

Research Article

Cost Index Predictions for Construction Engineering Based on LSTM Neural Networks

Jiacheng Dong, Yuan Chen , and Gang Guan

School of Civil Engineering, Zhengzhou University, Zhengzhou 450001, China

Correspondence should be addressed to Yuan Chen; chen_yuan@zzu.edu.cn

Received 15 February 2020; Revised 26 August 2020; Accepted 4 September 2020; Published 24 September 2020

Academic Editor: Reza Akhavian

Copyright © 2020 Jiacheng Dong 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.

In recent years, the cost index predictions of construction engineering projects are becoming important research topics in the field of construction management. Previous methods have limitations in reasonably reflecting the timeliness of engineering cost indexes. The recurrent neural network (RNN) belongs to a time series network, and the purpose of timeliness transfer calculation is achieved through the weight sharing of time steps. The long-term and short-term memory neural network (LSTM NN) solves the RNN limitations of the gradient vanishing and the inability to address long-term dependence under the premise of having the above advantages. The present study proposed a new framework based on LSTM, so as to explore the applicability and optimization mechanism of the algorithm in the field of cost indexes prediction. A survey was conducted in Shenzhen, China, where a total of 143 data samples were collected based on the index set for the corresponding time interval from May 2007 to March 2019. A prediction framework based on the LSTM model, which was trained by using these collected data, was established for the purpose of cost index predictions and test. The testing results showed that the proposed LSTM framework had obvious advantages in prediction because of the ability of processing high-dimensional feature vectors and the capability of selectively recording historical information. Compared with other advanced cost prediction methods, such as Support Vector Machine (SVM), this framework has advantages such as being able to capture long-distance dependent information and can provide short-term predictions of engineering cost indexes both effectively and accurately. This research extended current algorithm tools that can be used to forecast cost indexes and evaluated the optimization mechanism of the algorithm in order to improve the efficiency and accuracy of prediction, which have not been explored in current research knowledge.

1. Introduction

Due to unique industry characteristics, construction projects require a large amount of capital investment [1]. Accurate cost predictions are essential for the effective implementation of construction projects [2]. The cost indexes used in construction engineering projects are technical and economic indicators which reflect the impacts of market price fluctuations on the construction costs of engineering factors during certain periods of time. Construction engineering cost indexes have been widely implemented in cost predictions, tender compilations, and investment planning [3]. In addition, engineering cost indexes can be used to measure the fluctuations in the construction costs and also provide foundations for the bidding quotations, project predictions,

and settlements in the cost management processes [4]. Therefore, the accurate cost predictions have major benefits for investment planning, bidding, and feasibility analyses during the early stages of construction projects.

The research of engineering cost prediction has a long history. The initial prediction form must be based on a complete design drawing, which is unable to meet the requirements of the actual scenarios in terms of time and efficiency. With the development of information technology, cost prediction began to break away from the limitation of drawings, and gradually a solution emerged to establish an information model based on various algorithms. At present, the cost prediction methods available based on information technology can be divided into two branches: construction cost prediction and cost indexes prediction. The types of

construction cost prediction mainly include the prediction of bidding price [5] and market information price [6]. The prediction types of cost indexes mainly include the prediction of tender price index (TPI) [7] and Engineering News-Record (ENR) construction cost index (CCI) [8]. Among them, the construction cost prediction data mainly comes from expert judgment or historical data of similar projects [9], which lacks authority and persuasion. However, the data source of cost indexes prediction is generally a dynamic release from authoritative institutions. The calculation mode of construction cost is based on the engineering cost indexes. With reference to the concept of Computer Science, the cost indexes can be called the data provenance (lineage or pedigree) of construction cost. By directly processing the data provenance, the error and distortion of the data during the processing can be avoided to a certain extent [10]. Most of the current advanced research methods achieve prediction based on a certain regression analysis or assuming some combination patterns to appear repeatedly [11]. LSTM NN replaces the hidden layer neurons in the RNN with four logic units and establishes a long-time lag between input, feedback, and prevention of gradient explosion. This structure type realizes the continuity of the internal state error flow in the special memory unit, enables LSTM to capture the large dependence of the time step distance in the time series, and has a strong approximation ability for nonlinear and nonstationary time series, which will improve the prediction efficiency and accuracy of sequence data with temporal or spatial attributes.

2. Literature Review

2.1. Cost Prediction Methods. The methods of cost predictions can be divided into two categories: causal analysis and time series analysis [12]. With the development and advantages of artificial intelligence, machine learning algorithms have been applied in this field. As determined by the available research reports, the causal method, also known as causal analysis, needs to specify the relationships between the predictive variables and the dependent variables or between dependent variables and interpretative variables [12]. Therefore, causal analysis predictions are based on the interpretations of the relationships between the engineering cost indicators and other variables, which can then be used to predict project cost indexes [13]. In the present study, previous related reports were reviewed, in which causal methods had been successfully used to predict tender price indices (TPI), building prices, and construction costs [8]. Akintoye and Skitmore used OLS multiple regression analysis to construct a contract price model and provided a structural explanation for the trend changes in the TPI through the structural equation model [14]. Trost and Oberlender obtained the ranking of the factors influencing the accuracy of early cost estimation through factor analysis and multiple regression analysis [15]. Chen proposed a combination method based on transformed time series data, multiple regression analysis, Yule-Walker estimates, and incomplete principal component regression analysis to predict the company-level cost flow within a certain range successfully [16].

Statistical methods are also known as black box methods or time series methods which can be divided into the two categories of univariate and multivariate time series analyses [17]. Wong et al. utilized an autoregressive integrated moving average model (ARIMA) to predict the five main indicators of the Hong Kong labour market [18]. Hwang proposed two dynamic univariate time series models to successfully predict ENR CCI [19]. In contrast, multivariate time series analysis is based on multiple variables. The advantage of this type of analysis is that only quantitative data can be used, and objective predictions can be made without additional subjective judgment processes. Therefore, multivariate time series analysis methods theoretically have higher predictive abilities [20]. Hwang compared and analysed the prediction effects of autoregressive moving average models (ARMA) and vector autoregression (VAR) time series models on structural cost indicators. The comparison results indicated that univariate and multivariate time series analysis methods each had their own advantages, resulting in the ARMA (5, 5) having slightly higher precision [20]. Xu and Moon adopted a cointegration equation to establish a cointegrated VAR model, in which the deviations between the cointegrating relationships and the long-run stable relationships between the variables were considered. As a result, the ENR CCI was successfully predicted [12].

The branch of machine learning which has been most widely used in the field of engineering cost prediction is mainly composed of neural networks, support vector machines (SVM), and k-nearest neighbour (KNN) algorithms. Juszczak and Leśniak proposed a model based on the artificial neural networks (ANN) which are involved in radial basis functions (RBF) for the purpose of forecasting the indexes of site overhead costs. It was found that the prediction models had achieved satisfactory results [21]. Nam et al. proposed a hybrid model which combined artificial neural networks and wavelet transformation in its predictions of engineering cost index trends [22]. In the studies conducted by Cheng et al., a hybrid method was proposed which was based on Least Squares support vector machines (LS-SVM) and differential evolution (DE), referred to as ELSVM. The results of the aforementioned method indicated that it had the ability to successfully predict the fluctuations of ENR CCI [23]. Wang and Ashuri adopted a modified KNN algorithm to establish prediction models. It was found that although the models were better than the time series models, they still could not capture the feature of jump in CCI. Therefore, this study believed that further exploration of more advanced nonlinear machine learning algorithms was necessary [24].

2.2. Limitations of Traditional Prediction Methods and Advantages of LSTM. As discussed in previous literature review, traditional prediction methods have their own limitations in predicting the construction cost index. For example, the causal methods require many explanatory variables to be predicted and cannot reflect the uncertain price fluctuations [12]. The univariate time series methods

are only suitable for short-term predictions [11], and the multivariate time series methods are costly in terms of their analysis and prediction process [20]. The major drawback of the SVM and KNN algorithms is their high computational burden [25, 26]. LSTM was originally proposed by Hochreiter and Schmidhuber [27], which is an effective nonlinear recurrent network. LSTM has proven to be superior to most nonparametric prediction methods [28]. The advantages of LSTM can be specifically analysed by comparing with the limitations of other methods. It has been found that RNNs have the problem of gradient vanishing and lacking of long-term memory ability [29]. In the process of LSTM application in the field of cost prediction, LSTM replaces RNN neurons with memory cell states and controls the flow of information by adding an input gate, forget gate, and output gate. These nonlinear summation units of LSTM use the sigmoid function to calculate the memory state (previous network state) of the network as the input. If the output result reaches the threshold, then the output of the gate and the calculation result of the current layer are input to the next layer by means of matrix multiplication. If the threshold is not reached, then the output result is forgotten [30]. The weight of each layer of network and gate nodes will be updated during each backpropagation training process. This structural form grants LSTM more sophisticated transition abilities for addressing gradients [31], thereby compensating for the limitations of RNNs. Correspondingly, although LSTM has no advantage in dealing with highly nonlinear and long interval time series datasets [32], the training cost and duration of the LSTM model are generally low and easy to control when the number of hidden layer neurons is set reasonably.

3. Research Methodology

3.1. Research Questions and Methodology Consideration. This study poses two research questions: how LSTM NN can be applied to predict engineering cost indexes and how various factors can affect model performance including input features, time series length, and model structures.

Applying LSTM NN in predictions of construction engineering cost indexes and exploring the optimization mechanism are mainly based on the following considerations:

- (i) According to literature review and theoretical research analysis, the structure type, training cost, and calculation efficiency of LSTM NN are suitable for the processing of cost indexes data. However, the performance of LSTM NN in this field has not been explored in the previous research.
- (ii) The feature selection of an LSTM neural network model has a major influence on the prediction accuracy of the model. However, there is currently no standard selection criterion for the selection of the parameters of such a model.

3.2. Research Objectives and Research Methods. The aims of this research were to explore the various theories and

methods available for LSTM neural networks in the accurate predictions of construction engineering cost indexes and to evaluate the proposed model's prediction performances. The first objective of this research is to investigate the research gaps in the field of cost prediction and the limitations of the current forecasting methods. The corresponding research methods are literature review and theoretical analysis methods. The second research objective is to determine a set of indicators suitable for the prediction of China's cost indexes. Through literature review and expert argumentation, the content of the indicator set can reach a certain level of comprehensiveness. The third research goal is to verify the applicability of the LSTM NN, for which the case analysis method can be used to objectively judge the predictive performance of the model. The final research goal is to explore the optimization mechanism of the LSTM model. Different input features, time series lengths, and model structures are set by comparative analysis to improve prediction accuracy.

4. Selection of the Prediction Indicators and the Original Data Collection

4.1. Selection and Adjustment of the Forecasting Indicators. The indicators selection criteria of this study benefit from previous research of Zhang [33], which analysed various factors affecting Taiwan's engineering cost indexes. The indicators must reflect four aspects: economy, finance, stock market, and energy. In addition, the building materials market is a new consideration. Through statistical analysis of related studies and expert demonstrations, six indicators were identified, namely, GDP [3, 7, 13, 34], Floor Space Started [3, 13, 19, 20, 34, 35], Crude Oil Price [3, 13, 23, 34, 36], Prime loan rate [19, 20, 35, 36], Consumer Price Index [3, 7, 12, 13, 19, 20, 23, 34, 36, 37], and Money Supply [3, 7, 13]. Due to the fact that the price information of materials has a great influence on the domestic cost indexes, the jury believes that the relevant indicators should be added to the indicator set. Then, in order to evaluate the practicability of the proposed method, eight experts in the field of engineering and construction were interviewed, including two university professors, two cost consulting experts, two engineering managers of construction units, and two technicians of design units. The experts' review is divided into two rounds. The goal of the first round is to revise the initial indicators and delete the individual indicators, which involves a level of correlation that is too low for our purposes. The second round is implemented through expert questionnaire scoring, and the opinion concentration (\bar{E}_i) and dispersion (σ_i) are calculated at the same time. When the indicator satisfies $\bar{E}_i \leq 1$ or $\sigma_i \geq 3$, it can be selected. Subsequently, 16 indicators were selected as the final index set, as detailed in Table 1.

4.2. Preparatory and Data Collection Processes. The original data used for the training of forecasting model in this research were collected from several different data

TABLE 1: Coding, type, and data sources of the index set.

| Indicator name | Brief explanation | Data source |
|---------------------------------|--|--|
| Gross domestic product (GDP) | GDP refers to the final result of the production activities of all resident units in a country (or region) calculated in accordance with the national market price in a certain period of time, and is often recognized as the more accurate indicator to measure the economic status of the country | National Bureau of Statistics of China |
| Consumer price index (CPI) | CPI is a macroeconomic indicator that reflects changes in the price levels of consumer goods and service items generally purchased by households | |
| Money supply (MS) | Money supply refers to the sum of cash and deposits in circulation at a certain point in time; money supply is one of the main economic statistical indicators compiled and published by the central banks of various countries | |
| Floor space started (FSS) | Floor space started is the construction area of each house newly started in the report period; the commencement of the house shall be subject to the date when starting to break the ground and dig the trench (foundation treatment or permanent pile driving) | China entrepreneur Investment Club |
| Crude oil price (COP) | Crude oil price is the first purchase price of a barrel of crude oil | |
| Loan rate (LR) | Loan rate is the interest rate charged to borrowers when banks and other financial institutions issue loans; this refers to the annual interest rate of the Bank of China's loans | |
| Steel bar price (SBP) | Glass price (GLP) | Shenzhen housing and Construction Bureau, Glodon Company Limited |
| Concrete price (CP) | Wood price (WP) | |
| Cement price (CEP) | Block price (BP) | |
| Medium sand price (MSP) | Diesel price (DP) | |
| Gravel price (GRP) | Material cost index (MCI) | |
| Shenzhen engineering cost index | Shenzhen engineering cost index is an indicator that reflects the degree of impact of price changes on engineering costs over a certain period of time | |

resources, including the CEIC database (<https://insights.ceicdata.com>); National Data Network (<http://data.stats.gov.cn/>); Shenzhen Construction Cost Network (<http://www.szjs.gov.cn/>); and the Wide Timber Network (<https://www.gldjc.com/>). Based on the established indicator set, the data were collected from May 2007 to March 2019, totaling 143 months of datasets. Each month's data attributes correspond to 16 indicators. A total of 143 datasets which were collected from the various data sources were identified as appropriate for use in the following model training and evaluation processes.

Due to the fact that the original data only contained raw information which could potentially lead to problems related to noise, anomalous points, missing information, errors, and frequency differences, the data were preprocessed prior to being used for training the model. Boxplot is a method to describe data using five statistics in the data: minimum, first quartile, median, third quartile, and maximum [38]. Through the boxplot method, we found one outlier (cement price). Since the number of outliers has a little impact on the total number of samples and would not reduce the effective information, the method of deleting the dataset was used. If there are a large number of outliers, they must be treated as missing values, and the data must be filled in by means of average value or max likelihood. Due to the different dimensions and orders of magnitude, min-max

standardization was adopted to perform a certain linear transformation on the original data for the purpose of making the processed data fall into the range of $[0, 1]$. Then, following the completion of the data preparatory processing, the characteristics of the dataset and the statistical results were determined, as shown in Table 2.

5. Development and Comparison of the LSTM Neural Networks

5.1. Model Creation and Performance Evaluations. Based on normal machining learning theories and pervious research [11, 24, 39], the division of training data and test data generally takes datasets of the last 5%–10% time series as test sets. Since this study uses five months as the prediction data unit, the training set and test set are divided in a 4:1 ratio, i.e., the first 80% of the data blocks was used as the training set, with the remaining 20% used as the testing set. The accuracy of each model was assessed according to its ability to predict the engineering cost indexes based on the training data. Three common statistical error measurement methods were used to evaluate the accuracy and predictability results of the models, namely, mean square error (MSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) methods. Their equations are as follows:

TABLE 2: Statistics of the primary selection index set.

| Indicator name | Mean | Median | Maximum | Minimum | Std. Dev. |
|---------------------------------|-----------|-----------|-----------|----------|-----------|
| Shenzhen engineering cost index | 136.181 | 139.260 | 177.460 | 100.000 | 22.005 |
| Material cost index | 112.432 | 113.110 | 134.290 | 99.520 | 7.528 |
| Floor space started | 10300.070 | 10297.610 | 22178.700 | 2580.100 | 4547.512 |
| Crude oil price | 572.126 | 540.860 | 948.800 | 231.870 | 188.525 |
| Loan rate | 5.700 | 6.000 | 7.470 | 4.350 | 0.918 |
| Consumer price index | 124.960 | 127.260 | 144.080 | 100.000 | 11.624 |
| Money supply | 92.915 | 89.557 | 173.986 | 31.671 | 43.158 |
| Steel bar price | 4225.671 | 4250.000 | 6050.000 | 2400.000 | 868.063 |
| Concrete price | 385.115 | 360.720 | 540.920 | 314.780 | 57.445 |
| Cement price | 491.685 | 495.000 | 630.000 | 49.000 | 68.999 |
| Medium sand price | 97.531 | 95.000 | 146.000 | 54.000 | 34.170 |
| Gravel price | 111.252 | 120.000 | 160.000 | 60.000 | 26.289 |
| Glass price | 38.979 | 35.000 | 52.000 | 30.000 | 7.864 |
| Wood price | 1169.238 | 1200.000 | 1499.000 | 1000.000 | 139.684 |
| Block price | 279.203 | 290.000 | 290.000 | 233.000 | 14.054 |
| Diesel price | 7.325 | 7.350 | 9.360 | 5.470 | 1.110 |
| Gross domestic product | 14125.420 | 14035.170 | 25269.960 | 6015.410 | 5116.898 |

$$\begin{aligned}
 \text{MSE} &= \frac{1}{N} \sum_{t=1}^N [\hat{Y}(t) - Y(t)]^2, \\
 \text{MAE} &= \frac{1}{N} \sum_{t=1}^N |\hat{Y}(t) - Y(t)|, \\
 \text{MAPE} &= \frac{1}{N} \sum_{t=1}^N \frac{|\hat{Y}(t) - Y(t)|}{Y(t)} \times 100.
 \end{aligned} \tag{1}$$

5.2. LSTM Neural Network Development

5.2.1. Data Input. Due to the fact that cyclic calculations have unique types, the original data were required to be segmented by a time-step process, and the input of each time step was required to correspond to all of the index characteristics of a single time point. In addition, due to the powerful synthesis ability of the LSTM at different time steps, the index feature information of multiple time points could be gradually extracted and synthesized within the movements of the time steps, which allowed for the extracted feature vectors to be increasingly powerful in their expressions of the input data. Therefore, it was determined that during the data processing the data should be divided into blocks, with each block containing the input index characteristics of multiple time steps.

The data used in this study were divided into a set of input data representing every five-month period in order to predict the cost indexes of the following sixth month. The model structure of the LSTM was organized by groups, as shown in Figure 1. Each group contained five datasets, each of which included 16 indicators. Meanwhile, the dataset for the first month of each group contained 17 indicators, due to the fact that the engineering cost index value of the first month (T1) had been added to the dataset as a new feature. This was mainly due to the fact that the data for the first month had been utilized to predict the cost index for the

second month. Therefore, there was no output value in the first month. However, there was an output value (T2') beginning from the second month, and T2' was the comparison between the predicted value for the first month and the actual value for the second month, referred to as the calculated residual. The values of engineering cost indexes from the second month to the fifth month (T2–T5) had been placed on the output end, which was then compared with the predicted values as the comparison data. Due to the fact that the influence results were transmitted to the forecast of the next month, their input only included 16 indicators.

5.2.2. Model Training. During the process of developing deep neural networks of LSTM, a method of multilayered LSTM was applied for the structural combinations, as shown in Figure 2. LSTM' represents the given initial network model, [X1, . . . , X16, Y] represents 16 indicators and engineering cost index per month. The parameter learning processes for multiple LSTMs were conducted from the LSTM. In order to match the actual application scenarios, the mode structure had specifically deployed an analysis pattern, in which every month had its own LSTM module trained using the data of the current month. The results are output after the multilayer structure training is completed. Then, the results of loss and gradient calculation are input into the multilayer LSTM structure, and finally a cyclic network is formed.

MXNet is one of Amazon's most powerful deep learning frameworks. Currently, distributed machine learning platforms that support time series prediction based on LSTM include MXNet, PyTorch, and Caffe2. Compared with other deep learning frameworks, MXnet has the advantages of strong readability, ease of learning, high parallel efficiency, and memory saving [40]. It also supports multi-GPU training, multiple language interfaces, and multiple devices. There are two main reasons for choosing MXNet as the development framework to be used in this study. One is to ensure the high efficiency of the model in the actual

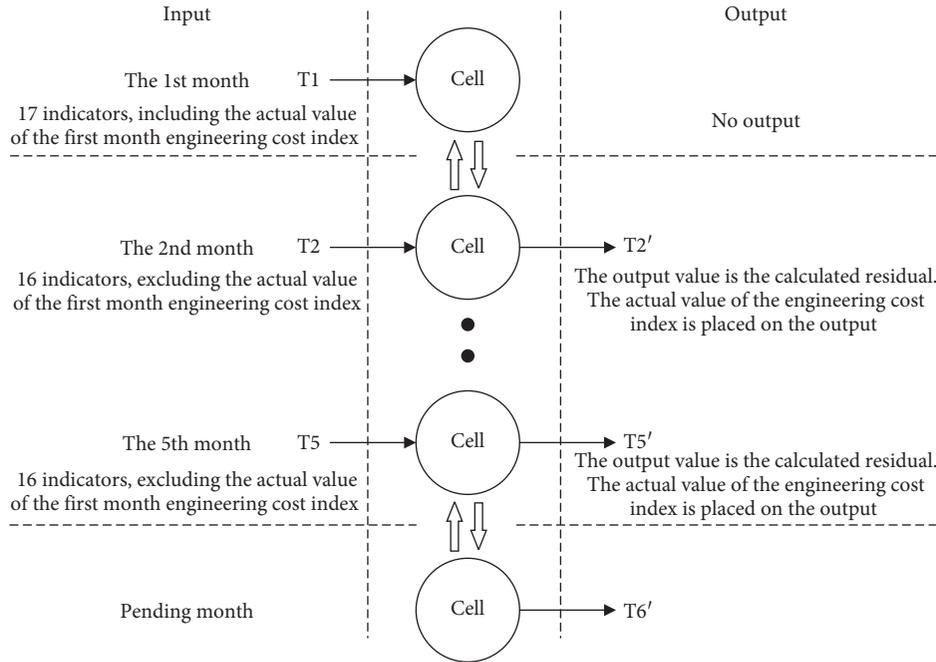


FIGURE 1: LSTM model data input structure.

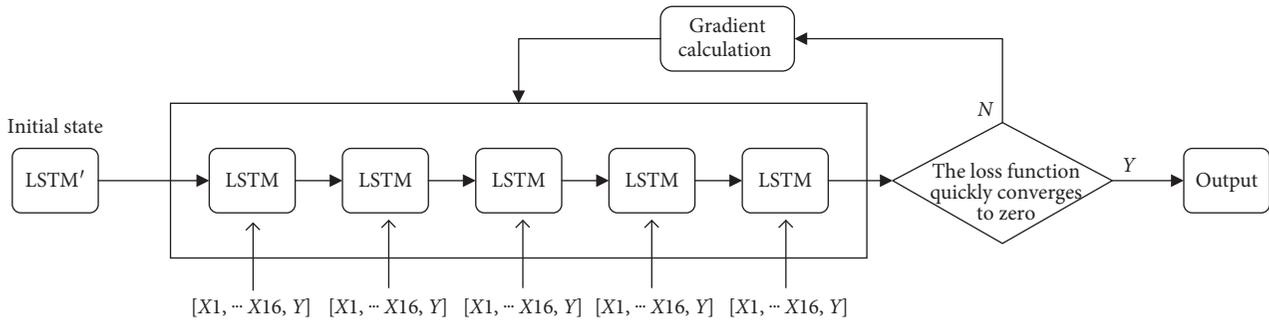


FIGURE 2: Training structure of the LSTM model.

application scenario, and another is to use the MXNet’s automatic gradient function and the packaged optimizer. In present study, two libraries in MXNet are used to build the entire LSTM model, namely, NDAarray and Autograd. The former is used to store and process data, while the latter can automatically derive the model parameters to achieve reverse gradient propagation. The parameters needed for model establishment based on MXNet include input gate, forget gate, output gate, candidate memory cells, weight parameters for the output layer, and migration parameters. The weight parameters were randomly initialized to a normal distribution, with a standard deviation of 0.01 and a mean of 0. The migration parameters were all initialized to 0, and gradients were created for all the network parameters. The network parameters were then connected to the network structure by an LSTM computing mode.

The LSTM model training process is shown in Figure 3. In the internal structure of LSTM NN, the calculation processes of the input gate, forget gate, output gate, and candidate memory cells consisted of the output of the input

layer and the last hidden state. For example, the 17 characteristic variables $x_t = [f_1, f_2, \dots, f_{17}]$ of the current month and the valid information retained in the previous month. The calculations of the current memory cell were controlled by the output of both the input gate and the forget gate. The forget gate controlled the reading of the cell state of the previous month, and the input gate controlled the inflow of the current candidate cell state. Finally, the current hidden state was calculated by the output gate and the current memory cell, and it had flowed into the next month’s calculations along with the current cell state.

The loss function L2 was calculated by comparing the predicted values calculated each month with the actual values placed in the output layer. During the training phase, L2 was taken as the optimization goal, and the L2 loss function was defined as follows:

$$L2 = \frac{1}{2} \sum_i |\text{label}_i - \text{pred}_i|^2 \tag{2}$$

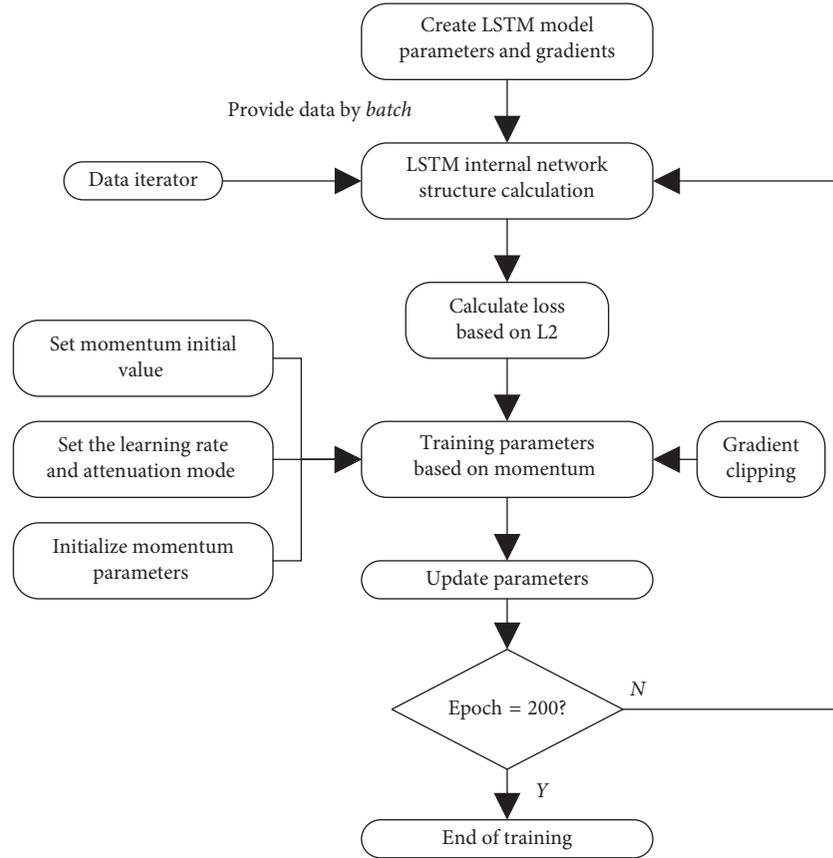


FIGURE 3: Training process of the LSTM model.

A momentum method was used as the model's optimization algorithm. By introducing the intermediate variables, the gradients in the irrelevant direction were cancelled both positively and negatively, which overcame the problems of slow convergence or even nonconvergence caused by the gradients swinging back and forth in the nondescending direction which had been encountered in the traditional gradient descent methods. The updated parameter formulas for each iteration of the momentum method were as follows:

$$\begin{aligned} v_t &\leftarrow \gamma v_{t-1} + \eta_t g_t, \\ \theta_t &\leftarrow \theta_{t-1} - v_t, \end{aligned} \quad (3)$$

where v_t represents the current momentum; η_t is the current learning rate; g_t indicates the current gradient; θ_t is the updated parameter; and γ denotes the momentum parameter. The intermediate momentum v_t needed to be initialized to v_0 when using the momentum method optimizer. After that, the model initialization state, data generator, learning rate, and learning rate attenuation mode could be set to perform the model parameter training. During the training stages, the calculated gradients were clipped to prevent gradient explosions during the process of backpropagation which would cause the model to diverge. The gradients after clipping were as follows:

$$\min\left(\frac{0}{\|g\|}, 1\right)g. \quad (4)$$

After the single batch gradient calculation was completed, the network parameters were updated by the momentum method optimizer. The Stochastic Gradient Descent Momentum (SGDM) was able to achieve faster parameter updates and the model had displayed improving converging ability. Finally, the trained model parameters were saved for future use during the prediction phase.

In addition, the detailed methods used in the training process included batch processing and flow training. With consideration given to the performance levels of the computers used in this study, the batch size of the training set was established as 5; the batch size of the test set was 1; and the learning rate was initialized to 0.1 after the in learning of 200 epochs was completed. Then, in order to avoid the model parameter values becoming too large, this method used a weight decay technique with a value of $5e-4$ and a gradient clipping parameter value of 0.1, which was able to achieve the effects of regularity for the L2 parameter. The model total time step (num_steps) was set as 5 and a data block consisted of datasets for every 5 months. The input value for the single step was the feature number (FeatureNumber = 17) which represented the 17 characteristic variables for the actual values of the engineering cost index. The output vector of each time step moving through a single fully connected layer contained a single value (num_outputs = 1) which indicated the only prediction of the engineering cost index for one month.

5.2.3. *Model Validation and Analysis Result.* After the development of the LSTM model, the decreases in the output for the loss function with the number of learning iterations could be used to determine the convergence and fitting effects of the model. As shown in Figure 4, the values of the loss function for the LSTM model had rapidly converged to near zero as the number of learning iterations increased during the training process. When the number of learning iterations reached 200, the value of loss function was observed to be almost zero, which indicated that the model had achieved a good convergence and could be used for the predictions of the test sets.

The differences observed in the fitting effects between the predicted and real values are displayed in Figure 5. These observed differences indicated that the trend patterns of the two curves were almost same and that the LSTM model had achieved improved prediction results.

As can be seen in Figure 6, the error values of the LSTM prediction model were extremely small, and it was slightly biased around the value of 0. Negative error means that the predicted value is lower than the actual value and the positive errors are the opposite. The maximum error value was only -2.03 , and the minimum error value was determined to be -0.7 . In addition, the mean absolute error (MAE) of this study's 27 test sets was only 0.96. Therefore, the prediction effects had met the prediction requirements. The mean square error (MSE) and the mean absolute percentage error (MAPE) were also selected to evaluate the prediction accuracy of the LSTM model. The calculation results are detailed in Table 3. The MAPE of this study's LSTM model was 0.71%, and the prediction accuracy had reached 99.29%, which was adequate to show the capacity of the LSTM neural networks to utilize the long-distance dependence information in the sequence data.

5.3. *Prediction Performance Comparison with SVM Model.*

For comparing the performance of the proposed model, this paper selects the current advanced SVM algorithm as the comparison object and trains the model based on the same dataset. The predicted results are shown in Figure 7, and the error results are shown in Table 4.

Through comparison, it is found that LSTM has advantages in terms of both prediction accuracy and parameter adjustment. The accuracy of the SVM model is 98.01%, while that of the LSTM model is 99.29%, and the fitting effect of the LSTM model is better. The LSTM model's fluctuation level of the absolute error and mean square error are smaller than that of the SVM model. In addition, the SVM model only involves two parameters, the penalty term "C" and the kernel function difference coefficient "gamma." However, there is no universally accepted method for determining these. The conventional approach is to take values based on experience within a certain range, then gradually narrow the range by comparing the MSE after training to determine the stronger parameters. Although LSTM involves many parameters, and generally the input value, output value, and hidden layer, the neuron number must be adjusted. The weights and thresholds are randomly assigned, and the parameters are

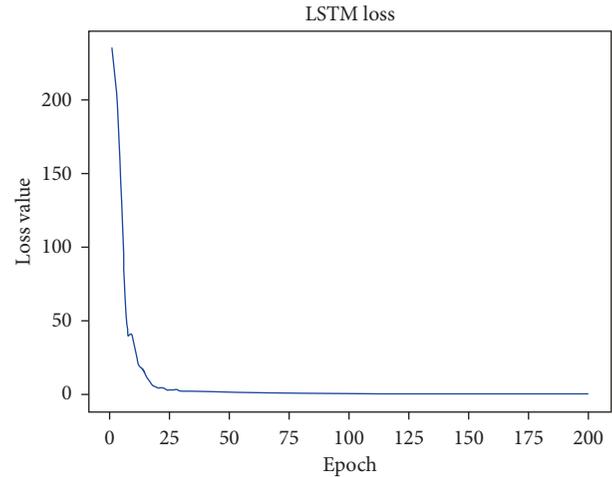


FIGURE 4: Trend of the loss function of the LSTM training set.

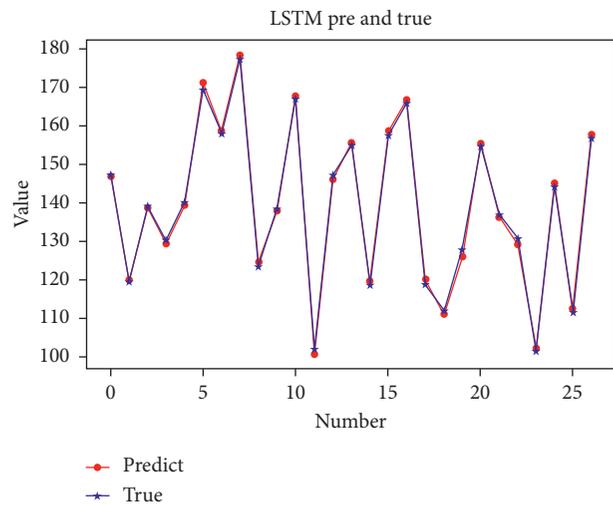


FIGURE 5: Prediction fitting of the LSTM model.

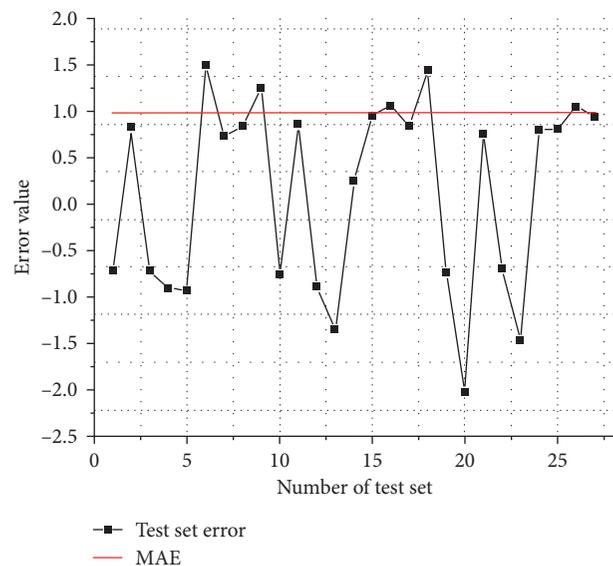


FIGURE 6: Prediction error trend of the LSTM model.

TABLE 3: Prediction error and accuracy results of the LSTM model.

| Model | Absolute maximum error | Absolute minimum error | MAE | MSE | MAPE | Model accuracy |
|-------|------------------------|------------------------|------|------|------|----------------|
| LSTM | 2.03 | 0.7 | 0.96 | 1.03 | 0.71 | 99.29 |

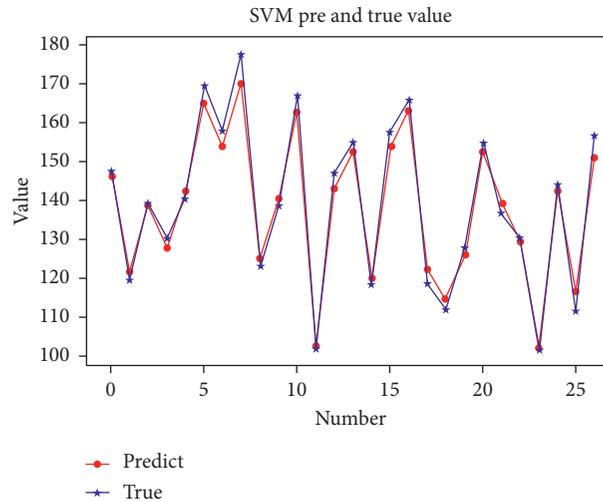


FIGURE 7: Prediction error trend of the SVM model.

TABLE 4: Prediction error and accuracy results of the SVM model.

| Model | Absolute maximum error | Absolute minimum error | MAE | MSE | MAPE (%) | Model accuracy |
|-------|------------------------|------------------------|------|-------|----------|----------------|
| SVM | 7.23 | 0.74 | 2.83 | 10.51 | 1.99 | 98.01 |

updated using SGDM. Taking these aspects together, the proposed prediction framework is shown to possess certain competitiveness.

5.4. Framework Application Scenarios and Steps. The proposed framework can be applied in forecasting the short-term or long-term trend of macroeconomic situation that has great influence to the cost and financial budget of a construction project, in terms of the real practical scenarios including policy making of government departments, the investment decision-making of real estate enterprises, the rationality of technical and economic indicators of design unit, and the dispute settlement between the client and the general contractor.

Take the issue of contract risk between the client and the general contractor as an example. In the bidding stage, the contracting company usually gives harsh bidding conditions for the price adjustment of building materials, which often makes the construction units in a passive situation. The proposed framework can avoid the risk of the construction party to a certain extent. The specific steps are as follows. Firstly, the construction units can quickly establish a training team within the validity period of the tender to collect the indicators of the current period and previous years and use it as the original training data. Secondly, the team members predict the monthly engineering cost indexes during the construction phase based on the proposed model and the construction period. Finally, judge the rationality of

the relevant requirements of the bidding documents according to the change range of the cost indexes between the completion period and the current period. If the requirements are reasonable, the construction units will normally participate in the bidding. Instead, they can apply to negotiate with general contractor or abandon the bid to minimize their own risks. In summary, the proposed framework has practical value to assist bidding decisions.

5.5. Analysis and Optimization of the Prediction Accuracy of the LSTM Model. The aforementioned research results showed that the proposed LSTM neural network model was suitable in prediction applications of construction engineering cost indexes. However, during the process of creating the LSTM model, it was found that there was no standard method available for sample selections, parameter settings, setup of time series lengths, and the designing of the model structure. Generally speaking, the setup of the model was in accordance with previous experience. However, it was accepted that the selection of the various samples and other model settings would potentially affect the prediction performance of the model. Therefore, it was necessary in this study to discuss the mechanisms and optimization of the parameter selections and model settings for the development process of the LSTM model.

5.5.1. Input Feature Analysis. There are many factors which may potentially affect the predictions of construction project

cost indexes. These factors can mainly be divided into four categories: economic, energy, construction market, and all indicators.

In the present study, in accordance with the aforementioned four groups of indicators, the following four models were established, and a basic model of all the indicators was used as a comparison model in order to explore the impacts of the input features on the engineering cost indexes. The prediction results of the other three models were obtained by modifying the input sample dimensions of the base model. The mean square error and prediction accuracy of the model were then successfully calculated. The results are shown in Table 5.

The absolute error values of the prediction results were calculated according to the prediction results of the 27 test sets. In order to compare the error values of the models and their stability, the absolute error values of the predictions of the four models were determined, as described in Figure 8.

As previously illustrated in Table 5, the prediction accuracy of Model M1 had reached 98.71%. Therefore, it was also confirmed to be appropriate to use the LSTM model of the economic indicators to predict the engineering cost indexes. In this study's comparison of Model M1 and Model M12, it was observed that the LSTM model with energy indicators added was more effective. However, although the prediction accuracy had been improved, the absolute error fluctuations of Model M1 were found to be similar to those of Model M12, which indicated that the energy indicators had only increased the amount of training information and not the amount of effective information. Then, by comparing Models M1, M12, and M123, it could be seen that the prediction accuracy of the LSTM model had gradually improved, and the performance results of the model had become increasingly more stable. These findings indicated that when the dimension of input data was small, appropriately increasing the input features of the model could potentially improve the overall prediction accuracy of the model.

This study then compared the four models in combination with Table 5 and Figure 8. It was found that the prediction accuracy of Model M3 was almost the same as that of Model M123 and the stability levels of the models were similar, both of which were better than those of Model M1 and Model M12. Therefore, using this study's experimental results, it was determined that the indicators related to the construction market had major impacts on the predictions of the construction engineering cost indexes. Then, by comparing Model M12 with Model M123, it was found that the prediction accuracy of the model had increased from 98.87% to 99.29%, which again showed that the indicators related to the construction market had major influences on the predictions of the engineering cost indexes. Furthermore, the results also indicated that the prediction accuracy of the model could be improved by the appropriate addition of effective input information. In addition, by outputting the loss values of the training set and test set, it was observed that the loss functions of the four models had all decreased rapidly, with good convergence and no occurrences of overfitting.

In summary, among the three types of indicators, economic indicators, energy indicators, and construction market indicators, the construction market indicators were found to have the most significant impacts on the predictions of the engineering cost indexes and could be used as effective information for the proposed model. It was observed that when increasing or decreasing the dimensions of the input features, the dimensions of the input data were small; appropriately increasing the effective information could potentially improve the prediction accuracy of the model. However, when the dimensions of the input data were larger, the prediction accuracy of the model could not be greatly improved. In such cases, even redundancy of the input information may occur, which could potentially reduce the accuracy of the model. Therefore, it was determined in this study that the economic, energy, and construction market indicators should be used as the input features for the proposed model, which would improve the prediction accuracy of the LSTM model. It was also believed that if the data collection was difficult, the construction market indicators could be directly used as the input features.

5.6. Time Series Length Analysis. The length of the time series may also affect the prediction accuracy of a model. The length of a time series is usually obtained from the analysis of specific problems, and there currently is no standard determination method. In the present study, 16 indicators were used as the input variables, and the data were processed into time series of lengths of 3, 5, 7, and 10, respectively. Then, the model was established and trained. The results are shown in Table 6. Since the time series lengths of each model had varied, it was necessary to redivide the training set and test set of each model. The test sets were extracted by a random function in order to ensure that the prediction accuracy of the test sets also represented the prediction accuracy of the model. Then, in accordance with the prediction results of the 27 indicators of each model test set, the absolute error values were calculated. The results are shown in Figure 9.

As can be seen in Table 6, the prediction accuracy of Model M123-d3 was lower than that of the other two models. Meanwhile, as shown in Figure 9, the absolute errors of Model M123-d3's predictions had fluctuated greatly, and the stability of the model's performance was obviously lower than that of the other three models. In this study's comparison results of Models M123-d5 and M123-d7, it was found that the accuracy levels had slightly decreased, which may have been caused by the different test sets. The stability levels of the aforementioned two models were also found to be similar. Then, by comparing the three models with the time series length of 5, 7, and 10, it could be seen that the prediction accuracy results were close, and the stability levels of the models' performances were also similar. That is to say, when the time series lengths had increased, the prediction accuracy of the models had first improved and then almost remained unchanged. Similarly, the loss functions of the four models still converged rapidly to 0, without any overfitting observed.

TABLE 5: Effects of the input features on the model’s accuracy.

| Model | Dimension | Category | Indicators | MSE | Model accuracy |
|-------|-----------|---------------------|---|------|----------------|
| M1 | 4 | Economy | LR, CPI, MS, GDP | 3.92 | 98.71 |
| M3 | 10 | Construction market | MCI, FSS, SBP, CP, CEP, MSP, GRP, GLP, WP, BP | 1.27 | 99.22 |
| M12 | 6 | Economy + energy | MCI, CPI, MS, GDP, COP, DP | 3.16 | 98.85 |
| M123 | 16 | All | All | 1.03 | 99.29 |

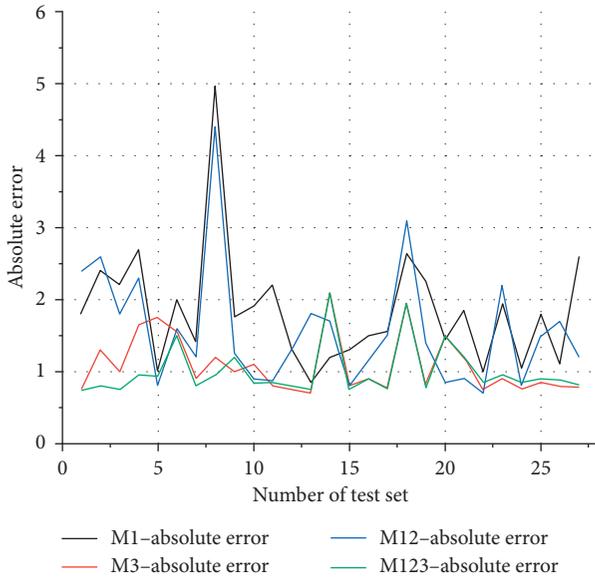


FIGURE 8: Predictions of the absolute error trends of the four models with different input features.

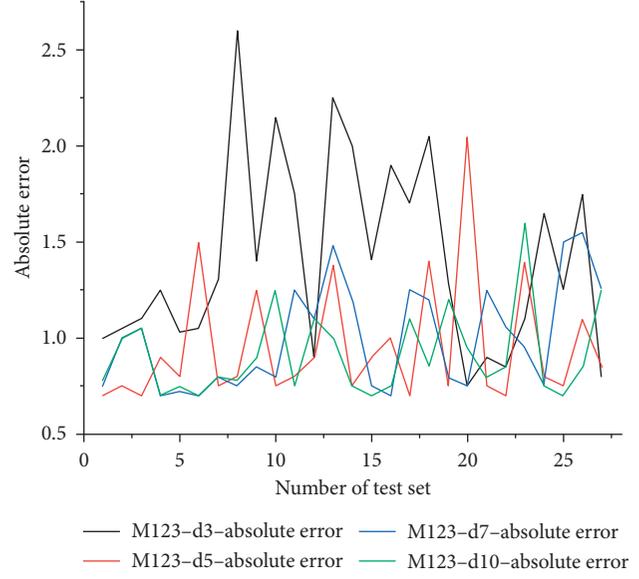


FIGURE 9: Predictions accuracy and absolute error trends of the models with different time lengths.

TABLE 6: Effects of the time series lengths on the accuracy rates of the models.

| Model | Time series length | Mean square error | Model accuracy (%) |
|----------|--------------------|-------------------|--------------------|
| M123-d3 | 3 | 2.27 | 98.95 |
| M123-d5 | 5 | 1.03 | 99.29 |
| M123-d7 | 7 | 1.10 | 99.22 |
| M123-d10 | 10 | 0.88 | 99.33 |

It was observed in this study that when the time series length was excessively short and the effective information provided by the samples was insufficient, the proposed LSTM model could not learn the transformation rules of the training samples, which led to a low accuracy rate of the model. However, because the further the data were taken from the predictive period, the smaller the prediction impacts on the data of the prediction period would be, the prediction accuracy of the model was not significantly improved when the time series length had been increased. Moreover, the longer the time series was, the more noise it would contain, which is not conducive to the accurate predictions of the model. In summary, the time series length has a certain influence on the prediction accuracy, but the training cost is more sensitive to its change. With the increase of time series length, the improvement of training cost is much higher than the prediction accuracy. Therefore, it is necessary to select the appropriate time series length to

improve the application efficiency of the model. In the present study, it was observed that the M123-d5 and M123-d10 Models exhibited the highest prediction accuracy. The accuracy rates of the two models were found to be analogous, although the training duration of Model M123-d10 was longer. Subsequently, the time series length was set as 5 in this research, in order to achieve an improved model performance.

5.7. Analysis of the Model Structure. For LSTM neural networks, the number of hidden layer neurons determines the structure of the neural network model. However, there is currently no unified method which can be applied to determine the number of neurons in a hidden layer. In this study, by comparing the prediction accuracy rates of the model under the conditions of various numbers of neurons in the hidden layer, the most suitable number of neurons was selected.

Therefore, on the basis that all 16 indicators were used as the input variables of the model and the time series length was set as 5, the number of hidden layer neurons was set to the value of 10 times between 10 and 150. The model’s training and the prediction results were successfully obtained, as shown in Table 7.

It was observed in this study that when the number of hidden layer neurons increased from 10 to 40, the mean square errors of the model’s predictions gradually decreased

TABLE 7: Effects of the model's structure on the accuracy results.

| Model | Model structure | Mean absolute error | Mean square error | Accuracy (%) |
|-----------|-----------------|---------------------|-------------------|--------------|
| M123-h10 | 16-10-1 | 1.11 | 1.37 | 99.18 |
| M123-h20 | 16-20-1 | 1.07 | 1.28 | 99.21 |
| M123-h30 | 16-30-1 | 1.07 | 1.23 | 99.21 |
| M123-h40 | 16-40-1 | 0.85 | 0.89 | 99.34 |
| M123-h50 | 16-50-1 | 0.96 | 1.02 | 99.29 |
| M123-h60 | 16-60-1 | 1.02 | 1.14 | 99.24 |
| M123-h70 | 16-70-1 | 1.05 | 1.26 | 99.21 |
| M123-h80 | 16-80-1 | 1.06 | 1.22 | 99.21 |
| M123-h90 | 16-90-1 | 1.01 | 1.15 | 99.27 |
| M123-h100 | 16-100-1 | 1.25 | 1.67 | 99.11 |
| M123-h110 | 16-110-1 | 1.05 | 1.16 | 99.24 |
| M123-h120 | 16-120-1 | 1.14 | 1.41 | 99.17 |
| M123-h130 | 16-130-1 | 1.16 | 1.45 | 99.15 |
| M123-h140 | 16-140-1 | 1.09 | 1.29 | 99.20 |
| M123-h150 | 16-150-1 | 1.13 | 1.48 | 99.18 |

until reaching a minimum value. In addition, the prediction accuracy rate of the model gradually increased until the maximum value was attained. Furthermore, as the number of hidden layer neurons continued to increase, the prediction accuracy rate of the model did not improve, and when the number of hidden layer neurons increased to approximately 100, then the accuracy of the model's results tended to fluctuate. It was determined in this study that too many or too few hidden layer neurons would potentially reduce the prediction accuracy of the model. For example, if the number of hidden layer neurons was too small, then the underfitting of the model led to increased prediction errors. Meanwhile, if the number of hidden layer neurons was too great, then the prediction accuracy of the model tended to not be improved, which had a tendency to lead to the occurrence of unstable phenomena and overfitting of the model.

6. Conclusions

In this article, we proposed a prediction model based on an LSTM neural network, which is suitable for the short-term engineering cost indexes prediction or other cost data with temporal or spatial properties. The proposed model will be applied to the feasibility study stage or bidding stage of the project. It can provide accurate industry trends so that all engineering participants can evaluate the project risk in a comprehensive manner in advance, which is helpful to formulate relevant response plans. This research makes significant contributions in terms of new emerging tools and new AI algorithm for the traditional field of construction cost index prediction. Firstly, the new emerging tools are originally applied in this area after reviewing previous research results. Although LSTM NN has been used in prediction problems in other application areas, there is a lack of explorative research to train the algorithm model by using specific construction date and evaluate the forecasting results for the theoretically suitability of cost index prediction. Secondly, the new AI algorithm of LSTM NN has the ability to sort out the limitations of existing methods in cost index prediction. Since

most of the traditional methods are not suitable for nonlinear fitting and have poor response to the timeliness of the data, LSTM NN has advantages in dealing with limitations of the gradient vanishing and the inability to address long-term dependence. Upon analysing the experimental results of the LSTM model, the following key findings are observed. (1) Sixteen prediction indicators can comprehensively and timely reflect the domestic economic, energy, and market conditions, which meet the requirements of capturing the fluctuation trend of the engineering cost indexes. (2) The proposed LSTM model has good fitting effect and small prediction error, which fully demonstrates the ability of the algorithm to utilize long-distance dependent information in sequence data. (3) Through the optimization mechanism in three aspects, the experience of model creation is successfully converted into principle standards, in which the optimization of the input features is the most critical. (4) Compared with other methods, the LSTM model possesses significant advantages in training cost, time series process, and short-term prediction accuracy. This model can be used to deal with similar time series, such as crowd or vehicle flow, and stock prices. Generally speaking, it was confirmed in this study that the LSTM neural networks were applicable and effective in regard to predictions of construction cost indexes. The obtained research findings of this study could potentially provide some guidance for subsequent researchers in selecting prediction algorithms and model parameters. However, the proposed method framework still has some limitations, such as the following. First, the data required for the research were mainly taken from four domestic databases, and the authenticity of these historical data lacks verification. Second, the index determination criteria used in this article lack authority, and different countries or organizations may involve various criteria. Finally, due to the limited amount of statistical data available in China at present, this study only validated the short-term prediction performance of LSTM. Based on the above limitations, our future work will focus on improving the structural layer of LSTM, in order to compensate for its disadvantages in the long-term prediction process.

Data Availability

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

Conflicts of Interest

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

References

- [1] S.-H. Ji, J. Ahn, H.-S. Lee, and K. Han, "Cost estimation model using modified parameters for construction projects," *Advances in Civil Engineering*, vol. 10, p. 2019, Article ID 8290935, 2019.
- [2] M. G. Meharie, Z. C. Abiero Gariy, R. N. N. Mutuku, and W. J. Mengesha, "An effective approach to input variable selection for preliminary cost estimation of construction projects," *Advances in Civil Engineering*, vol. 2019, p. 14, Article ID 4092549, 2019.
- [3] B. Ashuri, S. M. Shahandashti, and J. Lu, "Empirical tests for identifying leading indicators of ENR construction cost index," *Construction Management and Economics*, vol. 30, no. 11, pp. 917–927, 2012.
- [4] Y. Elfahham, "Estimation and prediction of construction cost index using neural networks, time series, and regression," *Alexandria Engineering Journal*, vol. 58, no. 2, 2019.
- [5] 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. 47, no. 5, pp. 1–8, 2019.
- [6] T. Biquiu, Z. Sai, and H. Jia, "Establishment of engineering cost prediction system based on BIM and ELM," *Construction Technology*, vol. 17, 2018.
- [7] J. M. W. Wong and S. T. Ng, "Forecasting construction tender price index in Hong Kong using vector error correction model," *Construction Management and Economics*, vol. 28, no. 12, pp. 1255–1268, 2010.
- [8] B. Ashuri and J. Lu, "Time series analysis of ENR construction cost index," *Journal of Construction Engineering and Management*, vol. 136, no. 11, pp. 1227–1237, 2010.
- [9] M. Innocent, J. S. Wasek, and A. Franz, "Predicting military construction project time outcomes using data analytics," *Engineering Management Journal*, vol. 30, no. 4, pp. 232–246, 2018.
- [10] S. Islam, "Provenance, lineage, and workflows," Master Thesis, Brown University, Providence, RI, USA, 2010.
- [11] G. B. Hua and T. H. Pin, "Forecasting construction industry demand, price and productivity in Singapore: the Box-Jenkins approach," *Construction Management and Economics*, vol. 18, no. 5, pp. 607–618, 2000.
- [12] J.-w. Xu and S. Moon, "Stochastic forecast of construction cost index using a cointegrated vector autoregression model," *Journal of Management in Engineering*, vol. 29, no. 1, pp. 10–18, 2011.
- [13] S. M. Shahandashti and B. Ashuri, "Forecasting engineering news-record construction cost index using multivariate time series models," *Journal of Construction Engineering and Management*, vol. 139, no. 9, pp. 1237–1243, 2013.
- [14] A. Akintoye and M. Skitmore, "Macro models of UK construction contract prices," *Civil Engineering and Environmental Systems*, vol. 10, no. 4, 1993.
- [15] 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.
- [16] H. L. Chen, "Developing cost response models for company-level cost flow forecasting of project-based corporations," *Journal of Management in Engineering*, vol. 23, no. 4, 2007.
- [17] T. Moon and D. H. Shin, "Forecasting construction cost index using interrupted time-series," *KSCE Journal of Civil Engineering*, vol. 22, no. 5, pp. 1626–1633, 2018.
- [18] J. M. W. Wong, A. P. C. Chan, and Y. H. Chiang, "Time series forecasts of the construction labour market in Hong Kong: the Box-Jenkins approach," *Construction Management and Economics*, vol. 23, no. 9, pp. 979–991, 2005.
- [19] S. Hwang, "Dynamic regression models for prediction of construction costs," *Journal of Construction Engineering and Management*, vol. 135, no. 5, pp. 360–367, 2009.
- [20] S. Hwang, "Time series models for forecasting construction costs using time series indexes," *Journal of Construction Engineering and Management*, vol. 137, no. 9, pp. 656–662, 2011.
- [21] M. Juszczuk and A. Leśniak, "Site overhead cost index prediction using RBF neural networks," in *Proceedings of the 3rd International Conference on Economics and Management (ICEM 2016)*, DEStech Transactions on Economics, Business and Management, pp. 381–386, Suzhou, China, 2016.
- [22] H. Nam, S. H. Han, and H. Kim, "Time series analysis of construction cost index using wavelet transformation and a neural network," in *Proceedings of the 24th International Symposium on Automation & Robotics in Construction (ISARC 2007)*, Chennai, India, September 2007.
- [23] M.-Y. Cheng, N.-D. Hoang, and Y.-W. Wu, "Hybrid intelligence approach based on LS-SVM and differential evolution for construction cost index estimation: a Taiwan case study," *Automation in Construction*, vol. 35, pp. 306–313, 2013.
- [24] J. Wang and B. Ashuri, "Predicting ENR'S construction cost index using the modified K nearest neighbors (KNN) algorithm," in *Proceedings of the Construction Research Congress*, San Juan, Puerto Rico, May 2016.
- [25] W.-J. Hwang and K.-W. Wen, "Fast kNN classification algorithm based on partial distance search," *Electronics Letters*, vol. 34, no. 21, pp. 2062–2063, 1998.
- [26] H. Wang and D. Hu, "Comparison of SVM and LS-SVM for regression," in *Proceedings of the 2005 International Conference on Neural Networks and Brain*, October 2006.
- [27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [28] Y. Tian and P. Li, "Predicting short-term traffic flow by long short-term memory recurrent neural network," in *Proceedings of the IEEE International Conference on Smart City/Socialcom/Sustaincom*, Chengdu, China, December 2016.
- [29] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Transactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.
- [30] W. Shihao, Z. Qinzhen, Y. Han, L. Qianmu, and Q. Yong, "A network traffic prediction method based on LSTM," Zte Communications, Shenzhen, China, 2019.
- [31] X. Ran, Z. Shan, Y. Fang, and C. Lin, "An LSTM-based method with attention mechanism for travel time prediction," *Sensors*, vol. 19, no. 4, p. 861, 2019.
- [32] M. Långkvist, L. Karlsson, and A. Loutfi, "A review of unsupervised feature learning and deep learning for time-series modeling," *Pattern Recognition Letters*, vol. 42, 2014.

- [33] Y. Zhang, "Forecasting the trend of construction cost indices for Taiwan employing support vector machine," *Master's thesis, Chao Yang University of Technology*, Taichung, Taiwan, 2007.
- [34] L. Jian, "Is the information available from historical time series data on economic, energy, and construction market variables useful to explain variations in ENR construction cost index?" in *Proceedings of the Construction Research Congress*, West Lafayette, Indiana, May 2012.
- [35] T. P. Williams, "Predicting changes in construction cost indexes using neural networks," *Journal of Construction Engineering and Management*, vol. 120, no. 2, pp. 306–320, 1994.
- [36] M.-T. Cao, M.-Y. Cheng, and Y.-W. Wu, "Hybrid computational model for forecasting taiwan construction cost index," *Journal of Construction Engineering and Management*, vol. 141, no. 4, p. 04014089, 2015.
- [37] T. Moon and D. H. Shin, "Forecasting model of construction cost index based on VECM with search query," *KSCE Journal of Civil Engineering*, vol. 22, no. 8, pp. 2726–2734, 2018.
- [38] J. W. Tukey, *Exploratory Data Analysis*, Addition-Wiley, Boston, MA, USA, 1997.
- [39] J. Wang and B. Ashuri, "Predicting ENR construction cost index using machine-learning algorithms," *International Journal of Construction Education and Research*, vol. 13, no. 1, pp. 47–63, 2017.
- [40] T. Chen, L. Mu, L. Yutian et al., "MXNet: a flexible and efficient machine learning library for heterogeneous distributed systems," *Journal of Computer Science*, <https://arxiv.org/abs/1512.01274>, 2015.