

Research Article

Study on Foundation Pit Construction Cost Prediction Based on the Stacked Denoising Autoencoder

Lanjun Liu ^{1,2}, Denghui Liu ³, Han Wu ², and Junwu Wang ²

¹School of Civil Engineering and Architecture, Wuhan Institute of Technology, Wuhan 430070, China

²School of Civil Engineering and Architecture, Wuhan University of Technology, Wuhan 430070, China

³China Construction First Group Corporation Limited, Beijing 100161, China

Correspondence should be addressed to Han Wu; wuhan20170620@163.com

Received 28 September 2020; Revised 20 November 2020; Accepted 24 November 2020; Published 3 December 2020

Academic Editor: Jun Shen

Copyright © 2020 Lanjun Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To accurately predict the construction costs of foundation pit projects, a model based on the stacked denoising autoencoder (SDAE) is constructed in this work. The influencing factors of foundation pit project construction costs are identified from the four attributes of construction cost management, namely, engineering, the environment, the market, and management. Combined with Chinese national standards and the practice of foundation pit project management, a method of the quantization of the influencing factors is presented. 60 deep foundation pit projects in China are selected to obtain 13 main characteristic factors affecting these project construction cost by using the rough set. Then, considering the advantages of the SDAE in dealing with complex nonlinear problems, a prediction model of foundation pit project construction costs is created. Finally, this paper employs these 60 projects for a case analysis. The case study demonstrates that, compared with the actual construction costs, the calculation error of the proposed method is less than 3%, and the average error is only 1.54%. In addition, three error analysis tools commonly used in machine learning (the determination coefficient, root mean square error, and mean absolute error) emphasize that the calculation accuracy of the proposed method is notably higher than those of other methods (Chinese national code, the multivariate return method, the BP algorithm, the BP model optimized by the genetic algorithm, the support vector machine, and the RBF model). The relevant research results of this paper provide a useful reference for the prediction of the construction costs of foundation pit projects.

1. Introduction

For developing countries, construction engineering not only provides the necessary infrastructure for economic and social development but is also the main driving force of GDP growth [1]. The capital investment of construction projects is large, and the investment recovery period is long. At present, there is a substantial difference between the predicted and actual construction costs in the construction industry, which hinders the efficient development of the entire industry. Around the world, a large number of construction projects have failed due to the cost overruns [2–4]. Accurate prediction of construction cost in the early stage is one of the main bases for project decision-making and investment cost control, and its research is of great significance.

Measures such as retaining structure, groundwater control, and environmental protection needed to ensure the safety and stability of underground space formed by underground excavation are called foundation pit engineering. Foundation pit projects are important subprojects of construction engineering that are characterized by a long construction period and many influencing factors with complex relationships [5]. The construction cost of foundation pit engineering accounts for a large proportion of the total cost, which is often between 20% and 50%. Compared with other partial projects, foundation pit projects are more likely to occur in the likelihood of large differences between the predicted and actual costs [6]. Therefore, it is of great significance to quickly and accurately predict the construction costs of foundation pit projects.

There are obvious differences in the definition of construction cost in different national codes and different construction contracts [7, 8]. In order to facilitate the follow-up research, the construction cost in this paper was the actual construction cost of the construction project. The research object of this paper, the foundation pit project construction cost, referred to all expenses incurred by the contractor during the construction of the foundation pit project. Total expenses generally include two parts, project productive expenses and enterprise nonproductive expenses [8]. Nonproductive expenses are the expenses incurred by construction enterprises for organizing and managing production and business activities, which are often closely related to the contractor's project management ability, while the project productive expenses are mainly the actual expenses incurred in the construction site of foundation pit engineering.

From the perspective of research objects, the related research is mainly focused on the prediction of construction engineering costs, but there has been little research on the prediction of the costs of foundation pit engineering or other construction subprojects. Forcada et al. [9] selected several important influencing factors, including the project type, organization type, and contract type, and predicted the rework cost of construction projects via the regression analysis method. Wang and Dai [10] selected only six indexes, including the building area, number of floors, building height, and type of roof, to effectively predict the construction cost of a 15-story building. However, they did not consider the relevant factors of project management, and the numbers of floors and building structures of the test set and sample set in the case analysis were basically the same. In Wang and Dai's case study, the construction cost of the project could be considered to have an approximately positive correlation with the building area, which was an important reason why the construction cost was accurately predicted via the use of only six indicators. Williams and Gong [11] carried out the text and data mining of project management materials to determine the main factors that affect construction costs. Although this method provided a new idea for the selection of indicators, the average calculation accuracy was found to be only 47.11%. The possible reason for this unideal result was that the main factors obtained by this method originated from project management data, but how these factors affect the project cost required more explanation and analysis. These scholars often selected only a few main influencing factors, but it is difficult for this strategy to reflect the complexity of construction projects.

In actual engineering practice, a construction project often includes several subprojects that are highly professional and varied [12]. Therefore, the basis of the scientific and effective prediction of construction costs is the scientific and effective prediction of the construction costs of divisional and subdivisional projects. To the best of the authors' knowledge, research on the prediction of foundation pit project construction costs based on deep learning has not yet been reported. Therefore, foundation pit engineering, which is a typical subproject, was selected as the research object in the present study.

The acquisition of key influencing factors of foundation pit project construction cost is another content that needs serious study [13]. By the means of questionnaire survey and subjective experience of experts, Lesniak and Juszczak [14] determined that the project type, project geographical location, and construction period as key influencing factors and used the back propagation neural network (BP) to predict the indirect cost of the project. By the case-based reasoning and measurement similarity, Kim and Kim [15] determined the key influencing factors of construction cost. Dong et al. [16] interviewed eight experts in the field of engineering and construction and determined 16 indexes that affected the construction cost. Among the above typical research results, the selection of characteristic factors is mostly based on the empirical method, the proportional method, and questionnaire survey method. These traditional methods have the disadvantages of strong subjectivity and lack of scientificity. In addition, the influence of the selection of characteristic factors on the prediction results has not been considered in the above studies. However, the variable selection method based on rough set has achieved good results in recent years, which could effectively overcome these shortcomings of traditional methods [17–20]. Su et al. [17] used the rough set method to effectively screen the key risk factors of subsea tunnel construction. Case study showed that the risk index system obtained by the rough set was more scientific than these traditional weight calculation methods. Zhang et al. [18] obtained the key factors and objective weights of landslide risks in mountain tunnel construction by using the rough set method. Xu et al. [19] used the variable precision fuzzy rough set (VPFRS) to screen the evaluation index system of the synergy effect of main and auxiliary industries in power grid. Research showed that rough set could eliminate redundant indicators and retain key indicators and effectively improve the efficiency and accuracy of evaluation. Barbagallo et al. [20] used the rough set method in the Rose Package to effectively identify and further screen out the index system for water supply reservoir management.

From the perspective of research methods, adopting appropriate methods to reflect the nonlinear relationship among influencing factors of foundation pit project construction cost and quickly and accurately predicting is the key to build an estimation model of foundation pit project construction cost. Because the influencing factors of building project construction cost are complex and the data collection of construction cost is not easy, the prediction of building project construction cost is a typical high-dimensional nonlinear problem. Moreover, the rapid prediction of building project construction cost often serves the cost management of enterprises and requires higher accuracy and time of prediction. At present, quota method is often used in engineering practice, such as Chinese National Standard (Standard method of measurement for public utilities works, GB 50857-2013; Standard for classification and measurement of construction cost index, GB/T 51290-2018; Code of bills of quantities and valuation for construction works, GB 50500-2013) and the construction contract of International Federation of Consulting

Engineers (FIDIC). Relevant researchers have already proposed some mathematical methods by which to predict the costs of construction projects. Trost and Oberlender [21] conducted a multiple regression analysis of 11 factors to predict project costs. Considering the poor calculation accuracy of multiple regression analysis, a score estimation program was also developed to evaluate the prediction accuracy. To accurately estimate the costs of construction projects, Ji and Ahn [22] proposed a prediction method based on the scenario-planning method. The results of their case study revealed that the estimation accuracy was between 4.23% and 4.86%. Cheng et al. [23] proposed a cost forecasting method based on the grey prediction model. The construction process was divided into three typical stages according to the cost prediction accuracy to take into account the complexity of construction cost prediction. Although these cost prediction methods implemented in these studies achieved certain results, it was difficult for them to scientifically and effectively deal with small sample numbers and nonlinear problems. In addition, these studies all assumed that the prices of labor, materials, and machines were static factors that did not change with time; thus, there was a substantial gap with actual engineering practice, in which the price-related indicators are dynamic. Therefore, sufficient prediction could not be achieved. Recently, some scholars have also applied artificial neural networks (ANNs) to construction cost prediction. Wang [24] divided construction engineering costs into three categories, namely, construction costs, structure costs, and outdoor engineering costs and used the BP network to predict the costs of construction projects. The results of a case study showed that, although the calculation error of this algorithm was large, it met the requirements of engineering practice. Gunduz et al. [25], respectively, used multiple regression analysis and the BP network, which is the most common ANN method, to predict the early costs of light rail transit and subway engineering development. In their research, 17 key factors that affect costs were selected, and 16 project datasets were selected as sample sets. The error of the multiple regression analysis was 2.32%, which was notably less than the error of the ANN (5.76%). The reason that the BP calculation accuracy of [24, 25] was not high, might be that the traditional BP model easily falls into the local extremum and has a diverse network structure [26].

Deep learning was developed from the traditional multilayer neural network, which has excellent nonlinear mapping and generalization abilities to represent complex high-dimensional functions [27]. The difference between the traditional multilayer neural network and deep learning mainly lies in the different training methods. Traditional ANNs are trained by supervised learning, while most deep learning algorithms combine unsupervised feature learning with supervised learning [28]. The denoising autoencoder (DAE) is a common deep learning network [29] that has two main forms. One is the stacked denoising autoencoder (SDAE), which has obvious advantages in dealing with high-dimensional nonlinear problems, and the other is the sparse encoder. When a neural network is used to classify and predict a large number of influencing factors of construction

costs, a lot of noise is easily produced due to the data of the influencing factors. To avoid the interference of noisy data as much as possible [30], the SDAE was selected in the present research to study the prediction of construction engineering costs.

Liu et al. [31] accurately predicted short-term power load by the SDAE method. The case study showed that SDAE had higher calculation accuracy than the BP and the DAE. Dai et al. [32] applied the SDAE model to data processing of dissolved gas concentration in transformer oil and transmission line temperature. The results showed that this method could effectively identify and repair outliers and missing information. However, the performance of SDAE algorithm was not further analysed in that paper. Dong et al. [33] used the SDAE to predict the short-term wind speed in wind power generation. The results of case study showed that SDAE had higher computational accuracy and faster computational efficiency than the artificial intelligence algorithms such as the BP. Chen et al. [34] extracted hyperspectral image features by the SDAE method. Compared with common methods such as the support vector machine (SVM), SDAE had better computing power. In order to deal with the variability and nonlinear correlation in the prediction of regional sharp power generation, Yan et al. [35] used the SDAE method to predict the regional wind power generation. Compared with other common wind power generation forecasting methods, SDAE had the highest forecasting authenticity.

By summarizing the existing research work, the following key issues deserve study in this paper. (1) Most of the current research results took architectural project construction cost as the research object. The research object was not meticulous enough, which led to the rough selection of influencing factors and large prediction error. It is a possible way to improve the construction cost prediction by selecting some subprojects such as foundation pit engineering for construction cost prediction. (2) In the current related research, the selection of characteristic factors was mostly based on the empirical method or the proportional method, and these selection methods were subjective and unscientific. These studies also did not consider the influence of the selection of characteristic factors on the selection of forecasting methods. (3) At present, the regression analysis, the grey theory, and the traditional artificial neural network are often used to predict the construction cost, but their prediction accuracies were not high and they took a long time. The SDAE based on deep learning network has strong learning and prediction ability. Training and learning sample data by establishing prediction model provides a new idea for accurately and quickly forecasting the construction cost of nonlinear engineering projects.

Based on the above analysis, this paper constructed the construction cost prediction model of foundation pit engineering by the SDAE. The main contributions of this paper are as follows. (1) According to the characteristics of the construction content and project management of foundation pit engineering, a more comprehensive system of the determination of influencing factors is constructed, and a corresponding quantification and normalization process is

put forward. This provides a research basis for the subsequent construction cost prediction of foundation pit engineering. (2) In order to overcome the shortcomings of strong subjectivity and lack of scientificity in traditional index selection methods, this paper uses rough set, a typical quantitative analysis method, to obtain the main characteristic factors affecting the project construction cost of foundation pit. In this paper, 60 deep foundation pit projects in Hubei Province of China are selected as cases, and 13 main characteristic factors affecting these project construction cost are screened out. In addition, the influence of key factors on prediction results is also preliminarily analysed. (3) The SDAE based on a deep learning network is selected, and a prediction model is established to train and learn the sample data to accurately predict the costs of foundation pit engineering projects with complex nonlinear characteristics. (4) A case study demonstrates that the proposed calculation model has good calculation accuracy. Compared with the Chinese national code, the BP, GA-BP, SVM, RBF, and multivariate return models, the calculation accuracy is higher and the prediction results are more stable.

The organizational structure of the remainder of this paper is as follows. Section 2 introduces the research materials and methods in detail, including the analysis of the influencing factors of the construction costs of foundation pit engineering and the prediction model based on the SDAE. Section 3 presents a case analysis, and the model and case analysis are discussed in Section 4. Finally, Section 5 presents the research conclusions and further research prospects.

2. Materials and Methods

2.1. Influencing Factors of Foundation Pit Project Construction Costs

2.1.1. Analysis of the Influencing Factors of Foundation Pit Project Construction Costs. There are many factors that affect the construction costs of foundation pit projects, and these factors can generally be divided into four categories, namely, factors related to engineering characteristics, the environment, the market, and management [36, 37].

The factors related to engineering characteristics reflect the structural characteristics of foundation pit engineering itself, including the building area, the type of pile foundation, the depth of the foundation pit, and the form of the foundation structure. Environment-related factors reflect the construction site environment, including the construction site conditions, the availability of construction water and electricity, and the difficulty of earthwork excavation.

The factors related to the market are the prices of materials, such as construction machinery, construction personnel, concrete, and steel bars [38]. To reflect the market fluctuations of these prices, these factors were quantified in the present research by the project cost index. The project cost index is the ratio that reflects the degree of change of the project cost in a certain period relative to the project cost in a certain fixed period. It reflects the changing

trend of the market price in the current period relative to the market price in the base period. With reference to China's national standard ("Standard for classification and measurement of construction cost index," GB/T 51290-2018), the calculation method of the project cost index is as follows:

$$A = \frac{P_a}{P_j} \times 100, \quad (1)$$

where A is the cost index, P_a is the current cost index, and P_j is the reference period cost index. The cost index of the base period is 100.

The factors related to management mainly reflect the management level of the contractor of the foundation pit project [39]. Owners, design units, testing units, and equipment suppliers were not taken into consideration in this study, as they have little influence on the construction costs of a foundation pit project without major changes or engineering accidents. In addition, the construction period, one of the three goals of the construction project, directly reflects the management achievements [40].

2.1.2. Selection and Quantification of Influencing Factors.

The factors that affect the construction costs of foundation pit projects can be divided into two categories, namely, quantitative and qualitative factors. In the present research, the data acquisition method of quantitative factors included field investigation and the consultation of project management data, while data on the qualitative factors were obtained by questionnaires [41].

According to the analysis results presented in Section 2.1.1, the quantification of the influencing factors used in this paper is as follows.

- (1) The depth of the foundation pit: the depth of the foundation pit has the most direct influence on the design and construction of foundation pit engineering. This index is a quantitative index, and its unit is meters (m).
- (2) The form of the foundation pit support: there are 8 common forms of foundation pit support, namely, the row pile support (1), the underground diaphragm wall support (2), the cement retaining wall (3), the soil nailing wall (4), the arch wall constructed by the reverse method (5), the undisturbed soil slope (6), the reinforced concrete row pile (7), and other foundation pit support forms (8). The numbers in parentheses reflect the respective scores of these forms of support.
- (3) The form of the infrastructure: according to the stress characteristics, there are five types of commonly used foundation structures, namely, the beam foundation, strip foundation, raft foundation, box foundation, and pile foundation structures. The dimensionless index for the beam foundation structure is 1. Similarly, the indexes of the strip, raft, box, and pile foundation structures are 2, 3, 4, and 5, respectively.

- (4) The type of pile foundation: a pile foundation is a deep foundation composed of multiple piles and pile caps connecting the tops of the piles, or a single pile foundation connected by columns and piles. The selection of the pile foundation has a great influence on the construction costs of foundation pit and building engineering projects. Common pile foundations mainly include prefabricated pipe piles, rotary bored piles, manually excavated piles, punched piles, various pile types, and nonengineering piles [42], the data scores of which are 1, 2, 3, 4, 5, and 6, respectively. The reason for the introduction of nonengineering piles is that the data for the pile foundation type cannot be filled in if the beam or raft foundation is adopted.
- (5) The quantity of the pile foundation: according to the Chinese national code ("Code of bills of quantities and valuation for construction works," GB 50500-2013), the engineering quantity of a pile foundation is mainly subject to the designed engineering quantity of concrete pouring. The engineering quantity of a pile foundation is a quantitative index, and its unit is m^3 .
- It is important to note that, in engineering practice, the pile foundation engineering quantity of a beam foundation or raft foundation is 0. To avoid the presence of "0" in the unsupervised learning stage, the pile foundation engineering quantities of 0 in the actual collected data were changed to 0.01.
- (6) The engineering geological conditions: the geological factors that affect building engineering mainly include topography, stratum lithology, geological structures, earthquakes, hydrogeology, natural building materials, and unfavorable physical and geological phenomena such as karst, landslide, collapse, sand liquefaction, and foundation deformation. According to China's national code ("Code of bills of quantities and valuation for construction works," GB 50500-2013), the difficulty of Earth and rock excavation is the most obvious embodiment of engineering geological conditions and has the most direct impact on the engineering cost. Therefore, in this work, the difficulty of earthwork excavation is divided into three levels, namely, very difficult, difficult, and general, and the dimensionless qualitative indexes of which are, respectively, 1, 2, and 3.
- (7) The construction area of the foundation pit: this is a quantitative index that can be calculated according to the design drawings, and its unit is m^2 .
- (8) The on-site construction conditions: site enclosure equipment, material stacking, temporary facilities, site water supply and drainage, temporary electricity utilization, etc., have significant impacts on the smooth progress of construction. In general, when the on-site construction conditions are favorable for the normal construction process, the

construction costs will decrease. This indicator is qualitative, and there are three situations, namely, complete compliance, basic compliance, and temporary noncompliance, the dimensionless indexes of which are 1, 2, and 3, respectively.

- (9) The meteorological characteristics: meteorological characteristics mainly refer to the influence of meteorology on the construction progress during the peak construction period of foundation pit projects. These characteristics can be divided into three situations, namely, those that have large, small, and no influence on the construction. The dimensionless indexes of these situations are 1, 2, and 3, respectively.
- (10) The off-site traffic conditions: because the construction sites of foundation pit engineering are generally located in city centers, the traffic conditions around the construction site have a certain influence on the construction progress and transportation costs of the project site. This is a qualitative indicator, and experts were invited to comprehensively evaluate the traffic flow in and out of the construction site, the road conditions, and the transportation routes. There are three kinds of evaluation results, which have great, small, and no influence on the construction progression. The dimensionless indexes of these results are 1, 2, and 3, respectively.
- (11) The labor cost index: this is a quantitative index. After selecting a suitable reference period, it is calculated with reference to equation (1).
- (12) The material cost index: the commonly used and expensive materials in foundation pit engineering are steel bars and concrete. Therefore, the steel bar cost index and concrete cost index are, respectively, introduced. After selecting a suitable reference period, they are calculated with reference to equation (1).
- (13) The machinery cost index: there are many machines used in foundation pit engineering, but the engineering costs are mainly affected by large-scale mechanical equipment. Therefore, the prices of only large-scale mechanical equipment, such as rotary bored pile machines, punching pile machines, and excavators, are considered in this study. After selecting a suitable reference period, the machinery cost index is calculated with reference to equation (1).
- (14) The management level of contractors: in general, the higher the contractor's management level, the more effectively the cost will be controlled. The management level of the contractor is determined according to the qualification of the construction unit, namely, excellent, good, medium, and poor. The dimensionless indexes of these results are 1, 2, 3, and 4, respectively.

- (15) The construction period: the construction period refers to the actual number of construction days from the commencement to the completion of a foundation pit engineering project. The unit is days (d).

According to the above analysis, influencing factors of foundation pit project construction costs can be summarized, as shown in Table 1.

2.1.3. Normalization of the Influencing Factors of Foundation Pit Project Construction Costs. To prevent the output of the self-coding network from reaching saturation or even prematurely falling into the local minimum due to the large differences in the absolute values of the input data, it is necessary to normalize the input and output vectors of the training data sample set \mathbf{X} in advance [43].

In this study, the input data is transformed by linear normalization [43]:

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}, \quad (2)$$

where x_i^* represents the normalized data, x_i represents the collected data, x_{\min} is the minimum value of this type of data, and x_{\max} is the maximum value of this type of data.

2.1.4. Selection of Key Influencing Factors Based on the Rough Set. In the initial index system established in Section 2.1.2, there might be repetitiveness and redundancy among these 16 indexes. Therefore, this paper used the rough set to screen the index system.

Rough set theory was put forward by the Polish mathematician Pawlak in 1982 [17]. Attribute reduction is one of the core contents of rough set theory. To understand the importance of an attribute or attribute set, you can remove this attribute or attribute set from the decision table to observe the change of decision attributes. If the decision attribute changes greatly after removing one attribute from the conditional attribute, then the conditional index has a high degree of importance in the index system. Otherwise, the conditional index has a low degree of importance. The basic theory of rough set and the use of ROSETTA software refer to references [18–20].

According to rough set theory, this paper analysed 16 influencing factors in Section 2.1.2. See Section 3.1 in this paper, for the acquisition and processing of 60 foundation pit engineering data in Hubei, China. The data in Section 3.1 was brought into the ROSETTA, and the original data was discretized and normalized to get the decision table [19]. Without changing the relationship between decision attributes and conditional attributes in the decision table, redundant attributes were removed, and the best attribute reduction was obtained.

The results showed that the optimal attribute reduction was [X11, X12, X14, X15, X16, X17, X21, X23, X31, X32, X33, X34, X41, X42]. The redundant attribute was [X13, X15, X22]. 13 main characteristic factors affecting these project construction cost by using the rough set were X11, X12, X14,

X16, X17, X21, X23, X31, X32, X33, X34, X41, and X42, which was the influencing factor system of case analysis in this paper.

2.2. Prediction Model of Foundation Pit Project Construction Costs Based on the SDAE

2.2.1. Introduction of the Automatic Encoder. The automatic encoder (AE) deep learning neural network algorithm and unsupervised algorithm are the theoretical basis of this paper. The AE algorithm adopts unsupervised learning and supervised fine-tuning. It uses the BP algorithm and makes the output value approximate to the input value to the greatest extent via layer-by-layer training.

The main steps of the self-coding neural network are as follows [29].

- (i) *Step 1.* Find the activation value of each layer of the network.

The activation value of the neurons in each layer is calculated by forward conduction and is taken as the input value of the next layer and transmitted forward in turn. The activation function is expressed by $f(z)$, and a_i^l represents the activation value of the i th neuron in the l th layer [44]:

$$a_i^l = f(z_i^l). \quad (3)$$

Additionally, ω_{ij}^l represents the weight between the j th neuron in the $(l+1)$ th layer and the i th neuron in the l th layer, b_j^{l+1} represents the offset term of the j th neuron in the $(l+1)$ th layer, and z_i^{l+1} represents the weighted sum of all inputs of the j th neuron in the $(l+1)$ th layer [44]:

$$z_i^{l+1} = \sum_{j=1}^n \omega_{ij}^l x + b_j^{l+1}, \quad (4)$$

where n represents the number of neurons in the l th layer and x represents the input value.

- (ii) *Step 2.* Update ω and b .

The residual error between the neurons in each layer and the output layer is obtained by the BP model, and ω and b are updated continuously by the gradient descent method to make the output increasingly more similar to the input. In the proposed method, ω and b are updated by the gradient descent method [26], and the equations are as follows:

$$\begin{aligned} \omega_{ij}^{l+1} &= \omega_{ij}^l - \alpha \frac{\partial}{\partial \omega_{ij}^l} J(\omega, b), \\ b_j^l &= b_j^l - \alpha \frac{\partial}{\partial b_j^l} J(\omega, b), \end{aligned} \quad (5)$$

where $J(\omega, b)$ is the cost function [25]:

TABLE 1: The complete and universal system of influencing factors.

Primary factors	Secondary factors	Type	Acquisition and calculation of data
X1: engineering attribute	X11: the depth of the foundation pit	Quantitative	Design documents or field investigation
	X12: the form of the foundation pit support	Quantitative	Design documents or field investigation
	X13: the form of the infrastructure	Quantitative	Design documents or field investigation
	X14: the type of pile foundation	Quantitative	Design documents or field investigation
	X15: the quantity of the pile foundation	Quantitative	Project management documents and field investigation
	X16: the engineering geological conditions	Qualitative	Field investigation and questionnaire survey
	X17: the construction area of the foundation pit	Quantitative	Design documents or field investigation
X2: environment attribute	X21: the on-site construction conditions	Qualitative	Field investigation and questionnaire survey
	X22: meteorological characteristics	Qualitative	Meteorological data and questionnaire survey
	X23: the off-site traffic conditions	Qualitative	Field investigation and questionnaire survey
X3: market attribute	X31: the labor cost index	Quantitative	Market research and equation (1)
	X32: the steel bar cost index	Quantitative	Market research and equation (1)
	X33: the concrete cost index	Quantitative	Market research and equation (1)
	X34: the machinery cost index	Quantitative	Market research and equation (1)
X4: manage attributes	X41: the management level of contractors	Qualitative	Field investigation and questionnaire survey
	X42: the construction period	Quantitative	Project management documents

$$J_{AE}(\omega, b) = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} a_{\omega, b}(x)^i - y^{i2} \right), \quad (6)$$

where m is the number of samples, x is the input, and y is the output.

The AE is not very effective in dealing with some noisy data, such as text, and its accuracy can even be decreased. The DAE proposed by Vincent [45] effectively eliminates noise interference and increases the robustness of the learned features. The inspiration of the ANN originated from the biological neural network, and the DAE also draws inspiration from reality.

2.2.2. Introduction of the SDAE. Erhan et al. [46] proposed a layer-by-layer unsupervised greedy learning algorithm in 2010. A stacked self-encoder is a superposition of multiple self-encoders in which the hidden layer of the previous layer is taken as the input layer of the next layer, and the parameters of the deep network are initialized by adopting unsupervised layer-by-layer pretraining, thereby improving the convergence speed and obtaining higher-level features. In this paper, softmax regression is used to construct a classifier to classify the features learned by SAE.

Taking the construction of a self-coding N -layer stack with N automatic encoders as an example, the general steps of the SDAE are subsequently introduced.

- (i) *Step 1.* The first AE corresponds to the first hidden layer Z_1 , the input layer is the original training data X , the output layer Y_1 is the reconstruction of the input layer, and the parameters are obtained by minimizing the reconstruction error:

$$\theta = \{\omega_{11}, b_{11}, \omega_{12}, b_{12}\}. \quad (7)$$

- (ii) *Step 2.* The second AE corresponds to the second hidden layer Z_2 , in which the upper hidden layer Z_1 is taken as the input layer and the output layer Y_2 is taken as the reconstruction of the input layer Z_1 . The parameters are obtained by minimizing the reconstruction error. In the same way, the i th AE corresponds to the i th hidden layer Z_i with the upper hidden layer Z_{i-1} as the input layer and the output layer Y_i as the reconstruction of the input layer Z_{i-1} , and the parameters are obtained by minimizing the reconstruction error.

- (iii) *Step 3.* Stack self-coding refers to the training of each self-encoder layer-by-layer from left to right, and the trained optimal parameters are used as the initialization parameters of the neural network. After pretraining, the parameters of all layers can be adjusted by the BP algorithm.

In order to facilitate readers to understand the structure of SDAE, this paper uses two DAEs to construct a two-layer stack self-coding, as shown in Figure 1. First of all, the first automatic encoder corresponds to the first hidden layer Z , the input layer X is the original training data, and the output layer Y is the reconstruction of the input layer X by minimizing the reconstruction error. Then, the second automatic encoder corresponds to the second hidden layer T , which takes the hidden layer Z of the previous layer as the input layer, reconstructs the input layer Z as the output layer S , and obtains the parameters by minimizing the reconstruction error. Finally, the output layer S is discarded, and the computing tools or classification tools needed for research are connected to the hidden layer T for output O [47].

2.2.3. The Data Flow Graph and Pseudocodes of SDAE. The data flow graph based on SADE classification prediction application is shown in Figure 2.

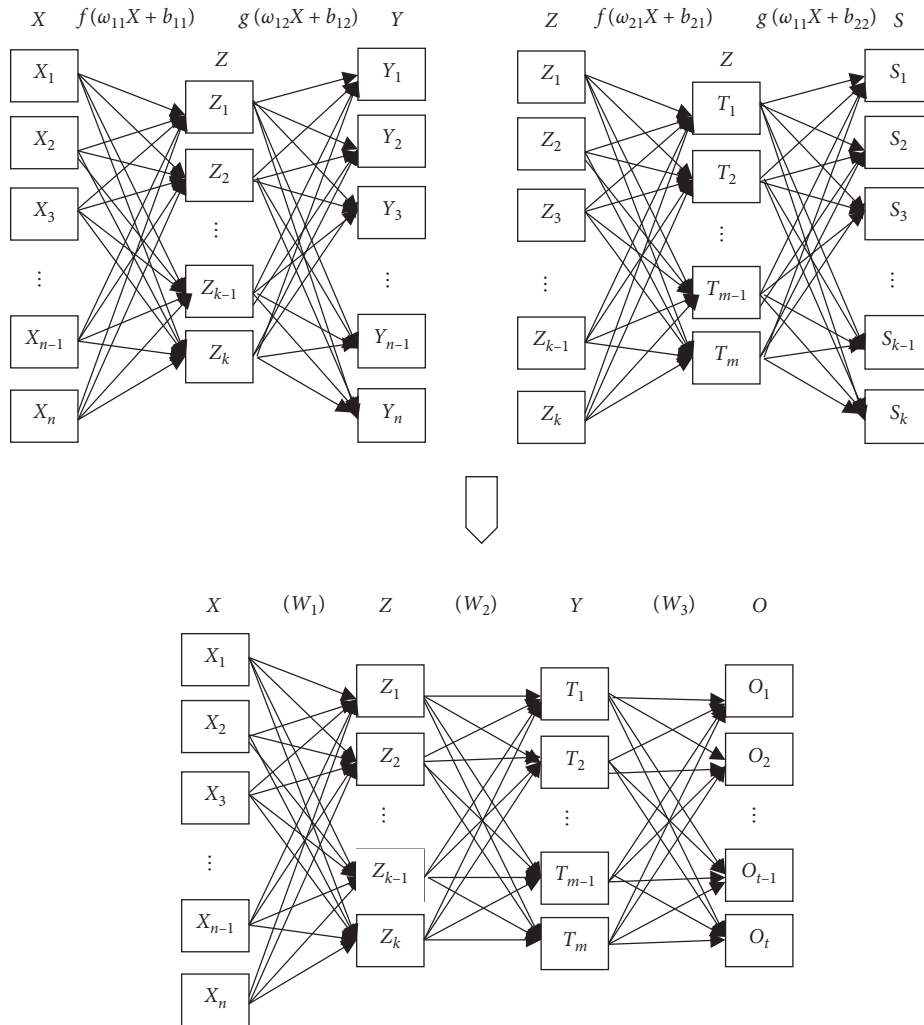


FIGURE 1: Structure diagram of SDAE composed of two DAEs.

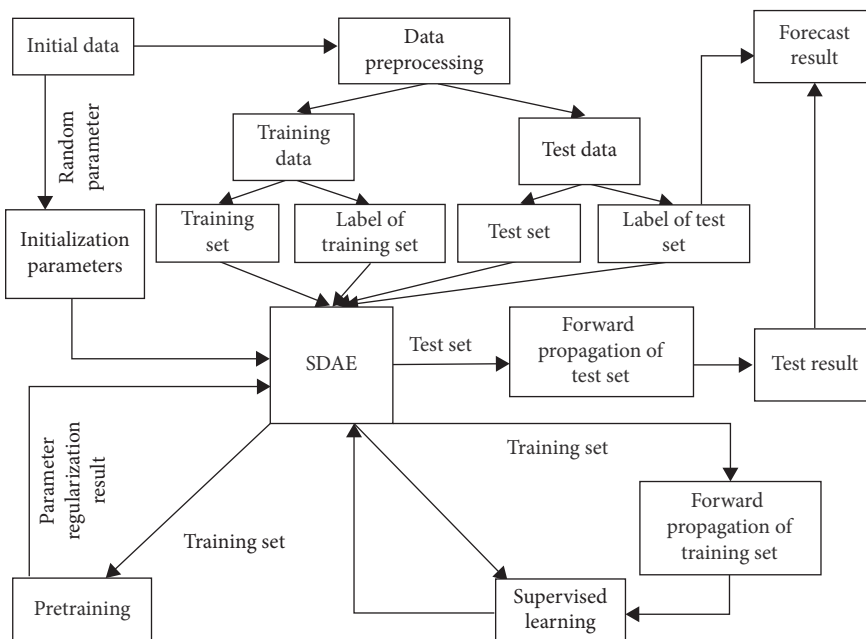


FIGURE 2: Data flow graph based on SADE classification prediction application.

- (i) Step 1 (data collection and preprocessing): collect the original data \mathbf{X} by various methods, preprocess the original data by equation (2) to get sample set \mathbf{X}^* , and divide the sample set \mathbf{X}^* to get training set \mathbf{X}_1^* and verification set \mathbf{X}_2^* . Because the deep learning algorithm belongs to the black box model, the predicted values in the training iteration process are random numbers. In this paper, the label data of the training set and the label data of the verification set are added to distinguish the prediction results.
- (ii) Step 2 (initialize parameters): set the maximum training times, learning efficiency, number of DEA networks, and initial values of weights. And bring the initialized parameters and data stream into the SDAE.
- (iii) Step 3 (unsupervised learning and forward flow of input data): its pseudocodes are in Table 2.
- (iv) Step 4 (supervised learning): its pseudocodes are in Table 3. Output the prediction result after reaching the calculation termination condition.

To facilitate the readers' understanding of the data flow graph in Figure 2, this section also detailed the pseudocodes of noise reduction encoder (Table 4), the unsupervised learning in Step 3 (Table 2), and supervised learning in Step 4 (Table 3).

(1) *Noise reduction encoder.* The algorithm 1-1 in Table 4 is pseudocode of sae_train algorithm. The first line of the code defines the input layer as 28 nodes and the three hidden layers as 100 nodes. Lines 2–6 are the first autoencoder, which is equivalent to an encoder. The second autoencoder is in the 7th–11th lines. The third autoencoder is in the 12th–16th lines, which is equivalent to a decoder.

(2) *Unsupervised learning.* Algorithm 1-2 in Table 2 is pseudocode of unsupervised pretraining process.

The first line indicates that the unsupervised learning process is pretraining layer by layer, and each layer is pretrained independently. Lines 2–5 indicate that adding pretrained weights takes the weight matrix learned in the pretraining stage as the initial value of network weights. Line 6 of the code assigns sae trained weights to nn network as initial values, which covers the previous random initialization and prepares for the next supervised learning.

(3) *Supervised learning.* Supervised learning is the core of the algorithm. Algorithm 1–3 in Table 3 is pseudocodes of supervised learning algorithms.

The first line of algorithm 1–3 code indicates that 1000 iterations should be performed in the supervised training phase. The fourth line indicates that the prediction network extracts the data of influencing factors and affected factors (construction cost) in the stage of supervised learning and training. Lines 6–8 indicate that the input is propagated

forward, and the results are obtained by the output layer after layer-by-layer weight assignment and feature extraction of all hidden layers.

3. Case Analysis

3.1. Acquisition and Normalization of Case Data. In this paper, 60 foundation pit projects in Hubei Province, China, were selected as a case study. The cost data of these projects were provided by two cooperative units (CCTEB Infrastructure Construction Investment Co., Ltd; China Construction First Group Corporation Limited). They provided the construction cost data of about 200 foundation pit projects, and the data of only 63 foundation pit projects was available. Finally, the authors randomly selected 60 foundation pit projects as case studies.

Only some engineering data of the projects are reported in Table 5 due to spatial constraints. In Table 5, y is the actual cost of each foundation pit project, and the unit is millions of RMB. The data of quantitative indicators were obtained by field research, market research, and the consultation of project management data. The reference period of the labor cost index, steel bar cost index, concrete cost index, and machinery cost index is January 1, 2018.

The data of the qualitative indicators were obtained by questionnaire surveys of 10 to 20 experts. The scoring result with the highest frequency was selected as the qualitative index score of the case project. In the questionnaire survey, experts were selected according to the criteria of being between 35 and 55 years old, holding a professional title above senior engineer or associate professor and having participated in the project construction for more than 6 months. With the assistance of SPSS 22 software, the reliability of the questionnaire survey results was analysed. The value of Cronbach's α was found to be 0.731; this exceeds the required minimum value of 0.6 [48], thereby indicating that the questionnaire survey results were reliable.

According to the content of Section 2.1.4, the 13 input vectors of the preliminary statistical training samples were normalized by equation (1) and were introduced into the SDAE via MATLAB software for calculation. The CPU of computer used in the case analysis was the Intel (R) Core (TM) i3-4170 @ 3.70 GHz, the memory was 6.00 GB, and the system was the Windows 7.

3.2. Prediction Results. In the process of modeling with the SDAE, the available data should be divided into two groups. The data of the training set is used for training, while the data of the test set is used for checking the model. Many researchers choose 90%: 10%, 80%: 20%, or 70%: 30% as the training and testing split ratio [49]. After normalization, the first 54 groups of data were taken as sample sets, and the remaining 6 groups were taken as test sets. Therefore, the ratio of training set data to test set data is 90%: 10%.

TABLE 2: The pseudocode of unsupervised pretraining process.

Algorithm 1-2: unsupervised pretraining algorithm pre_train

Entering: training data train_X
Output: The result of parameter regularization

1. sae = saetrain (sae, train_x, opts); (%) Construction of pre-training network
2. nn = nnsetup ([28 100 100 1]); (%) Set the network structure
3. nn.activation_function = "sigm"; (%) Activate function
4. nn.output = "sigm"; (%) Decoding
5. nn.learningRate = 1; (%) Learning rate in pretraining stage
6. nn.W{1} = sae.ae{1}.W{1}; (%) The weights trained by sae are assigned to nn network as initial values, covering the previous random initialization
7. nn.W{2} = sae.ae{2}.W{1};

TABLE 3: The pseudocode of supervised learning algorithms.

Algorithm 1-3: Nntrain

Input: training data and label of training data
Output: Adjusted model

1. opts.numepochs = 1000; (%) Maximum number of iterations
2. opts.batchsize = 60; (%) Parameter training is carried out with all training samples in each training
3. (%) nn.weightPenaltyL2 = 10;
4. nn = nntrain (nn, train_x, train_y, opts); (%) Training data extraction
5. nn = nnff (nn, test_x, zeros (size (test_x, 1), nn.size (end))); (%) Test data extraction
6. str = sprintf ("Predicted value 1 is (%)f", nn.a{end}(1) * (max1 - min1) + min1); (%) Denormalize the output predicted value
7. str1 = sprintf ("Predicted value 2 is (%)f", nn.a{end}(2) * (max1 - min1) + min1);
8. str2 = sprintf ("Predicted value 3 is (%)f", nn.a{end}(3) * (max1 - min1) + min1);

TABLE 4: The pseudocode of sae_train algorithm.

Algorithm 1-1: sae_train

Input: Input (training set or the result of the previous layer noise reduction decoder)
Output: The result of parameter regularization
Method: sae_train

1. sae = saesetup ([28 100 100]);
2. sae.ae{1}.activation_function = "sigm"; (%) Activate function (encoding)
3. sae.ae{1}.output = "sigm"; (%) Decoding
4. sae.ae{1}.learningRate = 1; (%) Learning rate
5. sae.ae{1}.inputZeroMaskedFraction = 0.; (%) De-noising effect of automatic coding
6. (%) sae.ae{1}.weightPenaltyL2 = 10; Regularized L2 factor
7. sae.ae{2}.activation_function = "sigm"; (%) Coding
8. sae.ae{2}.output = "sigm"; (%) Decoding
9. sae.ae{2}.learningRate = 1;
10. (%) sae.ae{2}.weightPenaltyL2 = 10;
11. sae.ae{2}.inputZeroMaskedFraction = 0.; (%) The denoise autoencoder trained in layers is equivalent to the dropout of hidden layers, which realizes the de-drying of automatic coder
12. sae.ae{3}.activation_function = "sigm"; (%) Coding
13. sae.ae{3}.output = "sigm"; (%) Decoding
14. sae.ae{3}.learningRate = 1;
15. (%) sae.ae{3}.weightPenaltyL2 = 10;
16. sae.ae{3}.inputZeroMaskedFraction = 0.;
17. opts.numepochs = 100; (%) Pre-training iterations
18. opts.batchsize = 60; (%) Parameter training is carried out with all training samples in each training

In view of the adjustability of the parameters in the SDAE model, the initial parameters were set as follows [30]: the maximum number of training iterations was 1000, the learning efficiency was 1.2, the number of DAEs was 3, and

the initial weight value was 5. The error function diagram is presented in Figure 3.

According to Figure 3, the predicted data converged between 80 and 100 iterations of the training process.

TABLE 5: Data for 60 case projects.

Factors	1	2	3	4	5	6	...	58	59	60
X11	3.4	2.7	5.5	5.0	6.3	7.5	...	10.3	12.5	9.3
X12	2	3	4	2	7	5	...	5	2	2
X14	3	4	6	6	5	3	...	6	4	6
X16	1	1	2	2	3	1	...	1	2	1
X17	736	631	1781	1361	2061	3863	...	3651	6311	4378
X21	2	2	1	1	2	3	...	2	2	1
X23	3	3	2	2	3	3	...	2	3	3
X31	1.18	1.18	1.45	1.18	1.32	1.32	...	1.32	1.18	1.32
X32	0.96	0.96	0.85	0.85	0.89	0.89	...	0.89	0.85	1.03
X33	0.93	0.93	1.12	1.12	1.37	1.37	...	1.37	1.12	0.94
X34	1.25	1.25	1.16	1.39	1.16	1.39	...	1.16	1.39	1.16
X41	2	2	3	3	1	1	...	2	2	4
X42	60	82	73	60	94	103	...	144	230	126
Y	5.17	4.83	7.61	11.36	12.01	16.37	...	18.37	23.11	24.74

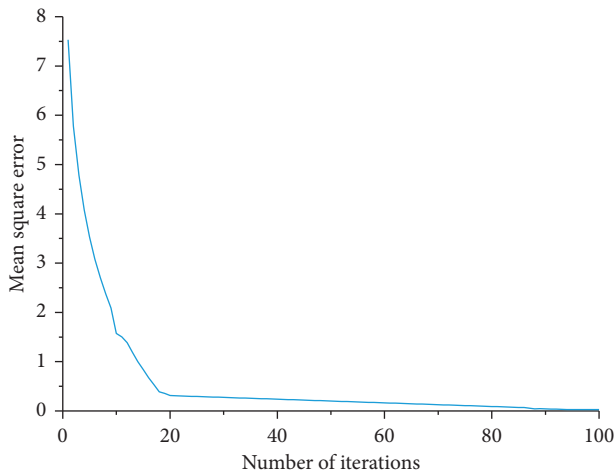


FIGURE 3: The chart of the error loss function.

MATLAB was then used to run the resulting code for a reverse data check, and the global error of pretraining (unsupervised learning) was found to reach 0.043 in the 90th training iteration and 0.025 in the 100th training iteration. In the supervised learning stage, it was found that the global error reached 0.00001 in the 653rd iteration, which met the prediction accuracy requirements.

The comparison between the predicted results and actual values of the foundation pit project construction costs is presented in Table 6. The average relative error of the six test sets was 1.54%.

In addition, 10-fold crossvalidation was conducted to test the accuracy of the algorithm [50]. The accuracy of 10 calculation results is exhibited in Figure 4 and was found to be very good. The errors of ten calculations were decreased. The average value of the maximum relative error was only 2.84%, and the average value of the minimum relative error was only 0.49%. In addition, the results of the 10 calculations were stable, which also proves the stability of the proposed algorithm.

4. Discussion

In this work, the SDAE was used to predict the construction costs of foundation pit projects. However, there were still two limitations in this study. (1) Different definitions of construction cost might have different influencing factors, which had a certain impact on the prediction results. If the cost definition was different from that, in the introduction of this paper, the influencing factors of foundation pit project construction cost would be likely to be different. (2) While the SDAE was successfully used to construct a prediction model of foundation pit project construction costs, many other deep learning methods could have been used.

4.1. Prediction Error Analysis of Different Forecasting Methods. At present, the commonly used cost forecasting methods are the calculation method based on national standard, the multivariate return analysis [21, 25], BP [25], GA-BP [51], SVM [34], and REF [52] models. In this study, the first 54 groups of data were selected as sample sets and the last 6 groups of data were selected as test sets and were also introduced into the models for calculation.

In this paper, 17 engineers were invited to calculate the construction cost of 60 foundation pit projects in the case analysis by using the Chinese national code (Code of bills of quantities and valuation for construction works, GB 50500-2013). The calculation took 24 days, and the results are shown in Figure 5. It could be seen that the calculation error of the GB 50500-2013 was very large, and the maximum error was 57.69%. The main reason might be that the calculation method based on the GB 50500-2013 roughly estimated that the construction cost was linear with the engineering quantity, while ignoring the influence of engineering changes on the construction cost. In addition, too long calculation time was another important deficiency of the calculation method based on the GB 50500-2013.

Using the return analysis function in Microsoft Excel 2016 software, the expression of multivariate return was calculated as follows:

TABLE 6: The prediction results of the proposed model.

Number of test set	55	56	57	58	59	60
Actual cost	7.38	4.19	31.48	18.37	23.11	24.74
Forecast cost	7.25	4.12	32.35	18.53	23.45	24.57
Relative error (%)	1.76	1.67	2.76	0.87	1.47	0.69

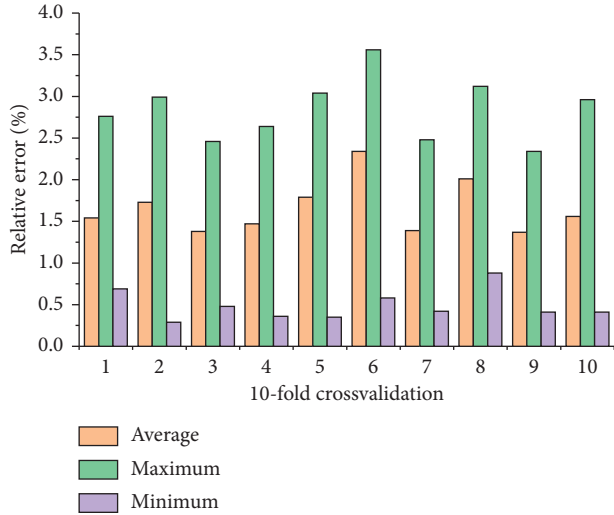


FIGURE 4: Error results of 10-fold crossvalidation.

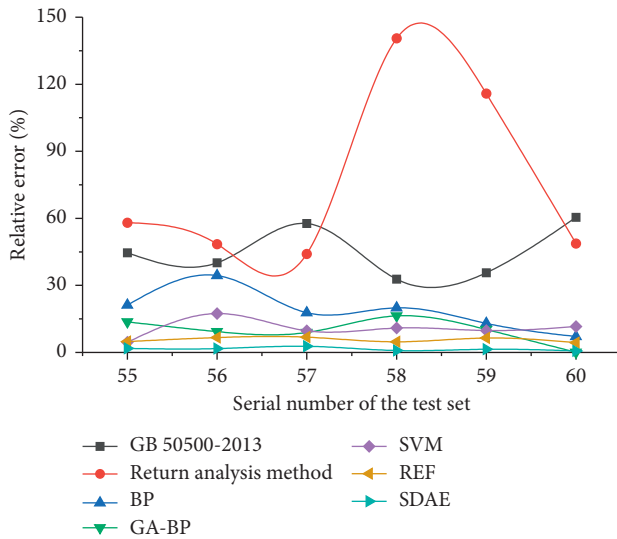


FIGURE 5: Relative errors of different computational models.

$$y = 4.278 + 1.375X_{12} - 0.000034X_{15} + 0.00493X_{17} + 3.04X_{41} - 0.1915X_{42}. \quad (8)$$

The data of the test set were introduced into equation (8), and the prediction results are presented in Figure 5. According to equation (8), it can be known that only five factors were related to the prediction results when using multivariate return analysis. The failure to make full use of

all index data is one of the important reasons for the low accuracy of the calculation results of this method [21].

In the BP algorithm, the selected training function was “traingda,” the activation function of the hidden layers was “logsig,” and the activation function of the output layers was “purelin.” The target error of training was set as 1×10^{-6} , and the maximum number of iterations was set as 1000. The learning rule of the network was the error gradient descent method. In the GA-BP algorithm, the number of individuals in the population was 50, the maximum genetic algebra was 1000, the number of binary digits was 20, and the generation gap was 0.9. In the cost prediction based on the SVM, the number of iterations was 100, and the population size and k value were 20 and 0.6, respectively. The calculation results are reported in Figure 5.

The maximum relative error of the SDAE model was only 0.0283, which is considerably less than the maximum relative errors of the other algorithms. The relative error is an important index in error analysis, and Table 7 presents the relative errors of several different calculation models. The relative error calculated by the proposed method was less than 3%, and the average error was only 1.54%. Among all the methods, the calculation error of the multivariate return analysis method was the largest, and the relative error was as high as 140.52%. The calculation errors of the BP, GA-BP, SVM, and RBF models were large, and the maximum relative errors were 34.43%, 16.39%, 13.60%, and 6.83%, respectively. These results also prove that the proposed method is effective and advanced in predicting the construction costs of foundation pit projects.

In addition, combined with the calculation results of other error analysis tools, it could be qualitatively considered that SDAE had the highest calculation accuracy in case analysis, and the calculation accuracy order of other methods was as follows: REF > SVM > GA-BP > BP > return analysis method. This sort of calculation accuracy was consistent with the previous research results [21, 25, 51], which also proved that the case analysis in this paper was scientific and correct.

In order to further compare and analyze the calculation errors of various calculation methods, the coefficient of determination (R^2), the root mean square error (RMSE), and the mean absolute error (MAE) were used to analyze the prediction error in the case study.

R^2 indicates the degree of correlation between the actual and predicted values. The closer R^2 is to 1, the higher the correlation; conversely, the closer R^2 is to 0, the lower the correlation [53]:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i^{\text{exp}} - y_i^{\text{pre}})^2}{\sum_{i=1}^N (y_i^{\text{exp}} - \bar{y}_i^{\text{exp}})^2}, \quad (9)$$

where y_i^{exp} is the actual result, y_i^{pre} is the predicted result, and \bar{y}_i^{exp} is the average value of the actual results.

The RMSE is an important standard used to measure the prediction results of machine learning models [54]. Its calculation method is as follows:

TABLE 7: Comparison of the three error representations of different models.

Error representations	R^2	RMSE	MAE
GB 50500-2013	0.3798	2.1701	6.8911
Return analysis method	0.2311	3.9010	9.1758
BP	0.6410	1.3725	3.1433
GA-BP	0.8276	0.8130	1.6050
SVM	0.8531	0.6203	1.2604
REF	0.8970	0.6977	1.0432
SDAE	0.9743	0.4110	0.3050

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i^{\text{exp}} - y_i^{\text{pre}})^2}. \quad (10)$$

The MAE is the average of absolute errors, which can better reflect the actual situation of errors in predicted values [53]:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i^{\text{exp}} - y_i^{\text{pre}}|. \quad (11)$$

The error results of different methods are shown in Table 7.

According to the calculation results presented in Table 3, the R^2 value of the SDAE model was the highest, namely, 0.9743, which is very close to 1. In other words, the predicted values calculated by the SDAE model were very close to the actual values. Compared with the GB 50500-2013, the multiple return analysis method, BP, GA-BP, SVM, and REF models, the proposed model had better prediction results. The RMSE of the SDAE model was 0.1689, which is notably less than the RMSEs of the other algorithms. The MAE of the SDAE model was 0.305, which is notably less than the MAEs of the other algorithms. Compared with the other common methods, the SDAE model exhibited a superior calculation accuracy.

4.2. Stability Analysis of Different Computational Models. Stability determines the reliability and generalization of the model in engineering application. In this paper, the standard deviation was used as a measure of the stability of SDAE model. Among the 54 training samples in Section 3.1, 20, 30, 40, and 50 samples were randomly selected as training sets, and the last 6 samples were also used as test sets. The standard deviations of different models are shown in Table 8.

It can be seen from Table 8 that the SDAE model showed a low standard deviation of prediction, regardless of the size of the training sample. The SDAE had the strongest stability with the increase of the training sample size. The larger the training sample size, the stronger the stability of SDAE model. Compared with the Chinese national code, the diversified return method, the BP, the GA-BP, the SVM, and the RBF, the standard deviation of SDAE was lowest, which showed that this model had stronger stability than other models.

4.3. The Influence of the Number of Input Variables on the Prediction Results. According to previous research results [13], when an artificial intelligence method is applied to the prediction of construction costs, the number of input variables has a notable influence on the accuracy of the prediction results. Therefore, the influence of the number of input variables on the prediction results was analysed. Considering that many factors affect the construction costs of foundation pit projects, only the following situations were analysed. Plan A was the use of the 16 influencing factors identified in Section 2.2.1. In Plan B, the influencing factor X11 (*the depth of the foundation pit*) was deleted. In Plan C, the influencing factor X12 (*the form of the foundation pit support*) was deleted. In Plan D, the influencing factors X11 and X12 were deleted. In Plan E, the influencing factor X13 (*the form of the infrastructure*) was deleted. In Plan F, the influencing factors X11, X12, and X13 were deleted. In Plan G, the X13, X15, and X22 were deleted. The index system of the Plan G was the same as that in case analysis. In Plan H, the influencing factors X11, X12, X13, and X21 (*the on-site construction conditions*) were deleted. In Plan I, the influencing factors X11, X12, X13, X21, and X22 (*meteorological characteristics*) were deleted. Finally, in Plan J, the influencing factors X11, X12, X13, X21, X22, X31 (*the labor cost index*), and X32 (*the steel bar cost index*) were deleted. The calculation results of these plans are shown in Table 9.

When impact factor X11 (Plan B) or X12 (Plan C) was deleted, the error of the calculation results increased obviously, whereas this did not occur when other single factors (such as Plan E) were deleted. In the example of reducing two influence factors at the same time (Plan D), the error of the calculation results increased obviously when X11 and X12 were deleted. However, when other influencing factors in addition to X11 and X12 were deleted, the calculation error did not increase obviously. For example, the maximum relative error of Plan I, in which influencing factors X11, X12, X13, X21, X22, X31, and X32 were deleted, was 6.0%, which is only slightly larger when only influencing factors X11 and X12 were deleted based on this analysis, it can be preliminarily considered that the influencing factors X11 and X12 have a substantial influence on the calculation accuracy. Comparing Plan A and Plan G, the calculation errors of the two index systems were very close. This could explain the rationality and efficiency of the index screening results in Section 2.1.4 of this paper. It should be emphasized that the analysis and discussion on

TABLE 8: Stability of calculation results under different sample numbers.

Model	20	30	40	50
GB 50500-2013	4.890	3.283	2.840	2.594
Return analysis method	13.370	7.292	6.382	4.900
BP	10.063	7.063	5.985	3.801
GA-BP	3.266	2.006	0.922	1.312
SVM	2.650	1.985	1.298	0.578
REF	1.783	1.357	1.034	1.231
SDAE	0.931	1.047	0.461	0.058

the number of input variables in this section was preliminary, not complete. The main reason was that there were too many input variables.

5. Conclusion

Foundation pit project construction costs are an important component of building project construction costs. The prediction of foundation pit project construction costs is the basis of not only cost planning but also of the cost decisions and planning of construction projects. In this paper, beginning from the four attributes of construction cost management (engineering, the environment, the market, and management), the influencing factors of foundation pit project construction costs were identified. Combined with China's national standards and the practice of foundation pit project management, a method of the quantization of the influencing factors was provided. Then, the SDAE was utilized to construct a prediction model of foundation pit project construction costs. Finally, 60 foundation pit projects in Hubei Province, China, were selected for a case analysis. The case study results demonstrated that, compared with the actual construction costs, the calculation error of the proposed method was less than 3%, and the average error was only 1.54%. In addition, three error analysis tools commonly used in machine learning (the determination coefficient, root mean square error, and mean absolute error) emphasized that the calculation accuracy of the proposed method was superior to those of the Chinese national code, the multivariate return method, the BP model, the BP model optimized by the genetic algorithm, the SVM model, and the RBF model. For 60 foundation pit projects in case analysis, deleting X13, X15, and X22 did not affect the prediction results. The result also proved the rationality and efficiency of the key impact indicators obtained by the rough set. On the basis of the research results in this paper, relevant researchers are encouraged to further find a complete and universal system of influencing factors affecting the project construction cost of deep foundation pit.

Data Availability

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

TABLE 9: Error analysis of the calculation results with different numbers of input variables.

Error representations	R^2	RMSE	MAE
Plan A	0.979	0.402	0.295
Plan B	0.943	0.473	0.323
Plan C	0.952	0.486	0.347
Plan D	0.901	0.520	0.483
Plan E	0.969	0.415	0.309
Plan F	0.894	0.423	0.334
Plan G	0.974	0.411	0.305
Plan H	0.863	0.569	0.518
Plan I	0.871	0.597	0.510
Plan J	0.858	0.566	0.528

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This study was supported by the Science and Technology Project of Wuhan Urban and Rural Construction Bureau, China (201943).

References

- [1] S. Durdyev and S. Ismail, "Role of the construction industry in economic development of Turkmenistan," *Energy Education Science and Technology Part a: Energy Science and Research*, vol. 29, no. 2, pp. 883–890, 2012.
- [2] M. Mukuka, C. Aigbavboa, and W. Thwala, "Effects of construction projects schedule overruns: a case of the Gauteng Province, South Africa," *Procedia Manufacturing*, vol. 3, pp. 1690–1695, 2015.
- [3] I. Zafar, T. Yousaf, and D. S. Ahmed, "Evaluation of risk factors causing cost overrun in road projects in Terrorism Affected Areas Pakistan - a case study," *KSCE Journal of Civil Engineering*, vol. 20, no. 5, pp. 1613–1620, 2016.
- [4] P. E. D. Love and D. J. Edwards, "Curbing rework in offshore projects: systemic classification of risks with dialogue and narratives," *Structure and Infrastructure Engineering*, vol. 9, no. 11, pp. 1118–1135, 2013.
- [5] J. P. Xu, Y. Tu, and Z. Q. Zeng, "A nonlinear multiobjective bilevel model for minimum cost network flow problem in a large-scale construction project," *Mathematical Problems in Engineering*, vol. 2012, Article ID 463976, 40 pages, 2012.
- [6] J. D. Wu, "Selection of the scheme of foundation pit support based on value engineering," *Construction Economy*, vol. 353, no. 3, pp. 88–90, 2012.
- [7] A. Gilchrist and E. N. Allouche, "Quantification of social costs associated with construction projects: state-of-the-art review," *Tunnelling and Underground Space Technology*, vol. 20, no. 1, pp. 89–104, 2005.
- [8] L. F. Cabeza, L. Rincon, V. Vilarino et al., "Life cycle assessment (LCA) and life cycle energy analysis (LCEA) of buildings and the building sector: a review," *Renewable & Sustainable Energy Reviews*, vol. 29, pp. 394–416, 2013.
- [9] N. Forcada, M. Gangolells, M. Casals et al., "Factors affecting rework costs in construction," *Journal of Construction Engineering and Management*, vol. 134, no. 8, Article ID 04017032, 2017.

- [10] B. Wang and J. Dai, "Discussion on the prediction of engineering cost based on improved BP neural network algorithm," *Journal of Intelligent & Fuzzy Systems*, vol. 37, no. 5, pp. 6091–6098, 2019.
- [11] T. P. Williams and J. Gong, "Predicting construction cost overruns using text mining, numerical data and ensemble classifiers," *Automation in Construction*, vol. 43, pp. 23–29, 2014.
- [12] Y. M. Wei, "Earned value method in the total cost variance analysis of land development and consolidation project," *China Land Science*, vol. 25, no. 5, pp. 73–78, 2011.
- [13] M. Attalla and T. Hegazy, "Predicting cost deviation in reconstruction projects: artificial neural networks versus regression," *Journal of Construction Engineering and Management*, vol. 129, no. 4, pp. 405–411, 2003.
- [14] A. Lesniak and M. Juszczak, "Prediction of site overhead costs with the use of artificial neural network based model," *Archives of Civil and Mechanical Engineering*, vol. 18, no. 3, pp. 973–982, 2018.
- [15] K. J. Kim and K. Kim, "Preliminary cost estimation model using case-based reasoning and genetic algorithms," *Journal of Computing in Civil Engineering*, vol. 24, no. 6, pp. 499–505, 2010.
- [16] J. C. Dong, Y. Chen, and G. Guan, "Cost index predictions for construction engineering based on LSTM neural networks," *Advances in Civil Engineering*, vol. 2020, 14 pages, Article ID 6518147.
- [17] M. Su, P. Wang, Y. Xue et al., "Prediction of risk in submarine tunnel construction by multi-factor analysis: a collapse prediction model," *Marine Georesources & Geotechnology*, vol. 37, no. 9, pp. 1119–1129, 2019.
- [18] G.-H. Zhang, Y.-Y. Jiao, L.-B. Chen, H. Wang, and S.-C. Li, "Analytical model for assessing collapse risk during mountain tunnel construction," *Canadian Geotechnical Journal*, vol. 53, no. 2, pp. 326–342, 2016.
- [19] X. M. Xu, Q. Wang, D. X. Niu et al., "Synergistic effect evaluation of main and auxiliary industry of power grid based on the information fusion technology from the perspective of sustainable development of enterprises," *Sustainability*, vol. 10, no. 2, Article ID 457, 2018.
- [20] S. Barbagallo, S. Consoli, N. Pappalardo, S. Greco, and S. M. Zimbone, "Discovering reservoir operating rules by a Rough Set approach," *Water Resources Management*, vol. 20, no. 1, pp. 19–36, 2006.
- [21] S. M. Trost and G. D. Oberlender, "Predicting accuracy of early cost estimates using factor analysis and multivariate regression," *Journal of Construction Engineering and Management*, vol. 129, no. 2, pp. 198–204, 2003.
- [22] S. H. Ji and J. Ahn, "Scenario-planning method for cost estimation using morphological analysis," *Advances in Civil Engineering*, vol. 2019, 10 pages, Article ID 4962653, 2019.
- [23] Y.-M. Cheng, C.-H. Yu, and H.-T. Wang, "Short-interval dynamic forecasting for actual S-curve in the construction phase," *Journal of Construction Engineering and Management*, vol. 137, no. 11, pp. 933–941, 2011.
- [24] X. J. Wang, "Forecasting construction project cost based on BP neural network," in *Proceeding of the 10th International Conference on Measuring Technology And Mechatronics Automation*, pp. 420–423, Changsha, China, February 2018.
- [25] M. Gunduz, L. O. Ugur, and E. Ozturk, "Parametric cost estimation system for light rail transit and metro trackworks," *Expert Systems with Applications*, vol. 38, no. 3, pp. 2873–2877, 2011.
- [26] K. M. Kiani and T. L. Kastens, "Testing forecast accuracy of foreign exchange rates: predictions from feed forward and various recurrent neural network architectures," *Computational Economics*, vol. 32, no. 4, pp. 383–406, 2008.
- [27] R. H. Mu and X. Q. Zeng, "A review of deep learning research," *KSII Transactions on Internet and Information Systems*, vol. 13, no. 4, pp. 1738–1764, 2019.
- [28] H. A. Owolabi, M. Bilal, L. O. Oyedele, H. A. Alaka, S. O. Ajayi, and O. O. Akinade, "Predicting completion risk in PPP projects using big data analytics," *IEEE Transactions on Engineering Management*, vol. 67, no. 2, pp. 430–453, 2020.
- [29] G. Zhang, Y. Liu, and X. Jin, "A survey of autoencoder-based recommender systems," *Frontiers of Computer Science*, vol. 14, no. 2, pp. 430–450, 2020.
- [30] E. K. Wang, X. Zhang, and L. Pan, "Automatic classification of CAD ECG signals with SDAE and bidirectional long short-term network," *IEEE Access*, vol. 7, pp. 182873–182880, 2019.
- [31] P. Liu, P. J. Zheng, and Z. Y. Chen, "Deep learning with stacked denoising auto-encoder for short-term electric load forecasting," *Energies*, vol. 12, no. 12, Article ID 2445, 2019.
- [32] J. Dai, H. Song, G. Sheng, and X. Jiang, "Cleaning method for status monitoring data of power equipment based on stacked denoising autoencoders," *IEEE Access*, vol. 5, pp. 22863–22870, 2017.
- [33] W. Dong, H. Sun, Z. Li, J. Zhang, and H. Yang, "Short-term wind-speed forecasting based on multiscale mathematical morphological decomposition, K-means clustering, and stacked denoising autoencoders," *IEEE Access*, vol. 8, pp. 146901–146914, 2020.
- [34] X. Chen, M. Li, and Y. Q. Xiao, "Stacked denoise autoencoder based feature extraction and classification for hyperspectral images," *Journal of Sensors*, vol. 2016, Article ID 3632943, 10 pages, 2016.
- [35] J. Yan, H. Zhang, Y. Liu, S. Han, L. Li, and Z. Lu, "Forecasting the high penetration of wind power on multiple scales using multi-to-multi mapping," *IEEE Transactions on Power Systems*, vol. 33, no. 3, pp. 3276–3284, 2018.
- [36] Y.-M. Cheng, "An exploration into cost-influencing factors on construction projects," *International Journal of Project Management*, vol. 32, no. 5, pp. 850–860, 2014.
- [37] V. Sharma, C. H. Caldas, and S. P. Mulva, "Identification and prioritization of factors affecting the overall project cost of healthcare facilities," *Journal of Construction Engineering and Management*, vol. 146, no. 2, Article ID 04019106, 2020.
- [38] J. Ngo, B. G. Hwang, and C. Y. Zhang, "Factor-based big data and predictive analytics capability assessment tool for the construction industry," *Automation in Construction*, vol. 110, p. 2020, Article ID 103042.
- [39] M. T. Hatamleh, M. Hiyassat, G. J. Sweis, and R. J. Sweis, "Factors affecting the accuracy of cost estimate: case of Jordan," *Engineering, Construction and Architectural Management*, vol. 25, no. 1, pp. 113–131, 2018.
- [40] Y. Wang and Q. J. Li, "Study on influencing factors of cost control of construction project based on structural equation," *Construction Economy*, vol. 41, no. 02, pp. 63–68, 2020.
- [41] X. Gao and P. Pishdad-Bozorgi, "A framework of developing machine learning models for facility life-cycle cost analysis," *Building Research & Information*, vol. 48, no. 5, pp. 501–525, 2020.
- [42] M. Yang and J. W. Jin, "Research progress on interaction of pile foundation with nearby existing subway tunnel," *Journal of Building Structures*, vol. 37, no. 8, pp. 90–100, 2016.
- [43] H. Wu and J. W. Wang, "Assessment of waterlogging risk in the deep foundation pit projects based on projection pursuit

- model,” *Advances in Civil Engineering*, vol. 2020, Article ID 2569531, 11 pages, 2020.
- [44] W. Liu, T. Ma, Q. Xie, D. Tao, and J. Cheng, “LMAE: a large margin Auto-Encoders for classification,” *Signal Processing*, vol. 141, pp. 137–143, 2017.
- [45] J. Zhang, Y. Zhang, L. Bai, and J. Han, “Lossless-constraint denoising based auto-encoders,” *Signal Processing: Image Communication*, vol. 63, pp. 92–99, 2018.
- [46] D. Erhan, Y. Bengio, A. Courville et al., “Why does unsupervised pre-training help deep learning,” *Journal of Machine Learning Research*, vol. 11, no. 3, pp. 625–660, 2010.
- [47] F. Xu and P. W. Tse, “Automatic roller bearings fault diagnosis using DSAE in deep learning and CFS algorithm,” *Soft Computing*, vol. 23, no. 13, pp. 5117–5128, 2019.
- [48] M. H. P. Passos, H. A. Silva, A. C. R. Pitangui, V. M. A. Oliveira, A. S. Lima, and R. C. Araújo, “Reliability and validity of the Brazilian version of the pittsburgh sleep quality index in adolescents,” *Jornal de Pediatria*, vol. 93, no. 2, pp. 200–206, 2017.
- [49] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks,” *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998.
- [50] J.-H. Kim, “Estimating classification error rate: repeated cross-validation, repeated hold-out and bootstrap,” *Computational Statistics & Data Analysis*, vol. 53, no. 11, pp. 3735–3745, 2009.
- [51] O. Kisi, M. Alizamir, and M. Zounemat-Kermani, “Modeling groundwater fluctuations by three different evolutionary neural network techniques using hydroclimatic data,” *Natural Hazards*, vol. 87, no. 1, pp. 367–381, 2017.
- [52] R. Sonmez, “Range estimation of construction costs using neural networks with bootstrap prediction intervals,” *Expert Systems with Applications*, vol. 38, no. 8, pp. 9913–9917, 2011.
- [53] M. A. Ahmadi, R. Haghbakhsh, R. Soleimani, and M. B. Bajestani, “Estimation of H₂S solubility in ionic liquids using a rigorous method,” *The Journal of Supercritical Fluids*, vol. 92, pp. 60–69, 2014.
- [54] H. H. Li, Y. D. Lu, C. Zheng et al., “Groundwater level prediction for the arid oasis of Northwest China based on the artificial bee colony algorithm and a back-propagation neural network with double hidden layers,” *Water*, vol. 11, no. 4, Article ID 860, 2019.