IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 9, 2020, accepted January 15, 2021, date of publication February 2, 2021, date of current version February 10, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3056568

Emerging Technologies of Deep Learning Models Development for Pavement Temperature Prediction

ABDALRHMAN MILAD[1], IBRAHIM ADWAN[1], SAYF A. MAJEED[2], NUR IZZI MD YUSOFF[1], NADHIR AL-ANSARI[3], AND ZAHER MUNDHER YASEEN[4]

[1]Department of Civil Engineering, Universiti Kebangsaan Malaysia, Selangor 43600, Malaysia
[2]Technical Computer Engineering, Al-Hadba'a University College, Mosul 41001, Iraq
[3]Department of Civil, Environmental and Natural Resources Engineering, Lulea University of Technology, 97187 Lulea, Sweden
[4]Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam

Corresponding authors: Abdalrhman Milad (miladabdalrhman@siswa.ukm.edu.my) and Zaher Mundher Yaseen (zahermundheryaseen@duytan.edu.vn)

This work was supported by Universiti Kebangsaan Malaysia under Grant GUP-2018-094.

ABSTRACT Air temperature is one of the critical factors influencing the bearing ability and performance of temperature-sensitive asphalt materials. This research investigates the relationship between air temperature at different depths and time to predict asphalt pavement temperature and evaluate asphalt performance. This paper discusses four deep learning-based regression models for calculating asphalt pavement temperature based on air temperature, depth from the asphalt surface, and time. Measurement of pavement temperature was made in the Gaza Strip. Monitoring stations were set up to measure asphalt pavement temperature and air temperature at different depths and times. The data were collected by hand measurement for the period from March 2012 to February 2013. The data is trained and validated using the Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM), and Gated Recurrent Unit (GRU). Bi-LSTM has an R^2 of 0.9555 for the generated dataset and outperforms other algorithms because of its superiority in feature extraction and multidimensional data processing. Through deep learning techniques, Bi-LSTM has demonstrated outstanding robustness and promising potential in predicting asphalt pavement temperature.

INDEX TERMS Geophysical monitoring, deep learning, Gaza Strip, pavement temperature, prediction, LSTM, GRU.

I. INTRODUCTION

Temperature is a critical factor in the design and analysis of pavements. There are three critical factors in pavement design. First, asphalt material properties change with temperature because of the viscoelastic behavior of the asphalt. Second, the asphalt pavement may experience fatigue damage because of thermal contraction and expansion [1]. Third, transverse top-down cracking may occur in asphalt pavements due to the extreme temperature drop in a day.

Researchers have developed regression models for predicting asphalt temperature by measuring pavement temperature using solar radiation and air temperature (AirT) [2]–[10]. These models produced false data for continuous temperature variations because they did not consider the environmental factors that may influence the surface temperature prediction, such as wind speed, wind direction, and relative humidity (RH). Solaimanian *et al.* introduced a system for predicting pavement temperature using AirT and solar radiation [11]. The model assumed a high-temperature equilibrium. However, Hermansson revised this model by considering a wind speed of 4 m/s [12]. Hermansson designed a new model for measuring summer pavement temperature by considering AirT, airspeed, and solar radiation and validated the model using one-month long-term pavement results [13]. However, the model validation did not consider the RH and wind direction for the summer data. Wang *et al.* used a different approach to design an analytical, numeric

The associate editor coordinating the review of this manuscript and approving it for publication was Danilo Pelusi.

model for predicting pavement surface temperature based on AirT, solar radiation, and RH. There is a 5°C difference between the predicted and actual value at very high temperatures [14].

The analytics-based models for pavement temperature prediction were developed and introduced later. Feng *et al.* designed a numeric model for solving surface energy equations to estimate surface temperature by considering solar radiation and heat flux. The model was validated using 24-hour data, which is moderately small. Validation of multiple datasets is critical to validate the model because limited datasets could produce biased results. The model also disregarded other environmental variables, such as RH and wind direction [15]. Yavuzturk *et al.* designed a model for predicting hourly temperatures anywhere on the pavement. The model considered thermal conditions, including temperature, solar radiation, geometry, pavement orientation, wind speed, and thermal properties of pavement. However, the approach for validating the model, which considered the thermal properties of asphalt concrete (AC), is problematic [10]. Wang *et al.* designed an algorithm for predicting a one-dimensional temperature profile in a multi-layered pavement using the ACs thermal properties and pavement surface temperature. The Kallas research validated this algorithm in 1964-1965. Despite this, the algorithm is no longer valid because of the tremendous improvement in the design and manufacture of AC mixes [16].

These findings were confirmed using field data calculated from the Kallas study during 1964–1965. This type of old data may not be valid in today's asphalt conditions as significant changes have been made in AC mix design and construction. The temperature-based model designed by Marshall or Hveem can't make an accurate prediction for AC, except for those designed using the Superpave mix design process [1].

A review of the different approaches revealed that, at present, there is no system for predicting the Asphalt Pavement Temperature (APT) at different depths and AirT. The effect of AirT at different depths and times on APT predictions is still uncertain. For this reason, this study considers AirT and depth in the development of a regression-based model for determining surface temperature.

Researchers throughout the globe are using neural networks and deep learning techniques to predict APT. Among the techniques often used to predict time-series patterns are Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (Bi-LSTM), and Gated Recurrent Unit (GRU).

Hochreiter and Schmidhuber proposed using LSTM to examine the features that vary with time. Thus, exploring the dynamic nature of rainfall-induced displacement is more appropriate [17], [18]. The benefit of using LSTM is fewer long-term dependence problems than Recurrent Neural Network (RNN). RNN has a straightforward structure but cannot use long-term details in the training process because of a lack of critical information. Even though LSTM has the same architecture as RNN, there are variations in the recurrent modules. Unlike the CNN approach, four layers communicate, often to address long-term dependency problems. To predict APT, the LSTM model is used to test the time-domain dynamic characteristics. The APT prediction was made by evaluating the AirT at different depths and times. The model was validated by comparing the predicted values with one-year measured values.

This research contributes to the field of APT prediction in the following ways. This study collected the data for AirT, depth of pavement at a different time from the entire Gaza Strip. It achieved a high determination coefficient (R2), minimum Mean Absolute Error (MAE), Minimum Mean Square Error (MSE), minimum Mean Absolute Percentage Error (MAPE), and minimum execution time of the model by using different deep learning-based prediction models to develop a robust APT prediction system and fine-tune the training parameters.

II. NEURAL NETWORK MODELS

A. CONVOLUTIONAL NEURAL NETWORK (CNN) MODEL

CNN and RNN are the two frequently used neural network architectures. CNN architecture is appropriate for two-dimensional data processing. The traditional convolutional neural network architecture is designed with a Convolution Layer (CL), activation function layer, pooling layer, and fully connected layer (FCL) [19]. Figure 1 shows the structure of a CNN model for predicting APT.

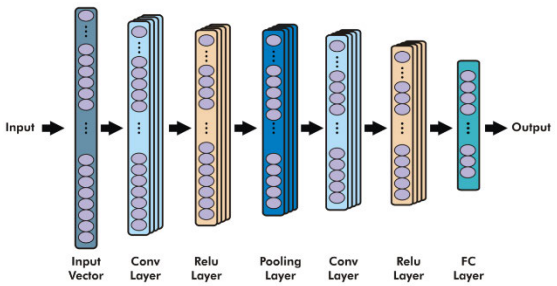

FIGURE 1. CNN structure for asphalt temperature prediction.

The convolution layers perform the convolution operation for extracting the local features. The convolutional operation for 2D data is given by equation (1) [12]:

$$S_{i,j} = \sum_{a=0}^{m} \sum_{b=0}^{n} I_{i-a,j-b} K_{m+a,n+b} \tag{1}$$

where I is the matrix of the input variable, K is the convolutional filter, a and b are the row and columns of the input matrix, and $S_{(i,j)}$ is the output of the convolutional operation at location (i, j). The pooling layer reduces the output neuron dimension by pooling or down-sampling and consolidating many values into one using max-pooling (MP). The FCL takes the output from the previous layer, process it using activation functions, and feed it as an input for the next layer.

In this study, the input parameters are time, AirT, and depth of the asphalt pavement, and the output is asphalt

temperature. This method has two convolutional layers, and each layer is followed by a batch normalization layer, ReLu layer, MP Layer of stride 2, and dropout layer with a dropout probability of 0.2. The performance consists of as many values as the number of filters and is connected to the FCL to predict APT [20].

B. LONG SHORT-TERM MEMORY (LSTM)

Hochreiter *et al.* proposed the LSTM model in 1997 [21], and later modified the model and elaborated the LSTM's forget gate [22]. LSTMs are a special type of RNN that can learn long-term dependencies and remember information for extended periods. A chain structure organizes the LSTM model [23]. Figure 2 shows the architecture of the LSTM neural network.

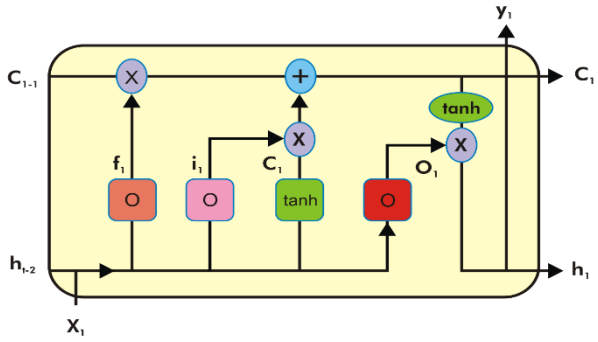

FIGURE 2. Structure of the LSTM network.

Figure 2 shows the cell in a box, which is the key to the LSTM. The horizontal lines across the section perform the computation. The cell's information state goes directly through the chain, with some linear interactions to keep the information running smoothly. The horizontal lines are used to add or discard the information to and from the cell. The sigmoid neuron layer is used to select the data. LSTM has three gates to protect and control cell status.

The **forget gate** is the first gate of LSTM that decides which information should be discarded from the cell. The outputs of the gate are $h_{t-1}$and x_t with 0 or 1 in the C_{t-1}cell state. The information is kept if the output 1 and discarded if the output is 0.

The **input gate** is the second gate that calculates the amount of information that should be added into the cell state and updated. It then uses a thin layer to generate a vector for editing the alternative c_{t-1} content. These steps decide the status of the cell. It multiplies the old state c_{t-1} with f_t, which is the output of the forget gate at time t. The c_{t-1} upgrades to c_t. The equations for f_t, c_{t-1}, and c_t are as follows (2)-(6).

$$f_t = \sigma_g(W_f x_t + U_f c_{t-1} + b_f) \tag{2}$$
$$i_t = \sigma_g(W_i x_t + U_i c_{t-1} + b_i) \tag{3}$$
$$O_t = \sigma_g(W_o x_t + U_o c_{t-1} + b_o) \tag{4}$$
$$c_t = f_t * c_{t-1} + i_t * \sigma_c\,(W_c x_t + b_c) \tag{5}$$
$$h_t = o_t * \sigma_h(c_t) \tag{6}$$

The Output gate is the third gate that decides the network output dependent on the cell status. The sigmoid function calculates the exportation of the cell. A thin component selects the cell state and multiplies it with the sigmoid function's output as shown in equations (7) -(8).

$$o_t = \sigma(W_o\,[h_{t-1}, x_t] + b_o) \tag{7}$$
$$h_t = o_t * \tanh(c_t) \tag{8}$$

where, W_o is the coefficient matrix for o_t, x_t is the input at time t, b_o is bias for o_t, h_t is the output of the current block. The size of the loss function and cross-entropy is given by (9).

$$J = -T_x * T_i = Y_i * log(h_i) \tag{9}$$

where T be the time period, Y_i is sample data, h_i is the LSTM output.

C. BIDIRECTIONAL LONG SHORT-TERM MEMORY (BI-LSTM)

Graves proposed Bi-LSTM, a reverse LSTM network that can better capture the bidirectional semantic dependence and is widely employed in natural language processing [24]. Bi-LSTM reverses the data, and the hidden layer synthesizes the forward and reverses information so that the network cells can simultaneously obtain context information. Figure 3 presents the Bi-LSTM structure.

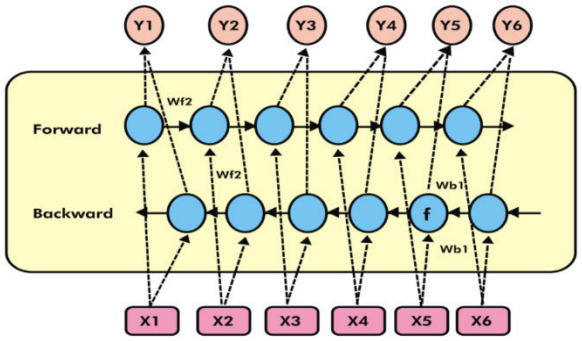

FIGURE 3. Diagram of the Bi-LSTM structure.

The Bi-LSTM estimation formula is as follows:

$$h_f = f(w_{f1} x_t + w_{f2} h_{t-1}) \tag{10}$$
$$h_b = f(w_{b1} x_t + w_{b2} h_{t+1}) \tag{11}$$

where h_f and h_b are the forward and reverse LSTM network output, respectively. The resultant hidden layer output is given by:

$$y_i = g(w_{o1} * h_f + w_{o2} * h_b) \tag{12}$$

D. GATED RECURRENT UNIT (GRU)

GRU is an LSTM-based model that optimizes the LSTM network structure while maintaining its performance. It has two gate systems, the update and reset gate, which solve the long delay in the time-series prediction problem [25].

The update gate tracks the amount of previous moment information passed to the current moment, while the reset gate controls the last information. Figure 4 presents the structure diagram of GRU.

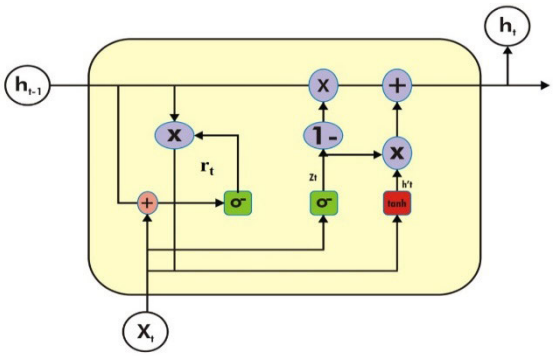

FIGURE 4. GRU structure diagram.

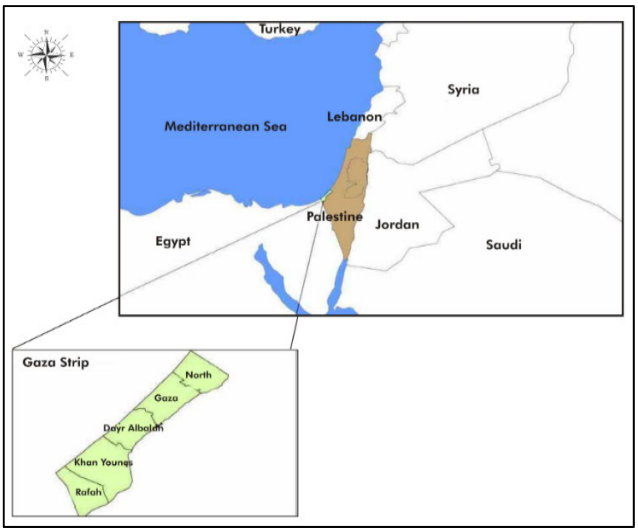

FIGURE 5. Location of the study area.

GRU performs better than LSTM in terms of computational cost and parameter generalization, and updates using the fixed number of training parameters on specific datasets [26].

In Figure 4, X_t is the input at time t, Z_t is the update gate output at time t, h_t and h_{t-1} are the output at time t and $t-1$, respectively, σ is the activation function, and r_t is the reset gate output at time t. Equations (13)-(17) give the calculations process of the memory unit.

$$r_t = \sigma(W_r * [h_{t-1}, X_t]) \tag{13}$$
$$Z_t = \sigma(W_z * [h_{t-1}, X_t]) \tag{14}$$
$$\hat{h}_t = \tanh(W_{\hat{h}} * [r_t * h_{t-1}, X_t]) \tag{15}$$
$$h_t = (1 - Z_t) * h_{t-1} + Z_t * \hat{h}_t \tag{16}$$
$$\hat{y}_t = \sigma(W_0 * h_t) \tag{17}$$

E. MODEL EVALUATION METHODS

The performance of each model was evaluated using MAE, MSE, MAPE, and R^2 [27], which were calculated using (18)-(21) [28], [29].

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_i - y_i\right| \tag{18}$$
$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2 \tag{19}$$
$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\hat{y}_i - y_i\right|}{\hat{y}_i} \tag{20}$$
$$R^2 = \left(\frac{\sum_{i=1}^{N}\left(y_i - \bar{y}\right)\left(\hat{y}_i - y\right)}{\sqrt{\sum_{i=1}^{N}\left(y_i - \bar{y}_i\right)^2}\sqrt{\sum_{i=1}^{N}\left(\hat{y}_i - y\right)^2}}\right)^2 \tag{21}$$

where $\hat{y}_i$ is the measured value, y_i is the predicted value, $\bar{y}_i$, is the mean of the measured value, and N is the number of sample sets.

III. METHODOLOGY

A. DATA DESCRIPTION

Measurement of the pavement temperature was made in the Gaza Strip. The monitoring stations were set up to measure pavement temperature and AirT at different depths (0, 2, 5.5, and 7 cm) and times in different seasons (winter, summer, spring, and autumn). The measurements were made during the period from March 2012 to February 2013. Figure 5 shows the geographical location of the study area.

This study developed an empirical statistical model to predict the APT for the typical Mediterranean weather condition in the Gaza Strip, which has specific climate characteristics. Gaza is located in the southwestern part of occupied Palestine on the eastern shore of the Mediterranean Sea within the tropic of cancer at latitude 31°30′0 ″N, 34°28′0.01″E. A total of 7,200 temperature measurements were made in the Gaza Strip. Figure 6 shows the time sequence distribution diagrams and data distribution histograms.

The distribution histogram shows that the data is evenly distributed. The time sequence distribution diagrams show that there are unique disruptions and noises in the data. There is no clear regulatory pattern in the time sequence, making it difficult to predict the APT.

B. DATA PRE-PROCESSING

The data contains outliers since measurements were made in the field. Data pre-processing is necessary to remove outliers. Standardization and normalization were performed to remove the outliers from the dataset and improve the trained model's performance [30]. Sections (1), (2), and (3) describe the pre-processing of the data.

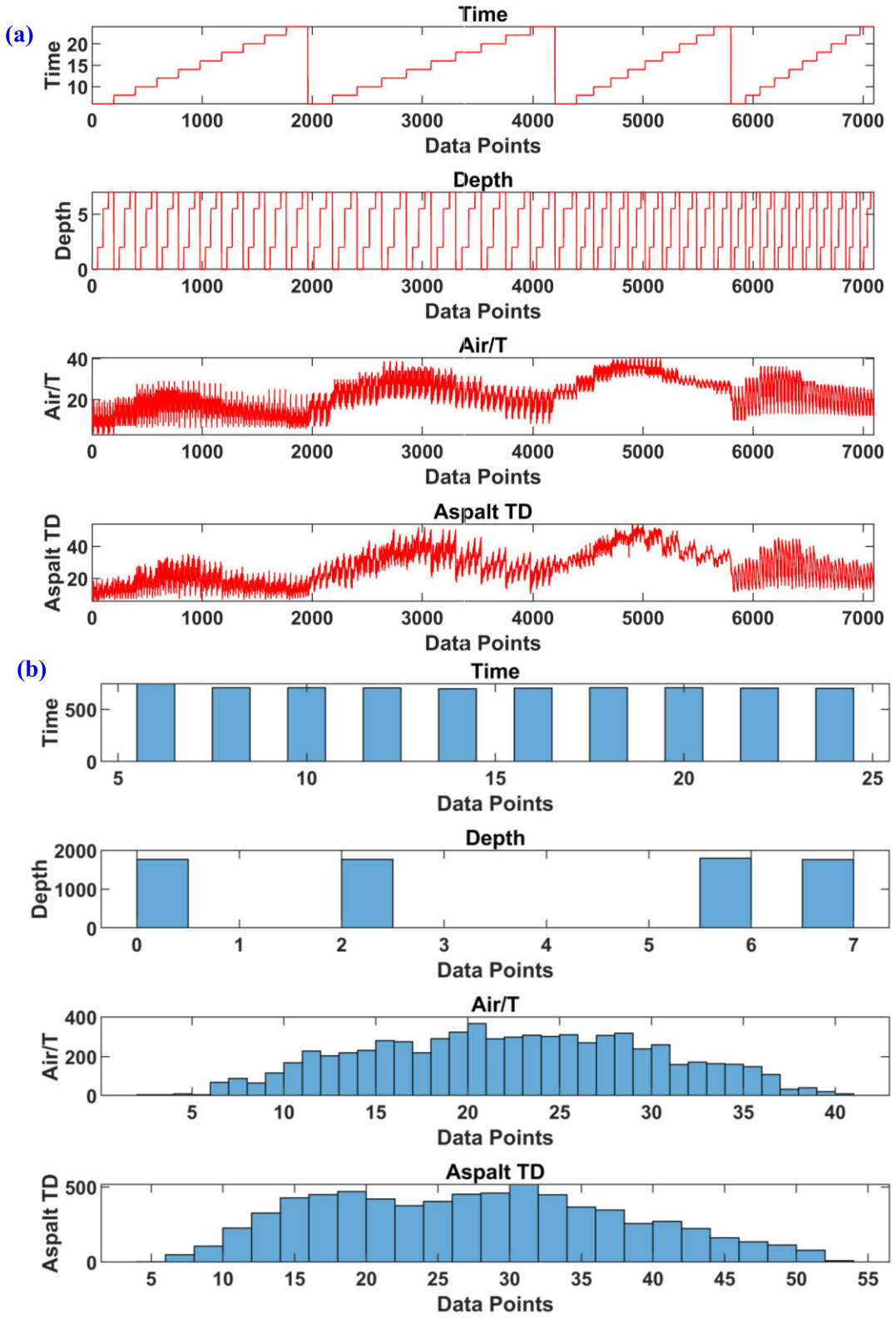

FIGURE 6. Distribution of each measured variable (a) Time sequence distribution diagrams (b) Data distribution histograms.

1) BOXPLOT ANALYSIS AND OUTLIER DETECTION

Boxplot is a data analysis method for detecting outliers in a dataset. The boxplot method is used when most of the parameters show variation at the higher end. A boxplot provides visualization of the outlier thresholds depending on the problem domain [31]. In this study, the boxplot analysis

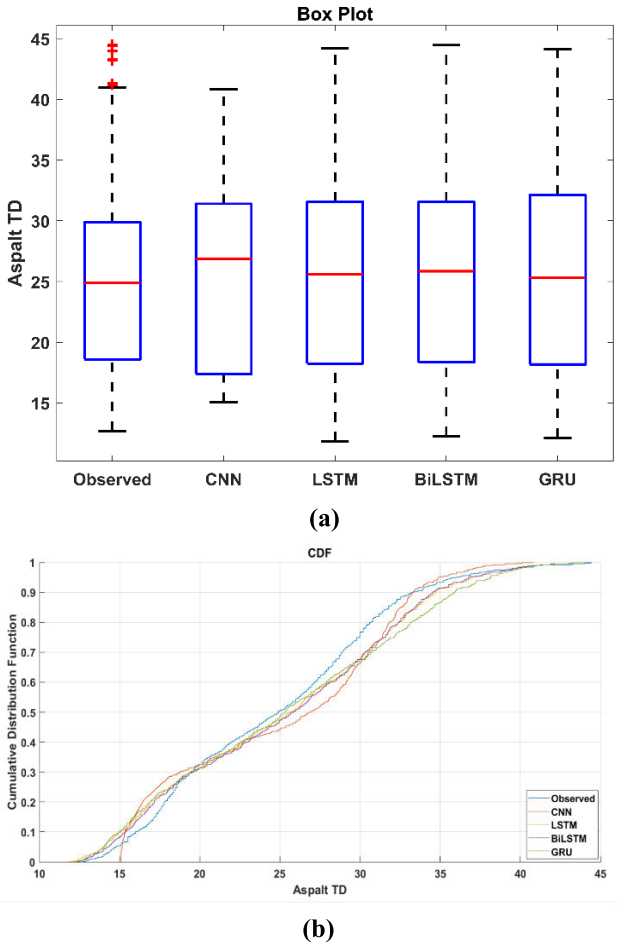

FIGURE 7. Representation of the measured and estimated asphalt temperature validation using CNN, LSTM, Bi-LSTM, and GRU models (a) Boxplot, and (b) Cumulative distribution function.

showed that most of the parameters lie outside the box and must be filtered out. The outliers were identified, assigned the highest threshold value, and removed manually to prevent data loss. The selection of the upper parameter threshold was not too lenient to avoid skewing the dataset and loosely penalize the values above the threshold value. Figure 7 shows the boxplot and cumulative distribution function of the observed and predicted APT of the validation data for CNN, LSTM, Bi-LSTM, and GRU models.

2) DATA STANDARDIZATION AND NORMALIZATION

Standardization is a technique of rescaling the data so that their standard deviation and mean are 1 and 0, respectively (22). This technique is beneficial if the data follows a Gaussian distribution and equalizes the range and data variability.

$$X' = \frac{X - \mu}{\sigma} \tag{22}$$

where μ is the mean of the feature values and σ is the standard deviation of the feature values.

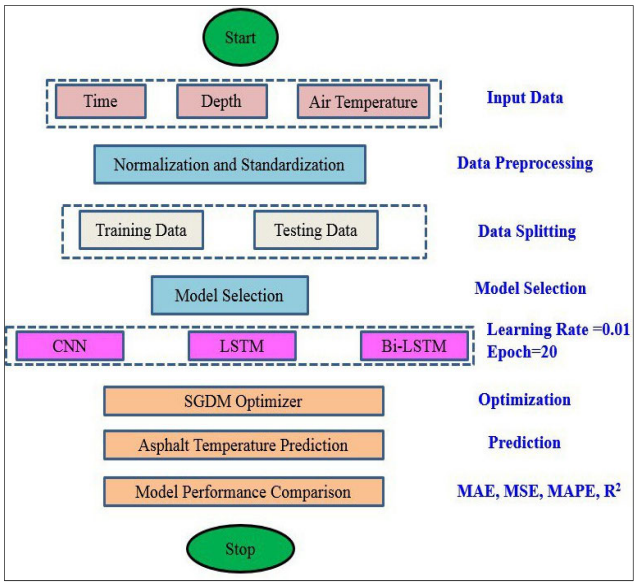

FIGURE 8. Flowchart for using the models to predict APT.

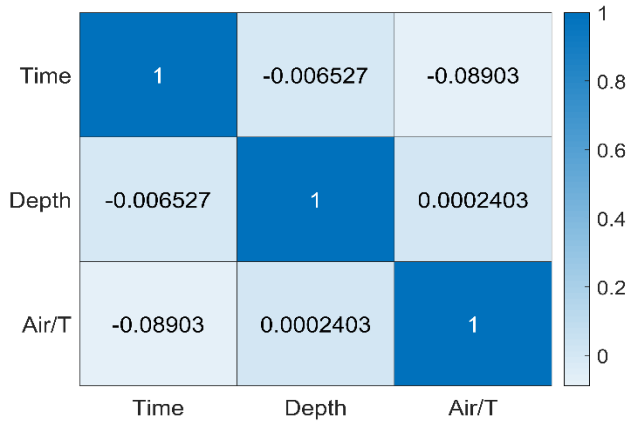

FIGURE 9. Pearson's correlation of the input parameters.

Normalization is a scaling technique that shifts and rescales the values within small ranges from 0 to 1 [32]. It is also called Min-Max scaling [33]. Normalization is often used in classification algorithms such as neural networks and nearest neighbor. The min-max normalization is mathematically expressed as (23).

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{23}$$

where, X_{max} and X_{min} are the maximum and the minimum values of the data variables, respectively.

C. METHODOLOGY FRAMEWORK

Figure 8 shows the methodology framework for predicting APT. The first step in this process is splitting the data into training and testing. In this study, the training dataset contains 6480 samples (80% of the dataset), and the testing dataset contains 720 samples (10% of the dataset) of collection

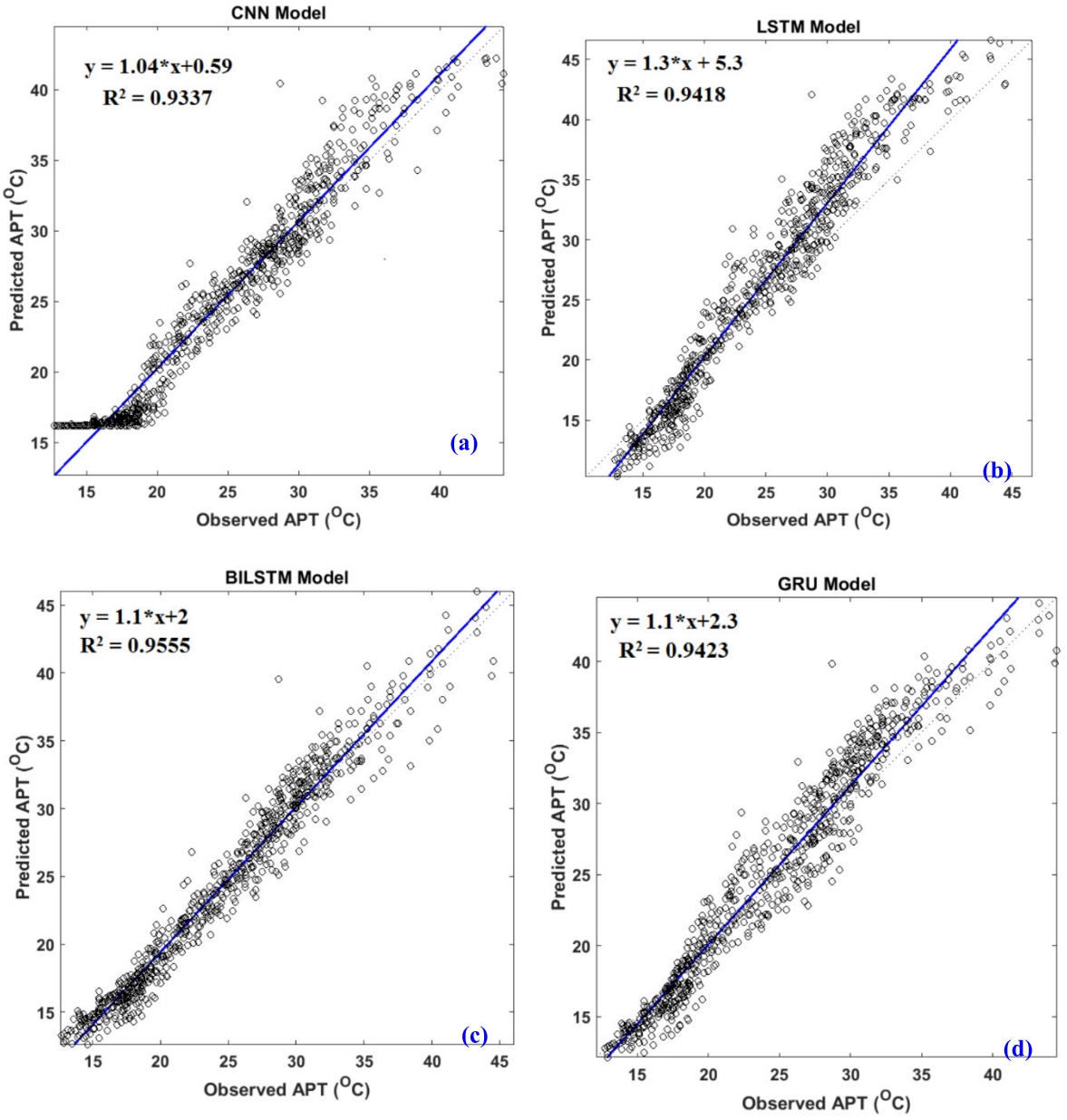

FIGURE 10. Distribution of the actual temperatures and the temperatures predicted by the (a) CNN, (b) LSTM, (c) Bi-LSTM, and (d) GRU models.

data. The model was able to predict advance day APT. This approach comprises four deep learning models, CNN, LSTM, Bi-LSTM, and GRU. Section 2 gives a detailed explanation of each algorithm. Table 1 presents the training parameters used to train the algorithms.

IV. RESULTS AND DISCUSSION

A. CORRELATION ANALYSIS

Pearson's correlation coefficient (PCC) is a proven metric for calculating the data's relationship probability. In this case, "X" and "Y" are field temperature measurements. The PCC between the pair of variables "X" and "Y" is given by (24).

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}_i)(y_i - \bar{y}_i)}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x}_i)^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2}} \quad (24)$$

TABLE 1. Training parameters.

Model	Learning Rate	Hidden Layers	Epoch	Optimizer	Learning rate drop factor	Learning rate drop period	Prediction Horizon
CNN	0.01	2	20	'sgdm'	0.1	5	1 day
LSTM	0.01	2	20	'sgdm'	0.1	5	1 day
Bi-LSTM	0.01	2	20	'sgdm'	0.1	5	1 day
GRU	0.01	2	20	'sgdm'	0.1	5	1 day

The PCC variables between time, depth, and AirT were calculated using correlation analysis, and the result shows a good correlation between the variables and can predict APT in advance.

TABLE 2. Performance of the trained model on validation data.

MODELS	MAE	MSE	MAPE	R^2
CNN	1.5082	3.5399	0.0623	0.9337
LSTM	1.6075	3.9172	0.0675	0.9418
BI-LSTM	1.3316	3.0132	0.0536	0.9555
GRU	1.6968	4.6708	0.677	0.9423

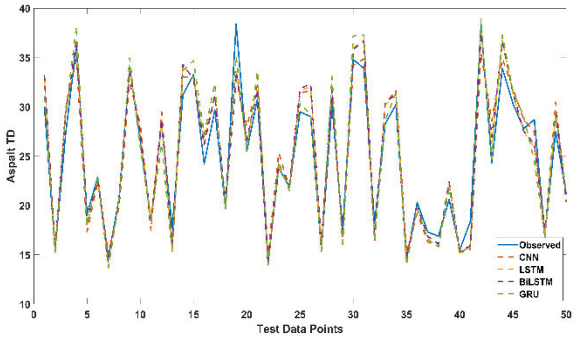

FIGURE 11. The measured APT and predicted APT for the CNN, LSTM, BiLSTM, and GRU models.

Figure 9 shows the PCC between time, depth, and AirT, which indicates a strong correlation between the variables. The PCC between depth and time is -0.006527. However, the PCC is -0.08903 and is lower than time and AirT, which were negatively correlated with the other parameters but positively correlated with one other. Therefore, all parameters were considered for LSTM modeling in this study. It also represents an absolute positive correlation between depth and AirT and thus can be used to develop an asphalt temperature prediction model. PCC can only be used as a comparison predictor because nonlinear relationships cannot be adequately tested.

B. MODEL PREDICTIONS

The results were implemented on an Acer aspire computer with 8GB RAM, Windows 10, 64 bit, and MATLAB 2020a version. Table 2 shows a comparison of the MAE, MSE, MAPE, and R^2 of the four models. It shows that Bi-LSTM has the best performance.

TABLE 3. Training time of CNN, LSTM, BI-LSTM, and GRU.

Models	Training Time (Sec)
CNN	14
LSTM	41
Bi-LSTM	42
GRU	130

Figure 10 shows the distribution of predicted and measured data for CNN, LSTM, Bi-LSTM, and GRU. Figure 10 (a) and (b) shows that the CNN and LSTM model were not sensitive to data change and have lower prediction accuracy. Figure 10 (c) and (d) show that the temperature

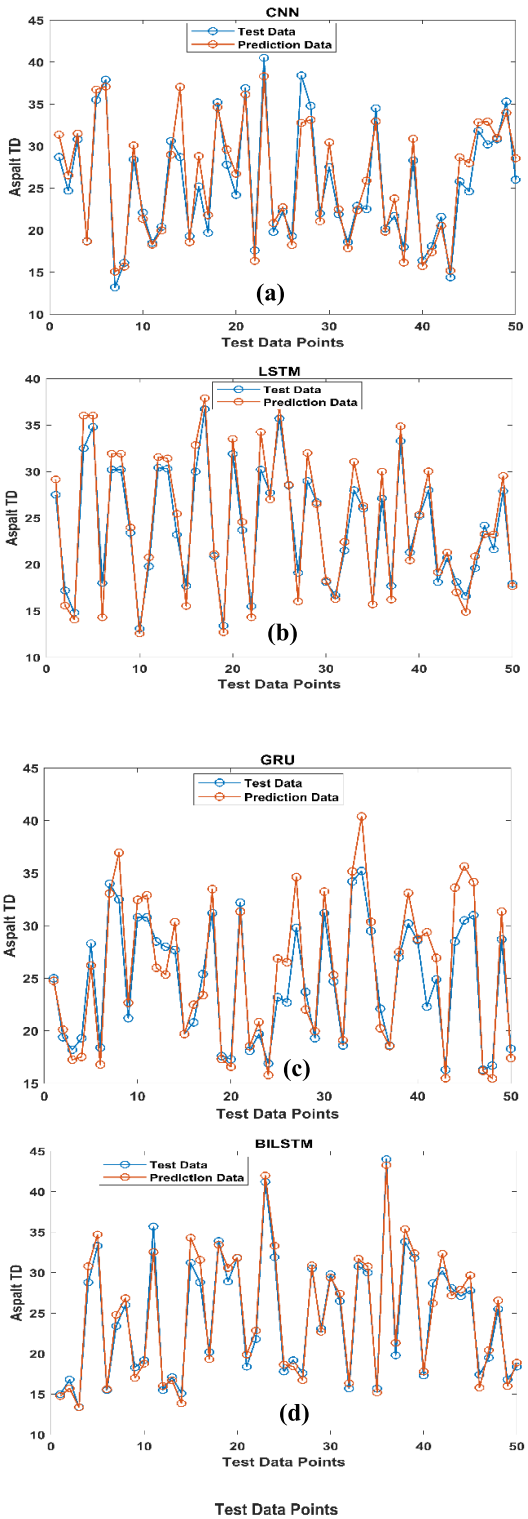

FIGURE 12. Comparison of the predicted and experimental values for (a) CNN, (b) LSTM, (c) GRU., and (d) BiLSTM.

predicted by BI-LSTM and GRU were close to the measured temperature.

Table 2 shows that the MAE, MSE, and MAPE for Bi-LSTM increased by 0.3652, 1.6576, and 0.6234 compared to the values for GRU. The MAE, MSE, and MAPE values for

Bi-LSTM increased by 0.2759, 0.904, and 0.0139 compared to those for LSTM, and 0.1766, 0.5267, and 0.0087, compared to those for CNN. The R^2 for Bi-LSTM is 0.9555 and is 0.0218, 0.137, and 0.0132 higher than those for CNN, LSTM, and GRU. These values show that Bi-LSTM is similar to GRU, and both models can accurately predict APT. Practical applications must consider the efficiency of training time. CNN has a much shorter training time than the other networks (refer to Table 3). However, the R^2 shows that Bi-LSTM has better performance. It is reliable with Bi-LSTM using a smaller number of gates. In conclusion, Bi-LSTM has the best performance and can be used in practical applications.

C. TRAINING AND VALIDATION RESULTS OF CNN, LSTM, BI-LSTM, AND GRU MODEL

Figures 11 and 12 show the measured APT and the APT predicted by CNN, LSTM, Bi-LSTM, and GRU for an epoch size of 20 and a batch size of 128. The x-axis represents the experimental data, and the y-axis represents the measured and predicted APT. Evaluation of model performance was based on MSE, MAE, MAPE, and R^2. The blue line represents the measured APT, and the red, yellow, violet, and green dotted lines represent the predicted APT for CNN, LSTM, Bi-LSTM, and GRU, respectively. The graph shows that the APT predicted by Bi-LSTM and GRU almost match the measured values, although large deviations caused some variations in the dataset. The peaks or discrepancies in the plots may be attributed to outliers. The prediction technique eliminates or plots these points by determining if such outliers were present in the training process. A comparison of Figures 11 and 12 with Table 2 show that the predicted values for CNN, LSTM, and GRU are higher and test data error than Bi-LSTM, particularly at high APTs, indicating that Bi-LSTM is stronger than CNN, LSTM, and GRU in predicting APTs.

V. CONCLUSION AND FINAL REMARKS

This paper is the first paper to make a comprehensive comparison of the performances of four deep learning prediction methods, CNN, LSTM, Bi-LSTM, and GRU. These methods predicted APT based on the AirT at different depths and times. Even though basic deep learning models such as CNN and LSTM can make good APT predictions, the performance Bi-LSTM and GRU are more superior based on the R^2values. The Bi-LSTM algorithm showed the best prediction capability after hyper-parameters tuning with R^2 values of 0.9337, 0.9418, 0.9555, and 0.9423 for CNN, LSTM, BI-LSTM, and GRU, respectively. The evaluation metrics showed that Bi-LSTM and GRU are superior to CNN and LSTM and have a minimum training time of 42 seconds, which is, equal to LSTM. This shows that the Bi-LSTM model is more suitable and applicable for real-time prediction of APT.

Future researchers should integrate loss balancing algorithms for multitask learning to enhance a difficult task's performance by using the experimental data collected for asphalt temperature prediction. This work has shown that Bi-LSTM is appropriate for predicting real-time APT based on the R^2 and computation time. However, some attempts can still be made in future work. Future studies should consider the relationships between parameters such as air temperature, solar radiation, wind speed, and RH in predicting APT.

CONFLICT OF INTEREST

The authors declare no conflict of interest to any party.

REFERENCES

[1] Z. H. Khan, M. R. Islam, and R. A. Tarefder, "Determining asphalt surface temperature using weather parameters," *J. Traffic Transp. Eng.*, vol. 6, no. 6, pp. 577–588, Dec. 2019.
[2] J. Chao and Z. Jinxi, "Prediction model for asphalt pavement temperature in high-temperature season in Beijing," *Adv. Civil Eng.*, vol. 2018, pp. 1–11, Jul. 2018.
[3] A. Asefzadeh, L. Hashemian, and A. Bayat, "Development of statistical temperature prediction models for a test road in Edmonton, Alberta, Canada," *Int. J. Pavement Res. Technol.*, vol. 10, no. 5, pp. 369–382, Sep. 2017.
[4] X. Shi, W. Lu, Y. Zhao, and P. Qin, "Prediction of indoor temperature and relative humidity based on cloud database by using an improved BP neural network in Chongqing," *IEEE Access*, vol. 6, pp. 30559–30566, 2018.
[5] I. D. Uwanuakwa, S. I. A. Ali, M. R. M. Hasan, P. Akpinar, A. Sani, and K. A. Shariff, "Artificial intelligence prediction of rutting and fatigue parameters in modified asphalt binders," *Appl. Sci.*, vol. 10, no. 21, p. 7764, Nov. 2020.
[6] Y. Li, L. Liu, and L. Sun, "Temperature predictions for asphalt pavement with thick asphalt layer," *Construct. Building Mater.*, vol. 160, pp. 802–809, Jan. 2018.
[7] B. Xu, H.-C. Dan, and L. Li, "Temperature prediction model of asphalt pavement in cold regions based on an improved BP neural network," *Appl. Thermal Eng.*, vol. 120, pp. 568–580, Jun. 2017.
[8] M. Alas and S. I. A. Ali, "Prediction of the high-temperature performance of a geopolymer modified asphalt binder using artificial neural networks," *Int. J. Technol*, vol. 10, no. 2, pp. 417–427, 2019.
[9] B. K. Diefenderfer, I. L. Al-Qadi, and S. D. Diefenderfer, "Model to predict pavement temperature profile: Development and validation," *J. Transp. Eng.*, vol. 132, no. 2, pp. 162–167, Feb. 2006.
[10] C. Yavuzturk, K. Ksaibati, and A. D. Chiasson, "Assessment of temperature fluctuations in asphalt pavements due to thermal environmental conditions using a two-dimensional, transient finite-difference approach," *J. Mater. Civil Eng.*, vol. 17, no. 4, pp. 465–475, Aug. 2005.
[11] M. Solaimanian and T. W. Kennedy, "Predicting maximum pavement surface temperature using maximum air temperature and hourly solar radiation," *Transp. Res. Rec.*, vol. 14, no. 17, p. 1, 1993.
[12] Å. Hermansson, "Mathematical model for calculation of pavement temperatures: Comparison of calculated and measured temperatures," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1764, no. 1, pp. 180–188, Jan. 2001.
[13] Å. Hermansson, "Simulation model for calculating pavement temperatures including maximum temperature," *Transp. Res. Rec., J. Transp. Res. Board*, vol. 1699, no. 1, pp. 134–141, Jan. 2000.
[14] T.-H. Wang, L.-J. Su, and J.-Y. Zhai, "A case study on diurnal and seasonal variation in pavement temperature," *Int. J. Pavement Eng.*, vol. 15, no. 5, pp. 402–408, May 2014.
[15] T. Feng and S. Feng, "A numerical model for predicting road surface temperature in the highway," *Procedia Eng.*, vol. 37, pp. 137–142, Jan. 2012.
[16] D. Wang, "Analytical approach to predict temperature profile in a multilayered pavement system based on measured surface temperature data," *J. Transp. Eng.*, vol. 138, no. 5, pp. 674–679, May 2012.
[17] X. Shiqing, "Present situation and advances of research on prediction methods of coal and gas," *Exp. Inf. Mining Ind.*, vol. 469, no. 5, pp. 21–23, 2008.
[18] Y. Leng, J. Peng, Q. Wang, Z. Meng, and W. Huang, "A fluidized landslide occurred in the Loess Plateau: A study on loess landslide in South Jingyang tableland," *Eng. Geol.*, vol. 236, pp. 129–136, Mar. 2018.
[19] F. A. Gers and E. Schmidhuber, "LSTM recurrent networks learn simple context-free and context-sensitive languages," *IEEE Trans. Neural Netw.*, vol. 12, no. 6, pp. 1333–1340, Nov. 2001.

[20] S. Lee, Y.-S. Lee, and Y. Son, "Forecasting daily temperatures with different time interval data using deep neural networks," *Appl. Sci.*, vol. 10, no. 5, p. 1609, Feb. 2020.
[21] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[22] F. A. Gers, J. Schmidhuber, and F. Cummins, "Learning to forget: Continual prediction with LSTM," *Neural Comput.*, vol. 12, no. 10, pp. 2451–2471, 2000.
[23] Y.-P. Ruan, Q. Chen, and Z.-H. Ling, "A sequential neural encoder with latent structured description for modeling sentences," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 26, no. 2, pp. 231–242, Feb. 2018.
[24] A. Graves and J. Schmidhuber "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," *Neural Netw.*, vol. 18, nos. 5–6, pp. 602–610, Jul. 2005.
[25] L. Wu, C. Kong, X. Hao, and W. Chen, "A short-term load forecasting method based on GRU-CNN hybrid neural network model," *Math. Problems Eng.*, vol. 2020, pp. 1–10, Mar. 2020.
[26] E. Diao, J. Ding, and V. Tarokh, "Restricted recurrent neural networks," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2019, pp. 56–63.
[27] M. Sajjad, Z. A. Khan, A. Ullah, T. Hussain, W. Ullah, M. Y. Lee, and S. W. Baik, "A novel CNN-GRU-based hybrid approach for short-term residential load forecasting," *IEEE Access*, vol. 8, pp. 143759–143768, 2020.
[28] D. V. Dao, H.-B. Ly, H.-L.-T. Vu, T.-T. Le, and B. T. Pham, "Investigation and optimization of the C-ANN structure in predicting the compressive strength of foamed concrete," *Materials*, vol. 13, no. 5, p. 1072, Feb. 2020.
[29] Y. Yu, J. Cao, and J. Zhu, "An LSTM short-term solar irradiance forecasting under complicated weather conditions," *IEEE Access*, vol. 7, pp. 145651–145666, 2019.
[30] S. Kandanaarachchi, M. A. Muñoz, R. J. Hyndman, and K. Smith-Miles, "On normalization and algorithm selection for unsupervised outlier detection," *Data Mining Knowl. Discovery*, vol. 34, no. 2, pp. 309–354, Mar. 2020.
[31] U. Ahmed, R. Mumtaz, H. Anwar, A. A. Shah, R. Irfan, and J. García-Nieto, "Efficient water quality prediction using supervised machine learning," *Water*, vol. 11, no. 11, p. 2210, Oct. 2019.
[32] M. M. Suarez-Alvarez, D.-T. Pham, M. Y. Prostov, and Y. I. Prostov, "Statistical approach to normalization of feature vectors and clustering of mixed datasets," *Proc. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 468, no. 2145, pp. 2630–2651, Sep. 2012.
[33] S. K. Panda, S. K. Bhoi, and M. Singh, "A collaborative filtering recommendation algorithm based on normalization approach," *J. Ambient Intell. Humaniz. Comput.*, pp. 1–23, Jan. 2020.

ABDALRHMAN MILAD was born in Tarhunah, Libya, in 1984. He received the Ph.D. degree in civil engineering from Universiti Kebangsaan Malaysia, in 2017. He was a Lecturer with the College of Civil Aviation Technology and Meteorology, Espiaa, Libya, in 2010. He is currently a Researcher with Universiti Kebangsaan Malaysia. His research interests include neural networks, deep learning, and pavement engineering.

IBRAHIM ADWAN is currently pursuing the Ph.D. degree with the Department of Civil and Structural Engineering, Universiti Kebangsaan Malaysia (UKM). His research interests include pavement engineering and pavement materials

SAYF A. MAJEED received the B.Sc. degree in electrical, electronics and communication engineering from the University of Mosul, Iraq, in 2004, and the M.Sc. degree in computer and communication engineering and the Ph.D. degree in electrical, electronics, and systems engineering from the National University of Malaysia (UKM), in 2008 and 2017, respectively. He is currently a Lecturer with the Department of Technical Computer Engineering, Al-Hadba'a University College, Mosul. He has authored several articles in various international conferences and journals. His current research interests include speech recognition, speech and audio processing, communications and multimedia computing, machine learning, deep learning applications, and pattern recognition.

NUR IZZI MD YUSOFF is currently a Senior Lecturer with the Department of Civil and Structural Engineering, Universiti Kebangsaan Malaysia (UKM). His research interests include pavement engineering and pavement materials.

NADHIR AL-ANSARI received the B.Sc. and M.Sc. degrees from the University of Baghdad, in 1968 and 1972, respectively, and the Ph.D. degree in water resources engineering from Dundee University, in 1976. He was with Baghdad University, from 1976 to 1995, then at Al AlBayt University, Jordan, from 1995 to 2007. He is currently a Professor with the Department of Civil, Environmental and Natural Resources Engineering, Lulea Technical University, Sweden. His research interests include geology, water resources, and environment. He served several academic administrative post, such as the Dean and the Head of the Department. His publications include more than 424 articles in international/national journals, chapters in books, and 13 books. He executed more than 60 major research projects in Iraq, Jordan, and U.K. One of his patent on Physical methods for the separation of iron oxides. He supervised more than 66 postgraduate students at Iraq, Jordan, U.K., and Australia universities. He is a member of several scientific societies, such as the International Association of Hydrological Sciences, the Chartered Institution of Water and Environment Management, and the Network of Iraqi Scientists Abroad. He is the Founder and the President of the Iraqi Scientific Society for Water Resources. He awarded several scientific and educational awards, among them is the British Council on its 70th Anniversary awarded him top 5 scientists in Cultural Relations. He is also a Member of the Editorial Board of ten international journals.

ZAHER MUNDHER YASEEN received the master's and Ph.D. degrees from the National University of Malaysia (UKM), Malaysia, in 2012 and 2017, respectively. He is currently a Senior Lecturer and a Senior Researcher in the field of Civil Engineering. He is major in civil engineering applications. In addition, he has an Excellent Expertise in machine learning and advanced data analytics. He has published over 190 research articles in international journals with a Google Scholar H-index of 33 and a total of 3,570 citations.