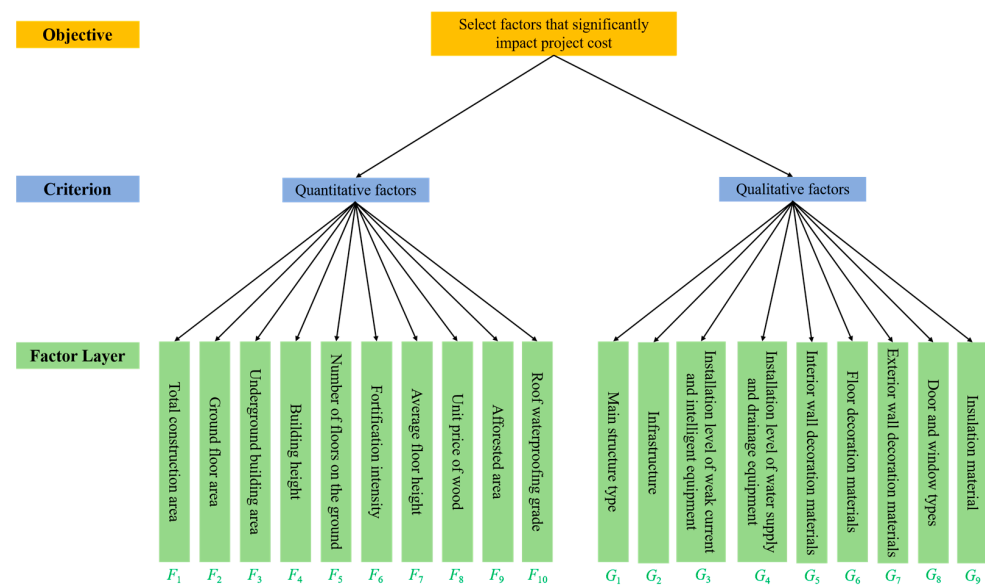


Figure 1. Hierarchical Analysis Structure Diagram.

2.2. BP Neural Network

Artificial Neural Network (ANN) is a computational model designed inspired by the human nervous system. Figure 2 shows a neural network structure with a single hidden layer, consisting of an input layer, a hidden layer, and an output layer. The training process of the BP neural network first requires the initialization of weight and bias parameters, which are randomly generated in the neural network. Then, the normalized feature parameters are inputted into the input layer, and a neural network is used for forward propagation to obtain the forecast output values. Compare the deviation between the forecast output value and the actual value and use the error back-propagation algorithm to adjust the weights and biases in reverse from the output layer to reduce the output error.

By repeatedly optimizing the weight and bias parameters, output errors can be reduced, and the accuracy and robustness of the forecast model can be improved.

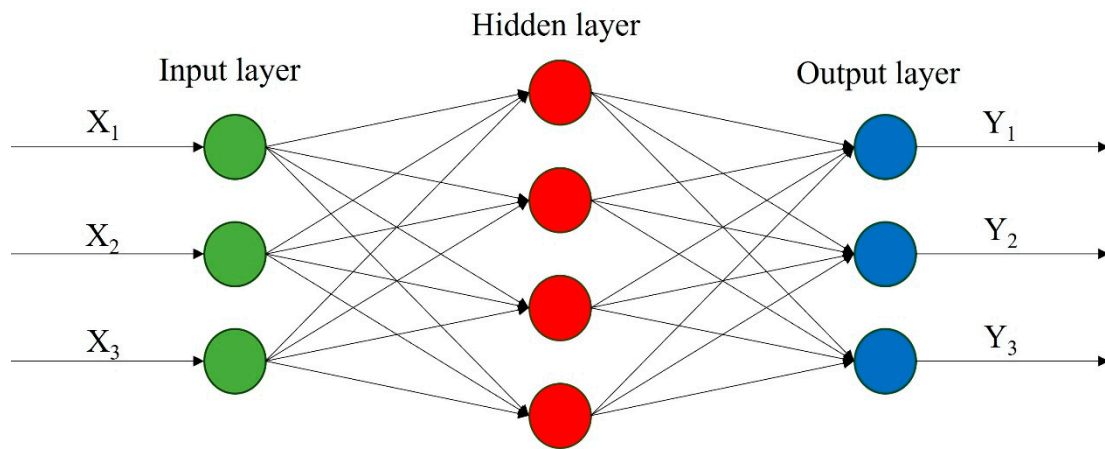


Figure 2. Neural Network Structure.

The calculation principle of a BP neural network is as follows:

$$Y = f_{\text{sig}} \left\{ b_0 + \sum_{k=1}^h [w_k f_{\text{sig}} (b_{ik} + \sum_{i=1}^m w_{ik} X_i)] \right\} \quad (1)$$

where b_0 and b_{ik} represent the bias values of the output layer and the k -th hidden layer neuron, respectively. w_k is the weight between the k -th hidden layer neuron and the output neuron, w_{ik} is the weight between the input variable and the hidden layer neuron k , X_i is the orthogonal value of the input variable i , and f_{sig} is the translation function.

Orthogonal the variable i to obtain X_i , and the orthogonalization principle is as follows:

$$X_i = 2 \times \frac{(i - i_{\min})}{(i_{\max} - i_{\min})} - 1 \quad (2)$$

The final forecast value (Y_p) is obtained by performing an arc cross transformation on Y :

$$Y_p = \frac{Y + 1}{2} \times (Y_{\max} - Y_{\min}) + Y_{\min} \quad (3)$$

2.3. GA

GA is an optimization algorithm based on natural selection and genetics principles. It simulates the process of biological evolution and gradually optimizes the quality of solutions by operations such as selection, crossover, and mutation. Figure 3 shows the flowchart of the GA. From Figure 3, it can be observed that the primary processes of a GA include setting parameters, generating initial population, calculating fitness, defining termination conditions, selection, crossover, mutation, generating new generation population, and outputting results. Setting parameters mainly includes setting population size, setting crossover probability, and setting mutation probability. The population size determines the number of individuals included in each generation of GA. A larger population can increase the breadth of the search space and improve the probability of the algorithm finding the global optimal solution. However, a large population can also increase computation time and resource costs. The crossover probability determines the probability of performing crossover operations in each generation. Cross operation generates new individuals by exchanging gene fragments of two individuals, which helps increase population diversity. A higher crossover probability can expand the search space, but it may also lead to the algorithm getting stuck in local optima. The mutation probability determines the probability of performing mutation operations in each generation. The mutation operation generates new individuals by randomly changing certain genes of individuals, which helps the algorithm

escape from local optima. However, excessively high mutation probability may lead to algorithm instability, while excessively low mutation probability may cause premature convergence of the algorithm. Generating an initial population provides a starting point for GA, which is a set of random initial solutions. Calculating fitness is to evaluate the quality of each solution and provide a basis for selecting operations. The termination condition is to determine whether the algorithm has achieved the expected goal or cannot continue to optimize, thereby ending the operation of the algorithm. Selection is the process of selecting outstanding individuals from the current group as parents to generate the next generation. Crossover is the process of generating offspring individuals with new characteristics by combining the excellent genes of the parent individual, thereby increasing the diversity of the population. Mutation is the random modification of an individual's genes with a certain probability, introducing new gene combinations that help break out of local optima and improve the algorithm's global search capability. Generating a new generation of population is the formation of a new population, providing a foundation for the next round of evolution. The output result provides the final result of the algorithm for users or subsequent analysis.

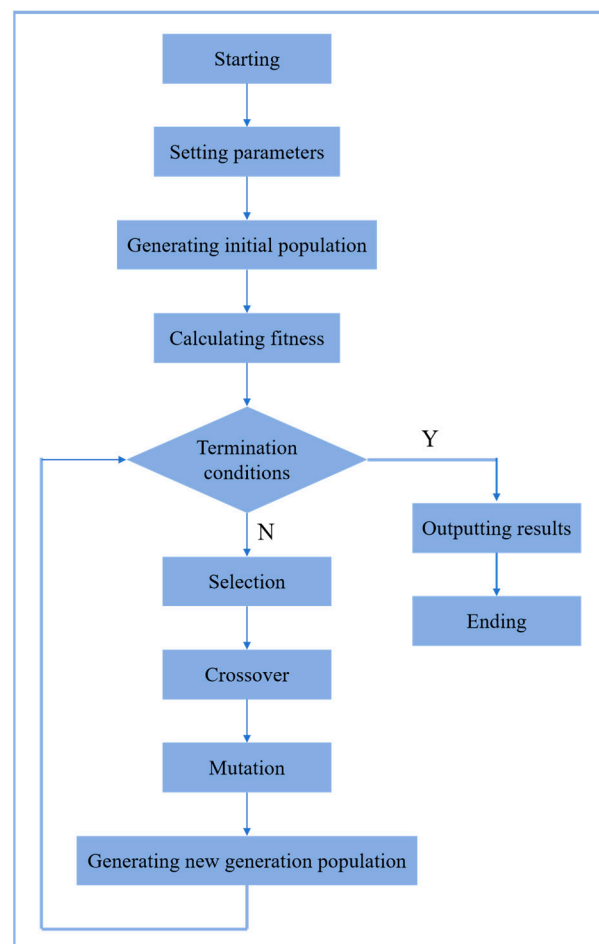


Figure 3. GA Flowchart.

2.4. GA-BP Neural Network

BP neural network is a deep learning model based on the backpropagation algorithm. In the process of establishing a forecast model by BP neural network, the main steps include data normalization, data segmentation, parameter initialization setting, model training, setting termination conditions, and updating w and b by the gradient descent algorithm. Data normalization aims to eliminate dimensional differences between different input parameters, keeping the data in the same order of magnitude, thereby improving the

training efficiency and forecast performance of neural networks. Data segmentation is the process of dividing a dataset into a training set and a testing set. The training set is used to train the neural network, while the testing set is used to evaluate the performance of the neural network. Parameter initialization setting is the process of setting hyper-parameters for a neural network and randomly generating its initial weights and biases. The initial weights and biases of a neural network directly affect its training effectiveness and predictive performance. Good initial weights and biases can accelerate the training process and improve training efficiency. Model training is the process of continuously adjusting the weights w and biases b of the neural network by forward information propagation and error backpropagation in order to minimize the error between forecasting and actual values. This is the core step of neural network learning, which gradually improves the performance of the neural network by continuous iteration and optimization. During the training process, it is necessary to set a termination condition to avoid overtraining or underfitting. The commonly used termination condition is that the RMSE is less than the predetermined value or reaches the maximum number of iterations. When the termination condition is met, the training process ends, and the resulting neural network model is the final forecast model. In the process of error backpropagation, the contribution of each neuron to the error (i.e., gradient) is calculated based on the error and loss function, and then the gradient descent method is used to update the weights w and bias values b of the neural network. This is a key step in neural network learning, which involves continuously updating weights and biases to gradually reduce the forecast error of the neural network and improve forecast performance.

The initial weights and biases of BP neural networks are usually randomly generated, which may cause the neural network to fall into local optima during training and fail to achieve global optima. GA is a population-based optimization method with strong global search capabilities. So, optimizing the BP neural network by GA can find better initial weights and biases, avoiding the BP neural network from getting stuck in local optima. The process of optimizing the initial weights and biases of the BP neural network by GA is shown in Figure 4. In Figure 4, the BP neural network passes the randomly generated initial weights and biases to the GA for encoding, thereby forming individuals in the GA. Then, the GA randomly generates a group of individuals based on the initial weights and biases as the initial population of the GA, evaluates the fitness of each individual in the initial population, selects excellent individuals for crossover and mutation operations based on the optimal fitness value, generates new individuals, iteratively executes the above steps until the stopping condition is met, and then decodes the optimal weights and biases from the optimized population and passes them to the BP neural network. Finally, the BP neural network establishes a project cost forecasting model based on the GA-optimized weights and biases.

The comparison of the advantages and disadvantages of GA-BP neural networks with other machine learning models is shown in Table 1. Establishing project cost forecast models based on GA-BP neural networks offers several advantages:

(1) Global Search and Optimization Capability

The global search capability of the genetic algorithm helps the BP neural network avoid getting stuck in local optimal solutions during training, thereby improving the model's prediction accuracy.

(2) Strong Robustness

The GA-BP neural network is robust to noisy data and outliers, making the model more stable and reliable when processing complex cost data.

(3) Learning and Approximation Abilities

The learning and approximation capabilities of the BP neural network enable the model to accurately capture the nonlinear relationships in cost data, thus improving prediction accuracy.

(4) Parameter Optimization

By optimizing the initial weights and biases of the BP neural network using the genetic algorithm, the model's training speed can be accelerated, enhancing prediction efficiency.

(5) Broad Applicability

GA-BP neural network models can be applied to different scales and types of construction projects, providing valuable support for project managers and decision-makers.

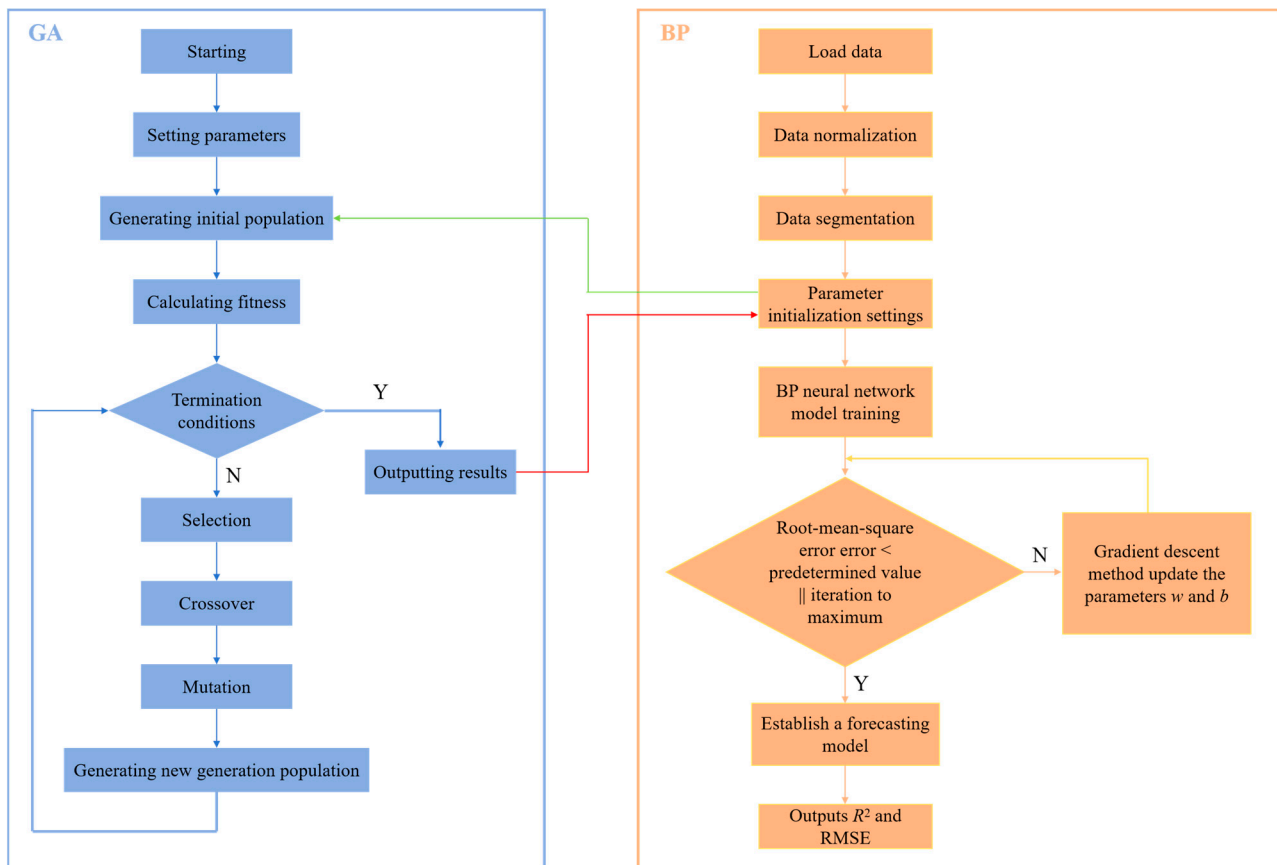


Figure 4. GA-BP neural network flowchart.

Table 1. The comparison of the advantages and disadvantages of GA-BP neural networks with other machine learning models.

Model	Advantages	Disadvantages
Support Vector Machine (SVM)	<ul style="list-style-type: none"> (1) Moderate computational cost, easy-to-interpret results. (2) Strong generalization ability, low generalization error rate. (3) Applicable to linear and nonlinear classification, as well as regression. 	<ul style="list-style-type: none"> (1) Sensitive to parameter tuning and kernel function selection. (2) Sensitive to missing data.
Decision Tree	<ul style="list-style-type: none"> (1) Conceptually simple, low computational complexity. (2) Highly interpretable, with easy-to-understand output results. (3) Broad applicability and strong extensibility. 	<ul style="list-style-type: none"> (1) Prone to overfitting. (2) Sensitive to imbalanced data.

Table 1. Cont.

Model	Advantages	Disadvantages
Naive Bayes	<ul style="list-style-type: none"> (1) Generative model, classifies by calculating probabilities. (2) Suitable for small-scale data and multi-class tasks. (3) Simple algorithm, convenient for incremental training. 	<ul style="list-style-type: none"> (1) Sensitive to the representation of input data. (2) Classification decisions have an inherent error rate.
BP Neural Network	<ul style="list-style-type: none"> (1) Simple implementation, low computational complexity, strong parallelism. (2) Has self-learning capabilities, capable of automatically extracting solution rules. (3) Possesses generalization and abstraction capabilities. 	<ul style="list-style-type: none"> (1) Vulnerable to local minima. (2) Lack of unified theoretical guidance for network structure selection. (3) Long training times, especially for complex problems.
GA-BP Neural Network	<ul style="list-style-type: none"> (1) Strong global search capability, aiding the neural network in better converging to the global optimal solution. (2) Good diversity, maintaining population diversity to avoid premature convergence. (3) Combines the learning and approximation capabilities of BP neural networks, improving the accuracy and efficiency of time series predictions. (4) Robust to noisy data and outliers. 	<ul style="list-style-type: none"> (1) Requires longer training times, especially for complex problems. (2) Complex parameter settings, such as population size, crossover probability, mutation probability, etc. (3) May get stuck in local optimal solutions, although genetic algorithms help mitigate this.

3. Characteristic Parameters of Project Cost Forecast Model

3.1. Factors Affecting the Project Cost

After extensively reviewing a large number of literature and books related to project cost, it is found that there are mainly 23 factors that affect project cost, as shown below: (1) total building area, (2) ground floor area, (3) underground building area, (4) average floor height, (5) building height, (6) unit price of wood, (7) roof waterproofing grade, (8) number of floors on the ground, (9) base structure, (10) pile foundation category, (11) main structure type, (12) seismic strength, (13) the green coverage, (14) earthwork processing, (15) seismic fortification intensity, (16) interior wall decoration materials, (17) exterior wall decoration materials, (18) concrete price, (19) installation level of water supply and drainage equipment, (20) installation level of weak current and intelligent equipment, (21) floor decoration materials, (22) types of door and window, and (23) insulation material. Based on the collected case data, we calculate the percentages of the 23 factors mentioned above in the project cost, respectively. We find that the following 19 factors account for a relatively high percentage of the project cost, namely: total construction area (F_1), ground floor area (F_2), underground building area (F_3), building height (F_4), number of floors on the ground (F_5), fortification intensity (F_6), average floor height (F_7), unit price of wood (F_8), afforested area (F_9), roof waterproofing grade (F_{10}), main structure type (G_1), infrastructure (G_2), installation level of weak current and intelligent equipment (G_3), installation level of water supply and drainage equipment (G_4), interior wall decoration materials (G_5), floor decoration materials (G_6), exterior wall decoration materials (G_7), door and window types (G_8), and insulation material (G_9). This indicates that these 19 factors are the main influencing factors. So, this study analyzes these 19 factors by the AHP.

3.2. Construction of Forecasting Model Indicator System

This paper selects the main structure type, total building area, above-ground building area, underground building area, foundation structure, building height, above-ground floors, and fortification intensity as input feature parameters for the neural network forecasting model. This section proposes the quantification principles for eight feature parameters, as follows:

(1) Total construction area

The total construction area refers to the total area of all floors of a building, including the area of all indoor spaces on the ground and the area of possible basements and attic spaces. The unit of total building area in this study is “ m^2 ”.

(2) Infrastructure

Infrastructure refers to the classification of basic structures used in construction projects to support and bear buildings. These basic structures are usually located between the building foundation and the building, serving as the load-bearing units between the two. The quantitative principles for this study are as follows: box foundation is 1, strip foundation is 2, independent foundation is 3, raft foundation is 4, and grid foundation is 5.

(3) Ground floor area

The above-ground building area refers to the total area of all parts of a building above-ground level, including the indoor space of all floors and possible other protruding parts such as bay windows or balconies. The unit of ground floor area in this study is “ m^2 ”.

(4) Underground building area

The underground construction area refers to the total construction area below the ground level of a building, including the sum of underground spaces such as basements and parking lots. The unit of underground building area in this study is “ m^2 ”.

(5) Main structure type

The main structural types include brick concrete structure, frame structure, shear wall structure, and frame shear structure. The quantitative principles for this study are as follows: the frame structure is quantified as 1, the frame shear structure as 2, the shear wall structure as 3, and the brick concrete structure as 4. Convert each structural type into a numerical value for comparison and analysis, and dimensionless methods can ensure accurate comparisons between these numerical values.

(6) Building height

Building height refers to the vertical distance from the ground or base of a building to its highest point. The unit of building height in this study is “ m ”.

(7) Number of floors on the ground

The number of floors above-ground refers to the number of floors above the ground level of a building, calculated from the ground level onwards. For example, if a building has 9 floors, quantify it as “9”.

(8) Fortification intensity

The fortification intensity refers to the degree of vibration intensity felt on the surface during an earthquake. The fortification intensity is usually divided into 12 levels according to the seismic intensity standard, represented by Roman numerals I to XII, with each level representing different seismic intensities and degrees of impact. For example, if the fortification intensity is level one, it is quantified as “1”.

3.3. Determine Characteristic Parameters Based on the AHP

(1) Determination of characteristic parameters

When establishing a project cost forecasting model based on a BP neural network, the selection of input feature parameters has a crucial impact on the performance of the forecasting model. Choosing too many input feature parameters may cause overfitting, low training efficiency, and high complexity in the forecast model. Choosing too few input feature parameters may result in underfitting and limited generalization ability of the forecast model. So, when establishing a forecast model based on a BP neural network, it is necessary to carefully select the number of input feature parameters. It is necessary

to ensure that the number of input features is sufficient to fully reflect the essence and complexity of the problem while avoiding problems such as overfitting and low training efficiency caused by too many input features. This paper uses the AHP to select input feature parameters for a neural network.

We have classified the 19 factors that have a significant impact on project cost into quantitative (Table 2) and qualitative (Table 3) factors. Quantitative indicators mainly include total construction area (F_1), ground floor area (F_2), underground building area (F_3), building height (F_4), number of floors on the ground (F_5), fortification intensity (F_6), average floor height (F_7), unit price of wood (F_8), afforested area (F_9), and roof waterproofing grade (F_{10}). Qualitative indicators mainly include main structure type (G_1), infrastructure (G_2), installation level of weak current and intelligent equipment (G_3), installation level of water supply and drainage equipment (G_4), interior wall decoration materials (G_5), floor decoration materials (G_6), exterior wall decoration materials (G_7), door and window types (G_8), and insulation material (G_9).

Table 2. Quantitative Hierarchical Structure Model.

Target Layer	Type Layer	Indicator Layer
Factors affecting project cost	Quantitative indicators	Total construction area (F_1)
		Ground floor area (F_2)
		Underground building area (F_3)
		Building height (F_4)
		Number of floors on the ground (F_5)
		Fortification intensity (F_6)
		Average floor height (F_7)
		Unit price of wood (F_8)
		Afforested area (F_9)
		Roof waterproofing grade (F_{10})

Table 3. Qualitative Hierarchical Structure Model.

Target Layer	Type Layer	Indicator Layer
Factors affecting project cost	Qualitative indicators	Main structure type (G_1)
		Infrastructure (G_2)
		Installation level of weak current and intelligent equipment (G_3)
		Installation level of water supply and drainage equipment (G_4)
		Interior wall decoration materials (G_5)
		Floor decoration materials (G_6)
		Exterior wall decoration materials (G_7)
		Door and window types (G_8)
		Insulation material (G_9)

(2) Establish a judgment matrix

After careful research and analysis, we have established a judgment matrix for quantitative and qualitative indicators, as shown in Tables 4 and 5.

Table 4. Quantitative indicator judgment matrix.

	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8	F_9	F_{10}
F_1	1.00	3.00	4.00	5.00	6.00	5.00	7.00	8.00	9.00	9.00
F_2	0.30	1.00	2.00	3.00	5.00	4.00	6.00	8.00	8.00	8.00
F_3	0.25	0.50	1.00	2.00	5.00	3.00	5.00	7.00	6.00	6.00
F_4	0.20	0.33	0.50	1.00	2.00	2.00	2.00	4.00	5.00	5.00
F_5	0.17	0.20	0.20	0.50	1.00	1.00	1.00	2.00	3.00	3.00
F_6	0.20	0.25	0.33	0.50	1.00	1.00	1.00	2.00	2.00	2.00
F_7	0.14	0.18	0.20	0.50	1.00	1.00	1.00	2.00	2.00	2.00
F_8	0.13	0.13	0.14	0.25	0.50	0.50	0.50	1.00	1.00	1.00
F_9	0.11	0.13	0.17	0.20	0.33	0.50	0.50	1.00	1.00	1.00
F_{10}	0.11	0.13	0.17	0.20	0.33	0.50	0.50	1.00	1.00	1.00

Table 5. Qualitative indicator judgment matrix.

	G ₁	G ₂	G ₃	G ₄	G ₅	G ₆	G ₇	G ₈	G ₉
G ₁	1.00	0.20	2.00	1.00	4.00	2.00	3.00	5.00	5.00
G ₂	5.00	1.00	7.00	5.00	8.00	7.00	9.00	8.00	8.00
G ₃	0.50	0.14	1.00	0.50	2.00	1.00	2.00	2.00	2.00
G ₄	1.00	0.20	2.00	1.00	4.00	2.00	3.00	5.00	5.00
G ₅	0.25	0.13	0.50	0.26	1.00	2.00	1.00	1.00	1.00
G ₆	0.50	0.14	1.00	0.50	0.50	1.00	2.00	2.00	2.00
G ₇	0.33	0.11	0.50	0.33	1.00	0.50	1.00	2.00	2.00
G ₈	0.20	0.13	0.50	0.20	1.00	0.50	0.50	1.00	1.00
G ₉	0.20	0.13	0.50	0.20	1.00	0.50	0.50	1.00	1.00

(3) Calculate the maximum eigenvalue and CI of quantitative indicators

According to the quantitative indicator judgment matrix, calculate the maximum eigenvalue and CI value of the quantitative indicator, as shown in Table 6. The normalization process is conducted on each column of the judgment matrix (Tables 4 and 5). Subsequently, the normalized matrix is summed row by row to obtain a new vector. This newly obtained vector is then normalized to derive the weight vector. By multiplying the weights of the scheme layer relative to the criterion layer by the weights of the criterion layer relative to the target layer, the combined weights of the scheme layer relative to the target layer can be obtained. Based on the judgment matrix, a characteristic polynomial $f(\lambda)$ can be constructed, where $f(\lambda) = |\lambda E - A|$, with A being the judgment matrix and E being the identity matrix. The characteristic equation is obtained by setting the characteristic polynomial equal to zero, i.e., $f(\lambda) = 0$. Solving the characteristic equation yields all the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, and the largest eigenvalue among them is the desired maximum eigenvalue (λ_{max}). For each eigenvalue λ_i , solving the linear system of equations $(A - \lambda_i E) \times x = 0$ yields the corresponding eigenvector α_i . Based on the maximum eigenvalue (λ_{max}) and the order (n) of the judgment matrix, CI can be computed using Equation (4).

Table 6. AHP Results of Quantitative Factors.

Eigenvalue	Eigenvector	Weight Value	Maximum Eigenvalue	CI
F ₁	3.256	32.559%		
F ₂	2.086	20.857%		
F ₃	1.549	15.495%		
F ₄	0.921	9.207%		
F ₅	0.512	5.119%		
F ₆	0.496	4.958%	10.292	0.032
F ₇	0.445	4.445%		
F ₈	0.255	2.546%		
F ₉	0.241	2.408%		
F ₁₀	0.241	2.408%		

(4) Consistency test of quantitative indicators

From Table 6, it can be seen that the weights corresponding to quantitative indicators 1–10 are 32.559%, 20.857%, 15.495%, 9.207%, 5.119%, 4.958%, 4.445%, 2.546%, 2.408%, and 2.408%, respectively. In order to ensure the rationality of the conclusions obtained using the AHP, it is necessary to conduct strict consistency checks on people's qualitative analysis judgments. The calculation formula for the consistency index (CI) is as follows:

$$CI = \frac{\lambda_{max} - n}{n - 1} \quad (4)$$

where λ_{max} is the maximum eigenvalue of the judgment matrix, n is the order of the judgment matrix (Dimension of the matrix). CI is a consistency indicator. If CI equals 0, it

indicates complete consistency; if CI is close to 0, it indicates satisfactory consistency; if CI is large, it indicates inconsistency. In addition, to verify whether the judgment matrix has satisfactory consistency, it is necessary to compare CI with RI together, that is, to test the coefficient CR:

$$CR = \frac{RI}{CI} \quad (5)$$

where RI is a random consistency index, which is a constant related to the order of the judgment matrix.

If the consistency ratio (CR) is less than 1.0, it is considered that the judgment matrix has passed the consistency test and is considered to have satisfactory consistency. If $CR > 1.0$, it is considered that the judgment matrix cannot pass the satisfactory consistency test and needs to be adjusted to ensure that it is more reasonable and consistent. According to the results of the quantitative indicator hierarchy analysis, the consistency test results are shown in Table 7.

Table 7. Summary of Consistency Test Results for Quantitative Analysis.

Maximum Eigenvalue	CI	RI	CR	Consistency Test Results
10.292	0.032	1.490	0.022	Pass

(5) Calculate the maximum eigenvalue and CI of qualitative indicators

Generally, the smaller the CR value, the better the consistency of the judgment matrix. If the CR value is less than 0.1, it is judged that the matrix satisfies the consistency test. However, if the CR value is greater than 0.1, it indicates that there is no consistency, and the judgment matrix should be adjusted appropriately before reanalysis. The calculated CI value for the 10th order judgment matrix is 0.032, and the RI value is 1.490 according to the table. Therefore, the calculated CR value is $0.022 < 0.1$, indicating that the judgment matrix in this study meets the consistency test and the calculated weights are consistent.

According to the qualitative indicator judgment matrix, calculate the maximum eigenvalue and CI value of the qualitative indicator, as shown in Table 8:

Table 8. AHP Results of Qualitative Factors.

Eigenvalue	Eigenvector	Weight Value	Maximum Eigenvalue	CI
G ₁	3.822	42.46%	9.357	0.045
G ₂	1.254	13.93%		
G ₃	1.254	13.93%		
G ₄	0.632	7.03%		
G ₅	0.566	6.29%		
G ₆	0.432	4.80%		
G ₇	0.427	4.75%		
G ₈	0.307	3.41%		
G ₉	0.307	3.41%		

From Table 8, it can be seen that the weights corresponding to qualitative indicators 1–9 are 42.46%, 13.93%, 13.93%, 7.03%, 6.29%, 4.80%, 4.75%, 3.41%, and 3.41%, respectively.

(6) Consistency testing of qualitative indicators

According to the results of the qualitative indicator hierarchy analysis, the consistency test results are shown in Table 9:

Table 9. Qualitative indicator hierarchy analysis results.

Maximum Eigenvalue	CI	RI	CR	Consistency Test Results
9.357	0.045	1.460	0.031	Pass

The calculated CI value for the ninth order judgment matrix is 0.045, and the RI value is 1.460 according to the table. The calculated CR value is $0.031 < 0.1$, indicating that the judgment matrix in this study meets the consistency test and the calculated weights are consistent with the standard.

(7) Determination of characteristic parameters

Figure 5 shows the weight calculation results of quantitative indicators based on the AHP. The numbers 1–10 in Figure 5 represent the total construction area, ground floor area, underground building area, building height, number of floors on the ground, fortification intensity, average floor height, unit price of wood, afforested area, and roof waterproofing grade. Among the quantitative factors, the total construction area has the highest weight, with a weight value of 32.56%, followed by the ground floor area (20.86%), underground building area (15.49%), building height (9.21%), number of floors on the ground (5.12%), fortification intensity (4.96%), average floor height (4.45%), unit price of wood (2.25%), afforested area (2.41%), and roof waterproofing grade (2.41%). Figure 6 shows the weight calculation results of qualitative indicators based on the AHP. The numbers 1–9 in the figure represent the main structure type, infrastructure, installation level of weak current and intelligent equipment, installation level of water supply and drainage equipment, interior wall decoration materials, floor decoration materials, exterior wall decoration materials, door and window types, and insulation material, respectively. Among the quantitative factors, the main structure type has the highest weight, with a weight value of 42.46%, followed by infrastructure (13.93%), installation level of weak current and intelligent equipment (13.93%), installation level of water supply and drainage equipment (7.03%), interior wall decoration materials (6.29%), floor decoration materials (4.80%), exterior wall decoration materials (4.75%), door and window types (3.41%), and insulation material (3.41%).

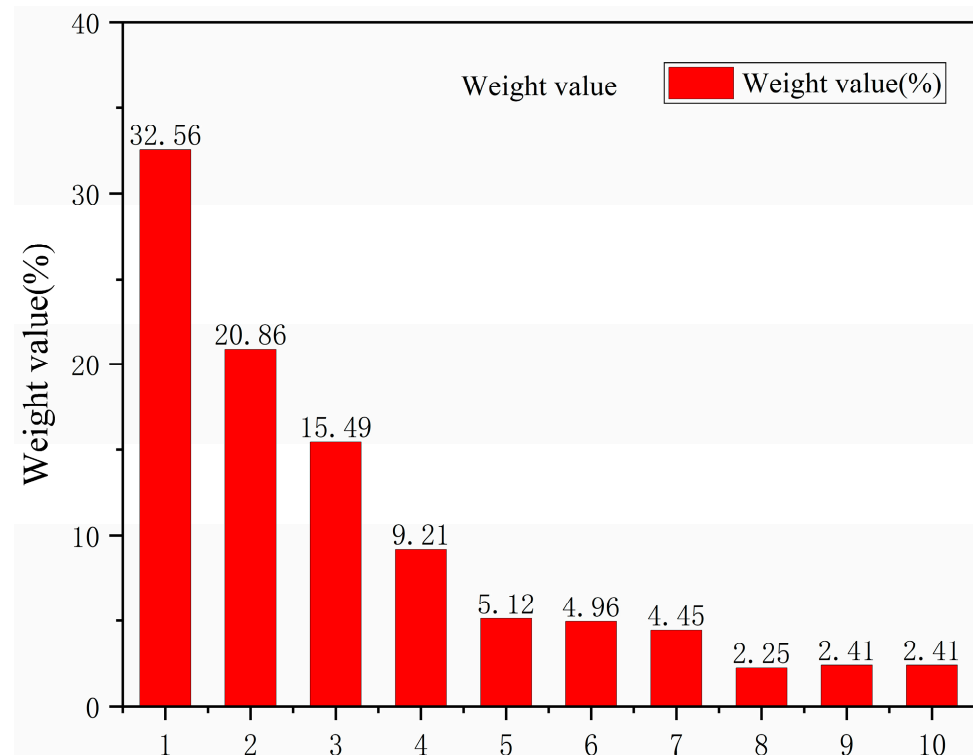


Figure 5. Quantitative Factor Weight Results.

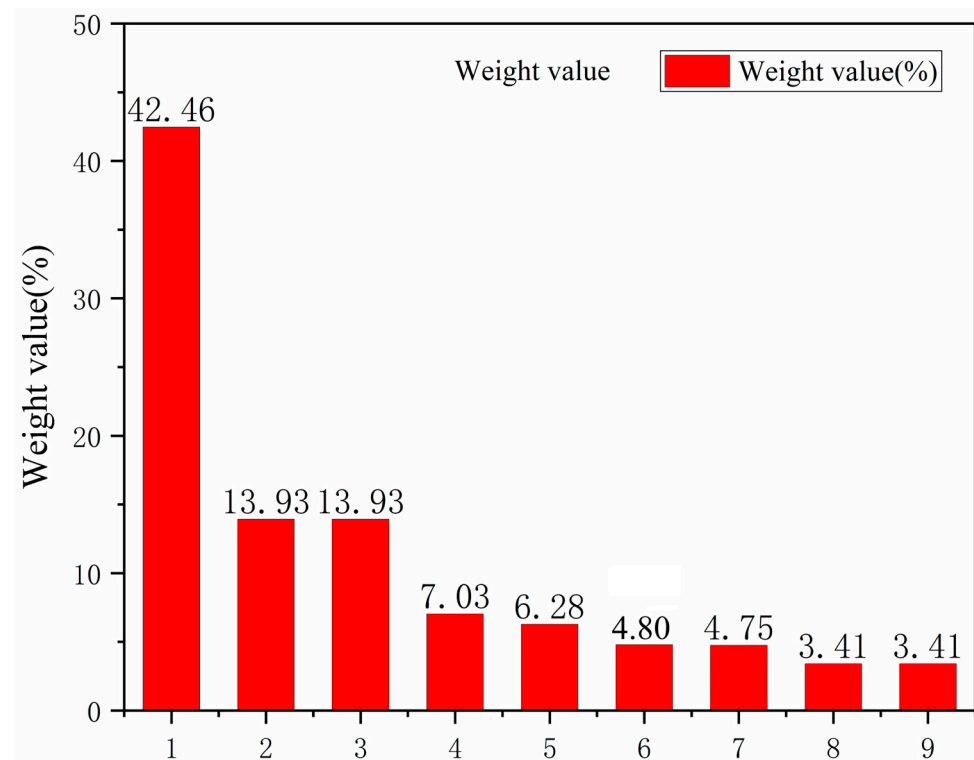


Figure 6. Qualitative Factor Weight Results.

When all influencing factors are considered in the project cost forecasting model, it may lead to increased model complexity, significantly heightened computational consumption, reduced model interpretability, and potentially cause overfitting on the training data, ultimately leading to a decrease in the accuracy of the forecasting model. On the other hand, if fewer influencing factors are taken into account in the project cost forecasting model, the model may fail to capture the characteristics of the relevant data and adequately reflect the mapping relationship between data, leading to underfitting and ultimately resulting in deteriorated model stability and reduced forecasting accuracy. On this basis, this project combines domestic and foreign research results and multiple experimental tests to select the first six quantitative influencing factors and the first two qualitative influencing factors as input feature parameters for the neural network. There are a total of eight feature parameters, namely: total building area, above-ground building area, underground building area, building height, above-ground floors, fortification intensity, and main structure type and foundation structure.

4. Project Cost Forecasting Model Based on BP Neural Network

4.1. Model Parameters Determination

The BP neural network is a deep learning algorithm that is primarily composed of an input layer, hidden layers, and an output layer.

(1) Input layer

The input layer is a part of the neural network structure, mainly responsible for receiving information and transmitting it to the next layer of neurons. Based on the analysis results in the previous section, eight factors, including main structure type, total building area, above-ground building area, underground building area, foundation structure, building height, above-ground floors, and fortification intensity, are used as input feature parameters for the neural network.

(2) Hidden layer

The significance of setting hidden layers in BP neural networks is to introduce nonlinearity, thereby increasing the nonlinear fitting ability of the entire network and enabling it to handle complex nonlinear problems. The hidden layer maps input data to output data through nonlinear transformation and feature extraction, enabling the model to learn complex relationships between data. Usually, setting up a hidden layer in BP neural networks can provide sufficient nonlinear expression ability. In addition, setting up a hidden layer can also improve training efficiency and model stability. In a single hidden layer BP neural network, the number of hidden layer neurons has a significant impact on the performance and effectiveness of the network. Setting too few hidden layer neurons may lead to insufficient network learning ability, difficulty in convergence during training, and poor model generalization ability. Setting too many hidden layer neurons may increase training time, difficulty in model interpretation and application, and increase the risk of overfitting. Therefore, choosing the appropriate number of hidden layer neurons is crucial for the model. The commonly used equation for determining the number of hidden layer neurons is as follows:

$$h = \sqrt{n + m} + a \quad (6)$$

where n represents the number of neurons in the input layer, m represents the number of neurons in the output layer, and a represents a random number between [1,10]. The number of hidden layer neurons in this paper is 6.

(3) Output layer

The output layer of a BP neural network is the last layer of the network, which plays a crucial role in the testing and training process of the model. The output layer is responsible for converting the network's calculation results into actual outputs, which can be classification labels, regression values, or other forms of forecast results. The output layer of this paper is the project cost. The paper collected 35 sets of case data from Guangdong Province from 2018 to 2023, and the principle of data preprocessing (normalization) is shown in Equation (2).

4.2. Project Cost Forecasting Model

Based on the training dataset, a project cost forecasting model is established using a BP neural network. The parameters of the BP neural network model are shown in Tables 10 and 11, and the BP neural network model is shown in Equation (7). Table 10 shows the weights between each input parameter and each hidden layer. In Table 10, neuron No. 1 represents the first hidden neuron. . . , neuron No. 6 represents the sixth hidden neuron. i represents the input parameters of the neural network; $i = 1$ represents the total construction area, $i = 2$ represents the infrastructure, and $i = 8$ represents the fortification intensity. w_{ik} represents the weight between the i -th input parameter and the k -th hidden neuron. Table 11 shows the bias values (b_{ik}) between each input parameter and each hidden layer, as well as the weight values (ω_k) and bias values (b_0) between each hidden layer and the output layer. b_{ik} represents the bias between the i -th input parameter and the k -th hidden neuron, ω_k represents the weight between the k -th hidden neuron and the output layer, and b_0 represents the bias between the hidden layer and the output layer.

Table 10. w_{ik} Calculation results.

	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$
Neuron No.1	−0.9583	−2.5885	−2.8053	−0.4698	2.3617	−1.8229	1.1068	1.0600
Neuron No. 2	−1.6767	1.1545	−0.5958	0.6207	0.8957	0.2246	1.9851	2.2353
Neuron No. 3	2.5474	1.0783	0.7598	1.5008	2.1600	−0.3604	−0.6380	−2.7527
Neuron No. 4	−0.2836	−1.3712	0.3448	2.8152	0.4663	−1.4571	−2.0033	−2.0745
Neuron No. 5	−2.5885	0.0750	−1.0985	1.4231	2.9844	2.6082	−0.1775	1.7775
Neuron No. 6	1.2332	−1.3970	0.5737	1.9075	−0.2269	−1.4624	−1.3009	−0.0294

$$\begin{aligned}
 Y = f_{\text{sig}} \left\{ \right. & \left[\begin{array}{l} -0.5288 \times f_{\text{sig}} \\ + 0.0321 \times f_{\text{sig}} \\ + 0.7480 \times f_{\text{sig}} \\ - 1.2799 \times f_{\text{sig}} \\ - 0.7199 \times f_{\text{sig}} \\ + 0.6307 \times f_{\text{sig}} \end{array} \right] \left[\begin{array}{l} [X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8] \\ [X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8] \\ [X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8] \\ [X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8] \\ [X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8] \\ [X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8] \end{array} \right] \left[\begin{array}{l} \begin{bmatrix} -0.9583 \\ -2.5885 \\ -2.8053 \\ -0.4698 \\ 2.3617 \\ -1.8229 \\ 1.1068 \\ 1.1068 \end{bmatrix} \\ \begin{bmatrix} -1.6767 \\ 1.1545 \\ -0.5958 \\ 0.6207 \\ 0.8957 \\ 0.2246 \\ 1.9851 \\ 2.2353 \end{bmatrix} \\ \begin{bmatrix} 2.5474 \\ 1.0783 \\ 0.7598 \\ 1.5008 \\ 2.1600 \\ -0.3604 \\ -0.6380 \\ -2.7527 \end{bmatrix} \\ \begin{bmatrix} -0.2836 \\ -1.3712 \\ 0.3448 \\ 2.8152 \\ 0.4663 \\ -1.4571 \\ -2.0033 \\ -2.0745 \end{bmatrix} \\ \begin{bmatrix} -2.5885 \\ 0.0750 \\ -1.0985 \\ 1.4231 \\ 2.9844 \\ 2.6082 \\ -0.1775 \\ 1.7775 \end{bmatrix} \\ \begin{bmatrix} 1.2332 \\ -1.3970 \\ 0.5737 \\ 1.9075 \\ -0.2269 \\ -1.4624 \\ -1.3009 \\ -0.0294 \end{bmatrix} \end{array} \right] \left. \begin{array}{l} - 2.4507 \\ + 1.8967 \\ - 0.2131 \\ - 0.3955 \\ - 0.4562 \\ + 2.7449 \end{array} \right\} + 0.0322 \quad (7)
 \end{aligned}$$

The coefficient of determination (R^2), also known as the determination coefficient or determinacy index, is commonly used to measure the degree of fit between a model and data [30]. Its value is between 0 and 1. When $R^2 = 1$, it indicates that the model completely matches the data. When $R^2 = 0$, it indicates that the model cannot explain any changes in the data. The closer the R^2 value is to 1, the higher the accuracy of the model's forecast

results, that is, the stronger the explanatory power of the independent variables of the function on the dependent variable. The calculation formula for R^2 is as follows:

$$R^2 = 1 - \frac{SSR}{SST} \quad (8)$$

where SSR represents the model fitting error, and SST represents the overall data dispersion.

Table 11. b_{ik} , ω_k , b_0 Calculation results.

	b_{ik}		ω_k	b_0
Neuron No. 1	−2.4507	Neuron No. 1	−0.5288	
Neuron No. 2	1.8967	Neuron No. 2	0.0321	
Neuron No. 3	0.2131	Neuron No. 3	0.7480	
Neuron No. 4	−0.3955	Neuron No. 4	−1.2799	0.0322
Neuron No. 5	−0.4562	Neuron No. 5	−0.7199	
Neuron No. 6	2.7449	Neuron No. 6	0.6307	

Based on the training dataset, a cost forecasting model is established using a BP neural network. The differences between the forecasting values and the actual values are compared, and these differences are quantitatively evaluated using the coefficient of determination (R^2). Figures 7 and 8 show the comparison between the forecasting and actual values of the training set and the testing set, respectively. In Figures 7 and 8, the horizontal axis represents the sample number, and the vertical axis represents the project cost. The coefficient of determination between the forecasting and actual values of the training set is 0.84, and the coefficient of determination between the forecasting and actual values of the testing set is 0.87. The coefficients of determination for both the training and testing sets are above 0.8, indicating a good fit between the project cost forecasting model and the case data. Root mean square error (RMSE) can quantitatively evaluate the difference between forecasting values and actual values. The smaller the RMSE, the higher the forecast accuracy of the model. So, we also calculate the RMSE in Figures 7 and 8, and the calculated results are 1590, 165, and 1907, 203, respectively. The calculation results of the RMSE for both the training set and the testing set are relatively high, indicating that the project cost forecast model established based on the BP neural network has optimization potential. In summary, we can conclude that the project cost forecasting model based on the BP neural network has good accuracy and robustness, but there is still room for further improvement.

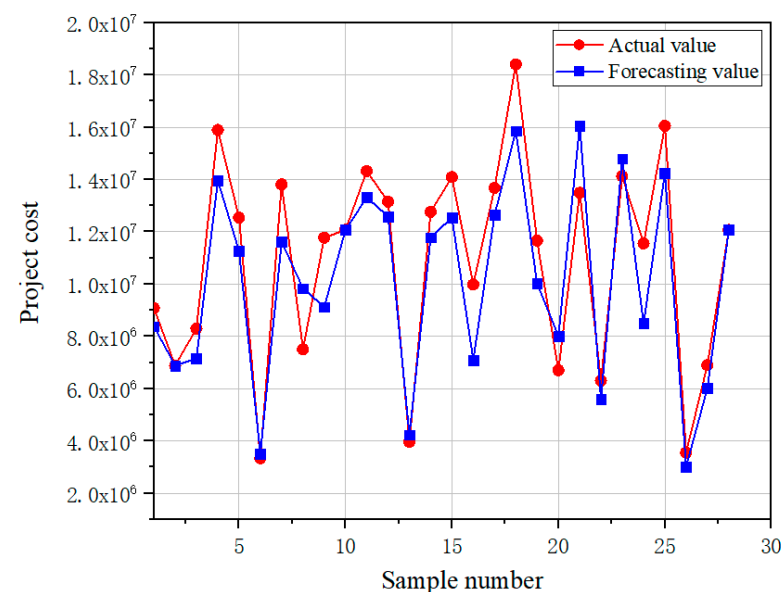


Figure 7. Comparison between forecasting and actual values in the training set ($R^2 = 0.84$).

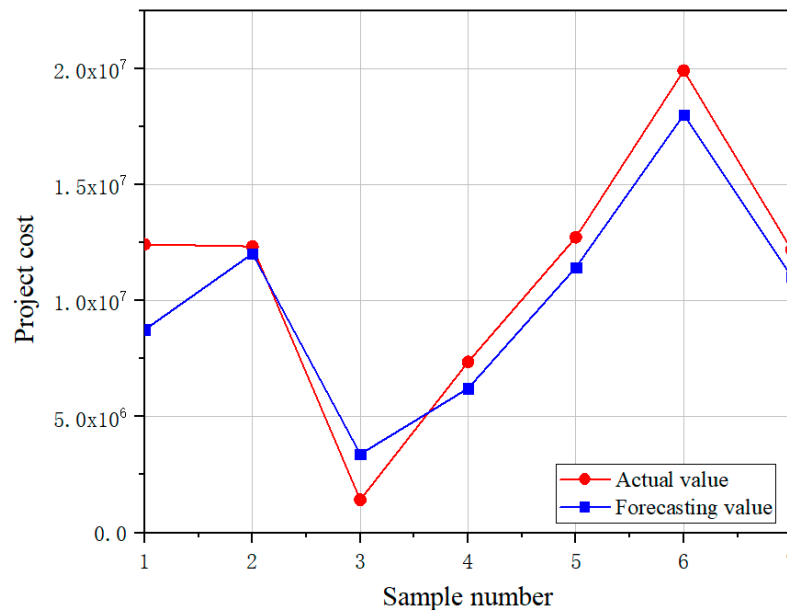


Figure 8. Comparison between forecasting and actual values in the test set ($R^2 = 0.87$).

5. Project Cost Forecasting Model Based on GA-BP Neural Network

Based on the training dataset, a project cost forecasting model is established using a GA-BP neural network. The parameters of the GA-BP neural network model are shown in Tables 12 and 13, and the BP neural network model is shown in Equation (9). Table 12 shows the weights between each input parameter and each hidden layer. In Table 12, neuron No. 1 represents the first hidden neuron. . . neuron No. 6 represents the sixth hidden neuron. i represents the input parameters of the neural network; $i = 1$ represents the total construction area, $i = 2$ represents the infrastructure, and $i = 8$ represents the fortification intensity. ω_{ik} represents the weight between the i -th input parameter and the k -th hidden neuron. Table 13 shows the bias values (b_{ik}) between each input parameter and each hidden layer, as well as the weight values (ω_k) and bias values (b_0) between each hidden layer and the output layer. b_{ik} represents the bias between the i -th input parameter and the k -th hidden neuron, ω_k represents the weight between the k -th hidden neuron and the output layer, and b_0 represents the bias between the hidden layer and the output layer.

Table 12. ω_{ik} result.

	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	
ω_{ik}	Neuron No. 1	1.0913	0.5400	0.0166	−0.3316	−0.8122	−0.8869	−0.5305	−1.4402
	Neuron No. 2	−1.3338	−0.4911	−0.5069	−0.2543	−0.6515	−0.4160	1.4377	−1.5021
	Neuron No. 3	−0.2722	0.2903	0.4347	0.7330	−1.5058	−0.3446	0.3022	0.5134
	Neuron No. 4	0.5409	−1.3090	−0.9199	−0.3177	−1.5802	0.1184	−1.0226	1.2999
	Neuron No. 5	−0.3503	−0.6654	0.5921	−0.6051	2.0075	−0.6208	−0.3136	0.1815
	Neuron No. 6	0.3658	0.6211	−0.5005	0.0679	−0.9905	0.2014	−0.8438	0.1362

Table 13. b_{ik}, ω_k, b_0 result.

	b_{ik}		ω_k	b_0
Neuron No. 1	−0.5123	Neuron No. 1	1.0577	0.7532
Neuron No. 2	−1.3311	Neuron No. 2	−1.3141	
Neuron No. 3	−0.2261	Neuron No. 3	1.2590	
Neuron No. 4	0.7719	Neuron No. 4	−1.3752	
Neuron No. 5	−0.6372	Neuron No. 5	−1.7507	
Neuron No. 6	1.3834	Neuron No. 6	0.0565	

The BP neural network model is shown in Equation (9).

$$\begin{aligned}
 Y = f_{sig} & \left\{ \left[\begin{aligned} & \left(\begin{aligned} & \left[\begin{array}{c} 1.0913 \\ 0.5400 \\ 0.0166 \\ -0.3316 \\ -0.8122 \\ -0.8869 \\ -0.5305 \\ -1.4402 \end{array} \right] \\ & [X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8] \end{aligned} \right) \\ & - 1.3141 \times f_{sig} \left(\begin{aligned} & \left[\begin{array}{c} -1.3338 \\ -0.4911 \\ -0.5069 \\ -0.2543 \\ -0.6515 \\ -0.4160 \\ 1.4377 \\ -1.5021 \end{array} \right] \\ & [X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8] \end{aligned} \right) \\ & + 1.2590 \times f_{sig} \left(\begin{aligned} & \left[\begin{array}{c} -0.2722 \\ 0.2903 \\ 0.4347 \\ 0.7330 \\ -1.5058 \\ -0.3446 \\ 0.3022 \\ 0.5134 \end{array} \right] \\ & [X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8] \end{aligned} \right) \\ & - 1.3752 \times f_{sig} \left(\begin{aligned} & \left[\begin{array}{c} 0.5409 \\ -1.3090 \\ -0.9199 \\ -0.3177 \\ -1.5802 \\ 0.1184 \\ -1.0226 \\ 1.2999 \end{array} \right] \\ & [X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8] \end{aligned} \right) \\ & - 1.7507 \times f_{sig} \left(\begin{aligned} & \left[\begin{array}{c} -0.3503 \\ -0.6654 \\ 0.5921 \\ -0.6051 \\ 2.0075 \\ -0.6208 \\ -0.3136 \\ 0.1815 \end{array} \right] \\ & [X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8] \end{aligned} \right) \\ & + 0.0565 \times f_{sig} \left(\begin{aligned} & \left[\begin{array}{c} 0.3658 \\ 0.6211 \\ -0.5005 \\ 0.0679 \\ -0.9905 \\ 0.2014 \\ -0.8438 \\ 0.1362 \end{array} \right] \\ & [X_1 X_2 X_3 X_4 X_5 X_6 X_7 X_8] \end{aligned} \right) \end{aligned} \right] + 0.7532 \right\} \quad (9)
 \end{aligned}$$

Based on the training dataset, a cost forecasting model is established using a GA-BP neural network. The differences between the forecasting values and the actual values are compared, and these differences are quantitatively evaluated using the coefficient of determination (R^2). Figures 9 and 10 show the comparison between the forecasting and

actual values of the training set and the testing set, respectively. In Figures 9 and 10, the horizontal axis represents the sample number, and the vertical axis represents the project cost. The coefficient of determination between the forecasting and actual values of the training set is 0.90, and the coefficient of determination between the forecasting and actual values of the testing set is 0.94. Compared to the determination coefficient of 0.84 for the training set and 0.87 for the testing set in the project cost forecast model based on the BP neural network, the determination coefficient of the project cost forecast model established by the GA-BP neural network has significantly improved. The calculation results of RMSE in Figures 9 and 10 are 1192, 036, and 1281, 422, respectively. Compared to the RMSE of 1590, 165 for the training set and 1907, 203 for the testing set in the project cost forecast model based on the BP neural network, the RMSE of the project cost forecast model established by the GA-BP neural network has significantly reduced. In summary, we can conclude that GA can significantly improve the accuracy and robustness of project cost forecasting models.

The computational resources employed in this study consist of a computer equipped with a 12th Gen Intel(R) Core (TM) i7-12700F processor operating at 2.10 GHz and 16.0 GB of RAM (with 15.8 GB available for use). This system runs on a 64-bit operating system based on an x64 processor. Regarding the time taken for training the model, the BP neural network-based cost forecasting model for construction projects in Guangdong Province was trained in 15 s. However, when the genetic algorithm was incorporated to optimize the initial weights and biases of the neural network (resulting in the GA-BP neural network), the training time increased to 38 s. In terms of trade-offs between model accuracy and computational cost, it is evident that the use of the genetic algorithm significantly improved the accuracy of the cost forecasting model. Specifically, the determination coefficient between the forecasting and actual values of the BP neural network testing set was 0.87, whereas after GA optimization, the determination coefficient of the GA-BP neural network testing set increased to 0.94. This improvement in model accuracy came at the cost of increased computational time, as mentioned earlier.

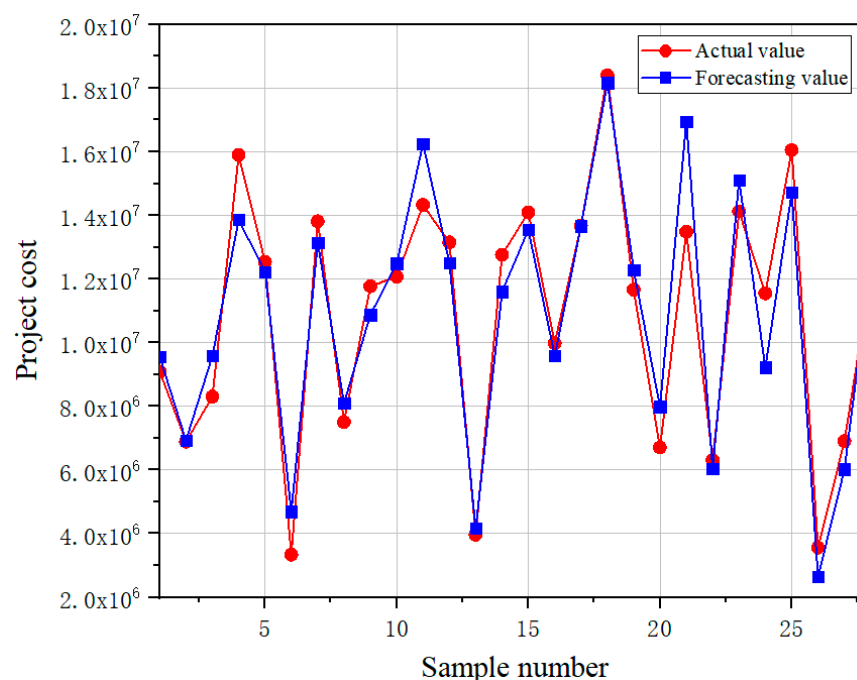


Figure 9. Comparison between forecasting and actual values in the training set ($R^2 = 0.90$).

However, it is worth noting that although the training time of the GA-BP neural network model is longer, a training time of 38 s is still acceptable for most practical applications, especially considering the significant improvement in model accuracy. In

addition, with the continuous advancement of computing technology, there may be more efficient algorithms and hardware in the future to further reduce training time while maintaining or improving model accuracy. So, while there is a trade-off between model accuracy and computational cost, the results of this study demonstrate that the use of a genetic algorithm for optimizing neural network weights and biases can yield significant improvements in model accuracy, making it a valuable tool for project cost forecasting in the construction industry. The computational resources required, and the training time involved are within reasonable limits given the current technological landscape.

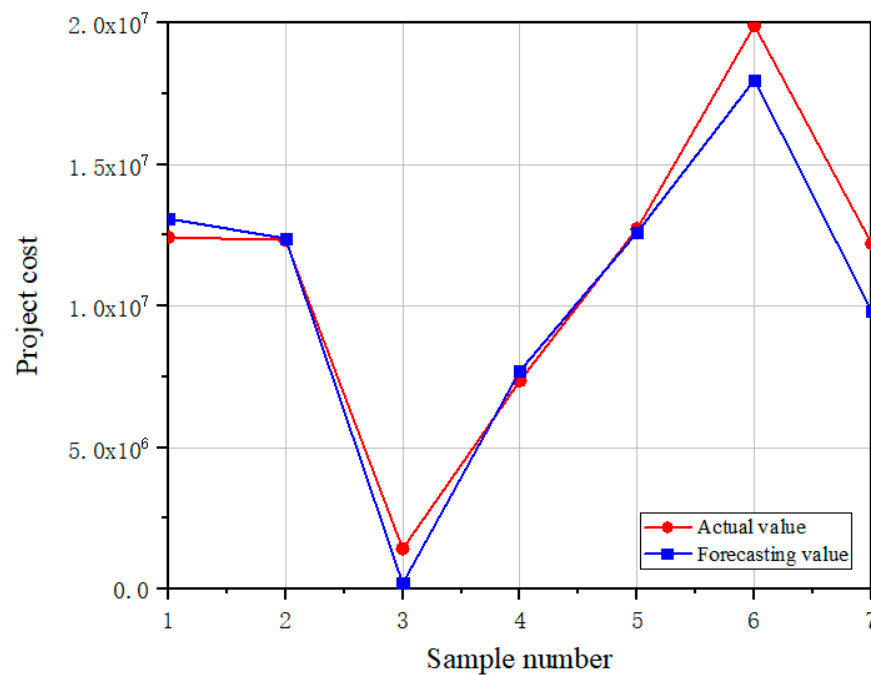


Figure 10. Comparison between forecasting and actual values in the test set ($R^2 = 0.94$).

6. Sensitivity Analysis of Characteristic Parameters in Project Cost Forecasting Model

6.1. The Significance of Parameter Sensitivity Analysis

Based on the sensitivity analysis theory of neural network parameters, a relative importance analysis is conducted on the eight characteristic input parameters of the neural network, namely the main structure type, total building area, above-ground building area, underground building area, foundation structure, building height, number of above-ground floors, and fortification intensity. Through sensitivity analysis theory, the key input parameters that have the greatest impact on project costs can be determined, namely the key factors. Key factors play an important role in project costs, so controlling these key factors during the construction process can better control project costs.

6.2. Basic Theory of Parameter Sensitivity Analysis

Parameter sensitivity analysis is a method used to evaluate how the output of a model is affected by changes in input parameters. Its basic principles mainly include the definition of parameter ranges, selection of sensitivity indicators, selection of appropriate analysis methods, execution of analysis, and interpretation of analysis results. Through sensitivity analysis, it is possible to determine which input parameters have a significant impact on the output results. This section is based on the weight and bias data of the GA-BP neural network and conducts sensitivity analysis on input parameters using the neural network parameter sensitivity analysis method proposed by Zhang and Goh [31]. The input parameters are the main structure type, total building area, above-ground building area, underground building area, foundation structure, building height, number of floors above-

ground, and fortification intensity. The main principles of relative importance analysis are as follows:

$$A_{.k} = \sum_{i=1}^5 |\omega_{ik} \times b_{ik}| \quad (10)$$

$$b_{ik} = \frac{|\omega_{ik} \times b_{ik}|}{A_{.k}} \quad (11)$$

$$C_{i.} = \sum_{k=1}^7 b_{ik} \quad (12)$$

$$D = \max(C_{i.}) \quad (13)$$

$$S_{i.} = \frac{C_{i.}}{D} \times 100\% \quad (14)$$

6.3. Parameter Sensitivity Analysis Results

Based on the basic theory of parameter sensitivity analysis, sensitivity analysis is conducted on the GA-BP neural network model of construction project cost in Guangdong Province. The sensitivity analysis results are shown in Figure 11. In Figure 11, 1 represents the number of floors above-ground, 2 represents the type of main structure, 3 represents the foundation structure, 4 represents the above-ground building area, 5 represents the total building area, 6 represents the underground building area, 7 represents the fortification intensity, and 8 represents the building height. It can be seen from Figure 11 that the most significant factor affecting the cost of construction projects in Guangdong Province is the number of above-ground floors, followed by the main structure type, foundation structure, above-ground building area, total building area, underground building area, fortification intensity, and building height.

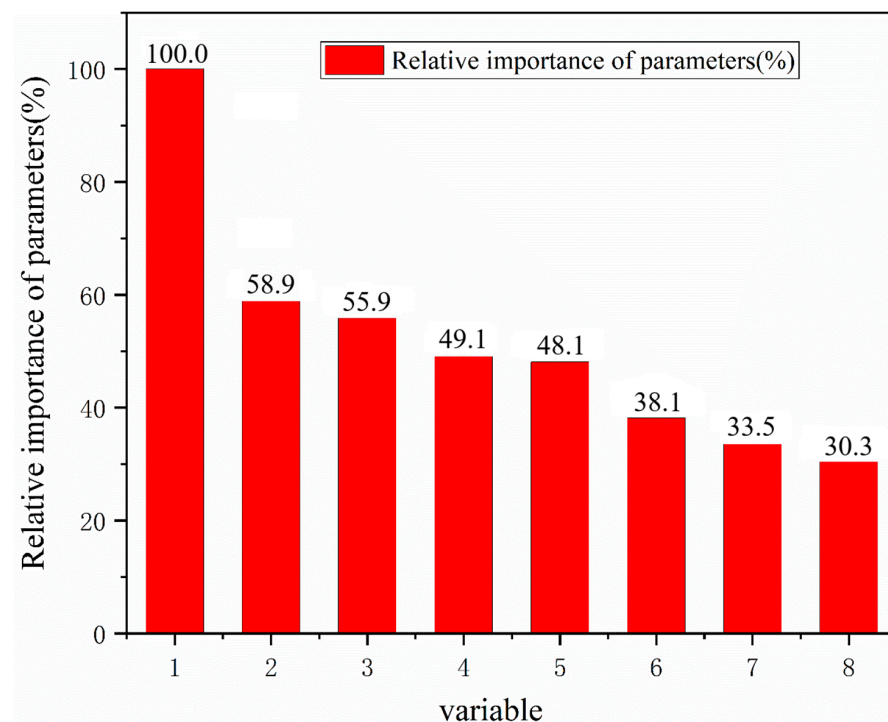


Figure 11. Bar Chart of Relative Importance of Feature Parameters.

7. Conclusions

This paper combines the Analytic Hierarchy Process, genetic algorithm, and BP neural network to establish a project cost forecasting model. The main conclusions of this study are as follows:

(1) Through literature review, this study found that the factors affecting project cost mainly include total building area, ground floor area, underground building area, average

floor height, building height, unit price of wood, roof waterproofing grade, number of floors on the ground, base structure, pile foundation category, main structure type, seismic strength, green coverage, earthwork processing, seismic fortification intensity, interior wall decoration materials, exterior wall decoration materials, concrete price, installation level of Water supply and drainage equipment, installation level of Weak current and intelligent equipment, floor decoration materials, types of door and window, insulation material. In order to establish a project cost forecasting model that is easy to use and promote, it is necessary to carefully select the above factors and discover the key influencing factors. This paper uses the AHP, a quantitative analysis method for multi-objective decision-making, to determine the importance of each influencing factor. The results show that the total building area, above-ground building area, underground building area, building height, number of floors above-ground, fortification intensity, main structure type, and foundation structure are key indicators that affect project cost.

(2) A neural network training and testing dataset is constructed using a case dataset. Based on the training set, a project cost forecasting model is established using the BP neural network. The results show that the coefficient of determination between the forecasting and actual values in the training set is 0.84, and the RMSE is 1590, 165. The determining factor between the forecasting and actual values in the testing set is 0.87, and the RMSE is 1907, 203. The above results indicate that the project cost forecasting model based on the BP neural network has good accuracy and robustness, but there is still room for further improvement. The initial weights and biases of the BP neural network are randomly generated, which may cause the network to fall into local optima during training and fail to achieve global optima. GA is a population-based optimization method with strong global search capability. So, optimizing the BP neural network by GA can find better initial weights and biases, avoiding the BP neural network from getting stuck in local optima. This paper establishes a project cost forecasting model using a GA-BP neural network. The results show that the coefficient of determination between the forecasting and actual values in the training set is 0.90, and the RMSE is 1192, 036. The coefficient of determination between the forecasting and actual values in the test set is 0.94, and the RMSE is 1281, 422. The accuracy and robustness of the neural network forecasting model have been significantly improved by GA optimization.

(3) Based on the parameter sensitivity analysis theory of neural networks, the relative importance of various factors affecting project cost can be quantitatively obtained. The results show that the most important factor affecting project cost is the number of above-ground floors (100%), followed by the main structure type (58.85%), foundation structure (55.87%), above-ground building area (49.10%), total building area (48.11%), underground building area (38.18%), fortification intensity (33.50%), and building height (30.38%). The results of parameter sensitivity analysis indicate the direction for controlling project costs.

The main contributions of this study are the establishment of a project cost forecasting model by integrating AHP, GA, and BP neural networks, as well as the quantitative assessment of the relative importance of various factors influencing project costs. The limitation of this study lies in the fact that the case data used in the paper are sourced from Guangdong Province, China. So, when assessing project costs in other countries or regions, an expansion of the dataset used in this paper is required, or alternatively, following the research approach of this paper, a local project cost forecasting model can be established based on local project case data. As an important component in the field of clean energy, offshore wind power has made significant progress in recent years and demonstrated broad development prospects. So, in the future, we will devote more energy to the field of offshore wind power, with the aim of establishing an intelligent forecasting model for the costs of offshore wind power projects.

Author Contributions: Conceptualization, C.L.; methodology, C.L.; software, X.X.; validation, X.X.; formal analysis, X.X.; investigation, C.L.; resources, X.X.; data curation, X.X., Z.C. and H.Z. (Haofeng Zheng); writing—original draft preparation, Y.X.; writing—review and editing, C.L. and H.Z. (Huiling Zhang); visualization, X.X.; supervision, H.Z. (Huiling Zhang); project administration, C.L.; funding

acquisition, H.Z. (Huilong Zhang) All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Fund of Guangdong Provincial Key Laboratory of Intelligent Equipment for South China Sea Marine Ranching (Grant NO. 2023B1212030003), the Non funded Science and Technology Research and Development Program of Zhanjiang City (Grant NO. 2024B01002), the program for scientific research start-up funds of Guangdong Ocean University (Grant NO. 060302072305), and the National Science Foundation of China (Nos. 42276070).

Data Availability Statement: All data used in the analysis are available on the website: <https://doi.org/10.6084/m9.figshare.27247551.v1>.

Conflicts of Interest: The authors declare that they have no conflict of interest in the work.

References

1. Ma, Z.; Liu, Z. BIM-based Intelligent Acquisition of Construction Information for Cost Estimation of Building Projects. *Procedia Eng.* **2014**, *85*, 358–367. [CrossRef]
2. Musarat, M.A.; Alaloul, W.S.; Liew, M.S. Incorporating inflation rate in construction projects cost: Forecasting model. *Heliyon* **2024**, *10*, e26037. [CrossRef] [PubMed]
3. Zhang, Y.F.; Mo, H.J. Intelligent building construction cost optimization and prediction by integrating BIM and elman neural network. *Heliyon* **2024**, *10*, e37525. [CrossRef] [PubMed]
4. Agarwal, S.S.; Kansal, M.L. Risk based initial cost assessment while planning a hydropower project. *Energy Strat. Rev.* **2020**, *31*, 100517. [CrossRef]
5. Butchers, J.; Williamson, S.; Booker, J.; Maitland, T.; Karki, P.B.; Pradhan, B.R.; Pradhan, S.R.; Gautam, B. Cost estimation of micro-hydropower equipment in Nepal. *Dev. Eng.* **2022**, *7*, 100097. [CrossRef]
6. Teegala, S.K.; Singal, S.K. Optimal costing of overhead power transmission lines using genetic algorithms. *Int. J. Electr. Power Energy Syst.* **2016**, *83*, 298–308. [CrossRef]
7. Hayati, D.W.; Chen, J.H.; Chen, Y.C.; Li, S.; Machsus, M.; Khoiri, M.; Wang, Q.-C.; Wei, H.-H. Smart cost estimation: Empirical case for extra-high voltage transmission towers. *Heliyon* **2024**, *10*, e31466. [CrossRef]
8. Pu, H.; Cai, L.; Song, T.; Schonfeld, P.; Hu, J. Minimizing costs and carbon emissions in railway alignment optimization: A bi-objective model. *Transp. Res. Part D Transp. Environ.* **2023**, *116*, 103615. [CrossRef]
9. Soleimani-Chamkhorami, K.; Garmabaki, A.H.S.; Strandberg, O.G. Life cycle cost assessment of railways infrastructure asset under climate change impacts. *Transp. Res. Part D Transp. Environ.* **2024**, *127*, 104072. [CrossRef]
10. Irfan, M.; Khurshid, R.B.; Anastasopoulos, P.; Labi, S.; Moavenzadeh, F. Planning-stage estimation of highway project duration on the basis of anticipated project cost, project type, and contract type. *Int. J. Proj. Manag.* **2011**, *29*, 78–92. [CrossRef]
11. Krishna, U.S.R.; Badiger, M.; Chaudhary, Y.; Gowri, T.V.; Devi, E.J. Optimizing Roads for Sustainability: Inverted Pavement Design with Life Cycle Cost Analysis and Carbon Footprint Estimation. *Int. J. Transp. Sci. Technol.* **2024**; in press.
12. Seo, J.; Park, H. Probabilistic seismic restoration cost estimation for transportation infrastructure portfolios with an emphasis on curved steel I-girder bridges. *Struct. Saf.* **2017**, *65*, 27–34. [CrossRef]
13. Zhou, G.; Etemadi, A.; Mardon, A. Machine learning-based cost predictive model for better operating expenditure estimations of U.S. light rail transit projects. *J. Public Transport.* **2022**, *24*, 100031. [CrossRef]
14. Rostami, J.; Sepehrmanesh, M.; Gharahbagh, E.A.; Mojtabei, N. Planning level tunnel cost estimation based on statistical analysis of historical data. *Tunn. Undergr. Space Technol.* **2013**, *33*, 22–33. [CrossRef]
15. Guan, Z.; Deng, T.; Jiang, Y.; Zhao, C.; Huang, H. Probabilistic estimation of ground condition and construction cost for mountain tunnels. *Tunn. Undergr. Space Technol.* **2014**, *42*, 175–183. [CrossRef]
16. Petrousatou, K.; Maravas, A.; Saramourtsis, A. A life cycle model for estimating road tunnel cost. *Tunn. Undergr. Space Technol.* **2021**, *111*, 103858. [CrossRef]
17. Elfahham, Y. Estimation and prediction of construction cost index using neural networks, time series, and regression. *Alex. Eng. J.* **2019**, *58*, 499–506. [CrossRef]
18. Chen, J.H.; Su, Y.M.; Hayati, D.W.; Wijatmiko, I.; Purnamasari, R. Improving Preliminary Cost Estimation in Indonesia Using Support Vector Regression. *Proc. Inst. Civ. Eng.-Manag. Procure. Law* **2019**, *172*, 25–33. [CrossRef]
19. Nourali, H.; Osanloo, M. Mining capital cost estimation using Support Vector Regression (SVR). *Resour. Policy* **2018**, *62*, 527–540. [CrossRef]
20. Wang, J.Z.; Zhang, J. Predicting in BIM Labour Cost with a hybrid approach Simple Linear Regression and Random Forest. *Earth Environ. Sci.* **2020**, *565*, 012108. [CrossRef]
21. Elmousalami, H.H.; Elshaboury, N.A.T.; Maxi, M.M.I.; Ibrahim, A.H.; Elyamany, A.H. Bayesian Optimized Ensemble Learning System for Predicting Conceptual Cost and Construction Duration of Irrigation Improvement Systems. *KSCE J. Civ. Eng.* **2024**, 100014.
22. Sonmez, R. Range estimation of construction costs using neural networks with bootstrap prediction intervals. *Expert Syst. Appl.* **2011**, *38*, 9913–9917. [CrossRef]

23. Yip, H.L.; Fan, H.; Chiang, Y.H. Predicting the maintenance cost of construction equipment: Comparison between general regression neural network and Box–Jenkins time series models. *Autom. Constr.* **2014**, *38*, 30–38. [[CrossRef](#)]
24. Wang, R.; Asghari, V.; Cheung, C.M.; Hsu, S.-C.; Lee, C.-J. Assessing effects of economic factors on construction cost estimation using deep neural networks. *Autom. Constr.* **2021**, *134*, 104080. [[CrossRef](#)]
25. Yang, G.; Xu, Y.; Huo, L.; Guo, D.; Wang, J.; Xia, S.; Liu, Y.; Liu, Q. Genetic algorithm optimized back propagation artificial neural network for a study on a wastewater treatment facility cost model. *Desalin. Water. Treat.* **2023**, *282*, 96–106. [[CrossRef](#)]
26. Ottaviani, F.M.; Marco, A.D. Multiple Linear Regression Model for Improved Project Cost Forecasting. *Procedia Comput. Sci.* **2022**, *196*, 808–815. [[CrossRef](#)]
27. İnan, T.; Narbaev, T.; Hazir, O. A Machine Learning Study to Enhance Project Cost Forecasting. *IFAC-PapersOnLine* **2022**, *55*, 3286–3291. [[CrossRef](#)]
28. Altuncan, I.U.; Vanhoucke, M. A hybrid forecasting model to predict the duration and cost performance of projects with Bayesian Networks. *Eur. J. Oper. Res.* **2024**, *315*, 511–527. [[CrossRef](#)]
29. Marco, A.D.; Ottaviani, F.M.; Bolognesi, F. Time series-based Project Cost Forecasting Framework. *Procedia Comput. Sci.* **2024**, *239*, 105–113. [[CrossRef](#)]
30. Nakagawa, S.; Johnson, P.C.D.; Schielzeth, H. The coefficient of determination R² and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *J. R. Soc. Interface* **2017**, *14*, 20170213. [[CrossRef](#)]
31. Zhang, W.; Goh, A.T.C. Assessment of Soil Liquefaction Based on Capacity Energy Concept and Back-Propagation Neural Networks. *Integr. Disaster Sci. Manag.* **2018**, 41–51.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.