Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Study on Reworks for Commercial Complex Projects Based on Deep Learning Neural Network

Wenyu Zhang

wenyu_970163.com

Hong Kong Polytechnic University

Xiang Ding

Article

Keywords: Commercial complexes, Reworks, Rework factors, Rework cost prediction, Deel learning neural network

Posted Date: February 24th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-7922408/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Commercial complexes have emerged as the main form in the modern urban development, the implementation process of which is characterized by diverse professional teams, multiple levels of construction, and more stakeholders. However, reworks and associated costs in commercial complex projects have not been addressed. This study focuses on reworks of the commercial complex projects including design and construction. The actual commercial complex projects with data of rework costs are based to identify controllable factors contributing to reworks and collect rework dataset including scores of the factors and rework costs. A deep learning neural network (DLNN) in multi-layer perceptron (MLP) form is utilized to predict rework costs, and is justified by comparing with other neural network applications. A case commercial complex project is applied to demonstrate the proposed MLP-based DLNN in predicting rework cost and analyze the importance of rework factors with SHAP values. With investigation into rework factors and prediction of rework costs, this study enables to achieve management and control measures on reworks in the design and construction processes of commercial complexes, thus promoting cost-effectiveness and sustainability in modern urban construction.

Introduction

The acceleration of urbanization has led to a massive influx of people into cities, fostering a stronger commercial atmosphere and presenting significant development opportunities for commercial real estate in China (Hong, et al. 2025). As a carrier for effectively enhancing socio-economic development, urban commercial complexes reflect their economic value in terms of the consumption industry chain and production cohesion (Kim 2011). Commercial complex projects involve large-scale construction, multi-disciplinary integration, more stakeholders, and fragmented organization structure (Mamaghani and Noorzai 2023), making them generally subject to uncertainties and reworks in the construction process. Therefore, the phenomenon of construction duration and cost overrun is particularly evident in the commercial complex project due to the reworks (Moon et al. 2015). As the owners responsible for both investment and operation of the commercial complex projects, they are very sensitive to the cost and duration of the project. Obviously, it is imperative for the owners to perform management on rework of commercial complex projects by identifying rework factors, predicting rework cost, and adopting preventive measures (Elmousalami et al. 2018; Tijani´c et al. 2020; Elsherbiny 2021). Nevertheless, few of studies on rework of commercial complex projects have been noticed. Some studies indicated that project type was linked to the concept of complexity, which correspondingly affects project performance including rework costs (Hwang et al. 2009). Hence, it is crucial to study the reworks of commercial complex projects by addressing their characteristics.

Construction sector has been criticized for its poor performance regarding cost and schedule overrun. Rework has been identified as one critical problem contributing to poor performance of the construction project, such as cost increase, delay, conflicts, and disputes (Love et al. 2022). The costs of reworks on residential, industrial and commercial building projects ranged from 2 to 6% of the contract values (Forcada et al. 2017). It has been reported that there exist different percentages for the rework costs on total project costs (Love et al. 2022). Matthews et al. (2022) reported that the rework cost amounts to 12% of the total project cost, whereas this percentage may be higher according to some studies (Love 2002; , Love and Sing 2013; Simpeh et al. 2015). The rework was found account for 5.06% of the completion cost of new buildings (Oyewobi et al. 2011). It was also found that cost growth and schedule overruns were significantly correlated with direct rework costs, meaning that reworks adversely affect project performances (Love 2002; Palaneeswaran et al. 2008). However, most of the existing studies on reworks regarding identifying and scoring rework factors are completely based on experience or knowledge of interviewee rather than actual projects with historical records. In addition, the reworks of commercial complex projects have not been addressed so far.

In addition to identifying and quantifying rework factors, prediction of rework costs have been also addressed by using machine learning (ML) techniques like random forest regression and artificial neural network (ANN) (Shoar et al. 2022;

Ma et al. 2021; Alaloul et al, 2018; Palaneeswaran et al. 2008). Other ANN applications include the prediction of rework costs caused by communication factors in construction projects (Trach et al. 2021) and the prediction of rework costs in high-rise buildings by addressing interrelationships among factors through the graph convolutional network (GCN) (Mostofi et al. 2024). However, deep learning network (DLNN) that is able to learn and grasp complex patterns behind reworks has not been applied to predict rework costs. In addition, no studies on predicting rework costs of commercial complex projects have been noticed.

This study focuses on rework and incurred cost of the commercial complex project containing design, construction and procurement processes. First, backgrounds about commercial complexes, reworks in construction, rework factors, rework costs and prediction methods are discussed. Second, identifying and scoring rework factors corresponding to actual rework costs are conducted based on actual projects and interview of participants, thus developing the sample dataset of reworks. Then, a multi-layer perceptron (MLP)-based DLNN is proposed to predict rework costs, and is justified by comparison with other ANN applications. Finally, a case commercial complex project is used to demonstrate the rework cost prediction and analyze the importance of rework factors with SHAP values.

Backgrounds

Comeriacial Complexes

The changes of lifestyle, pattern of motorized transportation and intensive activities propel people towards shopping centers, namely commercial complexes that reflect such a consumer culture of the modern era (Lee, 2022). Commercial complexes provide experiences at various levels, such as leisure-oriented cultural consumption, as well as traditional consumption. Additionally, they contain various facilities such as residences, businesses, cultural venues, and entertainment venues. The commercial complex is, thus, a space that symbolizes modern consumption culture and partly fulfilling social demands (Hong, et al. 2025) .

The entire construction process of a commercial complex is called commercial complex project in the paper. A commercial complex project can be identified if a project contains multiple trades operating within the construction area, unique trades that are not commonly used in a general construction project, and/or multiple stakeholders, i.e. multiple decision-makers (Mamaghani and Noorzai, 2023). In general, the commercial complex projects have the characteristics: 1) Multifunctional integration is a prominent feature of a commercial complex. In addition, multiple aspects such as traffic flow, functional layout, architectural aesthetics, and sustainability are considered. Therefore, careful design and planning are needed to ensure effectiveness in construction and coordination in operation. 2) The commercial complex contains underground infrastructure, above ground buildings, and the connection with rail transit facilities (Kim 2011). Hence, the commercial complex project involve multiple underground and above ground construction related to infrastructure and buildings, as well as mechanical, electrical, plumbing, and HVAC (Hong, et al. 2025). 3) As the investor and operator of a commercial complex, the owner may adjust function, quality, investment, and schedule of the project to adapt market requirements. Due to the complex characteristics, the commercial complex project faces risks related to not only complexities, technologies and geology but also dynamic adjustments on design, scale and finance, causing heavy reworks and expected costs (Moon et al. 2015).

Reworks in Construction

Mahamid (2024) reported that rework is a major challenge that leads to project failure in the construction industry. Palaneeswaran et al. (2005) concluded that rework leads to schedule delay, material waste excess and cost overrun. Wasfy (2010) found that rework increases project cost and schedule by values reaching to 30% and 70%, respectively.

Rework is simply defined as redoing an activity that is incorrectly implemented (Mahamid 2024). Actually, many definitions of rework can be found in literature. CII (2001) defined rework as activities that must be performed multiple

times on-site. Hwang et al. (2009) defined rework as unnecessary work due to not meeting required quality, stakeholders' order changes, and omissions or errors of stakeholders. Love 2002defined rework as an event or process that is caused by quality violations, unqualified quality problems, deviations, or errors. Forcada et al. (2017) noted that any additional work resulting from order changes, design errors, and volume changes should also be considered rework. Rework is known as non-value adding symptoms that affect the productivity and performance in construction projects, the cause of which is a discrepancy between the actual and planned state of the object and requires compliance (Shah et al. 2016).

Rework Factors

The main factors of reworks are considered changes, errors, omissions, core tasks, as well as team experience and poor coordination (Palaneeswaran et al. 2005, 2008). The investigation of Safapour et al. (2019) concluded that the main rework causes are errors and omissions, project scope changes, inadequate specification, lack of work skills, poor oversight. Khalesi et al. (2020) identified and ranked 49 rework factors related to engineering and reviews, material and equipment supply, human resource capability, construction planning and scheduling, leadership and communication. The most important causes include "design errors", "difference among plans and operational specifications", "non-compliance with specifications", "insufficient skill level", "unrealistic schedules", "poor communication" and "economic fluctuations". Al-Janabi et al. (2020) summarized 87 reasons for the rework and identified the five most common rework factors such as unfinished design at the time of the tender, poor communication between the client and the contractor/design consultant, insufficient labor skills, the client's lack of knowledge and experience in the design and construction process, insufficient contractor attention, and subcontractor to quality.

Trach et al. (2021) identified the rework causes in construction projects from the point of view of a client, designer and contractor, in which poor communication was mentioned as first cause. Owner's adjustment and design error/omission are considered to be root causes of rework and have a relatively greater cost impact than other factors (Hwang et al. 2009, 2014; Love and Edwards 2013). Hence, design inconsistencies are also considered major factors of reworks (Love et al. 2009). In addition, Sommerville (2007) investigated the causes and cost impact of rework on residential projects and pointed out that 50% of rework is attributed to design errors, 40% of rework is attributed to construction errors and 10% of rework is attributed to product defects. Ye et al. (2015) identified 39 major causes of reworks, some of which are associated with communication management, including poor coordination, poor way of passing instructions, and poor communication among participants. Love et al. (2019) indicated the management demand to use limited resources as the main contributor to field rework. Love et al. (2018) further revealed that interference of corporate senior manager in project decisions is a major reason of rework. Mahamid (2020) identified 51 manageable rework reasons based on collecting and analyzing 44 data from case studies.

Rework Costs

A lot of scholars have attempted to quantify or measure the degrees of reworks, thus helping owners or contractors to plan and perform control during both construction and design processes (Rogge 2001; Fayek et al 2004). Rogge (2001) developed the field rework index (FRI) to predict and track field rework. CII (2001) suggested total field rework factor (TFRF) measured with total direct cost of field rework over actual construction phase cost as a metric to measure the amount of field rework.

Shah et al. (2016) used the TFRF to give early warning of rework severity to contractors for adopting appropriate control measures. Fayek et al (2004) developed construction field rework index (CFRI) to track field rework by adding the total direct cost with indirect cost and dividing the result by the total field construction cost. Oyewobi et al. (2011) revealed that rework cost is about 5% in new building projects. Liu et al. (2018, 2020) pointed out that rework cost in residential projects is about 5% of the total project cost, and concluded that the main contributors to rework cost include

contractors (20.10%), design (18.91%) and client (14.73%). Forcade et al. (2017) revealed that rework results in cost overrun of 11% of the contract value of civil and building projects.

These indicators could quantify or measure the severity of reworks, however, their predictions rather than estimates afterwards based on actual results are more important for the owners or contractors to plan and implement management or control strategies (Trach et.al, 2021; Palaneeswaran et al. 2008) .

Rework Cost Prediction through ANN

Machine learning (ML) methods including ANNs are suitable for the problems that are difficult to solve using traditional mathematical methods and have been widely utilized in the construction industry for prediction and classification (Elmousalami et al. 2018; Tijani´c et al. 2020; Elsherbiny 2021; Liu et al. 2021) utilized three machine learning (ML) algorithms to predict the design error rework cost during the engineering and construction phase of industrial EPC projects, where the random forest (RF) algorithm outperformed other two algorithms. Forcada et al. (2017) adopted stepwise multiple regression to identify the main contributors to rework cost by utilizing the historical data rather than the values of rework costs gathered via questionnaire surveys.

As one of the ML techniques, ANN can learn from complex patterns and has proved its reliability and efficiency in the construction industry for cost forecasting, decision making, optimization, dispute resolution and risk assessment (Trach et.al, 2021; Liu et al. 2021; Palaneeswaran et al. 2008; Tijani´c et al. 2020). ANNs have also been applied to predict rework costs, though the studies are limited (Trach et al. 2021). Palaneeswaran et al. (2008) utilized general regression neural network (GRNN) and traditional backward propagation neural network (BPNN) to predict the impact of rework on schedule, cost, contractual claims. Alaloul et al. (2018) developed an ANN model to assess the coordination factors' influence on the construction projects delivery. Trach (et al. 2021) applied the neural network technique to predict the rework cost caused by communication factors in construction projects. Attia (2022) integrated BIM and ANN to predict reworks on the construction site. Mostofi et al. (2024) utilized the graph convolutional network (GCN) model to predict the rework cost in high-rise buildings by accommodating interrelationships among causes of rework, but the requirement of clarifying interrelationships among causes leading to difficulties in preparing input data.

Summary Analysis

Commercial complex projects typically involve multiple stakeholders, comprehensive design, and multiple functional construction, all of which increase the likelihood of reworks. Though some studies on commercial complexes have been performed recently, the attempt to explore the reworks of commercial complex projects has not been noticed so far. Many efforts have been paid to investigate rework factors in construction, but no studies on rework factors in both design and construction phases of commercial complexes have been conducted. Most of the previous studies on identifying and scoring rework factors as well as collecting dataset were achieved through questionnaire surveys that rely on knowledge, experience and subjective judgments rather than historical records of actual projects. ANNs have been applied to predict rework costs, no such applications in commercial complex projects have been found, not to mention the use of deep learning neural networks like MLP and CNN (Convolutional Neural Network).

This study will address the above issues in focusing on reworks of the commercial complex project during the entire process including design and construction. Actual projects and associated historical records of commercial complex projects will be based to identify rework factors and collect rework dataset. Then, a DLNN in the form of MLP will be proposed to predict rework cost of a commercial complex project by utilizing the developed rework dataset.

Identification and Scoring of Rework Factors

Commercial Complex Projects under Study

Commercial complexes typically encompass a variety of functions, such as shopping, office space, hotels, residences, and entertainment. A real estate company P of Hangzhou, China, mainly engages in the development and operation of commercial complexes. 18 commercial complex projects developed by Company P in Zhejiang province from 2016 to 2023 were selected to achieve identifying and scoring of rework factors for the commercial complex projects. The selected 18 commercial complex projects were developed by using the EPC (Engineering, Procurement, Construction), DB (Design and Building), or DBB (Design, bidding, Building) contracting forms, and the focus of this study is on the period from design to project completion.

In general, identifying and scoring rework factors for construction projects are achieved completely through questionnaire survey. In contrast, this study relies on actual commercial complex projects and their historical records of reworks and costs to identify and score the main controllable rework factors. The rework rate measured as the percentage of the rework cost to the planned project cost before signing EPC or design contract is adopted herein. Based on the historical records of Company P, the planned costs, rework costs, and corresponding rework rates of the 18 commercial complex projects are shown in Table 1.

Table 1 Rework Data Information Table of 18 Commercial Complex Projects

Projects	Building Area (m^2)	Planned Costs (Million RMB)	Rework Costs (Million RMB)	Rework Rate□%□
1	573301	416.56	1692.33	4.06
2	347863	330.21	1334.76	4.04
3	118101	1395.21	5899.38	4.23
4	118881	242.85	1396.24	5.75
5	395060	6057.12	62638.51	10.34
6	67758	141.84	1007.74	7.10
7	207846	4159.78	60289.72	14.49
8	348479	1018.07	4123.24	4.05
9	133704	1117.77	17111.43	15.30
10	362028	732.26	2142.53	2.93
11	280525	1318.91	3531.55	2.68
12	524226	825.56	11930.14	14.45
13	36962	481.68	8301.64	17.23
14	183361	277.48	7018.38	25.29
15	148173	132.18	879.76	6.66
16	212462	425.45	1476.67	3.47
17	351260	359.88	1696.45	4.71
18	278247	885.48	3806.91	4.30
Rework rate range	[2.68, 25.29]		Average rework rate	8.39

It is shown that Company P faces the problem of high rework costs in implementing the commercial complex projects. Only three projects have rework rates below 4%, three projects have rework rates above 15%, and the average rework rates are close to 8%. It is indicated that most of the commercial complex projects have challenges in cost management and control during the entire process containing design, construction and procurement.

Identification of Rework Factors

Based on previous studies on rework factors (Love et al. 2008, 2019; Forcada et al. 2017; Ye et al. 2015; CII 2001), 32 potential rework factors are identified, which can be controlled through management strategies. Additionally, some rework factors are mostly attributed to uncontrollable situations (such as weather, legal and political factors), which are not considered or identified in the study. With regards to the characteristics of the commercial complex project, more potential rework factors are considered, thus total 40 potential rework factors are initially identified.

Based on the 40 potential rework factors, five semi-structured interviews were conducted based on actual commercial complex projects. The participants interviewed all have experience participating in multiple commercial complex projects, including design, construction or procurement and management works. Based on the actual projects and interviews, 20 main controllable rework factors were identified as shown in Table 2. The factors are divided into three categories, such as overall, design, construction (including procurement.

Table 2 Identified Main Controllable Rework Factors in Commercial Complex Projects

No	Controllable rework factors	Category
1	Inadequate supervision of project managers by relevant parties	Overall
2	Too many financial approval procedures	
3	Inadequate timeliness of the owner's review and approval	
4	Inadequate function and quality match between stakeholders or phases, or too many dynamic adjustment of function and quality	
5	Insufficient number of project management personnel	
5-6	Inadequate management experience and capability in design for the project manager team	Design
6-7	Too many stakeholders involving design approval	
7-8	Inadequate internal communication within the design department	
8-9	Inadequate technical capability of the design department	
9-10	Inadequate supervision and management during the design phase	
10	Too long time for taking design approval	
11	Severity of issues about design completion degree, error and defect before construction	
12	The complexity and difficulty of the overall design	
13	Inadequate management experience and capability in construction (including procurement) for the project manager team	Construction and Procurement
14	Inadequate communication among contractors (including subcontractors and suppliers)	
15	Inadequate supervision and management on construction (including procurement)	
16	Insufficient skills of technical workers at the construction site	
17	Poor adjustability (resilience and flexibility) of construction scheduling	
18	Severe delay of construction works after design	
19	Poor construction continuity	
20	Too many changes or adjustments during the construction phase	

It is shown that some of the identified rework factors are directly related to the owner or the project manager term during the entire construction process of the commercial complex project. In addition, the design process contributes to 7 main rework factors, meaning that the design process is important to control the rework cost for the commercial complex project that contains design. It is also noted that the construction (including procurement) process heavily impact the reworks of the commercial complex project. Hence, the construction (including procurement) process is also crucial and needs to be emphasized in achieving preventive management and control measures on reworks. The identified rework factors could reflect the characteristics of commercial complex projects.

Rework Data Collection and Processing

The rework cost is represented in terms of rework rate in this study, which is measured by the percentage of rework cost to the planned cost of the project. Such a planned cost is the budget cost estimated by the owner of the commercial

complex project before signing the EPC contract or the design contract when adopting DBB mode. The rework dataset contains the scores of rework factors and the associated rework rate for each project, and the rework rate can be used to easily calculate the rework cost.

Once the 20 controllable rework factors in commercial complex projects have been identified through literature research and expert interviews, the rework dataset used to predict the rework cost needs to be collected. Generally, one of the primary challenges in predicting rework costs through some models such as ANN lies in the acquisition and availability of actual historical data. In contrast to relying solely on questionnaires, this study firstly collected input data based on the 18 actual projects and particularly the actual rework cost data. Then, the survey questionnaires were conducted to obtain the scores (from 0 to 10) of the 20 rework factors for each commercial complex project. The independent input data of the DLNN model consists of the scores of the 20 rework factors and the rework rate of the corresponding project, i.e., $X_i\ i = 1, \ldots, 20$ and Y, only one output variable. Note that the rework rate is calculated by dividing the actual rework cost of a commercial complex project by the planned cost.

18 commercial complex projects invested and operated by Company P from 2016 to 2023 were based to collect the rework dataset. First, the rework costs or rework rates were obtained based on the actual records for the 18 commercial complex projects. Then, the survey questionnaires on the scores (from 0 to 10) of the 18 rework factors were conducted among the engineers or management members of Company P participating in the projects. 10 participants for each project and total 180 participants were involved. The scores of the rework factors for each commercial complex project were obtained by averaging the 10 items of scores.

There may exist significant correlation among input variables (i.e., X_i denoting scores of rework factors or features) coupled with multicollinearity of them, thus leading to confusion in the predicted results and increasing the probability of errors. This study employs L1 regularization (Lasso Regression) and L2 regularization (Ridge Regression) methods to eliminate high multicollinearity and select the independent variables that make the most significant contribution to the dependent variable (Wu et al. 2009).

L1 regularization is a regularization method that reduces collinearity by adding a penalty term of absolute value. Lasso Regression can not only reduce the size of feature coefficients, but also compress some coefficients to zero, thereby achieving feature selection.

L2 regularization is another regression method that reduces the impact of collinearity by adding a regularization term. Ridge Regression reduces the impact of collinearity by adding a penalty term (the sum of squares of feature coefficients) to the loss function, thereby shrinking the size of feature coefficients.

In the Python platform used in the study, array operations and matrix operations of NumPy are respectively utilized to implement L1 regularization (Lasso Regression) and L2 regularization (Ridge Regression) and obtain the corresponding regression coefficients. NumPy is a fundamental library for scientific computing in Python, providing an efficient multi-dimensional array object ndarray, as well as a lot of mathematical functions and operations. It enables rapid processing of large amounts of data that is typically inputted, processed, and outputted in neural networks by using the form of NumPy arrays. For instance, the input dataset can be a two-dimensional or three-dimensional NumPy array, where each row represents a sample and each column represents a feature. The weights and biases of neural networks can also be represented and updated using NumPy arrays.

MLP-Based Rework Cost Prediction

Briefing of Multi-Layer Perceptron (MLP) Model

ANNs are a type of computational model composed of multiple neurons that interact with each other through connections to model and learn complex problems. The connection is represented by a weight matrix and the core is the neuron, which is designed to mimic biological neurons. A neural network consists one input layer, one output layer, and one ore more hidden layers. The input layer receives data from the external environment, with each node corresponding to an input feature. The output layer output the final results such as classification labels, prediction values, etc. Hidden layers are located between the input layer and the output layer. The nodes in each hidden layer receive input signals from the previous layer through weighted connections and undergo nonlinear transformation through an activation function.

The complexity of a neural network usually stems from the number of hidden layers and the number of neurons in each layer, leading to more costs of time and resources in implementing under CPU environment. Hence, only one hidden layer is considered in the traditional neural network model, but may limit deep earning and training from complex environments.

Compared to traditional neural networks, deep learning neural network (DLNN) in the forms including MLP and CNN have more hidden layers. The larger the network and the more layers it has, the more complex it becomes, requiring more resources and time for training. DLNN works best when paired with GPU-based architectures, achieving shorter training times compared to classical CPUs.

The MLP-based DLNN belongs to the feedforward neural networks with multiple layers of perceptron and possesses an activation function (Karlik and Olgac 2011) . MLP consists of a fully connected input layer, hidden layers and output layer, where every neuron in each layer is connected to every neuron in the preceding layer, as shown in Figure 1. The MLP-based DLNN enables to learn complex nonlinear relationships through flexible structure and adjustable layers and neuron counts.

The MLP-based DLNN is able to address a wide range of problems where learning is under supervision by backpropagation algorithm. The learning algorithm of the MLP-based DLNN is a method used to determine the connection weights between neurons, which involves three steps: transmitting input training signals, calculating relevant errors, and adjusting weights. This network propagates the input through the network from the input layer and hidden layers to the output layer and then defines the error and propagates this error back to the input layer by incorporating it into the learning formula. It employs gradient descent to minimize the squared error i.e., Mean Squared Error, MSE between the output values and desired values for those outputs. These error signals are then used to calculate the weight updates. This process represents the knowledge learned in the networks.

MLP-Based DLNN Implementation

The MLP-based DLNN can be implemented in Python platform through several libraries such as TensorFlow and NumPy. TensorFlow is an open-source machine learning framework developed by Google, which supports the construction, training, and deployment of DLNN models. The MLP-based DLNN model can utilize the following data structures provided by Python:

1. Array: It is used to store and manipulate weights, biases, and input and output data in the neural network. NumPy library is utilized to achieve efficient array operations.
2. Matrix and Vector: Many operations concern computations such as the multiplication of weight matrices and input vectors. These operations are implemented through NumPy and the matrix library in the deep learning framework TensorFlow.
3. Tensor: Tensor is a generalization of multi-dimensional arrays and is the most fundamental data structure in deep learning. The deep learning framework TensorFlow uses tensors as its core data structure.

This study used a MLP=based DLNN model with one input layer, four hidden layers, and one output layer. Regarding the 20 rework factors and the rework rate as the one output, the input layer has 20 neurons, the output layer has only one neuron, and each hidden layer has 25 neurons. Select optimizer based on parameters and set learning rate. The learning rate is 0.01, and the loss function is set to mean squared error (MSE). Choose three optimizers, grid search tuning, prevent model overfitting through regularization, initialize weights using Glorot uniform, and iterate 50 times.

In general, indicators such as the coefficient of determination (R^2), mean squared error (MSE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE) are often used to evaluate the performance of neural networks and determine the optimal set of hyperparameters. R^2 with the range from 0 to 1 is to measure the explanatory power of the model on the variation of the dependent variable. The closer R^2 is to 1, the stronger the explanatory power of the model. RMSE and MAPE are always positive numbers, indicating that the predicted values are consistent with the actual values. MAE is a positive value, with a range from 0 to any larger value.

This study used two data combinations, i.e., 70/15/15 and 80/10/10, to compare the generalization ability and prediction accuracy of DLNN. The 70/15/15 combination indicates that the input variable matrix is divided into a 70% training set, a 15% validation set, and a 15% test set, while the 80/10/10 combination indicates that the input variable matrix is divided into a 70% training set, a 15% validation set, and a 15% test set. The performance indicators of the two combinations of dataset are shown in Table 3, each of which are the average values of the corresponding indicators of the training, validation and test phases.

Table 3 MLP-Based DLNN **Performances for Two Data Combinations**

Data combinations	R^2	RMSE	MAE	MAPE
70%:15%:15%	0.9538	0.5627	0.4862	18.2382
80%:10%:10%	0.9538	0.5106	0.4329	17.7854

It shows that although the R^2 scores of the two models are the same, the 80/10/10 combination performs better in terms of RMSE and MAE, which means that its prediction accuracy on the test set is higher. The difference in MAPE is not significant, but it still indicates that the pattern is slightly better. Therefore, the 80/10/10 combination performs better in terms of prediction accuracy.

Comparison with Other ANN Applications

In order to justify the proposed MLP-based DLNN, its prediction perfromances based on the 80/10/10 data-combination are compared with the performances of some ANN applications in predicting costs or rework costs of building construction projects or infrastructure construction projects. The performances in terms of R^2, RMSE, MAE, and MAPE are listed in Table 4. These applications used back propagation neural network (BPNN) including MLP where learning is under supervision by backpropagation algorithm, or general regression neural network (GRNN). According to the meanings of the indicators, it is indicated that the performances of the applications in infrastructure project including road construction are better than the building project. On the other hand, the performances in the construction projects that may contain both building and infrastructure construction are shown between the two situations. The commercial complex project is more complicated in addition to containing both building and infrastructure construction, so the comparison demonstrates that the MLP-based DLNN model performs outstanding in predicting rework cost. Note that the MLP-based DLNN model used more hidden layers in contrast to the other BPNN models in Table 4, where the BPNN model of Alaloul et al. (2018) used two hidden layers and the others used one hidden layer. It is demonstrated that the

DLNN in the form of MLP leads to better performances in predicting rework costs by considering more numbers of hidden layers and neurons under GPU environment.

Table 4 Comparison of ANN Performance Metrics among Several Applications

Applications	Predicting Purposes	R^2		RMSE		MAE		MAPE	
		GRNN	MLP/ BPNN	GRNN	MLP/ BPNN	GRNN	MLP/ BPNN	GRNN	MLP/ BPNN
Palaneeswaran et al. (2008)	Rework cost for building project	0.5773	0.2628	0.7308	0.9644	0.4030	0.7520	□	□
	Rework cost for infrastructure work	0.9937	0.3099	0.0894	0.6050	0.0240	0.4140	□	□
Tijanic et al. (2020)	Cost for road project	0.9595	0.9688	□	□	□	□	13.06	27.47
Alaloul et al. (2018)	Cost for construction project	□	0.7700	□	0.8089	□	0.6471	□	□
This study	Rework cost for commercial complexes	□	0.9538	□	0.5106	□	0.4329	□	17.7854

Case Project Study

The MLP-based DLNN is applied to a case commercial complex project in predicting the rework rate. The importance of rework factors on the rework cost is analyzed based on SHAP values.

Rework Factor Scoring and Rework Cost Prediction

The case commercial complex project is located in Hangzhou of China, with the planned cost of 651.90 million RMB before signing the EPC contract. It includes a commercial shopping center, commercial street, offices, apartments, two basement levels, and a subway connection.

Based on the identified 20 main controllable rework factors for commercial complex projects, scoring of the rework factors for the case commercial complex project is conducted through the interview method as adopted in collecting rework dataset. The interviewed scores ranging from 1 to 10 are averaged to obtained the results used as the input data for the MLP-based neural network model, as shown in Table 5.

Table 5 Input Data for Rework Factors in Commercial Complex Project A

Input variables	Controllable rework factors	Scores
X_1	Inadequate supervision of project managers by relevant parties	7.75
X_2	Too many financial approval procedures	6.34
X_3	Inadequate timeliness of the owner's review and approval	7.12
X_4	Inadequate function and quality match between stakeholders or phases, or too many dynamic adjustment of function and quality	8.06
X_5	Insufficient number of project management personnel	5.78
X_6	Inadequate management experience and capability in design for the project manager team	8.14
X_7	Too many stakeholders involving design approval	8.21
X_8	Inadequate internal communication within the design department	7.28
X_9	Inadequate technical capability of the design department	6.05
X_{10}	Inadequate supervision and management during the design phase	8.22
X_{11}	Severity of issues about design completion degree, error and defect before construction	7.30
X_{12}	The complexity and difficulty of the overall design	7.50
X_{13}	Inadequate management experience and capability in construction (including procurement) for the project manager team	5.04
X_{14}	Inadequate communication among contractors (including subcontractors and suppliers)	5.15
X_{15}	Inadequate supervision and management on construction (including procurement)	7.25
X_{16}	Insufficient skills of technical workers at the construction site	5.18
X_{17}	Poor adjustability (resilience and flexibility) of construction scheduling	8.27
X_{18}	Severe delay of construction works after design	5.10
X_{19}	Poor construction continuity	8.40
X_{20}	Too many changes or adjustments during the construction phase	6.30

Based on the input data as shown in Table 6, i.e., X_1-X_{20} representing the scores of the 20 rework factors, the MLP-based DLNN model that has been trained well is used to predict the rework rate of the case project. The predicted rework rate of the case project is 7.07% and the rework cost is 46.09 million RMB. This prediction result has the performance as shown in Table 3 in contrast to the actual rework rate 6.20% and actual rework cost of 40.41million RMB of the case project

Importance Analysis of Rework Factors

In order to help achieve preventive management and control on reworks of the commercial complex project, the prediction results of the case project are analyzed to clarify the importance of the 20 factors in affecting rework costs.

SHAP (SHapley Additive exPlanations) values are based to perform such an analysis.

SHAP is a game-theory-based method for interpreting machine learning models, designed to quantify the specific contribution of each factor or feature to the prediction (Wang et al. 2024). In Python, SHAP calculation can be implemented using sharp library to obtain summary plots, illustrating the overall importance distribution of all features or factors.

The summary plot displays the distribution of SHAP values for all sample features, identifying the overall impact direction (positive/negative) and importance ranking of the features. The color of each point on the chart represents the magnitude of the feature value, with red indicating higher feature values and blue indicating lower feature values. Each point represents a row of data in the original dataset.

Feature Name (y-axis): The y-axis lists different features, which are ordered according to their contribution to the model output, with the most important features typically located at the top.

SHAP value (x-axis): The x-axis represents the SHAP value corresponding to each data point, indicating the specific impact of features on model predictions. A positive SHAP value indicates that the feature increases the model's predicted value, while a negative value indicates a decrease of the predicted value.

Color of points: The color of each point is encoded based on the original value of the feature, with blue typically indicating lower values and red indicating higher values. This color encoding aids in analyzing the relationship between the magnitude of a feature value and its SHAP value.

Distribution and density of points: The group of points corresponding to each feature displays the SHAP values of different samples in the dataset. The wider the distribution range, the greater the impact of that feature on the model output. The denser the dots, the more consistent the impact of changes in the feature values in that region on model predictions.

Figure 2 is a summary plot of SHAP values for the case project under study, which illustrates the impact values of the feature (X_1 to X_{20}) on the rework cost rate Y (i.e., model output). As can be seen from Figure 2, the features X_5, X_{13}, and X_4 are widely distributed, particularly X_5, and they have more significant SHAP values than others. Additionally, higher feature values (red dots) of X_5, X_{13}, and X_4 indicate their positive contributions to the prediction, while their lower values (blue dots) contribute negative SHAP values. It means that the inadequate design experience of the project manager team in the design phase (X_5), inadequate construction (including procurement) experience of the project manager team (X_{13}), and inadequate function and quality match between stakeholders or phases, or too many dynamic adjustment of function and quality (X_4) have more important and positive impacts on the rework cost. The SHAP values for poor adjustability (resilience and flexibility) of construction scheduling (X_{17}) are also widely distributed, but there are both red and blue points, indicating that these features have a bidirectional impact on the prediction, with both positive and negative effects. The SHAP values for the severity of issues about design completion degree, error and defect before construction (X_{11}) before construction mainly have positive values, and the red points (high feature values) are concentrated in the positive region, indicating that a high degree of design completion has a positive impact on model prediction.

The SHAP values for inadequate supervision and management on procurement and construction (X_{15}) are predominantly negative, with blue points (low feature values) clustered in the negative region. This indicates that lack of supervision and management on procurement and construction has a negative impact on the prediction for the case project. The SHAP values for the severe delay of construction works (X_{18}), too many changes or adjustments during the construction

phase (X_{20}), and too many financial approval procedures (X_2) indicate that these factors should be also emphasized in achieving preventive management measures for the case project.

Conclusions

Commercial complexes have become a trend with urban or urbanization development. Due to the characteristics of multifunctional integration and multiple professional team involvement in the entire construction process, the commercial complex project face a high risk of reworks and requires effective preventive management strategies. Hence, it is urgent to efefctive identify the rework factors and measure their impacts by predicting the rework costs in advance, which are the focuses of this study. Historical records about reworks and associated costs of 18 actual commercial complex projects were based to identify 20 main controllable rework factors through interview of engines and management professional participating in the corresponding projects. The rework data including scores of the factors was also collected based on the actual rework costs through interview. Then, a MLP-based DLNN was proposed to predict rework costs of the commercial complex projects based on the collected rework dataset. The comparison with other neural network applications demonstrates that the proposed DLNN obtains more satisfied performances by more effectively learning and grasping the complex patterns behind reworks.

This study provides an investigation into rework factors of the commercial complex projects based on the actual projects and historical records. In addition, it provides a MLP- neural network to predict rework costs before the design and construction process. Moreover, it helps the owner, the general contractor, or the project manager to plan and achieve preventive management strategies and measures to reduce reworks in the commercial complex project. The study contributes to the knowledge and methodologies in the domain of construction cost control, helping promote cost-effectiveness in construction and urban development by utilizing arterial intelligence methods like ANNs.

However, the identification and scoring of rework factors requires analysis and mining on the historical records, which may arise time consumption and difficulty issues. In addition, more appropriate number of hidden layers or other types of neural networks like CNN or RNN need to be explored to predict rework costs more effectively and accurately. Hence, further studies will address these issues in indenting rework factors, scoring factors and predicting rework costs by utilizing other DLNNs and other artificial intelligence techniques.

Funding Declaration

No special funding was available to support this research.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Declarations

Funding Declaration

No special funding was available to support this research.

Author Contribution

Wenyu Zhang wrote the manuscript text, perfomed analysis, developed DLNN model. and collaborately conducted data collection and survey. Xiang Ding collected historical data from actual projects and mainly conducted investigation.

Data Availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

References

1. Alaloul, W. S., Liew, M. S., Zawawi, N. A. W., Mohammed, B. S. & Adamu, M. An Artificial neural networks (ANN) model for evaluating construction project performance based on coordination factors. *Cogent Eng.* **5**, 1507657 (2018).
2. Al-Janabi, A. M., Abdel-Monem, M. S. & El-Dash, K. M. Factors causing rework and their impact on projects' performance in Egypt. *J. Civ. Eng. Manage.* **26** (7), 666–689. https://doi.org/10.3846/jcem.2020.12916 (2020).
3. Attia, R. Rework Prediction on a Construction Site through BIM & ANN Integration [Master's Thesis, the American University in Cairo]. AUC Knowledge Fountain. (2022).
4. Construction Industry Institute (CII). The field rework index: early warning for field rework and cost growth. Research summary 153-1, University of Texas at Austin, Austin, Tex. (2001).
5. Fayek, A. R., Dissanayake, M. & Campero, O. Developing a standard methodology for measuring and classifying construction field rework. *Can. J. Civ. Eng.* **31** (6), 1077–1089. https://doi.org/10.1139/l04-068 (2004).
6. Forcada, N., Gangolells, M., Casals, M. & Macarulla, M. Factors affecting rework costs in construction. *J. Constr. Eng. Manage.* **143** (8), 04017032. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001324 (2017).
7. Elsherbiny, A. *Prediction of design error rework cost in EPC industrial projects* (George Washington Univ, 2021).
8. Elmousalami, H. H., Elyamany, A. H. & Ibrahim, A. H. Evaluation of cost drivers for field canals improvement projects. *Water Resour. Manage.* **32** (1), 53–65. https://doi.org/10.1007/s11269-017-1747x (2018).
9. Hong, X., Li, S., Chen, T., Ji, X. & Song, X. Spatial performance evaluation and optimization of integrated aboveground and underground spaces in urban commercial complexes. - *J. Asian Archit.* **24** (2), 623–649 (2025).
10. Hwang, B. G., Zhao, X. & Goh, K. J. Investigating the client-related rework in building projects: The case of Singapore.Int. *J. Project Manage.* **32** (4), 698–708 (2014).
11. Hwang, B., Thomas, S., Hans, C. & Caldas, C. Measuring the impact of rework on construction cost performance. J. Constr. Eng. Manage., 10.1061/(ASCE)0733–9364(2009)135:3(187), 187–198. (2009).
12. Karlik, B. & Olgac, A. V. Performance analysis of various activation functions in generalized MLP architectures of neural networks. *Int. J. Artif. Intell. Expert Syst.* **1**, 111–122 (2011).
13. Khalesi, H. et al. Application of Hybrid SWARA–BIM in Reducing Reworks of Building Construction Projects from the Perspective of Time. *Sustainability* **12**, 8927 (2020).
14. Kim, S. A Study on the Characteristic Ways of Extension in Commercial Multi-Complexes Based on the Utilization of Common Spaces. Ph.D. Thesis, Seoul National University, Seoul, Korea, 9–22. (2011).
15. Lee, S. A Study on the Changing Architectural Properties of Mixed-Use Commercial Complexes in Seoul, Korea. *Sustainability* **14**, 2649. https://doi.org/10.3390/su14052649 (2022).
16. Love, P. E. D. & Sing, C. P. Determining the probability distribution of rework costs in construction and engineering projects. *Struct. Infrastruct. Eng.* **9** (11), 1136–1148. https://doi.org/10.1080/15732479.2012.667420 (2013).
17. Love, P. E. D., Matthews, J., Sing, M. C. P., Porter, S. R. & Fang, W. State of science: Why does rework occur in construction? What are its consequences? And what can be done to mitigate its occurrence? Engineering 18 (Nov): 246–258. (2022). https://doi.org/10.1016/j.eng.2022.05.010
18. Love, P. & Edwards, D. Curbing rework in offshore projects: Systemic classification of risks with dialogue and narratives. *Struct. Infrastruct. Eng.* **9** (11), 1118–1135 (2013).

19. Liu, S. et al. Application of artificial neural networks in construction management: Current status and future directions. *Appl. Sci.* **11** (20), 9616. https://doi.org/10.3390/app11209616 (2021).
20. Ma, G., Wu, M., Hao, K. & Shang, S. A DSM-Based CCPM-MPL representation method for project scheduling under rework scenarios. Adv. Civ. Eng. 2021: 1–16. (2021). https://doi.org/10.1155/2021/8878308
21. Mamaghani, O. A. & Noorzai, E. A framework to implement augmented reality based on BIM to improve operation and maintenance of mechanical facilities of commercial complexes. *Facilities* **41** (3/4), 229–247 (2023).
22. Mahamid, I. Impact of rework on material waste in building construction projects. *Int. J. Constr. Manag*. doi.org/10.1080/15623599.2020.1728607 (2020).
23. Mahamid, I. Rework investigation in residential building projects: Cost, factors and effects[J]. *Int. Rev. Appl. Sci. Eng.* **15** (1), 116–122 (2024).
24. Matthews, J., Love, P. E. D., Porter, S. R. & Fang, W. Smart data and business analytics: A theoretical framework for managing rework risks in megaprojects. *Int. J. Inf. Manage.* **65** (Aug), 102495. https://doi.org/10.1016/j.ijinfomgt.2022.102495 (2022).
25. Moon, S., Kim, Y. & Choi, Y. A Study on the Changing Architectural Properties of Void Space in Mixed-Use Commercial Complexes. *J. Korean Inst. Cult. Archit.* **52**, 120–121 (2015).
26. Mostofı, F., Tokdemir, O. B., Toğan, V. & Arditi, D. Predicting the cost of rework in high-rise buildings using graph convolutional networks. *J. Construct Eng. Manag*. **150** (8), 1–17 (2024).
27. Liu, Q., Ye, G., Feng, Y., Wang, C. & Peng, Y. Case-based insights into rework costs of residential building projects in China. *Int. J. Constr. Manag*. https://doi.org/10.1080/15623599.2018. 1484856 (2018).
28. Love, P. E. D. Influence of project type and procurement method on rework costs in building construction projects. *J. Constr. Eng. Manag.* **128** (1), 18–29 (2002).
29. Love, P. E. D., Mandal, P., Smith, J. & Li, H. Modeling the dynamics of design error induced rework in construction projects. *Constr. Manage. Econ.* **18** (5), 567–574 (2000).
30. Love, P. E. D., Smith, J. & Teo. P Putting into practice error management theory: Unlearning and learning to manage action errors in construction. *Appl. Ergon.* **69** (May), 104–111. https://doi.org/10.1016/j.apergo.2018.01.007 (2018).
31. Love, P. E., Smith, J., Ackermann, F. & Irani, Z. Making sense of rework and its unintended consequence in projects: the emergence of uncomfortable knowledge in practice. *Int. J. Project Manage.* **37** (3), 501–516. 10.1016/j.ijproman.2019.02.004 (2019).
32. Love, P. E. D. Influence of project type and procurement method on rework costs in building construction projects. J. Constr. Eng. Manage., 10.1061/(ASCE)0733–9364(2002)128:1(18), 18–29. (2002).
33. Liu, Q., Ye, G., Feng, Y., Wang, C. & Peng, Y. Case-Based Insights into Rework Costs of Residential Building Projects in China. *Int. J. Constr. Manag*. **20**, 347–355 (2020).
34. Oyewobi, L. et al. The effect of project types on the occurrence of rework in an expanding economy. *J. Civil Eng. Constr. Technol.* **2** (6), 119–124 (2011).
35. Palaneeswaran, E., Kumaraswamy, M., Ng, T. & Love, P. Management of rework in Hong Kong construction projects. Proceedings of Queensland University of Technology Research Week International Conference, July 4–5, 2005. (2005).
36. Palaneeswaran, E., Love, P. E. D., Kumaraswamy, M. M. & Ng, T. S. T. Mapping rework causes and effects using artificial neural networks. *Building Res. Inform.* **36** (5), 450–465 (2008).
37. Rogge, D. F., Cogliser, C., Alaman, H. & McCormack, S. *An investigation of field rework in industrial construction. Research Report 153 – 11* (Construction Industry Institute, University of Texas at Austin, 2001).
38. Safapour, E. & Kermanshachi, S. Identifying early indicators of manageable rework causes and selecting mitigating best practices for construction. *J. Manag Eng.* **35**, 04018060 (2019).

39. Shoar, S., Chileshe, N. & Edwards, J. D. Machine learning-aided engineering services' cost overruns prediction in high-rise residential building projects: Application of random forest regression. *J. Build. Eng.* **50** (Nov), 104102. https://doi.org/10.1016/j.jobe.2022.104102 (2022).
40. Simpeh, E. K., Ndihokubwayo, R., Love, P. E. D. & Thwala, W. D. A rework probability model: A quantitative assessment of rework occurrence in construction projects. *Int. J. Constr. Manage.* **15** (2), 109–116. https://doi.org/10.1080/15623599.2015.1033814 (2015).
41. Sommerville, J. Defects and rework in new build: an analysis of the phenomenon and drivers. *Struct. Surv.* **25** (5), 391–407. doi.org/10.1108/02630800710838437 (2007).
42. Shah, K. K., Shah, R. A. & Sharma, N. D. Development of field rework index to minimize the impact of rework for effective construction work. *Int. J. Sci. Dev. Res.* **1** (5), 370–375 (2016).
43. Trach, R., Lendo-Siwicka, M., Pawluk, K. & Połoński. M Analysis of direct rework costs in Ukrainian construction. *Arch. Civ. Eng.* **67** (2), 397–411. https://doi.org/10.24425/ace.2021.137175 (2021).
44. Tijanic, K., Car-Pušić, D. & Šperac, M. Cost Estimation in Road Construction Using Artificial Neural Network. *Neural Comput. Appl.* **32**, 9343–9355 (2020).
45. Wang, H., Liang, Q., Hancock, J. T. & Khoshgoftaar, T. M. Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. *J. Big Data*. **11** (1), 44 (2024).
46. Wasfy, M. Severity and Impact of Rework: A Case Study of a Residential Commercial Tower Peoject in the Eastern Province KSA. Master Thesis, KFUPM, Dhahran, KSA. (2010).
47. Wu, C. L., Chau, K. W. & Li, Y. S. Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. *Water Resour. Res.* **45**, 8 (2009).
48. Ye, G., Jin, Z., Xia, B. & Skitmore, M. Analyzing causes for reworks in construction projects in China. *J. Manag Eng.* **31** (6). https://doi.org/10.1061/(ASCE)ME.1943-5479.0000347 (2015).

Figures

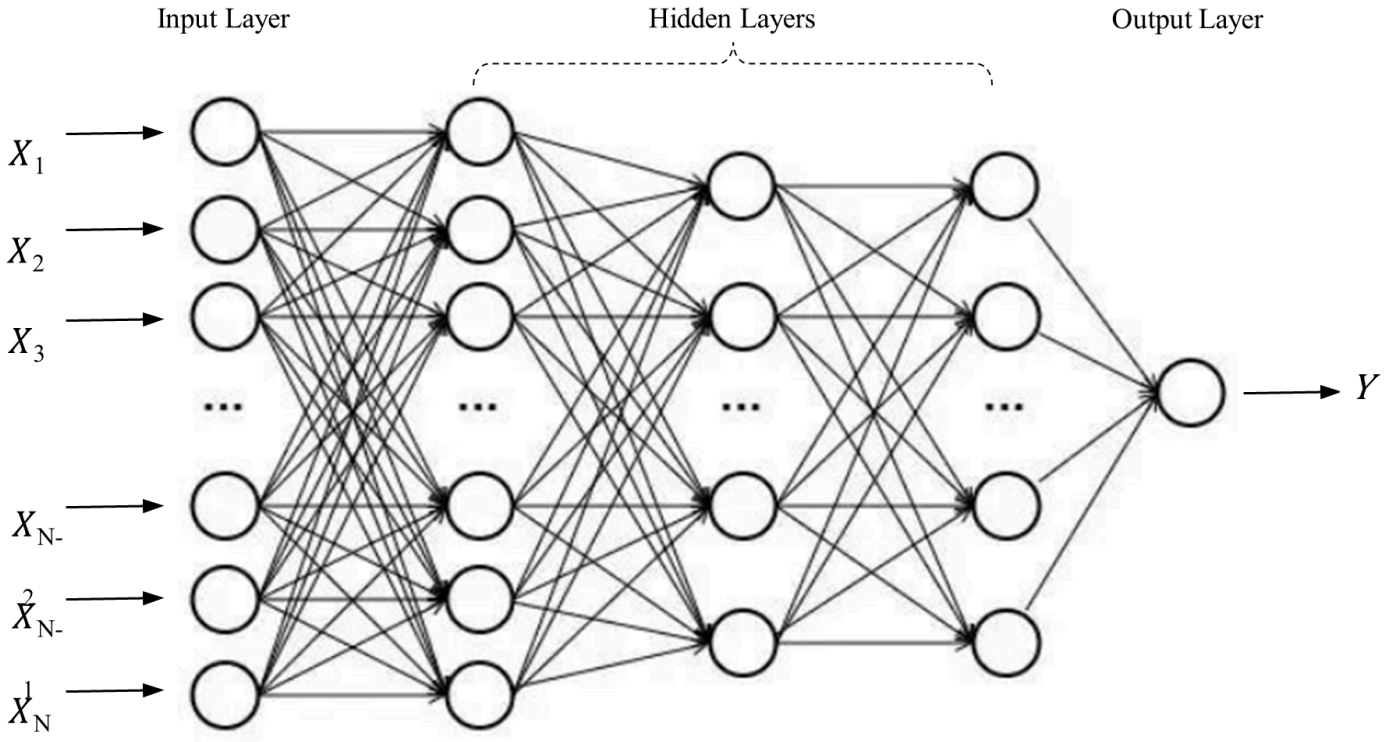


Figure 1

Description of DLNN in the form of MLP

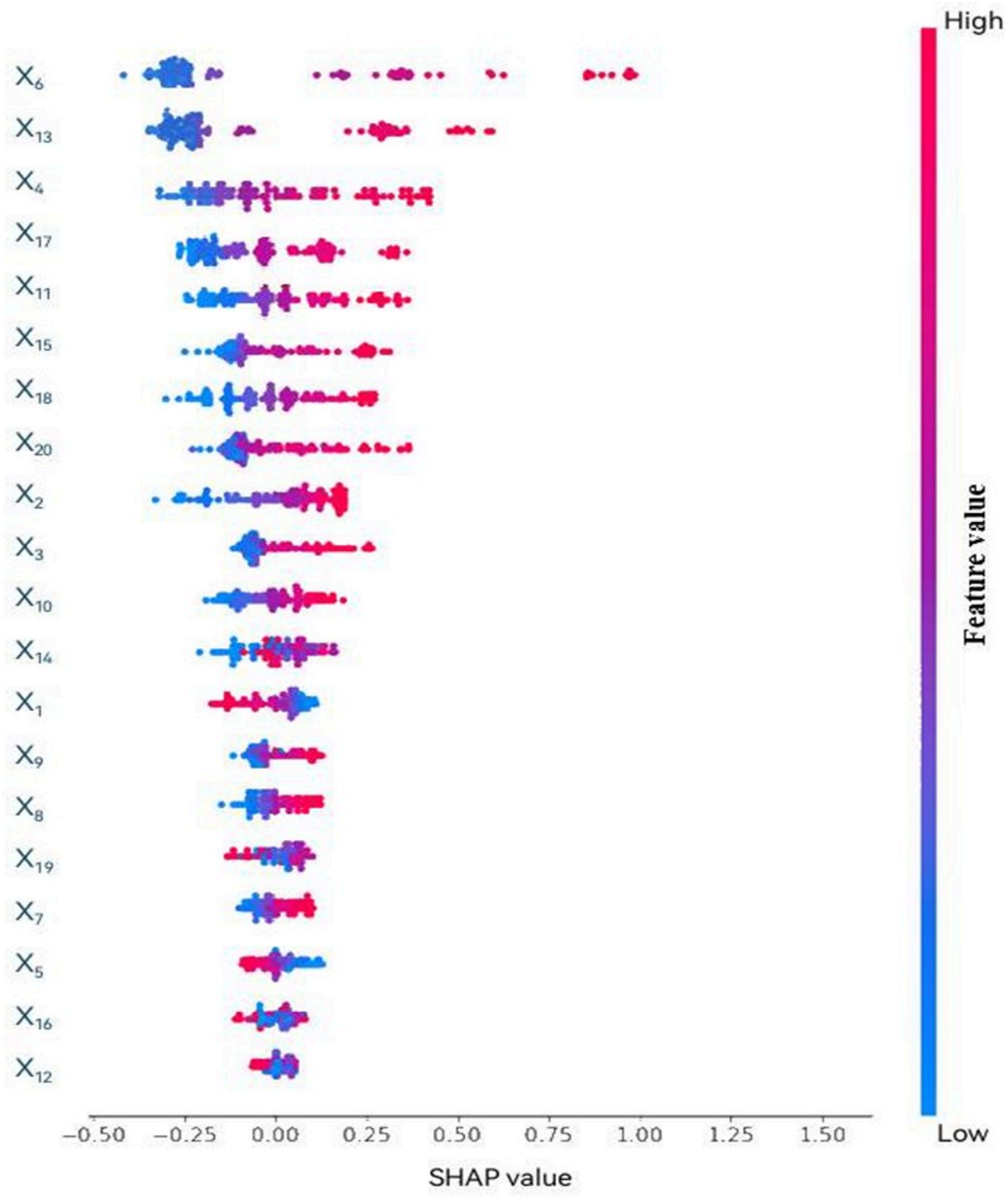


Figure 2

Impact Values of Factors on Rework Costs