1-30-2024

Research and comparison of pavement performance prediction based on neural networks and fusion transformer architecture

Hui Yao

Ke Han

Yanhao Liu

Dawei Wang

Zhanping You

Electronic Research Archive

Research article

Research and comparison of pavement performance prediction based on neural networks and fusion transformer architecture

Hui Yao[1,*], Ke Han[1], Yanhao Liu[1], Dawei Wang[2] and Zhanping You[3]

[1] Beijing Key Laboratory of Traffic Engineering, College of Metropolitan Transportation, Faculty of Architecture, Civil and Transportation Engineering, Beijing University of Technology, No.100, Pingleyuan, Chaoyang, Beijing 100124, China

[2] School of Transportation Science and Engineering, Harbin Institute of Technology, Harbin 150006, China

[3] Department of Civil and Environmental Engineering, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931-1295, United States

* **Correspondence:** Email: huiyao@mtu.edu.

Abstract: The decision-making process for pavement maintenance from a scientific perspective is based on accurate predictions of pavement performance. To improve the rationality of pavement performance indicators, comprehensive consideration of various influencing factors is necessary. To this end, four typical pavement performance indicators (i.e., Rutting Depth, International Roughness Index, Longitudinal Cracking, and Alligator Cracking) were predicted using the Long Term Pavement Performance (LTPP) database. Two types of data, i.e., local input variables and global input variables, were selected, and S-ANN and L-ANN models were constructed using a fully connected neural network. A comparative analysis of the predictive outcomes reveals the superior optimization of the L-ANN model. Subsequently, by incorporating structures such as self-attention mechanism, a novel predictive approach based on the Transformer architecture was proposed. The objective is to devise a more accurate predictive methodology for pavement performance indices, with the goal of guiding pavement maintenance and management efforts. Experimental results indicate that, through comparative analysis of three quantitative evaluation metrics (root mean square error, mean absolute error, coefficient of determination), along with visual scatter plots, the predictive model employing the fused Transformer architecture demonstrates higher robustness and accuracy within the domain of pavement performance prediction when compared to the L-ANN model. This outcome substantiates the efficacy and superiority of the model in terms of predictive performance, establishing it as a reliable

tool for accurately reflecting the evolution of asphalt pavement performance. Furthermore, it furnishes a theoretical reference for determining optimal preventive maintenance timing for pavements.

1. Introduction

As China's economy and technology continue to rapidly develop, infrastructure is also constantly improving and updating. Among these improvements, pavement engineering has achieved notable success. However, with the continuous aging of old pavement projects, the focus of pavement engineering implementation is not only on the construction of new pavements but also on the establishment of a complete and cyclical pavement maintenance system. In the realm of pavement maintenance management, conducting timely and appropriate maintenance can often result in lower costs compared to addressing issues after significant deterioration has occurred. Pavement surface performance data is a critical component of the pavement management system, and pavement surface performance prediction is the basis for maintenance decisions. In promoting the healthy construction and maintenance of asphalt pavement, it has become crucially important to accurately predict the development trend of asphalt pavement performance [1]. The establishment of a scientific and accurate pavement surface performance prediction model is essential for pavement maintenance and repair methods. By accurately grasping the degradation of key performance indicators of asphalt pavement, the optimal maintenance timing can be better determined and a scientific maintenance decision-making plan can be formulated to maximize the service benefits of pavement surfaces [2].

The deterioration of asphalt pavement performance refers to the situation where, during its actual service life, the original structure or function of the pavement undergoes changes due to the influence of various factors. These factors encompass a multitude of variables, such as pavement structure, traffic load, natural environment, and maintenance interventions, among others, and their interactions are intricate and diverse. Hence, characterizing the relationship between performance indicators describing pavement behavior and the factors that influence them is nonlinear and exceedingly complex.

In addressing the issue of predicting pavement performance, scholars both domestically and internationally have implemented a variety of research methodologies. Traditional prediction models utilized in the past include deterministic and probabilistic models. Zhu, Meilan et al. [3] conducted a mechanical and performance analysis of various pavement structures on Huaxi Road in Beijing. They established a predictive model for the pavement performance index of Huaxi Road based on actual test data and assessed the accuracy of related indicators such as PCI (Pavement Condition Index) and RQI (Roughness Quality Index). Jin, Yin-Li et al. [4] conducted a study using the Han-Ning Expressway network in Shaanxi Province, China as a case study to investigate the relationship between vehicle load and pavement performance. They utilized pavement evaluation data to obtain the Sectional Cumulative Equivalent Single Axle Load (SCESAL) and the Pavement Quality Index (PQI). Several regression models were developed to capture the relationship between SCESAL and PQI. Karam J developed a multiple linear regression model based on data collected from the LTPP database [5]. This model is used to predict the rut depth of hot mix asphalt concrete (HMA) pavement under specific structural, climatic conditions, traffic levels, and asphalt mixture volume characteristics. He, Zhimin et al. [6] collected observational data on five pavement performance indicators for the Beijing Sixth

Ring Expressway and performed linear regression analysis and established decay prediction models for these five asphalt pavement performance assessment indicators and conducted precision testing. The research methodologies of deterministic models generally only consider empirical predictive models between performance indicators and time series, failing to establish nonlinear relationships between pavement performance indicators and their influencing factors. This has resulted in a considerable gap between the deterministic models and actual pavement performance decay, with decreasing accuracy as the prediction period extends.

Conversely, probabilistic models were utilized for estimating the state distribution of pavement condition indicators at a given time, where Markov methods, semi-Markov methods, and residual curve methods are commonly employed. Probabilistic models account for the stochastic development of pavement conditions, but due to their prediction being based on probability transition matrices, their expression is less intuitive than direct predictions of pavement condition indicators. Furthermore, the determination of state transition probabilities is subjective. Abaza KA et al. [7] introduced a novel technology that utilizes reverse calculation based on discrete-time Markov models to estimate transition probabilities used in Markov-based pavement performance prediction models. A simplified road surface management model has been proposed for developing a flexible long-term repair plan for pavement surfaces. This model deploys a discrete-time Markov model to predict the deterioration of the performance of the original pavement surface and the repaired pavement surface [8]. The objective of the proposed model is to generate the optimal annual rehabilitation cycle within the specified analysis period. This goal is achieved by optimizing the cost-effectiveness index, which is defined as the ratio of expected performance improvement to annual rehabilitation costs. Pavement performance can be regarded as a function associated with time series. However, subdividing pavement surfaces into smaller segments and describing the pavement performance of each segment solely based on time is inadequate.

Pavement surface performance is influenced by numerous factors such as pavement-way structure, traffic load, and natural environment. These factors exhibit spatial variability in different pavement-way segments. As a result, using only time to describe pavement surface performance in small segments cannot capture this spatial variability. Characterizing the nonlinear relationship between performance indicators and their influencing factors that characterize road surface performance is exceedingly complex. Consequently, an increasing number of artificial intelligence methods are being applied to predict pavement performance indicators, including artificial neural networks, support vector machines, and genetic algorithms. For example, Yao L et al. [9] predicted the deterioration of pavement conditions in Jiangsu Province, including ruts, roughness, skid resistance, transverse cracking, and pavement fatigue, utilizing neural networks to develop five models to predict their performance. Abdualaziz Ali A et al. [10] investigated the integrated impact of poor pavement conditions on flexible pavement performance in two climate regions of the United States and Canada, employing multivariate linear regression (MLR) and artificial neural network (ANN) techniques for modeling to predict International Roughness Index (IRI) using Long-Term Pavement Performance (LTPP) databases to obtain pavement performance data. Based on the Grey GM (1,1) model (a single variable Gray prediction model with a first order difference equation) and modified by a Markov model, a Grey Markov combination model was established to accurately predict the performance of highway pavement [11]. Based on ThunderGBM's Ensemble learning model and Shapley's additive interpretation (SHAP) method, the international roughness index (IRI) of asphalt pavement is predicted. Using the SHAP method to explain its potential influencing factors and their interactions [12]. Using evaluation data from 33 runways, the ANN was trained, validated, and optimized through a series of

Heavy Weight Deflectometer (HWD) tests. By estimating the simplified pavement damage index in ANN, the impact of service life and air traffic on runway pavement performance was analyzed [13]. Artificial neural network (ANN) was used to estimate Marshall test parameters (OAC, stability, flow rate, voids, voids in mineral aggregates) using aggregate gradation as input for the prediction process. Test multiple ANNs to optimize neural network hyperparameters and generate accurate predictions. Test different Activation function, the number of hidden layers and the number of neurons in each hidden layer, and generate heat maps to compare the performance of each ANN [14]. Using principal component analysis (PCA), factor analysis, and cluster analysis, 71 types of unsupervised machine learning were used to identify the principal components and common factors of climate variables, and the dataset was classified into different groups. Then, two supervised machine learning methods, Fisher discriminant analysis and artificial neural network (ANN), were used to predict climate regions based on climate data [15]. In order to overcome the limitations of using friction coefficient to predict road anti slip resistance, they developed an improved neural network (GAI-NN) based on genetic algorithm [16]. Utilize many machine learning (ML) techniques to create more complex models for predicting pavement performance, and compare different ML models to evaluate their predictive capabilities [17]. Hossain MI et al. [18] utilized an artificial neural network (ANN) with a 7-9-9-1 architecture to predict the International Roughness Index (IRI) of flexible pavement by collecting climate and traffic data from the Long-Term Pavement Performance (LTPP) database. Rulian B et al. [19] developed an asphalt pavement performance model that considers traffic and climate loads, pavement age, initial roughness condition, and maintenance and repair (M&R) interventions based on the LTPP database. The model's predictions, which used an artificial neural network (ANN), effectively captured the effect of M&R interventions, and the predicted International Roughness Index (IRI) values corresponded well with observation. Liu J et al. [20] improved current asphalt mixture design by introducing machine learning (ML) models to predict pavement IRI. They selected 37 input features related to climate conditions, traffic, pavement-way structure, and pavement material properties, analyzed the impact of two different dimensionality reduction techniques on ML model performance, and trained different ML algorithms such as support vector regression (SVR), random forest, and artificial neural networks (ANN) to predict IRI. The performance indicators of these ML models were calculated and compared. Based on actual on-site data obtained from long-term pavement performance databases, flexibly modify many parameters of the ANN model, such as the number of neurons, hidden layers, and function types, to obtain more accurate prediction models. Compared with statistical modeling methods, the ANN method can be used to accurately predict pavement fatigue and rutting distress [21]. A multi input unified prediction model [22] based on artificial neural networks has been developed using a mixture of numerical and classification features from in-service pavement test sections in the United States. The input variables include pavement age, crack length and area, cumulative traffic load, two functional categories of the pavement, four climate zones, and maintenance effectiveness, with changes in the Pavement Condition Index (PCI) as the output. Luo, Z et al. [23] proposed an XGBoost-based road performance model for predicting IRI and introduced SHAP to enhance the interpretability of individual features in the model. Aranha, A.L et al. [24] analyzed the impact of different training datasets on machine learning models for road performance prediction. Kaya, O et al. [25] developed a road performance model based on statistical and artificial intelligence (AI) techniques. Xiao, M et al. [26] proposed an enhanced backpropagation neural network (BPNN) prediction model based on particle swarm optimization (PSO) algorithm. Saha, S et al. [27] proposed the development of a k-value prediction model based on artificial neural networks to improve

the sensitivity of subgrade to rigid pavement performance. Liu, G et al. [28] developed a novel artificial neural network (ANN) for predicting the lifecycle performance of rigid pavement surfaces. An increasing number of studies are employing the Long-Term Pavement Performance (LTPP) database to train deep neural network (DNN) models, aiming to learn the nonlinear and intricate relationships between multiple performance indicators (RD, IRI, AC, LC) of asphalt pavements and a variety of associated parameters (including maintenance and repair, climate, traffic, pavement structure, and characteristics) [29–33].

Currently, the Transformer model architecture is commonly employed to forecast long time-series data [34], through the utilization of self-attention mechanisms to learn complex patterns and dynamics from temporal data. Das A et al. [35] proposed a multi-layer perceptron (MLP)-based encoder-decoder model, dubbed as Time-series Dense Encoder (TiDE), for long-term time-series prediction. This model not only possesses the simplicity and speed of linear models but also can handle covariates and nonlinear dependencies. Furthermore, the Transformer is a versatile framework that can be applied to both univariate and multivariate time-series data, as well as temporal embeddings. Bai S et al. [36] proposed a bidirectional extended Short-term memory network model (Att-BiLSTM) based on attention mechanism, which uses the time series characteristics of pavement temperature and meteorological factors to improve the prediction performance. Guo, Feng et al. [37] proposed a Long Short-Term Memory (LSTM) model with attention mechanism, which efficiently and effectively learns time-series related features to better predict IRI.

The impact of changes in pavement performance indicators is multifaceted and comprehensive [38]. There are numerous factors that affect the performance of pavement usage, and selecting a judicious subset of these factors can significantly influence the size of the neural network performance. However, it is not advisable to include all factors as prediction parameters since selecting too many prediction parameters during neural network training can increase the difficulty of training and also lead to a higher probability of overfitting. The comprehensive set of selected influencing factors includes pavement base information, traffic load, pavement performance, temperature, and maintenance information, among others, which can optimize the neural network from a global perspective and achieve accurate prediction of asphalt pavement usage performance [9]. Pavement performance mainly includes four major indicators: Damage, smoothness, strength, and skid resistance. Foreign pavement management systems generally do not include pavement strength and skid resistance as indicators, with more analysis focused on smoothness and damage indicators. Different indicators serve different application purposes.

2. Methodology and objective

2.1. Artificial neural network (ANN)

The fully connected artificial neural network (ANN) is a network structure consisting of multiple layers, each composed of numerous neurons. A fundamental way to comprehend neural networks is to perceive them as composite functions that map input data to output results. Typically, neural networks consist of an input layer, hidden layers, and an output layer, as illustrated in Figure 1. The basic unit of this network is a neuron, which, through modeling and interconnections, sequentially inputs different variables. By employing specific learning algorithms, the network adjusts the weight matrices of its layers iteratively until convergence is reached, thereby enabling the network to adapt to the

requirements of its surrounding environment.

An artificial neural network (ANN) is a computational model based on artificial neurons and connection weights. They are widely used in various fields of study. The fundamental architecture of a neural network includes several key components:

1) Input Layer: This layer is responsible for receiving external input data, which serves as the initial information for the network's computations.

2) Hidden Layers: These layers process and transform the input data through a series of mathematical operations. Typically, there are multiple hidden layers in a neural network, allowing for complex representations and feature extraction.

3) Output Layer: The output layer receives the processed data from the hidden layers and produces the final results or predictions, which are then passed to the external environment.

4) Connection Weights: These represent the strength of connections between neurons in the network. During the model training process, the connection weights are continuously adjusted in order to minimize the discrepancy between the predicted outputs and the actual outcomes, thereby optimizing the network's performance.

5) Activation Function: An activation function is utilized to nonlinearly transform the inputs of a neuron, enabling the neuron to handle nonlinear problems effectively.

6) Loss Function: A loss function is employed to evaluate the discrepancy between the predicted outcomes of a model and the actual results. It serves as a guide for optimizing the model's parameters.

7) Backpropagation Algorithm: The backpropagation algorithm is a method used to compute the gradient of the model's error. By iteratively adjusting the model's parameters through gradient descent, the algorithm facilitates continuous optimization.

By combining and optimizing these components, neural networks can accomplish regression tasks. In this context, a neural network is developed to create a pavement surface prediction model that individually forecasts rutting, smoothness, longitudinal cracks, and alligator cracks. The structure and parameters of the model are adjusted to select the model with the highest predictive accuracy.
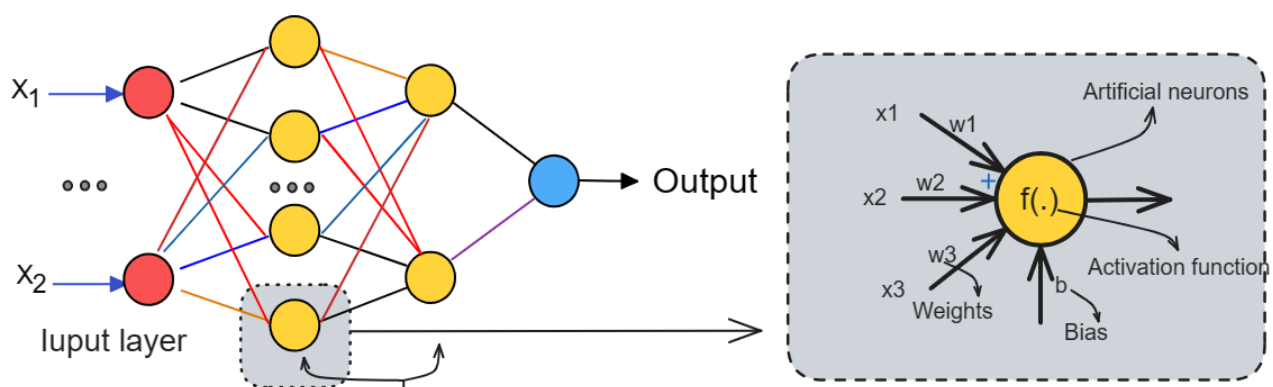

Figure 1. Artificial neural network structure.

2.2. Transformer algorithm

The Transformer model is a deep learning architecture that leverages attention mechanisms and commonly employs an encoder-decoder structure. The construction of the Transformer model involves

several essential components, including positional encoding, multi-head self-attention mechanism, residual connections, and layer normalization. In the following discourse, we will provide a comprehensive exposition of the technical intricacies involved in building a Transformer model.

The Transformer Encoder, as an integral part of the model, takes positional encoding information as input, with its central layer comprising a multi-head attention mechanism. The Add and Norm layers perform the operations of summation and subsequent normalization on the inputs and outputs of the Multi-Head Attention layer. Subsequently, this processed information is propagated to the Feed Forward layer. Finally, another round of Add and Norm operations is conducted, culminating in the generation of the ultimate output of the model [39].

The attention mechanism serves as the cornerstone of the Transformer model. The self-attention mechanism addresses the scenario where a neural network receives numerous vectors of varying sizes, with certain relationships existing between these vectors. However, during actual training, the network fails to fully exploit these interdependencies, resulting in highly suboptimal training outcomes. To tackle this issue of the inability of fully connected neural networks to establish correlations among multiple related inputs, the self-attention mechanism is employed. Essentially, the self-attention mechanism aims to enable the machine to discern the correlations between different parts of the entire input.

In the self-attention mechanism, the input matrix undergoes three distinct linear transformations, which map it into three sets of spatial vector matrices termed the query matrix Q, key matrix K, and value matrix V. Each row of the Q, K, and V matrices has a dimensionality of d_k, as depicted in Figure 2. The computation formula for the self-attention mechanism is defined by Eqs (1)–(4):

$$Q = Linear(X) = XW^Q \tag{1}$$

$$K = Linear(X) = XW^K \tag{2}$$

$$V = Linear(X) = XW^V \tag{3}$$

$$SelfAttention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{4}$$

where, X is the input matrix, W^Q, W^K, W^V signify weight matrices utilized for linear transformations. Moreover, the Softmax normalization function is employed to facilitate the computation of attention weights.
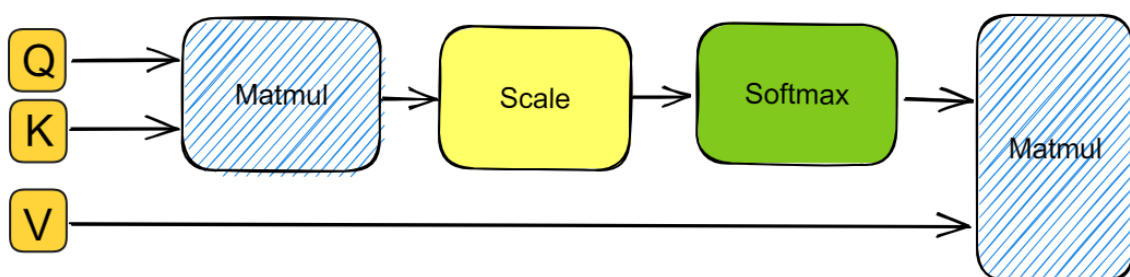

Figure 2. Self-attention mechanism [39].

The Transformer model is primarily composed of a self-attention mechanism and a fully connected layer as its submodules. The data is processed between the attention layer and the fully

connected layer through residual connections and normalization techniques.

The residual structure ensures that the output dimensions of the data after undergoing multi-head attention operations remain consistent with the input, enabling the utilization of residual connections. These connections address the issues of gradient vanishing and degradation of weight matrices. The specific approach for implementing residual connections is quite straightforward. It involves simply adding the input and the output of the multi-head attention layer together, as depicted by Eq (5) below:

$$X = X_{input} + SelfAttention(Q, K, V) \qquad (5)$$

The purpose of normalization is to standardize the hidden variables of the model into a standard normal distribution, thereby accelerating convergence. The specific calculation process is illustrated by Eqs (6)–(8) as follows:

$$\mu_j = \frac{1}{m} \sum_{i=1}^{m} x_{ij} \qquad (6)$$

$$\sigma_j^2 = \frac{1}{m} \sum_{i=1}^{m} (x_{ij} - \mu_j) \qquad (7)$$

$$LayerNorm(x) = \frac{x_{ij} - \mu_j}{\sqrt{\sigma_j^2 + \epsilon}} \qquad (8)$$

where, x_{ij} is the value located in the i-th row and j-th column of the output matrix, μ_j is the mean of the j-th column in the output matrix, and σ_j^2 is the variance of the j-th column in the output matrix.

The normalization process involves subtracting the mean of each column from its respective elements and dividing them by the standard deviation of that column. Consequently, normalized values are obtained. To prevent division by zero, a nonzero constant ϵ is added to the denominator in the equations.

2.3. Objective

We select four typical pavement performance indicators from the LTPP database: Rutting depth (RD), international roughness index (IRI), longitudinal crack (LC), and alligator crack (AC). Considering the comprehensiveness of these four data indicators, a combination of them is helpful for objectively evaluating pavement usage conditions based on the prediction results. By considering pavement performance indicators, pavement structure, environment, and maintenance measures as factors, two types of data inputs, namely partial input variables and overall input variables, are chosen to be combined with fully connected neural networks to obtain S-ANN and L-ANN models. The prediction results are then compared, and a pavement performance prediction method based on a fusion transformer decoder architecture is employed. Prediction models for rutting depth (RD), international roughness index (IRI), longitudinal crack (LC), and alligator crack (AC) are established based on the above four indicators, providing a novel approach to predicting pavement usage performance.

3. Pavement surface performance prediction based on two types of models

3.1. Data preparation

3.1.1. Data preprocessing

The Long Term Pavement Performance (LTPP) database is an innovative project aimed at fulfilling diverse pavement information requirements. We utilize existing knowledge in pavement technology and seek to develop models that explain the behavior of pavement surfaces. The LTPP database stores information regarding various design features, traffic and environmental factors, materials, construction quality, and maintenance activities, all of which impact pavement surface performance. Through the analysis and prediction of pavement surface performance using a substantial amount of stored data, improved predictive models can be established for pavement design and management. Furthermore, this database enables the understanding of the influence of different pavement design features on pavement surface performance, facilitating research on new materials, construction techniques, and maintenance methods in specific projects.

In this study, a 15 km highway segment in Texas was selected as the research area. Maintenance records revealed a pavement spacing interval of 154 m, which was utilized to collect data on four fundamental pavement surface performance indicators: rutting, roughness, cracking, and longitudinal cracking. Additionally, meteorological data, traffic parameters, and relevant pavement structure data were collected. However, it should be noted that there were instances of missing target year data, as well as missing data in traffic and weather parameters. The missing data, particularly in terms of longitudinal cracking (LC) and International Roughness Index (IRI), were addressed through interpolation, data imputation, or data deletion. Furthermore, under maintenance interventions, variations in pavement structure and materials led to abrupt changes in the trends of pavement surface performance. Maintenance interventions significantly influenced the deterioration of pavement surface performance. Additionally, certain illogical occurrences were observed in the data. For instance, errors were noted where IRI and rutting depth (RD) decreased over time in the absence of any maintenance actions. To address this issue, calibration adjustments were applied to data points between two consecutive maintenance action dates. Improving data quality is essential to ensure that survey results reflect the actual changes in pavement surface performance, rather than being influenced by poor data quality-induced variations.

The pavement structure in Texas consists of the layer of asphalt, a base layer, and a subgrade. Different material types and thicknesses will be fed into the model as different input variables. Within the experimental sections in Texas, the combinations of pavement structures varied. In fact, the pavement structure configurations differ across states, even when the number of layers is the same, as there are variations in thickness. In some pavement structures, a wearing course of approximately 1.1 cm thickness is added above the upper layer. If the wearing course is not considered separately, the asphalt layers are typically distributed in 2 layers, with a thickness ranging from 9.5 to 19 cm. The pavement structure configurations include a 4-layer structure (2 surface layers + 1 base layer + subgrade). The surface layer comprises the original surface layer and a bond coat layer. The base layer consists of two types: Treated base (TB) and granular base (GB). Additionally, information regarding the subbase layer is provided in Table 1.

Table 1. Pavement structure.

Material type	Name	Material code	Material thickness(cm)
AC	Friction Course	2	0.8–1.1
AC	Surface Course	1	0.7–2.5
AC	Binder Course	1, 13	1.1–8.4
TB	Treated Base	327, 350	6.9–15.2
GB	Granular Base	303, 307, 308, 309	6.3–15.6
TS	Treated Subbase	338	6.5–10.4
GS	Granular Subbase	308, 309	8.4–10
SS	Subgrade	103, 104, 114, 118, 202, 214, 215, 216, 217	130–204

Prior to network training, the following variables need to be preprocessed, as outlined in Table 2. The preprocessing tasks involve the following steps: First, categorical variables such as pavement functional material and pavement type are transformed into numerical codes using one-hot encoding. Second, due to the disparate magnitudes of the original variables, it is necessary to ensure that all data range within a common scale and improve convergence rates. This is achieved by applying min-max normalization to numerical data such as temperature and material thickness. By doing so, faster convergence rates can be achieved. Factors influencing pavement performance can be categorized into 24 factors, including basic information on pavement type, climate conditions, pavement age, maintenance information, traffic volume, and other factors. Considering the regional pavement conditions, multiple variables from Table 1 are employed as different input variables for the neural network prediction model. The output parameters of the prediction model consist of pavement surface performance indicators for the next year, including rutting depth (RD), International Roughness Index (IRI), alligator cracking (AC), and longitudinal cracking (LC).

Table 2. Neural network input variables.

Type	Function name	Variable type
	Friction Course Material	Nominal Variable
	Friction Course Thickness	Numerical Variable
	Surface Material	Nominal Variable(S-ANN)
	Asphalt Layer Thickness	Numerical Variable(S-ANN)
	Binder Course Material	Nominal Variable(S-ANN)
	Binder Course Thickness	Numerical Variable(S-ANN)
Basic Information	Base Material	Nominal Variable(S-ANN)
	Base Layer Thickness	Numerical Variable(S-ANN)
	Subbase Material	Nominal Variable(S-ANN)
	Subbase Layer Thickness	Numerical Variable(S-ANN)

Type	Function name	Variable type
Basic Information	Subgrade Material	Nominal Variable(S-ANN)
	ESAL	Numerical Variable
Traffic Load	AADT	Numerical Variable(S-ANN)
Climate	Temperature	Numerical Variable(S-ANN)
	Precipitation	Numerical Variable
	Humidity	Numerical Variable
Others	Pavement Age	Numerical Variable(S-ANN)
Maintenance Information	Maintenance Type	Nominal Variable
	Maintenance Material	Nominal Variable
	Material Thickness	Numerical Variable
	Rutting depth (RD)	Numerical Variable
Previous Year Pavement Performance	International Roughness Index (IRI)	Numerical Variable
	Alligator Cracking (AC)	Numerical Variable
	Longitudinal Cracking (LC)	Numerical Variable

3.1.2. Data analysis

The pavement information data is initially aggregated and subjected to statistical processing to conduct exploratory data analysis and identify significant aspects of these variables. The data is then visualized, as shown in Figure 3. Through visualizations, several observations can be made. The performance indicators exhibit a distribution centered on both sides, particularly for alligator cracking and longitudinal cracking, where the results on the right side tend to approach zero. This indicates that the overall occurrence of alligator cracking and longitudinal cracking is relatively low, implying better performance. On the other hand, rutting is predominantly concentrated on the right side, indicating poorer performance in this aspect. Moreover, higher values of International Roughness Index (IRI) are considered desirable, but the distribution is primarily skewed towards the left side, indicating poorer performance in terms of smoothness. Consequently, pavement maintenance measures must address these two situations and employ appropriate maintenance strategies to rectify the skewness.
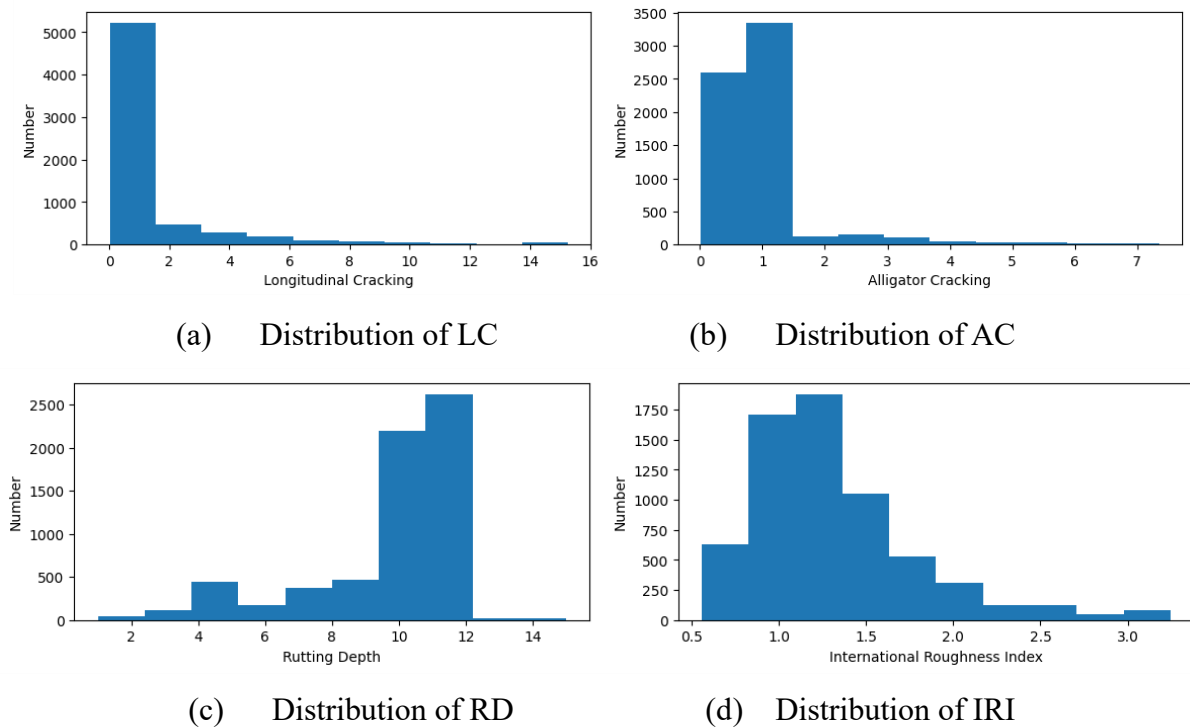
(a)   Distribution of LC

(b)   Distribution of AC

(c)   Distribution of RD

(d)   Distribution of IRI

Figure 3. Distribution of pavement indicators.

When comparing differences between multiple objects and their attributes, heatmaps are employed. Furthermore, the correlation between continuous feature variables is explored by calculating the correlation among all continuous numeric variables. A heatmap is used to visualize the potential correlations between inputs, as depicted in Figure 4. When strong correlations between pavement distress responses are evident, it suggests that the presence of one type of pavement distress may indicate the existence of another type. The four types of pavement distress: RD, IRI, AC, and LC exhibit strong correlations with one another, often co-occurring. Additionally, there is a significant correlation between LC, IRI, and temperature. Asphalt layer thickness is the most important factor affecting AC, while other related variables have a more evenly distributed impact on RD.

As shown in Figure 5, by comprehensively considering various influencing factors and utilizing the self-attention mechanism within the Transformer, correlations between these factors can be addressed. The subsequent integration with neural networks enables fitting and prediction. The aim is to develop a more precise predictive methodology for pavement performance indices, ultimately guiding pavement maintenance and management efforts.
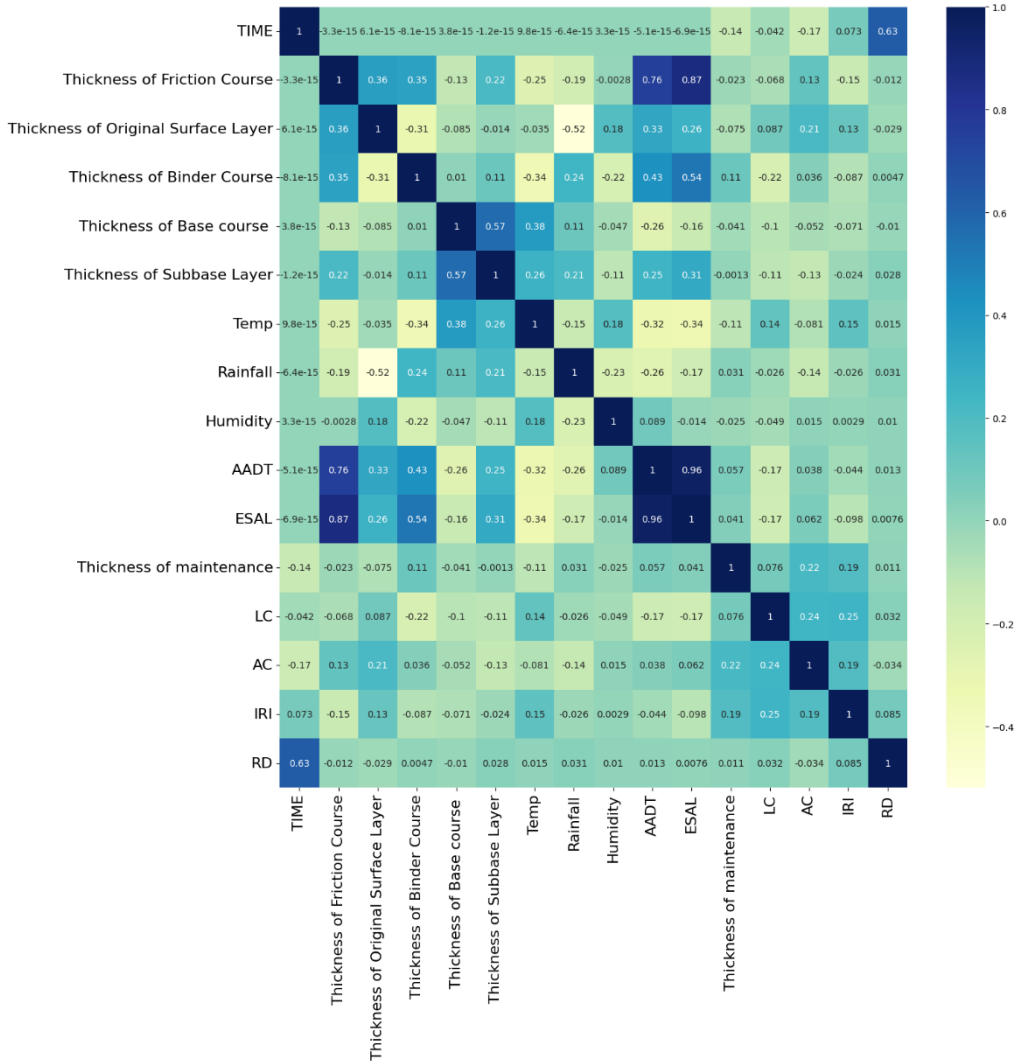
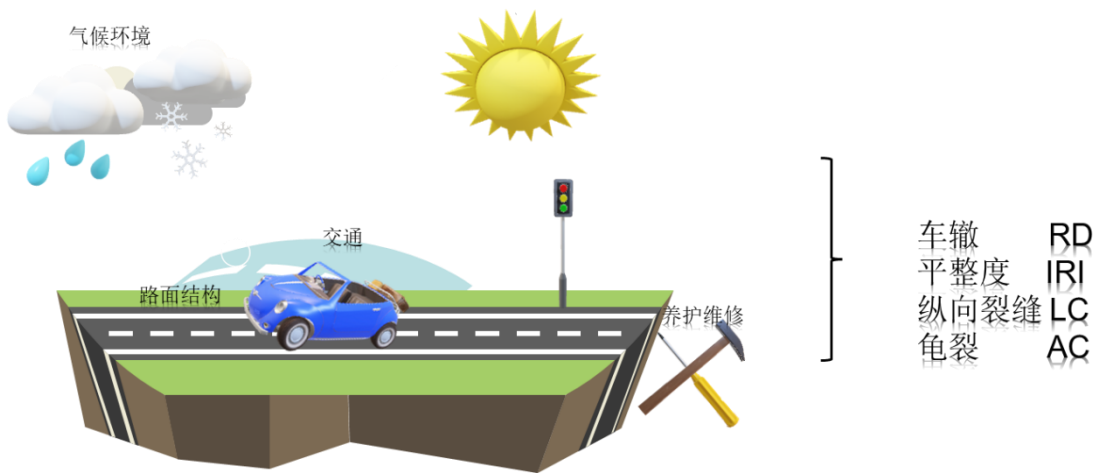**Figure 4.** Heatmap of pavement surface influencing factors.

Figure 5. Schematic diagram of influencing factors on pavement surface indicators.

3.2. The construction and optimization of the prediction model

3.2.1. Artificial neural network (ANN) prediction model

In this study, we focus on the development of a pavement surface performance prediction model based on neural networks, implemented on the Python 3.7 platform. The TensorFlow framework is utilized to construct the neural network model. Two variations of the model are considered: The S-ANN model, which incorporates 12 selected influencing factors from Table 2 as input neurons, and the L-ANN model, which includes all 24 influencing factors. The model structure is illustrated in Figure 6.

Optimal models are obtained by testing different structures, with all four models designed to have two hidden layers. For the key hyperparameters in the prediction models of rutting depth (RD), International Roughness Index (IRI), longitudinal cracking (LC), and alligator cracking (AC), an optimization process is performed.

After employing GridSearchCV and cross-validation for the initial selection and optimization of the hyperparameter 'hidden_layer_sizes' in the Artificial Neural Network (ANN) models, we considered three different hidden layer sizes: 16, 32 and 64. A total of 7 combinations were evaluated during the grid search. Subsequently, further refinement and fine-tuning were performed on the most promising parameter combinations. Table 3 summarizes the results, indicating that the optimal performance for the Large-scale ANN (L-ANN) model was achieved with a hidden layer size of (64, 64). On the other hand, the Small-scale ANN (S-ANN) model exhibited its best performance with a hidden layer size of (32, 32). These optimal configurations were determined through a systematic grid search followed by a detailed fine-tuning process.

Since the number of training samples is nearly equal for both models, ReLU activation function is applied to all hidden layers, and a dropout rate of 0.1 is utilized to deactivate a fraction of neurons to prevent overfitting. The loss is measured using mean squared error. By utilizing Keras' Model Checkpoint, the weights of the best model with the minimum validation loss are stored for future predictions. The dataset is divided into training and testing sets with an 8: 2 ratio. Considering a trade-off between model performance, stability, and computational efficiency, we set the learning rate (lr) to 0.0001, the number of epochs to 1000, and the batch size to 16, using the Adam optimizer as the optimal hyperparameter configuration for the prediction model. The model parameters are summarized in Table 4.

Table 3. Training comparison with different numbers of neurons for L-ANN.

Hidden-layer-sizes	R² (RD)	R² (LC)	R² (AC)	R² (IRI)
(16, 16)	0.8919	0.8308	0.8480	0.7905
(32, 16)	0.8916	0.8219	0.8735	0.8272
(16, 32)	0.9023	0.8533	0.8521	0.8205
(32, 32)	0.9034	0.8639	0.8784	0.8416
(64, 32)	0.9028	0.8661	0.8892	0.8382
(32, 64)	0.9025	0.8693	0.8996	0.8440
(64, 64)	0.9186	0.8744	0.9073	0.8519

Table 4. ANN model parameter settings.

Parameter name	Parameter value	Parameter name	Parameter value
Training Set:Test Set	8: 2	Loss Function	MSE
Number of Input Layers	12, 24	Activation Function	ReLU
Dropout	0.1	Optimizer	Adam
Number of Fully Connected Layers	2	Number of Training Epochs	1000
Number of Nodes in Fully Connected Layers	32, 64	Batch Size	16
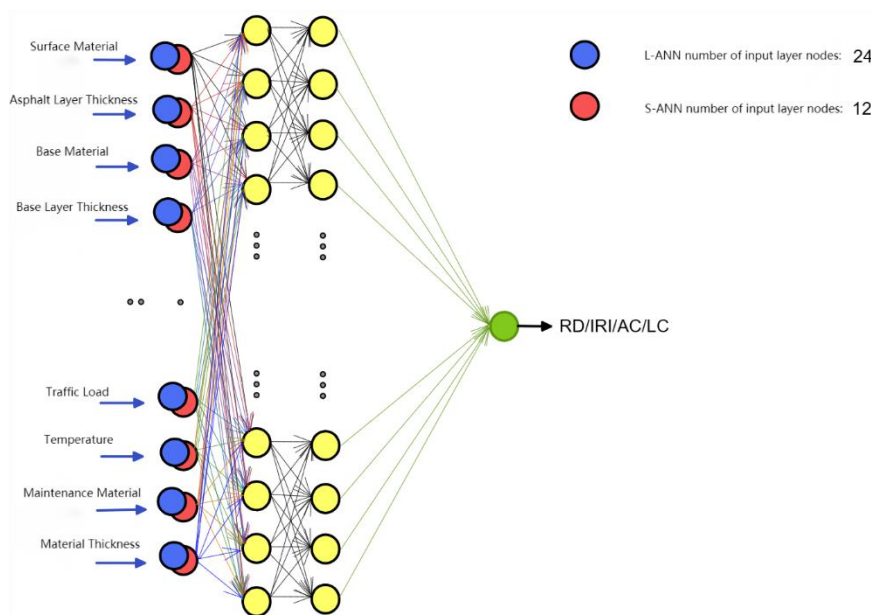
Figure 6. ANN prediction model architecture.

3.2.2. Transformer prediction model

The heatmap visually represents the strength of correlations between attributes, with darker colors indicating stronger correlations. From Figure 4, it can be observed that some input nodes exhibit strong correlations, while others have minimal or no correlations. The self-attention mechanism can be employed to capture the correlations between input variables. In this context, it captures the dependencies between elements in the sequence. By computing the attention scores between each element and others, self-attention can capture the dependencies between elements in the input sequence. These dependencies can be local or global in nature. By capturing these dependencies, the model gains a better understanding of the input features and improves its performance.

Taking into account the need to capture relationships among multiple features, a Transformer-based neural network architecture is defined for regression. The network comprises an Encoder module and a Multi-Layer Perceptron (MLP) module, with the self-attention mechanism serving as a feature extractor and the MLP functioning as the predictor. The L-ANN model is tested using the same Transformer architecture with identical structure.

The Transformer model is built using the TensorFlow framework, as shown in Figure 7.

1) The input section consists of 24 columns of data, of which 8 columns are standardized. Non-numeric values are replaced with numerical values through data imputation. The processed information is then converted into one-hot encoded features, resulting in a total of 47 dimensions. Initially, the input data is serialized by transforming the vector data into a diagonal matrix of size [47, 47]. The pavement information variables are filled with the value of 1 in the corresponding positions, while non-numeric and other variables are filled with 0. This matrix is used as the input for the neural network.

2) The sub-modules are integrated to construct a complete Transformer model. The designed Transformer model consists of two identical stacked sub-modules. The Encoder module primarily consists of a self-attention mechanism layer and an MLP module consisting of fully connected layers. Residual connections and normalization operations are applied within the Encoder module. The MLP in the Encoder module comprises a convolutional1D layer (conv1d), followed by a flattening operation. The flattened tensor passes through two fully connected layers (fc1 and fc2) with ReLU activation functions. The number of fully connected layers in the MLP remains consistent with the second part, i.e., two layers with 64 neurons in each hidden layer. Dropout is applied between the two fully connected layers. Following the Encoder module, the data proceeds to another MLP module, consisting of 16 neurons and one neuron, to transform the data dimensions into 1D for the output of pavement surface performance indicators.

Overall, this architecture is a variant of the Transformer model, where the Encoder layer incorporates self-attention and an MLP layer for the fully connected neural network. The Transformer model stacks such blocks and applies fully connected layers at the top for the final prediction.

Table 5 presents the parameters of the Transformer model. A single-head self-attention mechanism is chosen, with the dimensions of the weight matrices K, Q, and V set to $dk = 47$. The Transformer sub-module count is set to 2, and the number of nodes in the fully connected layers is set to 64. The dropout parameter is set to 0.1, which randomly removes a portion of the neurons to prevent overfitting in the fully connected layers. The model is trained for 1000 epochs using the Adam optimizer as the training optimizer. The ReLU activation function is applied to the fully connected layers. The training batch size is set to 32. In the experiment, the data is divided in an 8: 2 ratio, with the first 80% of the data used as the training set for model training and the remaining 20% used as the validation set to assess the training results. The mean squared error (MSE) is utilized as the model's loss function.

Table 5. Transformer model parameter settings.

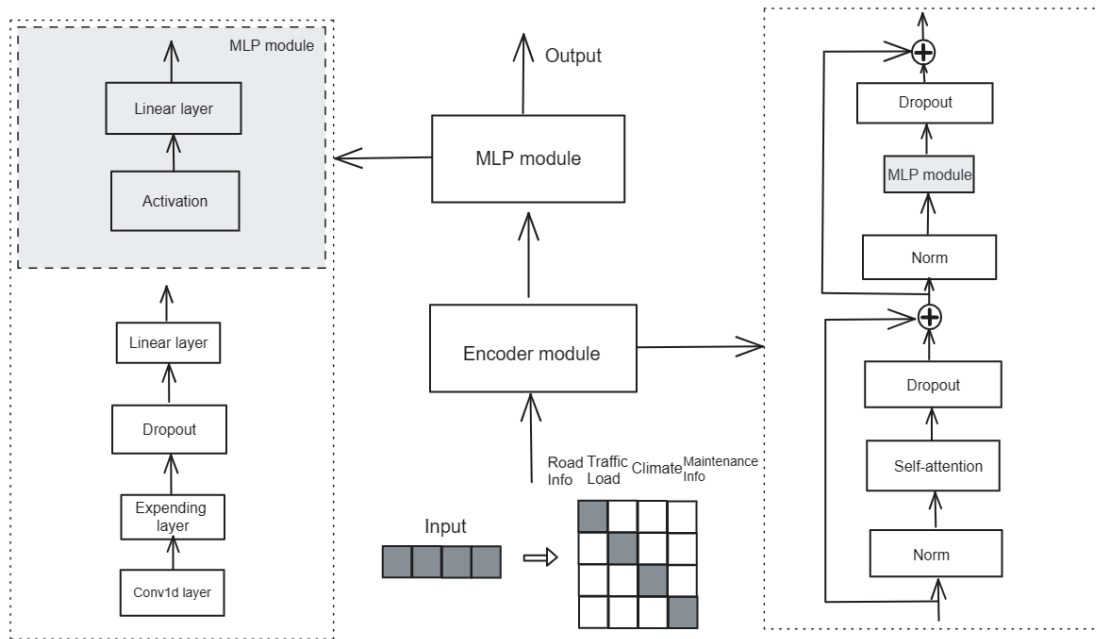| Parameter name	Parameter value	Parameter name	Parameter value
Training Set:Test Set	8: 2	Optimizer	Adam
Loss Function	MSE	Dropout Rate	0.1
Weight Matrix Dimension (dk)	47	Number of Training Epochs	1000
Activation Function	ReLU	Number of Nodes in Fully Connected Layer	64
Number of Submodules	2	Batch Size	32

Figure 7. Transformer prediction model architecture.

3.3. Research process and criteria for assessing model performance

We focus on the prediction of pavement performance, considering the continuous nature of the predictor variables. Therefore, it can be regarded as a nonlinear regression problem. To establish a training dataset for model construction, we incorporate comprehensive Long-Term Pavement Performance (LTPP) data.
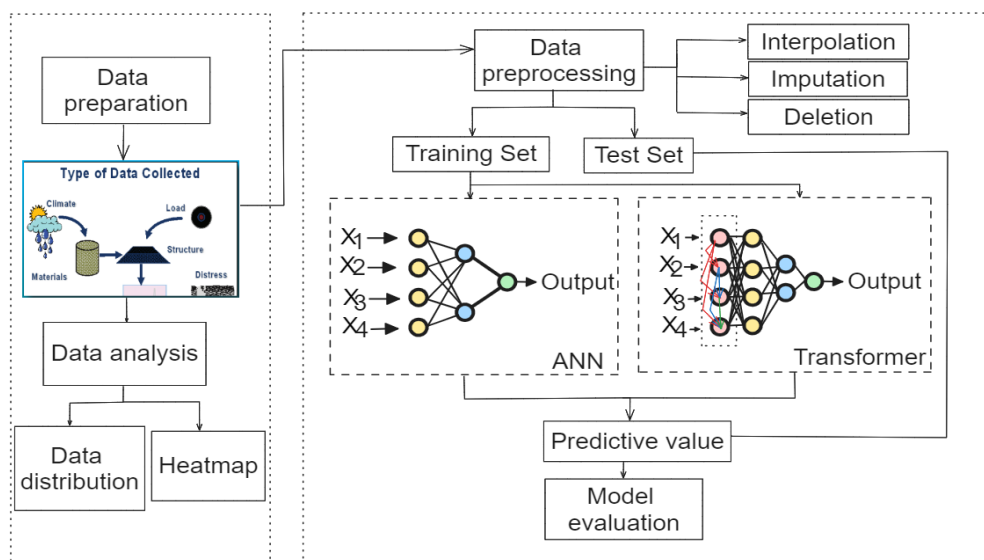

Figure 8. Technological approach.

In this research, the predictive capabilities of artificial neural network models with different input nodes are compared with those of the Transformer model. The objective is to explore the application

of the Transformer model in predicting asphalt pavement performance, aiming to provide more accurate predictions and deeper insights. Figure 8 illustrates the proposed research workflow, showcasing the comprehensive utilization of LTPP pavement performance data.

To quantitatively evaluate the predictive accuracy of the model and facilitate the comparative analysis of different machine learning models, several evaluation metrics are employed in this study. Specifically, the Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R-Square) are utilized to assess the performance of the machine learning models. The Mean Absolute Error (MAE) is employed to measure the difference between the predicted values and the actual values, while the Root Mean Square Error (RMSE) is utilized to enhance the sensitivity of the evaluation metric to large or small errors. The first two metrics focus on measuring the proximity between the predicted and actual values, where lower values indicate higher predictive accuracy. On the other hand, R-Square assesses the predictive ability of the model relative to the actual values, with larger values indicating higher predictive accuracy (with a maximum value of 1).

In summary, the selection of these three-evaluation metrics provides a comprehensive and accurate assessment of the performance of the neural network model. This approach ensures not only the evaluation of the accuracy of the model's predictions but also the assessment of its ability to capture the relationship between the predicted and actual values. The Eqs (9)–(11) for these metrics are as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{12}(y_i - \hat{y}_i)^2} \tag{9}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|^2 \tag{10}$$

$$R2 = 1 - \frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{\sum_{i=1}^{n}(\overline{y}_i - y_i)^2} \tag{11}$$

where, n is the number of samples in the pavement performance dataset; y_i is the measured value of pavement performance indicator; $\hat{y}_i$ is the predicted value of pavement performance indicator; $\overline{y}_i$ is the average value of pavement performance indicator.

4. Results and analysis

4.1. Validation and evaluation of the ANN prediction model

To validate the feasibility and effectiveness of the proposed models in this study, the performance of the S-ANN and L-ANN models in predicting pavement surface performance was compared using the same dataset. The evaluation metrics introduced in Eqs (9)–(11) were employed to quantitatively compare the performance of the machine learning models proposed in this study.

Figures 9 and 10 depict the training results of the S-ANN and L-ANN models for the prediction of rutting, smoothness, cracking, and longitudinal cracking. The comparison is based on the training and testing losses for 1000 epochs. From the figures, it can be observed that the losses for rutting and cracking reach almost stable values relatively quickly, indicating successful model training. For example, in the S-ANN model, the RD and IRI models both show a sharp decrease in losses to around 0.05 within the first 100 epochs, with the best model achieved before the 100th epoch, where the testing

loss reaches approximately 0.06. The behavior of the AC model is similar to the IRI model. However, the LC model takes more time to converge, stabilizing at losses of around 1.6 and 1.9 after 700 epochs, possibly due to specific data handling techniques used to address missing data. The loss values for the S-ANN and L-ANN models are presented in Table 6.

Table 6. ANN model Loss function.

Model type		loss	
		Train set	Test set
S-ANN	RD model	0.0324	0.0625
	IRI model	0.0376	0.0655
	AC model	0.0675	0.1179
	LC model	1.6099	1.9154
L-ANN	RD model	0.0582	0.1000
	IRI model	0.0068	0.0073
	AC model	0.0104	0.0175
	LC model	0.0568	0.1283

(a)   RD train and validation loss

(b)   IRI train and validation loss

(c)   AC train and validation loss

(d)   LC train and validation loss
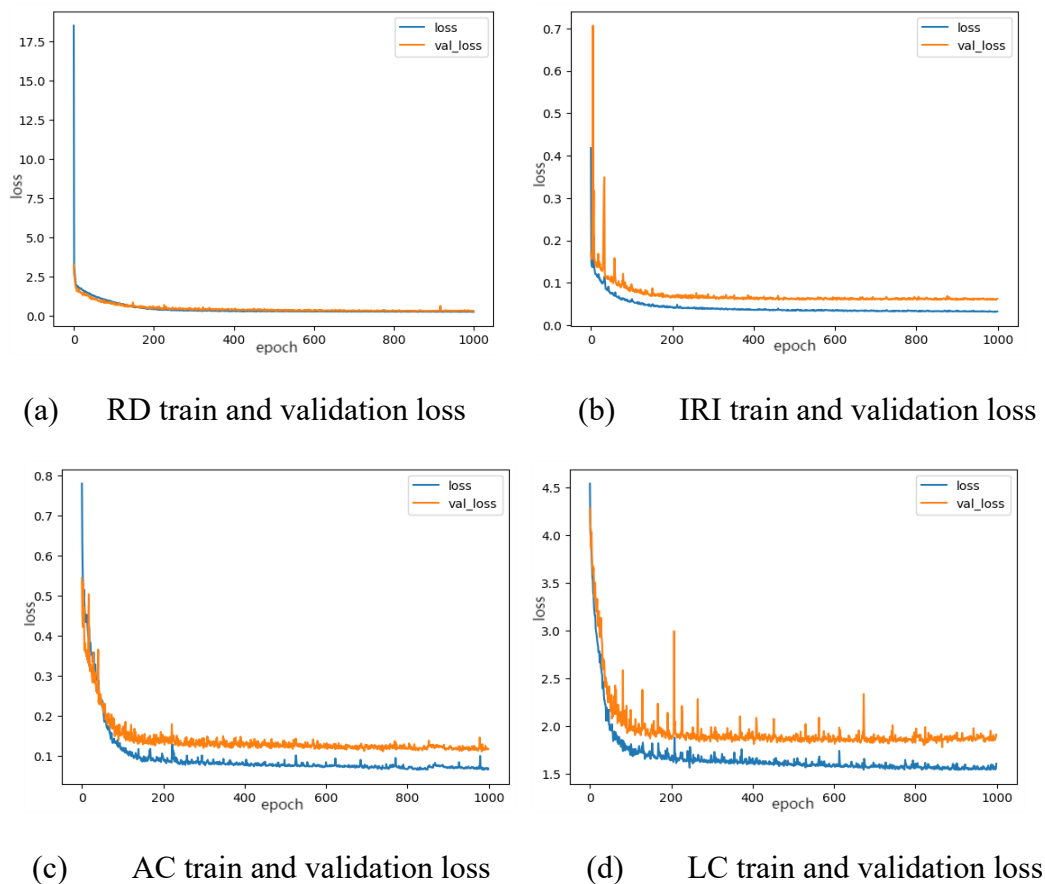
Figure 9. S-ANN model Loss function diagram.

(a) RD train and validation loss

(b) IRI train and validation loss

(c) AC train and validation loss

(d) LC train and validation loss
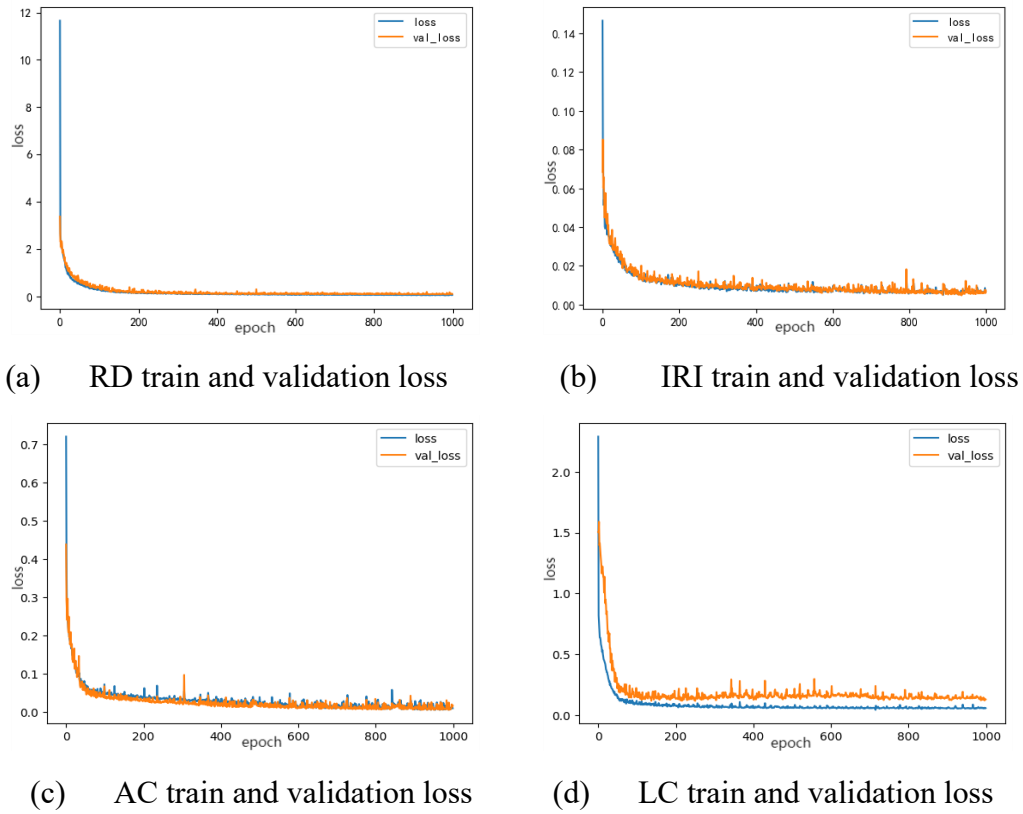
Figure 10. L-ANN model Loss function diagram.

(a) Predict RD scatter plots

(b) Predict IRI scatter plots

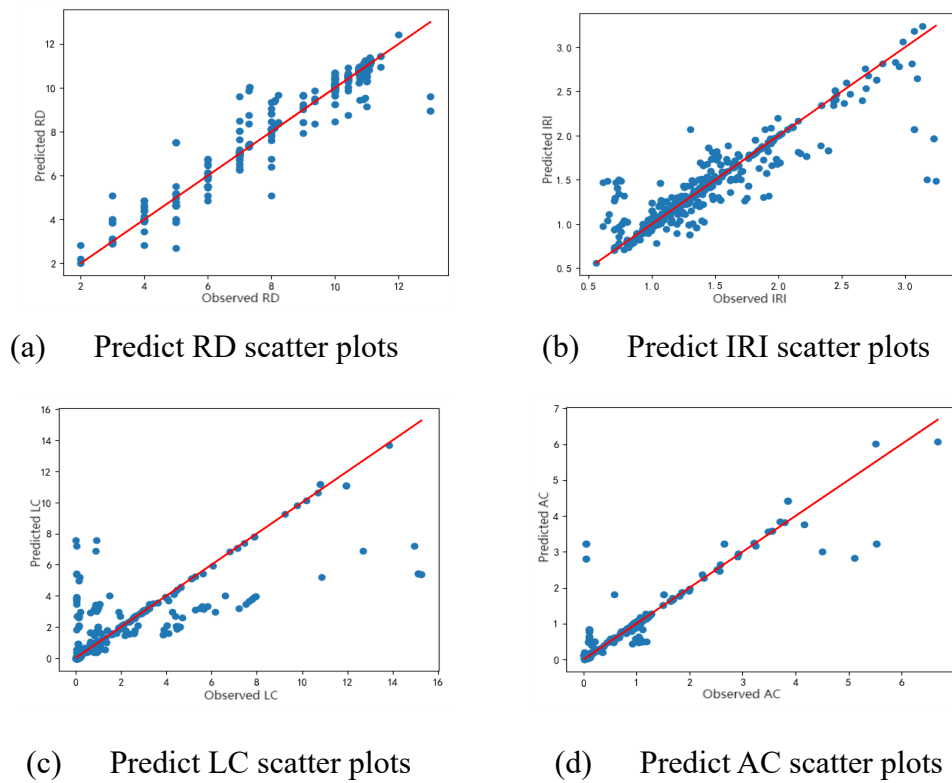(c) Predict LC scatter plots

(d) Predict AC scatter plots

Figure 11. S-ANN model scatter plots.
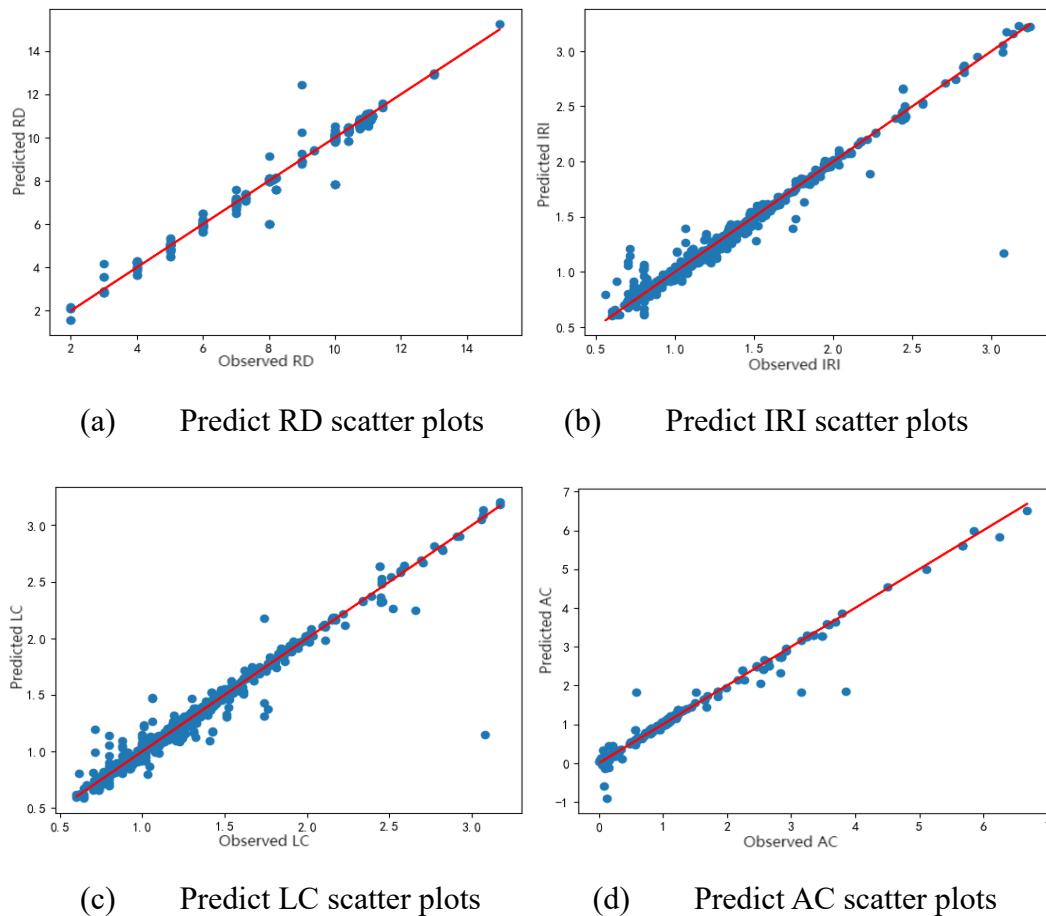
Figure 12. L-ANN model scatter plots.

Figures 11 and 12 present scatter plots of the predicted pavement surface performance indicators versus the target indicators on the testing dataset. The L-ANN model exhibits a favorable distribution of points around the diagonal line y = x, indicating a strong alignment between the proposed L-ANN model and the actual pavement surface performance. Most of the predicted performance values closely align with the target values, demonstrating superior predictive accuracy compared to the other S-ANN models.

Table 7 presents the RMSE, MAE, and R2 values for the two prediction models, S-ANN and L-ANN. Due to the presence of missing data, the LC and IRI models show relatively poorer performance. It can be observed that the L-ANN model exhibits stronger learning and generalization capabilities than the S-ANN model on both the testing and training datasets, resulting in higher levels of acceptable accuracy in predicting pavement surface performance indicators. For instance, the RD model in the S-ANN model has an RMSE of 0.2496 and an MAE of 0.2614 on the testing dataset, while in the L-ANN model, the RD model has an RMSE of 0.1862 and an MAE of 0.1132 on the training dataset, demonstrating superior learning capabilities compared to the S-ANN model. Overall, the L-ANN prediction model demonstrates significantly improved generalization capabilities. For the RD model, the L-ANN model achieves a reduction of 25.4% in RMSE and a decrease of 49.1% in MAE on the testing dataset compared to the S-ANN model. Furthermore, the R2 value increases by 8.8% for the L-ANN model. Similar improvements are observed for the other three performance indicator models, indicating that the predicted values of the L-ANN model closely approximate the actual values.

Table 7. Summary of evaluation indicators.

| Model type		MAE		RMSE		R-square	
		Train set	Test set	Train set	Test set	Train set	Test set
S-ANN	RD model	0.2261	0.2614	0.1740	0.2496	0.8556	0.8374
	IRI model	0.2845	0.2968	0.4763	0.5078	0.8298	0.7364
	AC model	0.2705	0.2894	0.2540	0.3434	0.8296	0.8111
	LC model	0.4890	0.5499	1.3436	1.3840	0.6465	0.6337
L-ANN	RD model	0.0867	0.1132	0.1595	0.1862	0.9288	0.9186
	IRI model	0.1093	0.1221	0.3208	0.3568	0.8915	0.8519
	AC model	0.0804	0.1035	0.1859	0.2459	0.9202	0.9073
	LC model	0.0903	0.1437	0.2261	0.3582	0.8873	0.8744

4.2. Validation and evaluation of the Transformer prediction model

To validate the feasibility and rationality of the pavement surface performance prediction using the L-ANN and Transformer models proposed in this study, a comparative analysis was conducted using the same set of sample data. Figure 13 illustrates the training and validation loss for each model at every training epoch, showing rapid changes followed by convergence and stabilization. It is worth noting that the models exhibit early convergence and achieve low loss values. As shown in Figure 7, the scatter plot of the predicted pavement surface performance indicators versus the target indicators for the Transformer model shows a closer alignment compared to the L-ANN model. This suggests that the Transformer model effectively captures nonlinear relationships, resulting in superior model fitting after training.
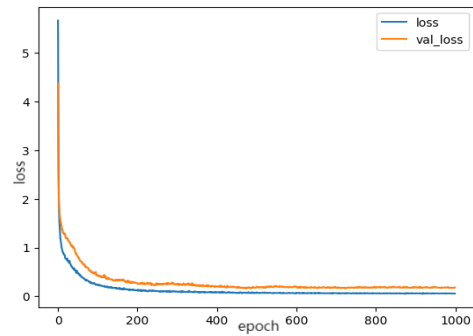

(a) RD train and validation loss (b) IRI train and validation loss
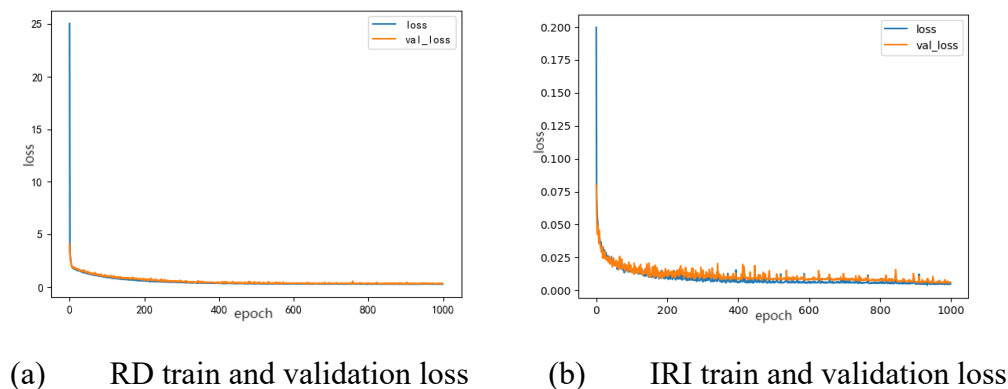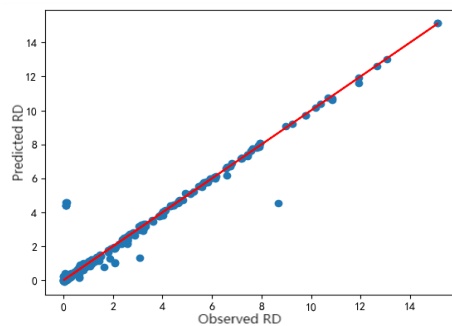
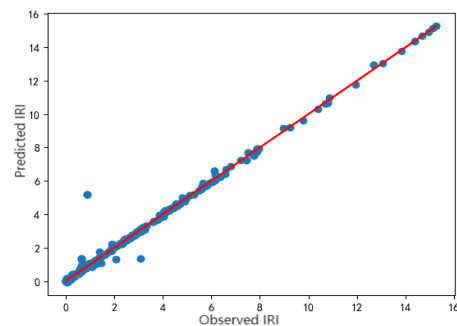**Figure 13.** Transformer model Loss function diagram.
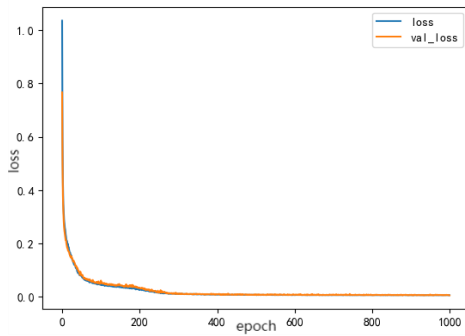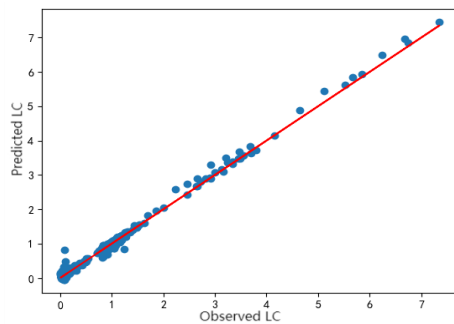
(c)   AC train and validation loss   (d)   LC train and validation loss

Figure 13. Transformer model Loss function diagram.
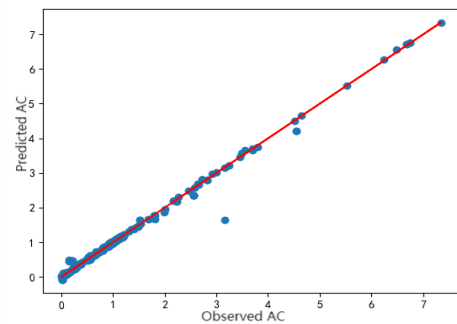

(a)   Predict RD scatter plots   (b)   Predict IRI scatter plots

(c)   Predict LC scatter plots   (d)   Predict AC scatter plots

Figure 14. Transformer model scatter plots.

Table 8 presents the relative importance scores of the RD, IRI, AC, and LC model indicators predicted by the two models. It can be observed that the Transformer model outperforms the fully connected neural network (L-DNN) in terms of performance prediction. Using the evaluation metrics proposed in Eqs (9)–(11), the performance of the deep learning models proposed in this study is further quantitatively compared. The MAE, RMSE, and R2 values of different prediction models are provided, taking into account the presence of missing data in LC and IRI, which contributes to relatively poorer performance in these two prediction categories, aligning with the nature of the data itself. The Transformer model demonstrates stronger learning and generalization abilities than the L-ANN model on both the training and testing sets, resulting in acceptable accuracy in pavement surface performance

prediction. Overall, the prediction accuracy on the training set is generally higher than that on the corresponding testing set. Specifically, the R2 values for L-ANN on the testing set are 0.9186, 0.8519, 0.9073, and 0.8744, while the R2 values for the Transformer model on the testing set are 0.9653, 0.9718, 0.9738, and 0.9642, all higher than those of the L-ANN model. Additionally, the other two performance indicators also surpass those of the L-DNN model. Therefore, the Transformer model exhibits superior learning ability and generalization performance in the prediction task.

Table 8. Summary of evaluation indicators.

Model type		MAE		RMSE		R-square	
		Train set	Test set	Train set	Test set	Train set	Test set
L-ANN	RD model	0.0867	0.1132	0.1595	0.1862	0.9288	0.9186
	IRI model	0.1093	0.1221	0.3208	0.3568	0.8915	0.8519
	AC model	0.0804	0.1035	0.1859	0.2459	0.9202	0.9073
	LC model	0.0903	0.1437	0.2261	0.3582	0.8873	0.8744
Transformer	RD model	0.0799	0.1215	0.1883	0.2985	0.9931	0.9653
	IRI model	0.0239	0.0260	0.0697	0.0798	0.9871	0.9718
	AC model	0.0209	0.0274	0.0724	0.0728	0.9942	0.9738
	LC model	0.0625	0.0970	0.2022	0.2355	0.9794	0.9642

The variation in model prediction results can be attributed to the use of the Transformer architecture, which combines self-attention mechanism and multi-layer perceptron (MLP) networks for regression prediction tasks. The self-attention mechanism enables the extraction of relevant features from the input data, while the MLP maps the output of the self-attention mechanism to the prediction target for forecasting. During the training process, the model automatically adjusts the parameters of the self-attention mechanism and MLP to minimize the discrepancy between the predicted results and the actual results.

5. Discussions and conclusions

After considering the impact of different input variables on the prediction model, we propose a method based on the Transformer architecture to improve the data quality of LTPP (Long-Term Pavement Performance) pavement performance prediction. Our aim is to enhance the quality of LTPP pavement performance prediction data. Several conclusions can be drawn as follows:

1) Based on the pavement structure features, weather features, traffic parameter features, and pavement performance in the LTPP database, with asphalt pavement performance indicators as the

research objective, a neural network model, S-ANN and L-ANN, is constructed with 12 and 24 input variable nodes, respectively. A comparison reveals that selecting a larger number of appropriate input variables can effectively improve the prediction accuracy of pavement performance. By considering all the major factors that influence pavement performance, the models have significant potential for accurately predicting pavement condition indicators. The selection of the S-ANN architecture is based on a preliminary analysis of pavement performance data, revealing that certain performance attributes exhibit relatively straightforward relationships and patterns. The compact structure of S-ANN is well-suited for capturing these direct relationships, effectively identifying fundamental patterns within the data. In pavement performance studies, there may be limitations in data, such as restricted environmental information or details about pavement infrastructure. In such situations, the simplified structure of S-ANN has proven to be robust under limited data conditions. On the other hand, the choice of the L-ANN architecture considers a more comprehensive dataset. In pavement performance prediction, a wealth of foundational and environmental information with potential interdependencies is encountered. Employing a deep model like L-ANN enhances the capacity to capture complex relationships and nonlinear features, thereby improving the model's fitting capability to the dataset. L-ANN's larger architecture, featuring 24 input variable nodes, provides a higher capacity to capture intricate and nonlinear patterns within the data. This is crucial in pavement performance modeling where the relationships between various factors can be complex and multifaceted.

2) To further improve the performance and interpretability of the prediction model, we introduce the Transformer algorithm. Traditional neural networks have failed to consider interactions between input variables, and the presence of influences between factors can lead to poor network performance or overfitting. The Transformer model structure, consisting of self-attention mechanism, normalization, and residual connections with L-DNN, is utilized to predict pavement performance. Through a comparative analysis of evaluation metrics, it has been demonstrated that the Transformer model exhibits stronger learning and generalization capabilities. The evaluation metrics on both the training and testing sets outperform those of the L-DNN prediction model. Compared to traditional fully connected neural network models, neural network models incorporating the self-attention mechanism offer greater flexibility and memory capacity. Self-attention enables the model to adaptively focus on different features in the input during the learning process. In the context of pavement performance, certain environmental conditions or road structures may have a more significant impact on performance. Self-attention allows the model to autonomously adjust its focus on specific features in different contexts, improving its understanding of the complexity of pavement conditions. Pavement conditions may be influenced by various factors, some of which may introduce uncertainty. Self-attention assists the model in handling changes and uncertainties in the input data, enhancing the model's adaptability to complex and dynamic environments. However, the self-attention mechanism may also introduce some challenges. First, it can increase the complexity of the model, requiring more computational resources. Additionally, since the self-attention mechanism processes different parts of the input data, it may necessitate a larger training dataset to learn the model parameters. Last, the self-attention mechanism, involving multiple input data interactions, may require longer training time. The attention mechanism can be effectively integrated into neural networks to enhance the performance and interpretability of regression prediction models. The proposed approach, based on the fused Transformer architecture, demonstrates superior prediction of LTPP pavement performance and holds potential for significant practical applications.

3) Predicting pavement performance is an essential component of pavement maintenance

management. The utilization of a pavement performance prediction model based on the Transformer algorithm enables efficient and accurate forecasting of future changes in pavement performance during pavement usage. This enhances the scientific basis of maintenance decision-making. An accurate prediction model provides more reasonable guidance for the planning of asphalt pavement maintenance and the allocation of maintenance funds.

Use of AI tools declaration

The authors declare that they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Conflict of interest

The authors declare there is no conflict of interest.

References

1. H. Yao, Y. Liu, X. Li, Z. You, Y. Feng, W. Lu, A detection method for pavement cracks combining object detection and attention mechanism, *IEEE Trans. Intell. Transp. Syst.*, **23** (2022), 22179–22189. https://doi.org/10.1109/TITS.2022.3177210

2. H. Yao, Y. Liu, H. Lv, J. Huyan, Z. You, Y. Hou, Encoder-decoder with pyramid region attention for pixel-level pavement crack recognition, *Comput.-Aided Civil Infrastruct. Eng.*, 2023. https://doi.org/10.1111/mice.13128

3. M. Zhu, K. Bi, H. Yu, X. Liu, W. Qiao, Research on pavement structure and performance prediction based on long-life test sectio, *J. Muni. Tech.*, **41** (2023), 58–65. https://doi.org/10.19922/j.1009-7767.2023.05.058

4. Y. Jin, T. Shen, Y. Tang, L. Li, Pavement performance prediction based on toll date: Case study in Shaanxi Province, China, in *18th COTA International Conference of Transportation Professionals*, **13** (2018), 545–554. https://doi.org/10.1061/9780784481523.055

5. J. Karam, H. Noorvand, Developing a rutting prediction model for HMA pavements using the LTPP database, *Int. J. Pavement Res. Technol.*, 2023. https://doi.org/10.1007/s42947-023-00340-9

6. Z. Hen, X. Sun, W. Yang, Q. Li, G. Xiao, S. Xu, Performance prediction model and accuracy analysis of asphalt pavement of sixth ring freeway in Beijing, *J. Muni. Tech.*, **38** (2020), 34–36.

7. K. A. Abaza, Back-calculation of transition probabilities for Markovian-based pavement performance prediction models, *Int. J. Pavement Res. Technol.*, **17** (2016), 253–264. https://doi.org/10.1080/10298436.2014.993185

8. K. A. Abaza, Simplified Markovian-based pavement management model for sustainable long-term rehabilitation planning, *Road Mater. Pavement Des.*, **24** (2023), 850–865. https://doi.org/10.1080/14680629.2022.2048055

9. L. Yao, Q. Dong, J. Jiang, F. Ni, Establishment of prediction models of asphalt pavement performance based on a novel data calibration method and neural network, *Transp. Res. Rec.*, **2673** *(*2019), 66–82. https://doi.org/10.1177/0361198118822501

10. A. A. Ali, U. Heneash, A. Hussein, S. Khan, Application of Artificial neural network technique for prediction of pavement roughness as a performance indicator, *J. King Saud Univ. Sci.*, 2023. https://doi.org/10.1016/j.jksues.2023.01.001

11. Y. Zhu, J. Chen, K. Wang, Y. Liu, Y. Wang, Research on performance prediction of highway asphalt pavement based on Grey–Markov model, *Transp. Res. Rec.*, **2676** (2021), 194–209. https://doi.org/10.1177/03611981211057527

12. Y. Song, YD. Wang, X. Hu, J. Liu, An efficient and explainable ensemble learning model for asphalt pavement condition prediction based on LTPP dataset, *IEEE Trans. Intell. Transp. Syst.*, **23** (2022), 22084–22093. https://doi.org/10.1109/TITS.2022.3164596

13. G. Liu, F. Niu, Z. Wu, Life-cycle performance prediction for rigid runway pavement using artificial neural network, *Int. J. Pavement Res. Technol.*, **21** (2020), 1806–1814. https://doi.org/10.1080/10298436.2019.1567922

14. K. Othman, Prediction of the hot asphalt mix properties using deep neural networks, *Beni-Suef Univ. J. Basic Appl. Sci.*, **11** (2022), 40. https://doi.org/10.1186/s43088-022-00221-3

15. Q. Dong, X. Chen, S. Dong, Classification of pavement climatic regions through unsupervised and supervised machine learnings, *J. Infrastruct. Preserv. Resilience*, **2** (2021), 5. https://doi.org/10.1186/s43065-021-00020-7

16. Z. Sun, X. Hao, W. Li, J. Huyan, H. Sun, Asphalt pavement friction coefficient prediction method based on genetic-algorithm-improved neural network (GAI-NN) model, *Can. J. Civ. Eng.*, **49** (2022), 109–120. https://doi.org/10.1139/cjce-2020-0051

17. M. Mers, Z. Yang, Y. A. Hsieh, Y. Tsai, Recurrent neural networks for pavement performance forecasting: review and model performance comparison, *Transp. Res. Rec.*, **2677** (2022), 610–624. https://doi.org/10.1177/03611981221100521

18. M. I. Hossain, L. S. P. Gopisetti, M. S. Miah, Prediction of international roughness index of flexible pavements from climate and traffic data using artificial neural network modeling, *Airfield Highw. Pavements 2017*, (2017), 256–267. https://doi.org/10.1061/9780784480922.023

19. B. Rulian, Y. Hakan, S. Salma, M. J. Z. Fanhmi, N. Yacoub, Performance model development for flexible pavements via neural networks, in *International Conference on Transportation and Development*, **5** (2022), 73–84. https://doi.org/10.1061/9780784484357.007

20. J. Liu, F. Liu, C. Zheng, E. Fanijo, L. Wang, Improving asphalt mix design considering international roughness index of asphalt pavement predicted using autoencoders and machine learning, *Constr. Build. Mater.*, **360** (2022), 129439. https://doi.org/10.1016/j.conbuildmat.2022.129439

21. M. M. Radwan, M. A. Abo-Hashema, H. P. Faheem, M. D. Hashem, ANN-based fatigue and rutting prediction models versus regression-based models for flexible pavements, in *3rd GeoMEast International Congress and Exhibition on Sustainable Civil Infrastructures*, (2020), 117–133. https://doi.org/10.1007/978-3-030-34196-1_9

22. M. Mahmood, U. Anuraj, S. Mathavan, M. Rahman , A unified artificial neural network model for asphalt pavement condition prediction, *Proc. Inst. Civil Eng.-Transp.*, **176** (2023), 14–24. https://doi.org/10.1680/jtran.19.00111

23. Z. Luo, S. Li, An interpretable prediction model for pavement performance prediction based on XGBoost and SHAP, in *Second International Conference on Electronic Information Engineering and Computer Communication (EIECC 2022)*, **12594** (2022), 187–194. https://doi.org/10.1117/12.2671361

24. A. L. Aranha, L. L. B. Bernucci, K. L. Vasconcelos, Effects of different training datasets on machine learning models for pavement performance prediction, *Transp. Res. Rec.*, **2677** (2023), 196–206. https://doi.org/10.1177/03611981231155902

25. O. Kaya, H. Ceylan, S. Kim, D. Waid, B. P. Moore, Statistics and artificial intelligence-based pavement performance and remaining service life prediction models for flexible and composite pavement systems, *Transp. Res. Rec.*, **2674** (2020), 448–460. https://doi.org/10.1177/0361198120915889

26. M. Xiao, R. Luo, Y. Chen, X. Ge, Prediction model of asphalt pavement functional and structural performance using PSO-BPNN algorithm, *Constr. Build. Mater.*, **407** (2023), 133534. https://doi.org/10.1016/j.conbuildmat.2023.133534

27. S. Saha, F. Gu, X. Luo, RL. Lytton, Development of an artificial neural network-based k-value prediction model to improve the sensitivity of base layer on rigid pavement performance, *Transp. Res. Rec.*, **2677** (2023), 1290–1308. https://doi.org/ 10.1177/03611981221143114

28. G. Liu, F. Niu, Z. Wu, Life-cycle performance prediction for rigid runway pavement using artificial neural network, *Int. J. Pavement Eng.*, **21** (2020), 1806–1814. https://doi.org/10.1080/10298436.2019.1567922

29. N. Wu, B. Green, X. Ben, Deep transformer models for time series forecasting: The influenza prevalence case, preprint, ArXiv:2001.08317.

30. Q. Zhou, E. Okte, I. L. Al-Qadi, Predicting pavement roughness using deep learning algorithms, *Transp. Res. Rec.*, **2675** (2021), 1062–1072. https://doi.org/10.1177/03611981211023765

31. J. Xin, M. Akiyama, DM. Frangopol, Sustainability-informed management optimization of asphalt pavement considering risk evaluated by multiple performance indicators using deep neural networks, *Reliab. Eng. Syst. Saf.*, **238** (2023), 109488. https://doi.org/10.1016/j.ress.2023.109448

32. S. Salma, Y. HaKan, B. Rulian, N. Jacob, Evaluating the effect of climate change in pavement performance modeling using artificial neural network approach, in *International Conference on Transportation and Development*, **5** (2022), 49–60. https://doi.org/10.1061/9780784484357.005

33. J. Lucey, A. Fathi, M. Mazari, Predicting pavement roughness as a performance indicator using historical data and artificial intelligence, in *International Airfield and Highway Pavements Conference 2019*, (2019), 10–18. https://doi.org/10.1061/9780784482476.002

34. J. Xin, M. Akiyama, DM. Frangopol, M. Zhang, Multi-objective optimisation of in-service asphalt pavement maintenance schedule considering system reliability estimated via LSTM neural networks, *Struct. Infrastruct. Eng.*, **18** (2022), 1002–1019. https://doi.org/10.1080/15732479.2022.2038641

35. A. Das, W. Kong, A. L. Leach, R. Sen, R. Yu, Long-term Forecasting with TiDE: Time-series Dense Encoder, preprint, arXiv:2304.08424.

36. S. Bai, W. Yang, M. Zhang, D. Liu, W. Li, L. Zhou, Attention-based BiLSTM model for pavement temperature prediction of asphalt pavement in winter, *Atmosphere*, **13** (2022), 1542. https://doi.org/10.3390/atmos13091524

37. F. Guo, Y. Qian, Intelligent pavement roughness forecasting based on a long short-term memory model with attention mechanism, *Airfield Highw. Pavements*, (2021), 128–136. https://doi.org/10.1061/9780784483503.013

38. X. Tong, Y. Dong, Y. Zhang, Pavement maintenance plan and practice based on pavement performance prediction, *J. Muni. Tech.*, **39** (2021), 30–34. https://doi.org/10.19922/j.1009-7767.2021.07.028

39. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, *Adv. Neural Inf. Process. Syst.*, **30** (2017). https://doi.org/10.48550/arXiv.1706.03762