Expressway Project Cost Estimation With a Convolutional Neural Network Model

XIAOJUAN XUE[1,2], YUANHUA JIA[1], AND YUANJIE TANG[1], (Member, IEEE)
[1]School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China
[2]CCCC First Highway Consultants Company Ltd., Xi'an 710075, China

Corresponding author: Yuanjie Tang (sd_tangtang@163.com)

ABSTRACT With the development of the economy, the prediction of expressway project costs has gained increasing research attention. In this study, based on the convolution neural network (CNN) algorithm, the prediction of the expressway construction cost was analyzed with respect to the conceptual design stage. By summarizing the existing research results, 10 new factors related to the bridge and tunnel are creatively introduced into the cost-prediction index of the expressway conceptual design stage. In addition, the data structure of the expressway project cost prediction is defined and a CNN model is established. Finally, the project information of 415 expressways in China collected in this study is used to verify the research results. The results of the case analysis show that the 10 new indexes related to the bridge and tunnel can improve the prediction accuracy of the model. In addition, the CNN model is more suitable for solving the high-dimensional nonlinear problem of expressway cost prediction than the conventional artificial-neural-network and regression-analysis models, and it can improve the prediction accuracy. The findings of this study can be used to devise financial plans in the early stage of expressway construction and facilitate cost management at the conceptual design stage to help investors acquire project funds in advance.

INDEX TERMS Expressway project cost prediction, convolutional neural network model, artificial neural network, regression analysis.

I. INTRODUCTION

Socioeconomic development and the inclusion of other disciplines in the field of engineering project management [1] have resulted in increasing research attention on project funding over the entire life cycle of a project. For several years, persistent cost undervaluation has reflected a lack of scientific management in project costs across the industry [2]. Therefore, excellent project cost planning is essential to scientifically minimize capital risks [3] and reasonably manage limited funds [4], [5]; these are critical factors in the early stages of any project [6], [7]. Because of the uncertainties in the initial stages of a project, an accurate prediction of the project's ultimate cost is difficult [8], while focusing considerably on the improvement of cost-estimation accuracy in the design stage, i.e., before funding and other factors have been determined, is counterproductive. The gradual application of machine learning in the field of project management has allowed for

the development of a data-driven model for efficient, simple, and accurate prediction of expressway project costs [9], with important theoretical and practical significance in this field of research.

Generally, machine learning can be classified as shallow learning and deep learning. Shallow learning methods, such as support vector machines, random forests, decision trees, and K-means methods, are not suitable for expressway cost-prediction problems, which usually involve many controlled factors; these method do no work well for high-dimensional nonlinear problems. In contrast, deep learning methods, such as artificial neural networks (ANNs) and convolutional neural networks (CNNs), have been widely used owing to the advantage of learning high-dimensional nonlinear models.

In recent decades, scholars have studied several project cost prediction problems, including highway cost forecasting. For example, Hegazy and Ayed [10] used a neural network (NN) to develop a model for predicting highway project costs. They considered 18 highway projects built in Newfoundland, Canada, and used the simplex method and

---

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaochun Cheng.

a genetic algorithm to optimize the weights for computing project cost. Adeli and Wu [11] proposed a regularized NN that depended only on the training data and not on the NN structure, learning parameters, or the number of iterations required to train the system to compensate for the overfitting problem. Their approach considered the problem of noisy data for achieving improved prediction accuracy. They used the model to estimate the cost of a reinforced-concrete pavement project. Al-Tabtabai *et al.* [12] used an ANN to determine the functional relationship between changes in the construction cost of a typical highway and the parameters that caused these changes; they applied their model to the prediction of the percentage increase in the project cost.

Wilmot and Mei [13] summarized the influencing factors for the individual cost of 2827 highway projects in Louisiana from 1981 to 1995 and applied these factors to a NN model. They predicted that highway construction costs in Louisiana would double between 1998 and 2015. Sodikov [14] determined that estimating costs during the conceptual stage of a project was difficult because of the lack of preliminary information, lack of a database on road engineering costs, lack of appropriate cost-estimation methods, and other uncertain factors. Although conventional tools such as regression analysis (RA) have been applied to this uncertainty problem, statistical studies have shown no reduction in the cost-estimation errors. Sodikov [14] determined the method to solve the problem of missing data, analyzed the impact of different variables on the project costs, and proposed an ANN cost-estimation model for developing countries.

Anderson *et al.* [15] discussed the costs of highway projects and the method to estimate and manage them before the planning and construction stages to avoid cost increases. Chou and O'Connor [16] developed a web-based application called "Preliminary Item-Level Cost Estimating System" (PILCES), which uses the similarity between historical data stored in a database and input items to predict the cost of a highway project at the conceptual stage. Kim *et al.* [17] developed a two-stage approach for predicting the cost of a highway project. In the first stage, they used a case-based reasoning model to collect eight features that affected project costs and analyzed 92 cases. In the second stage, they built an estimation model based on quantity and unit costs, collected cost estimates for major projects, and validated their approach against 78 cases. Their approach allows the user to effectively control the cost within a particular budget for an efficient feasibility study and preliminary design.

Attal [18] used an NN to identify the parameters that affect the cost and duration of highway construction. Asmar *et al.* [19] developed a method similar to the program evaluation and review technique that decomposes the project work content at the conceptual stage and predicts the highway project cost by analyzing historical bidding data to determine the most likely cost. Elbeltagi *et al.* [20] established an ANN model to predict the highway project cost at the conceptual design stage. They identified 18 influencing factors, which were divided into three categories: project-specific, project

participants, and environmental. Before using an NN model to predict data, the 18 factors were reduced to 11. The authors trained the model using the historical cost data of 67 projects in Libya. Further, they showed that their model could accurately predict the cost of Libyan highway projects at the conceptual stage and meet the requirement of an error less than 20% during the conceptual design phase. Swei *et al.* [21] proposed an RA-based cost-estimation method. The validity of their model was verified using 15 pavement bidding projects in five states of the United States. Tijanic *et al.* [22] established an ANN model for estimating road engineering costs and trained the model by using the road database of Croatia. The results showed that the NN is an effective method that can be used for cost predictions in the initial design stage.

In addition to predicting highway project costs, machine learning can be used for prediction in other fields such as meteorology, industry, agriculture, and other construction projects. Some researchers predicted the cost of residential projects using an NN model [23]–[26]. Rafiei and Adeli [27] used sensor collection techniques to propose a method for the global and local health assessment of structural systems. Chao and Skibniewski [28] applied the NN method to construction technology management. Hinton and Salakhutdinov [29] studied the conversion of high-dimensional data to low-dimensional data based on the use of an NN. Rafiei and Adeli [30] investigated NN-based feature extraction. Elfahham [31] forecasted the construction cost index by using NNs, linear regression, and autoregressive time series. Based on the early regularization method, Chandanshive and Kambekar [32] considered 78 construction projects in India as examples and used an ANN model to predict the early cost of the construction projects. The prediction results of the trained NN model showed that the model can well predict the cost of construction projects at the initial stage of construction. Abd *et al.* [33] determined 25 factors that affect the construction cost, employed them as input characteristic information, and used an ANN model to study the construction project cost in Iraq. A case analysis of the construction project cost proved the effectiveness of the ANN model in terms of cost prediction. Hashemi *et al.* [34] reviewed the cost of construction projects from 1985 to 2020 and concluded that, in almost all cases, the estimation and analyses of the factors affecting the costs of a project in the initial stage of the project is important, and ANN and RA are the most commonly used methods of project cost forecasting.

The continuous development of big-data technology in recent years has led to the gradual replacement of NN models with CNN models in an increasing number of applications. Cai *et al.* [35] used a recurrent NN (RNN) and a CNN to predict the load in commercial buildings. Chen *et al.* [36] used a CNN to model a tropical cyclone. They trained the model with satellite, infrared, brightness, temperature, and microwave rain rate data of 1097 tropical cyclones between 2003 and 2014 and used the data from 188 tropical cyclones from 2015 to 2016 to optimize the model. The results showed that their proposed model could improve the

prediction accuracy. Yang *et al.* [37] proposed a dual-CNN model for predicting residual life. Their model took advantage of the powerful feature extraction capabilities of a CNN model; the model did not require the extraction of features to maximize the retention of useful information. Yang *et al.* [38] used a CNN to learn important features related to rice yields from low altitude remote-sensing images and established a model to predict grain yield. Delgado *et al.* [39] proposed a prediction analysis and modeling system and used a 2.75-million-point power-transmission-project dataset for conducting a case study. Three of the most popular cost-estimation models (linear regression, support vector regression, and ANNs) based on big data have been brought to fruition. Liu *et al.* [40] obtained the speed time series from GPS data, transformed the collected data into two-dimensional data images, learned the traffic characteristics through a CNN model, and transformed the traffic incident-detection problem into a classification problem. The effectiveness of the model was proved through experiments.

Based on the aforementioned literature review, four conclusions can be drawn.

- First, ANN and RA are the most commonly used methods of project cost forecasting, with machine learning being repeatedly applied for the cost prediction of engineering projects. Project cost prediction is closely related to the development of machine learning algorithms.
- Second, existing research on highway cost prediction has not fully considered factors that affect costs; for example, the influence of factors in terms of the type and scale of the bridge and tunnel projects have not been considered (see the data structure of Section III for a detailed discussion).
- Third, the NN models used to predict the cost of engineering projects have deficiencies such as overfitting, require excessive parameters, and involve complicated calculations. Although improved studies have been conducted, these limitations still remain to some extent.
- Fourth, CNN models provide obvious advantages over other machine learning approaches in terms of reducing overfitting, excessive parameters, and complicated calculation and have been widely used in the fields of meteorology, industry, agriculture, etc. However, research on their applicability to expressway cost prediction is still lacking.

Therefore, to overcome the shortcomings of existing studies on expressway project cost prediction, this study analyzed the problem based on the CNN-based machine learning algorithm by considering more comprehensive influencing factors. Factors that affect expressway project costs were systematically summarized and 10 factors related to the bridge and tunnel were introduced. A data structure was then constructed to establish the CNN model. In addition, two main prediction models (ANN and RA) applied in the field of engineering projects were introduced as comparative models

to verify the effectiveness of the proposed CNN model based on the collected 415 expressway project data.

II. RESEARCH METHODOLOGY

This study focused on developing a model for predicting the expressway cost at the conceptual design stage, which comprises 30% of the complete project design [19]. In the case of expressway projects, this involves the determination of corridors and basic engineering scales for bridges and tunnels. The proposed model was developed in five steps. First, a literature review was performed to obtain the factors influencing the expressway project cost. Second, these influencing factors were selected or newly introduced. Third, the characteristics of the influencing factors were reconstructed into a two-dimensional matrix and the CNN method was used to establish an expressway project cost-prediction model. The expressway project data were then collected. Finally, the CNN and ANN models by Elbeltagi *et al.* [20] were trained against historical data, and the prediction results were compared. The results were used to verify (1) the effectiveness of the 10 newly introduced indicators in improving the prediction accuracy of the model, (2) the advantages of the application and prediction accuracy of the CNN model over the conventional ANN and RA model, and (3) whether the CNN model proposed in this study meets the requirement [20] of providing a result with a prediction error of less than 20%.

The contributions of this article are two-fold:

- First, 25 influencing factors for expressway project costs were systematically summarized, 10 of which are related to the bridge and tunnel and are innovatively introduced as influencing factors of the cost forecast at the conceptual design stage of the expressway.
- Second, the influencing factors were reconstructed to establish a CNN model for predicting the expressway project cost. To the best of the authors' knowledge, this is the first study to propose the application of a CNN model for predicting the cost of expressway projects.

III. DATA STRUCTURE

The cost of an expressway project is closely related to its overall scale, structural characteristics, economic development, raw materials, labor costs, etc. Significant research has been conducted on exploring the influencing factors of engineering project costs [41], [42]. Herbsman [43] and Wilmot and Cheng [44] suggested that, in addition to labor, materials, equipment, and their inflation, the bidding volume impacted the cost of highway projects. Koehn *et al.* [45] discussed the influence of land price, interest rates, and government regulations on project bidding prices. Hegazy and Ayed [10] proposed 10 factors that affected the cost of a highway project and used these as inputs to an NN model: the project category, project scope, project implementation time, construction season, project location, project duration, project scale, project capacity, hydrological conditions, and geological conditions. Wilmot and Mei [13] proposed 11 factors that affected the
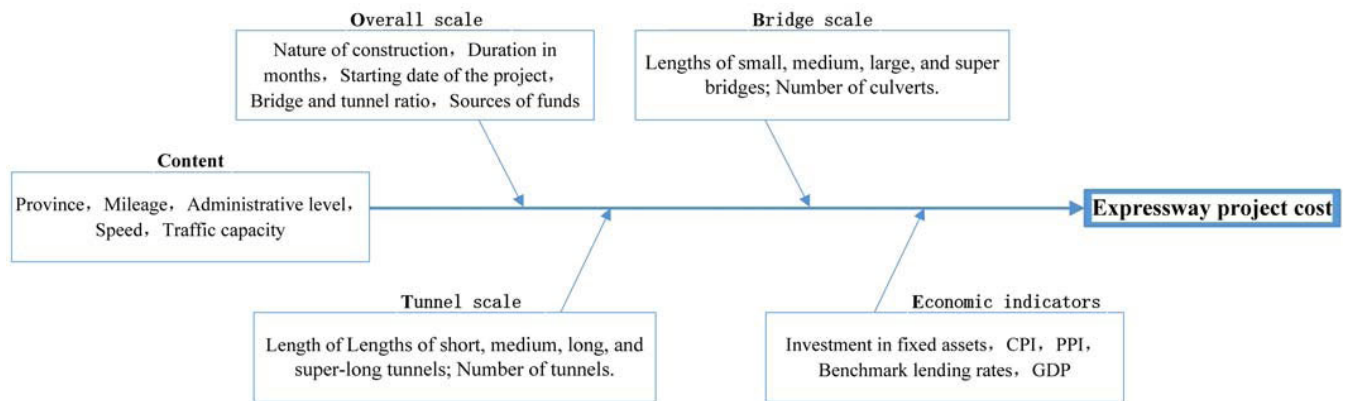
FIGURE 1. Logical structure of factors affecting the expressway cost.

cost of a highway project: the labor price, material price, equipment price, project bid quantity, contract terms, contract location, quarter of project implementation, annual bid scalar, change in bid quantity, change in project quantity, and change in specification standards. Elbeltagi *et al.* [20] summarized 18 factors that affected the cost of a highway project: the project type, construction detour, project location, construction year, project scope, project scale, project capacity, project duration, implementation season, geological type, capital status, transportation distance of materials and equipment, paving thickness, hydrological conditions, protection of public facilities, cost increase, project cooperation type, and contract interest. Rafiei and Adeli [30] proposed that the construction cost of a project was affected not only by the materials, labor, equipment and technology, project location, type, duration, and degree of material recycling but also by the fluctuations in economic variables such as the current assets, wholesale price index, and construction service level. These economic indicators may increase or decrease the cost but are difficult to characterize, and as such, are often ignored by conventional cost-prediction models.

Two types of problems could be formulated from the above discussion. First, some of the proposed factors, such as labor force, materials, machinery, paving thickness, hydrological conditions, and geological types, are difficult to accurately determine at the conceptual design stage of a project. In addition, a considerable amount of investment for determining these factors is considered impractical when the project implementation is uncertain. Second, most of the existing studies focused on factors related to project construction conditions and project scale, and the project-scale-related factors were often limited to mileage, capacity, number of lanes, etc. Existing studies ignore the two major categories of expressway bridge and tunnel scale, which have an important impact on the expressway cost. However, expressway construction projects are usually large-scale, and the terrain of the crossing area is complex and variable. Terrain and geological changes also have an impact on such bridge and tunnel projects.

Therefore, according to the systematization of the impact factors of the expressway project cost proposed by existing research, the data of China's expressway were used as a case study in this work, with the following influencing factors: content, overall scale, bridge scale, and tunnel size of the expressway project as well as economic development indicators for the starting year of the project. Among them, the content, overall scale, and economic indicators summarize the existing achievements, and these factors could easily be obtained during the early stages of expressway construction projects. The indicators of bridge scale and tunnel size are newly introduced among the 10 factors related to the bridge and tunnel and can be obtained at the conceptual design stage. In the second category (overall scale of the project), the authors considered the date of the project construction initiation as a factor, which enables the model to study the effect of the economic development level on expressway costs. In the case of the economic indicators, the current price was used for GDP and the expressway forecast value, so as to eliminate the influence of inflation factors and ensure the consistency of economic data forecasts.

Fig. 1 shows the logical framework for collating the factors affecting the construction cost of an expressway. The cost of the project is summarized in Table 1. The reasons for the selection of the 25 factors are listed in Table 2. These factors serve as the basis for the subsequent model construction.

IV. CNN MODEL FOR EXPRESSWAY PROJECT-COST ESTIMATION
A. INTRODUCTION TO THE CNN MODEL
The CNN is a deep learning model proposed by LeCun *et al.* [46]. Essentially, it is a multilayer perceptron, the main advantages of which are local connections and weight sharing. The CNN model not only reduces the number of weights but also makes the network easy to optimize. This helps to reduce model complexity and the risk of overfitting. The model generally includes a convolution layer, pooling layer, and fully connected layer, each of which is composed of many neurons. The shared weights are used to calculate the

TABLE 1. Factors affecting expressway costs.

Feature Number	Factors	Description	Data type
1	Province	Place code of project	Integer
2	Mileage	Recommended length of project	Integer
3	Administrative level	(1) National road, (2) provincial road, (3) special expressway	Integer
4	Speed	Design speed	Integer
5	Traffic capacity	Design of two-way lanes	Integer
6	Nature of construction	(1) New construction, (2) reconstruction	Integer
7	Duration in months	Construction to completion time	Integer
8	Starting date of the project	Month and year	Integer
9	Bridge and tunnel ratio	Ratio of bridge and tunnel length to the total length	Integer
10	Sources of funds	(1) BT, (2) BOT, (3) others	Integer
11	Length of small bridge	Meters	Integer
12	Length of medium bridge	Meters	Integer
13	Length of large bridge	Meters	Integer
14	Length of super bridge	Meters	Integer
15	Number of culverts	Number	Integer
16	Length of short tunnel	Meters	Integer
17	Length of medium tunnel	Meters	Integer
18	Length of long tunnel	Meters	Integer
19	Length of super-long tunnel	Meters	Integer
20	Number of tunnels	Number	Integer
21	Investment in fixed assets	Indicator of construction year	Integer
22	CPI	Indicator of construction year	Integer
23	PPI	Indicator of construction year	Integer
24	Benchmark lending rates	Indicator of construction year	Integer
25	GDP	Indicator of construction year	Integer

neurons, and the activation function is used as the input of the next layer for the nonlinear calculation output. The weights are obtained by training the dataset. The convolution layer extracts features, enhances the features of the original signal through the convolution operation, and reduces noise. The pooling layer compresses feature maps, extracts main features, and reduces dimensionality and overfitting [47], [48]. Considering these unique advantages, the CNN was used to establish the proposed expressway cost-prediction model with the following adjustments. Because the data features used in this study do not have high dimensions, the pooling layer was removed to avoid the deletion of necessary information. After feature extraction in the convolution layer,

the fully connected layer is used at the end of the network to create the final nonlinear combination of features and predict the cost.

1) CONVOLUTION KERNEL SIZE
The convolution layer uses a convolution kernel to share the parameters of the same layer, thus greatly reducing the number of model parameters. The receptive field of the convolution kernel is used to extract the input features of the model. In this study, the authors selected the convolution kernel according to four aspects, the first being the improvement in the depth of model learning. Although a larger convolution kernel implies a larger receptive field, features must

TABLE 2. Analysis of relationship between influencing factors and cost of expressway construction projects.

No.	Factor	Relationship between influencing factors and cost of expressway construction projects.
1	Province	Provincial factors have an impact on expressway construction cost according to the topography, geology, and economic development levels of the marked area. For example: (a) China features mountainous regions in the West and plateaus in the East, and most provinces are divided by mountains, rivers, and other naturally occurring landforms. However, engineering geology has a significant influence on the cost of an expressway project, and detailed information is difficult to obtain at the conceptual design stage. (b) China's economic development is imbalanced; the level of economic development varies from province to province, and expressway costs are closely related to the economic development level of the project location.
2	Mileage	The longer the mileage of an expressway project, the greater is the labor, material, and machinery required and the higher is the cost.
3	Administrative level	National-level expressways may aim to promote the development of a regional economy. Provincial-level expressways are used to promote the economic development of the province in which they are located; special-purpose expressways generally serve specific enterprises. The different levels determine the various factors to be considered in selecting a corridor belt for expressway construction, and this corridor-belt selection has a direct impact on expressway construction costs.
4	Speed	Different design speeds determine the selection of the design index, which significantly affects the expressway project cost.
5	Traffic capacity	The number of designed lanes directly determines the scale of an expressway project and has a significant impact on the cost.
6	Nature of construction	The cost of new projects is affected by the different quantities used in new reconstruction projects.
7	Duration in months	The longer the project construction time, the greater is the capital, labor force, and mechanical operation required; this significantly impacts the project cost.
8	Starting date of the project	The economic development level differs on a yearly basis, significantly influencing the project cost.
9	Bridge and tunnel ratio	The unit cost of bridge and tunnel structures is much higher than that of a common roadbed. The project scale is the main cost of an expressway project. The larger the ratio of bridges and tunnels to roadway, the larger is the cost.
10	Sources of funds	Different sources of funds represent different financing models, which directly affects the cost of funds for highway projects.
11	Length of small bridge	
12	Length of medium bridge	Different bridge lengths determine the amount of concrete and steel used, which directly affects the main costs of an expressway project. The longer the bridge, the higher is the cost.
13	Length of large bridge	
14	Length of super bridge	
15	Number of culverts	The quantity of concrete and steel is determined by the number of culverts, which directly affects the main cost of an expressway project.
16	Length of short tunnel	
17	Length of medium tunnel	Different tunnel lengths determine the amount of concrete and steel, and directly affect the main cost of expressway projects. The longer the tunnel, the higher is the cost.
18	Length of long tunnel	
19	Length of super-long tunnel	
20	Number of tunnels	The greater the number of tunnels, the greater is the cost. The greater the number of tunnels, the greater is the amount of portal construction, which directly affects the main cost of an expressway project.

21	Investment in fixed assets	The increase of investment in fixed assets indicates that the demand for labor, materials, and machinery may exceed the supply in the construction market, and this would lead to an increase in the unit price of labor, materials, and machinery, which, in turn, affect the cost of expressway projects.
22	CPI	The consumer price index (CPI) is used to reflect the change of the price level of consumer goods and services purchased by households, and reflects the labor cost over a certain period, which, in turn, affects the cost of expressway projects.
23	PPI	The industrial producer price index (PPI) reflects the changing trends and range of the ex-factory and purchase prices of all industrial producers, reflecting the cost of raw materials over a certain period, which, in turn, affects the cost of expressway projects.
24	Benchmark lending rates	The loan interest rate affects the capital cost of an expressway project. The construction period of an expressway project is approximately 3–5 years, and therefore, the 3–5-year benchmark loan interest rate is selected.
25	GDP	GDP represents the level of economic development in a certain period and affects the cost of expressway projects in a certain period.

be extracted in image learning. A 5 × 5 two-dimensional matrix was learned in this study, and the image dimension was considerably small. On using a larger convolution kernel, main feature information can be easily ignored and the depth of model learning can be reduced. Therefore, in this study, a small convolution kernel was selected for deep learning. The second aspect is to reduce computational complexity. Large convolution kernels require additional training parameters; this results in a sharp increase in the amount of computation and decreases computational performance. Third, the size of a convolution kernel must be greater than 1 to enhance the receptive field, and therefore, a 1 × 1 convolution kernel was not considered in this study. Fourth, 2 × 2, 3 × 3, and 4 × 4 convolution kernels were used for verification. After comparison of the operation of the parameters and the operation results, the input matrix used in this study was 5 × 5, and therefore, the use of a larger convolution kernel was not suitable. If a 2 × 2 convolution kernel is used, the number of parameters is reduced, computational performance could be improved, and learning depth of the model could be enhanced (see the case analysis of Section V for a detailed discussion). A convolution kernel is realized using a convolution operator and is formulated as

$$y = W * X + b, \tag{1}$$

where y is the output value of the convolution layer, X is the input feature information, W is the weight of the convolution layer, b is the bias of the convolution layer, and $*$ is the convolution operation.

2) ACTIVATION FUNCTION SIZE
CNN usually uses the activation function of a nonlinear mapping as an excitation function at the output of the convolution layer, and it is modelled as

$$h(y) = \begin{cases} y, & y > 0 \\ 0, & y \le 0. \end{cases} \tag{2}$$

3) OPTIMIZER
The learning process of CNN involves the constant update of the weights and biases of an NN to minimize the value of the loss function. The smaller the loss function, the stronger is the learning ability of the NN corresponding to the parameters and the better is the prediction performance of the NN. In this study, the mean square error was used as the loss function, and it is formulated as follows:

$$MSE = \frac{1}{n} \sum (y_i - \widehat{y}_i)^2, \tag{3}$$

where y_i and $\widehat{y}_i$ are the actual and predicted values, respectively.

An optimizer is required to search for the optimal parameters. In this study, the optimizers of the CNN were compared (gradient descent, momentum, and Adam optimizers), and the Adam optimizer was selected. A detailed verification of the case analysis is presented in section V.

B. DEVELOPMENT OF EXPRESSWAY COST PREDICTION MODEL
The 25 selected factors were used to establish the data structure, which was reconstructed into a matrix to establish the CNN model.

1) DATA DEFINITION
The data structure for the expressway cost-prediction model is defined as follows, along with the classification of the influencing factors:

$$X_i = (U_1, U_2, U_3, U_4, U_5),$$
$$U_{i,j} = (V_{i1}, V_{i2}, V_{i3}, V_{i4}, V_{i5})',$$

where U is a category, and $V_{i,j}$ is the factor of a category. The 25 factors were placed into five categories depending on how they affect the expressway project cost. Each category contains five detailed indicators, as listed sequentially in Table 3.

TABLE 3. Categorization of influencing factors for expressway project cost estimation.

Category	Description	Code of factors
U_1	Content of the expressway project	1–5
U_2	Overall scale of the expressway project	6–10
U_3	Bridge scale of the project	11–15
U_4	Tunnel size of the project	16–20
U_4	Economic development indicators for the starting year of the project	21–25

2) RESTRUCTURING THE MATRIX

Fig. 2 shows the reconstruction of the input matrix for the CNN based on the selected influencing factors.

$$\begin{bmatrix} v_{1,1} & v_{1,2} & v_{1,3} & v_{1,4} & v_{1,5} \\ v_{2,1} & v_{2,2} & v_{2,3} & v_{2,4} & v_{2,5} \\ v_{3,1} & v_{3,2} & v_{3,3} & v_{3,4} & v_{3,5} \\ v_{4,1} & v_{4,2} & v_{4,3} & v_{4,4} & v_{4,5} \\ v_{5,1} & v_{5,2} & v_{5,3} & v_{5,4} & v_{5,5} \end{bmatrix}$$

FIGURE 2. Highway feature reconstruction matrix.

3) CNN MODEL FOR EXPRESSWAY COST PREDICTION

For the proposed CNN model, the input is a 5×5 matrix, and the feature map has dimensions of $5 \times 5 \times 1$ (Fig. 3). As this study involved relatively small data dimensions, the model structure was simplified by removing the pooling layer to avoid rounding out important feature information. The upper layer of the feature information comprises 2×2 convolution kernels. Each of the C1 convolution layer, C2 convolution layer, and FC4 fully connected layer uses 64 different convolution kernels. Finally, the model outputs the predicted expressway cost.

C. MODEL EVALUATION

The mean absolute percentage error (MAPE) represents the error degree of the model. The smaller the MAPE value, the smaller is the error and the higher is the accuracy. This value, as formulated in (4), represents the amount by which the predicted value deviates from the true value.

$$MAPE = \frac{|p - A|}{A} \times 100\%, \tag{4}$$

where p is the predicted value, and A is the actual value.

V. CASE ANALYSIS

To verify the validity of the proposed model, the CNN model was used to learn case data via Python, and the same case data were simulated using the ANN model adopted by Elbeltagi *et al.* [20] via MATLAB. The RA model was used to train historical data in the SPSS environment. A laptop running 64-bit Windows 10 with 8 GB of RAM and an Intel i7-8550u CPU was used for programming.

A. DATA PREPARATION AND ANALYSIS

The case analysis was performed using information on 415 expressway projects collected through the credit information management system of the China expressway construction market (https://glxy.mot.gov.cn/). Data on expressway projects from 2005–2018 were collected, covering 31 provinces and direct municipalities. The relationship between each feature and the predicted value is shown in Fig. 4, depicting the relationship between the expressway project cost and each feature (described in Table 1) arranged from left to right and from top to bottom. The figure shows that most of the features have no obvious linear relationship with the predicted values, and the wide coverage of the case data allows for the easy development of a discrete distribution.

B. DATA NORMALIZATION

To make the data comparable, each feature value, depicted as in (5), was normalized using (6).

$$X_i = (X_{i1}, X_{i2}, X_{i3}, X_{i4}, X_{i5} \ldots \ldots \ldots X_{im})', \tag{5}$$

where X_i represents the sample value of feature i, and m represents the number of samples. After normalization, X' is obtained as follows:

$$X' = \frac{|X - X_{min}|}{|X_{max} - X_{min}|} \times 100, \tag{6}$$

where X_{min} and X_{max} are the minimum and maximum values of a feature, respectively.

C. CASE STUDY 1: COMPARISON BETWEEN RA, ANN, CNN MODELS

To illustrate the superiority of the CNN model over the conventional RA and ANN models, the same proportion of training and test datasets were used in random sampling for the same case data; the RA, ANN, and CNN models were used to forecast the cost of highway projects. Moreover, three simulation studies were performed.

1) PREDICTION USING THE RA MODEL

First, the most commonly used RA model, i.e., the multiple linear RA (MLRA), was evaluated. MLRA refers to the linear regression model with multiple explanatory variables; this is used to reveal the linear relationship between the explained variables and other multiple explanatory variables. The mathematical model of MLRA is given as

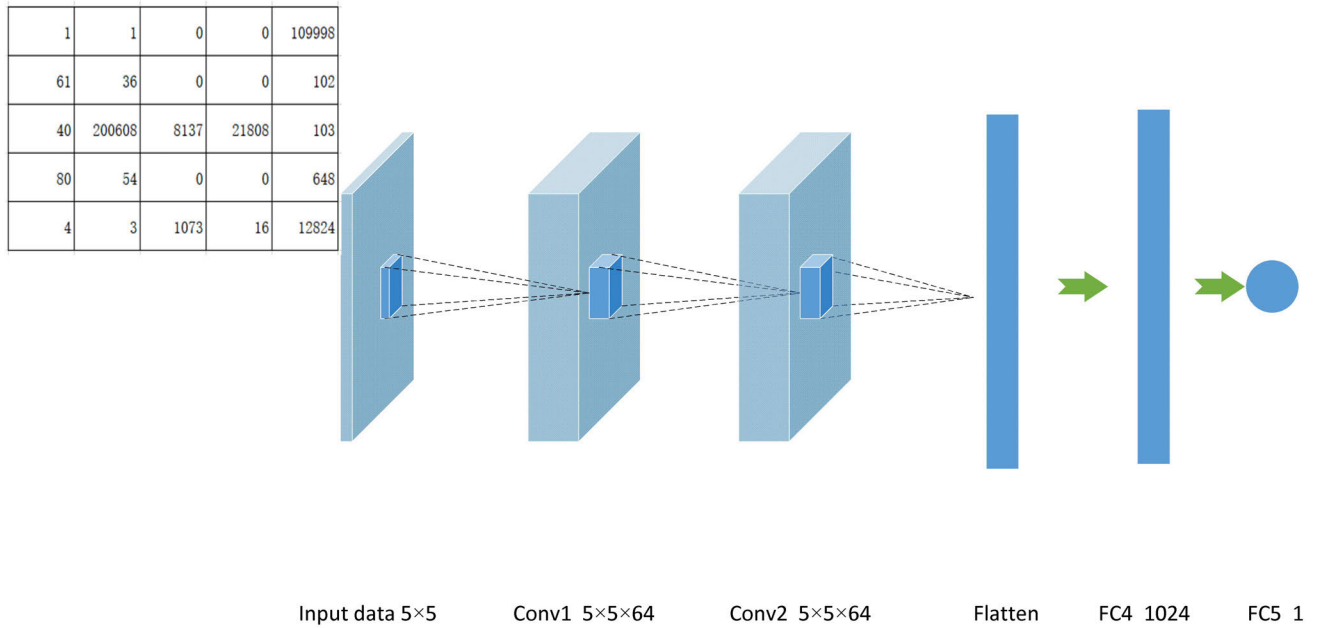$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_p x_p + e, \tag{7}$$

1	1	0	0	109998
61	36	0	0	102
40	200608	8137	21808	103
80	54	0	0	648
4	3	1073	16	12824

Input data 5×5   Conv1 5×5×64   Conv2 5×5×64   Flatten   FC4 1024   FC5 1

FIGURE 3. CNN model for highway cost prediction.

where p is the number of explanatory variables, y is the explained variable, x_i is the explanatory variable, and e is a random term.

To extract the main features as explanatory variables of the regression model, the Pearson product–moment correlation coefficient was used to calculate the correlation among all variables. The explanatory variables of the regression model were selected according to the correlation strength. The Pearson product–moment correlation coefficient is calculated as follows:

$$p_{xy} = \frac{Cov(X, Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}}. \quad (8)$$

The Pearson product–moment correlation coefficient is shown in Fig. 5, where the variables could not be presented because of space restrictions. The variables from left to right and from top to bottom are as follows: province; mileage; administrative level; speed; traffic capacity; nature of construction; duration in months; starting date of the project; bridge and tunnel ratio; sources of funds; lengths of the small, medium, large, and super bridges; number of culverts; lengths of the short, medium, long, and super-long tunnels; number of tunnels; investment in fixed assets; CPI; PPI; benchmark lending rates; and GDP.

Generally, the strong and weak relations among the variables represented by the Pearson product–moment correlation coefficient are divided as shown in Table 4.

Because this dataset is relatively discrete, six variables (mileage, length of the super bridge, lengths of the short, medium, and long tunnels, and number of tunnels) with

TABLE 4. Intensity division of pearson product–moment correlation coefficient.

p_{xy}	0.0–0.2	0.2–0.4	0.4–0.6	0.6–0.8	0.8–1.0
Strength of linear relationship	Extremely weak or irrelevant	Weak correlation	Moderate correlation	Strong correlation	Very strong correlation

TABLE 5. Model summaryb.

Model	R	R²	Adjust R²	Error of standard estimation	R² Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
					Change statistics					
1	.804ᵃ	.646	.643	.0799361	.646	224.763	3	370	.000	2.030

a. (constant), Mileage, Length of super bridge, Number of tunnels
b. Dependent variable: cost.

correlation coefficients ≥0.4 were utilized as explanatory variables. After multiple regression using the SPSS software, three variables were determined to have insignificant correlation and were removed from the regression equation. The remaining three variables (mileage, length of the super bridge, and number of tunnels) were used as explanatory variables for the multiple regression. The regression equation and coefficient correlation were significant, and the correlation parameters are listed in Tables V–VII.
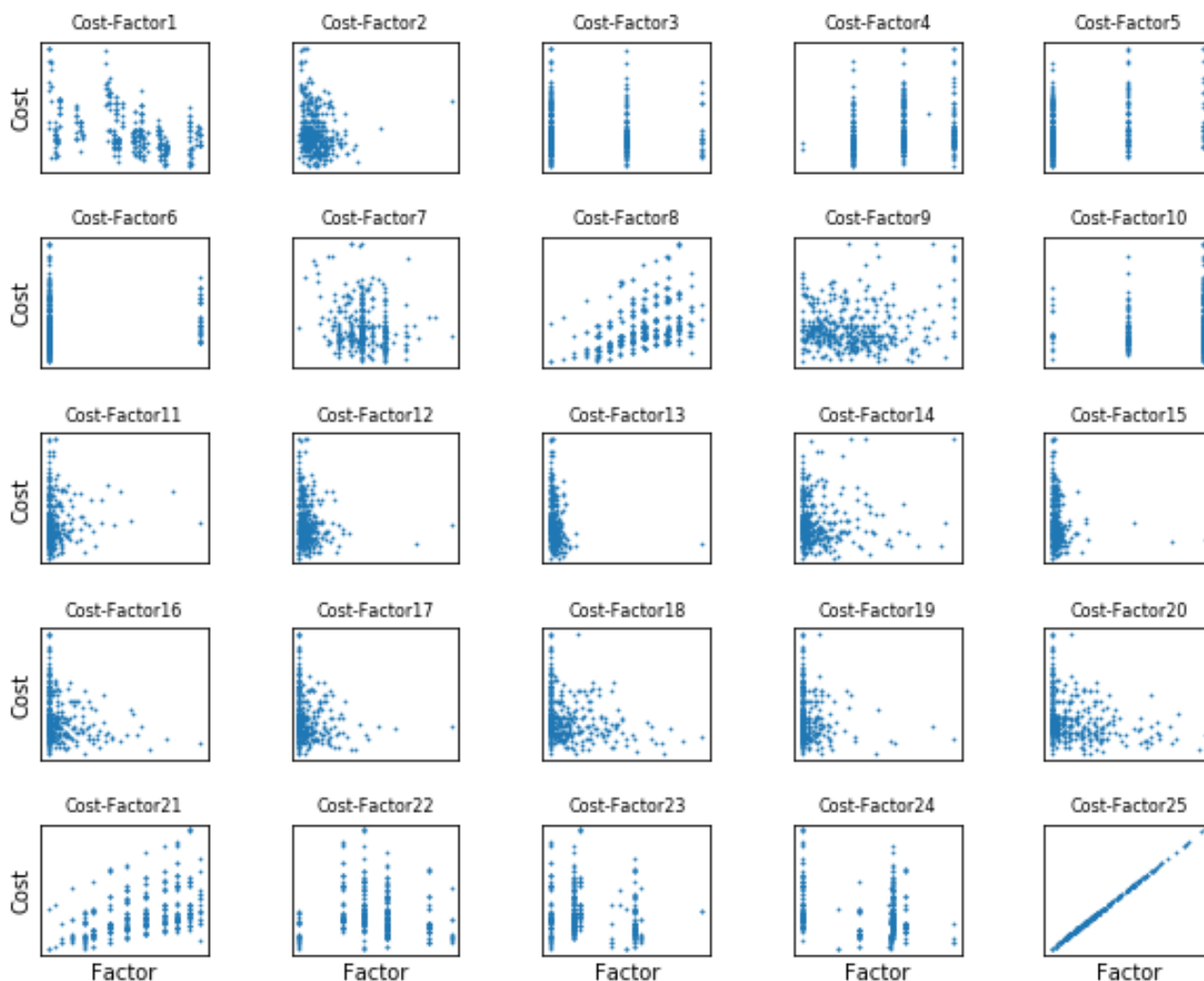
FIGURE 4. Relationship between data features and predicted values.

TABLE 6. Anovaa.

	Model	Sum of squares	df	Mean square	F	Sig.
	Regression	4.309	3	224.763		.000[b]
1	Residual	2.364	370	.006		
	Total	6.673	373			

a. (constant), Mileage, Length of super bridge, Number of tunnels
b. Dependent variable: cost.

The regression model was obtained from the data in Table 7.

$$y = 0.042 + 0.872x_1 + 0.197x_2 + 0.247x_p \qquad (9)$$

The regression model was used to predict the test data, shown in Fig. 6(a); the MAPE was 29.26%, and the evaluation parameters of the RA model are listed in Table 8 (RA).

TABLE 7. Coefficients of correlation.

Model		Coefficient of nonstandardization		Standard coefficient	t	Sig.	Collinear statistics	
		B	Standard error	Trial version			Tolerance	VIF
	(constant)	.042	.007		6.177	.000		
	Mileage	.872	.050	.557	17.176	.000	.909	1.100
1	Length of super bridge	.197	.028	.222	6.979	.046	.947	1.056
	Number of tunnels	.247	.024	.341	10.269	.000	.867	1.154

2) PREDICTION USING THE ANN MODEL

Elbeltagi *et al.* [20] developed an ANN-based prediction model, in which all the transfer functions of the hidden and output layers used Tansig, while Traimlm was used as the training function; the model was trained using MATLAB.
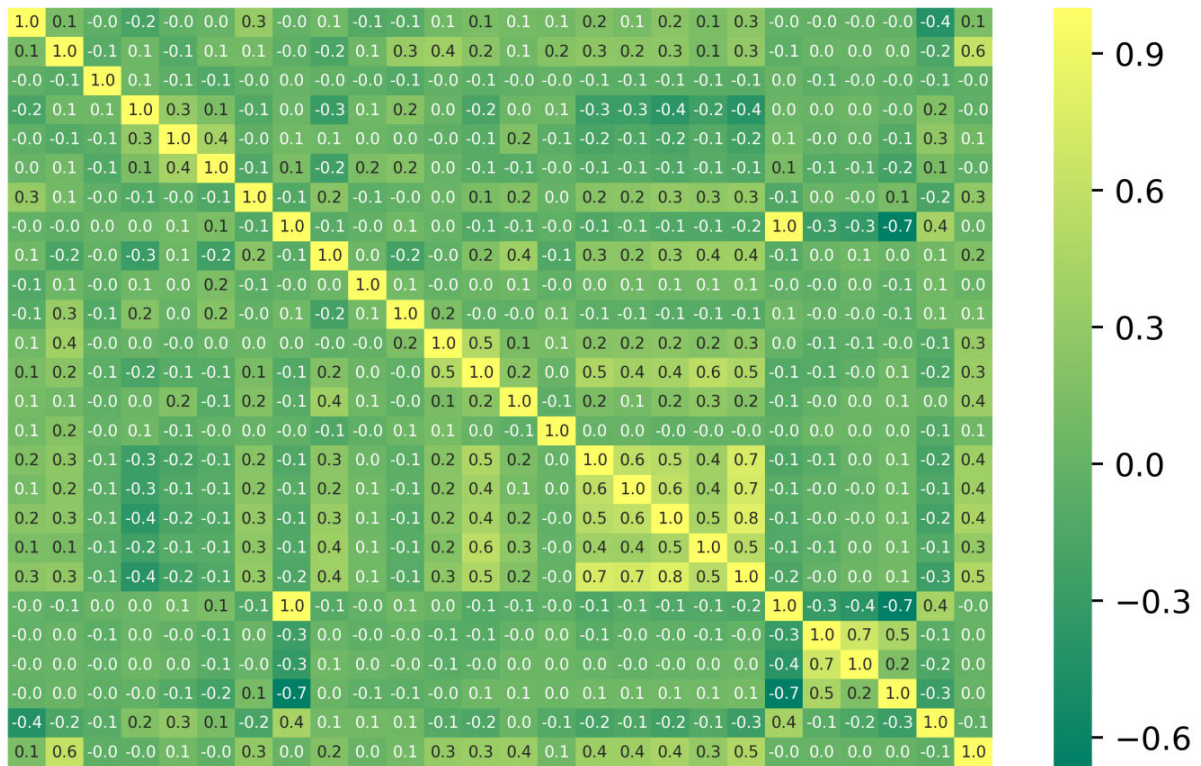
FIGURE 5. Pearson product–moment correlation coefficient of variables.

TABLE 8. Comparison of operating parameters.

Model	Number of training sets	Number of test sets	Learning rate	Number of training parameters	Iterations	MAPE	Runtime
CNN (with bridge and tunnel factors)	374	42	0.001	704	620	17%	21 s
ANN (with bridge and tunnel factors)	374	42	0.01	829	11	22.84%	1 s
CNN (without bridge and tunnel factors)	374	42	0.001	704	450	23.78%	12 s
ANN (without bridge and tunnel factors)	374	42	0.01	339	16	40.62%	1 s
RA	374	42	-	-	-	29.26%	-

Their results showed that the model had good applicability to the cost prediction of 67 highway projects in Libya. In the current study, we used the collected data to reproduce the ANN model in [20], with Tansig as the activation function and the Levenberg Marquardt algorithm as the optimizer. From the input data, 90% of the 415 case data were randomly extracted as the learning set, and the remaining 10% was used as the test set. There is no criterion for feature selection and the selection of the number of hidden-layer neurons in ANN-based models. However, the selection of these parameters has a significant influence on the prediction accuracy. To overcome these defects and ensure that the ANN model [20] achieves the best prediction result, two improvements were made in this study. First, the rationality of the ANN model in feature selection was ensured by calculating the distance correlation coefficients between the 25 features and the predicted values (as shown in Table 9). According to the correlation of the distance coefficient, 6–25 features were selected as input neurons in descending order. Second, as there was no clear optimal number of neurons in each NN layer [20], the number of
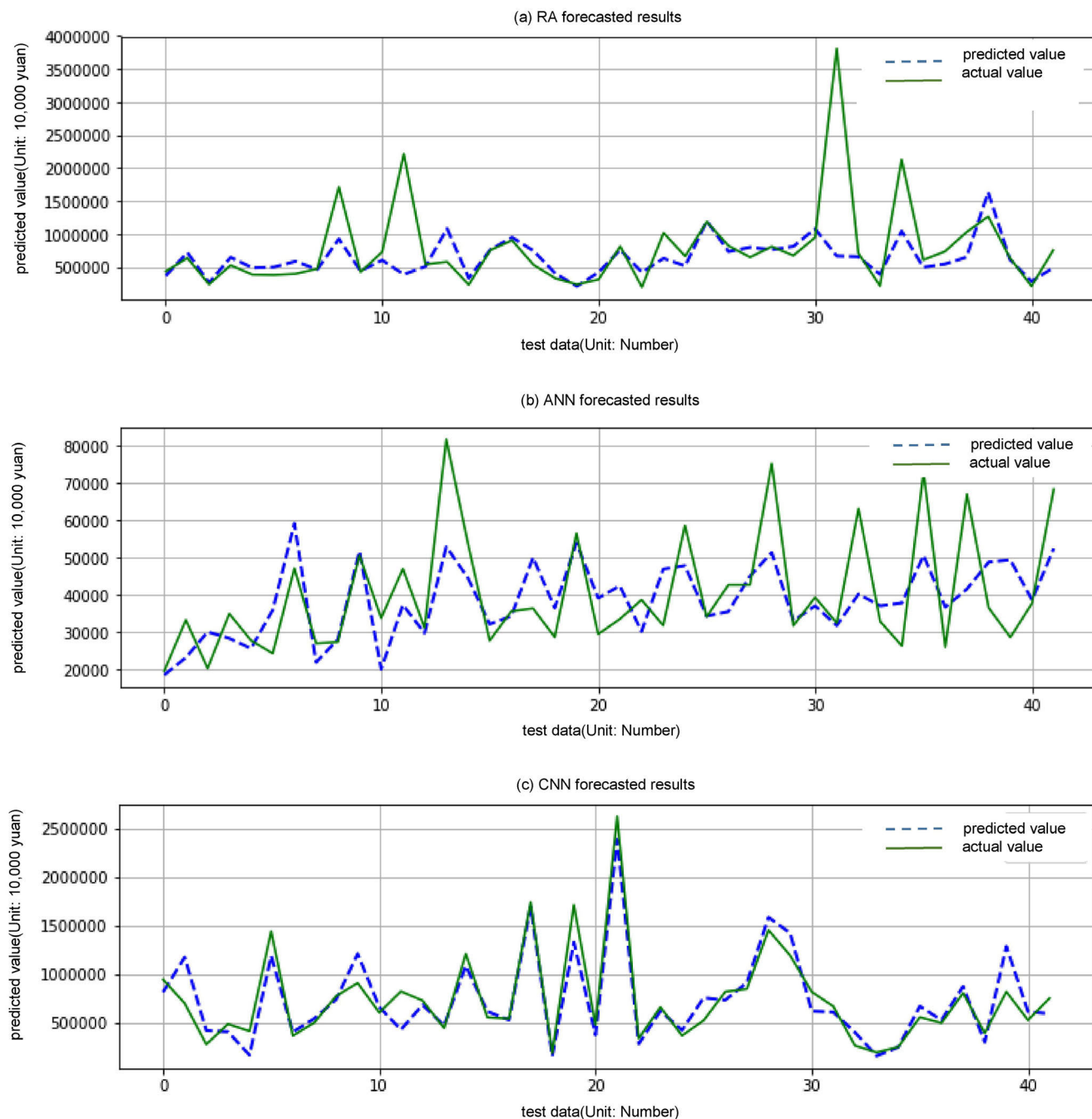
FIGURE 6. Forecasted results based on the AN (a), ANN (b), and CNN (c) models with bridge and tunnel factors.

neurons in the first hidden-layer is used to select the average input of the first hidden layer and the output of the first hidden neurons [49]. In addition, the number of neurons in the second hidden-layer is used to select a sequence of numbers between 6 and 50 with equal intervals of 2; for example, 6, 8, 10, 12, 14, 16, 18, 20, . . . , 50. Lastly, the user executed each ANN several times, and the ANN structure with the smallest error rate was selected as the MAPE value of each structure (as shown in Fig. 7). Fig. 7 shows the various network structures

corresponding to the first 6 and 25 input neurons of the distance correlation coefficients, arranged from left to right and from top to bottom, respectively. The ANN structure is labeled above each subgraph ("input neuron"–"first hidden layer neuron"–"second hidden layer neuron"–"output neuron"). For example, the first row, first column, first structure for 6–4–x–1 represents the first six distance correlation coefficients as input neurons; first hidden-layer neuron (4); second hidden-layer neuron (x), which is a sequence with
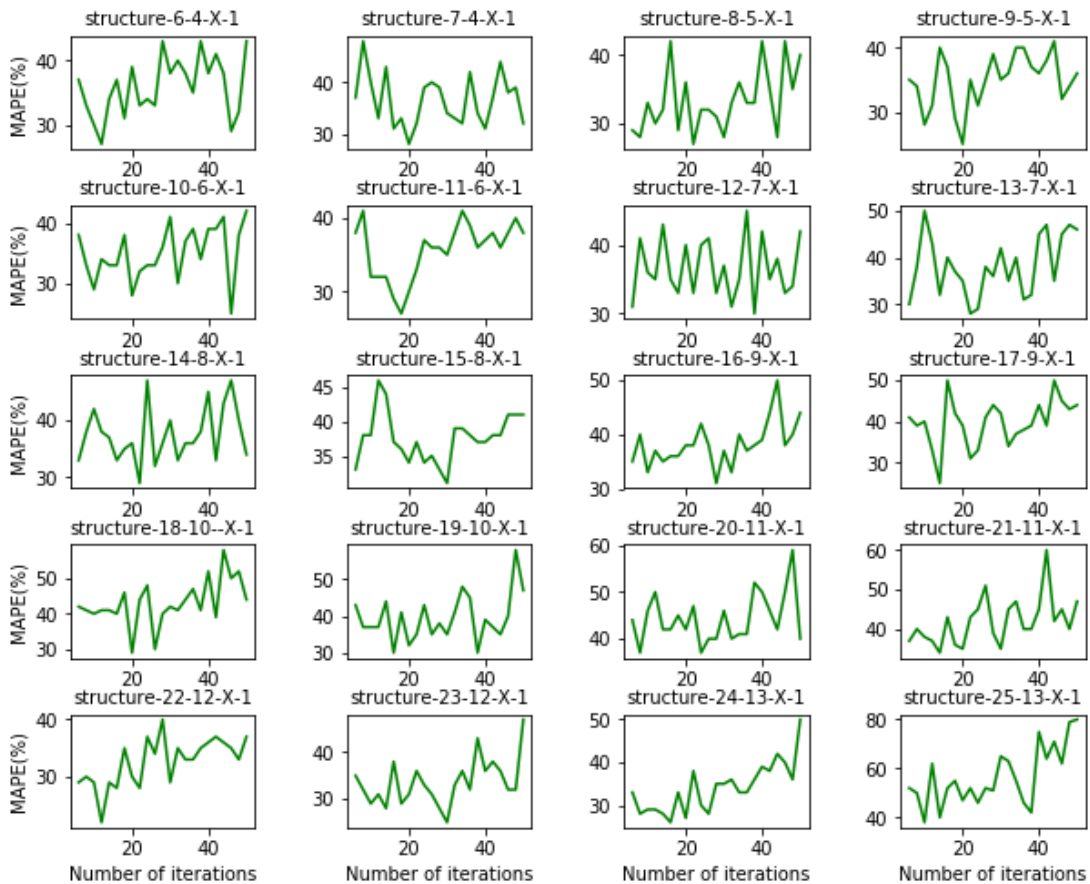
FIGURE 7. ANN structure and results of MAPE.

equal intervals of 2 between 6 and 50; and the output-layer neuron (1). The longitudinal axis represents the MAPE value corresponding to each ANN structure, and the horizontal axis represents the number of neurons in the second hidden layer, X.

The ANN-predicted MAPE for input neuron 22, with the first hidden layer as 12 and the second hidden layer 12, is 22.84%, which is optimal. The characteristics of the model input information are listed in Table 10, and the operation result parameters of ANN (with bridge and tunnel factors) are listed in Table 8. In addition, the prediction results are shown in Fig. 6(b).

In this study, the input information of an NN was divided into five categories. Among the 25 factors, the optimal input information of the 22–12–12–1 ANN model in Fig. 7 is described as follows. The *content of the expressway project* contains five factors, all of which are used as the input information of the ANN model. The *overall scale* of the expressway project contains five factors, of which *nature of construction* and *sources of funds* were deleted in this study. The *bridge scale of the project* contains five factors, all of which are used as input information for the ANN model. The *tunnel size of the project* contains five factors, all of which

are used as input information for the ANN model. In the fifth category, *economic development indicators for the starting year of the project, benchmark lending rates* was deleted. From the perspective of the impact of various factors on the project cost in practice, the three deleted factors have a certain impact on the highway project cost. That is, the *nature of construction* determines the project scale, while the *sources of funds* and *benchmark lending rates* determine the use cost of project funds. This may be because the values of these three factors are specific, and the project cost corresponding to these values varies greatly; an ANN model is not sensitive to such situations.

3) PREDICTION USING THE CNN MODEL
a: LEARNING RATE
For implementing the proposed CNN model, 90% of the case data were randomly selected as the training set, and 10% were selected for the test set. Fig. 8 shows that the model was sensitive to changes in the training and test datasets at a learning rate of 0.001, and the difference between errors of the training and test datasets converged after 620 iterations. Therefore, the learning rate of the model was set to 0.001.

TABLE 9. Correlation coefficients of expressway characteristics and cost distance.

Order	Factor	Feature Number	Distance correlation coefficient
1	Mileage	2	0.69
2	Length of large bridge	13	0.64
3	Number of tunnels	20	0.53
4	Length of long tunnel	18	0.48
5	Length of medium tunnel	17	0.46
6	Length of medium bridge	12	0.43
7	Length of short tunnel	16	0.43
8	Length of super bridge	15	0.39
9	Length of super-long tunnel	19	0.35
10	Duration in months	7	0.29
11	Number of culverts	14	0.27
12	Bridge and tunnel ratio	9	0.25
13	Province	1	0.20
14	GDP	25	0.13
15	Length of small bridge	11	0.11
16	Starting date of the project	8	0.10
17	Investment in fixed assets	21	0.10
18	Administrative level	3	0.10
19	Traffic capacity	5	0.08
20	Speed	4	0.08
21	CPI	22	0.08
22	PPI	23	0.07
23	Benchmark lending rates	24	0.06
24	Nature of construction	6	0.06
25	Sources of funds	10	0.06
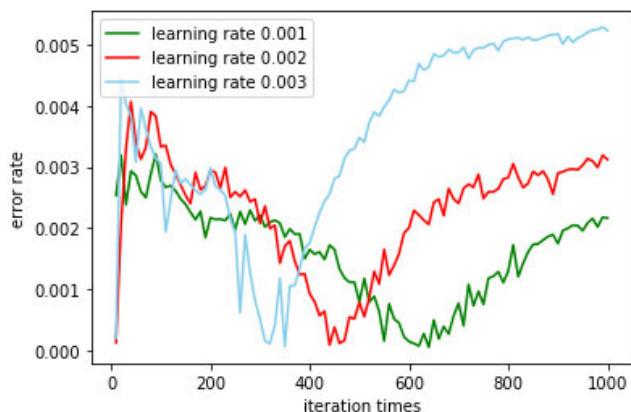
FIGURE 8. Prediction results based on the CNN and ANN models with respect to the bridge and tunnel factor.

b: CONVOLUTION KERNEL

The experimental models with 2×2, 3×3, and 4×4 convolution kernels were used to test the dataset. The test model with a 2×2 convolution kernel converges 620 times, with a run time of 21 s and MAPE of 17% [Fig. 6(c)]. The test model with the 3×3 convolution kernel converges 420 times, with a run time of 25.21 s and MAPE of 24.38% [Fig. 10(a)]. The test model with the 4×4 convolution kernel converges 330 times, with a run time of 29.92 s and MAPE of 26.89% [Fig. 10(b)]. The operation parameters of the three convolution kernels are shown in Table 11. The experimental results show that the 2×2 convolution kernel can reduce the number of parameters, improve the computational performance, enhance the depth of model learning, and improve the prediction accuracy.

c: OPTIMIZER

Three optimizers, i.e., the gradient descent, momentum, and Adam optimizers, were used to run the model. The loss function value of the test data is shown in Fig. 9. As the value of the loss function obtained using the Adam optimizer was the smallest, the Adam optimizer was used in this study.

TABLE 10. Input feature information of 22−12−12−1 ANN model.

Feature number	Category	Description	Factor	Distance correlation coefficient
1			Province	0.2
2		Content of the expressway project	Mileage	0.69
3	U_1		Administrative level	0.1
4			Speed	0.08
5			Traffic capacity	0.08
7		Overall scale of the expressway project	Duration in months	0.29
8	U_2		Starting date of the project	0.1
9			Bridge and tunnel ratio	0.25
11			Length of small bridge	0.11
12			Length of medium bridg	0.43
13	U_3	Bridge scale of the project	Length of large bridge	0.64
15			Number of culverts	0.27
14			Length of super bridge	0.39
16	U_4	Tunnel size of the project	Length of short tunnel	0.43
17			Length of medium tunnel	0.46
18			Length of long tunnel	0.48
19			Length of super-long tunnel	0.35
20			Number of tunnels	0.53
21		Economic development indicators for the starting year of the project	Investment in fixed assets	0.1
22			CPI	0.08
23			PPI	0.07
25			GDP	0.13
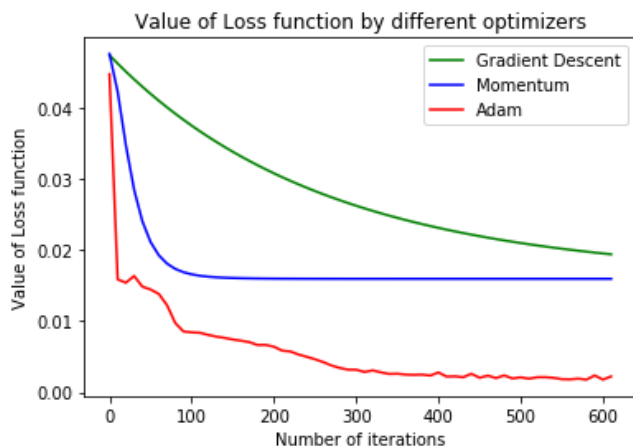
FIGURE 9. Loss-function values of test data obtained using different optimizers.

Therefore, the CNN model was selected with a learning rate of 0.001, 620 iterations, and the Adam optimizer. The results of the CNN (with bridge and tunnel factors) are shown in Fig. 6(c), and the parameters are listed in Table 8.

As Fig. 6 and Table 8 shown, the prediction error (17%) of the CNN-based model presented in this study is lower than those of the ANN (22.84%) model and the RA model (29.26%). In addition, the 20% error requirement at the conceptual design stage [20], [50] was satisfied, which demonstrated the advantages of the input features and the CNN model proposed herein. Second, under the same input conditions, the CNN model has more iterations than the ANN model, showing its superiority over the ANN model in terms of learning ability. Moreover, the CNN model reduces the overfitting problem in machine learning owing to its advantage of high-dimensional learning.

D. CASE STUDY 2: COMPARISON BETWEEN THE USE OF 25 AND 15 FACTORS

To verify the effectiveness of the 10 proposed factors related to the bridge and tunnel for improving the accuracy of model prediction, the CNN model was applied without the bridge and tunnel factors for a comparison with the simulations obtained in the first case study. To this end, 90% and 10% of the case data were randomly selected as the training and test sets, respectively. A result with an error rate of 23.78%

TABLE 11. Comparison of operating parameters of 2 × 2, 3 × 3, and 4 × 4 convolutional kernels.

Model	Size of convolution kernel	Number of training sets	Number of test sets	Learning rate	Number of training parameters	Iterations	MAPE	Runtime
CNN	2 × 2	374	42	0.001	704	620	17%	21 s
CNN	3 × 3	374	42	0.001	1344	420	24.38%	25.21 s
CNN	4 × 4	374	41	0.001	2240	330	26.89%	29.92 s
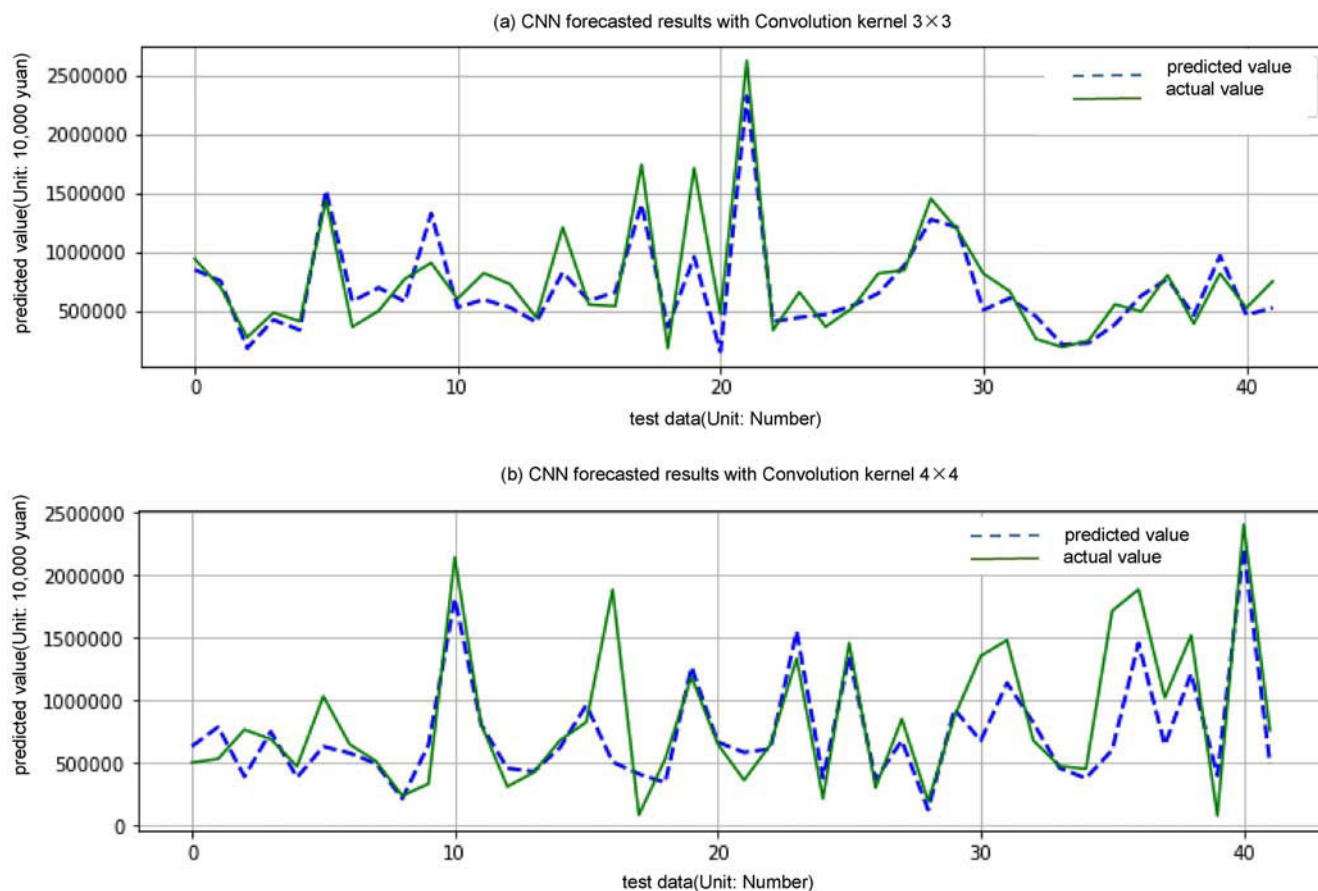
FIGURE 10. Prediction results of the CNN with (a) 3 × 3 and (b) 4 × 4 convolutional kernels, with bridge and tunnel factors.

was obtained after 450 iterations at a learning rate of 0.001. The parameters used in this CNN (without bridge and tunnel factors) are listed in Table 8, and the test results are shown in Fig. 11(a). Compared with the 17% error rate obtained using the 25-factor CNN model, the 15-factor CNN model achieves an error rate of 23.78%, which is relatively similar to that of the ANN model (22.84%). To verify the superiority of the CNN model under the same condition, the ANN model was executed again without the bridge and tunnel factors, and an error rate of 40.62% was obtained. The parameters of this ANN (without bridge and tunnel factors) are listed in Table 8, and the test results are depicted in Fig. 11(b). The results prove the superiority of the CNN model over the ANN model.

Moreover, the significant difference in the prediction results of the ANN model, caused by the removal of the bridge and tunnel factors, further highlights the effectiveness of the new factors in terms of improving the prediction accuracy of the CNN model.

Table 8 shows that, under the same input conditions, the CNN model completes more iterations than the ANN model, thus confirming the superior learning ability of the CNN model. Moreover, although the ANN model has the advantage of a small number of training parameters for learning low-dimensional nonlinear problems, its prediction accuracy is considerably less than that of the CNN model. Furthermore, the CNN model offers the advantage of a small
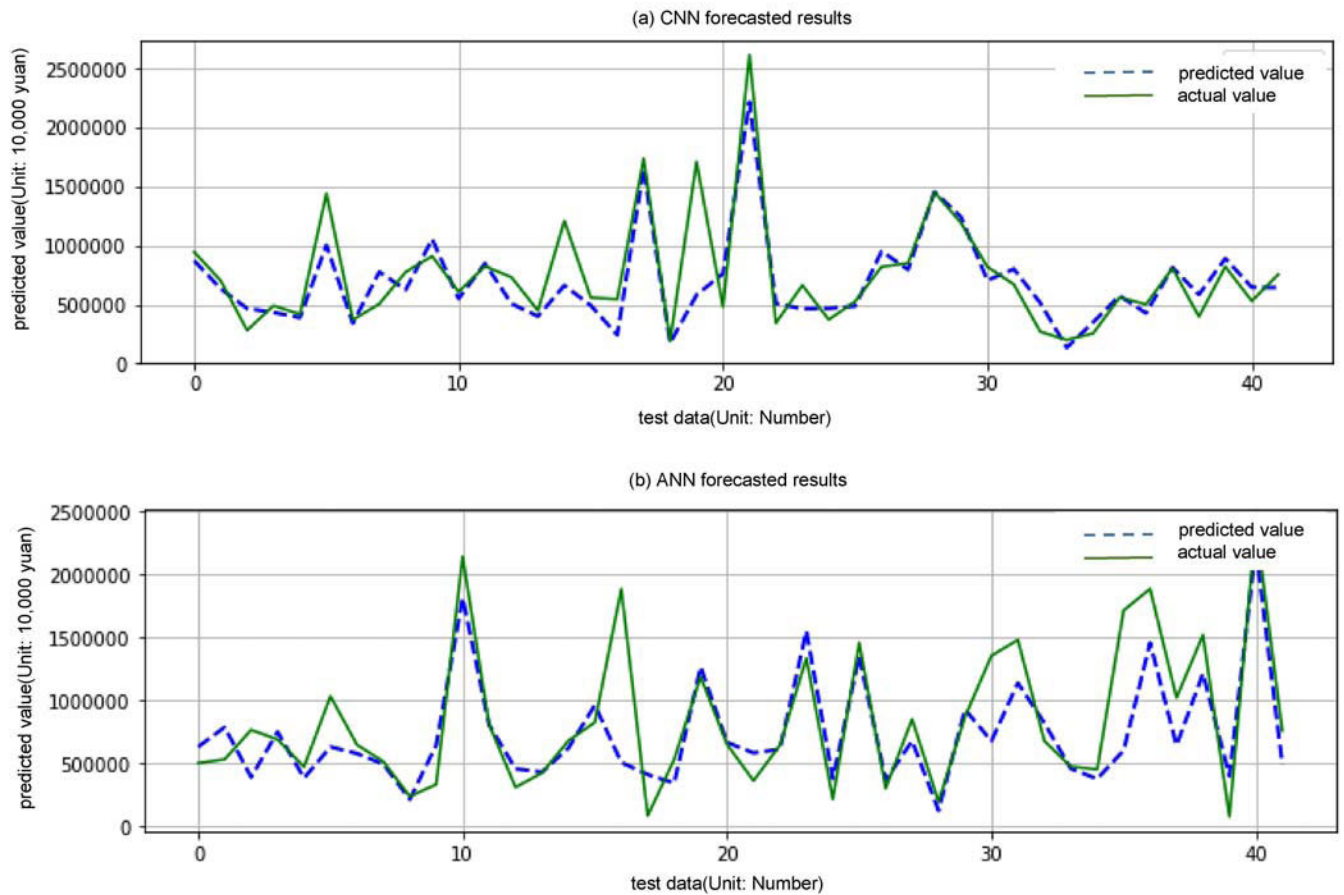
FIGURE 11. Prediction results based on the (a) CNN and (b) ANN models without the bridge and tunnel factors.

number of training parameters in high-dimensional nonlinear problems, and its accuracy is better than that of the ANN model, indicating that the CNN model is more suitable for high-dimensional nonlinear problems.

VI. CONCLUSION

In terms of project management, researchers have devoted increasing efforts toward the early prediction of highway project costs. At present, most of the cost-prediction methods are based on the ANN and RA models. However, these models suffer from a few disadvantages, such as over fitting, excessive parameters, and complex calculations.

Based on existing research results pertaining to factors influencing the cost of a highway project, 15 factors with high importance that are easy to obtain at the conceptual design stage were selected as the research object. Additionally, compared with previous studies, 10 innovative factors related to the bridge and tunnel were proposed; these have an important impact on the expressway project cost and can be obtained at the conceptual design stage. By reconstructing the characteristics of the influencing factors, the CNN model for expressway cost prediction was constructed. To the best of the authors' knowledge, this is the first study to demonstrate the application of a CNN model in expressway project cost prediction. Based on the data of 415 expressway projects collected in this study, the mainstream cost prediction models (ANN and RA) used in the field of construction engineering management were introduced for comparison. The case analysis results showed that the 10 new indexes related to the bridge and tunnel factors can improve the prediction accuracy of the model. A comparison of the CNN model with the conventional ANN and RA models showed that the CNN model is more suitable for highway cost prediction, which is a high-dimensional nonlinear problem, and it can improve the prediction accuracy.

This study can be used to devise financial plans in the early stage of expressway construction and to manage cost at the conceptual design stage, so as to enable investors to acquire project funds in advance. The 10 newly proposed factors related to the bridge and tunnel allow for a more comprehensive and multi-perspective analysis of the cost-influencing factors at the conceptual design stage of an expressway project. The information of 415 expressways collected in this study can also be used as a reference for conducting similar research in the future.

Although the influencing factors of the project cost include the *starting date of the project, investment in fixed assets, CPI, PPI, benchmark lending rates, GDP*, and other factors related to time series, the CNN model only studies the relationship between these index values and the highway project cost; it does not learn the characteristics of data from the perspective of overall time series. In the future, researchers can combine the advantages of the RNN-based machine learning model that learns time series with those of the CNN model.

REFERENCES

[1] A. Costin, A. Adibfar, H. Hu, and S. S. Chen, "Building information modeling (BIM) for transportation infrastructure–literature review, applications, challenges, and recommendations," *Autom. Construct.*, vol. 94, pp. 257–281, Oct. 2018, doi: 10.1016/j.autcon.2018.07.001.

[2] K. R. Molenaar, "Programmatic cost risk analysis for highway megaprojects," *J. Constr. Eng. Manag.*, vol. 131, no. 3, pp. 343–353, 2005, doi: 10.1061/(ASCE)0733-9364(2005)131:3(343).

[3] S. H. Han, D. Y. Kim, and H. Kim, "Predicting profit performance for selecting candidate international construction projects," *J. Constr. Eng. Manag.*, vol. 133, no. 6, pp. 425–436, 2007, doi: 10.1061/(ASCE)0733-9364(2007)133:6(425).

[4] Y. Tang, R. Liu, F. Wang, Q. Sun, and A. A. Kandil, "Scheduling optimization of linear schedule with constraint programming," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 2, pp. 124–151, Feb. 2018, doi: 10.1111/mice.12277.

[5] Y. Tang, R. Liu, and Q. Sun, "Schedule control model for linear projects based on linear scheduling method and constraint programming," *Autom. Construct.*, vol. 37, pp. 22–37, Jan. 2014, doi: 10.1016/j.autcon.2013.09.008.

[6] S. M. Trost and G. D. Oberlender, "Predicting accuracy of early cost estimates using factor analysis and multivariate regression," *J. Constr. Eng. Manag.*, vol. 129, no. 2, pp. 198–204, 2003, doi: 10.1061/(ASCE)0733-9364(2003)129:2(198).

[7] S. M. Ojo, "Cost of construction projects in Nigeria challenges and ways forward," in *Proc. West Africa Built Environ. Res. (WABER) Conf., 10th Anniversary Conf.*, Aug. 2019, pp. 5–7.

[8] B.-C. Kim and Y. H. Kwak, "Improving the accuracy and operational predictability of project cost forecasts: An adaptive combination approach," *Prod. Planning Control*, vol. 29, no. 9, pp. 743–760, Jul. 2018.

[9] D. J. Lowe, M. W. Emsley, and A. Harding, "Predicting construction cost using multiple regression techniques," *J. Constr. Eng. Manag.*, vol. 132, no. 7, pp. 750–758, 2006, doi: 10.1061/(ASCE)0733-9364(2006)132:7(750).

[10] T. Hegazy and A. Ayed, "Neural network model for parametric cost estimation of highway projects," *J. Constr. Eng. Manag.*, vol. 124, no. 3, pp. 210–218, 1998, doi: 10.1061/(ASCE)0733-9364(1998)124:3(210).

[11] H. Adeli and M. Wu, "Regularization neural network for construction cost estimation," *J. Constr. Eng. Manag.*, vol. 124, no. 1, pp. 18–24, 1998, doi: 10.1061/(ASCE)0733-9364(1998)124:1(18).

[12] H. Al-Tabtabai, A. P. Alex, and M. Tantash, "Preliminary cost estimation of highway construction using neural networks," *Cost Eng. (Morgantown)*, vol. 41, no. 3, pp. 19–24, 1999.

[13] C. G. Wilmot and B. Mei, "Neural network modeling of highway construction costs," *J. Constr. Eng. Manag.*, vol. 131, no. 7, pp. 765–771, 2005, doi: 10.1061/(ASCE)0733-9364(2005)131:7(765).

[14] J. Sodikov, "Cost estimation of highway projects in developing countries: Artificial neural network approach," *J. East. Asia Soc. Transp. Stud.*, vol. 6, pp. 1036–1047, 2005, doi: 10.11175/easts.6.1036.

[15] S. Anderson, K. Molenaar, and C. Schexnayder, "Guidance for cost estimation and management for highway projects during planning, programming, and preconstruction," National Cooperative Highway Research Program (NCHRP), Washington, DC, USA, Final Rep.574, 2006.

[16] J.-S. Chou and J. T. O'Connor, "Internet-based preliminary highway construction cost estimating database," *Autom. Construction*, vol. 17, no. 1, pp. 65–74, Nov. 2007, doi: 10.1016/j.autcon.2007.03.001.

[17] D. Y. Kim, B. Kim, and S. H. Han, "Two-staged early cost estimation for highway construction projects," in *Proc. 25th Int. Symp. Autom. Robot. Construct., ISARC*, Jun. 2008, pp. 490–495.

[18] A. Attal, "Development of neural network models for prediction of highway construction cost and project duration," M.S. thesis, Russ College Eng. Technol., Ohio Univ., Athens, OH, USA, 2010.

[19] M. E. Asmar, A. S. Hanna, and G. C. Whited, "New approach to developing conceptual cost estimates for highway projects," *J. Constr. Eng. Manag.*, vol. 37, no. 11, pp. 942–949, 2011, doi: 10.1061/(ASCE)CO.1943-7862.0000355.

[20] E. Elbeltagi, R. H. Abdel-Razek, and O. Hosny, "Conceptual cost estimate of Libyan highway projects using artificial neural network," *Int. J. Eng. Res. Appl.*, vol. 4, no. 8, pp. 56–66, 2014.

[21] O. Swei, J. Gregory, and R. Kirchain, "Construction cost estimation: A parametric approach for better estimates of expected cost and variation," *Transp. Res. B, Methodol.*, vol. 101, pp. 295–305, Jul. 2017.

[22] K. Tijanić, D. Car-Pušic, and M. Šperac, "Cost estimation in road construction using artificial neural network," *Neural Comput. Appl.*, vol. 32, no. 13, pp. 9343–9355, Jul. 2020.

[23] G. H. Kim, D. S. Seo, and K. I. Kang, "Hybrid models of neural networks and genetic algorithms for predicting preliminary cost estimates," *J. Comput. Civ. Eng.*, vol. 19, no. 2, pp. 208–211, 2005, doi: 10.1061/(ASCE)0887-3801(2005)19:2(208).

[24] G.-H. Kim, J.-E. Yoon, S.-H. An, H.-H. Cho, and K.-I. Kang, "Neural network model incorporating a genetic algorithm in estimating construction costs," *Building Environ.*, vol. 39, no. 11, pp. 1333–1340, Nov. 2004, doi: 10.1016/j.buildenv.2004.03.009.

[25] R. Sonmez, "Parametric range estimating of building costs using regression models and bootstrap," *J. Constr. Eng. Manag.*, vol. 134, no. 12, pp. 1011–1016, 2008, doi: 10.1061/(ASCE)0733-9364(2008)134:12(1011).

[26] J. Ahn, M. Park, H.-S. Lee, S. J. Ahn, S.-H. Ji, K. Song, and B.-S. Son, "Covariance effect analysis of similarity measurement methods for early construction cost estimation using case-based reasoning," *Autom. Construct.*, vol. 81, pp. 254–266, Sep. 2017, doi: 10.1016/j.autcon.2017.04.009.

[27] M. H. Rafiei and H. Adeli, "A novel unsupervised deep learning model for global and local health condition assessment of structures," *Eng. Struct.*, vol. 156, pp. 598–607, Feb. 2018, doi: 10.1016/j.engstruct.2017.10.070.

[28] L. C. Chao and M. J. Skibniewski, "Neural networks of estimating construction technology acceptability," *J. Constr. Eng. Manag.*, vol. 121, no. 1, pp. 130–142, 1995, doi: 10.1061/(ASCE)0733-9364(1995)121:1(130).

[29] G. E. Hinton, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006, doi: 10.1126/science.1127647.

[30] M. H. Rafiei and H. Adeli, "Novel machine-learning model for estimating construction costs considering economic variables and indexes," *J. Constr. Eng. Manag.*, vol. 144, no. 12, 2018, Art. no. 04018106, doi: 10.1061/(ASCE)CO.1943-7862.0001570.

[31] Y. Elfahham, "Estimation and prediction of construction cost index using neural networks, time series, and regression," *Alexandria Eng. J.*, vol. 58, no. 2, pp. 499–506, Jun. 2019.

[32] V. Chandanshive and A. R. Kambekar, "Estimation of building construction cost using artificial neural networks," *J. Soft Comput. Civ. Eng.*, vol. 3, no. 1, pp. 91–107 2019.

[33] A. M. Abd, N. A. Jasim, and F. S. Naseef, "Predicting the final cost of iraqi construction project using artificial neural network (ANN)," *Indian J. Scie. Technol.*, vol. 12, p. 28, Jul. 2019, doi: 10.17485/ijst/2019/v12i28/145640.

[34] S. Tayefeh Hashemi, O. M. Ebadati, and H. Kaur, "Cost estimation and prediction in construction projects: A systematic review on machine learning techniques," *Social Netw. Appl. Sci.*, vol. 2, no. 10, p. 1703, Oct. 2020.

[35] M. Cai, M. Pipattanasomporn, and S. Rahman, "Day-ahead building-level load forecasts using deep learning vs. Traditional time-series techniques," *Appl. Energy*, vol. 236, pp. 1078–1088, Feb. 2019, doi: 10.1016/j.apenergy.2018.12.042.

[36] B.-F. Chen, B. Chen, H.-T. Lin, and R. L. Elsberry, "Estimating tropical cyclone intensity by satellite imagery utilizing convolutional neural networks," *Weather Forecasting*, vol. 34, no. 2, pp. 447–465, Apr. 2019, doi: 10.1175/WAF-D-18-0136.1.

[37] B. Yang, R. Liu, and E. Zio, "Remaining useful life prediction based on a double-convolutional neural network architecture," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9521–9530, Dec. 2019, doi: 10.1109/TIE.2019.2924605.

[38] Q. Yang, L. Shi, J. Han, Y. Zha, and P. Zhu, "Deep convolutional neural networks for rice grain yield estimation at the ripening stage using UAV-based remotely sensed images," *Field Crops Res.*, vol. 235, pp. 142–153, Apr. 2019, doi: 10.1016/j.fcr.2019.02.022.

[39] J. Delgado, L. Oyedele, M. Bilal, A. Ajayi, L. Akanbi, and O. Akinade, "Big data analytics system for costing power transmission projects," *J. Constr. Eng. Manag.*, vol. 146, no. 1, pp. 386–474, 2020, doi: 10.1061/(ASCE)CO.1943-7862.0001745.

[40] X. Liu, H. Cai, R. Zhong, W. Sun, and J. Chen, "Learning traffic as images for incident detection using convolutional neural networks," *IEEE Access*, vol. 8, pp. 7916–7924, 2020.

[41] G. D. Oberlender and M. Trost, "Predicting accuracy of early cost estimates based on estimate quality," *J. Constr. Eng. Manag.*, vol. 127, no. 3, pp. 173–182, 2001, doi: 10.1061/(ASCE)0733-9364(2001)127:3(173).

[42] Y.-H. Huang and H.-Y. Huang, "A model for concurrent maintenance of bridge elements," *Autom. Construction*, vol. 21, pp. 74–80, Jan. 2012, doi: 10.1016/j.autcon.2011.05.014.

[43] Z. Herbsman, "Model for forecasting highway construction cost," *Transp. Res. Rec.*, vol. 1056, pp. 47–54, 1986.

[44] C. G. Wilmot and G. Cheng, "Estimating future highway construction costs," *J. Constr. Eng. Manag.*, vol. 129, no. 3, pp. 272–279, 2003, doi: 10.1061/(ASCE)0733-9364(2003)129:3(272).

[45] E. Koehn, R. Young, F. Seling, and B. Fallon, "Cost of environmental impact statements," *J. Constr. Div. ASCE*, vol. 105, no. 4, pp. 331–339, 1979.

[46] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998, doi: 10.1109/5.726791.

[47] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[48] S. Yu, S. Jia, and C. Xu, "Convolutional neural networks for hyperspectral image classification," *Neurocomputing*, vol. 219, pp. 88–98, Jan. 2017, doi: 10.1016/j.neucom.2016.09.010.

[49] T. Hegazy, O. Moselhi, and P. Fazio, "Developing practical neural network application using backpropagation," *J. Microcomput. Civil Eng.*, vol. 9, no. 2, pp. 145–159, 1994.

[50] R. L. Peurifoy and G. D. Oberlender, *Estimating Construction Costs*, 5th ed. New York, NY, USA: McGraw-Hill, 2002.

YUANHUA JIA received the M.S. degree in railway transportation and the Ph.D. degree in industrial economics from Beijing Jiaotong University, Beijing, China, in 1987 and 2002, respectively.

He has been serving as a Professor with the School of Traffic and Transportation, Beijing Jiaotong University, since 2002. His research interests include the coordination of traffic and the economy, logistics management, and transportation hub management.

XIAOJUAN XUE received the bachelor's degree in transportation engineering from Lanzhou Jiaotong University, Lanzhou, Gansu, China, in 2003, and the M.S. degree in transportation planning and management from Chang'an University, Xi'an, Shaanxi, China, in 2011. She is currently pursuing the Ph.D. degree in the field of traffic and transportation planning and management with Beijing Jiaotong University, Beijing, China.

She has worked as an Assistant Engineer with the Xi'an West Railway Station of Xi'an Railway Bureau, Xi'an, from 2003 to 2008. She has worked as an Engineer with CCCC First Highway Consultants Company Ltd., Xi'an, from 2011 to 2016, where she has been working as a Senior Engineer since 2016. Her research interests include project cost forecasts, the impact of transportation infrastructure on the economy, and the relationship between big data and the transportation economy.

YUANJIE TANG (Member, IEEE) received the M.S. and Ph.D. degrees in traffic and transportation engineering from Beijing Jiaotong University, Beijing, China, in 2010 and 2015, respectively.

From 2015 to 2017, he was a Postdoctoral Fellow with the Construction Engineering and Management Department, Purdue University, West Lafayette, USA. From 2017 to 2019, he was an Assistant Professor with the School of Traffic and Transportation, Beijing Jiaotong University, where he has been working as an Associate Professor since 2020. His research interests include project scheduling optimization and transportation infrastructure management.

Dr. Tang is an Associate Member of ASCE.

• • •