Expert Systems With Applications 175 (2021) 114844

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Deep sequence to sequence Bi-LSTM neural networks for day-ahead peak load forecasting

Neelam Mughees [a,*], Syed Ali Mohsin [b], Abdullah Mughees [c], Anam Mughees [d]

[a] School of Engineering and Technology, National Textile University, Faisalabad, 37610, Pakistan
[b] Department of Electrical Engineering, FAST-NUCES, Islamabad, 44000, Pakistan
[c] School of Electrical Engineering and Computer Science, NUST, Islamabad, 44000, Pakistan
[d] Department of Electrical Engineering, GC University, Faisalabad, 38000, Pakistan

ARTICLE INFO

Keywords:
Peak demand forecasting
Demand response programs
Bidirectional long-short term memory networks
Sequence to sequence regression
Deep neural networks

ABSTRACT

The power industry is currently facing the problem of an electricity supply–demand imbalance. The most inexpensive and efficient solution to alleviate this imbalance is to decrease electricity demand. Local electrical utilities should deploy demand response programs (DRP), and short-term peak demand forecasting (STPDF) plays a crucial role in their successful deployment. In residential sectors, peak demand forecasting is also critical because the energy policies, technological growth, and changing climate are further increasing the peak demand. Therefore, an accurate peak demand forecasting will help utility companies in avoiding blackouts and secure a continuous power supply by implementing subsidy schemes such as DRP. However, daily peak load is volatile, nonstationary, and nonlinear in nature, and hence it is hard to predict it accurately. This research work for the first time has attempted to design, implement, and test deep bidirectional long short-term memory based sequence to sequence (Bi-LSTM S2S) regression approach for "day-ahead" peak demand forecasting and has accomplished preliminary success. The day-ahead peak electricity demand forecasting model is designed and tested using the MATLAB software. For performance comparison, shallow Bi-LSTM S2S, shallow LSTM S2S, deep LSTM S2S, Levenberg-Marquardt backpropagation artificial neural networks (LMBP-ANN), and medium Gaussian support vector regression (MG-SVR) forecasting models are also developed and tested. Mean absolute percentage error (MAPE) and Root Mean Squared Error (RMSE) are used as performance metrics. It has been found out that in terms of both performance metrics, the proposed deep Bi-LSTM S2S day-ahead "peak" forecasting model has outperformed all the other models on both public holidays and normal days. The load pattern on public holidays is always different than on normal days, and there is always less data available in contrast to the normal days. Therefore, it is hard to accurately forecast their load.

1. Introduction

1.1. Motivation

Deployment of Demand response programs, formally known as short-time load management, mainly began in the summer of 1999 when many electrical utilities in the US had to deal with low generation electricity capacity, high electricity prices, and the aggressive pressure in the market to generate more profit per unit investment (Sioshansi & Vojdani, 2001). Globally, Demand Response is defined as "a tariff or program established to motivate changes in electric use by end-use customers, in response to changes in the price of electricity over time, or to give incentive payments designed to induce lower electricity use at times of high market prices or when grid reliability is jeopardized (Qdr, 2006)." The demand response programs can reduce the demand for electricity by changing the energy-usage habits of consumers.

Demand response programs curtailed load during the peak hours so new peak-power plants could have been an alternative. But by doing so there was going to be extra unused energy and fixed electricity costs during the off-peak hours too. Investors didn't want to invest money in such expensive power plants because even the high price of electricity during the peak hours didn't compensate for the huge capital costs as peak-hour power plants were only in operation during the peak hours (Sioshansi & Vojdani, 2001). Few hours of operation per year resulted in

* Corresponding author.
E-mail addresses: neelam.mughees@ntu.edu.pk (N. Mughees), ali.mohsin@nu.edu.pk (S.A. Mohsin), abdullah225mughees@gmail.com (A. Mughees), anammughees@gcuf.edu.pk (A. Mughees).

https://doi.org/10.1016/j.eswa.2021.114844
Received 25 December 2020; Received in revised form 2 March 2021; Accepted 2 March 2021
Available online 11 March 2021
0957-4174/© 2021 Elsevier Ltd. All rights reserved.

very little or no revenue. Hence, increasing the capacity was out of the question and the best utilization of present resources was necessary. The demand response concept became popular at that time as building more generating power plants to meet the increased demand was going to be much more costly.

Customers can respond to DR programs in three different ways (Albadi & El-Saadany, 2007). They can reduce their energy consumption during the peak demand hours or they can shift their peak load to off-peak hours. Lastly, customers can also respond by using backup generators to fulfill their demand during peak hours rather than consuming the electricity from the grid. All of these behaviors will result in flattened load curves and hence successful deployment of demand response programs. Backup generation is usually not a preferred method because it can cause pollution concerns if generation is not from environment-friendly energy resources. Price and demand forecasting plays an important part in the successful deployment of demand response programs. For example, if the utilities want to implement real-time pricing demand response programs, they would need to know the day-ahead or week-ahead demand so that they could decide beforehand what price of electricity they would need to charge for curtailment of peak load.

Therefore, load forecasting is crucial for utility for the management of demand response programs (Chan et al., 2012). Further, precise load prediction has become extremely vital because it assists utilities in effectively scheduling the loads and minimizing the surplus power production. Load forecasting that is based upon the length of time to be forecasted, known as forecasting horizon, is mainly of four different types (Hernandez et al., 2014):

1. Very short-term load forecasting (VSTLF): In this type, the load is usually forecasted from seconds to numerous hours. These models help in controlling the load flow.
2. Short-term load forecasting (STLF): In this type, the load is forecasted from few hours to weeks. These forecasting models help in balancing the generation and demand and offering incentive schemes such as demand response in the market.
3. Mid-term load forecasting (MTLF): These forecasting models forecast load from several months to a year. Electric utility asset planning is usually done with midterm load forecasting and,
4. Long-term load forecasting (LTLF): The last category involves forecasting load from several months to several years. Electric utility asset planning is usually done with long term load forecasting.

Furthermore, load forecasting can also be categorized based upon the aim of forecast (Hernandez et al., 2014):

1. One-value load forecasting (OVLF): In this category, only one value of the load is forecasted e.g., next day's peak load, next hours' load, next minute's load, next day's total load, and next year's total load.
2. Multiple-values load forecasting (MVLF): This category includes forecasting of multiple load values e.g. next 24-hours load, next hours load, next 48-hours load, and peak load with total load.

Traditionally, long-term forecasting was in demand, and it was never felt by the electrical utilities that short-term load forecasting could also prove to be helpful. However, with the advent of Real-Time Pricing DR techniques, the need for forecasting load for short time intervals i.e., week-ahead, day-ahead, hour-ahead, half hour-ahead, or even for a few minutes-ahead emerged (Chan et al., 2012). It became a necessity because utilities have to decide their real-time prices according to the accurately predicted load/demand. For example, if any Utility is charging $50 for electricity demand of 100 MW and the projected peak load for the next day rises, the utility's succeeding day price will also increase accordingly. It is obvious that utilities cannot first wait for demand to actually rise and then increase their charging price so load prediction is essential which is always very close to the actual demand.

Apart from managing Demand Response Programs, short term peak load forecasting is also crucial for the rectified power system in a way that it helps in planning and making reliable, secure, and cost-effective working policies (Srivastava, Pandey, & Singh, 2016; Hagan & Behr, 1987). Forecasting algorithms are dependent upon historical data to predict accurate future load. The historical data can be obtained from websites, directly from the utility officials, or smart meters installed at consumer premises as in (Quilumba et al., 2014), which has also considered the accuracy of prediction. Moreover, the smart meter data is diverse, complex, and huge and the traditional short-term load forecasting algorithms have not shown good prediction accuracies. For improving this accuracy, (Wu, Kong, Hao, & Chen, 2020) proposed a hybrid neural network model incorporating gated recurrent unit (GRU) and convolutional neural networks (CNN). The results confirmed that the hybrid model significantly improved the forecasting accuracy when compared with CNN, GRU, and backpropagation neural networks. However, in this research work, the historical load data of daily peak demand was obtained directly from the utility officials.

It can be seen that effective implementation of demand response programs is highly dependent upon electricity load and price forecasting. Therefore, this research work is one of its kind to check the effectiveness of the sequence-to-sequence regression approach using deep bi-directional long short-term memory networks (Bi-LSTM S2S) for forecasting the day-ahead peak electricity demand of a residential region being supplied by Faisalabad Electric Supply Company (FESCO). The research work has also aimed at addressing the challenge of effectively forecasting load on public holidays (special days).

Using the same historical data, the proposed model is also compared with its shallow variant and the deep and shallow variant of sequence to sequence unidirectional LSTM (LSTM S2S). The traditional approaches of artificial neural networks ANN and support vector regression SVR are also trained and implemented using the dame data to draw appropriate conclusions. During the hours of power shortfall, if an accurate peak demand forecast is available, FESCO can rely on it to offer incentive-based, time-based, and energy-management based DR programs, and find the target customers who will most likely participate in such programs. Demand response programs are the best way to reduce peak electricity demand, and thus tackling the problem of supply–demand imbalance. Utilities such as FESCO will be more confident about peak demand reduction and planning their electricity control operations. This demand forecasting will help FESCO in the successful deployment of demand response programs in the application area and thus achieving all the benefits that come with them. The contributions of this research work are summarized in sub-section 1.3 after conducting a detailed literature review in Subsection 1.2. Further, Section 2 describes the proposed methodology and background theory in detail, Section 3 illustrates the results and section 4 discusses them. Finally, section 5 concludes the research work and mentions future directions.

1.2. Literature review

In literature studies, a wide range of algorithms are being used for forecasting short-term demand, and they can be grouped into four different techniques, i.e., Artificial Intelligence, Statistical, Knowledge-Based Expert Systems, and Hybrid (Srivastava et al., 2016). Fuzzy logic, artificial neural networks, genetic algorithms, and support vector machines are widely used in artificial intelligence-based forecasting techniques. These networks require proper training, testing, and validation, and thus network stability problems can arise. To combat such resulting problems, continuous monitoring of these networks is required during the training process. Adaptive load forecasting, multiple regression method, stochastic time series, exponential smoothing, and iterative reweighted least square are some of the models which are used in forecasting load using statistical technique. Similar day approach method is also one of the traditional techniques for forecasting short term load. Forecasting using neural networks can also be combined with a fuzzy neural approach that works by fuzzy interfaces as in (Mori &

Kobayashi, 1995; Egrioglu et al., 2013).

Recently, deep neural networks-based models are gaining importance in the field of load forecasting. Deep learning is an artificial intelligence technique that involves deep network structures (Kim et al., 2015). Recurrent neural networks (RNN) proposed in (Hopfield, 1982) are a type of 'deep' networks that can consider the time-based correlation in a time-series (Aquino, et al., 2020; Chiang, Chen, & Huang, 2019; de Jesus Rubio, 2009; de Rubio & Systems, 2020; Hernández, Zamora, Sossa, Téllez, & Furlán, 2020; Meda-Campaña, 2018). It can model time-series more effectively as it can take in any length of historical data. Nevertheless, RNN suffers from the problems of gradient explosion and vanishing gradient during the training process. To overcome these problems, long short-term memory networks (LSTM) were proposed in (Hochreiter & Schmidhuber, 1997). Currently, LSTM is popular among researchers in fields such as natural language translation (Sutskever et al., 2014), voice recognition (Zazo et al., 2018), finance (Muncharaz, 2020), and speech synthesis (Zen & Sak, 2015).

Lately, research on LSTM for electrical load forecasting is being conducted. The aim is to effectively make use of the historical data and develop a forecasting model that is both reliable and efficient. However, long short-term memory networks are still very new to short-term load forecasting, and very few research papers implementing this technique are available on the internet. All the published conferences and journal papers are mostly after 2016 suggesting that the promising scope of LSTM networks is yet to reveal. The authors in (Bouktif et al., 2018) did medium to long-term forecasting and have developed many non-linear and linear machine learning models to select the best one as a baseline. The embedded and wrapper feature selection techniques selected the optimal features. The LSTM model performance was optimized by using the genetic algorithm to find an optimum number of layers and optimal time lags. The performance metrics Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) illustrated that the LSTM forecasting model captured the complex characteristics of the time series load data, and outperformed the machine learning algorithm that was optimized by tuning its hyperparameters.

In (W. He, 2017), the authors predicted 24 h ahead load on an hourly basis. They used timing, weather, and holiday data for the current hour and previous 24 h load as predictor variables. The time sequence length was 168 since data was used for 1 week. The proposed deep learning model had a convolutional neural network (CNN) part and recurrent neural network (RNN) employing LSTM cells part. The results depicted the superiority of their proposed deep model over traditional methods such as SVM and LR. A hybrid forecasting model based on empirical mode decomposition (EMD), similar days (SD) selection, and LSTM has been proposed in (Zheng et al., 2017). For analyzing the similarity between historical days and forecasting days the extreme gradient boosting based KM algorithm is utilized. The results illustrated the electrical load can be accurately forecasted.

LSTM for forecasting domestic load at a single household level has been proposed in (Kong et al., 2017) to deal with the uncertainty and high volatility issue. The proposed model is tested on public residential smart meter data set gathered by the Smart Grid Smart City project started by the Australian government. The results showed that the LSTM forecasting model outperformed extreme learning machine (ELM), backpropagation ANN, k-nearest neighbor (K-NN), and hybrid benchmark models for short term load forecasting. Likewise, (Y. Wang, Gan, et al., 2019) have used probabilistic load prediction technique to forecast load for individual homes. LSTM based deep model learned both the short- and long- term characteristics of the load profiles. Instead of mean square error (MSE), a pinball loss function is utilized to tune the parameters. Therefore, the authors have used quantiles to extended to probabilistic forecasting from the standard LSTM point forecasting. The model was tested on both commercial and residential customers and results depicted that the LSTM showed better performance over standard techniques.

In (F. He et al., 2019), a short-load forecasting model based on LSTM and Variational Mode Decomposition (VMD) has been proposed. VMD first decomposes the signal, take into account the related parameters, and uses a correlation coefficient for extending the input sequence. Bayesian Optimization Algorithm (BOA) has optimized the related parameters and LSTM forecasts the subsequences at the end. For comparison purposes, decomposition based and the individual models-based methods are used. The results demonstrated that the proposed hybrid LSTM + VMD model improved the forecasting accuracy in comparison to other methods. Another hybrid model incorporating LSTM and SNN was proposed in (M. Kim, Choi, Jeon, & Liu, 2019) for predicting next-week load. The authors incorporated seasonal, humidity, and temperature information for training the LSTM network and then the CNN layer forecasted next-week demand profile. The experiment was conducted on real-world data and the results demonstrated that the proposed hybrid LSTM + CNN model outperformed the previous methods of ARIMA, S2S LSTM, and LSTM. The hybrid combination of LSTM + CNN was also proved to be the best method for forecasting residential load consumption in (T.-Y. Kim & Cho, 2019).

Recently, (Somu et al., 2020) used LSTM and improved sine cosine optimization (SCO) method were used to forecast long, mid, and short-term load. A new Haar wavelet dependent mutation algorithm was proposed to diverge the SCO algorithm to obtain a global optimal solution. The LSTM hyperparameters were optimized using SCO algorithm and the model was tested on real-time power consumption data. The results in the form of MAE, MAPE, MSE, RMSE, and Theil statistics demonstrated that their model surpassed the other traditional models in terms of performance.

Bi-directional long short-term memory (Bi-LSTM) networks are a further improvement as they run information both ways i.e., from forward to backward and from backward to forward. Hence, they can learn from both future and past values. This property is particularly crucial in effectively forecasting electricity demand, which depends upon historical data. Presently, very few researchers have done STDF using the Bidirectional LSTM time series approach and regression approach and most of the journal papers available are from the current and the previous year. In 'deep' Bi-LSTM, there are two or more numbers of Bi-LSTM layers, whereas its 'shallow' variant only consists of one Bi-LSTM layer. In (Di Persio & Honchar, 2017), variations of standard LSTMs i.e., multi-layer LSTM, Bi-LSTM and sequence to sequence LSTM has been tested for short term load forecasting. The results indicated that Bi-LSTM outperformed the other methods on that specific time series load.

A multi-layer bidirectional RNN based on LSTM and gated recurrent unit (GRU) have been proposed in (Tang et al., 2019). In first layer, the LSTM units are arranged to run the data forwards and the GRU units run the data backwards, and vice versa happens in the second layer. The model is validated on a power company data and a competition data. Peak hourly load has been forecasted and the results showed that the proposed Bi-RNN outperformed SVR and BPNN models. Although the authors have tested the bidirectional RNNs it is important to note that their model doesn't employ Bi-LSTM.

In (S. Wang, Wang, Wang, & Wang, 2019), researchers have used a shallow Bi-LSTM network along with rolling update (RU) and attention mechanisms (AM) for short term load forecasting. The RU is used for real-time upgradation of the data and AM assigns influence weights to illustrate the efficient predictor variables characteristics. Finally, they forecasted the next half-hour load values using softmax layer and linear transformation layer. The authors' results indicated that the addition of RU and AM have improved the forecasting accuracy over traditional Bi-LSTM model. However, they didn't investigate the performance of deep Bi-LSTM model. A deep stacked STDF model consisting of CNN, Bi-LSTM, and LSTM layers have been proposed in (Hong et al., 2020). CNN extracts the spatial feature, Bi-LSTM uses data at both post and pre stages, and then LSTM extracts the temporal characteristics in the power data. They have forecasted load for a large vessel and took data from a training ship.

In (Atef & Eltawil, 2020), deep stacked LSTM and Bi-LSTM forecasting models for predicting the next hour load has been proposed. Two stacked architectures for both Bi-LSTM and LSTM models having two and three layers are compared with shallow LSTM model. A hyper-parameter optimization tool finds the best model and the results showed that deep models did not significantly improve the time series forecasting accuracy and takes twice the computational time when compared with shallow models. However, in each case, the Bi-LSTM model surpassed the LSTM model and SVR model in terms of RMSE, MAPE, and MAE.

It can be seen that almost all the previous research works except for one paper (Atef & Eltawil, 2020) has only focused on shallow models of Bi-LSTM networks for short term load forecasting. However, the authors in (Atef & Eltawil, 2020) have only tested the performance of deep configurations of Bi-LSTM for forecasting the next hour load. There is no attempt at forecasting the peak load and the authors have not mentioned whether they have used sequence to sequence regression approach or sequence to one. The authors found out that deep LSTM models didn't improve the forecasting accuracy. However, the research works done in other fields such as in speech recognition (Graves et al., 2013) and network-wide traffic state prediction (Cui et al., 2020) demonstrate that deep Bi-LSTM models or a combination of both LSTM and Bi-LSTM layers can significantly improve the forecasting accuracy of time series data.

1.3. Research gap and contributions

From the above-mentioned research works, we have identified three major research gaps. To the best of our knowledge, first major research gap is that there are no research papers that have checked the performance of shallow or deep Bi-LSTM models for forecasting the volatile, non-stationary, and non-linear "peak" load. The statistical properties such as autocorrelation, variance, and mean of peak load varies with time rather than staying constant. This makes it hard to accurately forecast the peak demand of electricity (Yu et al., 2019). In addition, the peak-load is specifically targeted because it influences the management of DRP the most as the main purpose of DRP is to reduce the peak load. We have found that there is one research paper available for forecasting the next hour load using deep Bi-LSTM, but no research work has used it for forecasting the daily "peak" load.

Secondly, there are no attempts at finding out if Bi-LSTM deep or shallow forecasting models can accurately forecast demand on special days or holidays. Public holidays always have a different load pattern or peak demand in contrast to normal weekdays, therefore, there can be big prediction errors when predicting load on a holiday. Moreover, according to (Ziel & Energy, 2018), the special days or holidays can be further grouped into two classes. The first class includes holidays that are always on the same date, for example, an Independence Day of a country. The second class includes those holidays that do not occur on the same date every year, for example, Eid holidays in Muslim countries. Therefore, this property of holidays along with the fact that there is always less data available for holidays in comparison to normal days also make it difficult to accurately forecast their peak demand.

Thirdly, no research work has used Bi-LSTM to address the issue of forecasting peak demand in developing countries where the power system is still underdeveloped and no smart meters have been installed to record large amounts of historical data. There must be forecasting models available for countries with grossly limited historical data. Based on the above three research gaps, the following contributions have been made in this research work:

1. It has successfully designed, implemented, and tested deep Bi-LSTM sequence to sequence regression approach for day-ahead "peak" demand forecasting for the first time.
2. It has successfully assessed if the proposed model is the best by comparing it with optimized shallow Bi-LSTM S2S, shallow LSTM S2S, deep LSTM S2S, LMBP-ANN, and MG-SVM forecasting models.
3. It is also the first attempt on forecasting day-ahead peak demand of a sector being supplied by FESCO as currently, there is no other forecasting model available.
4. It has efficiently addressed the challenge of accurately forecasting load on both classes of public holidays.
5. It has provided a solution for the management of demand response programs in Pakistan and similar developing countries.
6. The developed model is affective for grossly limited data, which means that it does not require large amounts of data for training.
7. This research work is useful for developing countries with small total load, especially at feeder level.

2. 2. Methodology

Historical daily peak electricity demand data for 30 months (May 2015–September 2017) is obtained from FESCO and has been published at (Mughees, 2020). FESCO is still under development and no smart meters have been installed at the customer premises to automatically record the consumption readings, therefore, the data obtained was manually written. It was initially in the form of daily peak current consumption I_p of a region being supplied by FESCO. It has been converted to daily peak active power consumption P or peak demand using the formula of the active power of three-phase AC system, given by (1):

$$P = \sqrt{3}V_L I_L \cos\theta. \tag{1}$$

where V_L is the line-to-line voltage, I_L is the line-to-line current, $cos\theta$ is the power factor and θ is the angle between phase voltage and line voltage of the power distribution system. Furthermore, power factor supplied by FESCO is maintained at 0.9 and V_L of the application area is 11 kV. These values were also same at the time of measurement. Electrical loads consume active power, and that is why it is usually forecasted and is known as the demand of electricity.

The obtained data is recorded in a manual form, so it has typographical errors resulting in missing values, outliers, and duplicate entries, which require data preprocessing before it could be used for accurate electricity demand forecasting. Moreover, demand forecasting is greatly influenced by some external factors. Short term demand forecasting is specifically affected by weather, time, economic, calendar, and random events. Therefore, considering these external factors input sequence X consists of 9 predictor variables or features, i.e., week of the year W, day of the month D_m, day of the week D_w, special day (holiday or not) S, maximum temperature T, maximum dew point temperature T_D, rainfall R, previous day peak electricity demand P_D, and previous week peak electricity demand P_W.

A statistical method has also been used to check the degree of correlation of all these predictor variables with the power demand. For each of the features, 818 time steps or observations from May 2015 to July 2017 are used for training, testing, and validation of the forecasting models, and then predicted demand by each model is compared with real demand data of public holidays, and different weeks of August, September, and October 2017. Mean absolute percentage error (MAPE) and root mean squared error (RMSE) (Aquino, et al., 2020; Chiang, et al., 2019; de Jesus Rubio, 2009; de Rubio & Systems, 2020; Hernández, et al., 2020; Meda-Campaña, 2018) are used as error evaluation metrics and are calculated for each case. Lower values of MAPE and RMSE refers to a more accurate and flexible model. MAPE is given by (2) and RMSE is given by (3):

$$MAPE = 1/N\sum_{i=1}^{N}\frac{P_A^i - P_P^i}{P_A^i}\times 100 \tag{2}$$

$$RMSE = \sqrt{1/N\sum_{i=1}^{N}\frac{\left(P_A^i - P_P^i\right)^2}{P_A^i}} \tag{3}$$

where P_A^i and P_P^i is the actual and forecasted electricity demand on a certain day i and N is the total number of days.

2.1. Data preprocessing

One of the significant problems that demand forecasting is facing is the presence of missing values in the observed historical load data. The problem of missing values mostly exists in statistical data analysis. The demand data that is obtained from FESCO is analyzed, and many recorded values are detected to be empty or null. Missing values are reconstructed using simple linear interpolation of neighboring values, linear state-space model, linear grey-box model, moving average method, and moving median method. However, the moving median method over a window size of 5 gave the best forecasting results in terms of MAPE and RMSE performance metrics.

Outlier values are extreme values which appear to be inconsistent with the rest of the recorded data. Outliers also considerably affect the quality of forecasting models. In the case of data obtained from FESCO, outliers are mined by applying Grubb's test, the generalized extreme Studentized deviate test, quartiles method, moving median, and moving average method. For filling the outliers, different techniques such as filling it with the center value determined by the mining method, filling it with the next, previous or nearest non-outlier value, filling it with linear interpolation, piecewise cubic spline interpolation, and shape-preserving piecewise cubic spline interpolation have been applied. However, moving median mining method and filling it with the next non-outlier value produced promising results. Furthermore, constant rows and duplicate entries of data are also removed. Normalization of data is achieved through the min–max normalization technique defined by the formula given in (4):

$$x^{`} = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \tag{4}$$

where $x^{`}$ is the normalized value. Fig. 1 illustrates the state of demand data before and after it is preprocessed using moving median method.

2.2. Deep Bi-directional long short term memory (Bi-LSTM) networks for the sequence-to-sequence regression

The deep learning concept has been roaming around for decades. In 1943, McCulloch and Pitts, for the first time, proposed the idea of 'deep learning' using the name 'cybernetics' (McCulloch & Pitts, 1943). However, the main hurdles in its application are (Shi et al., 2018): a) insufficient training and testing data, b) deficiency of effective training algorithm and, c) deficiency of high computing power resources for handling big network architecture.

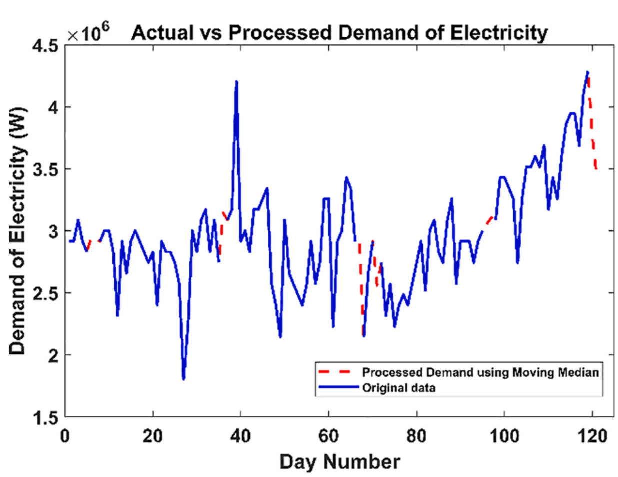

Fig. 1. Actual vs Processed Peak Demand of Electricity using Moving Median method.

In recent years, digitalization of advanced society and the production of high-performance computers have addressed these hurdles. Moreover, real-world application of deep learning was made possible when Geoffrey Hinton made a revolutionary strategy of greedy layer-wise pre-training for efficient training of deep networks. Long Short Term Memory (LSTM) networks are a type of deep learning neural networks known as recurrent neural networks (RNN) (Kong et al., 2017). They are capable of learning long term dependencies between the time steps of sequence raw data. Simple RNN can be treated as the processor of neural networks with sequential data as it has an internal memory that is capable of updating the condition of neurons with the prior network input (Almalaq & Edwards, 2017).

The well-known training algorithms that are used for RNN networks are backpropagation and real-time recurrent learning. The deep architecture of RNN is obtained by stacking RNN layers on top of each other. Nonetheless, the disadvantage of recurrent neural networks is that they stop working because of the problem of vanishing gradient descent (Marino et al., 2016). If the non-linear activation functions of neural networks are differentiable then a gradient descent optimization algorithm is used to reduce the training error (Mandic & Chambers, 2001). Gradient descent minimizes the error by taking its derivate and changing the weights in proportion to it.

Long Short Term Memory (LSTM) networks are a unique type of recurrent neural networks that are designed to combat the problem of vanishing gradient (Bouktif et al., 2018). They do so by developing a model that can keep information for a long time. LSTM network consists of memory cells that further contain self-loops as illustrated in Fig. 2. The figure shows three gates that are responsible for information flow in an LSTM cell i.e., input gate, forget gate, and output gate. The input gate is accountable for writing, forget gate for erasing, and output gate for reading the states of memory cells (Zheng et al., 2017). Storing of any sequential information, that is encoded on the memory cell's state is made possible by the self-loops in the LSTM network.

The equations governing the single cell operation of LSTM network are shown below:

$i_g = \sigma\left(i_{[t]}W_{ix} + o_{[t-1]}W_{im} + b_i\right)$ (5)

$f_g = \sigma\left(i_{[t]}W_{fx} + o_{[t-1]}W_{fm} + b_f\right)$ (6)

$f_g = \sigma\left(i_{[t]}W_{ox} + o_{[t-1]}W_{fo} + b_o\right)$ (7)

$u = \tanh\left(i_{[t]}W_{ux} + o_{[t-1]}W_{um} + b_u\right)$ (8)

$s_{[t]} = \left(f_g{}^{\circ}s_{t-1} + i_g{}^{\circ}u\right)$ (9)

$$h_{[t]} = \left(o_g{}^{\circ}\tanh(u)\right) \tag{10}$$

where i_g is the input of input gate, f_g is the input of forget gate, o_g is the

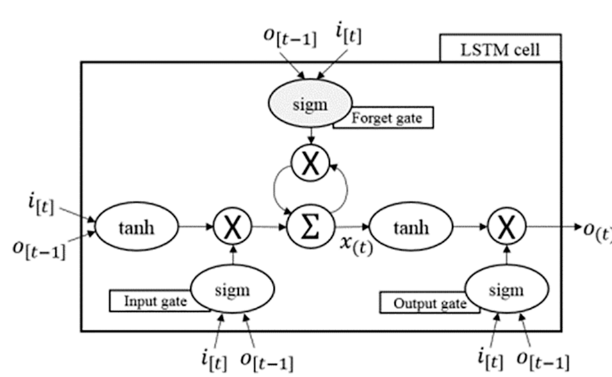

Fig. 2. Long Short Term Memory (LSTM) cell structure.

input of output gate, u corresponds to the update signal or input node, $s_{[t]}$ is the cell state value at time step t, $h_{[t]}$ is the cell's output value at time step t, σ refers to sigmoid activation function, *tanh* is the hyperbolic tangent activation function, W is the weight matrix, b is a bias term of a LSTM unit, and s_{t-1} is the preceding time step input. The sigmoid activation function σ is used because it transforms input value into a value between 0 and 1, therefore, it allows information to either flow completely through the gates or not at all. Moreover, the hyperbolic tangent activation function tanh is used to solve the problem of vanishing gradient as its second derivative gives a long range of values before it gradually decreases to zero.

Update signal modifies the memory state and whether it should do so is decided by the input gate. The input gate uses a sigmoid function that acts as a 'soft switch' and turns on or off depending upon the current input state of the cell and previous output state of the cell. The cell state will not be affected by the update signal if the value of the input gate is zero or close to zero. Long short term memory cells can be grouped to form a multi-layered or deep network architecture as shown in Fig. 3. There are many hidden cells present in every LSTM layer. The network with this architecture is usually used to predict the response value $\hat{y}[t]$ at time $t \in \mathbb{N}$ using all its previous inputs values $[i[0], i[1], \cdots, i[t]]$.

LSTM layer implemented in this research work is bidirectional which runs the input sequence in both forward and backward direction (S. Wang, Wang, et al., 2019). In every bidirectional LSTM layer, the number of memory cells are double. It has an advantage over unidirectional LSTM as it can learn from both the past and future values. This research work aims at testing this bi-directional capability of LSTM networks for short term load forecasting. The Bi-LSTM layers are stacked over each other to form a deeper Bi-LSTM forecasting model.

The Bi-LSTM layer structure adopted in this paper is illustrated in Fig. 4. Bi-LSTM layer working differs from LSTM layer in a way that it learns and stores information from the sequence input by running it in both forward and backward directions rather than in only forward direction. In the forward direction, memory cells in Bi-LSTM stores information from past values of peak demand forecasting features, and in a backward direction they are capable of storing information from the future values. This information is stored in separate hidden states ($\overleftarrow{h_t}$ and $\overrightarrow{h_t}$) which are concatenated to produce the final hidden state h_t, and then at any time t information from past and future are available.

The first LSTM unit in both the forward and backward layers take in the initial states (S_0 and h_0) of the deep neural network. In addition, it uses the first time step to calculate the first unit output h_1 and new cell state S_1. At any time step t, the cell unit takes in the upcoming time step of sequence and the current network cell and hidden states to calculate the new cell and the output state. The cell states are crucial as they learn the particulars from the preceding time steps. Moreover, each LSTM memory unit removes or add particulars from the cell states at every time step.

At any time step t, Bi-LSTM cell's forward and backward layer outputs are computed using the standard LSTM unit's operating Eqs. (5)–(10). Eq. (11) calculates each element of the final hidden state vector H_t of the Bi-LSTM layer:

$h_t = \partial(\overleftarrow{h_t}, \overrightarrow{h_t})$ (11)where ∂ represents the function that combines the forward and backward hidden state sequences. It can be a summation, a concatenating, a multiplication, or an average function. For Z time steps, the final output vector H_t can be written as $H_t = [H_{t-Z}, \cdots, H_{t-1}]$, where the last element H_{t-1} represents the forecasted peak demand, which would act as an input to the next iteration.

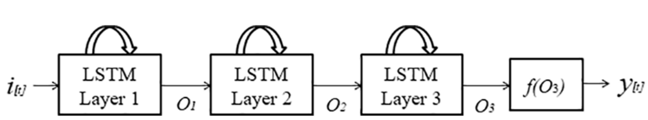

Fig. 3. LSTM multi-layered network architecture.

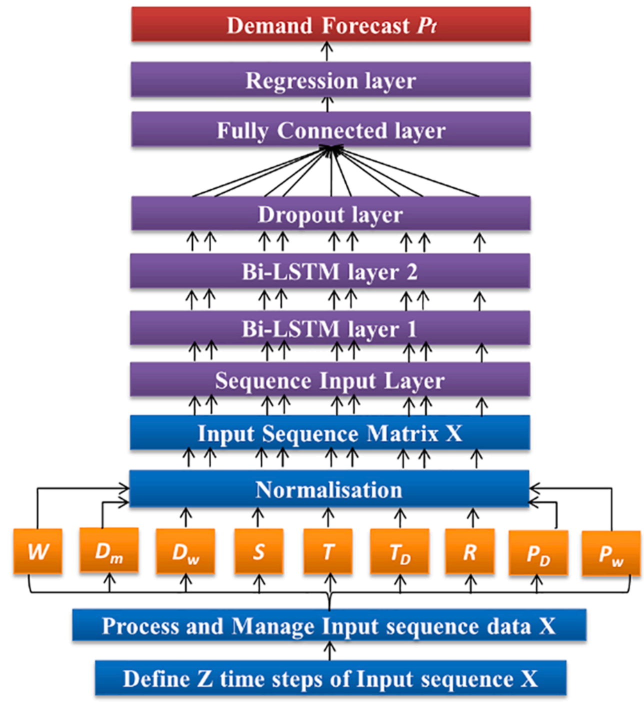

Fig. 4. Bi-LSTM layer structure.

The architecture of the proposed deep Bi-LSTM based sequence to sequence regression day-ahead demand forecasting model is based on six basic layers given in Fig. 5. The architecture starts with managing and processing the input features. The input features have values from the month of May 2015 to July 2017, which makes 818-time steps available for training. The value of Z time steps is thus 818. The techniques used for pre-processing have already been described in the previous subsection. After normalization, the input features are fed to the sequence input layer in the form of a matrix. Next, there are two Bi-LSTM layers to form the deep network and dropout layer to generalize the proposed forecasted model and avoid over fitting. Finally, the fully connected layer and regression layer work together to map the input features to the output of demand forecast. The training settings of the proposed deep Bi-LSTM architecture for peak load forecast is discussed in further detail in Section 2.3.

The main mathematical equations governing the proposed deep Bi-LSTM S2S 6-layer architecture are:

$X = [X_1, X_2, \cdots, X_n]$ (12)

$$H_{t1} = [H_{t-Z}, \cdots, H_{t-1}] \quad (13)$$

$$H_{t2} = [H_{t-Z}, \cdots, H_{t-1}] \quad (14)$$

$$F = WX_t + b \quad (15)$$

$$\text{loss} = \frac{1}{2S}\sum_{i=1}^{S}\sum_{j=1}^{V}(t_{ij} - y_{ij})^2 \quad (16)$$

where (12) represents the output X of the sequence input layer containing all the predictor variables, (13) represents the output H_{t1} of first Bi-LSTM layer at any time-step t with total Z time-steps, (14) represents the output $H_{t2}$of second Bi-LSTM layer, (15) demonstrates the output of the fully connected layer that multiplies the input sequence vector X_t by a weight matrix W and then adds a bias term b, and (16) represents the half-mean squared error loss function of the final layer of the proposed deep Bi-LSTM S2S network, i.e., regression layer. In equation (16), V represents responses number, S represents sequence length, t_{ij} is the target output, and y_{ij} is the deep Bi-LSTM network prediction.

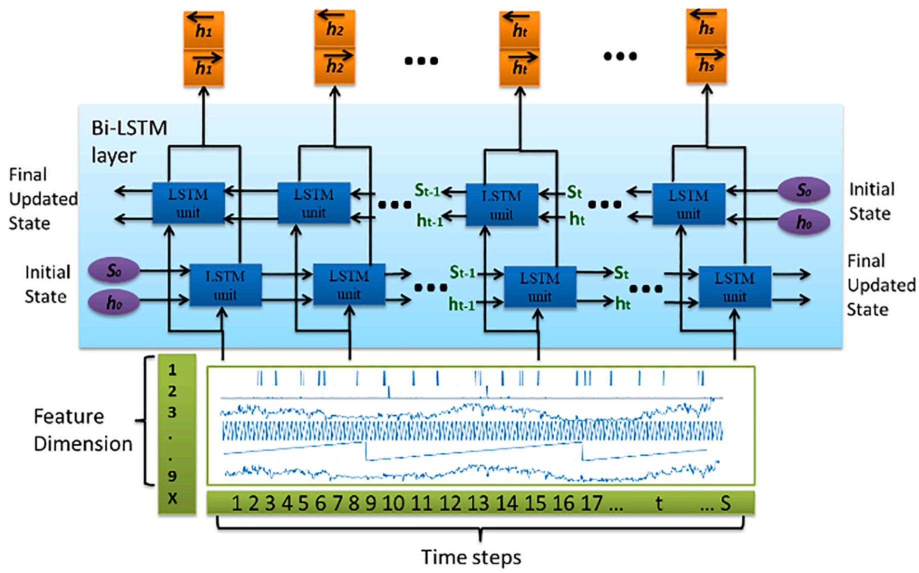

Fig. 5. Deep Bi-LSTM based sequence to sequence regression day-ahead peak demand forecasting model framework.

The input features considered in building this model are depicted in Fig. 6 and are defined as:

1. A sequence of past values of the week of the year W for Z time steps, values of W ranges from 1 to 53.
2. A sequence of values of the day of the month D_m for past Z time steps, values range from 1 to 31.
3. A sequence of past values of the day of the week D_w for Z time steps, the values range from 1 to 7.
4. The equivalent binary special day marks S for Z time steps, the values are either 0 or 1. Public holidays are considered as special.
5. Maximum temperature T sequence values for past Z time steps; T ranged from 9 to 47 for the considered Z time steps.
6. Maximum dew point temperature T_D sequence values for past Z time steps; T_D ranged from 4 to 30 for the considered Z time steps.
7. Rainfall R sequence values for past Z time steps, although it is a rare event in Faisalabad still its values ranged from 0 to 499 for the considered Z time steps.
8. A sequence of previous day peak electricity demand P_D for past Z time steps, P_D ranged from 1.54 MW to 5.83 MW for the considered Z time steps.
9. A sequence of previous week peak electricity demand P_W for past Z time steps, P_W ranged from 1.54 MW to 5.83 MW for the considered Z time steps.

The proposed framework is implemented on MATLAB software. All the required computations are performed on a laptop PC with 2.40 GHz Intel Core i3 processor and 6 GB RAM. The deep Bi-LSTM network is quite sensitive to the scale of data, so all the features are normalized by concatenating them horizontally. Each of the resulting normalized features has zero mean and unit variance. Forecasted response sequence values of power consumption value P will also be in their normalized form. Once the values are normalized to $\boldsymbol{W}, \boldsymbol{D_m}, \boldsymbol{D_w}, \boldsymbol{S}, \boldsymbol{T}, \boldsymbol{T_D}, \boldsymbol{R}, \boldsymbol{P_D}$, and $\boldsymbol{P_w}$ the input X to input sequence layer with feature dimension 9 takes the form of (17):

$$\mathbf{X} = [\boldsymbol{W}, \boldsymbol{D_m}, \boldsymbol{D_w}, \boldsymbol{S}, \boldsymbol{T}, \boldsymbol{T_D}, \boldsymbol{R}, \boldsymbol{P_D}, \boldsymbol{P_w}] \quad (17)$$

Each row of input sequence matrix X is fed to their corresponding

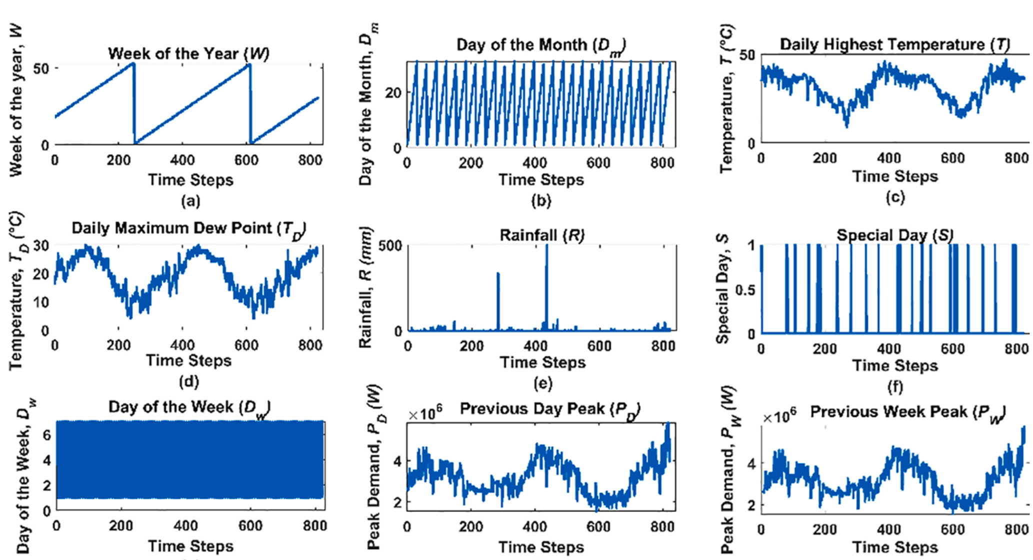

Fig. 6. Predictor variables for forecasting peak electricity demand.

two separate LSTM units in the Bi-LSTM layer. One LSTM unit is for the forward pass and the second LSTM unit is for the backward pass of each row of an input sequence. In this way, information from the past and future is stored in the Bi-LSTM layer. Hidden states are the outputs of the Bi-LSTM layer and they are responsible for memorizing the bi-directional long-term dependencies. Two Bi-LSTM networks are piled to form a deeper neural network model and improve the forecasting accuracy. The hidden states of second Bi-LSTM layer are fed to the dropout layer. Dropout layer is used to avoid network overfitting and performing poorly for new values. The outputs of the dropout layer are fed to the fully connected layer which maps the 10 outputs of the dropout layer to one single value i.e., the forecasted peak electricity demand of that particular time step. In the end, the regression output layer is connected to observe the loss function performance during the training of the network.

2.3. Training Settings

Input features' sequence values from the month of May 2017 to July 2018 are used which makes 818-time steps available for training. The value of Z is thus 818. Hence, for this particular research, after concatenation of 9 features, the dimension of input sequence X becomes 818×1 of data type cell where each row contains the normalized values of features for the corresponding time step. Lengths of sequences are set to 1 to avoid data padding in the specified mini-batch. The number of hidden units of both Bi-LSTM layers is set to 5. Dropout layer is made to have a dropout probability of 0.005. It avoids overfitting by randomly setting 1% of inputs to zero in this particular research case. The output size of the fully connected layer is set to 1 as the dimension of response value, i.e., peak electricity demand is 1.

Learnable parameters of first Bi-LSTM layer are set to recommended default values except for input weight learn rate factor which is set to 20 and input weight regularization factor is set to 1.205. These two values are set by manually optimizing the network performance. For second Bi-LSTM layer, input weight learn-rate factor which is also set to 20 but the input weight regularization factor is set to 1.2051. The number of hidden units in both networks is same i.e., 5. Adam optimizer is used for network training as it showed much better results than stochastic gradient descent momentum and root mean square propagation RMSprop optimizers for this particular problem. Maximum epochs are set to 40 and mini-batch size is set to 60.

To avoid gradient exploding, gradient threshold is set to 2. The loss function in the regression output layer is set to mean squared error MSE and training is stopped when MSE and root mean squared error RMSE are minimum. Some data was also used for validating the developed Bi-LSTM S2S forecasting model. The shallower Bi-LSTM S2S model that is used for performance comparison only has one Bi-LSTM layer with input weight learn rate factor of 20 and input weight regularization factor of 1.205. Furthermore, these two forecasting models are also compared with deep and shallow forecasting models of standard LSTM S2S using the same input sequence. Therefore, for better performance comparison, all the learnable parameters values of shallow and deep LSTM S2S models are set equal to corresponding values of shallow and deep Bi-LSTM S2S forecasting models.

3. 3. Results

The aim of this research work is to propose an original deep forecasting model for accurate and fast computation of day-ahead volatile, non-stationary, and non-linear "peak" electricity demand for both normal and special days using grossly limited data. For this purpose, a deep Bi-LSTM sequence to sequence regression model has been successfully designed, implemented, and tested. For performance comparison, the predicted day-ahead peak demands using the proposed deep Bi-LSTM S2S model are

compared with the prediction of optimized shallow Bi-LSTM S2S, shallow LSTM S2S, deep LSTM S2S, LMBP-ANN, and MG-SVR forecasting models. Further, the same historical data has been used for training, validation, and testing of the proposed and comparison models. The STPDF models are tested with actual peak demand data for 6 different weeks and 6 special days.

Week 1 and 2 consist of days from August 11, 2017, to August 24, 2017, week 3 and 4 consist of days from September 12, 2017, to September 25, 2017, and the last two weeks 5 and 6 consist of days from October 1, 2017, to October 14, 2017. Furthermore, special days consist of Pakistan Independence Day (14 August 2017), Eid Holidays (2–4 September 2017) and Ashura Holidays (30 September and 1 October 2017).

3.1. Deep Bi-LSTM based sequence-to-sequence regression (S2S) peak demand forecasting model

Figs. 7, 8, 9, and 10 demonstrate the comparison of actual peak demand with the forecasted day-ahead peak demand using the proposed deep Bi-LSTM based S2S regression model for normal and special days. In each figure, first row shows the real and predicted peak demand of electricity and second row shows the values of RMSE at each point. Table 1 displays the MAPE and RMSE values yielded by deep Bi-LSTM S2S model for 6 different weeks and 6 special days in the months of August, September, and October 2017. Special day(s) in the month of August is Independence Day, September are Eid holidays, and October are Ashura holidays. Average MAPE for 6 different weeks is found to be 4.84% and average RMSE is 0.22. For special days, the model yielded 0.14 average RMSE and 3.49% average MAPE.

3.2. Optimized peak comparison models (Shallow Bi-LSTM, Shallow LSTM, Deep LSTM, LMBP-ANN, and MG-SVR)

Table 2 displays the average MAPE and average RMSE values yielded by the other forecasting models for six different weeks and the six special days. Out of these models, best average MAPE of 4.88% and RMSE 0.22 for six different weeks is given by MG-SVM model. Moreover, for special days, the best performance out of these comparison models is given by LMPN-ANN model. Furthermore, the comparison of Table 2 and Table 3 shows that the proposed deep Bi-LSTM S2S model has outperformed all these models in terms of both MAPE and RMSE performance metrics. This comparison has been discussed in detail in the discussion Section 4.

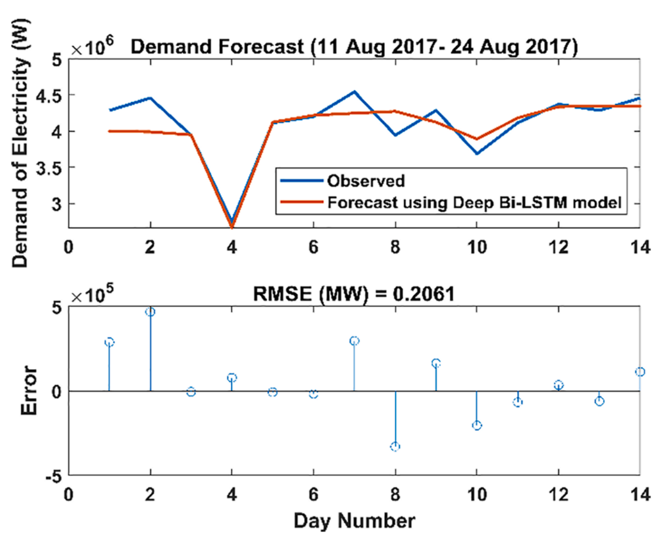

Fig. 7. Comparison of forecasted peak demand with observed peak demand using proposed deep Bi-LSTM S2S forecasting model for days from 11 Aug to 24 Aug 2017.

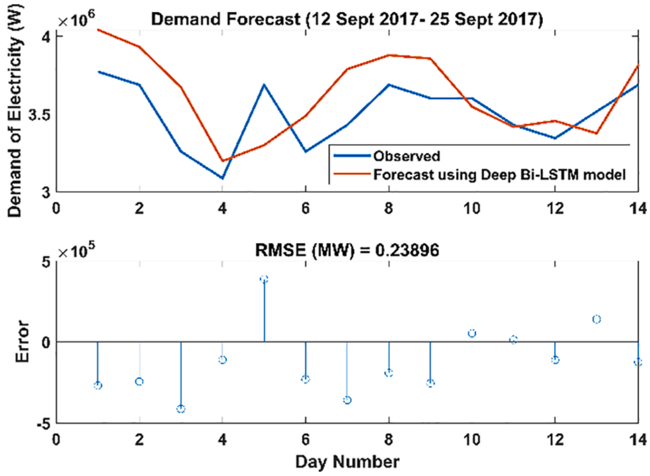

Fig. 8. Comparison of forecasted peak demand with observed peak demand using proposed deep Bi-LSTM S2S forecasting model for days from 12 Sept to 24 Sept 2017.

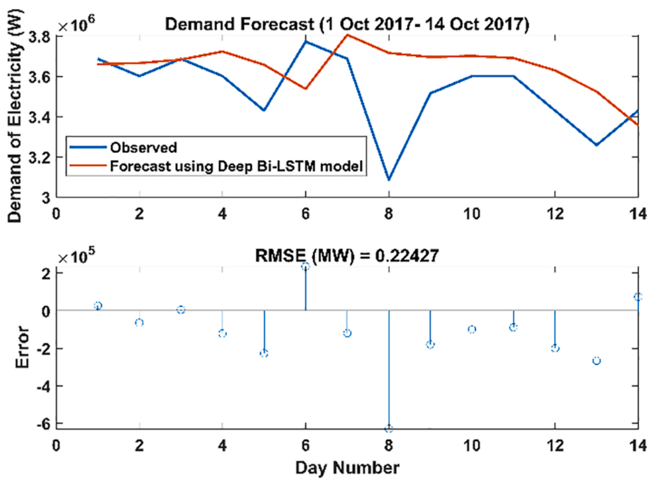

Fig. 9. Comparison of forecasted peak demand with observed peak demand using proposed deep Bi-LSTM S2S forecasting model for days from 1 Oct to 14 Oct 2017.

4. Discussion

This research work for the first time attempts to find out the effectiveness of deep Bi-LSTM S2S regression approach for forecasting day-ahead "peak" demand of both normal and "special" days using grossly limited data and has achieved model accuracy of up to 96.51% for special days as MAPE is only 3.49%. Figs. 11, 12, 13, and 14 compare the electricity demand predicted by proposed forecasting model and the comparison forecasting models with the actual electricity demand for six different weeks and public holidays. It can be seen that the proposed deep forecasting model is more closely following the actual electricity demand when compared with the other comparison models. Figs. 15 and 16 illustrate the performance comparison of the proposed deep Bi-LSTM S2S model with other comparison models in terms of both MAPE and RMSE.

It has been found out that in terms of RMSE, the proposed deep Bi-LSTM S2S model has outperformed MG-SVR by 68%, LMBP-ANN model by 64%, Shallow LSTM model by 70%, Deep LSTM model by 78%, and Shallow Bi-LSTM model by 75%, in the case of special days. For other days, the proposed model performed equal to MG-SVM model, outperformed LMBP-ANN model by 8.3%, shallow LSTM model by 12%, Deep LSTM model by 40.5%, and Shallow Bi-LSTM model by 12%, in terms of RMSE.

Moreover, in terms of performance metric MAPE, Bi-LSTM S2S model improved forecasting by 73% from MG-SVM, 68% from LMBP-ANN model, by 78% from shallow LSTM model, by 82% from Deep LSTM model, and by 81% from Shallow Bi-LSTM model, in case of special days. The performance for other days based on MAPE performance metric is also promising with the proposed model being better by 0.8% from LMBP-ANN, by 40% from shallow LSTM model, by 6% from Deep LSTM model, and by 81% from Shallow Bi-LSTM model. Also, in this case, the proposed model has shown comparable performance to MG-SVM model with it being better by only 2%.

Furthermore, Wilcoxon signed rank test (Woolson, 2007) is used for testing the statistical significance of the difference between the absolute errors of the proposed model and the comparison models. This test is chosen because the test data for the proposed and comparison models is same (paired) and is not normally distributed (nonparametric). Table 3 shows the p values obtained from conducting the Wilcoxon signed rank test.

Statistically speaking, in case of special days, the proposed deep Bi-LSTM S2S forecast model outperformed MG-SVM at 5% significance level as the p value is 2.96%. During the same days, it also outperformed LMBP-ANN, deep LSTM, and shallow Bi-LSTM models at 10% significance level as the p-values are 10.4%, 7.1%, and 7.1%, respectively. However, the improvement of proposed deep Bi-LSTM S2S forecast model over shallow LSTM is not statistically significant at 10% level as the p-value is 14.7%. These results indicate that that the shallow LSTM is indeed a strong competitor in case of special days, statistically. For normal days, the proposed deep Bi-LSTM S2S forecast model outperformed Deep LSTM at 5% significance level as the p value is only 0.54%. Nevertheless, the improvement of proposed deep Bi-LSTM S2S forecast model over other comparison models is not statistically significant during normal days at 5 or 10% levels. Thus, these comparison models can be considered as a formidable competitor in case of normal days, statistically.

Based on the above discussion, it can be concluded that the proposed deep Bi-LSTM S2S regression forecasting model proved to be more accurate and flexible for forecasting the non-linear, volatile, and non-stationary day-ahead "peak" electricity demand that must be supplied by FESCO for managing demand response programs. The proposed model has also effectively addressed the challenge of accurately forecasting load on "special" days. The good performance has been achieved even with less data. Predicting load demand on special days is a challenge because there are a few numbers of holidays in a whole year, therefore, there is always less data available to accurately train an artificial intelligence algorithm. Further, the holidays can either occur on a fixed date each year or can vary from year to year. However, It can be seen that all the other models have performed poorly when predicting load on holidays. The proposed deep model has achieved accuracy of 96.51% on holidays, whereas all the others models have shown accuracy between 80% and 89%.

There are two reasons to why the proposed deep Bi-LSTM S2S model has shown good performance on special days. Firstly, it is due to the sequence to sequence architecture of the Bi-LSTM layer, also known as encoder-decoder (Brownlee, 2017). The sequence to sequence architecture allows Bi-LSTM to learn using the hidden state of LSTM unit at every time step of the input sequence **X**. In other words, the deep Bi-LSTM model learned the temporal relationship between the peak demand on holidays and every predictor variable at every point of the input sequence. This is in contrast to sequence to one regression problems where the LSTM unit hidden state is only used after the whole sequence has been run through it. Secondly, for every input sequence, each Bi-LSTM layer uses the input sequence as it is for training the first LSTM layer and then it reverses the input sequence to train the next second LSTM layer. In our proposed model, there are two Bi-LSTM layers and thus four LSTM layers that are learning information from running

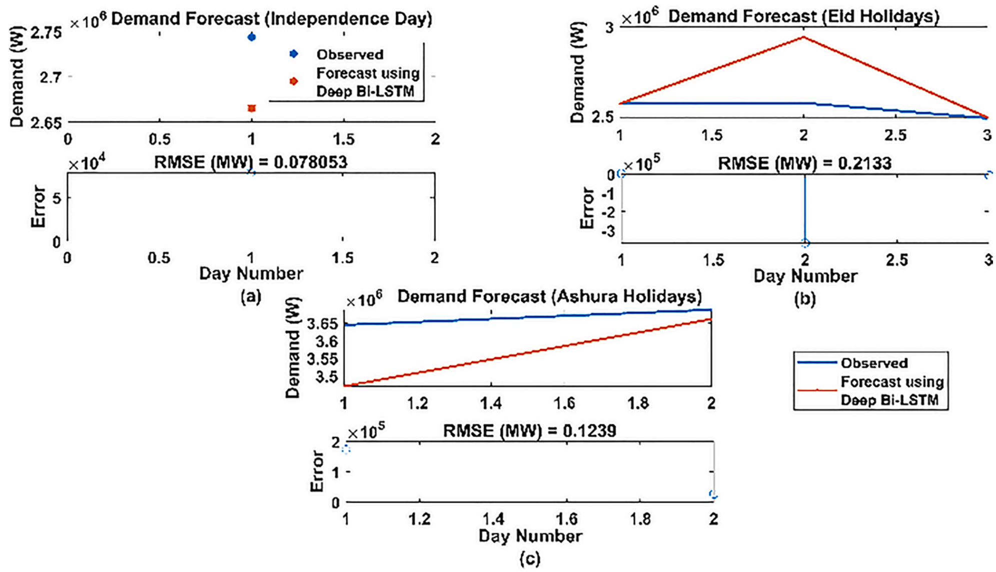

Fig. 10. Comparison of forecasted peak demand with observed peak demand using proposed deep Bi-LSTM S2S forecasting model for (a) Independence Day (b) Eid Holidays, and (c) Ashura Holidays.

Table 1
Deep Bi-LSTM S2S Model Performance (MAPE (%) and RMSE (MW)).

	MAPE (%) Normal Days	Special Days	RMSE (MW) Normal Days	Special Days
August	3.67	2.84	0.21	0.08
September	5.92	4.90	0.24	0.21
October	4.93	2.73	0.22	0.12
Average	4.84	3.49	0.22	0.14

Table 2
Forecasting Models for Comparison Performance (MAPE (%) and RMSE (MW)).

	MAPE (%) Normal Days	Special Days	RMSE (MW) Normal Days	Special Days
MG-SVM	4.74	13.34	0.22	0.44
LMBP-ANN	4.88	11.01	0.24	0.39
Shallow LSTM	5.42	15.83	0.25	0.47
Deep LSTM	8.03	19.78	0.37	0.65
Shallow Bi-LSTM	5.15	18.63	0.25	0.56

Table 3
P values of Wilcoxon signed rank test.

	Wilcoxon signed rank test(p value) Normal Days	Special Days
Deep Bi-LSTM vs MG-SVM	0.3821	0.0296
Deep Bi-LSTM vs LMBP-ANN	0.3085	0.1042
Deep Bi-LSTM vs Shallow LSTM	0.4305	0.1473
Deep Bi-LSTM vs Deep LSTM	0.0054	0.0711
Deep Bi-LSTM vs Shallow Bi-LSTM	0.1356	0.0711

the input backwards and forwards. This allows the deep Bi-LSTM model to fully learn the temporal dynamic behavior of peak load on holidays and with low computational complexity.

5. Conclusion

In this research work, an improved method for forecasting day-ahead non-linear, volatile, and non-stationary "peak" electricity demand on both normal days and special days using grossly limited data has been presented. In addition, it addresses the issues of managing demand

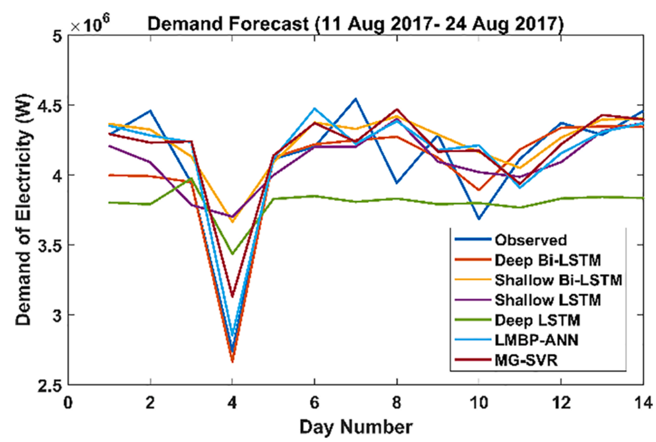

Fig. 11. Comparison of proposed and comparison models with actual peak demand for days from 11 Aug to 24 Aug 2017.

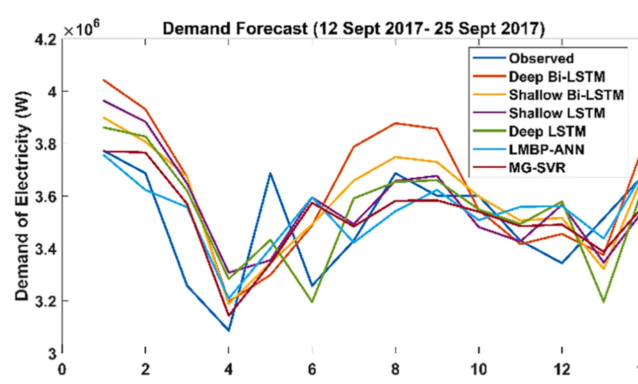

Fig. 12. Comparison of proposed and comparison models with actual peak demand for days from 12 Sept to 25 Sept 2017.

response programs and forecasting using limited data in developing countries such as Pakistan. Models were trained, validated and tested using the real daily peak demand data obtained from local utility

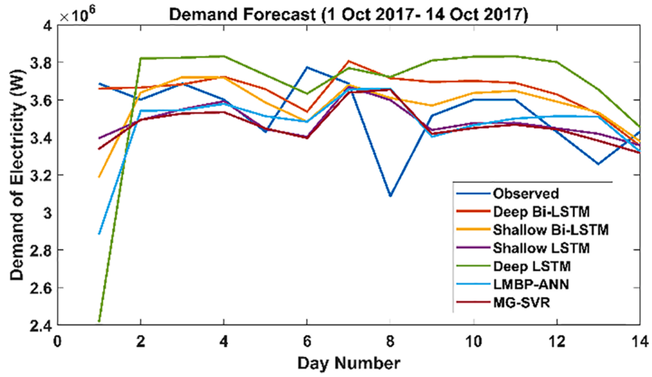

Fig. 13. Comparison of proposed and comparison models with actual peak demand for days from 1 Oct to 14 Oct 2017.

company FESCO. The research work has proposed deep Bi-LSTM sequence to sequence regression approach for STPDF for the first time. It has been successfully implemented in the MATLAB environment and has proved its effectiveness. It has always been a challenge to accurately forecast load on special days in contrast to normal days and that challenge has been effectively addressed by the proposed deep network.

Peak electrical demand is also dependent upon both the immediate past and future values of the prediction parameters. The already existing learning methods are not considering both values. However, the proposed deep Bi-LSTM network has learned from both the future and past values of prediction input parameters. This is also one of the reasons of why it has shown outclass performance in terms of special days, the other reasons have already been discussed in detail in section 4. Furthermore, this research will help FESCO in the successful deployment of demand response programs and the implementation of effective electricity operations. In this way, the proposed forecasting model has addressed the major electricity supply–demand imbalance problem in Pakistan and other developing countries. Due to unavailability of

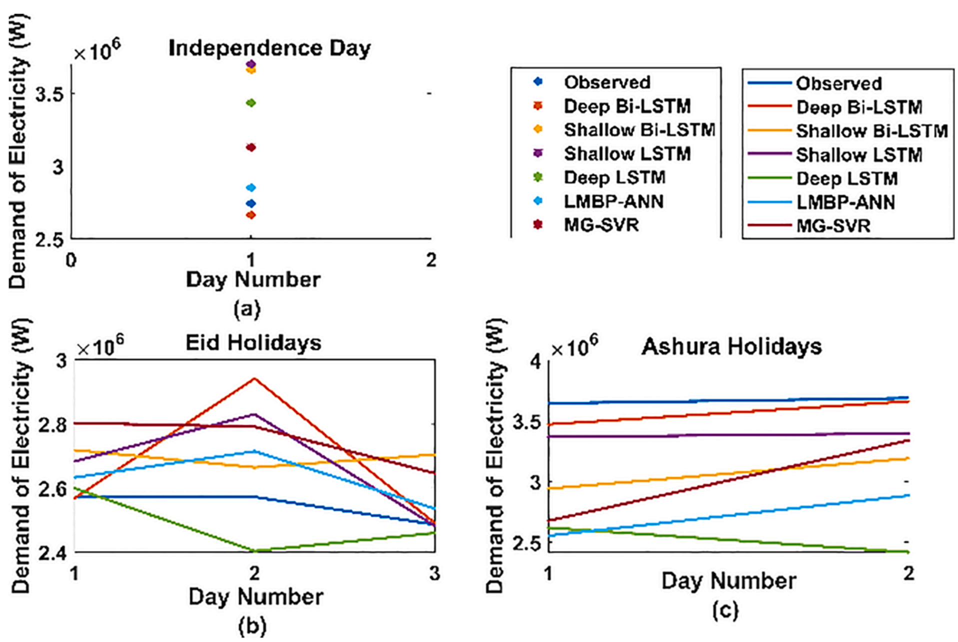

Fig. 14. Comparison of proposed and comparison models with actual peak demand for (a) Independence Day (b) Eid Holidays, and (c) Ashura Holidays.

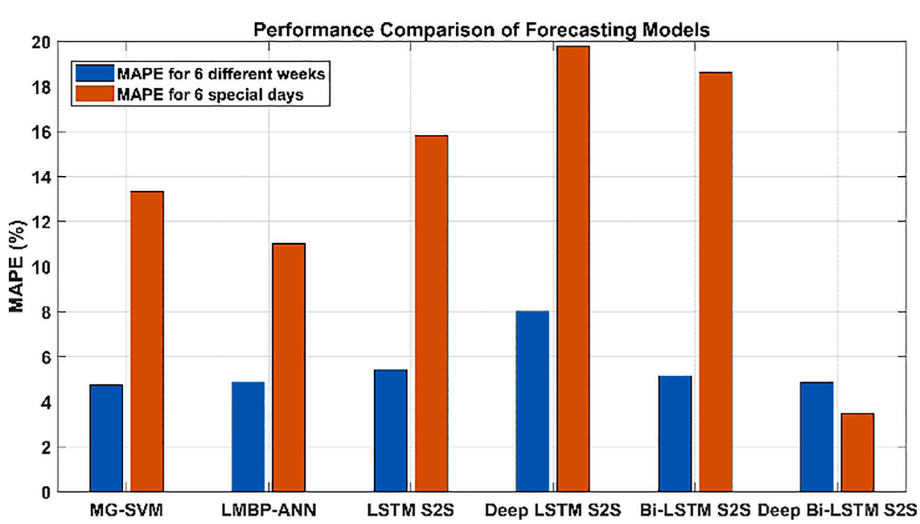

Fig. 15. Performance Comparison of Forecasting Models in terms of MAPE.

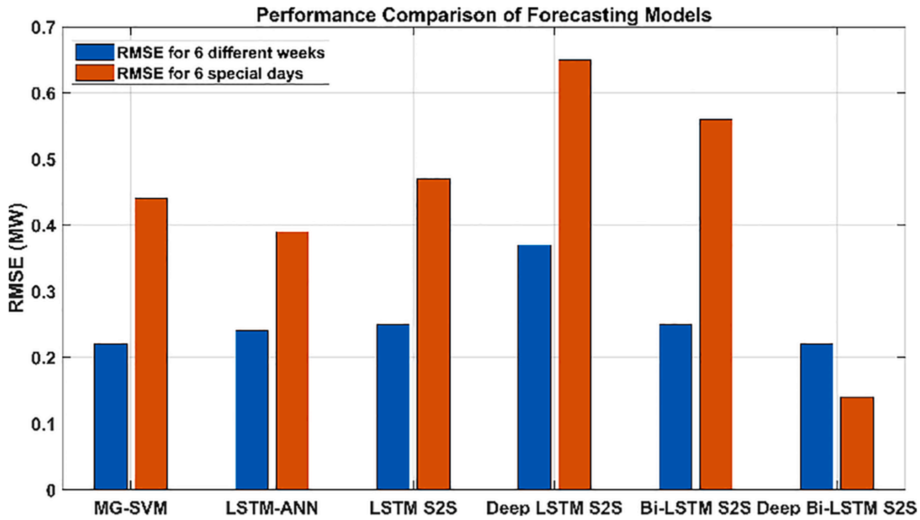

Fig. 16. Performance Comparison of Forecasting Models in terms of RMSE.

sophisticated equipment and smart meters, obtained electricity demand readings were manually recorded and were not computerized which introduced typographical and human error. Therefore, this research work has also treated the outliers and missing values in the historical load data.

Future work could focus on developing short term peak demand forecasting model at substation level, distribution level and for entire grid system of Pakistan. The training environment used in this research was CPU-based and although the computing time in each case was a few seconds researchers can try and use a GPU-computing environment which would make training the networks superfast. Furthermore, experts can also devise new methods for automatic tuning of hyper-parameters of the proposed deep Bi-LSTM S2S model, or make a performance comparison between already available popular methods. For instance, evolutionary algorithms have started to be widely used for such purpose (Baioletti, Milani, & Santucci, 2018; ElSaid, Wild, Jamiy, Higgins, & Desell, 2017).

CRediT authorship contribution statement

Neelam Mughees: Conceptualization, Methodology, Software, Data curation, Validation, Writing - original draft, Formal analysis. **Syed Ali Mohsin:** Supervision, Project administration. **Abdullah Mughees:** Visualization, Investigation. **Anam Mughees:** Validation, Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are thankful to Faisalabad Electric Supply Company (FESCO) for providing the required data and National University of Computer and Emerging Sciences, Chiniot-Faisalabad Campus for providing a platform for this research opportunity.

References

Albadi, M. H., & El-Saadany, E. F. (2007). Demand response in electricity markets: An overview. *IEEE Power Engineering Society General Meeting, 2007*, 1–5.

Almalaq, A., & Edwards, G. (2017). A review of deep learning methods applied on load forecasting. In *Machine Learning and Applications (ICMLA), 2017 16th IEEE International Conference On* (pp. 511–516).

Aquino, G., Rubio, J. D. J., Pacheco, J., Gutierrez, G. J., Ochoa, G., Balcazar, R., … Zacarias, A. (2020). Novel nonlinear hypothesis for the delta parallel robot modeling. *IEEE Access, 8*, 46324–46334.

Atef, S., & Eltawil, A. B. (2020). Assessment of stacked unidirectional and bidirectional long short-term memory networks for electricity load forecasting. *Electric Power Systems Research, 187*, Article 106489.

Baioletti, M., Milani, A., & Santucci, V. (2018). In *Learning bayesian networks with algebraic differential evolution* (pp. 436–448). Springer.

Bouktif, S., Fiaz, A., Ouni, A., & Serhani, M. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. *Energies, 11*(7), 1636.

Brownlee, J. (2017). Encoder-Decoder Long Short-term Memory Networks.

Chan, S. C., Tsui, K. M., Wu, H. C., Hou, Y., Wu, Y.-C., & Wu, F. (2012). Load/price forecasting and managing demand response for smart grids: Methodologies and challenges. *IEEE Signal Processing Magazine, 29*(5), 68–85.

Chiang, H.-S., Chen, M.-Y., & Huang, Y.-J. (2019). Wavelet-based EEG processing for epilepsy detection using fuzzy entropy and associative petri net. *IEEE Access, 7*, 103255–103262.

Cui, Z., Ke, R., Pu, Z., & Wang, Y. (2020). Stacked bidirectional and unidirectional LSTM recurrent neural network for forecasting network-wide traffic state with missing values. arXiv preprint arXiv. 1162.

de Jesus Rubio, J. (2009). SOFMLS: online self-organizing fuzzy modified least-squares network. *IEEE Transactions on Fuzzy Systems, 17*(6), 1296–1309.

de Rubio, J. J. J. I. T. o. N. N., & Systems, L. (2020). Stability analysis of the modified Levenberg-Marquardt algorithm for the artificial neural network training.

Di Persio, L., & Honchar, O. (2017). Analysis of recurrent neural networks for short-term energy load forecasting. *AIP Conference Proceedings, 1906*, Article 190006.

Egrioglu, E., Aladag, C. H., & Yolcu, U. (2013). Fuzzy time series forecasting with a novel hybrid approach combining fuzzy c-means and neural networks. *Expert Systems with Applications, 40*(3), 854–857.

ElSaid, A., Wild, B., Jamiy, F. E., Higgins, J., & Desell, T. (2017). Optimizing LSTM RNNs using ACO to predict turbine engine vibration. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion* (pp. 21–22).

Graves, A., Jaitly, N., & Mohamed, A. (2013). Hybrid speech recognition with deep bidirectional LSTM. *IEEE Workshop on Automatic Speech Recognition and Understanding, 2013*, 273–278.

Hagan, M. T., & Behr, S. M. (1987). The time series approach to short term load forecasting. *IEEE Transactions on Power Systems, 2*(3), 785–791.

He, W. (2017). Load forecasting via deep neural networks. *Procedia Computer Science, 122*, 308–314.

He, F., Zhou, J., Feng, Z., Liu, G., & Yang, Y. (2019). A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. *Applied Energy, 237*, 103–116.

Hernandez, L., Baladron, C., Aguiar, J. M., Carro, B., Sanchez-Esguevillas, A. J., Lloret, J., & Massana, J. (2014). A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings. *IEEE Communications Surveys & Tutorials, 16*(3), 1460–1495.

Hernández, G., Zamora, E., Sossa, H., Téllez, G., & Furlán, F. J. N. (2020). Hybrid neural networks for big data classification. *Neurocomputing, 390*, 327–340.

Hochreiter, S., & Schmidhuber, Jürgen (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.

Hong, C. W., Ko, M. S., Kim, H. R., Kim, S., & Hur, K. (2020). Short-term power load forecasting of a large vessel using deep stacking network architecture. *Transactions of the Korean Institute of Electrical Engineers, 69*(4), 534–541.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences, 79*(8), 2554–2558.

Kim, T.-Y., & Cho, S.-B. (2019). Predicting residential energy consumption using CNN-LSTM neural networks. *Energy, 182*, 72–81.

Kim, M., Choi, W., Jeon, Y., & Liu, L. (2019). A hybrid neural network model for power demand forecasting. *Energies, 12*(5), 931. https://doi.org/10.3390/en12050931

Kim, S., Choi, Y., & Lee, M. (2015). Deep learning with support vector data description. *Neurocomputing, 165*, 111–117.

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2017). Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid*.

Mandic, D., & Chambers, J. (2001). *Recurrent neural networks for prediction: Learning algorithms, architectures and stability*. Wiley.

Marino, D. L., Amarasinghe, K., & Manic, M. (2016). Building energy load forecasting using deep neural networks. In *Industrial Electronics Society, IECON 2016–42nd Annual Conference of the IEEE* (pp. 7046–7051).

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics, 5*(4), 115–133.

Meda-Campaña, J. A. (2018). On the estimation and control of nonlinear systems with parametric uncertainties and noisy outputs. *IEEE Access, 6*, 31968–31973.

Mori, H., & Kobayashi, H. (1995). Optimal fuzzy inference for short-term load forecasting. Power Industry Computer Application Conference, 1995. Conference Proceedings, 1995 IEEE, 312–318.

N. Mughees FESCO Daily Peak Load Data. https:// 2020 doi.org/10.17632/jh8t8hxf9b.1.

Muncharaz, J. O. (2020). Comparing classic time series models and the LSTM recurrent neural network: An application to S&P 500 stocks. Finance, Markets and Valuation, 6.

Qdr, Q. (2006). Benefits of demand response in electricity markets and recommendations for achieving them. US Dept. Energy, Washington, DC, USA, Tech. Rep.

Quilumba, F. L., Lee, W.-J., Huang, H., Wang, D. Y., & Szabados, R. L. (2014). Using smart meter data to improve the accuracy of intraday load forecasting considering customer behavior similarities. *IEEE Transactions on Smart Grid, 6*(2), 911–918.

Shi, H., Xu, M., & Li, R. (2018). Deep learning for household load forecasting A novel pooling deep RNN. *IEEE Transactions on Smart Grid, 9*(5), 5271–5280.

Sioshansi, F., & Vojdani, A. (2001). What could possibly be better than real-time pricing? Demand response. *The Electricity Journal, 14*(5), 39–50. https://doi.org/10.1016/S1040-6190(01)00207-X.

Somu, N., M R, G. R., & Ramamritham, K. (2020). A hybrid model for building energy consumption forecasting using long short term memory networks. *Applied Energy, 261*, 114131. https://doi.org/10.1016/j.apenergy.2019.114131

Srivastava, A. K., Pandey, A. S., & Singh, D. (2016). Notice of violation of IEEE publication principles: Short-term load forecasting methods: A review. *International Conference on Emerging Trends in Electrical Electronics & Sustainable Energy Systems (ICETEESES), 2016*, 130–138.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 3104–3112.

Tang, X., Dai, Y., Wang, T., & Chen, Y. (2019). Short-term power load forecasting based on multi-layer bidirectional recurrent neural network. *Transmission & Distribution, 13*(17), 3847–3854.

Wang, Y., Gan, D., Sun, M., Zhang, N., Lu, Z., & Kang, C. (2019). Probabilistic individual load forecasting using pinball loss guided LSTM. *Applied Energy, 235*, 10–20.

Wang, S., Wang, X., Wang, S., & Wang, D. (2019). Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting. *International Journal of Electrical Power & Energy Systems, 109*, 470–479.

Woolson, R. J. W. e. o. c. t. (2007). Wilcoxon signed-rank test. 1-3.

Wu, L., Kong, C., Hao, X., & Chen, W. J. M. P. i. E. (2020). A short-term load forecasting method based on GRU-CNN hybrid neural network model.

Yu, Z., Niu, Z., Tang, W., & Wu, Q. (2019). Deep learning for daily peak load forecasting a novel gated recurrent neural network combining dynamic time warping. *IEEE Access, 7*, 17184–17194.

Zazo, R., Sankar Nidadavolu, P., Chen, N., Gonzalez-Rodriguez, J., & Dehak, N. (2018). Age estimation in short speech utterances based on LSTM recurrent neural networks. *IEEE Access, 6*, 22524–22530.

Zen, H., & Sak, H. (2015). Unidirectional long short-term memory recurrent neural network with recurrent output layer for low-latency speech synthesis. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4470–4474).

Zheng, H., Yuan, J., & Chen, L. (2017). Short-term load forecasting using EMD-LSTM neural networks with a Xgboost algorithm for feature importance evaluation. *Energies, 10*(8), 1168.

Ziel, F., & Energy, C. (2018). Modeling public holidays in load forecasting: A German case study. *Journal of Modern Power Systems and Clean Energy, 6*(2), 191–207.

Neelam Mughees received the B.S degree from University of Engineering and Technology, Lahore, Pakistan, MS degree from National University of Computer and Emerging Sciences, Islamabad, Pakistan, and is currently pursuing the Ph.D degree in COMSATS University Islamabad, Pakistan, all in electrical engineering. Currently, she is working as a Lecturer of Electrical Engineering in National Textile University, Faisalabad, Pakistan. She is particularly interested in the field of Power Systems, Smart Grids, and applications of Machine Learning.

Syed Ali Mohsin He did B.S in Electrical Engineering from University of Engineering and Technology, Lahore, Pakistan in 1986. He got M.S. degree from The University of Texas, Austin, Texas, U. S. A. In 2008, he received his Ph.D degree from University of Engineering and Technology, Lahore, Pakistan (Split PhD Research Work at Purdue University, West Lafayette, Indiana, USA). He has been teaching in various institutions in and out of the country for over 25 years now. Currently, he is a Professor of Electrical Engineering at National University of Computer and Emerging Sciences, Islamabad, Chiniot-Fsd Campus, Pakistan. His research interests include Electromagnetics, Power Systems, and Microwaves. He is member of IEEE and various IEEE societies.

Abdullah Mughees He received his B.S degree from Govt. College University, Faisalabad, Pakistan and M.S degree in Electrical Engineering from National University of Computer and Emerging Sciences, Islamabad, Pakistan. Currently, he is doing PhD in Electrical Engineering from National University of Science and Technology, Islamabad. His research areas include Power Systems, Control Systems and Automation.

Anam Mughees She received the B.S degree and MS degree from University of Engineering and Technology, Lahore, Pakistan, and is currently pursuing the Ph.D degree from the same university, all in electrical engineering. Currently, she is working as a lab engineer in Electrical Engineering department in GC University Faisalabad, Pakistan. She is particularly interested in the field of Power Systems, Smart Grids, and applications of Machine Learning.