

RESEARCH

Open Access



ASENN: attention-based selective embedding neural networks for road distress prediction

Babitha Philip^{1,2†}, Zhenyu Xu^{3†}, Hamad AlJassmi^{1,2*}, Qieshi Zhang³ and Luqman Ali^{2,4,5}

[†]Babitha Philip and Zhenyu Xu have contributed equally to this work.

*Correspondence: h.aljassmi@uaeu.ac.ae

¹ Department of Civil and Environmental Engineering, United Arab Emirates University, Al Ain 15551, United Arab Emirates

² Emirates Center for Mobility Research (ECMR), United Arab Emirates University, Al Ain 15551, United Arab Emirates

³ CAS Key Laboratory of Human-Machine Intelligence-Synergy Systems, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, No. 1068 Xueyuan Avenue Shenzhen University Town, Shenzhen, China

⁴ AI and Robotics Lab (Air-Lab), United Arab Emirates University, Al Ain 15551, United Arab Emirates

⁵ College of IT, United Arab Emirates University, Al Ain 15551, United Arab Emirates

Abstract

This study proposes an innovative neural network framework, ASENN (Attention-based Selective Embedding Neural Network), for the prediction of pavement deterioration. Considering the complexity and uncertainty associated with the pavement deterioration process, two fundamental frameworks, SEL (Selective Embedding Layer) and MDAL (Multi-Dropout Attention Layer), are combined to enhance feature abstraction and prediction accuracy. This approach is significant while analyzing the pavement deterioration process due to the high variability of the contributing deterioration factors. These factors, represented as tabular data, undergo filtering, embedding, and fusion stages in the SEL, to extract crucial features for an effective representation of pavement deterioration. Further, multiple attention-weighted combinations of raw data are obtained through the MDAL. Several SELs and MDALs were combined as basic cells and layered to form an ASENN. The experimental results demonstrate that the proposed model outperforms existing tabular models on four road distress parameter datasets corresponding to cracking, deflection, international roughness index, and rutting. The optimal number of cells was determined using different ablation settings. The results also show that the feature learning capabilities of the ASENN model improved as the number of cells increased; however, owing to the limited combination space of feature fields, extreme depths were not preferred. Furthermore, the ablation investigation demonstrated that MDAL can improve performance, particularly on the cracking dataset. Notably, compared with mainstream transformer models, ASENN requires significantly less storage and achieves faster execution speed.

Keywords: Pavement deterioration, Tabular data, Road distress parameters, Prediction models, Deep learning

Introduction

Pavement management systems (PMS) are critical for cost-effectively managing highway networks and optimizing pavement performance during the estimated service life of pavements. Accurate performance prediction modeling is essential for the successful deployment of a PMS that is used to plan future maintenance and rehabilitation activities. Transportation agencies have encountered severe challenges because of the complexity associated with the pavement deterioration process, owing to the highly variable

nature of the elements, which degrade road networks. Furthermore, the global trend of increasing government fund allocation to the health and education sector has depleted highway budgets [1]. Consequently, many governments have adopted road user charges, such as tolls and fuel taxes, as a source of highway funding. For instance, the ASCE 2021 Infrastructure Report Card revealed that 43% of public roads in the United States (US) are in poor or mediocre conditions, demonstrating the adverse effects of road deterioration. These deteriorated roads annually cost almost \$130 billion for vehicle operating and repair expenses [2]. Despite various attempts, highway agencies still face a significant reduction in budget for operating road networks. Numerous efforts have been made to develop prediction models, including deterministic, probabilistic, and artificial neural network (ANN) models. The advantages of machine-learning (ML) methods in analyzing the pavement condition involves robust learning algorithms, enhanced performance, the ability to handle massive dataset, generalization ability, and so on [3]. The increase in the number of studies applying neural networks in various areas of pavement engineering, such as pavement design, construction, monitoring, and maintenance of pavements, suggests that these developed models can solve a variety of pavement-related issues [4].

Tabular models are excellent for deterioration prediction because roads, environment, and traffic data are frequently maintained in a tabular format and can easily extract rich semantic information. Tabular models commonly include both tree-based models [5–8] and deep learning (DL) models [9–11]. Tree-based models are known for their interpretability and are effective in handling categorical data, as they can display the prediction results through node splits. However, these models struggle to manage tabular data with continuous variables. Uncovering the relationships among continuous variables requires tree models to have a larger depth, often leading to overfitting and decreasing predictive performance. Recently, DL architectures have shown promising performance in various applications, such as image and natural language processing. DL architectures are designed to generate inductive biases that match data invariances and spatial dependencies. Finding comparable invariances in tabular data with varied properties, limited sample numbers, and extreme values is challenging. Desirable performance is achieved by a large number of parameters of the model, and no general and efficient components are designed to extract the related features from tabular data. Thus, the tabular data-processing abilities of deep networks have not yet been completely explored.

To solve the aforementioned problem, this study proposes a model with two special layers, a selective embedding layer (SEL) and a multi-dropout attention layer (MDAL). It combines them to form basic cells, which are then stacked to build attention-based selective embedding neural networks (ASENNs) for pavement deterioration prediction. The SEL has three components, filter, embedding, and fusion, to map input data to a more abstract feature space for efficient prediction. The operation of the layers starts by applying a feature selection mask to the input data to identify important factors, followed by type-wise embedding of the masked input data to exploit intra-information. Finally, the masked input data and embedding are fused to obtain more abstract representations. The MDAL consists of multiple identical structures, each of which includes a dropout layer followed by an attention module. The dropout layer obtains different combinations of feature fields by randomly dropping some input data fields. Motivated by SENets [12], the attention module is introduced to model the relationships between

different combinations of feature fields explicitly. Therefore, different attention-weighted combinations of input fields can be obtained using MDAL. To extract the embedded and attention-weighted information from the SEL and MDAL, several SELs and MDALs were combined as a basic cell, which uses two different paths to increase the diversity of features. One path includes an SEL and MDAL to complement the raw data information, and the other consists of two SELs to enhance the abstraction level of the current input. Multiple basic cells were stacked to build ASENNs for obtaining meaningful and abstractive features. As the number of cells increased, the feature learning capabilities and prediction accuracies improved. In this study, four ASENNs were developed corresponding to four major road distress parameters, namely cracking, deflection, international roughness index (IRI), and rutting, which are based on data acquired from road networks in the United Arab Emirates (UAE), from the Ministry of Energy and Infrastructures.

The remaining paper is organized as follows. Section "[Literature review](#)" discusses numerous studies that used neural networks in pavement performance analysis, and recent developments in the field of deep neural networks. Section "[Data collection and preprocessing](#)" describes the data and several preprocessing steps. Section "Methodology" includes a problem statement, followed by an explanation of the methodology adopted in this study. Section "[Development of Neural Network-based pavementdeterioration prediction model](#)" discusses the architecture adopted for model development. The results of the analysis are discussed in Sect. "[Results and discussion](#)", and the conclusions of the study are presented in Sect. "[Conclusion](#)".

Literature review

A significant portion of the existing literature on tabular models focuses on tree models and deep-learning networks. Decision-tree models, which have a high level of validity and interpretability in the decision-making process, can illustrate clear decision paths. Recently, various decision tree-based ensemble models, such as GBDT [13] and its variations, namely, XGboost [6], LightGBM [7], and Catboost [8], have become the tools of choice for tabular data mining. The main difference between these variations is tree symmetry. For asymmetric trees, LightGBM grows leafwise, requiring less execution time than XGBoost that grows levelwise. CatBoost minimizes the loss for all nodes at the same depth for symmetric trees, which can increase the computational speed and avoid overfitting. Although the variations have gaps in implementing the details, the gap in the overall performance is diminutive. In deep learning networks, multilayer perceptron (MLP) has been widely used for tabular datasets and has outperformed other developed models. With the development of DL, researchers have introduced various modules, such as residual and attention represented by ResNet [14] and transformer [15], for improving the structure to work with tabular data. For instance, a feature tokenizer is integrated into the transformer, making transformer-based models adaptive in the tabular domain [10]. Similarly, sequential attention was used to select the required features, demonstrating interpretability [9]. These transformer-based models can handle categorical features better than ensemble decision tree models and MLP, as they can map the features to a high-dimensional space and then use attention modules to obtain the connection among the features.

DL methods have been adopted in pavement studies in recent years. Ma et al. [16] proposed a method for pavement crack detection and tracking based on a library of pavement crack images through data augmentation using PCGAN (Pavement Crack Generative Adversarial Network). These images are further utilized to train the YOLO v3 Convolutional Neural Network model, optimizing hyperparameters through multiple training iterations for crack detection with an accuracy of 98.47%. Considering the ability of CNNs to handle spatial data efficiently, Jiang et al. [17] developed a framework to assess the risk of urban road collapse by combining several environmental and anthropogenic factors. The model employed the Synthetic Minority Over-Sampling Technique (SMOTE) for data augmentation. The CNN model thus developed identified the roads at high risk with an accuracy of 97%, and the study emphasized the importance of considering external factors such as environmental factors while analyzing the pavement data. The features of DL models over several existing ML models and image processing (IP) methods involve a data-driven end-to-end learning approach, elimination of hand-crafted features, higher accuracy, increased accessibility, and faster operation [3]. Thus, researchers have made efforts to explore the capability of the DL models in several directions of pavement management.

Haddad et al. [18] developed a rutting prediction model using deep neural network (DNN) techniques based on data obtained from a long-term pavement performance (LTPP) database. The predictive model was compared with state-of-the-art models and outperformed commonly used models in the literature. The generated model was used to assess and rank the relative influence of different factors on rutting, in addition to forecasting pavement rutting. The sensitivity analysis results confirmed the significant impact of traffic and weather conditions on rutting. Furthermore, rutting predictive curves for certain traffic, climate, and performance combinations have been developed to make rutting predictions available to all road agencies. However, the feature selection performed in the data pre-processing step involved removing the variables less correlated with the output to reduce the complexity of the analysis.

Gao et al. [19] proposed convolutional neural network (CNN)-long short-term memory (LSTM), an integrated architecture of the CNN and LSTM models to automatically detect application of maintenance and rehabilitation (M &R) treatment to a pavement area within a particular period and achieved an accuracy of 87.50%. The efficiency of CNN to categorize pavement segments to estimate IRI values from pavement surface images was explored by Abohamer et al. [20], and a comparison study was performed to investigate the efficiencies of an ANN model and multinomial regression (MNL) models based on 850 three dimensional images, which were sourced from the Louisiana DOT and Development (LaDOTD). It was found that the CNN model outperformed the ANN and MDL models during the training phase, achieving an accuracy of 93.4%, a coefficient of determination (R^2) value of 0.985 and an average error of 5.9%. Although, CNNs are considered to be a highly effective approach, these networks demand a considerable amount of computational power and memory resources [21].

Zhou et al. [22] developed an IRI progression model for asphalt concrete (AC) pavement using a recurrent neural network (RNN) algorithm. In addition, an ANN model was used for predicting the IRI drop following maintenance/rehabilitation. The RNN model was trained and tested using time-series data on fatigue cracking, transverse

cracking, rutting, climatic conditions, and traffic factors from the LTPP database. Conversely, the ANN model was inputted with the pavement structure, type of maintenance/rehabilitation activity, degree of maintenance/rehabilitation, and IRI value before maintenance/rehabilitation. The combination of both models enabled the prediction of IRI values over the service life of AC pavements. However, RNNs can be difficult to train [23] and may require longer execution time [24].

The pavement deterioration process being a complex nature, the development of pavement distress prediction models requires large and extensive data of good quality from reliable sources. Pavement data collection has been transformed into automated process with the advancements in technology. However, pavement data analysis is gaining more research focus [25]. The successful application of ANN and DL is remarkable. To the best knowledge of the authors, the current literature focuses on a limited number of distress parameters. Additionally, there is a lack of lightweight, less computationally complex models for the pavement distress prediction. Therefore, this research aims to introduce the ASENN model for predicting pavement deterioration, leveraging selective embedding and attention mechanisms informed by environmental and traffic data. Notably, our proposed model outperforms existing state-of-the-art models in terms of accuracy and efficiency, requiring less memory and achieving faster execution times.

Data collection and preprocessing

The dataset used in the proposed work was acquired from the Ministry of Energy and Infrastructure (MoEI), the highway agency in the United Arab Emirates (UAE). Road data collected and stored by the Road Department at MoEI from 2013 to 2019, along 32 road sections in the country were selected for analysis. The dataset used in this study can be broadly classified into road distress, pavement, traffic, and environmental data. A detailed description of this data is mentioned in a related work [26].

The data was divided into training and testing sets with a ratio of 80:20. An example of the dataset is provided below in Table 1.

Road Distress Data

The major distress parameters considered in this study were (i) cracking (%), (ii) deflection (mm/100), (iii) IRI (m/km), and (iv) rutting (mm). Each of these parameters is measured independently using different measuring devices by the highway agency. Cracking is measured using a laser crack measurement system, deflection using a falling weight deflectometer, IRI using a laser profilometer, and rutting using a laser rutting measurement system. Data related to these parameters from the year 2013 to 2019 are collected for this study.

Initially, the data related to each road distress parameter were combined year-wise, as shown in Table 2 (Before pre-processing). Furthermore, the road distress parameters corresponding to the year of data collection are grouped in ' $t-1$, t ' format for the development of the prediction model. An illustration of this process is presented in Table 2.

Pavement data

Each road network has specific characteristics, which may vary in type, age, and structural properties. In this study, the collected pavement features included the road type

Table 1 Example of each type of data used in the analysis

Input variables		Value
Pavement data	Road type	Arterial
	Maintenance type	Partial
	Age of the road section	1
	Year of data collection	2015
Environmental data	Temperature	26
	Humidity	49
	Atmospheric pressure	1019
Traffic data	Traffic count (heavy vehicles)	4,203,429
	Traffic count (light vehicles)	487,421
	Direction of traffic	Forward
Road distress parameter	Cracking at $t-1$	0
	Deflection at $t-1$	49
	IRI at $t-1$	1
	Rutting at $t-1$	2.674
Label (Output)		Value
Road distress parameter	Cracking at t	0
	Deflection at t	58.8
	IRI at t	1.2
	Rutting at t	3.07625

Table 2 Conversion of road distress data into ' $t-1$, t ' format

Before preprocessing				After preprocessing	
IRI-2013	IRI-2014	IRI-2015	IRI-2016	IRI at $t-1$	IRI at t
2.536	2.894	3.549	4.982	2.536	2.894
				2.894	3.549
				3.549	4.982

(arterial, collector, expressway, and freeway), maintenance type (major, surface, partial), age of the road section, and year of data collection. Among these features, the maintenance type and age of the road section were not directly available from the highway agency database. Hence, these two features were calculated by observing year-wise road distress values. The maintenance type adopted in each road section was determined by comparing the road distress values in consecutive years, and the age of the road section was calculated as the number of years passed from the most recent maintenance. In the major treatment, all road distresses were corrected. However, this type of treatment is relatively expensive. In the surface treatment, surface layers of the pavement are maintained, leaving the deep distresses, especially deflection unmaintained. Partial treatment, on the other hand, is applied to the topmost layers of the pavement, by which IRI, rutting, and shallow cracks are corrected, however the deflection and deep cracks are not treated.

Traffic data

The traffic data included counts for light and heavy vehicles. Vehicles under Class 1–3 and 4 (FHWA vehicle classification) are considered light and heavy vehicles,

respectively. The average number of vehicles passing over a road section in a year was considered for the analysis. Additionally, the direction of traffic flow was noted as ‘Forward’ for the South-North/West-East direction of the traffic flow and ‘Backward’ for the North-South/East–West direction of the traffic flow.

Environment data

Data on the climatic conditions of the road sections were collected using online resources. Environmental data included in this study were humidity (%), temperature (°C), and atmospheric pressure (in millibars). Based on the location of the road section and time of data collection, which is available in the database provided by MoEI, corresponding environmental data were collected from online resources.

Methodology

The overall framework is illustrated in Fig. 1. The pavement, environmental, and traffic data were added to the road distress parameter data (cracking, deflection, IRI, and rutting) at $t-1$, respectively, which were then sent to the ASENNs for predicting the pavement deterioration data at t . SELs and MDALs are the essential components of ASENNs, described in detail below. The data from different sources undergo pre-processing steps such as one-hot encode, normalization, and so on while developing the database.

In this study, we address the problem of pavement deterioration prediction as a regression task. Our objective is to develop a predictive model that accurately estimates the degree of pavement deterioration over time. The goal is to create a reliable mapping function $F(x)$ to predict the deterioration y at t , where x refers to Input variables.

The input data x consists of various features, including both numerical and categorical attributes. These features collectively capture the factors affecting pavement conditions. Numerical features include road characteristics, environmental conditions, and traffic parameters. Categorical features, on the other hand, represent qualitative attributes like road type, construction materials, and maintenance history.

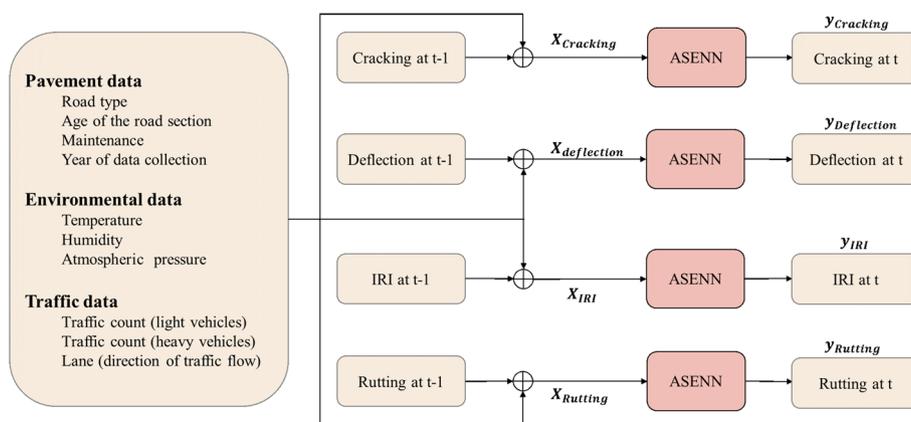


Fig. 1 Overall framework

Selective embedding layer (SEL)

To capture the underlying information from raw data X and obtain abstract features by mining information from a new embedding feature space, we propose SEL, as shown in Fig. 2. Specifically, the SEL comprises the following three components: filter, embedding, and fusion, which determines the important attributes from X , embeds the filtered features X' to obtain embedding features X'' according to their data types and fuses X' and X'' to obtain the abstract features X''' .

Filter

Feature selection was used to extract relevant and useful features for the prediction task. Given the variability of tabular data, numerous approaches for selecting features have been developed. Traditional tree-based models utilize information entropy, information gain rate, and Gini coefficient for selecting features. In contrast to tree models, fully connected networks (FCNs) select features using a data-driven approach instead of specific functions. Moreover, a feature selection mask can be formed using Entmax sparsity mapping [27] and element-wise multiplication with the features to achieve feature selection. In the proposed work, Entmax is chosen due to its inherent ability to yield sparse outputs. In the proposed methodology, filter modules are designed based on the Entmax function, enabling the precise extraction of relevant and useful features from influencing factors. This choice ensures that only the most impactful features, essential to the task at hand are considered, reducing the risk of redundant or less significant feature interference.

In this study, a filter with combined FCNs and a mask selection approach was used. Given an input vector $X \in R^k$, the filter used fully connected (FC) layers and a softmax activation function to learn a mask $M \in R^k$, which was then multiplied with X to obtain the filtered feature $X' \in R^k$. The selector details are given in Eqs. (1, 2).

$$M = \text{Softmax}(FC_{f2}(\text{LeakyReLU}(FC_{f1}(X)))) \tag{1}$$

$$X' = X \odot M. \tag{2}$$

Here, FC and \odot represent the FC layer and element-wise multiplication, respectively. Note that the sparse auto-encoder architecture is applied to the filter, meaning FC_{f1} maps X into a small dimension (for example, $k/2$) and FC_{f2} recovers it to dimension k . Thus, the filter can mine the importance of each field from the low-dimensional space

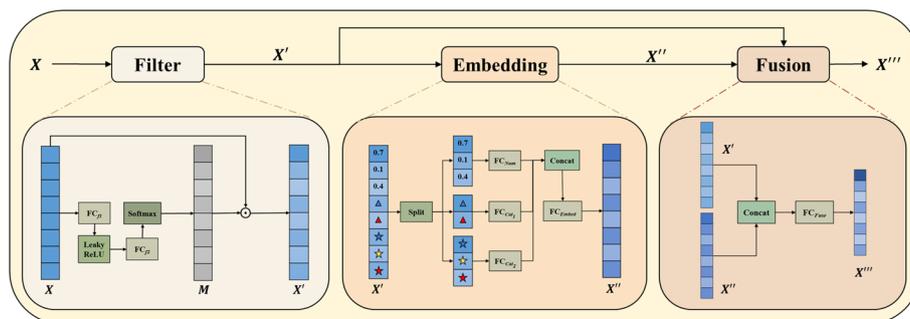


Fig. 2 Overall structure of the Selective Embedding Layer

by FC layers, and use the softmax function for obtaining the feature selection mask M , which weights each field of X .

Embedding

Each field of filtered feature X' is embedded by feature type to exploit intra information. Specifically, the FC layers implement the embedding operation, which are used to store intra-information and integrate the information in the global field. According to the given information of the dataset, the filtered features X' can be split into numerical and categorical features, which means $X' = \{X_{Num}, X_{Cat_1}, X_{Cat_2}, \dots, X_{Cat_k}\}$. For all numerical features, X_{Num} and FC_{Num} were used to build the embedding E_{Num} . For each type of categorical feature X_{Cat_i} , FC_{Cat_i} was used to build the embedding E_{Cat_i} . In this case, FC_{Cat_i} was dedicated for embedding the i th category of the features. Finally, all embeddings were concatenated as the input of the FC_{Fuse} , and important global information was obtained in the embedding features X'' . We formalize this process using Equations (3–6).

$$\{X_{Num}, X_{Cat_1}, X_{Cat_2}, \dots, X_{Cat_k}\} = \text{split}(X), \quad (3)$$

$$E_{Num} = FC_{Num}(X_{Num}), \quad (4)$$

$$E_{Cat_i} = FC_{Cat_i}(X_{Cat_i}), \quad (5)$$

$$X'' = FC_{Embed}(\text{concat}(\{X_{Num}, X_{Cat_1}, X_{Cat_2}, \dots, X_{Cat_k}\})) \in R^k. \quad (6)$$

Fusion

The filtered features X' and embedding features X'' were fused to obtain abstract features X''' . We consider that the filter selects important factors and retains the original information, while embedding realizes the abstraction of the internal information and obtains the information from the new embedding space. To exploit the information of the filtered features and embedding features, X' and X'' were concatenated as the input of an FC layer FC_{Fuse} , then the output was

$$X''' = FC_{Fuse}(\text{Concat}(X', X'')) \in R^k. \quad (7)$$

Multi-dropout attention layer (MDAL)

DL models with attention mechanisms achieve satisfactory results. For example, transformer-based models utilize multihead self-attention to achieve optimal performance in various tasks. However, this is guaranteed by large model parameters and long-term training. To reduce the parameters, speed up training, and introduce attention mechanisms, a multidropout attention layer (MDAL) was proposed, as shown in Fig. 3a.

In the MDAL, X_{i_0} was passed through the dropout layers with different marks for obtaining different field combinations D_i , which were then transformed into attention weights

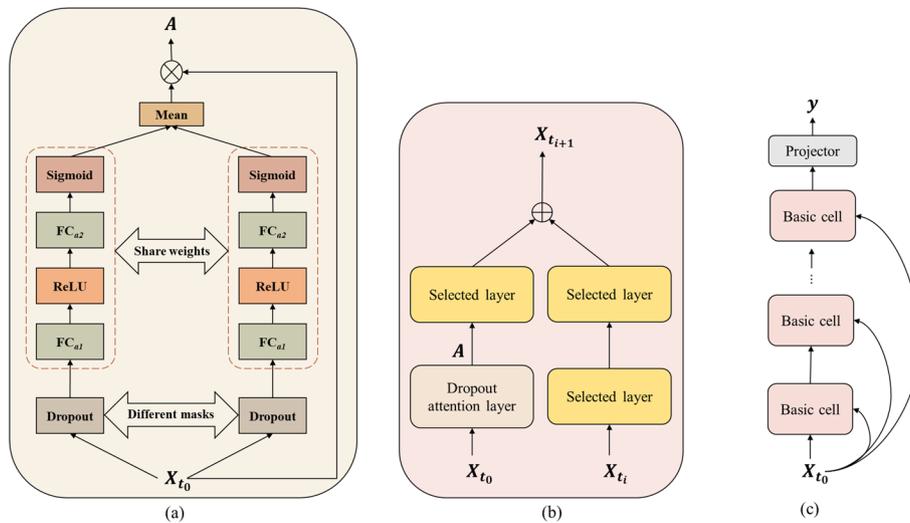


Fig. 3 **a** Multi-dropout attention layer (MDAL); **b** Details of the basic cell; **c** The architecture of our ASEN

A_i through the proposed attention module. Output A was the product of X_{t_0} and the average of all the weights A_i . Formally, we specify this layer as follows:

$$D_i = \text{Dropout}_i(X_{t_0}), \quad (8)$$

$$\text{Attention}(X) = \text{Sigmoid}(FC_{a2}(\text{ReLU}(FC_{a1}(X)))), \quad (9)$$

$$A_i = \text{Attention}(D_i), \quad (10)$$

$$A = X_{t_0} \odot \text{Mean}(A_1, A_2, \dots, A_i). \quad (11)$$

Here, the dropout layers were used to discard some fields of the X_{t_0} and realize a combination of the remaining fields. Because multi-dropout sampling can speed up the convergence and achieve lower error rates than methods without it, different masks and dropout rates were applied to the dropout layers, while the parameters of the attention module were shared.

Basic cell and attention-base selective embedding neural network (ASENN)

The SEL and MDAL were combined to form the basic cells, which were then stacked to build ASENNs for pavement deterioration prediction, as depicted in Fig. 3b and c. The basic cell combines abstract and attention-weighted features for obtaining meaningful representations for predictions and is composed of two main paths. One is to stack two SELs to improve the abstraction of the features of, and the other is to use an MDAL and SEL to obtain attention-weighted field combination information. The outputs of both paths were added and passed to the next cell. Here, we define the output of the i th basic cell as:

$$X_{t_{i+1}} = SL(SL(X_{t_i})) \oplus SL(\text{MDAL}(X_{t_0})), \quad (12)$$

where \oplus denotes element-wise addition.

To find meaningful and abstractive features, multiple basic cells were stacked to build the ASENNs. In this study, the projector was used to map $X_{t+1} \in R^k$ to a scalar $y \in R$, where y is the pavement deterioration data at t .

Development of neural network-based pavement deterioration prediction model

In this study, ASENNs used the same architecture for all four datasets. The neural numbers of the FC layers in the ASENNs were set according to the input variables listed in Table 3. In the MDAL, dropout rates of 0.1 and 0.2 were used. The loss function was the mean squared error. The model optimizer was Adam, with 1×10^3 learning rates. To store the best model, we trained our models for 1×10^4 epochs and saved the model with minimum test loss for prediction. The APIs provided by Pytorch were used to build the ASENNs.

Each dataset contained 17 input variables, meaning $X \in R^k, k = 17$. Because the input variables Lane, Road type, and Maintenance type were categorical variables, one-hot encoding was applied to transform them into numerical variables that are suitable for embedding. In addition, a logarithmic operation was applied to the pavement deterioration data at t and $t-1$. The distribution of pavement deterioration data at t shows inconsistent orders of magnitude for the label types, as shown in Fig. 4. Applying logarithm helps alleviate this phenomenon, which can simply model the nonlinear relationship. In addition, the absolute value of the data was reduced, which accelerated network convergence.

Results and discussion

Experiments were conducted on four datasets to demonstrate the efficacy of the selected nets. The root mean square error (RMSE), Coefficient of Determination (R2), mean absolute error (MAE), and mean absolute percentage error (MAPE) comparisons of all the models are presented in Tables 4, 5, 6, 7. The execution time and parameters of the compared models are shown in Figure 5 to demonstrate the benefit of being lightweight. Furthermore, two ablation studies were conducted to determine the optimal number of selected blocks and validate the efficiency of the MDAL. All results revealed that the selected nets exhibited promising performance.

Comparison with existing models

The four datasets were compared with the existing models listed below.

Table 3 The neural number in each component

Component	Filter		Fusion	MDAL	
Fc layer	FC_{f1}	FC_{f2}	FC_{Fuse}	FC_{a1}	FC_{a2}
Neural numbers	9	17	17	9	17
Component	Embedding				
Fc layer	FC_{Num}	FC_{Cat_1}	FC_{Cat_2}	FC_{Cat_3}	FC_{Embed}
Neural numbers	8	3	2	4	17

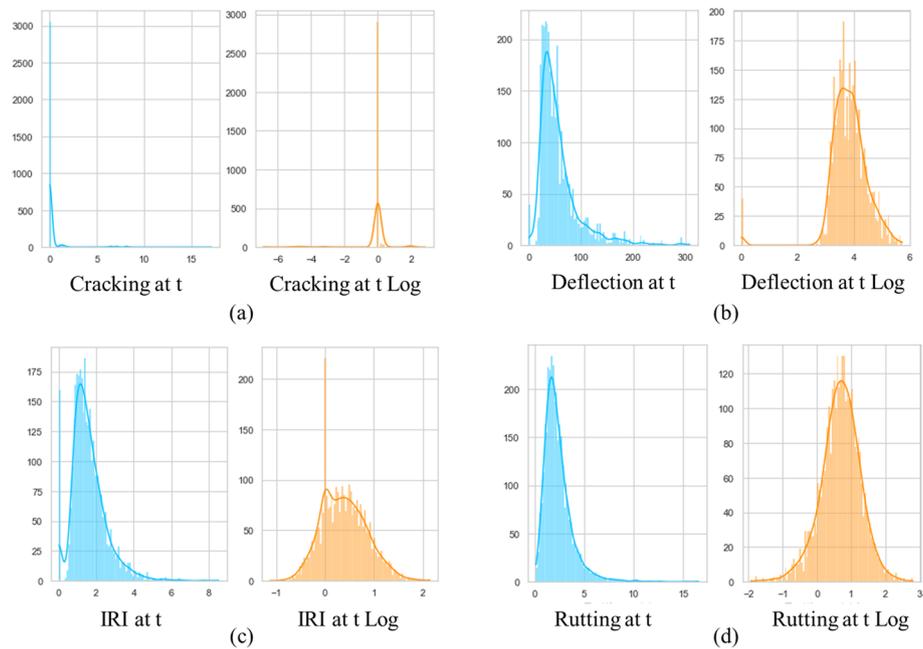


Fig. 4 The distributions of pavement deterioration data at t before and after applying logarithm: **a** Cracking, **b** Deflection, **c** IRI and **d** Rutting

Table 4 The compared results on the Cracking Dataset

Method	RMSE	R^2	MAE	MAPE
Random forest [5]	0.782	0.401	0.236	0.694
XGBoost [6]	0.771	0.417	0.233	0.687
LightGBM [7]	0.763	0.430	0.247	0.726
CatBoost [8]	0.760	0.433	0.228	0.627
MLP [3]	0.758	0.437	0.233	0.681
ResNet [14]	0.762	0.431	0.243	0.757
FT-transformer [10]	0.759	0.435	0.227	0.626
TabNet [9]	0.766	0.425	0.250	0.650
Ours	0.728	0.481	0.218	0.573

Table 5 The compared results on the deflection dataset

Method	RMSE	R^2	MAE	MAPE
Random Forest [5]	0.017	0.999	0.005	0.012
XGBoost [6]	0.010	0.999	0.003	0.010
LightGBM [7]	0.021	0.999	0.007	0.017
CatBoost [8]	0.024	0.999	0.007	0.029
MLP [3]	0.110	0.986	0.027	0.147
ResNet [14]	0.038	0.998	0.025	0.213
FT-transformer [10]	0.009	0.999	0.006	0.022
TabNet [9]	0.060	0.995	0.044	0.182
Ours	0.002	0.999	0.001	0.006

Table 6 The compared results on the IRI dataset

Method	RMSE	R^2	MAE	MAPE
Random forest [5]	0.031	0.998	0.012	0.070
XGBoost [6]	0.025	0.999	0.008	0.038
LightGBM [7]	0.032	0.998	0.014	0.041
CatBoost [8]	0.041	0.998	0.014	0.036
MLP [3]	0.125	0.983	0.048	0.177
ResNet [14]	0.089	0.991	0.049	0.328
FT-transformer [10]	0.025	0.999	0.013	0.048
TabNet [9]	0.123	0.984	0.083	0.464
Ours	0.024	0.999	0.005	0.022

Table 7 The compared results on the rutting dataset

Method	RMSE	R^2	MAE	MAPE
Random forest [5]	0.021	0.999	0.008	0.020
XGBoost [6]	0.012	0.999	0.005	0.016
LightGBM [7]	0.054	0.996	0.014	0.052
CatBoost [8]	0.053	0.996	0.013	0.035
MLP [3]	0.106	0.987	0.026	0.141
ResNet [14]	0.040	0.998	0.022	0.114
FT-transformer [10]	0.008	0.999	0.006	0.016
TabNet [9]	0.048	0.997	0.035	0.145
Ours	0.002	0.999	0.001	0.006

1) Tree-based models

- Random forest [5]: An ensemble decision tree model that can rank the importance of variables in a regression task. The Random Forest model is configured with default parameter settings. It utilizes a forest consisting of 100 trees, and the quality of a split is evaluated using the squared error function.
- XGBoost [6]: The most popular GBDT implementation of GBDT used in regression and classification applications. The XGBoost is configured with default parameter settings in [6].
- LightGBM [7]: The GBDT implementation uses histogram-based algorithms, grows trees leaf-wise, and outperforms XGBoost regarding training speed. The LightGBM is configured with default parameter settings. The maximum number of tree leaves are 31, the learning rate is set to 0.1, and the model is trained with a default of 100 boosted trees.
- CatBoost [8]: GBDT implementation uses an innovative method to deal with categorical features. The parameter configuration includes 1000 iterations, a tree depth of 10, a learning rate of 0.1, and the root mean squared error (RMSE) as the chosen loss function.

2) Deep learning models

- MLP [3]: The MLP that uses four FC layers with LeakyReLU activation function. The neural numbers of the FC layers are 128, 128, 128, and 1, respectively.
- ResNet [14]: A ResNet variation that uses the simplified architecture and applies to tabular learning. The parameter configuration for the model is as follows: The model consists of a total of 3 blocks, each having an input and output size of 128. The first linear layer in each block has an output size of 256. The dropout rates for the first and second dropout layers in each block are given by 0.25 and 0 respectively. The final output size of the model is 1.
- FT-transformer [10]: The transformer-based model that adds a Feature Tokenizer to transform all features into embeddings. The FT-transformer is configured with default parameter settings in [10].
- TabNet [9]: The transformer-based model uses sequential attention to reason the decision steps. In our study, we adopted the default settings suggested by the authors, including a decision prediction layer width of $n_d=8$, attention embedding width of $n_a=8$, and a total of $n_{steps}=3$ architecture steps.

The RMSE, R^2 , MAE, and MAPE results are presented in Tables 4, 5, 6, 7. Overall, ASENNs achieved the best results on the four datasets, followed by the FT-transformer. Specifically, ASENNs exhibit the lowest RMSE and MAE values, signifying its predictive accuracy. Furthermore, it outperforms other models with the highest R^2 value and the lowest MAPE, highlighting its optimal data fitting and minimal relative error. Besides, it can be seen that ASENNs outperform MLP with a much lower RMSE, MAE, MAPE and higher R^2 value, which show the superior performance of our model architecture in deterioration prediction.

For ASENNs, the training and test losses are shown in Fig. 5, and scatter figures of the predicted and actual values are presented in Fig. 6. The training and testing loss curves show less divergence; therefore, the model was not subjected to overfitting. The curve show that the best model was obtained before the 200th epoch, as there was no decrease in the loss after the 200th epoch. In addition, the predicted and actual points were distributed around the $y = x$, meaning that the predictions of our models can match the ground truth.

Model size and execution time

The model size and execution time are shown in Fig. 7. Here, execution time refers to the prediction time for a test set. A summary library provided by PyTorch was used to measure the storage capacity of the model. The execution time of the proposed model was recorded as 14.5 ms, while the memory occupied by the selected nets was 0.57 MB. Compared to the FT transformer, ASENn required 6.7 times less storage space. Compared with TabNet, the proposed model execution rapidly. Although MLP and ResNet occupy small memory and have fast execution speeds, the proposed models

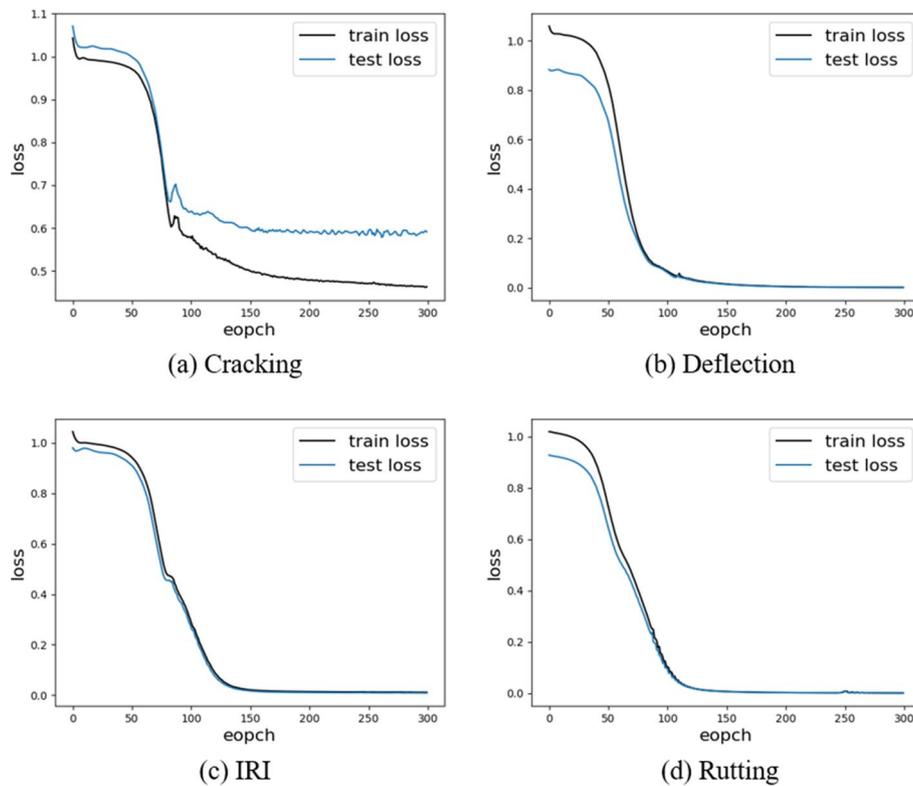


Fig. 5 The training and test loss of Selected Nets on four datasets: **a** cracking, **b** deflection, **c** IRI and **d** rutting

outperformed both in terms of accuracy. Therefore, the proposed selected nets algorithm considers not only accuracy but also storage and computing resources.

Numbers of selected blocks

The number of basic cells determines the abstraction ability of ASENNs. To choose the appropriate number of cells for prediction, cells = {4, 6, 8, 10} were used for the ablation study, as shown in Table 8. With the increase in the number of cells from 4 to 10, the model showed a stronger ability to abstract and achieved better results. Changing the number of cells exhibited no significant effect on the performance of the model when using the deflection dataset, indicating sufficient accuracy of the proposed model prediction results. Overall, a cell value of 8 achieved promising performance among the settings. It can also be seen from the table that the model showed comparable performance while maintaining the value of the cell at 4, which can help reduce the computing and memory resources within the tolerable range. In addition, extreme depths (cells = 10) were found unsuitable for prediction because they may be caused by the limited combination of feature fields.

Effectiveness of MDAL

To verify the effectiveness of MDAL, MDAL was replaced with SEL in the basic cell for ablation. In this case, the number of basic cells was set to 8 and the same training dataset was used to ensure validity. ASENN with MDAL outperformed the one without MDAL

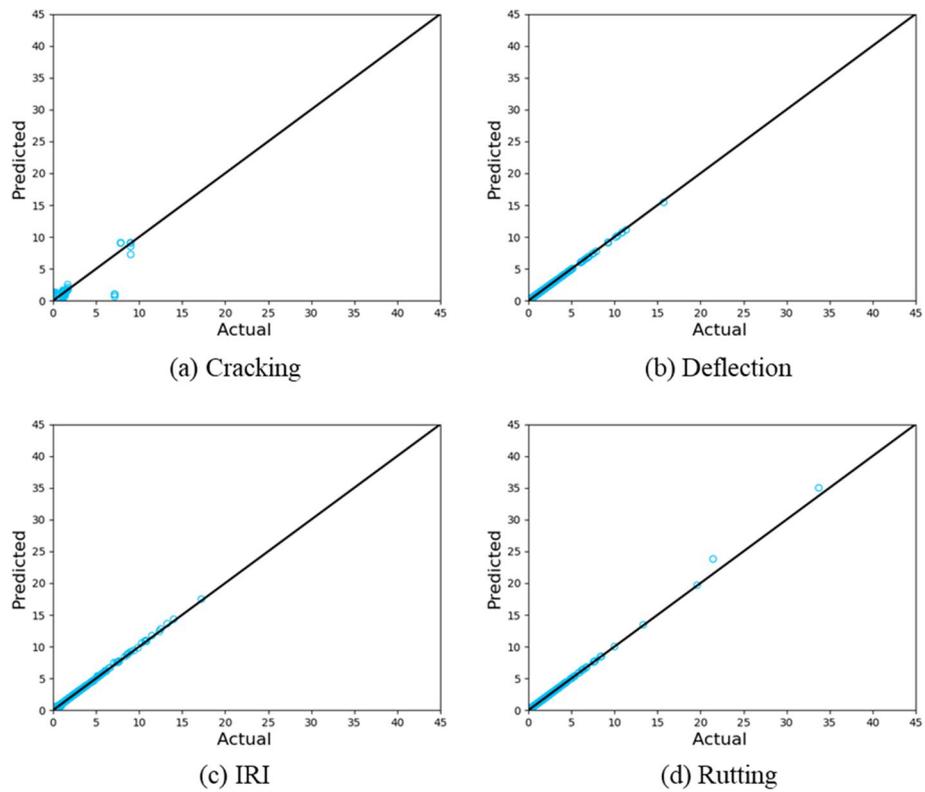


Fig. 6 The scatter figures of predicted and actual values: **a** cracking, **b** deflection, **c** IRI and **d** rutting

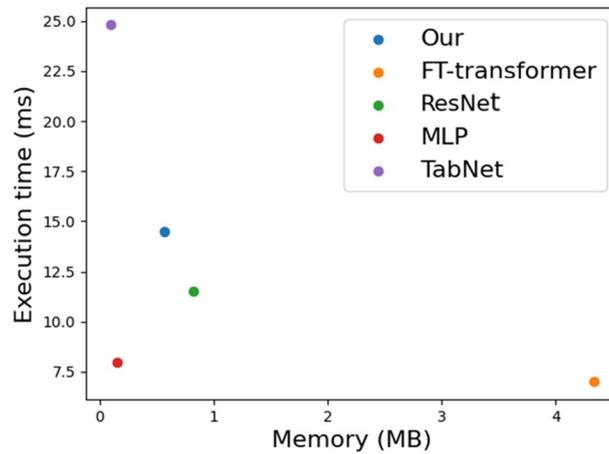


Fig. 7 Model size and execution time of different models

Table 8 Effect of neuron numbers ablation on performance

cells	Cracking	Deflection	IRI	Rutting
4	0.757	0.0028	0.0259	0.0048
6	0.725	0.0036	0.0116	0.0036
8	0.725	0.0027	0.0109	0.0014
10	0.733	0.0026	0.0068	0.0042

Table 9 Effect of MDAL ablation on performance

	Cracking	Deflection	IRI	Rutting
Without MDAL	0.748	0.0048	0.021	0.004
With MDAL	0.727	0.0027	0.010	0.002

on the four datasets. Without the MDAL, the abstraction performance decreased, as shown in Table 9.

Conclusion

Predicting the future pavement performance is an essential component of a pavement management system. However, the complexity associated with pavement deterioration process arising from the contributing heterogeneous factors poses a severe challenge in performance prediction. The ASENN model proposed in this study addresses the limitations of the existing models in terms of requiring less memory and faster execution time without comprising the accuracy of the model. In this study, an ASENN was proposed for road distress prediction, considering the impact of a wide variety of factors causing pavement deterioration. The ASENN model combines the SEL and MDAL components, stacked one over the other to form the basic cell based on the real big data collected from the major road networks in the UAE from the year 2013 to 2019.

The SEL framework was designed to identify abstract features from tabular data using filtering, embedding, and fusion approaches and the MDAL framework was proposed to achieve multiple attention-weight combinations from the raw data. Experiments on cracking, deflection, IRI, and rutting datasets were performed, which showed the effectiveness and efficiency of ASENNs in pavement deterioration regression tasks. In addition, the number of basic cells and effectiveness of MDAL were explored by ablation studies, indicating that good results can be achieved with a relatively less number of basic cells. As the number of blocks increases, the ability of the network to extract abstract features also increases. The proposed ASENN model showed high accuracy with an average value of RMSE of 0.1894, an R^2 of 0.8701, and a MAPE of 0.1518 for predicting cracking, deflection, IRI, and rutting models. Compared to SOTA models like Tabnet and ResNet, the proposed work offers enhanced precision and stability across all tasks.

However, extreme depths were not preferred due to the limited combination space of feature fields. For MDAL, attention-weighted combinations of raw data increase the diversity of features and enhance the performance. The ASENN model could be extended further to obtain more ML techniques in the domains of pavement prediction, management, and maintenance. In the future, we aim to enhance our data preprocessing methodologies, aiming for improved dataset refinement. We will also investigate advanced deep-learning prediction strategies.

Acknowledgements

The authors acknowledge the Ministry of Energy and Infrastructure, UAE for providing access to road data.

Author contributions

The authors HA-J, BP, ZX, and QZ, collectively contributed to the conceptualization, methodology, software development, validation, and data curation of this study. BP and ZX performed the formal analysis. HA-J provided the necessary resources for the research. The original draft of the manuscript was prepared by BP, ZX, and LA, and later reviewed and

edited by all authors. BP and ZX were responsible for visualization. The project was supervised by HA-J and QZ with project administration handled by HA-J. All authors read and approved the final manuscript.

Funding

This research was funded by the UAE University, Grant No. 31R291.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 26 June 2023 Accepted: 17 October 2023

Published online: 28 October 2023

References

- Rogers M, Enright B. Highway engineering. 3rd ed. Hoboken: John Wiley & Sons, Incorporated; 2016.
- American Society of Civil Engineers, 2021 Report Card for America's Infrastructure: Solid Waste, 2022.
- Sholevar N, Golroo A, Esfahani SR. Machine learning techniques for pavement condition evaluation. *Autom Constr*. 2022;136(Feb): 104190. <https://doi.org/10.1016/j.autcon.2022.104190>.
- Yang X, Guan J, Ding L, You Z, Lee VCS, Mohd Hasan MR, Cheng X. Research and applications of artificial neural network in pavement engineering: a state-of-the-art review. *J Traffic Transp Eng*. 2021. <https://doi.org/10.1016/j.jtte.2021.03.005>.
- Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32. <https://doi.org/10.1023/A:1010933404324>.
- Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. p. 785–94. <https://doi.org/10.1145/2939672.2939785>.
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Liu TY. LightGBM: a highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst*. 2017;2017(Dec(Nips)):3147–55.
- Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: unbiased boosting with categorical features. *Adv Neural Inf Process Syst*. 2018;2018(Dec(Section 4)):6638–48.
- Arik S, Pfister T. TabNet: attentive interpretable tabular learning. In: 35th AAAI Conference on Artificial Intelligence, AAAI 2021. 2021;8A:6679–87.
- Gorishniy Y, Rubachev I, Khurlov V, Babenko A. Revisiting deep learning models for tabular data. In: 35th conference on neural information processing systems (NeurIPS 2021). 2021, 23(NeurIPS), pp. 18932–18943. <https://doi.org/10.48550/ARXIV.2106.11959>
- Yao L, Dong Q, Jiang J, Ni F. Establishment of prediction models of asphalt pavement performance based on a novel data calibration method and neural network. *Transp Res Rec*. 2019;2673(1):66–82.
- Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-excitation networks. *IEEE Trans Pattern Anal Mach Intell*. 2020;42(8):2011–23.
- Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat*. 2001;29(5):1189–232. <https://doi.org/10.1214/aos/1013203451>.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE computer society conference on computer vision and pattern recognition, 2016–Decem. 2016. p. 770–8. <https://doi.org/10.1109/CVPR.2016.90>.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: 31st conference on neural information processing systems (NIPS 2017). Long Beach, CA, USA, 2017.
- Ma D, Fang H, Wang N, Zhang C, Dong J, Hu H. Automatic detection and counting system for pavement cracks based on PCGAN and YOLO-MF. *IEEE Trans Intell Transp Syst*. 2022;23(11):22166–78.
- Jiang J, Wang F, Wang Y, Jiang W, Qiao Y, Bai W, Zheng X. An urban road risk assessment framework based on convolutional neural networks. *Int J Dis Risk Sci*. 2023. <https://doi.org/10.1007/s13753-023-00498-7>.
- Haddad AJ, Chehab GR, Saad GA. The use of deep neural networks for developing generic pavement rutting predictive models. *Int J Pavement Eng*. 2021. <https://doi.org/10.1080/10298436.2021.1942466>.
- Gao L, Yu Y, Hao Ren Y, Lu P. Detection of pavement maintenance treatments using deep-learning network. *Trans Res Rec*. 2021;2675(9):1434–43. <https://doi.org/10.1177/03611981211007846>.
- Abohamer H, Elseifi M, Dhakal N, Zhang Z, Fillastre CN. Development of a deep convolutional neural network for the prediction of pavement roughness from 3D images. *J Transp Eng Part B Pavements*. 2021;147(4):4021048.
- de Venancio PVAB, Lisboa AC, Barbosa AV. An automatic fire detection system based on deep convolutional neural networks for low-power, resource-constrained devices. *Neural Comput Appl*. 2022;34(18):15349–68.
- Zhou Q, Okte E, Al-Qadi IL. Predicting pavement roughness using deep learning algorithms. *Transp Res Rec*. 2021;2675(11):1062–72. <https://doi.org/10.1177/03611981211023765>.

23. Oruh J, Viriri S, Adegun A. Long short-term memory recurrent neural network for automatic speech recognition. *IEEE Access*. 2022;10:30069–79.
24. Vidal C, Malysz P, Naguib M, Emadi A, Kollmeyer PJ. Estimating battery state of charge using recurrent and non-recurrent neural networks. *J Energy Storage*. 2022;47: 103660.
25. Bayat R, Talatahari S, Gandomi AH, Habibi M, Aminnejad B. Artificial neural networks for flexible pavement. *Information*. 2023;14(2):62.
26. Philip B, Jassmi HA. A Bayesian approach towards modelling the interrelationships of pavement deterioration factors. *Buildings*. 2022. <https://doi.org/10.3390/buildings12071039>.
27. Peters B, Niculae V, Martins AFT. Sparse sequence-to-sequence models. In: *ACL 2019—57th annual meeting of the association for computational linguistics, proceedings of the conference*. 2020. p. 1504–19. <https://doi.org/10.18653/v1/p19-1146>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
