

APPLIED RESEARCH

A Novel Temporal Feature Selection Based LSTM Model for Electrical Short-Term Load Forecasting

KHALID IJAZ¹, ZAWAR HUSSAIN^{1,2}, JAMEEL AHMAD³, SYED FAROOQ ALI⁴,
MUHAMMAD ADNAN³, AND IKRAMULLAH KHOSA⁵, (Member, IEEE)

¹Department of Electrical Engineering, School of Engineering, University of Management and Technology, Lahore 54770, Pakistan

²School of Science and Engineering, Lahore University of Management Sciences, Lahore 54792, Pakistan

³Department of Computer Science, School of Systems and Technology, University of Management and Technology, Lahore 54770, Pakistan

⁴Department of Software Engineering, School of Systems and Technology, University of Management and Technology, Lahore 54770, Pakistan

⁵Department of Electrical and Computer Engineering, COMSATS University Islamabad, Lahore Campus, Lahore 54000, Pakistan

Corresponding author: Zawar Hussain (zawar.hussain@umt.edu.pk)

This work was supported by University of Management and Technology Lahore, Pakistan.

ABSTRACT An accurate electrical Short-term Load Forecasting (STLF) is an eminent factor in the power generation, electrical load dispatching and energy planning for the power supply companies, specifically in developing countries. This paper proposes a novel temporal feature selection-based Long Short-term Memory (LSTM) model developed by the combination of standard Artificial Neural Network (ANN) layer and LSTM for electrical short term load forecasting. The LSTM model has excellent capability of predicting the stochastic nature of an hour ahead electrical loads. The standard ANN layer consisting 11 neurons is used as an input to LSTM cells. Such a combination of ANN layer with LSTM was never proposed before. The proposed model accommodates variations in weather as well as temporal inputs like humidity, holidays, and date-time features in the hourly load data of the power supply company situated in Johor, Malaysia. This paper gives the insights of hyper parameter tuning to capture the more generalized electrical load patterns in the dataset without compromising the time complexity of the proposed model. The proposed approach was compared with five existing approaches, namely: ANN, LSTM model 1, LSTM model 2, LSTM model 3 and Convolutional Neural Network-LSTM (CNN-LSTM) using hourly load dataset of Johor. The experimental results demonstrate that the proposed approach outperformed the existing approaches in terms of root mean square error, mean absolute percentage error and Diebold-Mariano statistical inference test within 95% confidence interval.

INDEX TERMS Artificial neural network, deep learning, load forecasting, long short-term memory.

I. INTRODUCTION

Short-term Electrical Load Forecasting (STLF) is used by electric utilities to forecast a few minutes, hours or one week ahead electrical load [1]–[3]. STLF also plays a protagonist role in the secure, economic, and reliable operation of energy companies and is widely used in the power generation scheduling, fuel purchase scheduling, security analysis, and adjustment of tariff rates [4], [5]. An in-accurate electrical STLF causes irregular power flows and system's congestion

The associate editor coordinating the review of this manuscript and approving it for publication was Pierluigi Siano¹.

which degrades the security and protection of electrical power system giving rise to imbalanced generation planning. Therefore, electricity generation, transmission and distribution networks governed by electric companies over the world need an accurate estimate of STLF for reliable and economical short-term operations of power systems [6]. STLF also helps in economic dispatching of electrical load, voltage stability of high voltage alternating current (HVAC) and high voltage direct current (HVDC) lines, and for predicting highly disruptive blackouts [7].

The U.S. Energy Information Administration (EIA) collects, analyzes, and disseminates independent and impartial

energy information to promote sound policymaking, efficient markets, and public understanding of energy and its interaction with the economy and the environment. The annual energy outlook report published in 2021 (AEO2021) presents the long-term trends in electricity demand [8]. Electricity demand grows modestly throughout the projection period. This report further suggests that the U.S. electricity use percentage growth rate, with three-year rolling average, is less than one percent from 2020 to 2050 in the reference case. On the contrary, the developing countries require four times higher electricity demand than that in developed countries. The main reason of the increased electricity demand is the abrupt change in meteorological factors of the developing countries [9]. The need for robust and high performance forecasting methods is inevitable to predict a variational electricity demand in such a scenario. The electricity market of some developing countries still relies on the conventional statistical methodologies like Regression analysis, autoregressive integrated moving average (ARIMA) for STLF and Power Market Survey (PMS) to carry out future power generation planning [10]. For instance, the researcher has presented conventional Bagged Regression Tree, artificial neural network (ANN) and extreme gradient boosting (XGBoost) algorithm to elucidate the problem of STLF in Pakistan [11], [12]. However, the conventional methodologies are not suitable to handle the dynamics of power systems under nonlinear behavior of the meteorological and temporal data such as humidity, temperature, load during festivals and holidays for the STLF problem [13]. Therefore, the planning authorities of electric utilities in developing countries should be motivated to take advantage of the artificial intelligence (AI) based computational intelligence methods for high-performance STLF.

There is a dire need for a rigorous work, yet to be done, to motivate the electric utilities to deploy deep learning methodologies for planning of power generation in developing countries like Pakistan, India and Malaysia. However, in developed countries, several deep learning methods using ANN, long short-term memory (LSTM) and convolutional neural network-long short-term memory (CNN-LSTM) to solve STLF problem are found in the literature. To overcome the above-stated research gap, this study aims to exploit the potential, strengths and weaknesses of different deep learning models on a real-time data set for the motivation of electric utilities in the developing countries. For this purpose, this paper also presents a novel temporal feature selection-based LSTM model for one-hour ahead load prediction.

The initial step in implementing the methodology of the proposed framework is the accumulation of the raw data related to STLF. The accuracy of STLF heavily rely on the historic electrical load, meteorological and temporal factors. Therefore, the input exogenous variables have been carefully selected from the raw data in this study for highly accurate deep learning forecasting models. The input datasets are then pre-processed to convert into meaningful multi-variate time-series electrical load data. The STLF data is

then segregated into training, validation, and test data. After training and developing a proposed forecasting model based on Adam optimizer and hyper-parameters tuning such as low learning rate, high batch size and small number of neurons or LSTM units in the hidden layer, the evaluation will be performed to assess the accuracy of the DL forecasting models using key error indices, such as mean absolute percentage error (MAPE), root-mean square error (RMSE), absolute percentage error (APE) and statistical analysis such as Diebold-Mariano test. Finally, the results reveal that the proposed feature selection-based LSTM model performs comparatively better than the other DL methodologies such as conventional ANN, LSTM model 1, 2, 3 and CNN-LSTM.

II. LITERATURE REVIEW

The stochastic nature of weather-sensitive electrical load and multi-variate temporal data such as festivals, holidays and week-days make the STLF task very challenging and demanding [14]. A lot of research work has been consumed to unravel the STLF difficulties in the presence of non-linear electrical load data. Different researchers have presented various regression models, such as ARIMA and seasonal-ARIMA (SARIMA) in [15], [16]. ARIMA and SARIMA use lagged average values of STLF time series data to capture seasonal effects by using auto-correlation function (ACF) and partial auto-correlation function (PACF) analysis [17]. Moreover, a single linear regression and a multiple linear regression models have also been discussed in [18]. However, the statistical regression algorithms are not capable of extracting the non-linear electrical load patterns and temporal variations [19], [20].

To resolve the above stated issues for STLF, Principal Component Analysis (PCA) based on dimensionality reduction algorithm can be applied to strengthen the performance of statistical regression models [21]. However, the difficulty in the appropriate selection of the coefficients of co-variance matrix in PCA may lose key seasonal impact of temperature influences on the electrical load data [22]. In contrast, Singular Value Decomposition (SVD) is efficient in extracting both seasonal and random components and is more robust than PCA [23]. However, SVD is computationally expensive due to the calculation of its unitary matrix [24].

Machine Learning (ML) models have been implemented later to get rid of the time complexity problem of SVD in STLF problem. ML algorithms have delineated advancement in the performance of STLF by improving accuracy in dealing with non-linearity of the electrical load data, and accurate forecasting of the peaks of electrical load than the dimensionality reduction models and statistical regression models [25]–[27]. ML methods mainly consist of the ANNs which can handle the non-linear nature of the weather-sensitive loads during the prediction of electrical load forecasting [28]. The conventional ANN algorithm experiences overfitting problem when larger number of neurons and the hidden layers are used to extract the highly variational temporal and meteorological features in non-linear electrical load pattern.

Some other major drawbacks, such as complex hyper-parameters tuning problem due to highly diversified input data and vanishing gradient problem limit the applications of the ANN models [29].

The growth in Deep Learning (DL) has improved the accuracy of STLF models using highly differentiated input data in contrast to conventional ML algorithms. Recurrent neural networks (RNNs) are the modified architectures of feed-forward neural network (FF-NN), which apply their internal state to proceed the variable-length sequences of inputs. RNNs show better reliability and stability to build an STLF model to extract non-linear inputs and output relationship in an electrical load data. Similarly, RNN with selected auto-regressive features was also presented to improve the efficiency of STLF algorithm [30]. Unfortunately, the vanishing gradients issue persists in RNN. Moreover, RNNs were inadequate in capturing long term dependencies. RNN architectures are then modified later and a new variant of RNN has been implemented which is named as LSTM to overcome the vanishing gradient problem [31], [32].

LSTM also boosted the capability of learning long-term dependencies between weather-sensitive and temporal features present in electrical load curve patterns using the special gated mechanism [33]. LSTM also achieves better accuracy in the presence of a large amount of multi-dimensional input data to map the input-output non-linear relationship during training in each batch size, which effectively improves the performance of STLF. The multi-task learning (MTL) based LSTM model was also deployed to improve the generalization capability and load forecasting efficiency [34]. However, the implementation of large number of hidden layers, neurons and complex parameters sharing architecture to develop the input output non-linear relationship during training instigate overfitting. Similar day characteristics based hybrid Empirical Mode Decomposition (EMD)-LSTM was also constructed to minimize the STLF errors [35]. However, these LSTM based models overlooked the local trends in an electrical load pattern during clustering and constituting of Intrinsic Mode Functions (IMFs). Moreover, LSTM based models fail to learn local trends in an electrical load pattern due to uni-directional (forward) processing of the sequential data. Hence, the LSTM does not lessen the forecasting error to a desired extent [36].

The STLF forecasting accuracy has been further improved by deploying the new Hybrid DL models which can extract the local trends in multi-variate time-series electrical load patterns efficiently. Every constituent of the model adds advantages in the STLF forecasting problem. CNN-LSTM has already gained a well-desired reputation in under-considered STLF problem, in which CNN captures the local trends in an electrical load pattern influenced by the weather and temporal features in high-dimensional multi-variate electrical time series data. Whereas the LSTM model predicts the electrical load with better prediction accuracy [37]–[40]. However, hybrid CNN-LSTM model still combat with the overfitting issue due to the enormous number of hidden layers [41].

The STLF forecasting accuracy has been improved by deploying the new model using a combination of Convolutional Neural Network (CNN) with Fuzzy time series. However, a large number of parameters and convolutional layers in CNN also reduces the generalization capability [42].

A. ORIGINAL CONTRIBUTIONS

For our STLF problem, a proposed LSTM model has been implemented on a Malaysian dataset as mentioned in [42]. It has been observed from the careful study of electrical load data of Johor city of Malaysia that the holidays and working days are unavoidable forecasting agents in the forecasting engines which affect the electrical demand load. Fig. 1 demonstrates the electrical load profile from 04 January to 10 January for the year of 2009. The electrical demand load is higher on the working days, which are from 05 January to 09 January for the year of 2009. The maximum electrical demand load reaches to 65080 MW in these working days. However, the maximum electrical demand load for the Saturday is below 51000 MW. Because on Saturday, some banks and business companies of Johor city remains closed. Similarly, 04 January of the same year is Sunday, which is weekly off-day [43]. From Fig. 1, it can be easily concluded that the maximum demand load of Sunday for the Johor city reduces to 47,000 MW which is comparatively lower than that of the electrical demand load of other working days. The temporal and meteorological features have been incorporated in a dataset due to the above-discussed relationship between the working days, holidays, and the electrical demand load. Therefore, the new features added in the Malaysian dataset are holidays, working days, weekdays, and on-off days. The above stated temporal features have not been used in the previous work which supports unique contribution during pre-processing in this research study.

The other main contribution lies in the implementation of a novel temporal feature selection-based LSTM model which is a combination of ANN layer and cascaded LSTM cells. The ANN layer extracts the important temporal features. The captured features representing unique predictor matrix are then transferred to the LSTM cell for prediction. This type of combination of ANN layer with cascaded LSTM cells are not presented before for one-hour ahead electrical load forecasting. Furthermore, a new concept of hyper-parameters tuning of proposed LSTM model on the multi-dimensional exogenous variables is also presented, in which a higher batch size and very small learning rate during training is applied for better generalization capability. As the multivariate time features are changing very abruptly with the time, so it is necessary to develop a robust LSTM model based on temporal feature-selection for one-hour ahead electrical load forecasting. The Proposed LSTM model tends to overcome the overfitting and underfitting problems by using appropriate number of hidden layers and neurons in the novel temporal feature-selection based LSTM model so that the non-linear electrical load curve pattern can be easily captured during training.

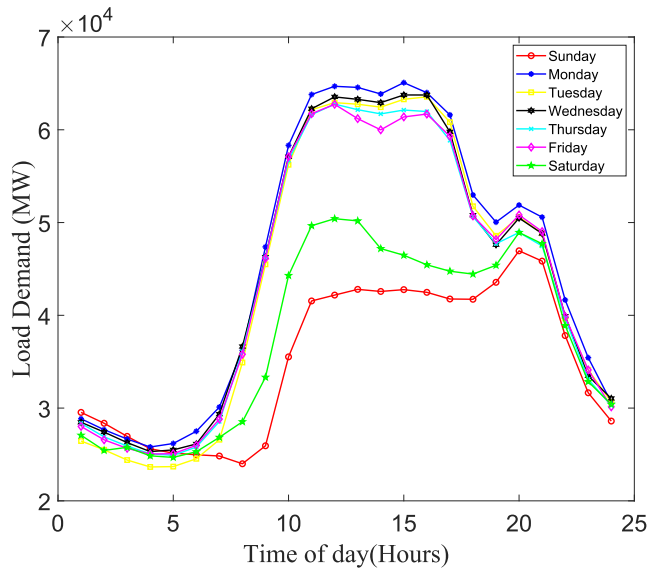


FIGURE 1. One week daily electrical load demand from January 04,2009 to January 10, 2009.

The proposed temporal feature selection-based LSTM model for one-hour ahead STLF proclaims the better forecasting accuracy than ANN at the cost of slightly increased time complexity. LSTM uses input gate, forget gate and output gate to capture, store and output the long-term information, temporal associations and their long-term dependencies in electrical load curve pattern which increases the time complexity of the LSTM architecture. LSTM effectively removes the vanishing gradient problem than ANN. Fortunately, the advancement in high-performance computing (HPC) servers and machines disentangle the issue of time complexity. Therefore, the modern research ruminates attention towards the better performance of forecasting algorithms, such as LSTM, for the reliability of real-time power system operations. However, the computational complexity and forecasting accuracy of the proposed LSTM model is much better than the LSTM model 1, LSTM model 2, LSTM model 3, and CNN-LSTM. At the end, the proposed LSTM proves to be a promising STLF forecasting engine.

According to the above-stated facts, the main contributions of this research paper can be recapitulated as follows:

- 1) A new combination of ANN and LSTM model is proposed in which ANN layer is infused as a temporal feature selection layer followed by the LSTM layer. A unique predictor matrix is formulated in which the significance of feature elements depends upon its corresponding weights generated by the ANN layer. This type of DL composition is presented for the first time for one-hour ahead electrical load forecasting.
- 2) The LSTM layer is used to capture the long-term and short-term dependencies between the important feature elements and input-output relationship of the electrical load dataset. The proposed model uses appropriate hyper-parameters for tuning, training, and improving

validation loss curve and deploying LSTM with suitable number of neurons in the hidden layer to avoid overfitting.

- 3) The non-linear spatio-temporal forecasting variables, which were not used before, are added in existing Malaysian dataset to improve prediction accuracy. Moreover, this study uses higher batch size during training on these spatio-temporal forecasting variables. To the best of the author's knowledge, the higher batch size was never used before for this type of STLF issue.
- 4) The proposed LSTM model can evenly be used on other electrical load datasets with small tuning of hyperparameters due to less complex architecture and small number of DL model parameters.
- 5) A comprehensive qualitative and quantitative analysis have been accomplished between the proposed temporal feature selection-based LSTM model, ANN, LSTM model 1, LSTM model 2, LSTM model 3, and CNN-LSTM. The evaluation metrics such as mean absolute percentage error (MAPE), root mean square error (RMSE), mean percentage error (MPE), time efficiency and Absolute Percentage Errors (APEs) show that the proposed LSTM model outperformed the six aforementioned models.
- 6) The Diebold-Mariano test, which is a statistical analysis, also demonstrates the superiority of the proposed LSTM model over the aforementioned DL architectures.

The remaining sections are organized as follows. Section III provides short-term load forecasting problem formulation. Section IV presents the proposed system methodology. Section V provides system configuration for DL simulations. Section VI provides discussion on results and Section VII concludes the paper.

III. SHORT-TERM LOAD FORECASTING PROBLEM FORMULATION

A. PROBLEM STATEMENT

The one-hour ahead electrical load forecasting problem can be developed as follows. For a given electrical STLF time-series dataset $X_t = \{f_1, f_2, \dots, f_n, l\}$, which consists of $(n + 1)$ historical data series, where $l = \{l_1, l_2, \dots, l_n\}$ represents the electrical load historical data, i are the samples taken at different time-stamp and $f_j = \{f_1^i, f_2^i, \dots, f_n^i\}$ is the historical data of the total n temporal and weather features that influences electrical load.

The ultimate target is to precisely forecast the electrical load at a future time $(t + T)$ indicated by $\hat{l}(t + T)$. The forecasted value $\hat{l}(t + T)$ is acquired by a neural-network defined by the function $F(\cdot)$. The proposed LSTM model determines the function $F(\cdot)$, which depicts the input-output relationship between the features from the feature set f_j and the electrical load l in such a way that the difference between the forecasted load $\hat{l}(t + T)$ and the actual load $l(t + T)$ at time instant $t + T$ should be minimized. The whole one-hour

ahead electrical load prediction problem can be expressed as shown in Eq. (1) and Eq. (2):

$$\hat{l}(t + T) = l(t + T) \tag{1}$$

$$e = |\hat{l}(t + T) - l(t + T)| \tag{2}$$

B. PERFORMANCE OF NEURAL NETWORK

The above stated problem can be resolved by implementing a high performance forecasting neural network. The performance of neural network is usually demonstrated by the mean square error (MSE) between the target $\hat{l}(t + T)$ and actual $l(t + T)$ load. The structure of a neural network includes input layer, hidden layer, and the output layer. The training of the neural network is governed by the weights of the neurons within these layers. So, the intention of optimal neural network during training is to find the set of neuron weights, which not only reduce the MSE between target and actual load but also eliminates the difference between training and testing error. Therefore, the error function $E(w)$ in terms of MSE defines the performance of neural network and can be mathematically expressed in Eq. (3):

$$\min E(W) = \frac{1}{N} \sum_{i=1}^N (\hat{l}(t + T, W) - l(t + T))^2 \tag{3}$$

where W is the set of neuron weights that minimize global MSE. Different gradient-based optimization algorithms can be used to tune the neuron weights such as Gradient Descent (GD) [44], Stochastic Gradient Descent (SGD) and Gradient Boosting Decision Tree (GBDT) [45]. The above methods lack the tendency to tune neurons weights in local minima and saddle points due to high dimensional temporal and weather factors. Moreover, the gradient based optimization algorithm cannot converge to the optimum point quickly. Adam algorithm has the capability to overcome the above mentioned problems by using two momentum concepts as shown in Eq. (4) and Eq. (5), which controls the learning rate during weights updating process [46], [47].

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{4}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{5}$$

The algorithm updates exponential moving averages of the gradient m_t and the squared gradient v_t where the hyper-parameters $\beta_1, \beta_2 \in [0, 1]$ control the exponential decay rates of these moving averages.

C. HYPER-PARAMETERS OF NEURAL NETWORK

The two most prominent hyper-parameters in designing optimal neural network are the number of neurons in the hidden layer and the number of hidden layers. A model with insufficient number of neurons in hidden layers experiences underfitting which propagates large training error. An adequate number of neurons in the hidden layer consequences in good generalization capability and the model combats with overfitting. However, the large number of hidden layers encounter overfitting problem. The proposed LSTM model consists of only one hidden layer with only ten neurons in the hidden layer to eradicate the overfitting problem.

IV. PROPOSED SYSTEM ARCHITECTURE AND METHODOLOGY

The proposed system framework is shown in Fig. 4. Raw data is collected and segregated which mainly consists of historic electric load, weather data containing humidity and temperature variation, and temporal data for days of the week and on/off days. The data is then pre-processed, on/off days and holidays are binary encoded, and then converted into a useful multi-variate STLF time series data. The whole multivariate STLF dataset is partitioned into training, validation, and the test datasets. The above three datasets are normalized for the better convergence and performance of optimization algorithm using Eq. (6).

$$X_{sc} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{6}$$

where X_{sc} is the scaled input data ranges from 0 to 1. X_{min} and X_{max} are the minimum and maximum values respectively in the input predictor matrix.

The training data is then transformed into a useful input predictor matrix. The training input predictor matrix and validation data are consigned to the LSTM training module. The training data is used to train the neural network model while the validation set helps in fine-tuning of the hyper-parameters and to remove the underfitting and overfitting issue.

Fig. 5 represents the proposed LSTM based deep learning architecture. The proposed LSTM model is based on feature extraction block, LSTM forecasting block and output layer. The input layer of proposed LSTM model behaves same as the input layer of ANN in feature extraction block. The neurons and weights of ANN layer facilitate to extract non-linear relationship between exogenous variables and actual load. Hence, the proposed LSTM model uses ANN input layer in feature extraction block to capture seasonal and temporal features with high degree of variability. Consider the ANN input layer consists of n -neurons in the proposed LSTM model.

Let X_j be the input of j^{th} neuron in the input layer and W_j be the weight associated with the j^{th} neuron, then the output of the j^{th} neuron a_j can be computed as $a_j = X_j W_j$. Generally, X_t represents the input vector of LSTM forecasting block and can be represented by Eq. (7).

$$X_t = [X_1 W_1 \ X_2 W_2 \ \dots \ X_j W_j \ \dots \ X_n W_n] \tag{7}$$

The hidden layer in proposed LSTM model constitutes of LSTM forecasting block which implements LSTM cascaded cells to capture long-term dependencies in electrical load pattern. The operation of LSTM unit has already been discussed in section IV. Hence, the new proposed LSTM model is a combination of ANN input layer (containing 11 neurons) and LSTM layer containing 8 cells. The last output layer consists of only one neuron.

The proposed model is trained on the optimal set of hyper parameters during tuning which provides the single-valued electrical load output, and the pre-trained LSTM model has

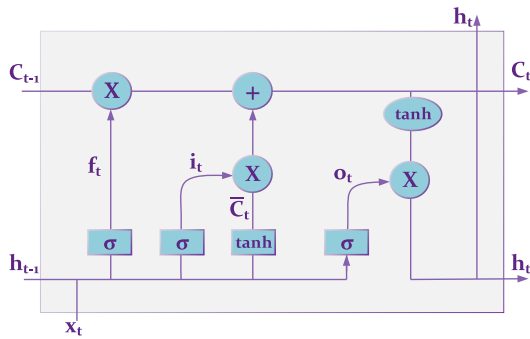


FIGURE 2. LSTM Cell.

been developed for the testing purposes using Adam optimizer. The LSTM forecasting module loads the proposed pre-trained LSTM model to forecast the electrical load of one hour resolution after inverse normalization. Then, the error metrics such as RMSE, MAPE and MPE are assessed. The learning rate is a hyper parameter that controls the dynamics of the model in response to the estimated error, each time the model weights are updated. Choosing the optimal learning rate is a challenging task as a too small value may result in a longer training time whereas a value too large may result in suboptimal set of weights. In this research study and the proposed LSTM model, learning rate is adjusted to 0.003 to learn all the features accurately and precisely. The superiority of the proposed LSTM model has been validated by comparing with the baseline ANN model, LSTM model 1, LSTM model 2, LSTM model 3 and CNN-LSTM using the same STLF framework.

A. DATASET

The proposed method is implemented on the hourly electric load data of the electric utilities of the city of Johor in Malaysia extending from 01 January 2009 to 31 December 2010 [42]. The electrical load data has 17520 values based on two years of hourly load data. The data has been segmented into training, validation, and test set with the ratio of 60%, 20% and 20% respectively as shown in Table 1. The temporal and weather data has been accumulated from www.worldweatheronline.com and www.timeanddate.com for years 2009 and 2010. The input features comprise of Temperature, Humidity, Weekdays, Holidays, Years, Months, Week, Days, Hour, Minutes and Seconds. We had two options i.e., either removing the minutes and seconds or simply making them zero. Both these options generate same results. We opted for the latter option as it increases the scalability i.e., will work well for higher resolutions of future datasets. Moreover, this study also focuses on the development of reliable deep learning algorithm to predict the electrical load pattern using hourly load data. Therefore, a sparse column vectors have been added in the dataset by making variables of minutes and seconds zero so that a robust deep learning model can be built considering the sparsity of the data. It can be observed that the feature extraction with sparse vectors

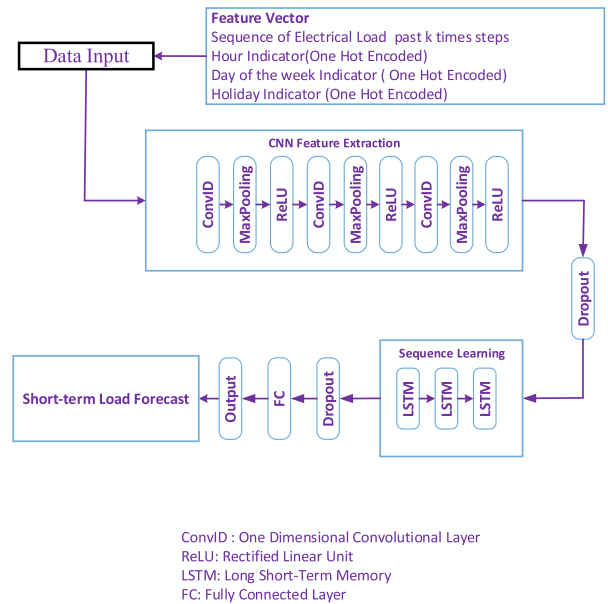


FIGURE 3. CNN-LSTM architecture [48].

generates a unique predictor matrix during training which improves the robustness and reliability of the proposed LSTM architecture. The proposed LSTM architecture can be used on other electrical load dataset for the STLF application due to the above discussed advantage. The complete historical electrical load data is then organized for a one-hour ahead prediction. A data vector load to the proposed LSTM model consists of eleven dimensions. The characteristics of eleven-dimensional input features can be described as follows:

- 1) Humidity: The humidity data is also gathered with a resolution of three hours and the total 2920 humidity readings can be found from the available dataset. The missing entries are then again linearly interpolated to make compatible with the electrical load data. Fig. 6 also represents the high randomness in the humidity.
- 2) Temperature: The temperature data is collected with a resolution of three hours and the total 2920 temperature readings can be found from the available dataset for two years. The missing entries are then linearly interpolated to make compatible with the electrical load data. Fig. 7 clearly depicts the highly variational and non-linear nature of the temperature.
- 3) Electrical Load: The electrical load data in Mega Watt is recorded with a resolution of one hour. Fig. 8 delineates the highly diversified and stochastic nature of electrical load data between the different months of two years.
- 4) Holidays/Date-time Features: The holiday index is encoded in a binary value 0 and 1. 0 represents the close days (holidays) and 1 represents the on-days (working days). The other date-time features such as minutes, hours seconds and weeks are extracted from the time stamp.

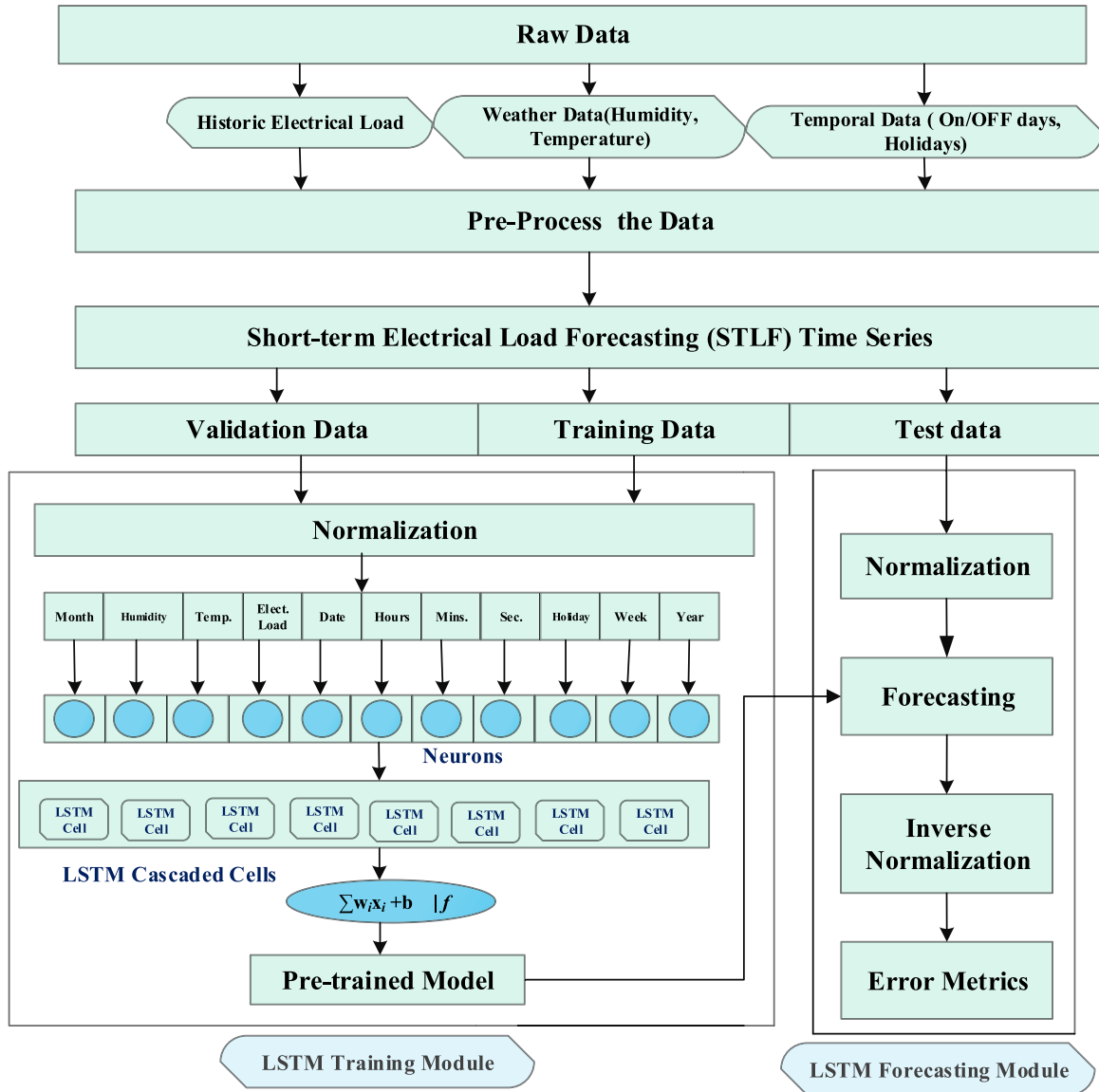


FIGURE 4. System model for research methodology.

V. SYSTEM CONFIGURATION WITH SIMULATION SETUP

The proposed method has been authenticated in Anaconda Navigator and Jupyter Notebook in which Python 3.7 is used. Simulation was accelerated by Nvidia graphics processing unit (GPU), GeForce Nvidia GTX1080Ti on a PC with an Intel core i7, 2.3GHz CPU, 16GB RAM and 1 TB hard disk running on a Windows operating system.

For the STLF problem, the proposed LSTM, ANN, LSTM model 1, LSTM model 2, LSTM model 3 and CNN-LSTM have been simulated in a popular open-source deep learning platform named Tensorflow with a version of 2.2.0, which supports Python APIs and CUDA extension [49] Keras has also been integrated with Tensorflow [50] because it also delivers additional high-level APIs, user friendly modular

interface, and important machine learning packages such as scikit-learn for Python [51] The other relevant libraries such as Dense, Activation, Optimizers and LSTM have also been imported from Keras to implement hidden layers, activation function such as Sigmoid, implement Adam optimizer and LSTM layer module respectively. Early stopping technique is used with a patience technique to stop the training when the deep neural network (DNN) model suffers from serious overfitting. The small patience value such as 30 can stop the training process earlier [52]. However, the small patience value may fail to clearly identify the impact of overfitting in training and validation loss curve for the fair comparison between different DL architectures. Therefore, the patience is set to 50 to perceive the impact of overfitting in training

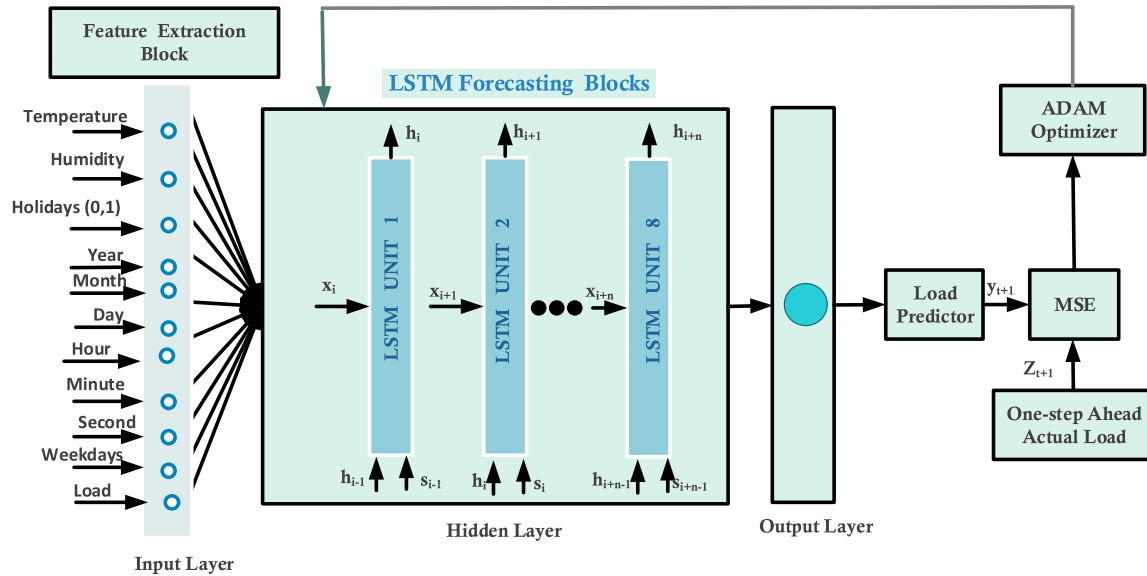


FIGURE 5. Proposed LSTM-based deep learning architecture.

TABLE 1. Segregation of data into Training, Validation and Test set.

Segregation Type	Period	Resolution	Number of input parameters	Number of samples
Training Set	01/01/2009 1:00 a.m. to 15/03/2010 1:00 a.m.	1 hr	11	10512
Validation Set	15/03/2010 2:00 a.m. to 08/08/2010 1:00 a.m.	1 hr	11	3504
Test Set	08/08/2010 2:00 a.m. to 01/01/2011 0:00 a.m.	1 hr	11	3504

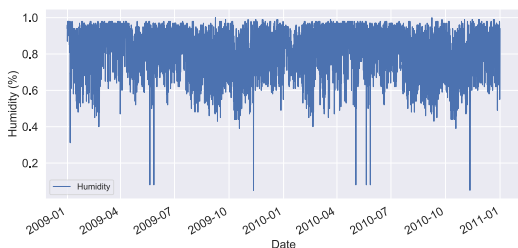


FIGURE 6. Percentage humidity for year 2009-2010.

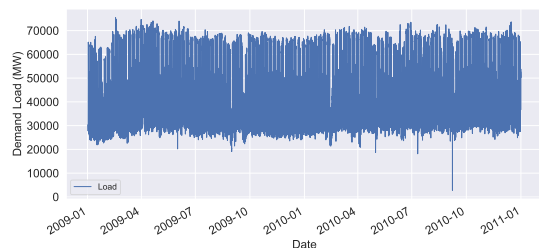


FIGURE 8. Load (in MW) for year 2009-2010.

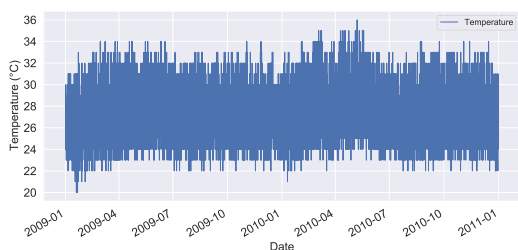


FIGURE 7. Temperature (degree Celsius) for year 2009-2010.

and validation loss curve of ANN, the proposed LSTM model, LSTM model 1, LSTM model 2, LSTM model 3 and CNN-LSTM.

The hyper-parameters used for the fine tuning of STLF algorithms are listed in Table 2. The number of training

epochs and Batch size are also the key hyper-parameters that have significant impact on the generalization ability and forecasting accuracy of DNN. The higher batch size trains the data based on large number of samples within the training dataset for one epoch. Considering the higher number of samples during training in a single epoch, the higher batch size can be proved useful to extract the more generalized pattern. For the STLF problem, Batch size is set to 500 due to the above-mentioned reason. The number of training epochs are adjusted to 1000 to capture the input-output relationship between features and electrical load accurately. To ensure fair comparison, we chose the same hyperparameters of LSTM. However, in our proposed approach, sigmoid activation function is used but CNN-LSTM uses ReLU activation function. Unlike others, the proposed approach is a hybrid model of

TABLE 2. Hyperparameter settings of various Deep Learning architectures.

Model	Hyper-parameter
ANN, LSTM Model 1,2,3 and Proposed LSTM	Epoch: 1000 Monitor: MSE Learning Rate: 0.003 Batch size: 500 Optimizer: Adam Activation Function: Sigmoid Loss function: MSE Input Vector Size: 1×11 Patience : 50
CNN-LSTM	Epoch: 1000 Monitor: MSE Learning Rate: 0.003 Batch size: 500 Optimizer: Adam Activation Function: ReLU Loss function: MSE Input Vector Size: 1×11 Patience : 50

ANN and LSTM. The first layer of proposed model is ANN with simple neurons that are used as an input to LSTM layer containing 8 cells. The usage of sigmoid function in the proposed approach reduces the problem of vanishing gradient. ReLU, used in CNN-LSTM, does have some limitations especially the case where large weight updates can result in the negative value of the summed input to the activation function, regardless of the input to the network. This implies that any node with such a problem will always output 0 activation value.

The learning rate is also a pre-eminent hyper-parameter that controls the dynamics of the DNN model in response to the estimated error, each time the DNN model weights are updated. The error function based on highly diversified and dimensional features accomplish the fine-tuning of optimal learning rate very demanding. A low learning rate may result in a longer training time due to slower convergence of optimal set of neurons weights. Whereas a high learning rate fails to converge to a minimum value of the error function and may result in sub-optimal set of neurons weights. Contemplating the above stated concepts, the learning rate was initially set to 0.9 for the STLF error function as described in Eq. (2) and the response of training and validation loss curve was observed using gradient-check method. The learning rate was then fine-tuned by decreasing previous learning rate values during different observations using gradient-check method. Finally, the DNN model provided the best training-validation loss curve and optimal set of neurons weights at a learning rate of 0.003. Adam optimizer has been used to minimize the loss during training. Adam is a replacement optimization algorithm for stochastic gradient descent for training deep learning models as already described in the performance of neural network section. Adam combines the characteristics of the AdaGrad and RMSProp algorithms to provide an optimization algorithm that can handle sparse gradients on noisy problems.

The neural structures of ANN, LSTM model 1, LSTM model 2, LSTM model 3, CNN-LSTM and the proposed LSTM are described below:

A. ANN CONFIGURATION

The neurons always play a crucial role in the fine-tuning of the neural network model. The input layer of fully connected ANN model has 11 neurons because the input vector to the ANN model is eleven-dimensional. The ANN model comprises of only 1 hidden layer to avoid overfitting. According to the criteria of selecting hidden neurons in a hidden layer as discussed in Section III the neurons are initially set to 10 to avoid underfitting. The ANN generates the high prediction accuracy when the number of hidden layer neurons are adjusted to 20. More than 12 neurons in the hidden layer depicts serious overfitting in the experiments during hyper-parameter tuning. Therefore, the neurons in fully connected hidden layer of ANN are optimally set to 8 so that ANN can be able to capture the electrical load pattern. The output layer of ANN constitutes only 1 neuron to output a single valued one-hour ahead electrical load forecasted value.

B. LSTM MODEL 1, Model 2 AND MODEL 3 CONFIGURATION

LSTM model 1, LSTM model 2 and LSTM model 3 have 11 neurons in the input layer. All variants have 8 LSTM cells in the hidden layer and 1 neuron in the output layer. However, LSTM model 1, LSTM model 2 and LSTM model 3 are built on 24, 48 and 72 lagged values of electrical load respectively.

C. CNN-LSTM CONFIGURATION

The first 1-dimensional convolutional layer in CNN-LSTM has 128 kernels. The large number of kernels are used in the CNN-LSTM model to capture all patterns of an electrical load which are governed by the meteorological and temporal attributes present in the electrical load dataset. The kernel size is also adjusted to 1×1 so that CNN-LSTM extracts local trends of every individual electrical load sample. The 1×1 filter also uses the advantage of removing flattening layer in CNN-LSTM model which makes the architecture simpler than any other CNN-LSTM model. The output of 1-dimensional maxpooling layer with pooling size 1 is applied after the 1-dimensional convolutional layer. After that, the LSTM layer is employed as a fully connected layer with 8 LSTM cells followed by the output layer of 1 neuron. In this study, the CNN-LSTM model is intended to implement with 1 convolutional layer and smaller number of LSTM cells to remove overfitting and to make the CNN-LSTM model uncomplicated.

D. THE PROPOSED LSTM CONFIGURATION

The input layer of fully connected LSTM model has also 11 neurons because the eleven-dimensional input vector has been loaded to the proposed LSTM model. The hidden layer of proposed LSTM model has been set to 1 to avoid overfitting. The number of hidden layer neurons are fixed

to 8 as shown in Fig. 5 to escape from overfitting and then fine-tuned the other hyper-parameters such as learning rate to validate the prediction accuracy of the proposed LSTM. Similarly, the output layer of the presented LSTM model consists of only 1 neuron to squash a single valued one-hour ahead electrical load forecast value

E. EVALUATION METRICS AND ERROR FUNCTION

After getting our predicted load of LSTM and ANN APE, MAPE and RMSE, introduced in Eq. (8) to Eq. (11) which are the most important features to use when comparing the outcomes of the two models. These error metrics are defined as follows:

$$MPE = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i) \quad (8)$$

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

$$APE = \left| \frac{(\hat{y}_i - y_i)}{y_i} \right| \times 100\% \quad (10)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N (APE_i) \quad (11)$$

where N is the number of samples, \hat{y}_i is forecasting value, and y_i is actual load value. MPE, RMSE APE and MAPE provide short-term performance of these models. Smaller +ve values indicate the closeness of match between actual value y_i and estimated value \hat{y}_i .

VI. RESULTS AND DISCUSSION

The proposed LSTM model is compared with ANN, LSTM model 1, LSTM model 2, LSTM model 3 and CNN-LSTM to investigate the performance of the aforesaid models for electrical STLF. Training and validation loss curves, time complexity, evaluation indices such as MPE, RMSE, MAPE, APEs and Diebold-Mariano test are used to prove the validity of proposed LSTM model for this purpose as described in next sections:

A. ERROR PERFORMANCE

To compare the performance of our proposed LSTM architecture with ANN, LSTM model 1, LSTM model 2, LSTM model 3 and CNN-LSTM, the error metrics have been calculated as mentioned in Table 3. From these metrics, the mean percentage error (MPE) of our proposed architecture is 0.05, which is comparatively smaller than the MPE of an ANN architecture which is 1.02. Hence, the proposed LSTM model shows superiority over ANN in terms of STLF accuracy. Similarly, the MPE of LSTM model 1, LSTM model 2 and LSTM model 3 is -1.08 , -0.60 and -0.30 respectively as shown in Fig. 9. The negative values of MPE indicate that the conventional three LSTM models forecast the higher electrical load values than the actual electrical load values for most of the test data. Similarly, the MPE of CNN-LSTM is 0.40, which is higher than the proposed LSTM model. Hence, the proposed LSTM model shows improved MPE than the other under-considered models for the given STLF study.

Similarly, the proposed LSTM model shows significant improvement for other evaluation metrics such as RMSE and MAPE over an ANN. The MAPE and RMSE of the proposed LSTM model are 0.64 and 451.12 respectively as shown in Fig. 9 and Fig. 10. While the MAPE and RMSE of conventional ANN are 1.18 and 728.49 respectively, which are worse than the proposed LSTM. Hence, the proposed LSTM offers high STLF accuracy than ANN. It can be seen from the Table 3 that the MAPE of LSTM model 1, LSTM model 2, LSTM model 3 and CNN-LSTM are 1.56, 0.96, 0.86 and 2.78 respectively which are again worse than the proposed LSTM. Since the proposed LSTM model captures the diversified features more precisely than the ANN and other models by eliminating vanishing gradient and overfitting problem, the one-hour electrical forecasted load maps the actual electrical load accurately. Therefore, the proposed LSTM displays superior performance in terms of evaluation indices in our defined STLF problem.

B. TRAINING LOSS AND VALIDATION LOSS

A complete forward and backward propagation of the entire training dataset in a neural network once to learn the algorithm is termed as an epoch. Fig. 11 illustrates the training and validation loss curve of different neural network models. All neural models have 8 neurons in the hidden layer. The validation loss in proposed LSTM decreases rapidly as compared to ANN during first 80 to 100 epochs. Hence, the proposed LSTM model seeks to optimize the model parameters such as weights quickly than ANN as depicted in Fig. 11 (a) and Fig. 11(b). The conventional ANN learns slowly than the proposed LSTM as mentioned above due to the complex hyper-parameter tuning process in ANN. It can also be observed from the loss curve that the proposed LSTM and ANN tries to converge after 400 epochs.

Furthermore, overfitting and underfitting phenomena can be examined by observing the loss curve of error rate with respect to number of epoch. Overfitting represents a phenomenon in which distance between training and validation loss curve increases as the number of epochs increases. The gap between training and validation loss curve remains smaller and almost same during 1000 epochs which represents neither underfitting nor overfitting in the proposed LSTM model. Conversely, the gap between training and validation loss curve becomes wider in ANN which represents overfitting in the conventional ANN model. ANN experiences consecutive overfitting in the presence of high dimensional features due to sufficient neurons in the hidden layer to capture all the input-output relationship between electrical load and features which eventually decreases the generalization capability and hence produces overfitting. However, the small number of neurons in ANN generates high forecasting errors due to under-fitting. It can be realized from the extensive experiments that if the neurons of the hidden layer are set to 8, then the conventional ANN encounters from overfitting. Conversely, if the hidden layer neurons are set to 7 in ANN, then the ANN fails to find the optimal set

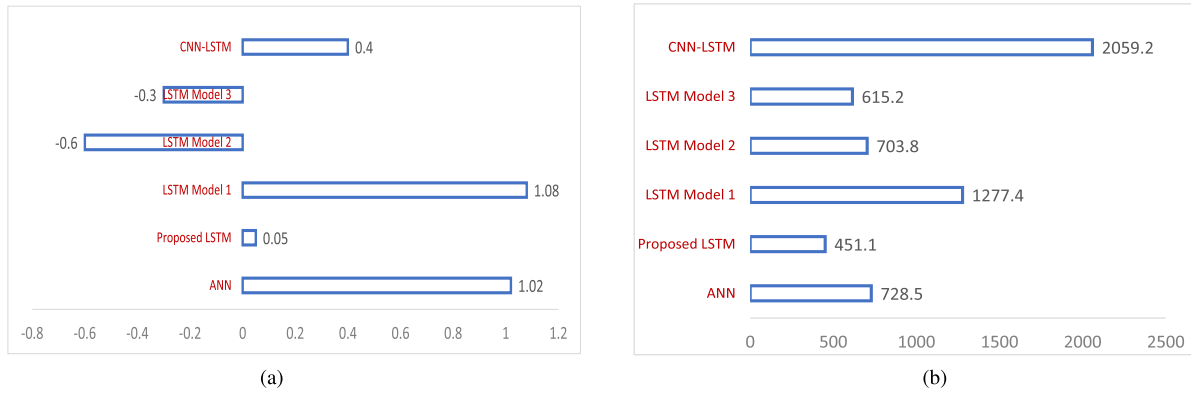


FIGURE 9. (a) Comparison of Mean Percentage Error (MPE) of Deep Learning architectures and (b) Comparison of Root Mean Square Error (RMSE) of Deep Learning architectures.

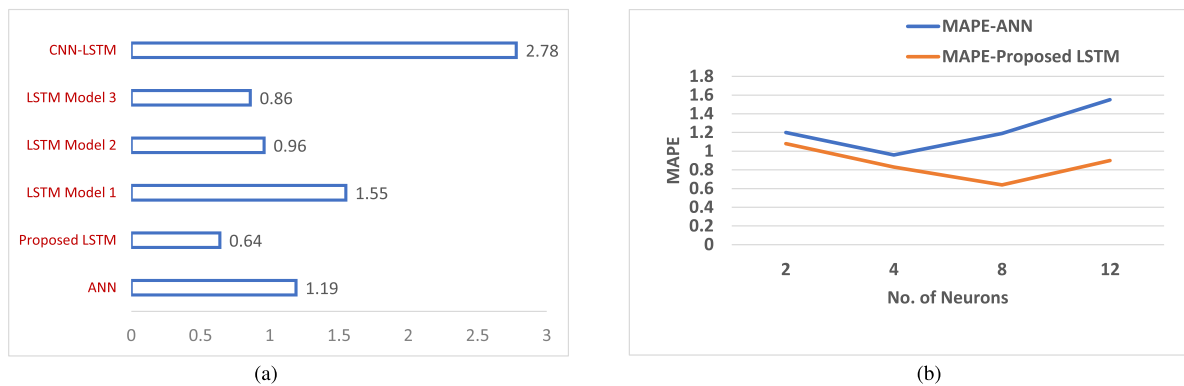


FIGURE 10. (a) MAPE for various Deep Learning models (b) MAPE variation with number of neurons.

of weights. Consequently, ANN generates higher forecasting errors with 7 neurons in the hidden layer as demonstrated in Fig. 10(b). The proposed LSTM removes all the above limitations and provides higher forecasting accuracy with 8 neurons in the hidden layer as shown in Fig. 10.

Furthermore, ANN validation loss curve experiences volatile behavior after 100 epochs which again represents the failure of optimizing the model’s parameter accurately. However, the proposed LSTM validation loss curve delineates smooth learning. Hence, the performance of ANN degrades consistently with the increase in number of epochs in STLF scenario.

LSTM model 1, model 2 and model 3 get stuck in local optima and encounter with saddle points as shown in the loss curves of LSTM model 1, model 2 and model 3 in Fig. 11 (c), Fig. 11 (d) and Fig. 11(e). The lagged values in the above mentioned LSTM models restrict the smooth learning of the algorithm. Therefore, the optimization phenomena in the lagged LSTM models remains unsuccessful to converge in the under-considered STLF problem. In Hybrid CNN-LSTM, the convolutional layer is introduced before the LSTM layer to capture the local trends between electrical load and the features. The Hybrid CNN-LSTM model learns the input-output relationship between various features and

electrical load impressively on the training dataset but lacks the generalization capability on the test dataset due to the addition of convolutional layer, use of enormous number of filters, large number of neurons in LSTM layer and complexity of the model as shown in loss curve of Hybrid CNN-LSTM. Hence, it may be concluded that the proposed LSTM model provides highly accurate forecasting algorithm by mitigating hyper-parameters tuning, overfitting and underfitting issues.

C. TIMING COMPLEXITY OF STLF MODELS

The proposed LSTM requires 247.9 seconds to train with 1000 epochs which is worse than the ANN. The ANN consumes 74.73 seconds to train the electrical load data as shown in shown in Table 3. Similarly, the test time of the proposed LSTM model is 0.81 seconds which is again slightly low-grade than ANN, which requires 0.31 seconds to forecast the one-hour ahead electrical load on the same dataset. The proposed LSTM back propagation algorithm takes much more complex computations to avoid vanishing and exploding gradient descent problem due to three gated mechanisms such as input gate, output gate and forget gate. Therefore, the training and test time consumed by the LSTM is higher than the ANN. Conversely, the ANN lacks the gated mechanisms

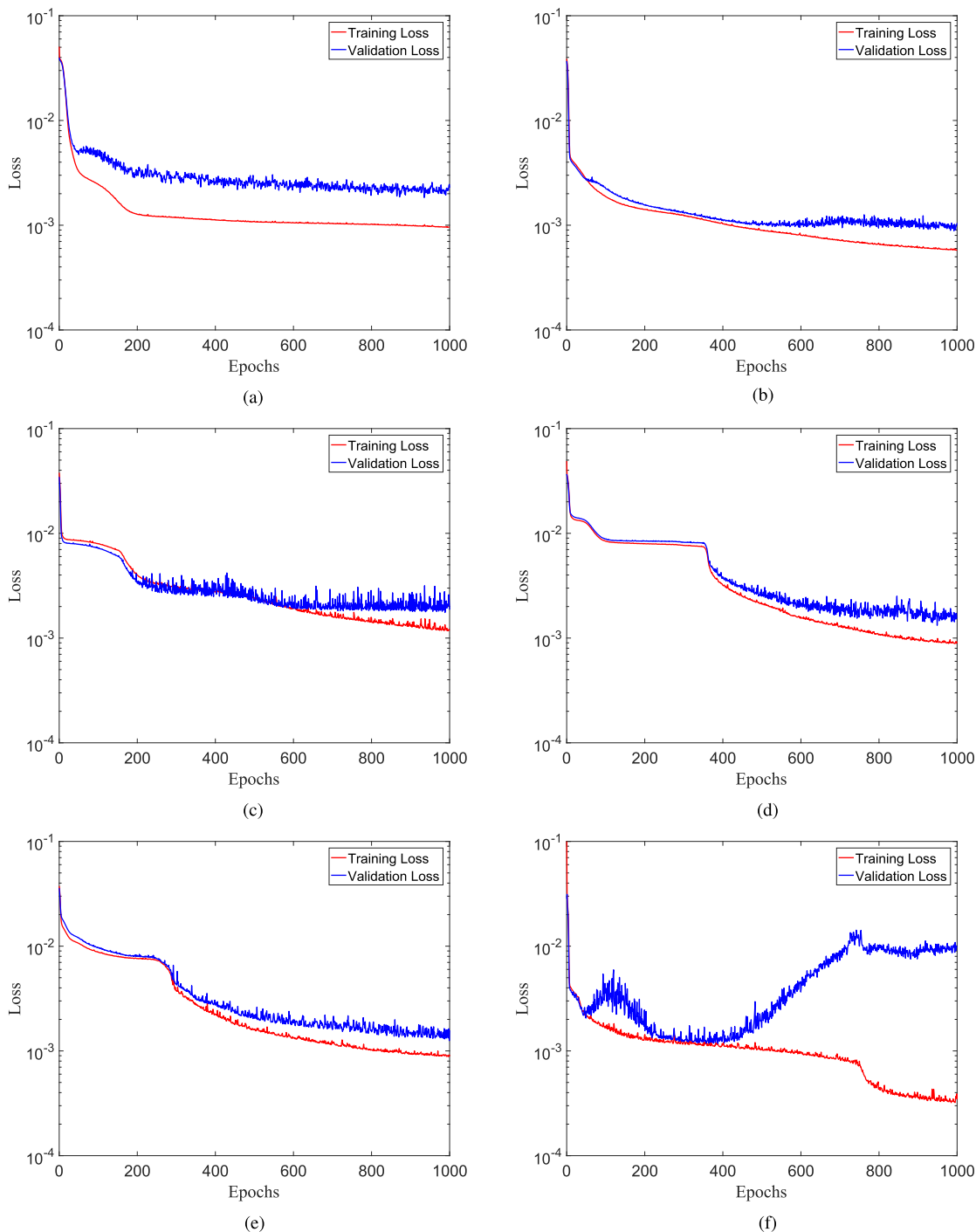


FIGURE 11. Training and validation loss curves for (a) ANN (b) Proposed LSTM (c) LSTM Model 1 (d) LSTM Model 2 (e) LSTM Model 3 (f) CNN-LSTM.

and therefore the training time of ANN is less than the proposed LSTM model. However, ANN fails to extract stochastic nature of electrical load in the presence of highly variational temporal features due to the lack of gated mechanisms and the overfitting issue which consequences in the higher MAPE value than the proposed LSTM as shown in Fig. 10.

Moreover, the presence of high computing machines makes the efficient algorithm such as the proposed LSTM implementable and give the privilege to the researchers to concentrate on the trade-off between highly accurate prediction and the computational time. The proposed LSTM achieves higher forecasting accuracy and thus decreases MAPE to a satisfactory extent than the conventional

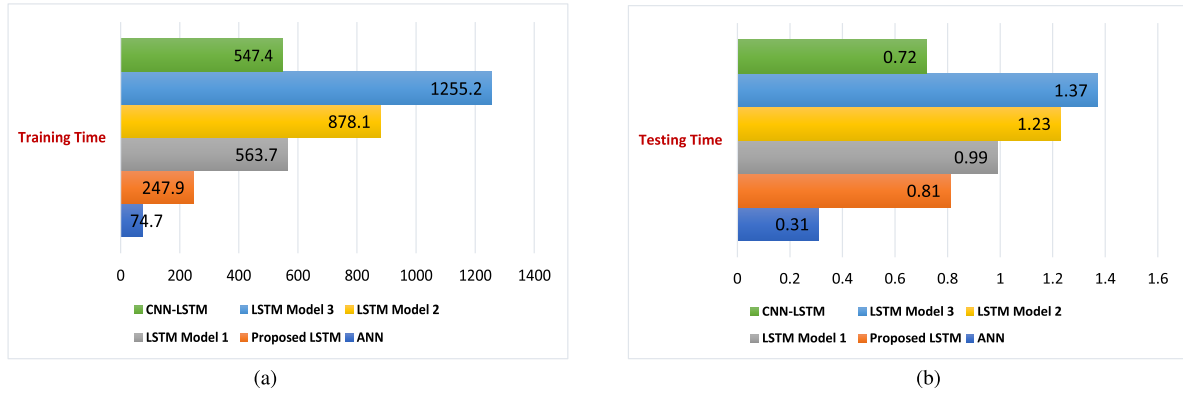


FIGURE 12. (a) Training Time (sec) for Deep Learning architecture and (b) Testing Time (sec) for Deep Learning architectures.

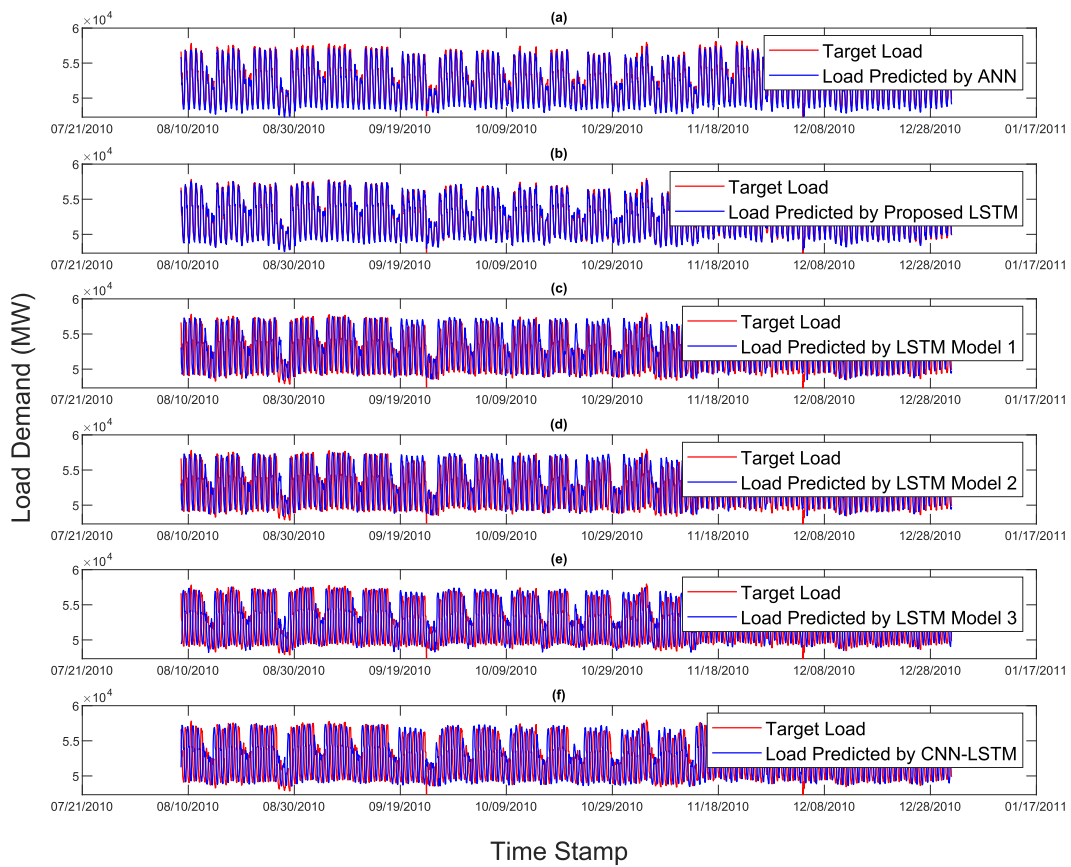


FIGURE 13. Full load prediction for August 08, 2010 to December 31st, 2010 using various Deep Learning architectures.

TABLE 3. Overall performance results.

Model	RMSE	MAPE	MPE	Training Time (sec)	Testing Time (sec)
ANN	728.5	1.19	1.02	74.7	0.31
Proposed LSTM	451.1	0.64	0.05	247.9	0.81
LSTM Model 1	1277.4	1.55	1.08	563.7	0.99
LSTM Model 2	703.8	0.96	-0.6	878.1	1.23
LSTM Model 3	615.2	0.86	-0.3	1255.2	1.37
CNN-LSTM	2059.2	2.78	0.4	547.4	0.72

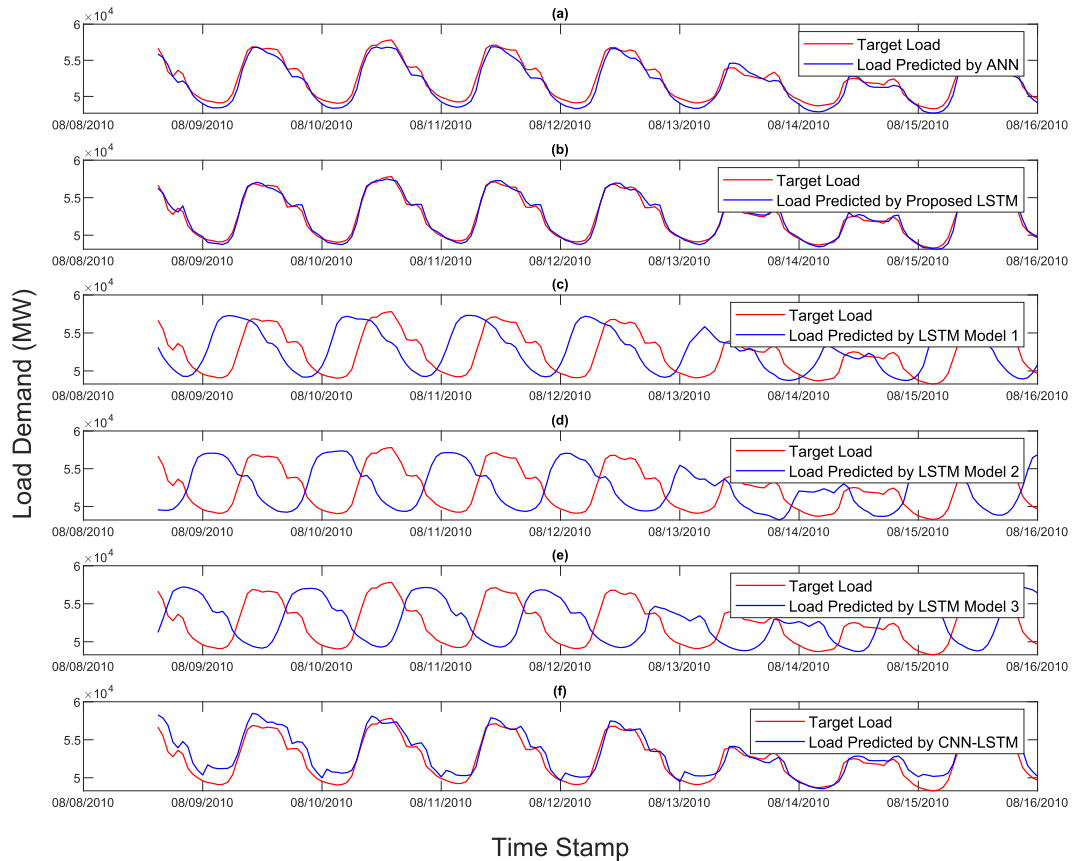


FIGURE 14. Weekly load prediction using various Deep Learning architectures.

ANN at the cost of increased time complexity as shown in Fig. 12.

LSTM model 1, LSTM model 2 and LSTM model 3 consume 563.7 sec, 878.1 sec and 1255.23 sec respectively to train with the same 1000 epochs. It is observed that the lagged values diversified the dimensions of the input matrix which enhance the training time of LSTM model 1, LSTM model 2 and LSTM model 3 to a larger extent. Similarly, the test time of LSTM model 1, LSTM model 2 and LSTM model 3 is 0.99 sec, 1.23 sec and 1.37 sec respectively which is worse than the proposed LSTM. Eventually, the LSTM models based on fixed-time lagged values are impractical to implement for one-hour ahead electrical STLF. Similarly, CNN-LSTM model is fully trained on an under-considered electrical load data after 547.42 sec. The computational time complexity of CNN-LSTM is almost double than the proposed LSTM model because each constituent in hybrid model requires sufficient time to accomplish the complete operation. Moreover, the convolution process in a convolutional layer expends enormous time which raises the time complexity of the CNN-LSTM model.

From the above discussion, it may be concluded that the proposed LSTM model captures the short and long-term dependencies efficiently with an improved time complexity as compared to LSTM model 1, LSTM model 2, LSTM

model 3 and CNN-LSTM. ANN also consumes less training and test time than LSTM model 1, LSTM model 2, LSTM model 3 and CNN-LSTM which depicts that ANN is better than the above lagged load based LSTM models and CNN-LSTM in terms of computational complexity. Furthermore, it may be deduced that the proposed LSTM achieves significantly better STLF performance than ANN at the cost of slight increase in the test time.

D. LOAD FORECASTING PERFORMANCE

The closest match between predicted and target load trends are presented in Fig. 13 to show the STLF performance for 146 days from 08 August 2010 to 01 January 2011 between proposed LSTM model and five state-of-art models such as conventional ANN, LSTM model 1, 2, 3 and CNN-LSTM. It can be realized from Fig. 13 (a-b) that the proposed LSTM model yields the strongest match between the predicted and target loads, especially during peak load and valley load than conventional ANN. The other state-of-art models like LSTM model 1, 2, 3 and CNN-LSTM observe difficulties in capturing the trends of peak loads and hence are not suitable for this electrical load dataset as shown in Fig. 13 (c)-(f).

The performance analysis has been extended further to identify the suitable DL model, which will capture the abrupt

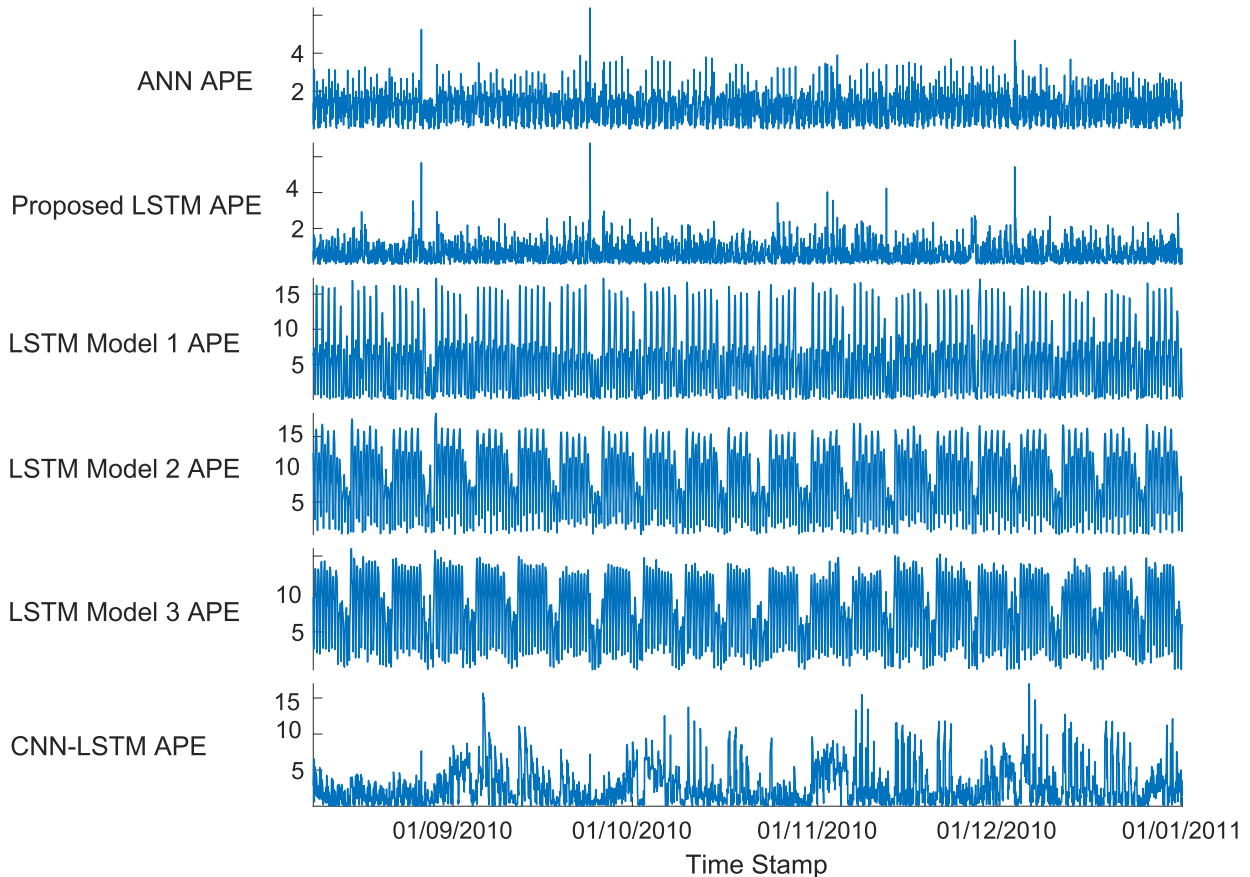


FIGURE 15. Performance comparison of APEs for various architectures and models.

variations in electrical load pattern accurately. For this purpose and better visualization, STLF performance of one week from 09 August 2010 to 16 August 2010 has been considered. Fig. 14 (a)-(f) suggests that the proposed LSTM model captures the sharp variations, randomness, and notches in an electrical load trend comparatively better than the other state-of-art deep learning models. The forecasting performance based on average MAPE of above-defined one week of the proposed LSTM and other state-of-art deep learning methodologies has also been considered. The proposed LSTM attained average MAPE in the best case with the mean value 0.64 and the variance value 0.02, better than the 1.23 ± 0.05 received by the conventional ANN. Similarly, the average MAPEs received by LSTM model 1, 2, 3 and CNN-LSTM are 5.7, 8.4, 8.08 and 1.82 respectively with the corresponding variance values 1.86, 3.53, 1.97 and 0.83. The above improved average MAPE of proposed LSTM implies that forecasting curve can precisely repeat the actual load curve as compared to ANN, LSTM model 1, 2, 3 and CNN-LSTM due to accurate mapping of input-output relationship between electrical load and highly variational temporal features in proposed LSTM model.

Additionally, the absolute APEs are also considered as the useful error metrics against actual electrical load samples

in the test data to select the best STLF forecasting engine. Fig. 15 illustrates that the proposed LSTM model propagates less APEs as compared to ANN, LSTM model 1, 2, 3 and CNN-LSTM. As discussed earlier that the proposed LSTM model finds the closest match between actual and forecasted electrical load curve due to extraction of local trends convincingly in electrical load pattern therefore the proposed LSTM model generates less APEs. As the APEs signify forecasting errors against each actual load sample, therefore APEs also provide the intuition about the stability of the STLF engine. The proposed LSTM generates less forecasting errors for all samples of the test dataset, which reveals the stability of presented model than the other deep learning models.

E. STATISTICAL ANALYSIS

From the above discussion, it can be deduced that the proposed feature selection-based LSTM model attains better forecasting performance than the conventional ANN, LSTM model 1, 2, 3 and CNN-LSTM. However, significant improvement in the performance of electrical load forecasting must be required by the proposed LSTM model for the recommendation of a real-time STLF DL algorithm in the power system operations of under-developed countries. For this purpose, the Diebold-Mariano test has been performed

TABLE 4. DM test results.

DM Test	S_1	Results
Test between Proposed LSTM and ANN	32.73	Accept H_1 , decline H_0
Test between Proposed LSTM and LSTM model 1	48.12	Accept H_1 , decline H_0
Test between Proposed LSTM and LSTM model 2	72.18	Accept H_1 , decline H_0
Test between Proposed LSTM and LSTM model 3	69.84	Accept H_1 , decline H_0
Test between Proposed LSTM and CNN-LSTM	34.14	Accept H_1 , decline H_0

to determine the remarkable improvement by the proposed LSTM model over the competing DL models.

Let, H_0 represents the null hypothesis in Diebold-Mariano test, which will be accepted only when all DL models including the proposed LSTM model, have approximately the same STLF performance. The other two alternative hypotheses are H_1 and H_2 . H_1 will be accepted when the proposed LSTM model delineates significant improvement in STLF accuracy than the other DL models, and H_2 will be accepted when the other DL models show significant improvement than the proposed LSTM model. Let, S_1 be the statistic which is used to test the above hypothesis such as H_0 , H_1 and H_2 , and the confidence level be 95% [53]. H_1 will be accepted and H_0 will be declined when S_1 is greater than 1.96. H_2 will be accepted and H_0 will be declined when S_1 is smaller than -1.96 . When S_1 is within -1.96 to 1.96 , H_0 will be accepted and there will be no significant difference in forecasting accuracy between the proposed LSTM and all other DL models. The proposed LSTM model shows remarkable improvement in forecasting accuracy than the conventional ANN, LSTM model 1, 2, 3 and CNN-LSTM, as shown from DM Test results given in Table 4.

VII. CONCLUSION

A new temporal feature selection-based LSTM model was presented for short-term load forecasting. The proposed LSTM model was composed of ANN input layer and LSTM hidden layer. ANN input layer operated as a feature extraction module and LSTM hidden layer worked as a forecasting module. The proposed LSTM model was also based on optimally selecting the number of hidden layers, neurons, and the tuning of the hyper-parameters to avoid overfitting. The proposed framework achieved the better forecasting accuracy using the following design concepts of neural network in the proposed LSTM model:

- 1) The proposed approach used artificial neural network layer that extracted the desired features while maintaining the same computational complexity as convolutional layer of CNN.
- 2) Many hidden layers can cause overfitting in deep learning models. So, only one LSTM hidden layer was used in the proposed LSTM model to avoid overfitting.
- 3) An adequate number of neurons and LSTM cells in the hidden layer may also cause overfitting. The proposed LSTM model used only eight LSTM cascaded cells in the hidden layer to avoid overfitting. Similarly, LSTM cells less than six could cause underfitting because

model would fail to record the electrical load pattern with respect to all eleven input features.

- 4) A large batch size was used to enhance the better generalization capability with suitable number of epochs as discussed above.
- 5) A learning rate is optimally selected to 0.003 with the Adam optimizer to maintain optimal learning step size to extract all the local trends in the electrical load data in the presence of highly dimensional features.

The evaluation metrics of the proposed LSTM model were compared with ANN, LSTM model 1, LSTM model 2, LSTM model 3 and CNN-LSTM on a Malaysian dataset which showed superiority over all the other deep learning models in terms of root-mean square error, mean-percentage error and mean-absolute percentage error. The experimental studies demonstrated that the proposed LSTM provided the best forecasting performance by matching the forecast and actual load especially during peak and valley load duration. The Diebold-Mariano test also suggested that the proposed LSTM model significantly improved the STLF accuracy. Based on the above study, an implementation of the proposed LSTM model was recommended in the power system operations to the electric utilities of under-developed countries for the STLF problem.

One of the main challenges faced in STLF is to predict the electrical loads of special days such as special holidays due to the following reasons:

- (1) The electrical load pattern of special holidays and non-working days is unique, the aggressive deep learning model is required, which captures the variation of local trends in electrical load pattern. Moreover, the designed deep learning model should base on three levels of deep learning competence such as modeling, tuning and applications.
- (2) Holidays occur infrequently, even irregularly, so the historical data are insufficient. Therefore, the training sets are also very low for these days. This problem can be removed by using data augmentation and generative adversarial networks (GANs) to enhance the data and training sets.
- (3) Since enormous factors effect the consumption of electrical load of holidays in different ways, it is difficult to determine the impact of individual factors. This issue can be resolved by determining the most significant factors present in the electrical load dataset, which make positive impact on the electrical load forecasting. Algorithms based on similar day characteristics and clustering can be the suitable candidates to resolve the above mentioned issue. These issues will be addressed in our future work by designing and

implementing the hybrid deep learning model based on three levels of competence and GANs to forecast one-hour and one-day ahead electrical load of special days.

REFERENCES

- [1] G. Dudek, "Short-term load forecasting using neural networks with pattern similarity-based error weights," *Energies*, vol. 14, no. 11, p. 3224, May 2021.
- [2] G. A. Darbellay and M. Slama, "Forecasting the short-term demand for electricity," *Int. J. Forecasting*, vol. 16, no. 1, pp. 71–83, Jan. 2000.
- [3] I. Shah, H. Iftikhar, S. Ali, and D. Wang, "Short-term electricity demand forecasting using components estimation technique," *Energies*, vol. 12, no. 13, p. 2532, Jul. 2019.
- [4] T. K. Chau, S. S. Yu, T. Fernando, H. H.-C. Iu, and M. Small, "A load-forecasting-based adaptive parameter optimization strategy of STATCOM using ANNs for enhancement of LFOD in power systems," *IEEE Trans. Ind. Informat.*, vol. 14, no. 6, pp. 2463–2472, Jun. 2018.
- [5] A. Ahmad, N. Javaid, M. Guizani, N. Alrajeh, and Z. A. Khan, "An accurate and fast converging short-term load forecasting model for industrial applications in a smart grid," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2587–2596, Oct. 2017.
- [6] K. S. Khan, S. M. Ali, Z. Ullah, I. Sami, B. Khan, C. A. Mehmood, U. Farid, J. Saleem, and J.-S. Ro, "Statistical energy information and analysis of Pakistan economic corridor based on strengths, availabilities, and future roadmap," *IEEE Access*, vol. 8, pp. 169701–169739, 2020.
- [7] A. J. Wood, B. F. Wollenberg, and G. B. Sheblé, "Short-term demand forecasting," in *Power Generation Operation and Control*. Hoboken, NJ, USA: Wiley, Nov. 2014, pp. 566–619.
- [8] *Annual Energy Outlook 2021*. Accessed: Feb. 3, 2021. [Online]. Available: <https://www.eia.gov/outlooks/aeo/electricity/sub-topic-01.php/>
- [9] A. A. Mir, M. Alghassab, K. Ullah, Z. A. Khan, Y. Lu, and M. Imran, "A review of electricity demand forecasting in low and middle income countries: The demand determinants and horizons," *Sustainability*, vol. 12, no. 15, p. 5931, Jul. 2020.
- [10] D. Saleem, T. U. R. Khan, and S. Abbas, "Electrical load forecasting techniques employed in power sector of Pakistan," in *Proc. Int. Conf. Eng. Emerg. Technol. (ICEET)*, Feb. 2020, pp. 1–7.
- [11] A. R. Khan, S. Razzaq, T. Alquthami, M. R. Moghal, A. Amin, and A. Mahmood, "Day ahead load forecasting for IESCO using artificial neural network and bagged regression tree," in *Proc. 1st Int. Conf. Power, Energy Smart Grid (ICPESG)*, Apr. 2018, pp. 1–6.
- [12] A. Y. Kharal, A. Mahmood, and K. Ullah, "Load forecasting of an educational institution using machine learning: The case of NUST, Islamabad," *Pakistan J. Sci.*, vol. 71, pp. 252–257, Dec. 2019.
- [13] M. Jawad, A. Rafique, I. Khosa, I. Ghous, J. Akhtar, and S. M. Ali, "Improving disturbance storm time index prediction using linear and nonlinear parametric models: A comprehensive analysis," *IEEE Trans. Plasma Sci.*, vol. 47, no. 2, pp. 1429–1444, Feb. 2019.
- [14] M. Jawad, S. M. Ali, B. Khan, C. A. Mehmood, U. Farid, Z. Ullah, S. Usman, A. Fayyaz, J. Jadoon, N. Tareen, A. Basit, M. A. Rustam, and I. Sami, "Genetic algorithm-based non-linear auto-regressive with exogenous inputs neural network short-term and medium-term uncertainty modelling and prediction for electrical load and wind speed," *J. Eng.*, vol. 2018, no. 8, pp. 721–729, Aug. 2018.
- [15] V. Ş. Ediger and S. Akar, "ARIMA forecasting of primary energy demand by fuel in Turkey," *Energy Policy*, vol. 35, no. 3, pp. 1701–1708, Mar. 2007.
- [16] H. Musbah and M. El-Hawary, "SARIMA model forecasting of short-term electrical load data augmented by fast Fourier transform seasonality detection," in *Proc. IEEE Can. Conf. Electr. Comput. Eng. (CCECE)*, May 2019, pp. 1–4.
- [17] S. N. Dodamani, V. J. Shetty, and R. B. Magadam, "Short term load forecast based on time series analysis: A case study," in *Proc. Int. Conf. Technol. Advancements Power Energy (TAP Energy)*, Jun. 2015, pp. 299–303.
- [18] B. Yildiz, J. I. Bilbao, and A. B. Sproul, "A review and analysis of regression and machine learning models on commercial building electricity load forecasting," *Renew. Sustain. Energy Rev.*, vol. 73, pp. 1104–1122, Jun. 2017.
- [19] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, "A review on time series forecasting techniques for building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 74, pp. 902–924, Jul. 2017.
- [20] U. B. Tayab, A. Zia, F. Yang, J. Lu, and M. Kashif, "Short-term load forecasting for microgrid energy management system using hybrid HHO-FNN model with best-basis stationary wavelet packet transform," *Energy*, vol. 203, Jul. 2020, Art. no. 117857.
- [21] F. Jun-long, X. Yu, F. Yu, X. Yang, and L. Guo-liang, "Rural power system load forecast based on principal component analysis," *J. Northeast Agricult. Univ. English Ed.*, vol. 22, no. 2, pp. 67–72, Jun. 2015.
- [22] F. M. Bianchi, E. De Santis, A. Rizzi, and A. Sadeghian, "Short-term electric load forecasting using echo state networks and PCA decomposition," *IEEE Access*, vol. 3, pp. 1931–1943, 2015.
- [23] N. Abu-Shikhah and F. Elkarmi, "Medium-term electric load forecasting using singular value decomposition," *Energy*, vol. 36, no. 7, pp. 4259–4271, Jul. 2011.
- [24] S. Arora and J. W. Taylor, "Short-term forecasting of anomalous load using rule-based triple seasonal methods," *IEEE Trans. Power Syst.*, vol. 28, no. 3, pp. 3235–3242, Aug. 2013.
- [25] N. Shabbir, L. Kutt, M. Jawad, M. N. Iqbal, and P. S. Ghahfaroki, "Forecasting of energy consumption and production using recurrent neural networks," *Adv. Electr. Electron. Eng.*, vol. 18, no. 3, pp. 190–197, Oct. 2020.
- [26] N. Shabbir, L. Kutt, M. Jawad, R. Amadihanger, M. N. Iqbal, and A. Rosin, "Wind energy forecasting using recurrent neural networks," in *Proc. Big Data, Knowl. Control Syst. Eng. (BdKCSSE)*, Nov. 2019, pp. 1–5.
- [27] W. Ahmed, H. Ansari, B. Khan, Z. Ullah, S. M. Ali, C. A. Mehmood, M. B. Qureshi, I. Hussain, M. Jawad, M. U. S. Khan, A. Ullah, and R. Nawaz, "Machine learning based energy management model for smart grid and renewable energy districts," *IEEE Access*, vol. 8, pp. 185059–185078, 2020.
- [28] A. A. Mamun, M. Sohel, N. Mohammad, M. S. Haque Sunny, D. R. Dipta, and E. Hossain, "A comprehensive review of the load forecasting techniques using single and hybrid predictive models," *IEEE Access*, vol. 8, pp. 134911–134939, 2020.
- [29] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—A novel pooling deep RNN," *IEEE Trans. Smart Grid*, vol. 9, no. 5, pp. 5271–5280, Mar. 2018.
- [30] L. Li, C. J. Meinrenken, V. Modi, and P. J. Culligan, "Short-term apartment-level load forecasting using a modified neural network with selected auto-regressive features," *Appl. Energy*, vol. 287, Apr. 2021, Art. no. 116509.
- [31] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 841–851, Jan. 2019.
- [32] M. Yuvaraju and M. Divya, "Residential load forecasting by recurrent neural network on LSTM model," in *Proc. 4th Int. Conf. Intell. Comput. Control Syst. (ICICCS)*, May 2020, pp. 395–400.
- [33] A. Sherstinsky, "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network," *Phys. D, Nonlinear Phenomena*, vol. 404, Mar. 2020, Art. no. 132306.
- [34] J. Liu, Z. Zhang, X. Fan, Y. Zhang, J. Wang, K. Zhou, S. Liang, X. Yu, and W. Zhang, "Power system load forecasting using mobility optimization and multi-task learning in COVID-19," *Appl. Energy*, vol. 310, Mar. 2022, Art. no. 118303.
- [35] H. Zheng, J. Yuan, and L. Chen, "Short-term load forecasting using EMD-LSTM neural networks with a xgboost algorithm for feature importance evaluation," *Energies*, vol. 10, no. 8, p. 1168, Aug. 2017.
- [36] P. Lv, S. Liu, W. Yu, S. Zheng, and J. Lv, "EGA-STLF: A hybrid short-term load forecasting model," *IEEE Access*, vol. 8, pp. 31742–31752, 2020.
- [37] B. Farsi, M. Amayri, N. Bouguila, and U. Eicker, "On short-term load forecasting using machine learning techniques and a novel parallel deep LSTM-CNN approach," *IEEE Access*, vol. 9, pp. 31191–31212, 2021.
- [38] X. Liu, Z. Zhang, and Z. Song, "A comparative study of the data-driven day-ahead hourly provincial load forecasting methods: From classical data mining to deep learning," *Renew. Sustain. Energy Rev.*, vol. 119, Mar. 2020, Art. no. 109632.
- [39] N. Somu, G. Raman MR, and K. Ramamritham, "A deep learning framework for building energy consumption forecast," *Renew. Sustain. Energy Rev.*, vol. 137, Mar. 2021, Art. no. 110591.
- [40] S. H. Rafi, Nahid-Al-Masood, S. R. Deeba, and E. Hossain, "A short-term load forecasting method using integrated CNN and LSTM network," *IEEE Access*, vol. 9, pp. 32436–32448, 2021.
- [41] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, Apr. 2016.

- [42] H. J. Sadaei, P. C. de Lima e Silva, F. G. Guimarães, and M. H. Lee, "Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series," *Energy*, vol. 175, pp. 365–377, May 2019.
- [43] *Calendar for Year 2009 (Malaysia)*. Accessed: Aug. 26, 2021. [Online]. Available: <https://www.timeanddate.com/calendar/?year=2009&country=69/>
- [44] P. Ray, D. P. Mishra, and R. K. Lenka, "Short term load forecasting by artificial neural network," in *Proc. Int. Conf. Next Gener. Intell. Syst. (ICNGIS)*, Sep. 2016, pp. 1–6.
- [45] S. Liu, Y. Cui, Y. Ma, and P. Liu, "Short-term load forecasting based on GBDT combinatorial optimization," in *Proc. 2nd IEEE Conf. Energy Internet Energy Syst. Integr. (EI2)*, Oct. 2018, pp. 1–5.
- [46] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. ICLR*, San Diego, CA, USA, 2015, pp. 1–15.
- [47] Z. Zhang, "Improved Adam optimizer for deep neural networks," in *Proc. IEEE/ACM 26th Int. Symp. Quality Service (IWQoS)*, Jun. 2018, pp. 1–2.
- [48] M. Alhoussein, K. Aurangzeb, and S. I. Haider, "Hybrid CNN-LSTM model for short-term individual household load forecasting," *IEEE Access*, vol. 8, pp. 180544–180557, 2020.
- [49] S. Shi, Q. Wang, P. Xu, and X. Chu, "Benchmarking state-of-the-art deep learning software tools," in *Proc. 7th Int. Conf. Cloud Comput. Big Data (CCBD)*, Nov. 2016, pp. 99–104.
- [50] *Tensorflow: An end-to-end Open Source Machine Learning Platform*. Accessed: Aug. 9, 2021. [Online]. Available: <https://www.tensorflow.org/>
- [51] *Keras: The Python Deep Learning Library*. Accessed: Aug. 9, 2021. [Online]. Available: <https://keras.io/>
- [52] G. Vrbancic and V. Podgorelec, "Transfer learning with adaptive fine-tuning," *IEEE Access*, vol. 8, pp. 196197–196211, 2020.
- [53] F. X. Diebold and R. S. Mariano, "Comparing predictive accuracy," *J. Bus. Econ. Statist.*, vol. 13, no. 3, p. 253, Jul. 1995.



KHALID IJAZ received the M.Sc. degree in electrical engineering from UMT, Lahore. He is currently pursuing the Ph.D. degree with COMSATS University Islamabad, Lahore Campus, Pakistan. Since August 2018, he has been working as a Lecturer with the Electrical Engineering Department, School of Engineering, UMT, where he was also a Lab Engineer, from April 2013 to July 2018. He was also associated with the Telecom industries and worked at PTCL and WorldCall, Pakistan for

five years. He has an in-depth knowledge and hands on experience of mobile switching center (MSC) and CDMA technology. His research interests include machine learning, deep learning, and reinforcement learning applied to time-series and imaging data.



ZAWAR HUSSAIN received the bachelor's degree in electrical engineering and the master's degree in electronics engineering with specialization in communication systems engineering from the Hochschule Bremen University of Applied Sciences, Germany. He is currently pursuing the Ph.D. degree in electrical engineering with the Lahore University of Management Sciences (LUMS), Lahore, Pakistan. Since October 2014, he has been working as an Assistant Professor

with the Department of Electrical Engineering, School of Engineering, UMT, Lahore. During his academics, he was associated with several institutes as a Research Assistant, including the Laboratory of Optoelectronics and Light-Wave Technology and Institute of Aerospace and Technology, Hochschule Bremen. Moreover, he also worked with leading German automotive industry, including Siemens and BMW. His research interests include wireless communication, RF microwave communication, machine learning, and artificial intelligence.



JAMEEL AHMAD received the Ph.D. degree in electrical engineering from UET Lahore, and the M.Sc. degree in electrical engineering from the University of Southern California (USC), Los Angeles, USA, in 2005. He is currently working as an Assistant Professor with the Department of Computer Science, University of Management and Technology (UMT), Lahore, Pakistan, where he was also worked with the Department of Electrical Engineering, SEN, for about ten years and

remained as the Director of final year projects with the BSEE Department for four years. He worked at Qualcomm and Broadcom Corporation, San Diego, California, USA, from June 2007 to June 2010, on 2G and 3G mobile communication systems. His research interests include distributed control and optimization of smart microgrids for energy internet, cloud computing, parallel and distributed computing, machine learning techniques applied to face recognition, energy, wireless systems, and connected and autonomous vehicles. He is a reviewer of numerous journals and conferences.



SYED FAROOQ ALI received the M.S. degree in CS from The Ohio State University, Columbus, USA, the M.S. degree (Hons.) in CS from LUMS, Lahore, Pakistan, and the Ph.D. degree in CS from UMT, Pakistan. During his M.S. degree, he was on LUMS Fellowship. He did his Ph.D. course work and the Ph.D. comprehensive exam at The Ohio State University. He is currently working as an Assistant Professor with UMT. His research interests include computer vision, digital image

processing, and medical imaging. He is a reviewer for various IEEE conferences and journals.



MUHAMMAD ADNAN received the bachelor's degree in electrical engineering (specialization in electronics and communication) from UCET, Bahuaddin Zakariya University, Multan, in 2002, the M.S. degree in telecommunication engineering from UET, Peshawar, in 2005, and the Ph.D. degree in information and communication systems from the Beijing University of Posts and Telecommunication, China, in 2010. His research interests include information and communication, energy, and artificial intelligence. He is a reviewer of numerous journals and conferences. He has been a member of the Faculty of Electrical Engineering, UMT, since October 2013. Prior, he served PTCL-Etisalat Academy, from 2004 to 2010, and NFC Institute of Engineering and Technology, Multan, from 2011 to 2013.

and artificial intelligence. He is a reviewer of numerous journals and conferences. He has been a member of the Faculty of Electrical Engineering, UMT, since October 2013. Prior, he served PTCL-Etisalat Academy, from 2004 to 2010, and NFC Institute of Engineering and Technology, Multan, from 2011 to 2013.



IKRAMULLAH KHOSA (Member, IEEE) received the Ph.D. degree in electronics and telecommunications from the Politecnico di Torino, Italy, in 2015. He is currently working as an Assistant Professor with the Electrical and Computer Engineering Department, COMSATS University Islamabad, Lahore Campus. His research interests include artificial intelligence, data analysis, machine learning, and pattern recognition.

...