Structural Durability & Health Monitoring

Tech Science Press

https://doi.org/10.32604/sdhm.2026.081046

ARTICLE

Combining Principal Component Analysis and Multilayer Perceptron to Establish a Construction Quality Prediction Model

Ching-Lung Fan*

Department of Civil Engineering, Republic of China Military Academy, Kaohsiung, Taiwan

*Corresponding Author: Ching-Lung Fan. Email: p93228001@ntu.edu.tw

Received: 22 February 2026; Accepted: 03 June 2026; Published: 24 August 2026

ABSTRACT: Quality attainment in public construction projects is paramount for effective project management. This study employs Principal Component Analysis (PCA) in the initial phase to discern noncollinear critical defects from inspections across 1015 projects. The identified components are categorized into three aspects: Inspection Records and Occupational Safety and Health (Aspect I), Concrete Quality (Aspect II), and Construction Team Quality Management (Aspect III). Subsequently, a Multilayer Perceptron (MLP) network, trained on 13 PCA-identified critical defects, transforms input data into a probability, indicating project quality. The MLP model exhibits exceptional performance with 91.3% accuracy, 90.2% precision, and 96.1% recall. This hybrid machine learning approach effectively extracts and correlates defect-related information with construction quality, providing a robust tool for accurate quality prediction in construction projects.

KEYWORDS: Principal component analysis; multilayer perceptron; construction quality; defects

1 Introduction

The construction industry is essential to the development and maintenance of civil infrastructure, and billions of dollars are invested in construction projects annually [1]. In 2022, the Taiwanese government allocated NT$392.7 billion to public construction projects, accounting for more than half of the construction industry's total output value and 2.1% of the country's gross domestic product. Such investment highlights the crucial role of public construction in national development and public welfare. Major government-led public construction projects are not only important indicators of governance capacity but also directly or indirectly affect people's daily lives. Timely completion, meeting planned objectives, and ensuring construction quality are all critical to their overall success.

However, despite substantial public investment and the strategic importance of these projects, persistent deficiencies in construction quality management remain a serious concern in Taiwan. Based on the authors' statistical analysis of available construction inspection records released by the Public Construction Commission (PCC) for the period 2020–2024, more than 70% of inspected projects had incomplete quality control records, and nearly 60% lacked properly maintained construction logs. The analysis also showed that concrete-related defects were common, with improper pouring and compaction accounting for 26.1% of the recorded defects and improper curing accounting for 18.0%. These findings indicate that deficiencies in inspection implementation and construction process control are not isolated cases but recurring, systemic problems in engineering practice. In 2021, a concrete-pouring collapse on the MRT San-Ying Line resulted in three fatalities and two injuries, and the accident was associated with improper construction procedures and missing inspection records. In 2022, the ceiling collapse at the Bade Civil Sports Center, which was triggered by a mild earthquake, was attributed to deficiencies in quality management. More recently, a 2024 post-earthquake investigation of the New Taipei MRT Circular Line revealed that the contractor had deviated from the approved design drawings, resulting in misalignment of the steel box girder and track deformation. Taken together, these facts suggest the need for effective data-driven quality assessment methods.

Experts and scholars have established various indicators or criteria for assessing project quality requirements. Institutional organizations have also adopted construction standards and have conducted reasonable and comprehensive evaluations of contractor construction quality from the perspective of quality management [2]. For example, Lee et al. [3] analyzed the functionality and occurrence of defects in bridge structures and road pavement materials. On the basis of their analysis and relevant standards, they developed a quality performance index for assessing the quality of a project. Jogdand and Deshmukh [4] developed a building construction quality index by using the analytic hierarchy process to calculate the weights of construction quality attributes. The index could be used to evaluate project quality and served as a benchmark for future quality improvement initiatives. Kim et al. [5] combined the balanced scorecard framework with the Delphi method to determine 44 indicators. Through importance–performance analysis, they prioritized the indicators by importance with the goal of improving the quality of facility maintenance. Although project managers primarily focus on costs, target timelines, and the successful completion of construction processes, project success is also strongly related to project quality [6].

In this study, inspection data from the Public Construction Management Information System (PCMIS) were used as variables to predict the construction quality of projects on the basis of various project attributes. The PCMIS lists a total of 499 item defects. These defects may be correlated; hence, some information may be redundant. Weighting these defects would result in imbalanced weights for defects representing the same aspect or that are otherwise related, resulting in unreliable evaluation outcomes. Moreover, some defects might have negligible effects on quality. In quality assessments, using human judgment to intuitively select numerous defects can result in these defects being insufficiently representative and thus result in a biased evaluation. Data analysis involving high-dimensional quality records often encounters multicollinearity, leading to biased or unstable model coefficients. To mitigate this impact and enhance interpretability, Principal Component Analysis (PCA) is employed in this study. PCA transforms the original correlated variables into a concise set of uncorrelated, linearly independent principal components, thereby capturing maximum variance while successfully filtering out redundant information and noise contained in the dataset. PCA can help identify the most important features in the decision-making process, reduce the range of possibilities, and eliminate noise contained in redundant datasets [7].

PCA and factor analysis (FA) are both machine learning techniques for dimensionality reduction that can be classified as multivariate simplification strategies. The objective of both methods is to decompose the original variables to extract latent categories by aggregating strongly correlated variables in a single category, producing categories that are not strongly correlated. In the two methods, new indicators or variables are synthesized from linear combinations of the original variables. PCA can therefore be used when conducting a comprehensive assessment with multiple indicators to consolidate the information in these indicators into several principal components. These components can then be weighted and summed to obtain a composite evaluation index. Researchers have used PCA and FA for evaluating performance in engineering tasks. For example, Lin et al. [8] employed PCA to assess key performance indicators in value management during construction. They identified three principal components and their interrelationships and selected 18 key performance indicators from a pool of 47 potential indicators. Similarly, Chen et al. [9] employed FA to evaluate engineering value and extracted 23 performance indicators from 32 initial indicators. These indicators were categorized into four groups: satisfaction, team composition and capability, job plan, and team member participation. PCA is known for its ability to reduce dimensionality, compress data, and reduce noise. In machine learning, it is often used in combination with other algorithms to enhance classification and regression performance. For example, researchers have integrated PCA with cluster analysis to evaluate the performance of asphalt as it ages [10], monitor structural health [11], and investigate fatal construction accidents [12]. Additionally, PCA has been combined with neural networks to detect and identify damage [13–15] and construction project quality [16].

Artificial neural networks (ANNs) are mathematical models that mimic the structure of the human brain and perform distributed parallel calculations [17]. The fundamental unit of an ANN is the perceptron, which mimics a neuron, the basic building block of the human brain. Perceptron networks can be categorized as single-layer perceptron (SLP) and multilayer perceptron (MLP) networks on the basis of their architecture. An MLP network primarily comprises artificial processing units (also known as artificial neurons or cells), an input layer, one or more hidden layers, and an output layer. The output of each processing unit serves as the input for other units in the network. The MLP network is a well-known neural network architecture that is commonly used for tasks such as pattern recognition, classification, and regression [18].

ANNs have been widely applied in studies related to the condition assessment of various infrastructure systems, such as detecting internal cracks in pipelines [19], predicting floor response spectra [20], detecting bridge damage [21], identifying corrosion in steel structures [22], estimating the load-carrying capacity of reinforced concrete bridge piers [23], predicting the fatigue life of asphalt concrete pavements [24], identifying damage in plate structures [25], predicting pavement performance [26,27], and evaluating seismic damage in tunnels [28]. The present study focuses on predicting construction quality from inspection records and project attributes, rather than on deterioration mechanisms, continuous sensing, or structural health monitoring. The cited SHM studies are referenced only to demonstrate the general applicability of PCA and neural networks for processing high-dimensional correlated data.

In other construction-related applications, ANN has also been employed to predict construction quality and schedule [29], analyze construction contract risks [30], predict construction crew productivity [31], estimate construction project duration [32], and predict project costs [33]. However, ANNs have limitations; their effectiveness may be limited for problems involving numerous variables. Constructing an ANN model may require more computation time than constructing other machine learning networks, resulting in delays. To address this issue, the network can be optimized by reducing the number of variables and processing elements [34] and by incorporating other algorithms to enhance the model's performance [35,36]. Other challenges encountered in the implementation of ANNs include determining appropriate parameter settings, interpreting black-box models, and avoiding becoming trapped in local optima.

Although previous studies have demonstrated the applicability of PCA and ANNs in a wide range of engineering problems, research specifically focused on construction quality prediction remains relatively limited. Existing studies have mainly applied PCA to dimensionality reduction, anomaly detection, or performance evaluation, whereas ANN-based models have been widely used for prediction and classification tasks in infrastructure, cost estimation, safety assessment, and condition evaluation. However, these studies generally exhibit three limitations when extended to construction quality prediction. First, feature selection is often insufficiently targeted, particularly when a large number of quality-related variables or defect items are involved. As a result, redundant or highly correlated variables may be directly incorporated into the prediction model, reducing analytical efficiency and obscuring the truly influential factors. Second, existing prediction frameworks rarely integrate specific process-related defects encountered in civil engineering practice into a unified analytical structure. Third, although ANN-based models provide strong predictive capability, their practical usefulness is often constrained by limited interpretability, making it difficult to explain which specific construction defects are most closely associated with quality outcomes.

This study contributes to the literature in three aspects. First, it introduces a more targeted feature selection mechanism for construction quality prediction by extracting representative defect information from a large set of inspection variables. Second, it explicitly incorporates specific civil engineering construction defects into the prediction framework, thereby strengthening the practical connection between statistical modeling and on-site quality management. Third, it improves the interpretability of ANN-based quality prediction by organizing the identified critical defects into three aspects that are readily understandable and applicable to project owners, supervisors, and contractors. Therefore, this study lies in strengthening proactive quality management during construction, thereby reducing the risk of future structural deficiencies. Accordingly, its contribution is more appropriately positioned as a construction quality assessment, while the early identification of critical defects may also indirectly support durability control and long-term infrastructure health performance.

This paper substantially expands upon the earlier study [16] by providing a more rigorous methodological framework, clearer statistical justification for PCA-based defect selection, and a more comprehensive model-validation procedure. Although the underlying dataset and baseline accuracy align with those of the initial proof-of-concept, the current study introduces critical incremental scientific advances that transform the earlier exploratory work into a theoretically and practically meaningful contribution. Specifically, rather than treating the data as a static predictive target, this manuscript develops a systematic K-means-based quality-risk grouping approach to mathematically address severe class imbalance, provides a thorough justification for PCA component retention, incorporates post-hoc SHAP interpretability to decode the "black-box" neural network, and performs an in-depth misclassification analysis to identify systematic error patterns. These methodological expansions bridge the gap between simple machine learning implementation and actionable engineering knowledge, providing an explainable and rigorous quality assessment framework that was entirely absent in the preliminary study. Furthermore, the current manuscript incorporates additional performance metrics, baseline model comparisons, calibration analysis, feature-importance analysis, misclassification analysis, and a more detailed discussion of practical implications and future research directions.

2 Methodology

By performing a numerical weighted average, PCA can be used to identify the maximum variance of variables to transform strongly correlated variables into a smaller set of new independent variables, known as principal components, which explain most of the variance. These selected principal components can be used to interpret the original data. Many defects (variables) can affect construction quality, but their contributions vary. Hence, replacing the entire space of defect samples with a few mutually independent matrices of their linear combinations is not only efficient but also enables analysis of the correlations between each defect and construction quality. The first principal component explains the most variance across the entire space of defect samples (499 items). The second principal component must be orthogonal to the first principal component, and it explains most of the remaining variance. Subsequent principal components are selected similarly.

The PCMIS construction inspection data used in this study comprises data for a total of 1015 cases. The dataset includes information on defects, scores, grades, contract sums, project categories, and construction progress. The objective of the present study was to identify critical defects hidden in the PCMIS database and to develop a model capable of accurately predicting construction quality. Furthermore, the developed model was used to analyze the relationship between project attributes and construction quality. The objective of this analysis was to provide insights regarding defect prevention for project organizers, supervisors, and contractors. The theoretical and methodological approaches employed in this study included PCA, cluster analysis, and the MLP. The procedures and mathematical formulas involved are explained in the following subsections.

2.1 Principal Component Analysis

In PCA, a few noncorrelated principal components are created as linear combinations of the original correlated variables. For m original variables, PCA can produce up to m principal components. Typically, however, only n principal components ($n << m$) that explain most of the variance are selected, simplifying the model without losing information. Because these principal components are independent, the resulting model is less likely to contain bias. The linear combinations of the original variables for producing the principal components are as follows:

$$\begin{aligned} P_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1m}X_m \\ P_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2m}X_m \\ &\vdots \\ P_n &= a_{n1}X_1 + a_{n2}X_2 + \cdots + a_{nm}X_m \end{aligned} \tag{1}$$

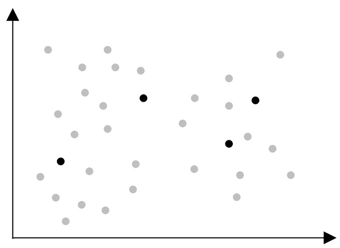
In the equation, $P_1, P_2, \ldots, P_n$ represent the n principal components and a_{ij} represents the weight of the jth variable for the ith principal component. Moreover, $a_{i1}^2 + a_{i2}^2 + \cdots + a_{im}^2 = 1$, and $a_{i1}a_{j1} + a_{i2}a_{j2} + \cdots + a_{im}a_{jm} = 0$. These conditions ensure that the first principal component, P1, explains the largest percentage of the variance in the original data. The second principal component P_2, which is independent of P_1, has the second-largest variance and explains most of the remaining variability that is not explained by the first principal component. This pattern continues for the subsequent principal components. Two methods can be employed to perform PCA: one using the variance–covariance matrix S, and the other utilizing the correlation matrix R. Assuming that the variables $X_1, \ldots, X_m$ have a covariance matrix S, the variance of the first principal component and the coefficients of its linear combination correspond to the largest eigenvalue λ_1 of S and its associated eigenvector. The covariance matrix represents the degrees of correlation between variables and is obtained by directly computing the covariance between each pair of variables. Larger covariance indicates a stronger correlation between the variables.

2.2 Cluster Analysis

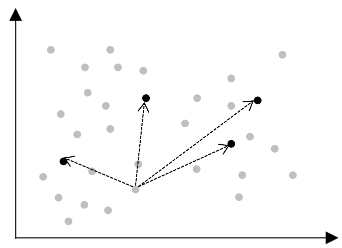
Cluster analysis is a technique for partitioning data into several clusters, with high homogeneity within each cluster and substantial differences between clusters. Clustering can be performed to establish groupings and identify common characteristics within groups. Cluster analysis can reduce the distance between variables in a dataset and emphasize the similarities within each cluster, thereby facilitating decision-making. The most commonly used algorithm for cluster analysis is K-means. In this algorithm, data are partitioned into k clusters by first randomly selecting k seed points as the initial centers for these clusters. Each remaining sample p is assigned to its nearest cluster, and the centers of each cluster are then recalculated. The distances between each sample and the new cluster centers are compared. Samples are reassigned, and cluster centers are recalculated iteratively until the sum of squared errors (SSE) reaches its minimum value. The formula used in the K-means algorithm is as follows:

$$\mathrm{SSE} = \sum_{i=1}^{k} \sum_{p \in c_i} |p - m_i|^2 \tag{2}$$

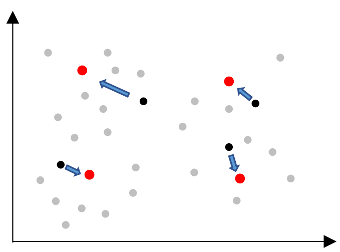
In the equation, k represents the number of clusters, p denotes a sample in the space, and m_i represents the average value of the samples in cluster c_i. The SSE is the sum of the squared distances between each sample and the center of its cluster. The steps of the algorithm are as follows (Fig. 1): (a) The initial centroids for the k clusters are randomly selected. (b) The distances between each sample p and the centroids of the clusters are calculated. A shorter distance indicates that p is more similar to the data in that cluster. Subsequently, p is assigned to the nearest cluster. (c) The average values of the centroids for each cluster are recalculated after assignment, and the previous seed centroids are replaced. (d) Steps (b) and (c) are repeated iteratively by reassigning each data point to a cluster and updating the centroids. This process is continued until the SSE is minimized. When the boundaries of the clusters no longer change, and the specified criteria are satisfied, the clustering process is complete.



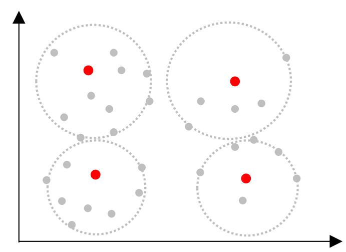
(**a**) Randomly select *k* seed points.



(**b**) Calculate the distances between each *p* and the nearest *k* seeds.



(**c**) Assign a new centroid for each cluster and replace the previous seeds.



(**d**) Repeat these calculations until the SSE is minimized.

Figure 1: Four steps of the K-means clustering algorithm.

In this study, to ensure that samples with similar inspection grades were handled consistently and to prevent large differences in sample sizes across inspection grades, the frequency of the 18,246 defects and the scores of the 1015 projects were subjected to cluster analysis using the K-means algorithm. The data were divided into four groups according to the original grades A, B, C, and D. From best to worst quality, these groups were Level A, Level B, Level C, and Level D (Table 1).

Table 1: Four levels of construction quality and percentage of the dataset.

Quality	**Good**		**Poor**	
Clustering group	Level A	Level B	Level C	Level D
Project number	455 (44.8%)	157 (15.5%)	309 (30.4%)	94 (9.3%)
Inspection grade	A	B	C	D
Score	90–100	80–89	70–79	<70
Project number	6 (0.6%)	780 (76.8%)	227 (22.4%)	2 (0.2%)

In cluster analysis, the choice of K directly affects the rationality and interpretability of the clustering results. In this study, the value of K was quantitatively evaluated using a combination of the Elbow Method and the Silhouette Coefficient. Based on the defect frequency and inspection score data from 1015 projects, the K-means algorithm was performed for K = 2 to 8 (with the random seed fixed at 42 and repeated 10 times, using the average results), and the SSE and the average silhouette coefficient were calculated for each value of K. According to the Elbow Method, the SSE curve showed a clear elbow at K = 4, indicating that the rate of SSE reduction decreased substantially beyond this point. Specifically, when K increased from 3 to 4, the SSE decreased by 15.2%, whereas increasing K from 4 to 5 decreased the SSE by only 3.8% (Fig. 2). This finding suggests that K = 4 represents an appropriate trade-off between model simplicity and explanatory power.

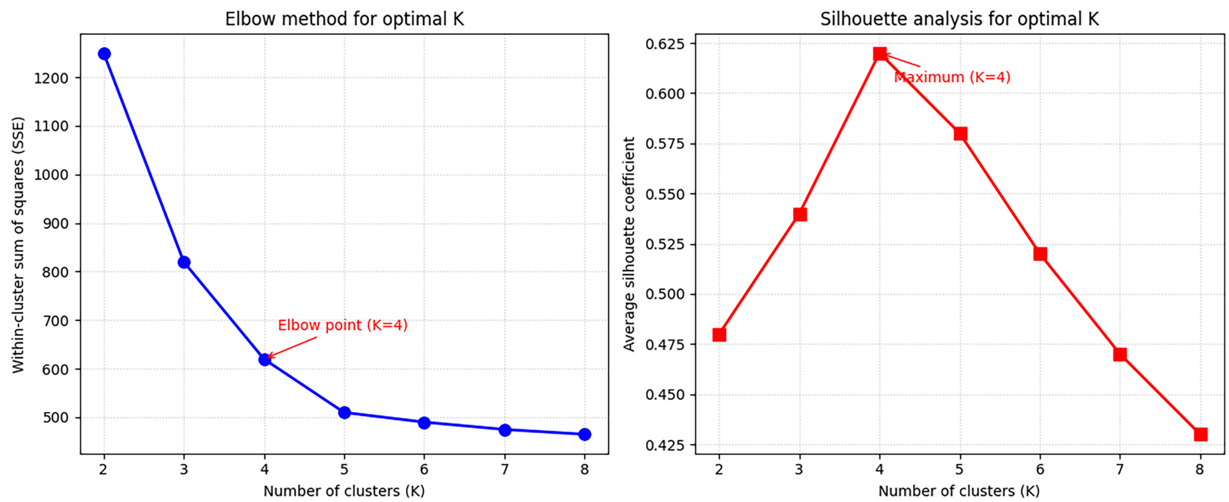


Figure 2: Elbow and Silhouette analyses indicating that K = 4 provides a stable and interpretable grouping of construction quality risk for subsequent prediction and inspection management.

The Silhouette Coefficient, which measures the balance between within-cluster compactness and between-cluster separation, ranges from −1 to 1. In this study, the average silhouette coefficient reached 0.62 at K = 4 (Fig. 2), which was markedly higher than that for K = 3 (0.54) and K = 5 (0.58). In addition, all samples had silhouette coefficients greater than 0.3, indicating a well-defined clustering structure. Beyond the statistical criteria, the reasonableness of K = 4 was further validated through interpretation in the context of construction quality inspection practice. Specifically, the four cluster centers and their distributions across key quality indicators exhibited distinct inspection-related patterns, which correspond well to the four-level quality categorization commonly used in engineering practice. Therefore, both the statistical evidence and the engineering interpretation support selecting K = 4 as the final number of clusters.

The practical role of K-means clustering in this study extends beyond data preprocessing. It also provides a data management mechanism for reorganizing highly imbalanced inspection-grade data into more stable, interpretable quality groups. In the original inspection system, most projects were concentrated in grades B and C, whereas grades A and D contained very few cases. Such an uneven distribution reduces the stability of supervised learning and limits the practical usefulness of the original grade labels for inspection management. By clustering projects based on both inspection scores and defect frequency, K-means produces quality groups that reflect not only the final inspection score but also each project's observed defect burden. This grouping enables projects with similar quality-risk characteristics to be analyzed and managed together.

From an inspection management perspective, K-means-based grouping can transform raw inspection records into actionable quality-risk categories. Projects classified into lower-quality clusters can be prioritized for follow-up inspection, enhanced supervision, or corrective actions. Moreover, the clustering results help inspection agencies identify groups of projects with similar defect patterns, thereby supporting risk-based inspection planning rather than relying solely on the original A–D inspection grades. Therefore, K-means improves the practical usability of the inspection database by stabilizing the prediction target, enhancing interpretability, and supporting more systematic inspection-resource allocation.

2.3 Multilayer Perceptron

MLPs consist of an input layer, one or more hidden layers, and an output layer. The input layer receives the predictor variables, while the hidden layers learn nonlinear relationships between the inputs and the target output. The number of hidden layers and neurons can be adjusted according to data complexity; however, overly complex architectures may increase computational cost and the risk of overfitting, whereas insufficient neurons may limit the model's learning capacity. The output layer then converts the learned representations into the final prediction, often expressed as a class probability.

During feedforward computation, each neuron calculates a weighted sum of its inputs and bias, followed by a nonlinear activation function such as sigmoid, tanh, or ReLU. This process propagates information layer by layer until the output is obtained. For each neuron, the weighted input is computed using the inputs x_i, weights w_i, and bias b_i, as expressed in Eq. (3). The parameters of the activation function are the weights connecting each input-layer neuron to each hidden-layer neuron [37]. The activation function determines whether the result of the summation of impulses, also known as the weighted sum of the inputs, is sufficient to generate an output [17]. Two commonly used activation functions in MLP are the sigmoid function and rectified linear unit (ReLU), as described in Eqs. (4) and (5).

$$y_i = \sum_i^n w_i \cdot x_i + b_i \tag{3}$$

$$\text{Sigmoid} = \frac{1}{1 + e^{-x}} \tag{4}$$

$$\text{ReLU} = \max(0, x)$$
$$\text{ReLU} = \begin{cases} 0 & \text{for} \quad x < 0 \\ x & \text{for} \quad x \geq 0 \end{cases} \tag{5}$$

Once the MLP learning model has been established, backpropagation is used to minimize the error; it is often combined with stochastic gradient descent (SGD). In recent deep learning networks, SGD has been shown to be highly effective for optimization [38]. Among first-order gradient-based optimization algorithms, the adaptive moment estimation (Adam) optimizer is widely used. Adam facilitates gradient convergence in the correct direction by computing first- and second-moment estimates of the gradients [39]. Backpropagation is a supervised learning method that requires both features and labeled data (ground truth values) as inputs. The features are fed into the neural network and processed through the hidden layers until the output layer produces a result. The error between the output and actual labels is then calculated using a loss function. Finally, the neural network's weights and biases are updated to minimize the loss function. A commonly used loss function, the cross-entropy loss (Eq. (6)), measures the overall error between the predicted and true probability distributions. The loss function evaluates the model's fit to the samples; a smaller value indicates greater similarity between the predicted and actual values and therefore a better fit.

In Eq. (6), t_i and y_i represent the true label and the probability of the predicted label, respectively, whereas n is the total number of data points. The optimizer is a numerical method for iteratively updating the weights and biases during batch training until the combination that minimizes the loss is found.

$$\text{Cross-Entropy} = -\sum_i^n t_i * \ln(y_i) \tag{6}$$

3 Research Data and Framework

Haponava and Al-Jibouri [40] noted that, in most relevant studies, performance was measured only after project completion rather than during the construction phase. The Public Construction Inspection Mechanism was established in Taiwan in 1993 to improve the quality management of public construction projects. Under this mechanism, government agencies organize construction surveillance units composed of experts and scholars to inspect randomly selected projects. Each project is typically examined by three inspection committee members, who review relevant documents, conduct on-site inspections, and assess project quality based on construction inspection criteria, applicable regulations, and contractual requirements. After the inspection, a meeting is held to consolidate the identified defects, scores, and recommendations. The final inspection grade is determined by averaging the scores assigned by the committee members and is classified into four categories: Grade A (90–100 points), Grade B (80–89 points), Grade C (70–79 points), and Grade D (<70 points). Four types of defects have been identified: Management (M), covering 113 defects; Quality (Q), covering 356 defects; Schedule (S), covering 10 defects; and Design (D), covering 20 defects. The total number of defects is 499. After the inspection is completed, the organizing agency must upload the relevant inspection data for that day to the PCMIS.

In this study, a dataset of 1015 inspection projects obtained from the PCMIS was analyzed. The data included information on defects, scores, inspection grades, contract sums, project categories, and construction progress. The researchers analyzed these variables to examine the data and draw conclusions. The statistics of the project attribute data are as follows: (1) Defect types: The frequency of management defects was 8282, quality defects was 9836, schedule defects was 78, and design defects was 50. (2) Scores: 6 projects had scores above 90, 780 projects had scores between 80 and 89, 227 projects had scores between 70 and 79, and 2 projects had scores below 70 (Fig. 3). The total sample size was 1015 projects (population size). (3) Contract sums were categorized into three groups: NT$100–1000 million (372 projects), NT$1000–5000 million (282 projects), and above NT$5000 million (361 projects). (4) The seven project categories were airport engineering (AE), civil engineering (CE), harbor engineering (HE), new construction engineering (NE), other engineering (OE), power and air conditioning engineering (PE), and renovation engineering (RE). The proportions of AE, CE, HE, NE, OE, PE, and RE projects, respectively, were 3.2% (32 projects), 8.8% (89 projects), 0.9% (9 projects), 11.6% (118 projects), 30.1% (306 projects), 17.4% (177 projects), and 28.0% (284 projects; Fig. 4). (5) Construction progress was calculated as the percentage of completed tasks and quantities. The average progress was 46.78% with a standard deviation of 27.4. The maximum progress was 100%, and the minimum progress was 0%.

This study aimed to predict the construction quality of these 1015 projects. Cluster analysis was conducted on two variables extracted from the quality-related inspection data: defect frequency and scores. The purpose of the cluster analysis was to group projects with similar characteristics together. Hence, the original four inspection grades were regrouped with four clusters: Level A, Level B, Level C, and Level D. Projects in Level A and Level B were grouped as high-quality projects (612 projects), whereas those in Level C and Level D were grouped as low-quality projects (403 projects). The goal of the prediction was to sort the projects into these two categories. Further analysis was also conducted to identify specific defects affecting construction quality and to examine the correlation between quality and project attributes. Such an investigation could provide valuable information for implementing effective construction management practices.

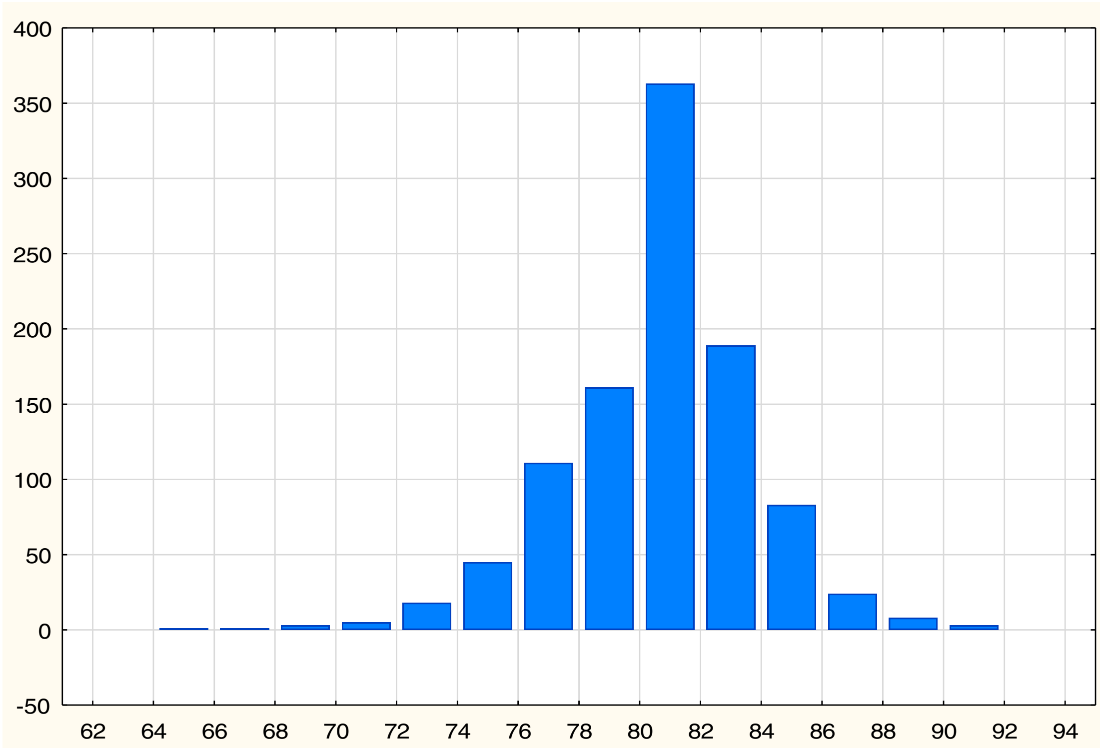


Figure 3: Histogram of project numbers by inspection score.

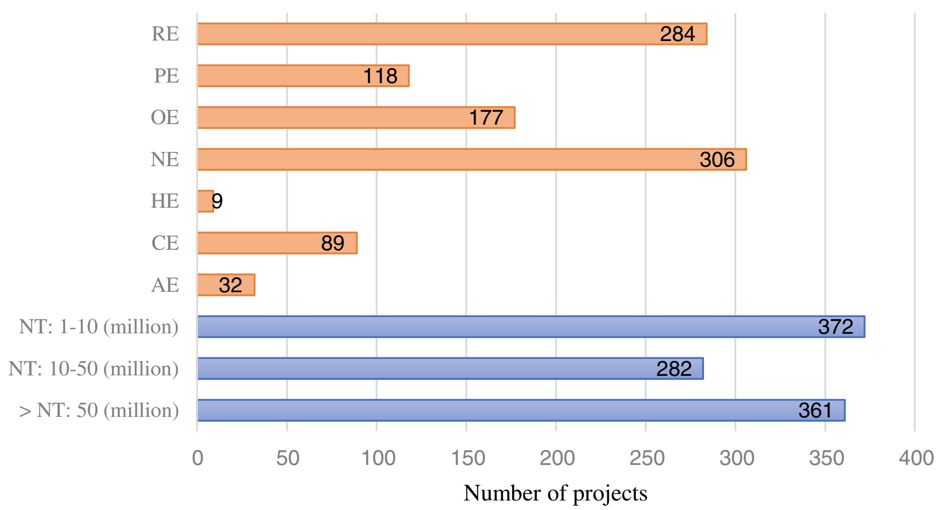


Figure 4: Statistics of contract sums and project categories.

Although the PCMIS inspection system provides official grades A–D and numerical scores, the original grade distribution was highly imbalanced, with only 6 grade-A projects and 2 grade-D projects among the 1015 inspected projects. In contrast, most projects were concentrated in B and C. Using the original four-grade system directly as the prediction target would therefore create an extremely sparse and unstable multi-class learning problem. To reduce this imbalance and define project quality based on both inspection scores and defect burden, K-means clustering was applied to defect frequency and inspection scores. The clustering results produced four groups with a more balanced distribution, which were subsequently consolidated into two broader categories, namely high-quality (Good) and low-quality (Poor) projects, for the final classification task. Accordingly, the role of cluster analysis in this study was not to replace the official inspection grading system, but to provide a more stable and practically interpretable target structure for predictive modeling.

The main research framework and process of this study are presented in Fig. 5 and described as follows:

- PCA: Principal components were retained based on the scree plot and cumulative explained variance, and critical defects related to construction quality were selected using an absolute component loading threshold of 0.4. These retained defect were then used as input variables for the construction quality prediction model.
- Cluster Analysis: Cluster analysis was performed on the defect frequency and scores to group projects with similar characteristics into the same cluster. Clustering reduces the sample size of projects with different grades and provides output variables for the construction quality prediction model.
- MLP: The MLP was used as a nonlinear binary classifier to distinguish high-quality from low-quality projects. The data were divided into an independent stratified test set (20%) and a development set (80%) for 5-fold stratified cross-validation and grid-search-based hyperparameter tuning. The final model used 22 input features and adopted a two-hidden-layer architecture (40, 30) with ReLU hidden activations, a sigmoid output, the Adam optimizer, a learning rate of 0.001, a batch size of 32, and 30 training epochs. Model robustness was further examined across multiple random seeds.
- Model evaluation: The model for predicting construction quality was assessed using a confusion matrix to calculate its accuracy, precision, and recall. This evaluation indicated the effectiveness of the model constructed using PCA and the MLP. Additionally, the factors contributing to model misclassifications were explored, and the relationships between project attributes (input variables) and quality were analyzed. These findings provide valuable insights for project managers of construction teams attempting to effectively enhance the quality of public-sector construction.

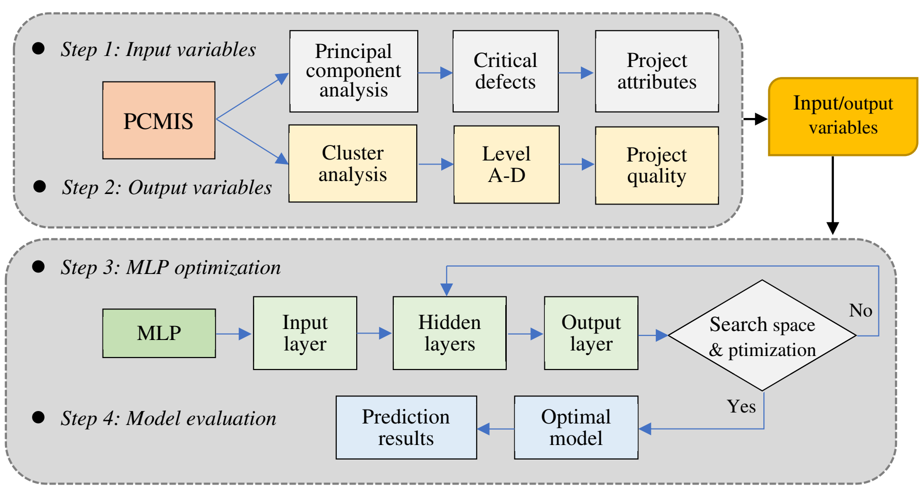


Figure 5: Research framework of the proposed PCA–MLP construction quality prediction model.

4 Results and Discussion

In this study, PCA was applied to the 499 defect variables in the PCMIS database to reduce redundancy and multicollinearity before model construction. The defect variables were binary indicators, with 1 indicating the presence of a defect and 0 indicating its absence. Because these variables were already measured on the same 0/1 scale, they were not standardized prior to PCA. The covariance matrix was used in the original analysis to preserve the variance structure of the defect-occurrence data in their original coding.

The PCA results were used to identify the most representative defects affecting construction quality. To improve transparency in the dimensionality reduction process, principal component retention was determined based on the scree plot and cumulative explained variance rather than a single numerical threshold. Fig. 6 presents the scree plot of the principal components extracted from the defect dataset. The plot shows that the eigenvalues declined rapidly in the early components and then gradually leveled off, indicating that the majority of the variance in the original dataset was concentrated in a limited number of components. Based on the scree plot and cumulative explained variance, 10 principal components were retained, and the retained component structure explained 84.223% of the total variance, suggesting that most of the information contained in the original 499 defect variables was preserved after dimensionality reduction. After component retention, defect variables with absolute component loadings exceeding 0.4 were selected, yielding 13 critical defects for subsequent analysis.

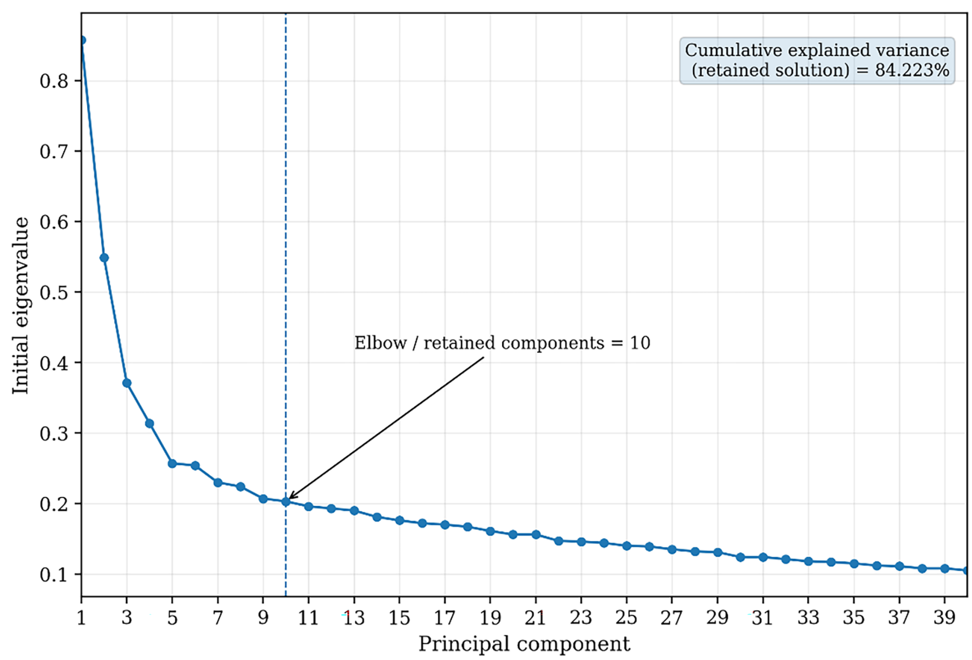


Figure 6: Scree plot showing the retention of 10 principal components, preserving 84.223% of defect-information variance for critical defect selection.

4.1 Critical Defects

After the principal components were retained, defect variables were screened based on their associations with the retained components. In this study, an absolute component loading threshold of 0.4 was adopted to identify critical defects with meaningful contributions to the retained components. This threshold was used for variable selection, rather than for principal component retention. Based on this procedure, 13 critical defects were identified as input variables for the comprehensive evaluation of construction quality. These defects were further interpreted as three aspects: Inspection Records and Occupational Safety and Health (Aspect I), Concrete Quality (Aspect II), and Construction Team Quality Management (Aspect III). Table 2 reports the component loadings and extracted communalities of these 13 critical defects.

Table 2: Component loadings and extracted communalities of the 13 critical defects identified by PCA.

Aspect	Defect Code	Critical Defect	Component Loadings			Extracted Community
			P_1	P_2	P_3	
Aspect I	M47	Absence of inspections of construction operations, failure to conduct material and equipment inspections, and failure to complete inspection records	0.534	−0.304	0.019	0.998
	M76	Failure to complete quality control self-inspection forms	0.517	−0.247	0.088	0.999
	Q286	Lack of proper fall prevention facilities, such as guardrails, covers, safety nets, and safety harnesses	0.471	0.116	−0.017	0.994
	Q60	Uncleared garbage and waste materials that can affect the environment	0.438	0.088	−0.049	0.991
	M51	Failure to complete manufacturing supervision reports or no record of implementing these reports	0.417	−0.307	0.112	0.994
	M48	Failure to supervise construction contractors in implementing on-site safety and health, traffic maintenance, and environmental protection measures	0.415	−0.088	0.001	0.994
	M75	Failure to complete construction logs, failure to adhere to specified formats, or incomplete records	0.413	−0.372	0.205	0.999
	Q283	Missing records regarding the inspection and review of materials and equipment	0.446	0.007	−0.352	0.999
Aspect II	Q2	Improper pouring and compaction of concrete resulting in the formation of cold joints, honeycombing, or voids	0.266	0.614	0.21	0.997
	Q5	Presence of debris (such as wires, iron pieces, or formwork) on the surface of the concrete	0.311	0.588	0.265	0.997
Aspect III	M61	Other deficiencies related to the organizing agency or project management firm	0.218	0.236	0.56	0.998
	M29	Other defects in supervisory unit quality management	0.368	0.156	0.412	0.998
	M113	Other defects in contractor quality management	0.305	0.227	0.498	0.999

KMO and Bartlett's test were further conducted for the 13 critical defects retained after PCA-based variable screening. The overall KMO value was 0.723, indicating acceptable sampling adequacy. Bartlett's test of sphericity was significant (χ^2 = 1367.25, df = 78, $p < 0.001$), confirming that the correlation structure among the retained defect variables was suitable for subsequent latent-structure interpretation.

The first principal component (P_1) comprised eight defects: M47, M76, Q286, Q60, M51, M48, M75, and Q283. These defects are all related to labor safety and health, or the completion of construction inspection checklists. Therefore, this principal component was named "Inspection Records and Occupational Safety and Health" (Aspect I). Of the variables, an absence of inspection of construction operations, failure to conduct material and equipment inspections, and failure to complete inspection records (M47) were found to significantly affect construction quality (component loading = 0.534). The second principal component (P_2) comprised two defects: Q2 and Q5. Both are related to deficiencies in the quality of concrete construction. This principal component was named "Concrete Quality" (Aspect II). Improper pouring and compaction of concrete, resulting in the formation of cold joints, honeycombing, or voids (Q2), is a deficiency that significantly affects construction quality (component loading = 0.614). The third principal component (P_3) comprised three defects: M61, M29, and M113. These deficiencies are related to the quality management practices of the organizing agency, supervisory unit, and contractors involved in the construction team. This principal component was named "Construction Team Quality Management" (Aspect III). Other deficiencies related to the organizing agency or project management firm (M61) also play a crucial role in determining the construction quality (component loading = 0.56).

The PCA results for the PCMIS data were used to characterize defects and identify defect categories. Some defects were strongly correlated and could be grouped into a single aspect. PCA can facilitate the identification of primary latent factors and unimportant secondary factors, thereby simplifying the data structure. It also provides a comprehensive criterion for variable selection (the largest individual differences in component loadings among the principal components). The 13 critical defects identified through PCA significantly affect construction quality and are influential factors in determining grades, scores, and on-site management performance. Construction teams that must undergo government inspections could reference these findings to better manage defects. In particular, the following three defect categories are described as a reference for owners, supervisors, and contractors:

- Inspection Records and Occupational Safety and Health (Aspect I): The first principal component of critical defects can be divided into two subcategories. The first subcategory pertains to defects related to inspection records, including the following: "Absence of inspections of construction operations, failure to conduct material and equipment inspections, and failure to complete inspection records (M47)," "Failure to complete quality control self-inspection forms" (M76), "Failure to complete manufacturing supervision reports or no record of implementing these reports" (M51), "Failure to complete construction logs, failure to adhere to specified formats, or incomplete records" (M75), and "Missing records regarding the inspection and review of materials and equipment" (Q283). An inspecting committee will consider incomplete construction inspection reports inadequate, which affects the inspection score. The second subcategory pertains to defects related to labor safety and health. It includes the following: "Lack of proper fall prevention facilities, such as guardrails, covers, safety nets, and safety harnesses" (Q286), "Uncleared garbage and waste materials that can affect the environment" (Q60), and "Failure to supervise construction contractors in implementing on-site safety and health, traffic maintenance, and environmental protection measures" (M48). When inspecting a construction site, the inspection committee first checks the safety measures and protections provided for workers. Site safety is the foundation of site management. Detailed inspections should be conducted daily, both before and during construction, to prevent occupational incidents and ensure that project management strategies are effective.

- Concrete Quality (Aspect II): The critical defects of the second principal component are "Improper pouring and compaction of concrete resulting in the formation of cold joints, honeycombing, or voids" (Q2) and "Presence of debris (such as wires, iron pieces, or formwork) on the surface of the concrete" (Q5). These defects are commonly observed in concrete construction projects and can significantly affect construction quality. Avoiding these defects is crucial for supervisors.
- Construction Team Quality Management (Aspect III): The critical defects of the third principal component are "Other deficiencies related to the organizing agency or project management firm" (M61), "Other defects in supervisory unit quality management" (M29), and "Other defects in contractor quality management" (M113). These defects are all related to the quality management operations of the organizing agency, supervisory unit, and contractor. Clearly, whether the construction team effectively implements management tasks has a decisive effect on construction quality.

By addressing and managing these defects, construction teams can enhance performance and improve construction quality to meet the requirements and expectations of government agency inspectors.

4.2 Training and Evaluation of Predictive Models

Before model development, the data were first preprocessed. First, after checking the completeness of all variables, it was confirmed that the final analytical dataset used in this study contained no missing values; therefore, no imputation or case deletion was performed. Second, regarding scale transformation, the defect variables used for PCA were coded as binary occurrence indicators (0/1) and measured on the same scale; therefore, no additional standardization was applied before PCA. By contrast, the continuous project-attribute variables used as inputs to the MLP were normalized using min–max normalization prior to model training, in order to reduce the influence of differences in variable scale on model learning.

4.2.1 Hyperparameter Search Space and Optimization Strategy

To improve the robustness of model validation, this study adopted a repeated validation strategy to reduce the potential influence of random data partitioning and random weight initialization. The network architecture and training hyperparameters of the MLP, including the number of hidden layers, the number of neurons per hidden layer, the learning rate, and the batch size, were systematically optimized via a grid search combined with 5-fold stratified cross-validation, with validation performance as the primary criterion for model selection. To avoid overly optimistic evaluation, 20% of the data were first reserved as an independent test set using stratified sampling, and these samples were excluded from model training, cross-validation, and hyperparameter tuning. The remaining 80% of the data were used as the development set for 5-fold cross-validation and hyperparameter search. After the optimal hyperparameter combination was identified, the model was repeatedly trained and evaluated across multiple random seeds to assess its prediction stability. All candidate configurations were trained under the same training protocol to ensure a fair comparison across models, and the number of training epochs was fixed at 50. Early stopping was disabled at this stage to avoid inconsistencies in training duration caused by premature termination. Final model performance was evaluated using Accuracy, Precision, Recall, and F1-score, and the results were reported as the mean ± standard deviation across repeated experiments. The search space considered in this study was defined as follows: the number of hidden layers was set to 1, 2, or 3; the number of neurons in each hidden layer was set to 20, 30, 40, or 50; the learning rate was set to 0.1, 0.01, 0.001, or 0.0001; and the batch size was set to 16, 32, 64 or 128.

4.2.2 Gradient Experiments on the Number of Hidden Layers and Neurons

Fig. 7 illustrates the trend in validation accuracy across different hidden-layer depths (1, 2, or 3 layers) and neuron configurations. The results show that, for the single-hidden-layer model, validation accuracy increased from 86.2% to 89.5% as the number of neurons increased from 20 to 50. However, as the number of neurons increased further, validation performance began to decline, suggesting the model may have started to overfit. For the two-hidden-layer model, the (40, 30) configuration achieved the highest validation accuracy of 91.8%, outperforming the single-layer architecture overall.

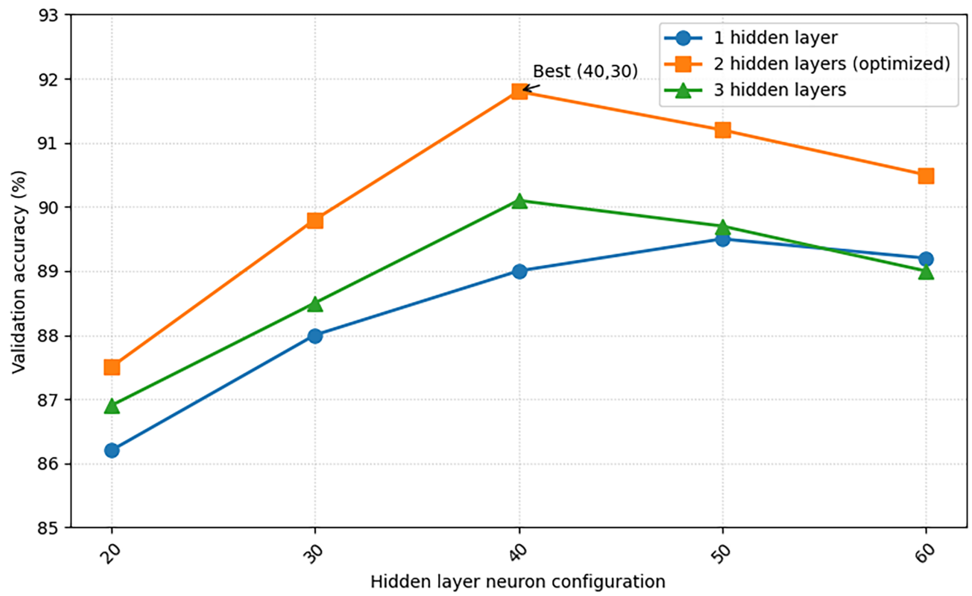


Figure 7: Model performance comparison under different hidden layer structures and neuron counts.

For the three-hidden-layer model, the highest validation accuracy was approximately 90.1% under the (30, 20, 10) configuration. However, the training time increased by approximately 40%, and the validation loss curve exhibited larger fluctuations, indicating that, under a small-sample setting, a three-layer architecture was more prone to instability and overfitting (Fig. 8). Based on these results, this study selected the two-hidden-layer architecture as the main model structure and adopted (40, 30) as the optimal neuron configuration. To ensure a fair comparison across different architectures, both the ablation analysis and hyperparameter sensitivity analysis were conducted with a fixed training length of 50 epochs, without early stopping. For the final model, however, the maximum number of training epochs was set to 30 based on the convergence behavior observed on the validation set, to reduce computational cost and avoid overfitting in the later stages of training.

Table 3 summarizes the performance metrics on the independent test set, including accuracy, precision, recall, and F1-score, together with the number of model parameters for several representative configurations. The results show that the two-hidden-layer architecture (40, 30) achieved the best overall performance while maintaining a moderate number of parameters, indicating that it is well suited to the task of construction quality prediction.

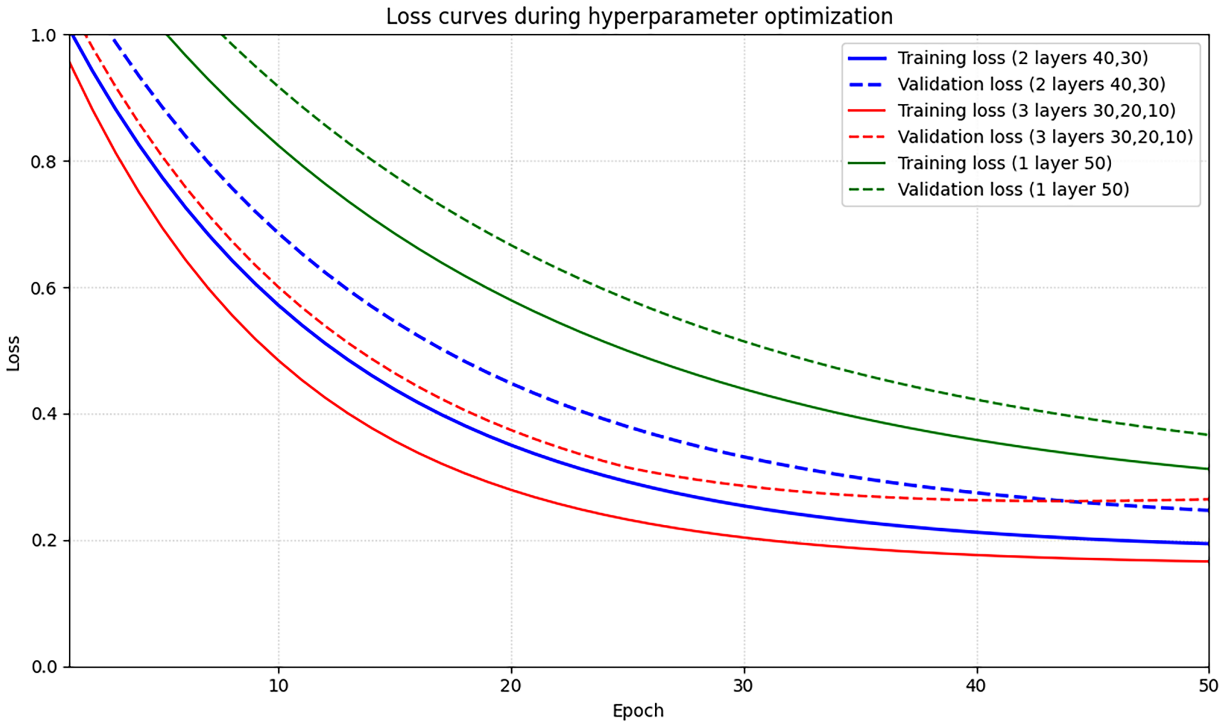


Figure 8: Loss curve comparison of MLP configurations under hyperparameter optimization.

Table 3: Performance comparison of MLP with different hidden-layer architectures.

Hidden-Layer Architecture	Number of Parameters	Accuracy (%)	Precision (%)	Recall (%)	F1-Score
1 layer (50)	1201	89.2	87.6	94.1	0.907
2 layers (40, 30)	2221	91.3	90.2	96.1	0.936
2 layers (60, 40)	3301	90.5	89.1	95.3	0.921
3 layers (30, 20, 10)	1951	89.8	88.3	94.7	0.914

4.2.3 Optimization of Learning Rate and Batch Size

Under the fixed two-hidden-layer architecture (40, 30), this study further conducted a grid search to optimize the learning rate and batch size. As shown in Fig. 9, when the learning rate was set to 0.001, the model converged relatively quickly and achieved lower final loss values, with the training loss reaching approximately 0.21 and the validation loss approximately 0.24. By contrast, when the learning rate was 0.01, the loss curves showed noticeable oscillation, indicating unstable optimization. When the learning rate was reduced to 0.0001, convergence became excessively slow, and the validation loss remained above 0.35 even after 50 epochs. For batch sizes 32 and 64, both performed similarly; however, 32 yielded slightly higher validation accuracy (91.8% vs. 91.5%) and better computational efficiency per training epoch. When the batch size was 16, the larger gradient noise led to unstable convergence. In contrast, a batch size of 128 yielded inferior generalization performance due to a low update frequency, resulting in a validation accuracy of approximately 90.3%. Based on these results, the final model was configured with a learning rate of 0.001 and a batch size of 32.

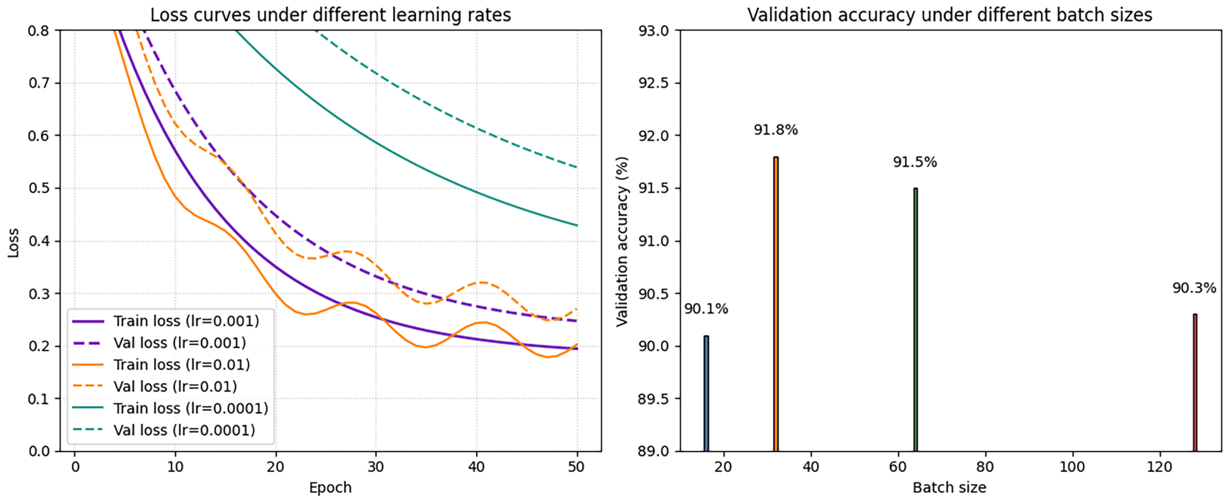


Figure 9: Sensitivity analysis of learning rate and batch size for MLP training.

4.2.4 Optimizer Experiment

Under the same MLP architecture—consisting of an input layer with 22 neurons, two hidden layers with 40 and 30 neurons, respectively, and an output layer with 1 neuron—all models were trained for 50 epochs with a fixed batch size of 32 and the cross-entropy loss function. Each optimizer was evaluated over five repeated runs, and the average results were reported. To ensure a fair comparison, the same stratified data-partitioning strategy adopted throughout this study was used: 20% of the data were reserved as an independent test set, whereas the remaining 80% were used as the development set for model training and validation. Under this setting, the performances of three widely used optimizers, namely Adam, SGD, and RMSprop, were compared. Among them, Adam showed the fastest convergence, reaching the best validation loss at approximately epoch 18, whereas SGD required about 38 epochs and still achieved inferior final performance. Fig. 10 presents the validation loss curves of the three optimizers during training. Adam exhibited the most rapid loss reduction, decreasing from 0.68 to 0.28 within the first 10 epochs, and then converging smoothly to approximately 0.22 without obvious oscillation. RMSprop converged slightly more slowly than Adam, becoming relatively stable after around 20 epochs, with a final validation loss of approximately 0.26; however, its curve still showed minor fluctuations. By contrast, SGD converged the most slowly, still showing a slight downward trend after 40 epochs, with a final validation loss of approximately 0.36, and its training curve displayed more pronounced oscillation.

The characteristics of the construction quality prediction dataset provide a natural explanation for the Adam optimizer's superior performance. First, the input variables include 13 critical defects (binary discrete variables), 7 project categories (discrete variables), and continuous variables such as construction progress and contract sums, resulting in a dataset with mixed data types and a relatively sparse feature structure. Because Adam adaptively adjusts the learning rate for each parameter, it can handle sparse gradients more effectively, whereas SGD uses a uniform learning rate for all parameters and is therefore less efficient for sparse features. Second, the construction quality dataset exhibits a certain degree of class imbalance (612 high-quality projects vs. 403 low-quality projects) and potential nonlinear interactions among variables. By combining momentum (first-order moments) with adaptive learning-rate adjustment (second-order moments), Adam is better able to maintain stable update directions in a complex loss landscape, thereby reducing the risk of slow convergence or oscillation that may occur with SGD.

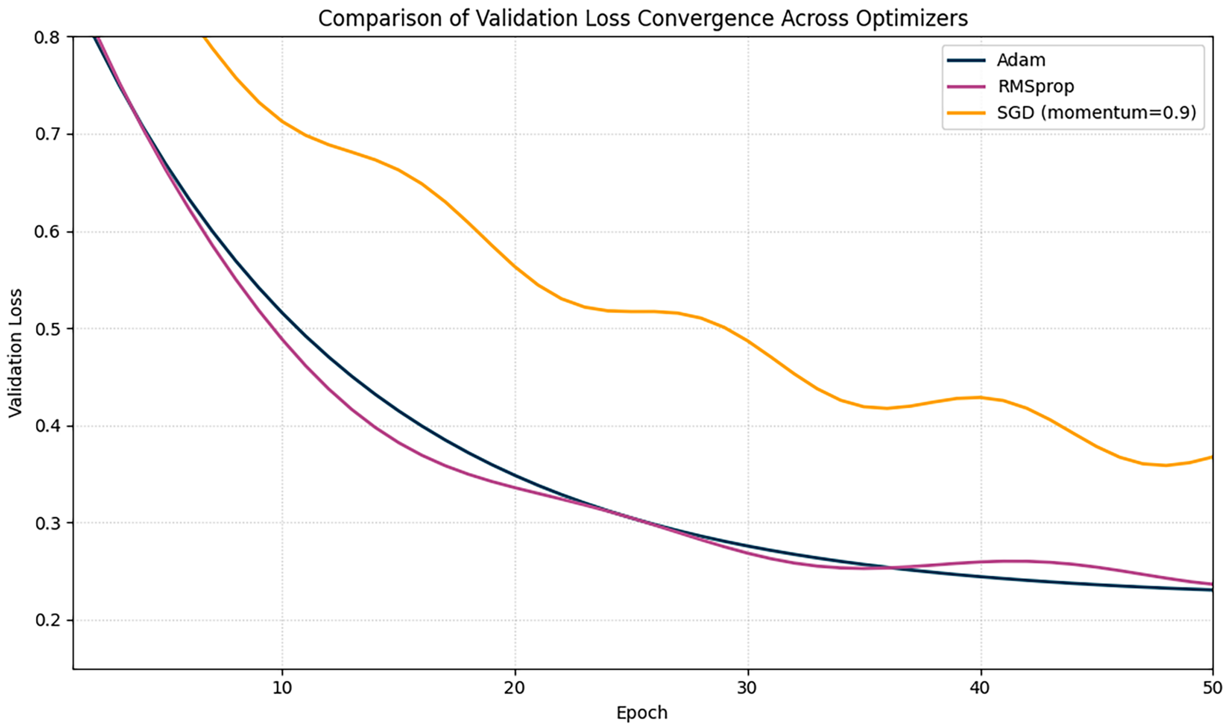


Figure 10: Comparison of validation loss convergence curves for Adam, RMSprop, and SGD.

4.2.5 Model Evaluation

Generally, data can be either discrete or continuous, and the nature of the data affects model selection, data transformation, and the interpretation of prediction results. Table 4 summarizes the final set of 22 input features used in the proposed construction quality prediction model. These predictors comprise three groups: (i) 13 critical defects (Defect_M47–M113), encoded as binary variables indicating whether the corresponding PCMIS defect code was recorded at least once for a given project (1 = yes, 0 = no); (ii) two continuous project attributes, namely contract sum and construction progress; and (iii) seven one-hot encoded project-category variables (Category_AE–RE).

In this study, the MLP was implemented as a nonlinear binary classifier to distinguish high-quality from low-quality projects. The final model adopted an architecture with an input layer of 22 neurons, two hidden layers with 40 and 30 neurons, respectively, and an output layer with 1 neuron. The hidden layers used the ReLU activation function, whereas the output layer used the sigmoid activation function to generate the probability that a project would be classified as high quality. The network was optimized using the Adam optimizer with a batch size of 32. During model development, repeated experiments across multiple random seeds showed stable performance on the development set. As summarized in Table 5, the model achieved an overall accuracy of 92.1 ± 1.2% on the development set. For the Good class, the precision, recall, and F1-score were 94.0 ± 1.5%, 95.0 ± 1.2%, and 0.945 ± 1.3%, respectively, whereas for the Poor class, the corresponding values were 90.0 ± 2.5%, 87.0 ± 2.0%, and 0.885 ± 2.2%. The macro-average and weighted-average F1-Scores were 0.915 ± 1.8% and 0.921 ± 1.6%, respectively, indicating stable, balanced classification performance during model development.

Table 4: Summary of the 22 input features used in the classification model.

No.	Feature Name	Type	Description
1	Defect_M47	Binary (0/1)	Defect variables are binary indicators denoting whether the corresponding PCMIS defect code was recorded ≥1 time for the project (1 = yes, 0 = no).
2	Defect_M76	Binary (0/1)	
3	Defect_Q286	Binary (0/1)	
4	Defect_Q60	Binary (0/1)	
5	Defect_M51	Binary (0/1)	
6	Defect_M48	Binary (0/1)	
7	Defect_M75	Binary (0/1)	
8	Defect_Q283	Binary (0/1)	
9	Defect_Q2	Binary (0/1)	
10	Defect_Q5	Binary (0/1)	
11	Defect_M61	Binary (0/1)	
12	Defect_M29	Binary (0/1)	
13	Defect_M113	Binary (0/1)	
14	Contract_Sum	Continuous	Contract amount (NT$ million)
15	Construction_Progress	Continuous	Percentage of construction completion (%)
16	Category_AE	Binary (0/1)	Airport Engineering
17	Category_CE	Binary (0/1)	Civil Engineering
18	Category_HE	Binary (0/1)	Harbor Engineering
19	Category_NE	Binary (0/1)	New Construction Engineering
20	Category_OE	Binary (0/1)	Other Engineering
21	Category_PE	Binary (0/1)	Power and Air Conditioning Engineering
22	Category_RE	Binary (0/1)	Renovation Engineering

Table 5: Detailed metrics for the development set (mean ± standard deviation).

Class	Precision (%)	Recall (%)	F1-Score
Good	94.0 ± 1.5	95.0 ± 1.2	0.945 ± 0.013
Poor	90.0 ± 2.5	87.0 ± 2.0	0.885 ± 0.022
Macro Avg	92.0 ± 2.0	91.0 ± 1.6	0.915 ± 0.018
Weighted Avg	92.4 ± 1.8	91.8 ± 1.4	0.921 ± 0.016
Accuracy		92.1 ± 1.2	

The confusion matrix is commonly used to evaluate the predictive performance of supervised machine learning models because it facilitates a direct comparison between model predictions and actual labels. Based on the confusion matrix, Accuracy, Precision, and Recall were calculated using Eqs. (7)–(9). In the present study, Table 6 presents the overall confusion matrix for all 1015 projects, rather than only for the independent test subset. Across the full dataset, the total number of correctly classified instances (TP + TN) was 927, corresponding to an overall accuracy of 91.3%. In addition, the precision for the "Good" class was 90.2%, and the recall was 96.1%, indicating that the model showed strong overall classification performance in identifying construction quality categories.

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + FP + TN} \tag{7}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{8}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{9}$$

Table 6: Confusion matrix summarizing the overall classification results for the full dataset.

Actual	Predicted	
	Good	**Poor**
Good	588	24
Poor	64	339

To provide a more comprehensive evaluation of classification performance, this study further employed Specificity, ROC-AUC, and Balanced Accuracy in addition to Accuracy, Precision, and Recall. Assuming that high-quality projects were treated as the positive class, Specificity (TN/(TN + FP)) measures the model's ability to correctly identify low-quality projects, and its value reached 0.934. Furthermore, based on the model's probability scores, the ROC curve was generated by plotting the true positive rate (TPR) against the false positive rate (FPR) under different decision thresholds. The corresponding AUC was 0.976, indicating excellent discriminative ability. In addition, Balanced Accuracy, defined as (Recall + Specificity)/2, is particularly suitable for imbalanced classification problems because it accounts for predictive performance across both positive and negative classes. The obtained Balanced Accuracy of 0.948 further confirms that the proposed model achieved stable and balanced classification performance across the two quality categories.

Fig. 11 presents the reliability diagram used to evaluate the model's calibration performance. The horizontal axis represents the mean predicted probability within each bin, and the vertical axis represents the corresponding observed fraction of positive instances. A perfectly calibrated model would produce a calibration curve lying on the diagonal reference line. As shown in the figure, the calibration curve closely follows the diagonal across most probability bins, indicating that the predicted probabilities are generally consistent with the observed outcomes. The overall expected calibration error (ECE) is 0.031, indicating a low calibration error and supporting the reliability of the model's probability estimates.

To validate the effectiveness of the proposed PCA–MLP hybrid model in the task of construction quality prediction, this study selected several commonly used baseline models in civil engineering research, including a single MLP, Logistic Regression, Support Vector Machine (SVM), Decision Tree, Random Forest, and XGBoost, for comparison. To ensure a fair comparison, all baseline models were also tuned using a consistent hyperparameter optimization strategy (e.g., grid search with a fixed search range). Each model configuration was repeated 10 times, and the results are reported as mean ± standard deviation. As shown in Table 7, the proposed PCA–MLP achieved the best overall classification performance. Its accuracy was 1.3 percentage points higher than that of XGBoost, and its recall reached 96.1%, exceeding the other models by 1.6–6.0 percentage points, indicating a stronger ability to reduce false negatives (FNs). As a linear model, Logistic Regression performed worse than the nonlinear models (Random Forest, XGBoost, and PCA–MLP), suggesting that the relationship between the input features and the target labels involves complex nonlinear patterns. By contrast, the single MLP without PCA achieved only 85.6% accuracy and required longer training time (48.2 s), indicating that PCA-based dimensionality reduction effectively suppressed redundant and noisy features, thereby improving both training efficiency and generalization performance. Although Random Forest and XGBoost also performed well, their recall values (93.8%–94.5%) remained lower than that of PCA–MLP, suggesting that neural networks are more advantageous in capturing nonlinear feature interactions.

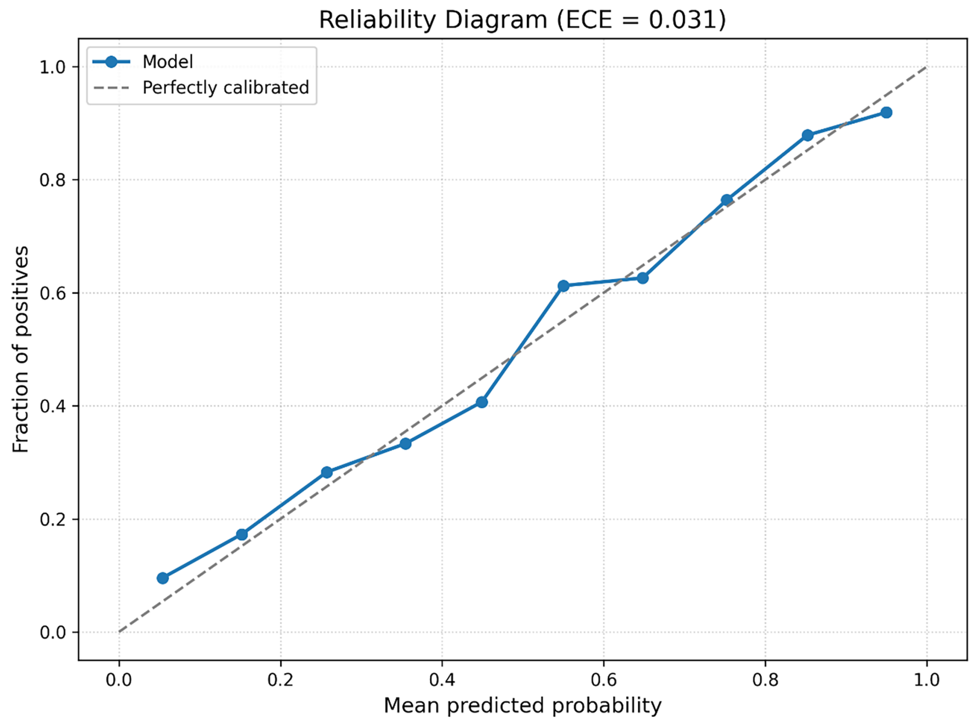


Figure 11: Reliability diagram of the MLP with expected calibration error (ECE).

Table 7: Comparison of different models for construction quality prediction (mean ± standard deviation).

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score	Training Time (s)
Single MLP	85.6 ± 1.4	84.2 ± 1.6	90.1 ± 1.2	0.868	48.2 ± 2.1
Logistic Regression	86.5 ± 1.0	85.3 ± 1.2	91.0 ± 0.9	0.879	8.2 ± 0.4
SVM	87.2 ± 0.9	86.0 ± 1.1	92.3 ± 0.8	0.889	12.4 ± 0.5
Decision Tree	88.2 ± 1.1	86.9 ± 1.3	92.8 ± 0.8	0.897	5.6 ± 0.2
Random Forest	89.1 ± 0.8	87.9 ± 0.9	93.8 ± 0.7	0.907	8.7 ± 0.3
XGBoost	90.0 ± 0.7	88.8 ± 0.8	94.5 ± 0.6	0.916	10.2 ± 0.4
PCA-MLP	91.3 ± 0.7	90.2 ± 0.8	96.1 ± 0.6	0.936	15.6 ± 0.6

From a practical engineering perspective, construction quality prediction mainly requires two capabilities: high recall and rapid prediction. First, regarding high recall, misclassifying a low-quality project as high quality (i.e., a false positive) may lead to insufficient follow-up quality control and potential safety risks. Therefore, reducing false negatives was given priority in evaluating model performance. In this regard, PCA–MLP achieved the highest recall (96.1%), indicating a stronger ability to identify low-quality projects and thus a lower risk of missed detections. Second, regarding rapid prediction, once a model has been trained, its inference time for new samples should be sufficiently short to support practical decision-making. The results show that all models achieved millisecond-level prediction times per sample, indicating that they are all feasible for practical application. Among them, PCA–MLP required 15.6 s for training, which remains acceptable overall. Therefore, the proposed model provides a favorable trade-off between recognition performance (high recall) and computational cost (training and inference time).

4.3 Result Analysis

MLP is a typical black-box model. Although it has strong nonlinear fitting capability, its internal decision-making process lacks intuitive interpretability. In civil engineering, practitioners require not only accurate predictions but also a clear understanding of the model's decision logic, so that the critical defects affecting construction quality can be identified and used to guide on-site management. Therefore, incorporating interpretability analysis is essential to enhancing the model's engineering applicability. In recent years, Shapley Additive Explanations (SHAP) has become one of the mainstream approaches for machine-learning interpretability because of its unified additive feature-attribution framework and solid theoretical foundation in game theory. For example, in predicting the seismic performance of reinforced concrete bridges, Luo et al. [41] used SHAP analysis to identify the key contributions of ground-motion intensity parameters (e.g., Sa10) and geometric parameters (e.g., pier height and span length) to bridge seismic response. Following this line of research, the present study employed SHAP to interpret the proposed PCA–MLP construction quality prediction model, quantify the contribution of the 13 critical defects to the prediction results, and reveal the relative importance of each defect in construction quality assessment (Fig. 12).

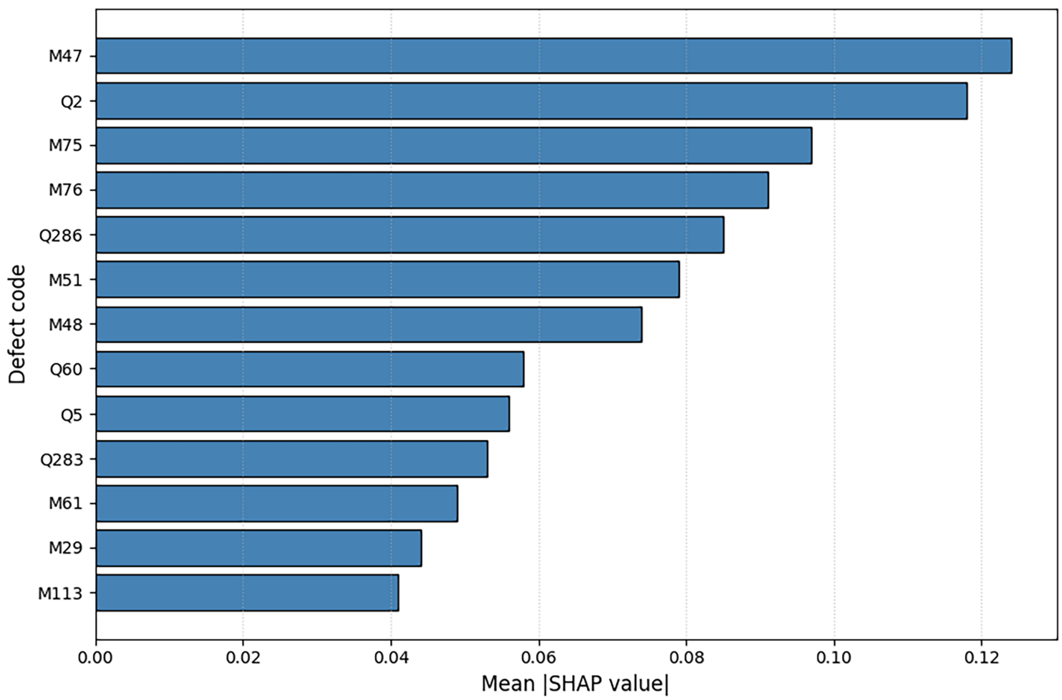


Figure 12: Global SHAP feature importance for 13 critical defects, identifying priority defect indicators for inspection management.

Among all features, M47 and Q2 ranked first and second, respectively, and their mean absolute SHAP values were substantially higher than those of the other features. This result indicates that the completeness of inspection records and the quality of concrete construction are the two most critical factors in determining construction quality. This finding is highly consistent with engineering practice: incomplete inspection records often imply failure of procedural control, whereas concrete-quality defects directly reflect the level of construction workmanship. In addition, defects such as M75, M76, and Q286 were also relatively important, suggesting that construction quality is strongly influenced by systematic issues in quality management. In other words, document management, procedural control, and occupational safety constitute three major pillars of construction quality.

From a management perspective, these SHAP results can be translated into specific inspection actions. When M47 is flagged, agencies should immediately verify whether construction-operation inspections, material/equipment inspections, and inspection records are complete and traceable. When Q286 is flagged, inspection teams should prioritize fall-prevention facilities, including guardrails, covers, safety nets, and safety harnesses, because this defect reflects both occupational-safety risk and weak site control. For Q2 and Q5, supervisors should strengthen concrete pre-pour checklists, compaction monitoring, and post-pour surface inspections to reduce honeycombing, cold joints, voids, and embedded debris. Thus, SHAP interpretation not only explains the model but also converts defect importance into a risk-based inspection protocol.

A total of 64 projects were classified as being a false positive (FP; i.e., the model predicted that a poor-quality project was good quality), and 24 projects were classified as being a false negative (FN; i.e., the model predicted that a good-quality project was poor quality). Both FPs and FNs are incorrect predictions; the input variables could be examined to explore the factors contributing to these types of misclassification.

(1) Defects

The input variables were 13 critical defects. To analyze the 64 FP projects and 24 FN projects, the average frequency of each critical defect was calculated for 88 projects (Fig. 13). Subtracting the average frequencies for FP and FN yielded a difference. A defect with a larger difference had different effects on the FP and FN predictions. In particular, the difference values were greater than 0.15 for the following defects: "Absence of inspections of construction operations, failure to conduct material and equipment inspections, and failure to complete inspection records" (M47), "Failure to complete construction logs, failure to adhere to specified formats, or incomplete records" (M75), and "Uncleared garbage and waste materials that can affect the environment" (Q60; Fig. 14).

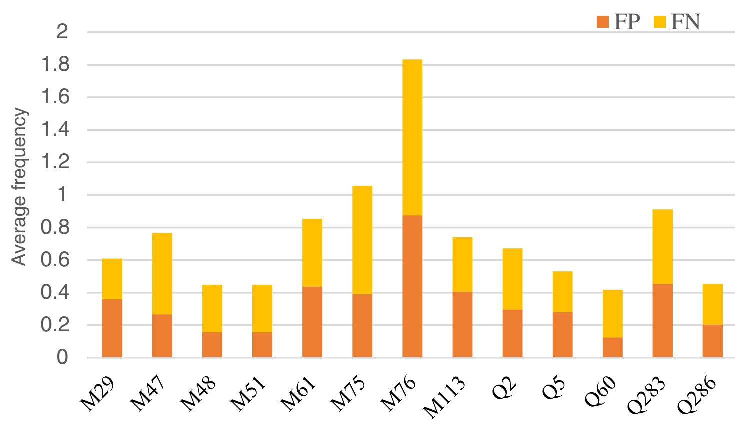


Figure 13: Average frequency of critical defects for the FP and FN cases.

Figure 14: FP vs. FN difference values for critical defects.

(2) Contract sums

Fig. 15 illustrates the numbers of FPs and FNs classified by the projects' contract sum: NT$1–10 million, NT$10–50 million, and greater than NT$50 million. For the FPs, the proportions of these three types of projects were similar to the proportions in the total population of 1015 projects, indicating that the contract sum did not significantly affect FP predictions. However, the proportions significantly differed for the FNs. The NT$1–10 million range contained fewest projects (6 projects), with the NT$10–50 million and >NT$50 million ranges containing more projects (9 projects each). However, the number of FNs was small; determining whether the contract sum affects FN explanations was thus difficult.

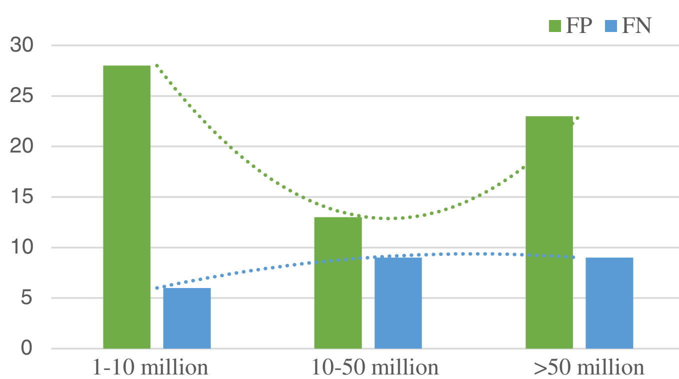


Figure 15: Number of FP and FN projects by total contract sum.

(3) Construction progress

The projects were categorized into two groups on the basis of the construction progress: less than and greater than 50% completion. The proportions of projects for which the prediction was a FP within these two progress categories were similar to the proportions in the population. Specifically, the proportions of FP projects with less than and greater than 50% completion were 59% and 41%, respectively (Fig. 16a); these proportions were 56% and 44% in the population. These results suggested that FP misclassifications were not affected by the construction progress. However, the proportion of FN projects with lower than 50% progress was 75% (Fig. 16b). This indicates that projects with progress lower than 50% may be misjudged as having poor construction quality when, in reality, these projects may ultimately be high quality. Furthermore, the input variables for score and project category were not discovered to strongly affect FP and FN misjudgments. Fig. 17 reveals that the proportions of FP and FN misjudgments across the seven project categories were similar.

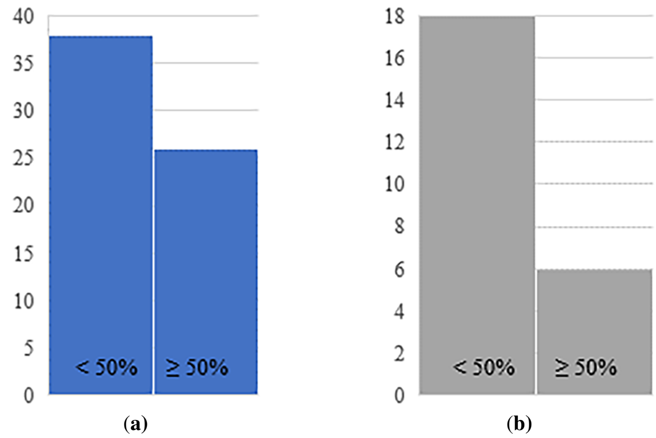


Figure 16: Number of (**a**) FP and (**b**) FN misjudgments by construction progress.

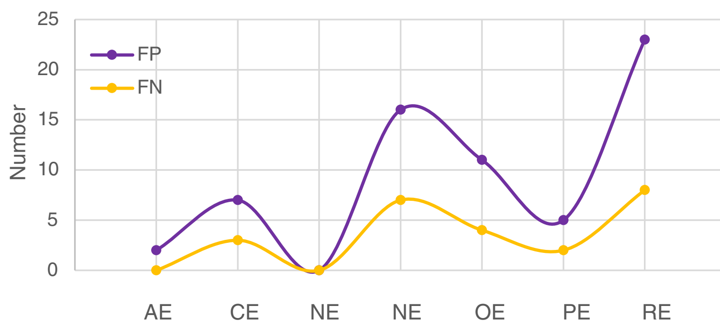


Figure 17: Number of FPs and FNs for each project category.

The FP and FN results provide several actionable implications for construction management. Because FP cases indicate poor-quality projects that were previously predicted to be good-quality, they represent a more critical management risk than FN cases. Such projects may be overlooked by inspection agencies and lack sufficient follow-up supervision. Therefore, projects classified as good quality by the model should still be subject to secondary review when key red-flag defects are present. In particular, the FP–FN comparison revealed that M47, M75, and Q60 had relatively large differences between the two types of misclassification. These defects are associated with incomplete construction-operation inspections, incomplete construction logs, and uncleared garbage or waste materials affecting the construction environment. Accordingly, the presence of these defects should trigger additional document review, on-site verification, or corrective action, even when the model prediction is favorable.

FN cases also provide useful management information. In this study, 75% of FN cases occurred in projects with construction progress below 50%, suggesting that early-stage projects may be more easily judged as poor quality because their records, construction processes, or site conditions are still developing. Therefore, for projects at an early construction stage, model predictions should be interpreted alongside construction progress. A poor-quality prediction for an early-stage project should not necessarily be regarded as a final quality failure; rather, it should be used as an early-warning signal for targeted follow-up inspection.

(4) Analysis of poor construction quality

Fig. 18 illustrates the frequency of the 13 critical defects in the projects defined by actual quality (not predicted quality). The proportion of poor-quality projects with defects M47, M48, M51, Q60, and Q286 was at least 2.5% higher than that of good-quality projects. That is, approximately 70% of the poor-quality construction projects had one of these five critical defects. In particular, 84.2% of the projects with "Failure to complete manufacturing supervision reports or no record of implementing these reports" (M51) were of poor quality, highlighting that documenting inspection reports is a critical means of quality supervision and control. Therefore, inspection committee members should emphasize record-keeping regarding construction inspections. If relevant inspection information is not documented, project quality cannot be assured. By contrast, the proportions of poor-quality projects with Q2 and Q5 defects were less than 40%, suggesting that although many projects may have these two defects, they are not significantly related to overall quality.

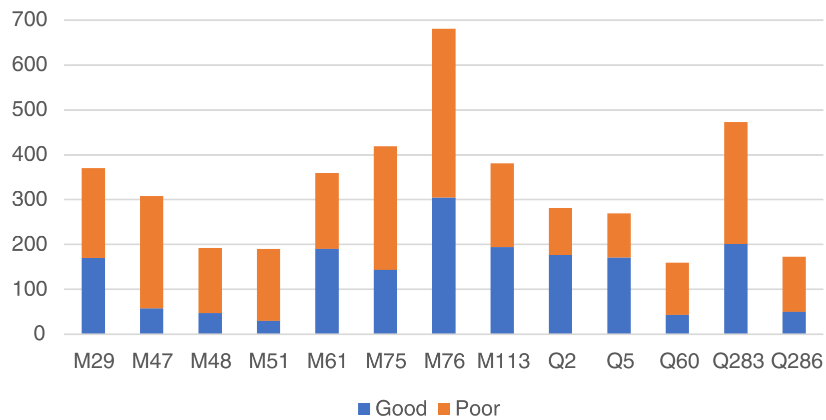


Figure 18: Proportion of good- and poor-quality projects with each critical defect.

Regarding project categories, 50.7% of renovation engineering (RE) projects were poor, whereas 88.9% and 78.1% of harbour engineering (HE) and airport engineering (AE) projects, respectively, were good (Fig. 19a). This indicates that the construction quality of RE projects is generally worse than that of HE and AE projects. Moreover, regarding inspection grades, 61.7% of Grade C projects had poor quality, and 66.3% of Grade B projects had good quality (Fig. 19b). The three contract sum categories had similar proportions of projects with good and poor construction quality; poor projects accounted for 35.5%–41.7% of the projects for each category (Fig. 19c). The proportion of poor projects also did not differ greatly for projects with construction progress of less than vs. greater than 50% (40.4% and 39.2%, respectively; Fig. 19d).

4.4 Limitations and Future Work

A limitation of this study is that the empirical dataset was obtained from Taiwanese public construction projects. Although the PCMIS database provides a large-scale and institutionally standardized source of inspection records, the resulting model may still reflect Taiwan-specific regulatory requirements, inspection procedures, project management practices, and defect-recording conventions. Therefore, the generalizability of the proposed PCA–MLP framework to other regions or countries should be interpreted with caution.

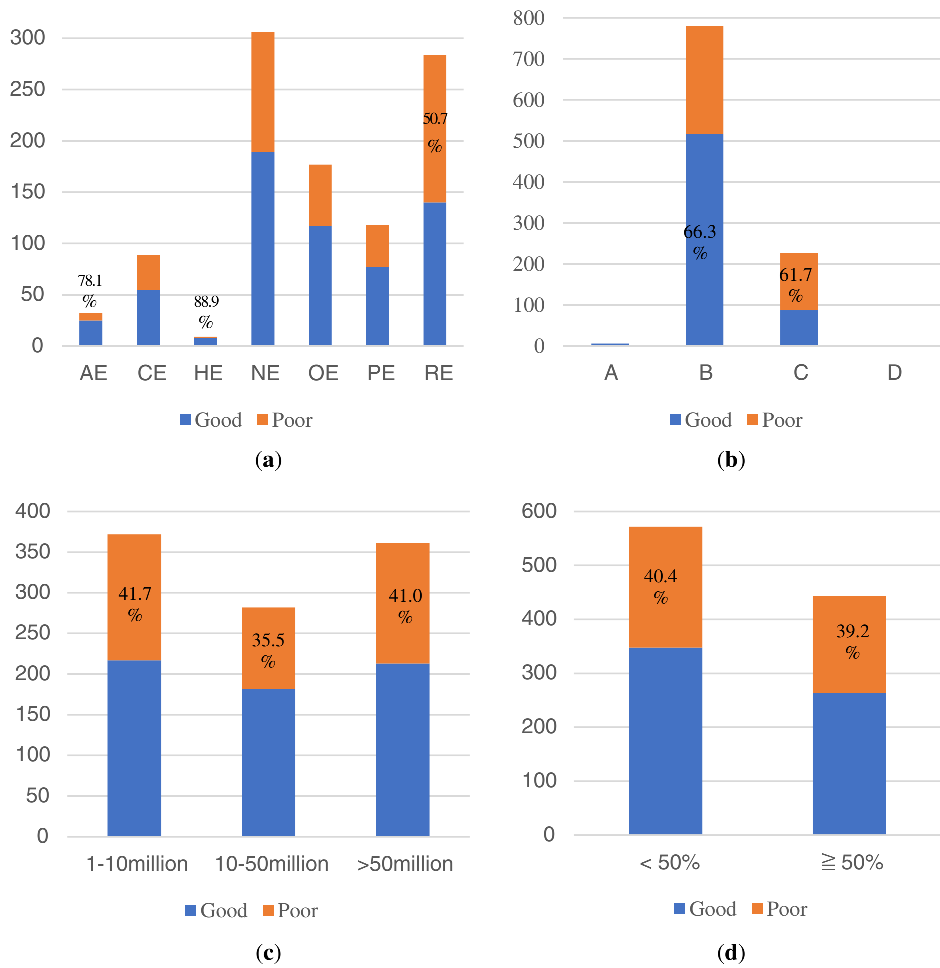


Figure 19: Proportion of projects of good and poor quality by (**a**) project category, (**b**) inspection grade, (**c**) contract sum, and (**d**) construction progress.

In addition, the PCA–MLP model in this study relied primarily on defect records from construction inspections. Other potentially informative data sources, such as site-monitoring data (e.g., real-time sensor measurements of concrete temperature, humidity, compaction, and strain) and raw-material inspection data (e.g., reinspection results for cement, aggregates, and reinforcing steel), were not incorporated. These multi-source datasets contain dynamic information related to defect formation and may further improve the model's predictive accuracy and generalizability. Moreover, integrating experimental and theoretical models has been shown to be effective for predicting the mechanical performance of defective components in civil engineering. For example, in studies of concrete-filled steel tube (CFST) columns, the introduction of PMC strengthening and the consideration of internal void defects, together with axial compression tests and theoretical analysis, enabled accurate prediction of the load-carrying capacity of defective members [42].

Inspired by these developments, future research can be extended in three directions. First, multi-source data fusion may be introduced by combining site-monitoring data, such as vibration, temperature, and strain measurements collected during construction, and material testing data, such as concrete compressive strength and steel yield strength, with the existing 13 critical defects as complementary model inputs. Multi-modal learning or feature-level fusion methods may then be applied to further improve prediction accuracy. Second, defect–performance relationship modeling may be established through scaled experimental models of construction defects, such as concrete honeycombing, cold joints, and reinforcing-bar spacing deviations, to quantify the relationship between defect severity and structural performance, including load-carrying capacity and durability. On this basis, experimentally derived degradation patterns could be embedded into machine learning models as prior knowledge, thereby extending the analysis from simple defect occurrence to the estimation of performance impact. Third, a small-scale comparative test may be conducted using a pilot dataset from an international partner agency or region, such as the Ministry of Land, Infrastructure, Transport, and Tourism of Japan. Transfer learning or domain adaptation techniques could then be explored to adapt the model parameters to regional engineering characteristics, thereby establishing cross-regional credibility and ensuring broader applicability beyond the Taiwanese context.

5 Conclusion

By combining PCA with MLP, valuable information can be extracted from large datasets to discover key factors affecting results. In this study, a combined model was used to analyze construction inspection reports from the PCMIS. PCA was performed to calculate the component loadings of defects and identify latent correlations, and an MLP network was used to transform the input data into probabilities indicating construction quality; finally, a model for predicting construction quality was established. In the first stage of this study, PCA was used to identify 13 critical defects, which were then grouped into three aspects: Inspection Records and Occupational Safety and Health, Concrete Quality, and Construction Team Quality Management. In the second stage, an MLP was trained on these project attribute variables; its weights were iteratively updated to minimize the difference between the desired and actual outputs. The goal of this iterative training process was to achieve model convergence and to ultimately predict construction quality.

In the construction sector, quality is influenced by many complex and uncertain factors, making effective defect control a persistent challenge for project teams. By reducing defect redundancy and identifying 13 critical defects, the proposed PCA–MLP framework provides a more focused basis for construction quality assessment. These findings suggest that practitioners should prioritize these critical defects in project quality-control practices. More broadly, the proposed framework may serve as a useful digital tool for proactive quality control in public construction projects and support the development of more intelligent, data-driven quality management systems on smart construction sites. The proposed model may serve as a useful analytical reference for construction teams to identify quality-related defect patterns under conditions similar to those in the present dataset. However, its broader generalizability remains to be further verified through external, longitudinal, and cross-regional validation.

Acknowledgement: Not applicable.

Funding Statement: The author received no specific funding for this study.

Availability of Data and Materials: Data available on request from the authors.

Ethics Approval: Not applicable.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Mostafa K, Hegazy T. Review of image-based analysis and applications in construction. Autom Constr. 2021;122(3):103516. doi:10.1016/j.autcon.2020.103516.
2. Minchin RE, Hammons MI, Ahn J. A construction quality index for highway construction. Constr Manag Econ. 2008;26(12):1313–24. doi:10.1080/01446190802621036.

3. Lee YJ, Rhee SK, Kim DS, Lee C. Assessment of expressway construction using Quality Performance Index (QPI). KSCE J Civ Eng. 2013;17(2):377–85. doi:10.1007/s12205-013-1930-4.
4. Jogdand PME, Deshmukh SS. Development of building quality measurement tool: building construction quality index (BCQI). Int Res J Eng Technol. 2017;4(1):1020–6.
5. Kim Y, Kim MS, Kim JH. Development of key performance indicators for the improvement of university facility management services in Korea. J Asian Archit Build Eng. 2018;17(2):313–20. doi:10.3130/jaabe.17.313.
6. Ebrahim Abu El-Maaty A, Akal AY, El-Hamrawy S. Management of highway projects in Egypt through identifying factors influencing quality performance. J Constr Eng. 2016;2016(2):4823630. doi:10.1155/2016/4823630.
7. An Y, Li H, Su T, Wang Y. Determining uncertainties in AI applications in AEC sector and their corresponding mitigation strategies. Autom Constr. 2021;131(3):103883. doi:10.1016/j.autcon.2021.103883.
8. Lin G, Shen GQ, Sun M, Kelly J. Identification of key performance indicators for measuring the performance of value management studies in construction. J Constr Eng Manag. 2011;137(9):698–706. doi:10.1061/(asce)co.1943-7862.0000348.
9. Chen WT, Chang PY, Huang YH. Assessing the overall performance of value engineering workshops for construction projects. Int J Project Manag. 2010;28(5):514–27. doi:10.1016/j.ijproman.2009.08.005.
10. Zhao P, Gao D, Ren R, Han K, Yang Z, Meng W, et al. Short-term aging performance evaluation of asphalt based on principal component and cluster analysis. J Test Eval. 2021;49(1):590–602. doi:10.1520/jte20180781.
11. Park S, Lee JJ, Yun CB, Inman DJ. Electro-mechanical impedance-based wireless structural health monitoring using PCA-data compression and *k*-means clustering algorithms. J Intell Mater Syst Struct. 2008;19(4):509–20. doi:10.1177/1045389x07077400.
12. Chiang YH, Wong FK, Liang S. Fatal construction accidents in Hong Kong. J Constr Eng Manag. 2018;144(3):04017121. doi:10.1061/(asce)co.1943-7862.0001433.
13. Dackermann U, Li J, Samali B. Dynamic-based damage identification using neural network ensembles and damage index method. Adv Struct Eng. 2010;13(6):1001–16. doi:10.1260/1369-4332.13.6.1001.
14. Li J, Dackermann U, Xu YL, Samali B. Damage identification in civil engineering structures utilizing PCA-compressed residual frequency response functions and neural network ensembles. Struct Control Health Monit. 2011;18(2):207–26. doi:10.1002/stc.369.
15. Zahedi Tajrishi F, Mirza Goltabar Roshan A, Zeynalian M, Vaseghi Amiri J. A new collection of compressed damage indices for multi-damage detection of cold formed steel shear walls based on neural network ensembles. Can J Civ Eng. 2016;43(12):1034–43. doi:10.1139/cjce-2015-0417.
16. Fan CL. AI-enhanced defect identification in construction quality prediction: hybrid model of unsupervised and supervised machine learning. Procedia Comput Sci. 2023;230(1):112–9. doi:10.1016/j.procs.2023.12.066.
17. Alaloul WS, Liew MS, Wan Zawawi NA, Mohammed BS, Adamu M. An artificial neural networks (ANN) model for evaluating construction project performance based on coordination factors. Cogent Eng. 2018;5(1):1507657. doi:10.1080/23311916.2018.1507657.
18. Liu S, Chang R, Zuo J, Webber RJ, Xiong F, Dong N. Application of artificial neural networks in construction management: a systematic review. Appl Sci. 2021;11(20):9616. doi:10.3390/app11209616.
19. Kashfi S, Rouzegar J, Tavakolpour-Saleh A. Piping structural health monitoring: ANN methods for internal semi-elliptical crack detection. Int J Press Vessels Pip. 2025;214:105417. doi:10.1016/j.ijpvp.2024.105417.
20. Annamdasu ML, Challagulla SP, Kontoni DN, Rex J, Jameel M, Vicencio F. Artificial neural network-based prediction model of elastic floor response spectra incorporating dynamic primary-secondary structure interaction. Soil Dyn Earthq Eng. 2024;177:108427. doi:10.1016/j.soildyn.2023.108427.
21. Corbally R, Malekjafarian A. A data-driven approach for drive-by damage detection in bridges considering the influence of temperature change. Eng Struct. 2022;253(3):113783. doi:10.1016/j.engstruct.2021.113783.
22. Khayatazad M, Honhon M, De Waele W. Detection of corrosion on steel structures using an artificial neural network. Struct Infrastruct Eng. 2023;19(12):1860–71. doi:10.1080/15732479.2022.2069272.
23. Xu B, Wang X, Yang CW, Li Y. Machine learning–aided rapid estimation of multilevel capacity of flexure-identified circular concrete bridge columns with corroded reinforcement. J Struct Eng. 2024;150(3):04024002. doi:10.1061/jsendh.steng-12718.

24. Houlík J, Valentin J, Nežerka V. Predicting the fatigue life of asphalt concrete using neural networks. J Mater Civ Eng. 2026;38(4):04026013. doi:10.1061/jmcee7.mteng-20929.
25. Ho LV, Trinh TT, De Roeck G, Bui-Tien T, Nguyen-Ngoc L, Abdel Wahab M. An efficient stochastic-based coupled model for damage identification in plate structures. Eng Fail Anal. 2022;131(4):105866. doi:10.1016/j.engfailanal.2021.105866.
26. Li J, Zhang S, Wang X. Physics-informed neural network with fuzzy partial differential equation for pavement performance prediction. Autom Constr. 2025;171:105983. doi:10.1016/j.autcon.2025.105983.
27. Li M, Huang L, Pidwerbesky B. An ANN-based approach for nondestructive asphalt road density measurement. J Transp Eng Part B Pavements. 2024;150(3):04024033. doi:10.1061/jpeodx.pveng-1354.
28. Wu X, Bao X, Shen J, Chen X. Graph neural network–based framework for predicting seismic damage in shield tunnels with contact loss defects. Autom Constr. 2025;180(3):106535. doi:10.1016/j.autcon.2025.106535.
29. Fan CL. Optimization and performance evaluation of machine learning classifiers for predicting construction quality and schedule. Autom Constr. 2025;179(1):106470. doi:10.1016/j.autcon.2025.106470.
30. Hamdy K, AbdelRasheed I, Essawy YAS, Gamal ElDeen A. Automated risk analysis for construction contracts using neural networks. J Leg Aff Dispute Resolut Eng Constr. 2024;16(4):04524023. doi:10.1061/jladah.ladr-1149.
31. Patel P, Patel DV, Lad VH, Patel KA, Patel DA. Predicting construction crew productivity for concrete-pouring operations. J Leg Aff Dispute Resolut Eng Constr. 2024;16(2):04524001. doi:10.1061/jladah.ladr-1034.
32. Fan SL, Yeh IC, Chi WS. Improvement in estimating durations for building projects using artificial neural network and sensitivity analysis. J Constr Eng Manage. 2021;147(7):04021050. doi:10.1061/(asce)co.1943-7862.0002036.
33. Jezzini Y, Assaad RH, El-adaway IH. Modeling framework to quantify and gauge project cost risks due to construction material price volatilities using predictive probabilistic deep-learning algorithms and stochastic risk modeling. J Constr Eng Manage. 2025;151(7):04025071. doi:10.1061/jcemd4.coeng-16055.
34. Cheung SO, Tam CM, Harris FC. Project dispute resolution satisfaction classification through neural network. J Manage Eng. 2000;16(1):70–9. doi:10.1061/(asce)0742-597x(2000)16:1(70).
35. Mhady AA, Gurgun AP, Budayan C, Koc K. New hybrid models integrating the firefly optimization algorithm with the artificial neural networks and adaptive neuro-fuzzy inference systems to improve estimation at completion. J Constr Eng Manage. 2025;151(11):04025160. doi:10.1061/jcemd4.coeng-16487.
36. Fan CL. Predicting the construction quality of projects by using hybrid soft computing techniques. Comput Model Eng Sci. 2025;142(2):1995–2017. doi:10.32604/cmes.2025.059414.
37. Ghassemieh M. Evaluation of stiffened end-plate moment connection through optimized artificial neural network. J Softw Eng Appl. 2012;5(3):156–67. doi:10.4236/jsea.2012.53023.
38. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90. doi:10.1145/3065386.
39. Zhang L, Shen J, Zhu B. A research on an improved Unet-based concrete crack detection algorithm. Struct Heal Monit. 2021;20(4):1864–79. doi:10.1177/1475921720940068.
40. Haponava T, Al-Jibouri S. Proposed system for measuring project performance using process-based key performance indicators. J Manage Eng. 2012;28(2):140–9. doi:10.1061/(asce)me.1943-5479.0000078.
41. Luo L, Li Y, Dai H, Sun H, Jia M, Yuan H, et al. Seismic performance prediction and interpretation of RC bridges under vehicle-bridge interaction: from VBI system simulation to ensemble learning surrogate models. Eng Struct. 2025;341:120816. doi:10.1016/j.engstruct.2025.120816.
42. Luo L, Sun H, Jia M, Peng B, Li X, Yuan H, et al. Experimental study and theoretical prediction of axial compression behavior in PMC-reinforced CFST columns with void defects. Eng Struct. 2024;313:118258. doi:10.1016/j.engstruct.2024.118258.