LSTM+MA: A Time Series Model for Predicting the Pavement IRI

Tianjie Zhang1, Alex Smith2, Huachun Zhai3, and Yang Lu4\*

1. **Computing PhD Program, Boise State University, Boise, ID, 83725, USA Email:** [**tjzhang@u.boisestate.edu**](mailto:tjzhang@u.boisestate.edu)
2. **Department of Computer Science, Boise State University, Boise, ID, 83725, USA. Email:** [**alexsmith194@u.boisestate.edu**](mailto:alexsmith194@u.boisestate.edu)
3. **Idaho Asphalt Supply, Nampa, ID, 83687, USA, Email:** [**hzhai@idahoasphalt.com**](mailto:hzhai@idahoasphalt.com)
4. **Department of Civil Engineering, Boise State University, Boise, ID, 83725, USA, Email:** [**yangfranklu@boisestate.edu**](mailto:yangfranklu@boisestate.edu)

**\*Corresponding Author**

***Abstract*—** **Accurately prediction of pavement performance is essential for transportation agencies to appropriately allocate resources road maintenance and rehabilitation. The International Roughness Index (IRI) is one of the most commonly used pavement performance indicators to reflect the surface roughness. However, existing research on IRI prediction mainly focuses on using linear regression or traditional machine learning, which cannot take into account the history effects on IRI caused by climate, traffic, pavement construction and intermittent maintenance. In this work, a Long Short-Term Memory (LSTM)-based model, LSTM+MA, is proposed to predict the IRI of pavements using the time-series data extracted from the Long-Term Pavement Performance (LTPP) dataset. Effective preprocessing methods and hyperparameter fine-tuning are selected to improve the accuracy of the model. The performance of the LSTM+MA is compared with other state-of-the-art models, including Logistic Regressor (LR), Support Vector Regressor (SVR), Random Forest (RF), K-nearest-neighbor Regressor (KNR), Fully connected Neural Network (FNN), XGBoost (XGB), Recurrent Neural Network (RNN) and LSTM. The results show that selected preprocessing methods can help the model learn quickly from the data and reach high accuracy with small epochs. Also, it shows that the proposed LSTM+MA model significantly outperforms other models, with coefficients of determination () of 0.965 and Mean Square Error (MSE) of 0.030 in the test datasets. Moreover, an overfitting score is proposed in this work to represent the severity degree of the overfitting problem and it shows that our proposed model does not suffer severely from overfitting.**

***Index Terms*—** **Deep Learning, International Roughness Index, LSTM, LTPP**

# I. INTRODUCTION

Pavement roughness is defined as the deviation of a pavement surface from an actual surface with characteristic dimensions that affect vehicle dynamics, ride quality, dynamic loads, and drainage [1]. It is considered one of the most significant indicators of pavement condition as it dramatically affects driving comfort and safety, especially for high-speed-limit roads [2]. Various factors, including construction quality, maintenance, distress, climate, and traffic, all affect pavement roughness [3]. International Roughness Index (IRI) is widely recognized as a key indicator of pavement functional condition and ride quality, which is extensively utilized for evaluating road conditions [4]. It is costly and time-consuming for transportation agencies to manually survey and monitor pavement roughness [5]. Therefore, predicting the IRI during pavement life becomes essential for decision-making and rehabilitation planning in transportation infrastructure management.

A lot of research and attempts have been made to estimate the roughness based on pavement and traffic conditions. Traditionally, many researchers have tried to evaluate IRI using a linear regression method. However, the relationship between IRI and other parameters is highly nonlinear [6], which makes linear regression a rather rough measure of IRI. Machine learning-based approaches are more advanced and become more popular to solve nonlinear problems. For example, Deng et al. [7] modified a feedforward network with particle swarm optimization to predict the rutting performance of asphalt pavement using material properties, structure parameters, and traffic conditions from collecting data from Idaho State. Wang et al.[8] developed a hybrid grey relation analysis and Support Vector Regressor (SVR) to predict pavement performance, obtaining a Root Mean Square Error (RMSE) of 0.298. However, overfitting becomes a major limitation of machine learning-based approaches. The model can easily learn from the training data too well, leading to poor performance in unseen data. Ensemble models, which combine predictions from multiple individual models to make a final prediction, become popular in IRI evaluation because it takes various decisions from several approaches, which can efficiently ease the overfitting issue [9]. The idea behind ensemble learning is to leverage the diversity and collective intelligence of each individual model to improve overall predictive accuracy and robustness [10]. In recent years, there have been numerous research works using ensemble learning in predicting pavement performance. For example, Wang et al. [11] developed an Adaboost regression model to improve its performance in predicting IRI by using the inputs including pavement thickness, service age, average annual daily truck traffic (AADTT), and crack length from Long-Term Pavement Performance (LTPP) dataset. The proposed model reached an of 0.95 and Mean Square Error (MSE) of 0.0088. Gong et al. [2] used a Random Forest (RF) Regression model to estimate the IRI value of flexible pavement by considering the distress, traffic, maintenance and structure data from LTPP. It outperformed the Linear Regression model ( is 0.62) and obtained an more than 0.95 in both training and testing datasets. Damirchilo et al. [6] used XGboost to avoid overfitting of the training and handle the missing value in LTPP datasets. It obtained a of 0.7 outperforming the RF and SVR. Song et al. [5] proposed a ThunderGBM-based ensemble learning model coupled with the Shapley Additive Explanation method to predict the IRI of asphalt pavements. 2699 observations were extracted from the LTPP database to train the model and the model achieved a satisfactory result with a value of 0.88 and RMSE of 0.08.

However, these machine learning-based methods did not consider the time-dependent characteristics of pavement roughness. The pavement surface roughness evolves over time and is subject to a variety of factors, e.g., climate, traffic flow, material properties, although the initial IRI is based on the quality of construction. After construction, the IRI of the pavement changes with the climate, traffic pavement maintenance and rehabilitation over time. In other words, IRI is a time-accumulated index that the history of all these factors could influence the present service status. Thus, models considering the history effect become a direction to precisely predict the IRI for pavement management decision making. Recurrent Neural Networks (RNNs) are a class of neural networks designed explicitly for sequential data processing, making them well-suited for time-series prediction tasks [12]. Unlike other deep learning algorithms like Convolutional Neural Networks [13, 14] or traditional Feed-forward Neural Networks [15], RNNs can capture dependencies and patterns in the temporal dynamics of the data, allowing them to model the evolving behavior of pavements over time. By utilizing the historical IRI measurements and other relevant variables in the LTPP dataset, we can train an RNN model to learn the underlying patterns and make accurate predictions of future IRI values. Very limited existing research has explored using RNN to predict pavement performance. For example, Zhou et al. [1] proposed an RNN-based model to predict the IRI on the asphalt pavement based on the LTPP datasets. The loads, temperature, precipitation, evaporation, rutting and cracking are considered as input parameters for the analysis. The results showed that the presented RNN model reached a of 0.93. Han et al. [16] proposed a modified RNN model for falling weight deflectometer (FWD) back calculation. The model showed a stronger generalization ability than traditional Artificial Neural Networks.

However, a traditional RNN model may suffer from the vanishing gradient problem. It happens during the backpropagation through time when gradients diminish exponentially [17]. This makes it challenging for traditional RNNs to capture long-term dependencies in the LTPP data, which are critical in time-series prediction tasks. Moreover, a traditional RNN has a simple memory mechanism that cannot selectively remember or forget information over longer sequences. This limitation can hinder its ability to effectively model complex patterns in the data and may not fully leverage historical information when predicting IRI values.

Therefore, in this work, we proposed a Long Short-Term Memory (LSTM) based model, LSTM+MA, which combines two LSTM blocks and a multi-head attention layer to predict the IRI performance of the pavement. In this model, one LSTM network is utilized as an encoder to capture important information from the entire sequence. Then, the cell state and hidden state from this LSTM model are transferred to the other LSTM network which works as a decoder to generate a sequence. The output from the encoder and decoder are both utilized to fully capture the historical information from the data. Then, they are combined together as input to the multi-head attention mechanism. After that, a fully-connected layer is added to predict the current IRI value of a pavement. Data preprocessing and hyperparameter fine-tuning are used in this work to improve the models’ performance. In addition, the presented LSTM+MA model is compared with other state-of-the-art models, including Logistic Regressor (LR), SVR, RF, K-nearest Neighbor Regressor (KNR), Fully-connected Neural Network (FNN), XGBoost (XGB), RNN and LSTM. The results show that our presented model outperforms other models as it gains the highest among all the tested models.

# II. Methodology

## A. Data Engineering

To train the models, 25,167 samples are collected from the LTPP dataset. This dataset is a comprehensive and widely used resource in the field of transportation engineering. It is a collection of data gathered from pavement sections across the United States and Canada. The primary objective of the LTPP program is to monitor and evaluate the long-term performance of various pavement types under different conditions. In this work, predictors in traffic, climate, pavement construction and maintenance are obtained from LTPP and utilized for training as shown in Table 1. Climate-related factors, including precipitation, temperature, and Freeze-Thaw cycles, are considered in the training procedure. Pavement construction and maintenance factors include initial IRI value, age, maintenance type, and transverse crack length. While for the traffic, the Average Annual Daily Traffic (AADT) is considered in the train process.

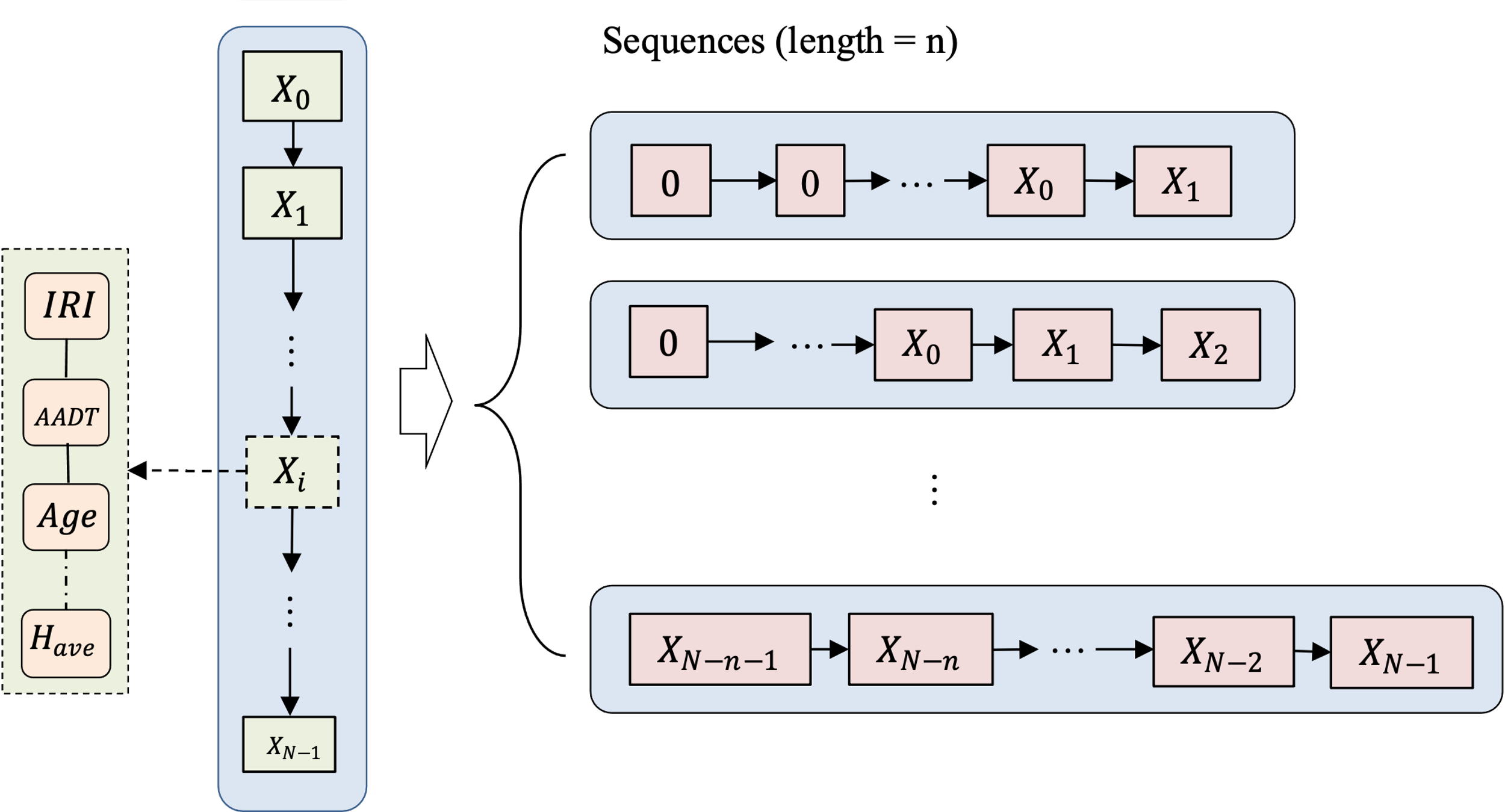
TABLE 1. Predictor utilized in the prediction and its description

|  |  |
| --- | --- |
| **Predictor** | **Description** |
| **Traffic** | |
| AADT | Average Annual Daily Traffic |
| **Pavement construction and maintenance** | |
|  | The initial IRI value after the pavement was built. |
| Age | Years since the road was constructed. |
|  | Different types of maintenance interventions, such as crack sealing or overlay. |
|  | Length of transverse crack. |
|  | The surface type: asphalt or concrete. |
| **Climate** | |
|  | Average annual temperature |
|  | Freeze-thaw cycle: number of days in a year when the maximum air temperature is greater than 0°C (32°F) and minimum air temperature is less than 0°C (32°F) on the same day. |
|  | Average annual precipitation. |
|  | Amount of snow that will be received in an average year. |
|  | Average annual humidity. |

A higher volume of traffic can accelerate pavement deterioration, leading to increased roughness. By incorporating AADT as a predictor, the model can capture the influence of traffic on IRI. The initial IRI () is the most important parameter as it serves as a baseline and can influence the future evolution of the pavement condition [18]. Pavement age since the start of measuring IRI is a critical factor in understanding the long-term performance of the pavement. It helps quantify the effects of ageing and deterioration processes on the roughness of the pavement. Cracking is one of the most common distresses observed in pavements. Including transverse crack length as a predictor helps the model consider the extent of cracking as a factor affecting roughness. Longer cracks may result in more significant roughness development over time. Maintenance activities can greatly influence pavement condition and roughness. Different types of maintenance interventions, such as crack sealing or overlay, can have varying effects on IRI. By incorporating maintenance type as a predictor, the model can learn how different maintenance actions impact the roughness of the pavement. Climate-related factors have a significant impact on the pavement's performance and condition. Changes in temperature, precipitation, humidity, snowfall, and freeze-thaw cycles can lead to cracking and deterioration.

## B. Data Preprocessing

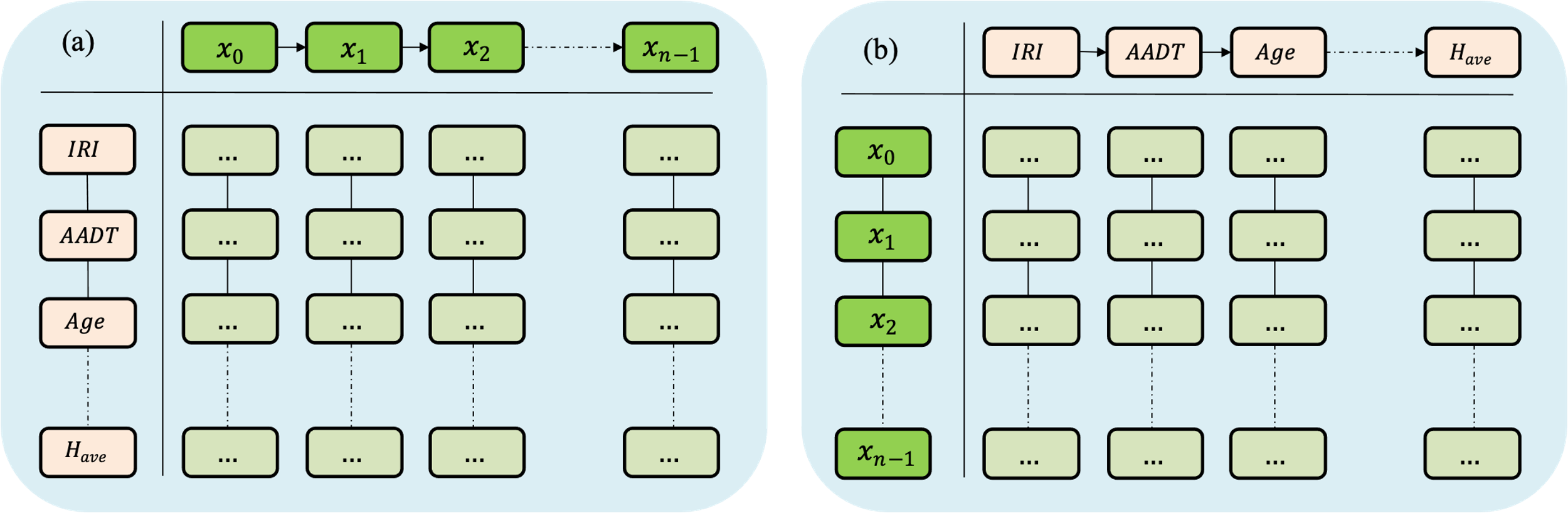
We need to transfer the data into sequences as which is required by LSTM models. For a road with N-years recordings: , it would be divided into several sequences (length is ) rather than put the whole recording into the model. As shown in Figure 1, sequences would be generated from this road. Zero would be padding as default if there are not enough recordings in a single sequence.



**Figure 1.** The procedure of transferring a road information into several sequences. stands for one measurement in a road. All the predictors including IRI, AADT, Age and so on would be measured each time.

For example, if a road had 5 IRI measurements (through 2018 to 2022), four sequences would be generated with 1, 2, 3, and 4 historical elements with the goals of predicting this road’s 2nd, 3rd, 4th, and 5th year’s IRI value, respectively. These sequences contain the initial measurements and are continuous. This method’s ability to process variable-length sequences makes it more flexible and adaptable to real-world pavement data, where the length of historical data may vary for different road sections. However, as many sequences are subsets of other sequences and in order to prevent data leakage, we separated our train and test datasets by which roads were in them, not by randomly splitting the samples. The samples are arranged in order of least recent to most recent. Every sample has a feature representing the number of months since the road was constructed to when the measurement was made. There is also an empty sample appended to the end of every sequence that contains only the number of months since the construction to predict the IRI.

We also transposed the original input matrix to increase the training speed and accuracy of our models as shown in Figure 2.



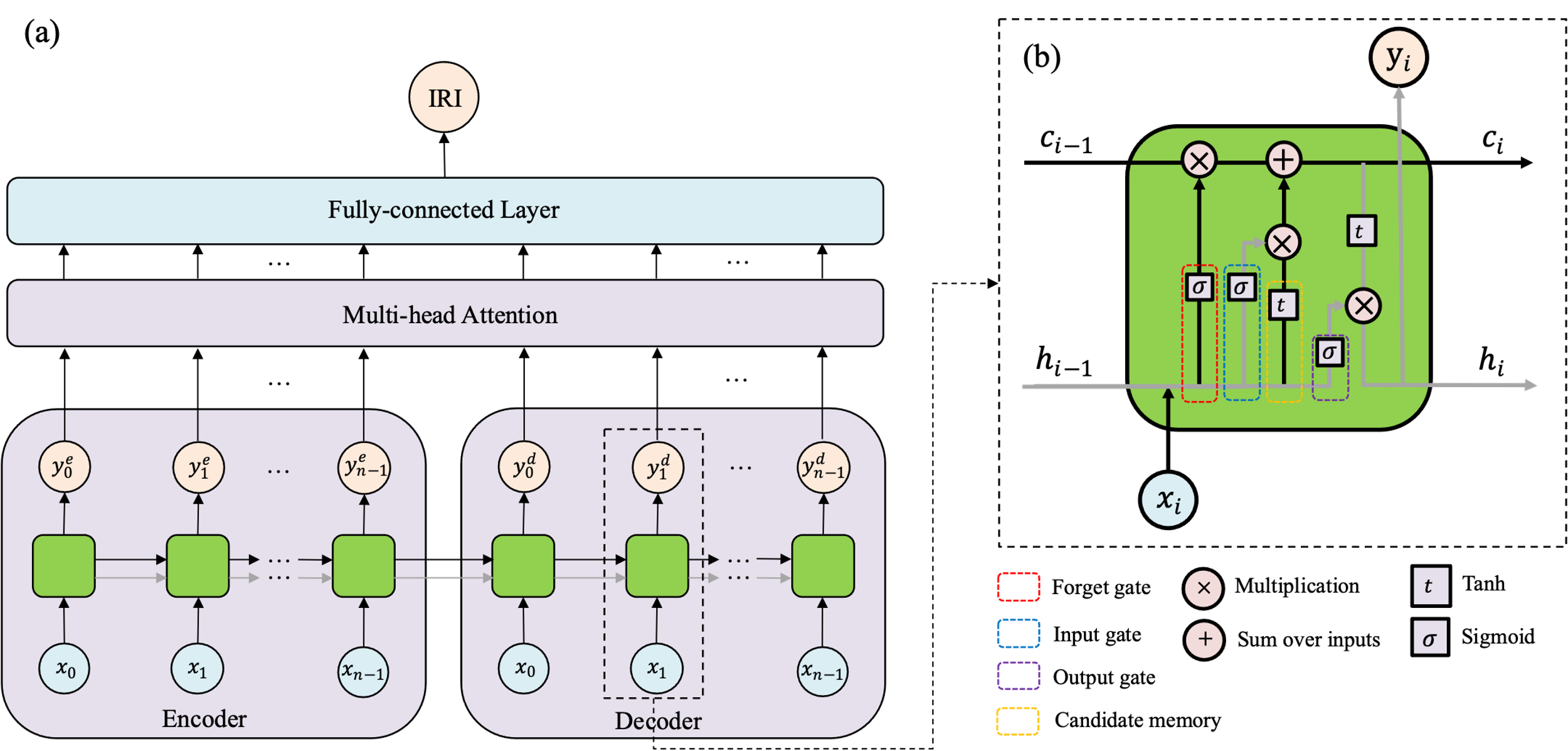
**Figure 2.** Dataset shape visualization: (a) Conventional time-series data; (b) Transposed time-series data.

As one of the main problems with processing and predicting this kind of data is retaining the impact of inputs from earlier in the input sequence, transposing the input matrix allows the model to process the entire series of events at the same time. In the transposed data, the model is not given a sequence of events. It is given a series of properties that have only an implied sequence and an explicit temporal order for all the data points. This greatly reduces the requirements on the model to remember and helps both models like the RNN which have problems with vanishing gradients and models like LSTMs and the LSTM+MA which already have other measures in place to help.

At the same time, some other preprocessing was applied to the data. The mean value of the left and right wheel path IRI was used as the IRI value of a road. Min-max scaling was applied to each parameter, which is particularly useful when the features have different scales or ranges, as it helps to normalize the data and bring all features to the same scale (0 to 1). It is commonly employed in machine learning algorithms that are sensitive to the scale of the input features.

## C. LSTM+MA

The proposed LSTM+MA model is a sequence-to-sequence architecture that combines LSTMs and attention mechanisms as shown in Figure 3. It is designed for predicting the IRI of pavement based on various predictors from traffic, climate, pavement construction and maintenance.



**Figure 3.** The details of the proposed LSTM+MA model. (a) The architecture of the LSTM+MA model. It is an encoder-decoder structure. Both the encoder and decoder are a single LSTM network. The output from the encoder and decoder are combined and input into a multi-head attention mechanism. A fully-connected layer is used to get the information from the attention module and predict the final IRI value. stands for the input vector and is the output vector. (b) The structure of the LSTM layer. is the cell state and is the hidden state.

The presented LSTM+MA follows an encoder-decoder architecture, which is commonly used for sequence-to-sequence tasks (Figure 3 (a)). The encoder processes the input sequence and captures its temporal information, while the decoder generates the output sequence based on the information provided by the cell state () and hidden state () from encoder. LSTM is a type of RNN that addresses the vanishing gradient problem and allow for capturing long-term dependencies in sequential data [19]. The encoder and decoder both utilize LSTM layers, enabling the model to handle variable-length input and output sequences effectively. The LSTM component can be broken down into a few specific components that are responsible for different parts of the model behavior. The input gate, forget gate, output gate and candidate memory in the LSTM component can be explained using Equation (1), (2), (3) and (4).

(1)

(2)

(3)

(4)

Where is the sigmoid function, tanh stands for the tangent hyperbolic function, is the input gate, is the forget gate, is the output gate, is the candidate memory, is the weight matrix and is the bias. The input, forget, and the candidate memory allow the model to learn what data to retain in the cell matrix as it processes the sequence. The output gate updates the hidden state of the model. These components are then combined to generate the hidden and cell states as shown in Equation (5) and (6).

(5)

(6)

Where is the output of the LSTM layer and the hidden layer for the next LSTM layer in the sequence. It can be calculated as the Hadamard product, or element wise product of the output gate and the current cell state. is the current cell state and is the sum of the Hadamard product of the forget gate and the last cell state and the product of the input gate and the candidate gate. This step forgets old information and stores the new information that the model has.

The attention mechanism enhances the model's ability to focus on relevant parts of the input sequence when generating the output. This is especially useful in long sequences, as it helps the model to attend to the most informative time steps and ignore noise or less relevant information.

Multi-head attention [20] is used as the attention mechanism in this work because it allows the model to attend to different positions within the input sequence simultaneously. It computes multiple attention heads, each capturing different patterns and relationships in the data. Moreover, it helps the model capture complex dependencies and relationships between predictors, which can be valuable in pavement IRI prediction, where the performance is influenced by multiple factors. The attention mechanism embedded improves the model's interpretability by highlighting the importance of different parts of the input sequence during the prediction process. After applying the attention mechanism, the model uses a fully-connected layer to transform the output to the desired prediction format (IRI values). The output is reshaped to match the appropriate dimensions for the final prediction.

There are some advantages of our proposed LSTM+MA model: (1) The LSTM layer is a specialized variant of RNNs, designed to address the vanishing gradient problem and facilitate the modeling of long-term dependencies in sequential data. By incorporating memory cells and gating mechanisms, LSTMs can effectively retain and update information over extended time steps, making them particularly well-suited for time series prediction tasks. The cell state () is controlled by the forget gate, input gate, and candidate memory for each time step. By applying element-wise multiplication to the forget gate and the previous cell state, the LSTM forgets irrelevant information. The input gate and the candidate memory determine which new information is incorporated into the updated cell state. The resulting cell state retains valuable historical context while selectively integrating new information, facilitating the modeling of long-term dependencies in the data. The hidden state () is the LSTM layer's output for the current time step, representing the processed and summarized information that is passed to the next time step. It is calculated by applying element-wise multiplication to the cell state and the output of the output gate. The hidden state captures relevant information from the current input and retains valuable historical context from previous time steps, enabling the LSTM to carry forward essential features as it processes the sequence. (2) The attention mechanism in the model allows it to focus on the entire context of the encoded and generated sequence to predict the future state more accurately as well as make the model be able to consider a much larger context. The multiheaded attention layer further enhances the model's capacity to capture diverse patterns and relationships between predictors, which can lead to better predictive performance.

D. Evaluation Metrics

The performance of models is evaluated by MSE and . MSE measures the average of the squared differences between the predicted results and actual values. It can provide a measure of how well the model's predictions align with the true values, with higher values indicating greater errors. It can be calculated using Equation (7).

(7)

Where *n* is the number of total evaluated data, is the *i*-th data point, is the prediction of the *i*-th data.

stands for the coefficient of determination. It measures the proportion of the total variation in the prediction that can be explained by the model. It is used to measure the accuracy of a model’s fit and can be calculated using Equation (8).

(8)

Where SSR is the sum of squares regression, SST stands for the sum of squares total, is the average value of the data.

To evaluate and compare the performance between different models statistically, each model is trained and tested three times on randomly split train and test datasets. The average value and standard deviation would be calculated based on these tests. By doing this, the average value can show the general accuracy of the model while the standard deviation from different tests can represent the robustness of the model.

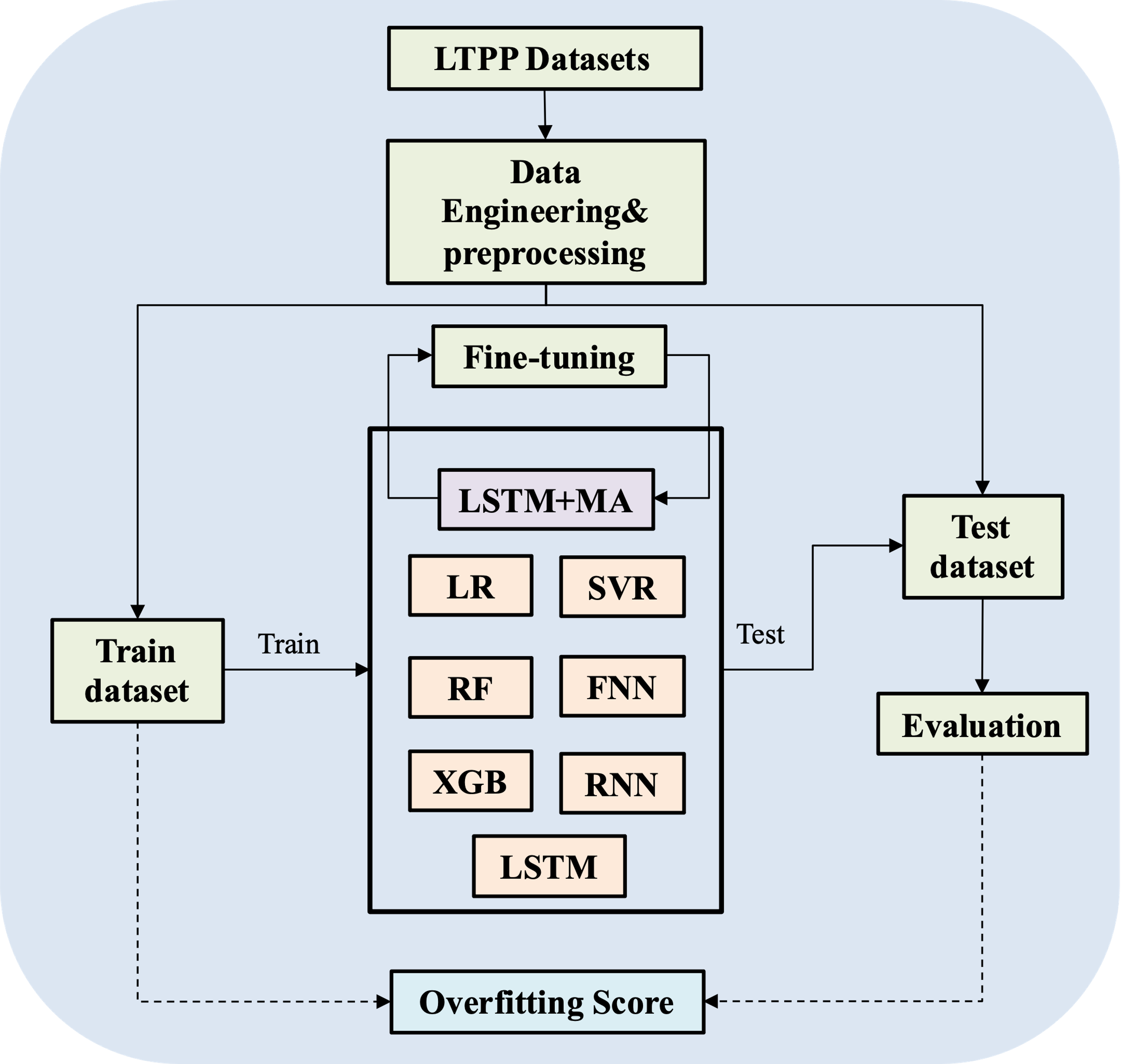
Overfitting problems are also considered in this work. Overfitting occurs when the regression models learn the training data too well, resulting in poor generalization to unseen data. The difference between the training and testing performance is considered as an overfitting score in this work as shown in Equation (9). It is utilized as a metric to quantify the overfitting issue in the selected approach.

(9)

Where stands for the overfitting score, is the value when model applying in the train set, is the value when model applying in the test set. It can provide a measure of the overfitting degree in the model, with higher values indicating greater overfitting problem.

E. Overall Procedure

The aim of the work is to propose a model that can predict the IRI value with higher accuracy based on the predictors extracted from the LTPP dataset. First, data engineering is applied on the LTPP data to seek some useful and related predictors for the IRI prediction. Then, data preprocessing is applied to process the data. The proposed model LSTM+MA would be compared with other state-of-the-art models to show its high performance as shown in **Figure** 4.



**Figure 4.** The overall procedure of the work.

Seven models, which are popular used in other research, are trained and tested on this dataset and compared with our proposed method. Models include LR, SVR, RF, KNR, FNN, XGB, RNN and LSTM. The details of these models are shown in TABLE 2. Each model would be run three times based on different train-test split. By doing this, the average and standard deviation of these three running could be calculated which can make the experiments statistically meaningful. A hyperparameter fine-tuning process would be conducted in the LSTM+MA model to find the best performance it can reach. In this approach the possible parameters, including the input sequence length, the learning rates, the leaning-rate decay rate, the LSTM layer dropout, the number of LSTM layers, the batch size, and the number of training epochs, are selected to determine the optimal permutation of input parameters for the model’s performance.

TABLE 2. The description of all the models used in this work to compare with. SL stands for supervised learning, EL stands for ensemble learning and ANN stands for artificial neural network.

|  |  |  |
| --- | --- | --- |
| **Model** | **Type** | **Description** |
| LR | SL | Logistic Regression models the probability of an input belonging to a particular class using a logistic (sigmoid) function. |
| SVR | SL | It is an extension of Support Vector Machines (SVM) for regression, where it tries to find a hyperplane that best fits the data while minimizing the error. |
| RF | EL | Random Forest is based on the concept of constructing multiple decision trees during training and then combining their predictions to make a final prediction. |
| KNR | SL | KNN Regressor is a non-parametric supervised learning algorithm used for regression tasks. |
| FNN | ANN | A Feedforward Neural Network is an artificial neural network that consists of a series of linear layers. |
| XGB | EL | It is a variant of gradient boosting that builds multiple weak learners (often decision trees) sequentially, where each new learner corrects the errors made by the previous ones. |
| RNN | ANN | RNN is a type of neural network designed to handle sequential data, making it suitable for time-series analysis. Unlike traditional feedforward neural networks, RNNs have connections that form cycles, allowing them to maintain hidden states and process sequences of data. |

Machine learning models are implemented using Scikit-learn and PyTorch, two machine learning libraries for Python. The data processing methods and models are all implemented in Python and computed under the following machine speculations: Ubuntu 22.04.2 LTS, AMD Ryzen 9 5900, NVIDIA RTX 3080 with 16GB RAM, 16GB of system RAM.

# III. Results

Figure 5 shows the difference of MSE and score during the training and testing procedure when using original dataset and the transposed dataset for the model LSTM+MA.

**A graph of a graph

Description automatically generatedA graph of a graph

Description automatically generated**

1. (b)

**Figure 5.** Graphs that compare the MSE and scores in the training and testing procedure with the original dataset and the transposed input dataset for the LSTM+MA model: (a) MSE; (b) scores.

It shows that no matter if you use original data or the transposed dataset, the predicting MSE is decreasing with the increasing of running epochs and the score increases with more running iterations. However, when the transposed and original time-series datasets are compared by training models with each dataset and looking at the MSE and over the training sequence, it can be seen that the transposed data leads to a significant improvement in training time and accuracy as shown in Figure 5. By utilizing the transposed data, the MSE drops dramatically and quickly to a very low state around epoch 25 which is even lower than the model in epoch 200 when using original data. In other words, the model learns from the data much quicker when the LTPP data is transposed. Moreover, the method with transposed data as preprocessing performs much smoother in MSE and . The model trained on the original data shows a sawtooth shape on the relationship between MSE (or ) with epochs. It means the model cannot learn very stable from the original data.

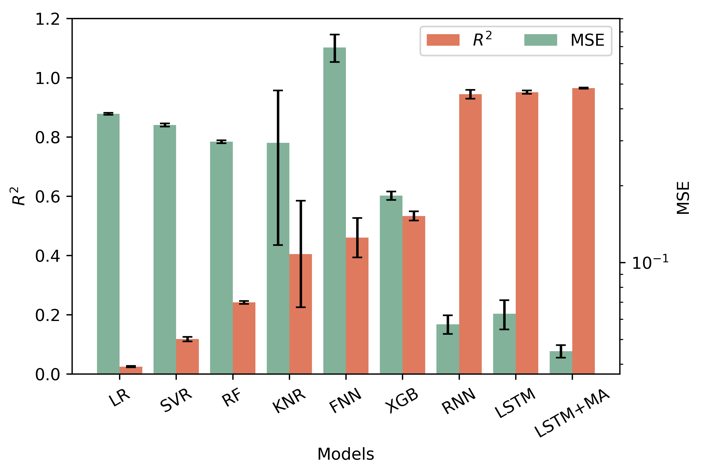
It is interesting to find that the test data closely mirrored the performance of the models on the training data no matter using original dataset or processed dataset, which means that the model demonstrates robustness to noise and generalize well to unseen data.

After preprocessing the data, a grid-search approach is utilized to determine the optimal parameters for the LSTM+MA model. The input sequence length, the learning rates, the leaning-rate decay, the LSTM layer dropout, the number of LSTM layers, the batch size, and the number of training epochs were selected to be optimized. We discover that the optimal sequence length for the LSTM-MA model is 5 elements and the optimal number of hidden layers for the encoder and decoder is 3. The optimal parameters from the hyperparameter tuning are shown in Table 3.

TABLE 3. Hyperparameter tuning results

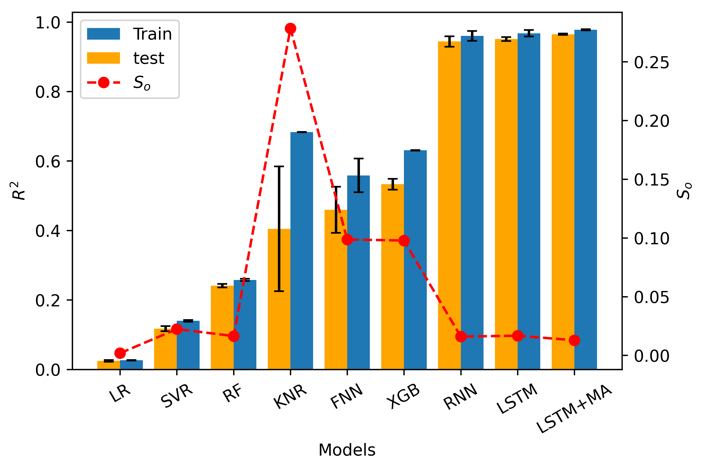
|  |  |
| --- | --- |
| **Parameter Name** | **Value** |
| Learning Rate | 0.01 |
| Number of hidden Layers | 3 |
| Input Sequence Length | 5 |
| Learning rate decay | 0.8 |
| Batch size | 512 |
| Epochs | 200 |
| Dropout | 0.25 |

After determining the best performance of our proposed model, we compare this model to other state-of-the-art models including LR, SVR, RF, KNR, FNN, XGB, RNN, and LSTM. Figure 6 shows the and MSE values from the models’ comparison results.



**Figure 6.** The and MSE value of all the models in test dataset. The models are organized in ascending order based on the which is achieved on the test dataset. The red bar stands for the and the green bar stands for the MSE. The black error bar stands for the standard deviation;

Figure 6 is organized in ascending order by value of each model achieved on the test dataset. It shows that our proposed LSTM+MA model outperforms other models in the LTPP dataset as it gets the highest value (0.965) and the lowest MSE (0.030). It is noteworthy that the models which consider impacts over time, including RNN, LSTM and LSTM+MA, have higher accuracy than the models which don’t consider the time-accumulated influence. The high performance of RNN, LSTM and LSTM+MA proves that the pavement IRI is a time-dependent factor and using time-series model can perform better on IRI prediction. The black error bar is the standard deviation calculated by running models in randomly split train-test datasets. A small standard deviation means the model performs stable in different data inputs. In other words, it can represent the robustness of the model. It is interesting to find that the KNR gets the apparent larger standard deviation than other models both in and MSE. This is because the KNR algorithm is predicted based on the K nearest points which means it is highly related to the interaction among the data points. Our proposed model gets the lowest standard deviation (0.0026) in value, which means it obtains a stable performance and is much more robust than other models.



**Figure 7.** The model performance on the train dataset and test dataset. The blue bar is the value of the model on the train dataset while the yellow bar is the score on test dataset. The red dash line is the overfitting score.

Overfitting is a serious problem in the long-term pavement performance prediction. It occurs when the model learns the training data too well, resulting in poor generalization. Therefore, to obtain the performance difference on the train and test dataset of each model, the score from the training and testing performance are considered and shown in Figure 7. The red dash line represents the overfitting score () among all the models. It shows that KNR is the most overfitted model compared to others. Although LR obtains the lowest overfitting problem with a of 0.0018, the performance of LR model is the worst among all the models. The LSTM+MA gets the second lowest (0.0128) except LR and at the same time, it obtains the highest value. In other words, our proposed model keeps high accuracy as well as a low overfitting problem in predicting the IRI value based on the LTPP dataset.

# IV. Discussion

The proposed LSTM+MA model adjusting the LSTM to make it combine with multi-head attention reaches the highest value compared to other models. It leads to a higher performance than using a plain LSTM, which means that the embedded attention mechanism does improve the model information extraction and interpretation capacity.

The model that can handle time-series data shows a dominant performance than other models. This is because the pavement IRI, as well as friction, rutting and deflection, are all time-series data, which means the traffic, climate and other factors would have an accumulated influence on these performances over time. In other words, our presented model is not only limited in predicting the pavement IRI, but also can be used to predict other pavement performance including friction, deflection and transverse profile as they are all time-series data. Moreover, it can also be used in other fields like stoke price predicting or weather prediction where the data is highly time related.

Although there are important improvements revealed by this study, there are also limitations. First, the choosing of predictors are not considered thoroughly. We must point out that we do not discuss the dependency between each predictor which might introduce redundancy and overfitting into the model. However, these problems could be solved if we consider using a feature/component ablation section to choose the most related predictors. Also, we could apply Principal Component Analysis (PCA) to reduce the dimensionality of data to avoid noise and redundancy.

# V. Conclusion

A LSTM based model, LSTM+MA, is proposed in this work to predict the pavement IRI based on time-series data extracted from the LTPP database. Traffic, climate, pavement construction and maintenance are considered as the most important factors to the contribution of IRI and are utilized as predictors to train the models. The original data are transferred to sequences first in order to improve usage of data in the model. Data transpose are utilized to allow the model to process the entire series of events at the same time. It demonstrates that transposing the data matrix can apparently accelerate the convergence speed and helps the model to learn from data more quickly and adequately.

The presented LSTM+MA model combines two LSTM layers as an encoder and decoder to extract the time-series information from the dataset. A multi-head attention layer is connected with the LSTM layer to allow the model to attend to different positions within the input sequence simultaneously, which can improve the model's interpretability by highlighting the importance of different parts of the input sequence. Hyperparameter fine-tuning are utilized to obtain the best performance of our proposed approach. State-of-the-art models including LR, SVR, RF, KNR, FNN, XGB, RNN and LSTM are implemented and compared with our presented method. Results show that our proposed LSTM+MA achieves a higher accuracy than other models as it obtains the highest value (0.965) and the lowest MSE (0.030). In addition, the overfitting problem is studied and an overfitting score is proposed in this work to quantify the degree of overfitting of models. It shows that the LSTM+MA model obtains the second lowest overfitting score (0.0128) except LR. The proposed method keeps high accuracy and low overfitting degree in predicting the IRI value based on the LTPP dataset.

Our proposed LSTM+MA mainly addresses the difficulty on precisely predicting the time-dependent evolution of pavement indexes, like IRI. Specially, IRI values should increase during the regular service period due to deterioration. However, IRI values can decrease as well. This is due to the maintenance work, like seal coating, chip sealing, micro surfacing, etc. To precisely predict future IRI values based on existing dataset is challenging. The uncertainty can affect the allocation of pavement maintenance resources. Our proposed method is one of initial attempts to shed light on this challenging area. It is not only a potential application for pavement agency but also a significant improvement for the Intelligent Transportation System.

References

[1] Q. Zhou, E. Okte, and I. L. Al-Qadi, "Predicting pavement roughness using deep learning algorithms," *Transportation Research Record,* vol. 2675, no. 11, pp. 1062-1072, 2021.

[2] H. Gong, Y. Sun, X. Shu, and B. Huang, "Use of random forests regression for predicting IRI of asphalt pavements," *Construction and Building Materials,* vol. 189, pp. 890-897, 2018.

[3] M. Mazari and D. D. Rodriguez, "Prediction of pavement roughness using a hybrid gene expression programming-neural network technique," *Journal of Traffic and Transportation Engineering (English Edition),* vol. 3, no. 5, pp. 448-455, 2016.

[4] M. R. Kaloop, S. M. El-Badawy, J. W. Hu, and R. T. Abd El-Hakim, "International Roughness Index prediction for flexible pavements using novel machine learning techniques," *Engineering Applications of Artificial Intelligence,* vol. 122, p. 106007, 2023.

[5] Y. Song, Y. D. Wang, X. Hu, and J. Liu, "An efficient and explainable ensemble learning model for asphalt pavement condition prediction based on LTPP dataset," *IEEE Transactions on Intelligent Transportation Systems,* vol. 23, no. 11, pp. 22084-22093, 2022.

[6] F. Damirchilo, A. Hosseini, M. Mellat Parast, and E. H. Fini, "Machine learning approach to predict international roughness index using long-term pavement performance data," *Journal of Transportation Engineering, Part B: Pavements,* vol. 147, no. 4, p. 04021058, 2021.

[7] Y. Deng and X. Shi, "An accurate, reproducible and robust model to predict the rutting of asphalt pavement: neural networks coupled with particle swarm optimization," *IEEE Transactions on Intelligent Transportation Systems,* vol. 23, no. 11, pp. 22063-22072, 2022.

[8] X. Wang *et al.*, "A hybrid model for prediction in asphalt pavement performance based on support vector machine and grey relation analysis," *Journal of advanced transportation,* vol. 2020, pp. 1-14, 2020.

[9] C. C. Ikeagwuani and D. C. Nwonu, "Statistical analysis and prediction of spatial resilient modulus of coarse-grained soils for pavement subbase and base layers using MLR, ANN and Ensemble techniques," *Innovative Infrastructure Solutions,* vol. 7, no. 4, p. 273, 2022.

[10] G. I. Webb and Z. Zheng, "Multistrategy ensemble learning: Reducing error by combining ensemble learning techniques," *IEEE Transactions on Knowledge and Data Engineering,* vol. 16, no. 8, pp. 980-991, 2004.

[11] C. Wang, S. Xu, and J. Yang, "Adaboost algorithm in artificial intelligence for optimizing the IRI prediction accuracy of asphalt concrete pavement," *Sensors,* vol. 21, no. 17, p. 5682, 2021.

[12] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Physica D: Nonlinear Phenomena,* vol. 404, p. 132306, 2020.

[13] T. Zhang, D. Wang, A. Mullins, and Y. Lu, "Integrated APC-GAN and AttuNet Framework for Automated Pavement Crack Pixel-Level Segmentation: A New Solution to Small Training Datasets," *IEEE Transactions on Intelligent Transportation Systems,* vol. 24, no. 4, pp. 4474-4481, 2023.

[14] T. Zhang, D. Wang, and Y. Lu, "ECSNet: An Accelerated Real-Time Image Segmentation CNN Architecture for Pavement Crack Detection," *IEEE Transactions on Intelligent Transportation Systems,* 2023.

[15] R. Abd El-Hakim and S. El-Badawy, "International roughness index prediction for rigid pavements: an artificial neural network application," *Advanced Materials Research,* vol. 723, pp. 854-860, 2013.

[16] C. Han, T. Ma, S. Chen, and J. Fan, "Application of a hybrid neural network structure for FWD backcalculation based on LTPP database," *International Journal of Pavement Engineering,* vol. 23, no. 9, pp. 3099-3112, 2022.

[17] Y. Dong *et al.*, "Forecasting pavement performance with a feature fusion LSTM-BPNN model," in *Proceedings of the 28th ACM international conference on information and knowledge management*, 2019, pp. 1953-1962.

[18] X. Chen, H. Zhu, Q. Dong, and B. Huang, "Optimal thresholds for pavement preventive maintenance treatments using LTPP data," *Journal of Transportation Engineering, Part A: Systems,* vol. 143, no. 6, p. 04017018, 2017.

[19] H. Sak, A. Senior, and F. Beaufays, "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition," *arXiv preprint arXiv:1402.1128,* 2014.

[20] A. Vaswani *et al.*, "Attention is all you need," *Advances in neural information processing systems,* vol. 30, 2017.