IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received July 26, 2018, accepted August 15, 2018, date of publication August 29, 2018, date of current version September 21, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2867681

Empirical Mode Decomposition Based Deep Learning for Electricity Demand Forecasting

JATIN BEDI AND DURGA TOSHNIWAL, (Member, IEEE)
Department of CSE, IIT Roorkee, Roorkee 247667, India

Corresponding author: Jatin Bedi (jbedi@cs.iitr.ac.in)

This work was supported by the Ministry of Human Resource Development (MHRD), Government of India.

ABSTRACT Electricity is of great significance for national economic, social, and technological activities, such as material production, healthcare, and education. The nationwide electricity demand has grown rapidly over the past few decades. Therefore, efficient electricity demand estimation and management are required for better strategies planning, energy utilization, waste management, improving revenue, and maintenance of power systems. In this paper, we propose an empirical mode decomposition (EMD)-based deep learning approach which combines the EMD method with the long short-term memory network model to estimate electricity demand for the given season, day, and time interval of a day. For this purpose, the EMD algorithm decomposes a load time series signal into several intrinsic mode functions (IMFs) and residual. Then, a LSTM model is trained separately for each of the extracted IMFs and residual. Finally, the prediction results of all IMFs are combined by summation to determine an aggregated output for electricity demand. To demonstrate the applicability of the proposed approach, it is applied to electricity consumption data of city Chandigarh. Furthermore, the performance of the proposed approach is evaluated by comparing the prediction results with recurrent neural network (RNN), LSTM, and EMD-based RNN (EMD+RNN) models.

INDEX TERMS Deep learning, electricity demand prediction, empirical mode decomposition, energy analytic, long short term memory network.

I. INTRODUCTION

Electricity plays a significant role in the social, economic and technological development of a country [1]. In the past few decades, there is a tremendous rise in the total electricity demand of country India [2]. So, there is an emerging need for making reliable and continuous improvements in the energy analytic systems. Energy analytic systems are responsible for the proper operation, scheduling, management of future requirements, utilization monitoring and electricity demand forecasting. One significant characteristic of energy analytic systems is the ability to predict electricity demand over a wide range of time periods. It is of great help for agents who develop electricity buying and selling activities, energy suppliers and for electricity network managers who predict future demand. Based on time horizon of prediction, the demand forecasting can be broadly classified into three different classes [3] namely short-term electricity demand forecasting, medium-term electricity demand forecasting and long-term electricity demand forecasting.

- ***Long Term Demand Forecasting***: has a forecasting range of 1 to 50 years. It plays a crucial in future strategies planning, power system installation planning, and construction of new generation units.
- ***Medium Term Demand Forecasting***: has a forecasting range of 1 to 12 months. It is very beneficial for better planning of future requirements and strategies [4].
- ***Short Term Demand Forecasting***: has a forecasting range of hours, days or week ahead. It is very useful for efficient handling of daily operations, generation capacity scheduling, purchase plans and evaluations [4].

One of the significant features of electricity is that it is difficult to be stored once it has been generated. Therefore, efficient systems are required to accurately forecast the electricity load. Both overestimation and underestimation of electricity load can cause serious issues [5]. The overestimation of electricity load could lead to high operational cost, waste of available resources, electricity market fluctuation and high distribution losses. Similarly, underestimation of demand can lead to financial losses and unmet demands.

In the past few years, a number of research studies are done for optimal forecasting of electricity demand. Several machine learning algorithms such as Support Vector Machine (SVM) [6], Decision Trees [7], Artificial Neural Network (ANN) [8], Recurrent Neural Network (RNN) [9] have been used by the researchers for electricity demand forecasting. Out of these algorithms, neural network were found as promising tools with a satisfactory level of accuracy. However, they suffer from various issues including slow learning rate, over-fitting, deciding optimal hyper-parameters value etc. Furthermore, in the recent years, various hybrid models have also gained a lot of interest to improve overall prediction accuracy. But there is still a need for more robust solutions to analyze and forecast the electricity load patterns effectively.

In this paper, we propose a hybrid model using Empirical Mode Decomposition (EMD) and Long Short Term Memory (LSTM) network to improve the prediction accuracy of electricity demand forecasting models. EMD is a data denoising technique that has been used in different fields including medical [10], traffic congestion prediction [11], tourism management and industry [12]. The outline of the paper is as follows: Section II explains the related work done in the field of energy demand forecasting. Section III provides a description of deep learning concepts used in this study. The methodology of proposed EMD based deep learning approach (EMD+LSTM) is explained in section IV and conclusion is stated in section V.

II. RELATED WORK

Electricity data is defined as the time series data with demand observations at equal time intervals [13]. Time series prediction aims to determine the future event based on a series of given historical events. Electricity load prediction is a hard non-linear problem that can be solved only by complex formulations. In literature, various tools and techniques were developed to solve these complex nonlinear problems. Some of them are as follows:

In general, electricity demand forecasting models can be divided into three major classes: physical or data-driven models, statistical or machine learning methods and hybrid methods. Physical models [14] were used for modeling the thermal behavior/equations between variables. Some of them include natural ventilation, passive solar, climate and air conditioning system. These models are typically influenced by various factors such as temperature, HVAC components and type of material [15].

In contrast to physical models, statistical methods do not require any other prior physical information. These methods work by building a prediction model based on some learning algorithm. The process of building a prediction model consists of two phases: training phase and testing phase. During the training phase, learning algorithm builds a model by delineating the hidden relationships present in the data. Subsequently, in the testing phase, the trained model is used for predicting the future demand. Initially, simple linear regression and multivariate regression were used for the task of prediction. In the year 1980, Parti and Parti [16] proposed a system to predict buildings energy consumption using linear regression. Aydinalp-Koksal and Ugursal [17] presented a regression based approach to estimate energy consumption of residentials in Canada. Later in the year 2012, Aranda *et al.* [18] developed a regression model to forecast load demand of the banking sector in Spain.

ANN gained a lot of popularity and have been widely used for the task of prediction in different domains since the year 2000. Kalogirou [19] presented a review describing the application of ANN in energy analytics. Escriva *et al.* [20] introduced a multi-layer neural network model with three layers to predict electricity demand in University of Valencia, Spain. In the year 2016, Biswas *et al.* [21] implemented a neural network approach to predict load for the residential houses. Several input variables including outdoor temperature, solar radiation were used to model the complex relationship. The prediction results of the proposed neural network approach were compared with other existing techniques and found promising. ANN have good approximation capabilities but there are certain problems associated with them such as over-fitting, high training time. These problems are rectified by the SVM models. Various research studies have motivated the use of SVM as a prediction tool by comparing the results of SVM with ANN and other regression models [14], [22].

In the recent years, various hybrid and deep neural network models were also introduced to optimally forecast electricity demand. Gunay [23] predicted the gross electricity demand of Turkey by combined application of ANN and multivariate regression model. The multi-linear regression model was used to decide the significant variables from the initially given set of six variables. These statistically significant variables were then fed as input to the ANN for estimating the annual electricity demand. On the basis of prediction results, the author concluded that the approach could be easily used to make an accurate prediction for electricity demand. Rahman *et al.* [24] presented an approach for electricity demand prediction of commercial and residential buildings using Recurrent Neural Network (RNN). This approach was used to predict hourly consumption of safety building in Utah and residential buildings in Texas. Yaslan and Bican [25] combined EMD with SVM to improve the prediction accuracy of SVM time series prediction model.

Each type of forecasting models has its own pros and cons. Physical methods are very simple to use but have low generalization capability and high computational complexity. Statistical methods lack at handling the nonlinear hidden features of data. Neural network models (ANN, MLP and RNN) have powers to handle the nonlinear complex relationships but can be used only for handling short-term data dependencies. In the present work, we propose a deep learning based hybrid model to forecast electricity demand of Chandigarh, India. The approach combines EMD algorithm with LSTM network to improve the prediction accuracy of existing time series prediction models. This paper addresses the following major things:

- How to handle complex nonlinear relationships and long term historical dependencies present in the data.
- An efficient way to combine EMD with Deep learning algorithm (LSTM) for noise-free data training.

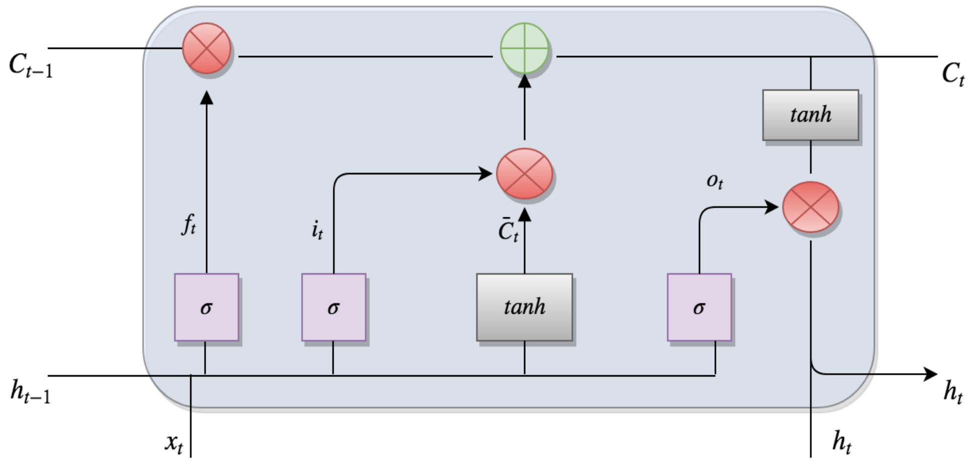


FIGURE 1. Architecture of the LSTM memory block.

- An efficient way to provide support for dynamic learning with the help of Deep LSTM network models.

III. THEORETICAL BACKGROUND

A. RECURRENT NEURAL NETWORK AND LONG SHORT TERM MEMORY NETWORK

Recurrent Neural Networks (RNN) [26] have become very popular for modeling complex time series prediction problems as they have the tendency to retain knowledge acquired through subsequent timestamps. This preserved knowledge play a crucial role in the accurate prediction of time series events. The information is retained in RNN with the help of network loops, where a node uses the previous timestamp output and current input to determine the current output. Even though RNN network have good approximation capabilities, they are not fit for handling long-term dependencies of data [26]. This is because of the exponential decay in error gradient as it propagates through the network (vanishing gradient problem) [26], [27].

LSTM network [28] introduced by Hochreiter and Schmidedhuber overcome the issues with RNN network. The problem of vanishing gradient are solved by replacing nodes in the RNN with memory cells and gating mechanism [29]. Figure 1 shows the basic architecture of a LSTM memory cell. Depending on the input, LSTM memory cell can remember or forget cell state, scale input. The overall support in a cell is provided by three gates: Input, forget and output. Given an input x_t to cell and previous timestamp output h_{t-1}. The forget gate f_t gate determines the input for cell state C_{t-1} using sigmoid function and it is given by:

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \tag{1}$$

Input gate i_t, determines the values to be updated to C_t using Equation 2

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \tag{2}$$

Output gate o_t, regulates output values of cell and is given by Equation 3 where $\bar{C}_t$ denotes the output of nonlinear tanh layer.

$$\begin{aligned} C_t &= f_i \odot C_{t-1} \oplus i_t \odot \bar{C}_{t-1} \\ o_t &= \sigma(W_o[h_{t-1}, x_t] + b_0) \end{aligned} \tag{3}$$

Finally, the output value of LSTM memory cell is given by

$$h_t = o_t \odot C_t \tag{4}$$

The set of weights (W_f, W_b, W_o) & biases (b_f, b_i, b_o) in the LSTM network model are determined by the Back-propagation algorithm.

B. EMPIRICAL MODE DECOMPOSITION (EMD)

EMD [30], [31] is a nonlinear analysis approach for non-stationary time series data. It is an unsupervised data-driven decomposition that does not require any prior system. EMD algorithm works by decomposing a non-stationary time series

Algorithm 1 EMD (Empirical Mode Decomposition)

1: *Given a time series signal S(t), identify local maxima and minima.*
2: *Calculate upper $S_u[t]$ and lower $S_l[t]$ envelope by interpolation of local maxima and minima.*
3: *Computer the mean of upper and lower envelopes*

$$m_t = \frac{S_u[t] + S_l[t]}{2} \tag{5}$$

4: *Subtract mean from time series signal $h(t) = S(t) - m(t)$.*
5: *Repeat Step 2 to 4 until one of the stopping criteria is reached:*
- *S(t) reaches zero.*
- *Both Constraints are satisfied.*
- *Max. number of iterations reached.*

6: *Treat h(t) as new IMF and calculate the residual signal r(t) as: $r(t) = S(t) - h(t)$.*
7: *Use r(t) as new S(t) and repeat step 1 to 6, until all IMFs are obtained.*

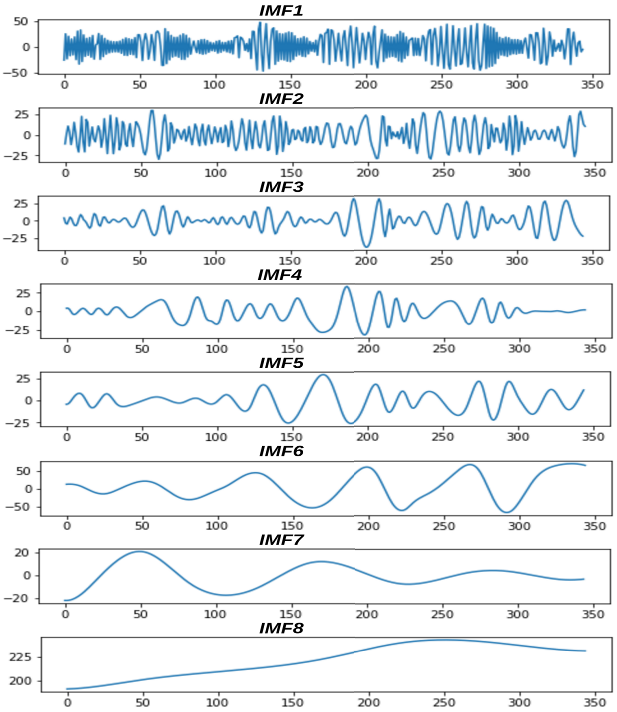


FIGURE 2. Example of IMFs obtained after applying EMD over electricity data of a season.

signal into a set of Intrinsic Mode Functions (IMFs) along with a residue. Figure 2 shows the example of IMFs obtained by applying EMD algorithm over a time series data signal. EMD of a signal can be achieved by means of a process called sifting. The sifting procedure requires two constrained to be satisfied:

- Each IMF must have the same number of zero crossings and extrema.
- Each IMF has symmetric envelopes.

The step by step procedure of EMD algorithm is given in Algorithm 1:

Now, the original time series signal can be expressed as:

$$S(t) = \sum_{i=1}^{n} h_i + r_n \tag{6}$$

There are two important features of EMD that make it appropriate for the task of time series analysis or prediction.

- Reconstruction property i.e. superimposition of all IMFs would lead to original time series signal without any kind of data loss.
- EMD algorithm is better at capturing trend for the non-stationary signals. Electricity load time series is a kind of non-stationary signal that consists of various components as it gets affected by a number of factors such as weather conditions, location, timing etc.

IV. METHODOLOGY

In this section, we explain the methodology of the proposed EMD based hybrid approach. Figure 3 shows the flowchart of the proposed approach *(EMD+LSTM)*.

A. INPUT DATA

The study uses the electricity consumption data of the city Chandigarh, India. Chandigarh is a Union Territory (UT) with a land area of 114 km square. It does not have any own power generation units and receives power from several

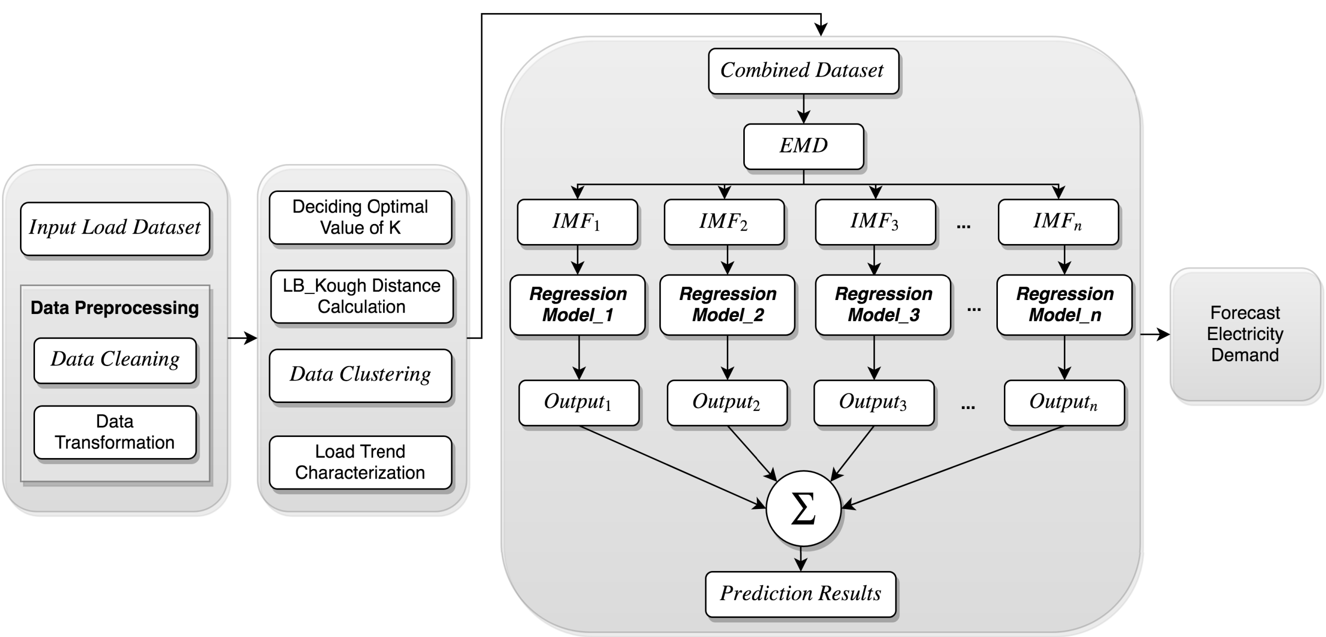


FIGURE 3. Flowchart of the proposed *(EMD+LSTM)* approach.

sources [32]: 67% of its power from PSEB Mohali, 10% from BBMB Dhulkote and 23 % of its power through Nalagarh. The average per capita electricity consumption in Chandigarh is 1168 units and it is growing at a rate of 52 million units per year.

The electricity consumption data of city Chandigarh is recorded every 15 minutes, which means there are 96 data points for one day. The dataset is available for a period of 5 years (from January 2013 to January 2018). Every 15 minutes interval (data point) in the dataset is assigned with a label starting from $Block_1$ for $00:00\,AM - 00:15\,AM$ to $Block_{96}$ for $11:45\,PM - 00:00\,AM$.

B. DATA PREPROCESSING

Electricity consumption data collected from the real-world sensors or smart meters is highly likely to contain different discrepancies including noise, outliers and missing values. These inconsistencies in the data necessitate the use of data preprocessing to make it appropriate for the task of data analysis. Data preprocessing [33], [34] includes various techniques such as data cleaning, aggregation, data transformation and data sampling. In this work, we use the following data preprocessing steps to improve the generalized performance of the proposed approach:

- ***Data Cleaning***: There are few (1 %) missing values in electricity consumption data of UT Chandigarh. These values are replaced by the mean electricity consumption value of that month.
- ***Data Aggregation***: Data aggregation [35] is carried out to combine consumption files with different formats into a single usable format.
- ***Data Transformation***: Data transformation [35] is the process of bringing out data to a particular scale. It plays a key role while computing similarity among time series data of different days, months and years. For the purpose of data transformation, we subtract the mean value of a time series from each timestamp value of that time series signal.

C. DATA CLUSTERING

Data clustering [34], [35] is the process of finding hidden patterns in the data. It belongs to the class of unsupervised learning algorithms which work by dividing the set of input data points among different groups based on some similarity measures. The data points are assigned to different groups in such a way that the data points belonging to same group should be highly similar to each other and dissimilar to the points belonging to other groups. Data clustering algorithms can be broadly divided into three classes namely Partitioning based clustering, Hierarchical clustering, and Density based clustering algorithms [36]. In this paper, we use K-means clustering algorithm to find out the groups of months in the electricity data that follows similar consumption patterns.

K-means is the partitioning based clustering algorithm [35] that works by grouping the time series data objects into classes based on a distance based similarity measure i.e. Euclidean distance. However, Euclidean distance is not a good similarity measure in case of time series data as it is not capable of dealing with distortions in time series signals. Dynamic Time Warping (DTW) [37] is a similarity measure that can handle distortions in the time domain by allowing elastic transformation of the time series. It is one of the efficient methods to calculate optimal match in time series data by warping in the time domain. Given two time series sequences A $< a_1, a_2, a_3 \ldots a_n >$ and B $< b_1, b_2, b_3 \ldots b_m >$, DTW algorithm initiates by computing a cost matrix $C \in R^{n*m}$ for the alignment of two sequences:

$$C \in R^{n*m} : c_{i,j} = ||a_i - b_j|| \quad i \in [1, n], j \in [1, m] \tag{7}$$

Then the algorithm tries to find an optimal path $P < p_1, p_2 \ldots . p_k >$ with $p_z = (p_x, p_y)$, $x \in [1, n]$, $y \in [1, m]$, $z \in [1, k]$ which passes through low cost areas of cost matrix C. The optimal path P must satisfy the following constraints:

- The start and end points of path P must be first and last points of aligned sequences A and B.
- The indices of sequences must be considered in monotonically increasing order.

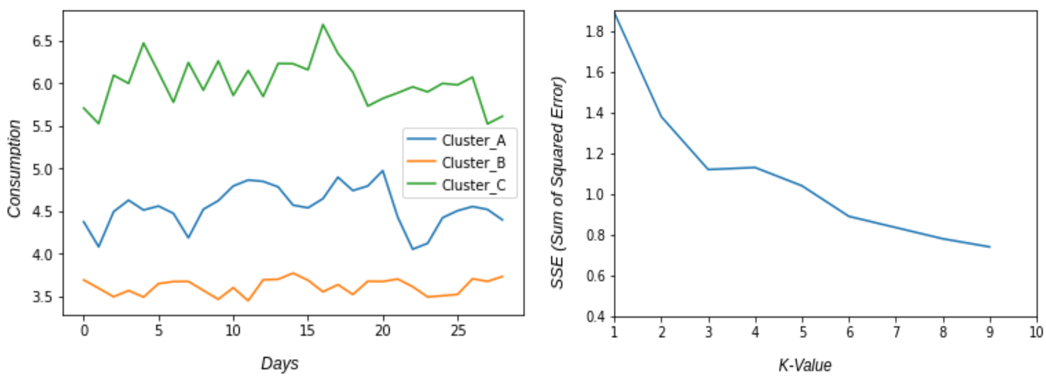
DTW distance algorithm provides an optimal match between time series sequences, but the high runtime complexity (quadratic) of the algorithm is a major problem. In this paper, we use a lower bound for DTW Distance measure i.e. *LB_Kough* distance measure [38] given by equation 8. Table 1 shows the LB_Kough distance metric value for time series data of all months in a year.

$$\begin{aligned} LB\_Kough(X, Y) &= \sum_{i=1}^{n} (C_i - U_i)^2 + (C_i - L_i)^2 \\ (UpperBound) : U_i &= max(q_{i-r} : q_{i+r}) \\ (LowerBound) : L_i &= min(q_{i-r} : q_{i+r}) \end{aligned} \tag{8}$$

Another major aspect with k-means clustering algorithm is to determine the value for user-defined parameter k (number of clusters). Several methods such as Gap statistics [39], Silhouette index [39] and Elbow method [39] can be used to determine optimal value of k. In the present study, we use Elbow method to determine the value for parameter k. In Elbow method [40], we generate a plot of k-value against SSE as shown in Figure 4b. The optimal k value belongs to the point where SSE changes drastically in the plot. From figure 4b, it can be stated that the optimal k value for electricity consumption dataset of Chandigarh is 3. Figure 4a shows the results of k-means clustering algorithm over consumption data of city Chandigarh. Furthermore, Figure 5 shows months of the year belonging to each of the three clusters such as Cluster_A represents *April*, *May*, *June*, Cluster_B represents *July*, *August*, *September* and Cluster_C represent *October*, *November*, *December*, *January*, *February*, *March*. These three clusters basically represent three seasons (*Spring*, *Autumn and Monsoon*) of a year.

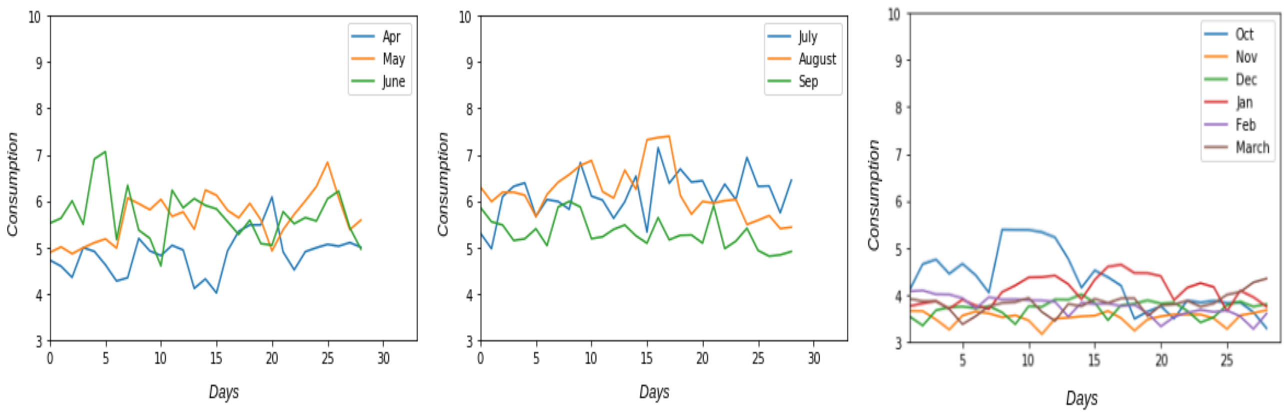
TABLE 1. ***LB_Kough*** **distance values between the time series data of months.**

LB_Kough	*January*	*February*	*March*	*April*	*May*	*June*	*July*	*August*	*September*	*October*	*November*	*December*
January	$--$	1.34	1.57	3.10	3.23	2.94	1.94	1.90	2.04	1.10	0.89	0.87
February	0.97	$--$	1.34	2.20	2.33	2.42	2.97	3.1	2.89	1.37	0.92	1.02
March	1.5	0.93	$--$	2.0	2.31	2.19	2.68	2.13	2.93	1.68	1.72	1.3
April	3.10	2.09	2.03	$--$	1.28	1.04	1.61	1.90	2.14	3.18	3.59	3.29
May	3.12	2.18	2.00	0.39	$--$	0.27	0.89	1.26	1.92	1.71	1.75	2.23
June	2.14	2.11	1.99	0.93	0.89	$--$	1.83	1.90	2.19	2.93	2.63	2.58
July	2.39	2.09	2.21	1.92	1.81	1.73	$--$	0.91	1.02	1.57	1.89	2.93
August	1.82	2.05	1.94	1.14	1.69	1.06	0.33	$--$	0.61	1.99	2.87	2.37
September	1.65	1.95	2.36	1.29	1.464	1.38	0.55	0.26	$--$	1.27	1.96	2.12
October	0.93	0.95	1.03	1.79	1.94	1.62	1.37	1.40	2.60	$--$	0.70	0.77
November	0.61	0.83	1.12	1.99	2.23	2.68	2.34	3.02	3.12	1.09	$--$	0.12
December	0.23	0.97	1.11	2.73	2.64	2.13	3.08	2.79	2.0	0.93	0.16	$--$



(a) (b)

FIGURE 4. **Clusters analysis. (a) K-means clustering results. (b) Plot of k-value against SSE (Elbow Method).**



(a) (b) (c)

FIGURE 5. **Within clusters assignments. (a) Cluster_A. (b) Cluster_B. (c) Cluster_C.**

D. ELECTRICITY LOAD TREND CHARACTERIZATION

The purpose of load trend characterization is to provide a better understanding of data by capturing variations at different levels. Depending on the time horizon of prediction, it can be done on various levels such as seasonal trend characterization, monthly trend characterization, daily and hourly

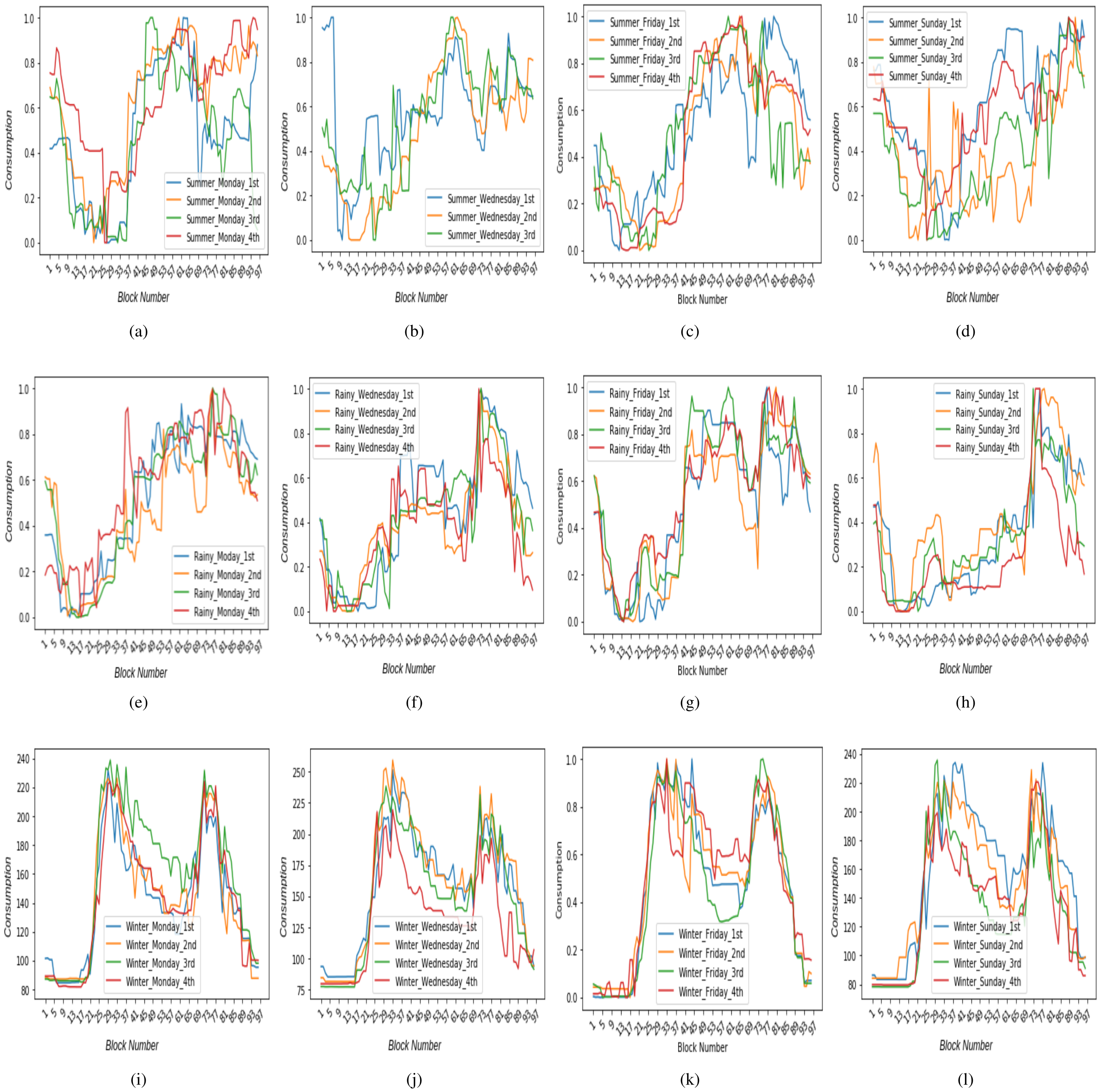


FIGURE 6. **Variations in electricity demand (in MWs) patterns for selected days in all three Seasons. (a) Summer_Monday. (b) Summer_Wednesday. (c) Summer_Friday. (d) Summer_Sunday. (e) Rainy_Monday. (f) Rainy_Wednesday. (g) Rainy_Friday. (h) Rainy_Sunday. (i) Winter_Monday. (j) Winter_Wednesday. (k) Winter_Friday. (l) Winter_Sunday.**

trends characterization. In this study, the goal is to predict average and peak electricity demand for the season, day and time intervals specified by the user. For this purpose, load characterization is performed at two different levels:

- ***Seasonal Analysis***: The output of K-means clustering algorithm (as shown in figure 4a) clearly reflects the electricity load trends over three seasons of a year.
- ***Daily/Time-span Analysis:*** To get a deeper insight in variations of daywise electricity demand in all three seasons, we randomly select four days (*Monday*, *Wednesday*, *Friday*, *Sunday*) from a week. Further, from electricity consumption data (last 5 years) of city Chandigarh, we randomly select four entries[1] corresponding to each selected day (*Monday*, *Wednesday*, *Friday*, *Sunday*) and every season (*Spring*, *Autumn* & *Monsoon*). Figure 6 shows the variations in the electricity demand for each of the

[1]The four entries are chosen for data plotting purpose only. Electricity consumption dataset of all 5 years is used for the task of demand prediction.

TABLE 2. Comparison of average demand (in MWs) prediction results.

Season_Day	Normalized(N) / Un-Normalized(UN)	Average Demand Prediction Error							
		RNN		LSTM		EMD+RNN		EMD+LSTM	
		RMS Train Error	RMS Test Error	RMS Train Error	RMS Test Error	RMS Train Error	RMS Test Error	RMS Train Error	RMS Test Error
Summer_Mondays	N	0.16	0.22	0.153	0.201	0.13	0.15	**0.12**	**0.14**
	UN	24.82	34.09	24.43	33.46	18.27	23.29	**16.97 (8.58%)**	**21.91(11.08%)**
Summer_Wednesdays	N	0.12	0.13	0.09	0.11	0.08	0.11	**0.06**	**0.09**
	UN	25.13	28.98	18.59	23.52	14.27	19.12	**11.83 (5.50%)**	**16.47 (7.66%)**
Summer_Fridays	N	0.19	0.20	0.18	0.19	0.18	0.19	**0.13**	**0.17**
	UN	27.39	28.12	20.18	21.23	19.21	20.29	**13.91 (6.41%)**	**15.29 (7.04%)**
Summer_Sundays	N	0.12	0.17	0.09	0.13	0.08	0.10	**0.07**	**0.08**
	UN	23.58	30.37	17.67	24.28	14.71	19.02	**11.42 (6.21%)**	**15.89 (8.66%)**
Rainy_Mondays	N	0.13	0.15	0.12	0.14	0.09	0.11	**0.08**	**0.10**
	UN	18.72	21.27	17.60	19.28	12.99	15.39	**11.39 (5.71%)**	**14.27 (6.91%)**
Rainy_Wednesdays	N	0.12	0.14	0.10	0.11	0.08	0.10	**0.07**	**0.09**
	UN	23.33	25.49	18.78	20.98	14.17	18.04	**12.73 (6.23%)**	**16.90 (7.24%)**
Rainy_Fridays	N	0.23	0.25	0.18	0.20	0.18	0.19	**0.12**	**0.14**
	UN	22.39	25.12	16.30	19.90	17.29	19.98	**10.37 (5.37%)**	**12.10 (5.99%)**
Rainy_Sundays	N	0.13	0.14	0.10	0.11	0.09	0.10	**0.08**	**0.09**
	UN	18.16	19.02	15.00	16.21	13.46	15.64	**10.17 (5.59%)**	**12.84 (6.99%)**
Winter_Mondays	N	0.09	0.12	0.07	0.10	0.07	0.11	**0.07**	**0.08**
	UN	11.87	13.78	8.55	12.10	9.01	12.46	**8.03 (5.53%)**	**9.76 (6.72%)**
Winter_Wednesdays	N	0.09	0.10	0.07	0.09	0.07	0.09	**0.07**	**0.08**
	UN	14.51	15.64	11.82	14.30	11.57	14.52	**10.03 (6.77%)**	**12.42 (8.39%)**
Winter_Fridays	N	0.14	0.15	0.14	0.15	0.13	0.13	**0.10**	**0.11**
	UN	14.82	16.47	14.67	15.72	13.29	14.13	**9.42 (6.34%)**	**11.47 (7.72%)**
Winter_Sundays	N	0.09	0.11	0.07	0.11	0.07	0.08	**0.05**	**0.06**
	UN	10.28	12.00	7.98	11.36	7.71	9.83	**5.10 (5.13%)**	**7.23 (6.32%)**
	Average % Consumption forecasting (Average Demand) Error Using EMD+LSTM Model							**6.11%**	**7.56%**

selected day (*Monday*, *Wednesday*, *Friday*, *Sunday*) in all three seasons (*Spring*, *Autumn & Monsoon*).

To develop a prediction model for estimating the electricity demand based on season, day and timestamp of the day, we divide the electricity demand of each day into different time intervals namely *Time_Interval*_1 (00:00 am to 06:00 am), *Time_Interval*_2 (06:00 am to 12:00 pm), *Time_Interval*_3 (12:00 pm to 06:00 pm) and *Time_Interval*_4 (06:00 pm to 00:00 am). The load trend characterization plays a critical role in building efficient prediction models, future strategies planning, capacity scheduling and purchase planning.

E. PROPOSED EMD BASED LSTM MODEL

The next step is to build prediction models to approximate future demand by using the available dataset. The clustering algorithm identified three seasons and we have picked four days for a week, a total of twelve data chunks are available. Therefore, we train 12 LSTM network models (one for each day in every season) to estimate future electricity demand for the user specified time interval. The proposed approach works on the principle of divide and conquer i.e. by dividing a task into smaller sub-tasks and then integrating the solutions to provide the optimum results.

In order to develop models for forecasting average and peak electricity demand for a user-specified time_interval and its subsequent time_interval of a day, steps 1 to 4 are applied on the data of each selected day (*Monday*, *Wednesday*, *Friday*, *Sunday*) in all seasons (*Spring*, *Autumn & Monsoon*).

- **Step 1:** The time series dataset is divided into IMFs and residual by applying EMD algorithm.
- **Step 2:** Create an input matrix of features for input to LSTM models.
- **Step 3:** Do train and test data Splitting.
- **Step 4:** Train and build LSTM models to predict results for each extracted IMF and residue.
- **Step 5:** Combine all prediction results by summation to generate ensemble output.

Basically, the LSTM model requires input data to be of three dimensions (S, W, F) where S defines the number of input samples, W denotes sequence length and F defines the number of features in each sequence. The simplest method to build up a LSTM model is to provide an input vector (a_1, $a_2 .... a_n$) to the model for predicting the output. This method is based on the assumption that timestamp values are independent of each other. However, this assumption is not valid in our case. Another way of training the LSTM models is to implement window based learning. This method allows

TABLE 3. Comparison of peak demand (in MWs) prediction results.

Season_Day	Normalized(N) / Un-Normalized(UN)	Peak Demand Prediction Error							
		RNN		LSTM		EMD+RNN		EMD+LSTM	
		RMS Train Error	RMS Test Error	RMS Train Error	RMS Test Error	RMS Train Error	RMS Test Error	RMS Train Error	RMS Test Error
Summer_Mondays	N	0.19	0.19	0.17	0.18	0.16	0.17	**0.15**	**0.16**
	UN (% Error)	28.73	33.95	26.30	31.33	20.49	24.09	**17.59 (7.81%)**	**20.60 (9.15%)**
Summer_Wednesdays	N	0.12	0.13	0.09	0.14	0.09	0.10	**0.08**	**0.10**
	UN (% Error)	26.87	28.47	19.57	29.52	16.08	22.02	**14.37 (5.97%)**	**23.12 (9.60%)**
Summer_Fridays	N	0.22	0.22	0.18	0.20	0.15	0.16	**0.13**	**0.14**
	UN (% Error)	29.17	29.68	22.14	25.80	19.22	20.21	**13.15 (5.40%)**	**14.83 (6.19%)**
Summer_Sundays	N	0.11	0.15	0.10	0.14	0.08	0.10	**0.08**	**0.07**
	UN (% Error)	22.11	30.09	20.10	28.42	14.01	20.70	**13.87 (6.69%)**	**19.26 (8.29%)**
Rainy_Mondays	N	0.15	0.19	0.13	0.15	0.10	0.12	**0.08**	**0.10**
	UN (% Error)	23.10	25.37	20.07	22.09	16.67	19.80	**14.80 (5.50%)**	**17.41 (6.47%)**
Rainy_Wednesdays	N	0.13	0.13	0.10	0.11	0.09	0.11	**0.07**	**0.08**
	UN (% Error)	26.95	27.55	19.30	23.26	16.17	23.55	**10.42 (5.84%)**	**14.80 (7.46%)**
Rainy_Fridays	N	0.20	0.21	0.19	0.21	0.14	0.16	**0.10**	**0.11**
	UN (% Error)	21.98	23.24	18.83	21.91	16.12	17.98	**13.80 (5.12%)**	**14.94 (6.75%)**
Rainy_Sundays	N	0.15	0.17	0.16	0.17	0.13	0.16	**0.10**	**0.12**
	UN (% Error)	18.75	21.23	19.24	20.21	16.00	20.84	**11.83 (4.88%)**	**17.97 (7.41%)**
Winter_Mondays	N	0.15	0.19	0.12	0.17	0.10	0.13	**0.08**	**0.11**
	UN (% Error)	19.26	20.22	17.20	19.52	14.64	17.82	**11.47 (6.00%)**	**15.05 (7.88%)**
Winter_Wednesdays	N	0.09	0.12	0.08	0.12	0.07	0.10	**0.06**	**0.09**
	UN (% Error)	16.89	22.57	14.76	22.04	13.41	18.94	**11.13 (5.89%)**	**16.76 (8.07%)**
Winter_Fridays	N	0.15	0.15	0.14	0.15	0.14	0.15	**0.12**	**0.14**
	UN (% Error)	19.48	19.78	17.97	19.46	16.01	17.29	**13.28 (6.97%)**	**16.47 (8.64%)**
Winter_Sundays	N	0.11	0.18	0.09	0.17	0.09	0.12	**0.07**	**0.11**
	UN (% Error)	14.85	26.92	12.84	25.33	11.67	19.88	**8.88 (5.90 %)**	**15.03 (8.30%)**
	Average % Consumption forecasting (Peak Demand) Error Using EMD+LSTM Model							**5.99%**	**7.85%**

the LSTM model to directly deal with previous timestamp values (Lagged values). It works by dividing a time series signal of length L into ($L - Out\_Size - Input\_Window\_Size$) patches of length ($Out\_Size + Input\_Window\_Size$) where $Input\_Window\_Size$ represents the size of input window and Out_Size denotes the length of the output window. Further, within each window a local normalization step is applied to stabilize the learning process of LSTM window model. In this paper, the LSTM network models are trained with $Input\_Window\_Size$ of 16 and Out_Size of 2. After each prediction timestamp, LSTM window based network model shifts both input and output windows by two steps causing a overlap with the prior windows. In this way, the proposed model provides support for dynamic learning. After proper model training and parameters tuning, it can be used to predict demand for the test data windows. There are several things related to the LSTM models that need to be considered carefully:

1) HYPERPARAMETERS TUNING

One of the important steps while building a model is the selection of parameters and the proper parameters tuning. There are several parameters associated with LSTM model such as $Number\_of\_hidden\_layers$, $Input\_neurons$, Out_window_size, $Input\_window\_size$, $number\_of\_epochs$, $regularization\_weight$ and $learning\_rate$ that need to be determined accurately. The size of Out_window and $Input\_window$ parameters depends on the time horizon of prediction. The $Input\_neurons$ parameter is determined by the dimensions of the input data. Some other parameters viz. stopping criteria, learning rate are determined by executing several runs of LSTM training. There is no global method to estimate hyper-parameters. The selection of hyper-parameters is completely data dependent.

2) MODEL REGULARIZATION

The regression models may overfit due to the presence of noise and small perturbations in the training data [24]. The Weight decay model regularization is adopted to minimize model over-fitting. This method constrains the model by penalizing the weight vectors for being large.

3) TRAIN AND TEST SPLITTING

In all three seasons, the overall electricity consumption dataset of each day is divided into a ratio of 60:10:30 for the purpose of model training, validation and testing respectively.

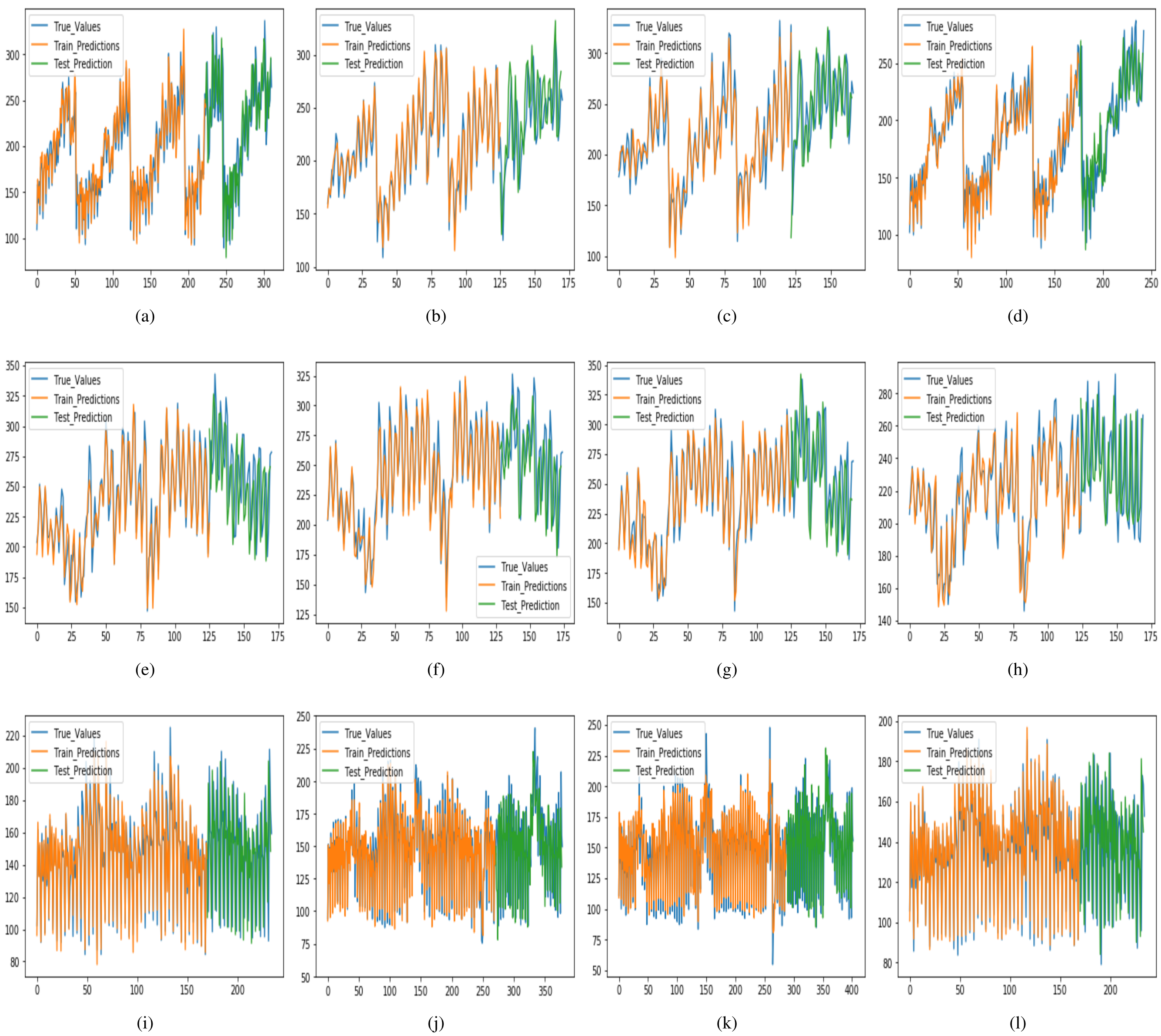


FIGURE 7. **Average electricity demand (in MWs) prediction results using *EMD+LSTM* for selected Days in all three seasons (*Mon: Monday, Wed: Wednesday, Fri: Friday, Sun: Sunday, Avg: Average, Pred: Prediction, Res: Results, Sumr: Summer). (a) Summer_Mon_Avg_Load_Pred_Res. (b) Sumr_Wed_Avg_Load_Pred_Res. (c) Sumr_Fri_Avg_Load_Pred_Res. (d) Sumr_Sun_Avg_Load_Pred_Res. (e) Rainy_Mon_Avg_Load_Pred_Res. (f) Rainy_Wed_Avg_Load_Pred_Res. (g) Rainy_Fri_Avg_Load_Pred_Res. (h) Rainy_Sun_Avg_Load_Pred_Res. (i) Winter_Mon_Avg_Load_Pred_Res. (j) Winter_Wed_Avg_Load_Pred_Res. (k) Winter_Fri_Avg_Load_Pred_Res. (l) Winter_Sun_Avg_Load_Pred_Res.**

4) PERFORMANCE MEASURES

Prediction accuracy of a model is defined as the ability to predict with minimum error. In this study, we have used the following measures to evaluate the prediction accuracy of the proposed model:

- ***RMSE* [34]**: It is defined as

$$RMSE = \sqrt{\frac{1}{2n}\sum_{t}\sum_{i=1}^{n}(R_{i,t} - \bar{R}_{i,t})^2} \tag{9}$$

  where $R_{i,t}$ represents the real value for timestamp t and $\bar{R}_{i,t}$ represents the predicted value at timestamp t.

- ***Absolute Percentage Error***: It is defined as the percentage (%) deviation of predicted values from the real time demand observations.

F. RESULTS AND DISCUSSION

In addition to EMD based LSTM model *(EMD+LSTM)*, several other regression models namely *RNN*, *LSTM*, EMD based RNN model *(EMD+RNN)* are also trained to predict average and peak electricity demand for a particular time_interval (specified by the user) and its subsequent time_interval of a day. These regression models also use the concept of lagged values (previous timestamps) to

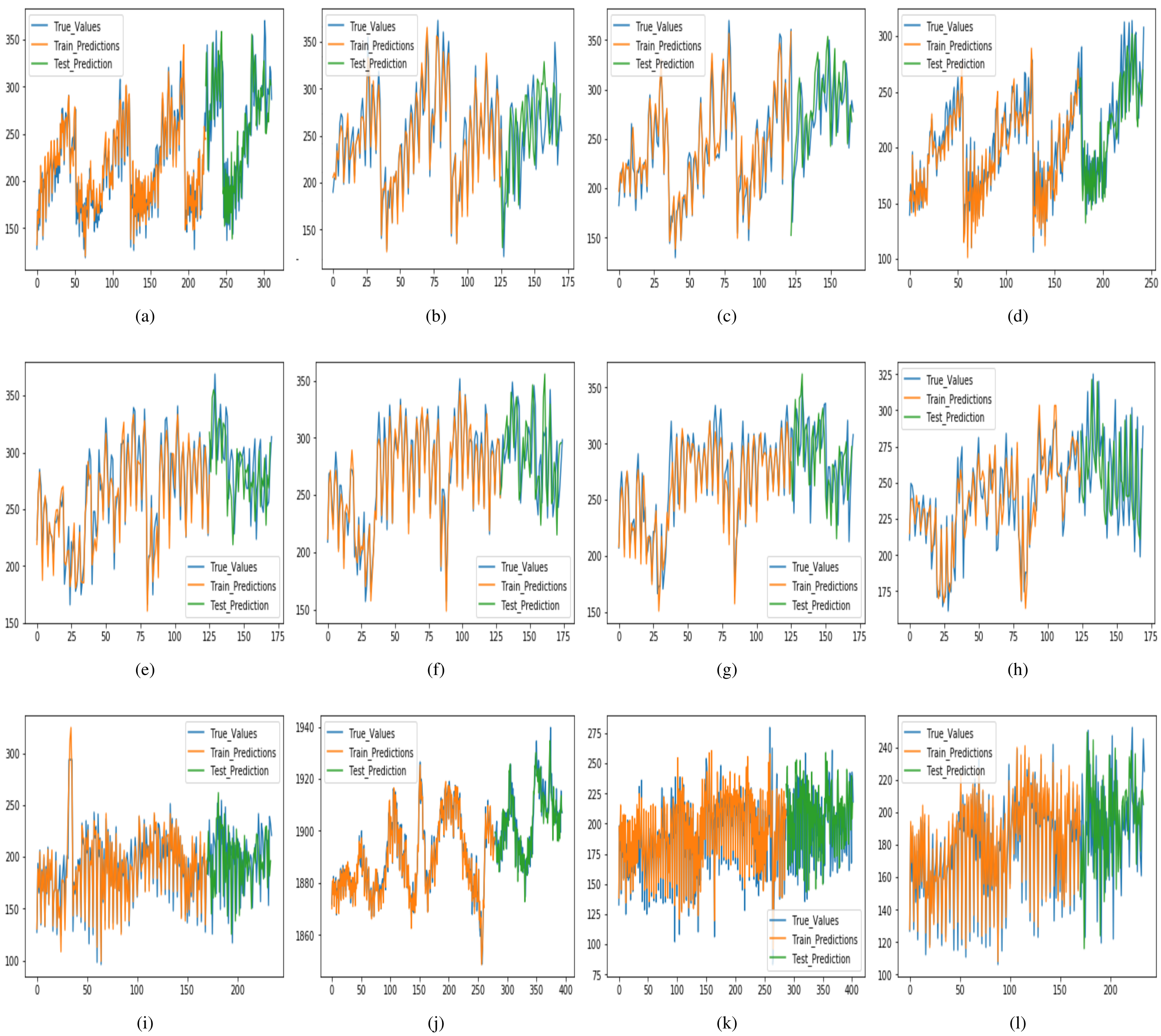


FIGURE 8. Peak electricity demand (in MWs) prediction results using *EMD+LSTM* for selected days in all three seasons (*Mon: Monday, Wed: Wednesday, Fri: Friday, Sun: Sunday, Avg: Average, Pred: Prediction, Res: Results, Sumr: Summer). (a) Sumr_Mon_Peak_Load_Pred_Res. (b) Sumr_Wed_Peak_Load_Pred_Res. (c) Sumr_Fri_Peak_Load_Pred_Res. (d) Sumr_Sun_Peak_Load_Pred_Res. (e) Rainy_Mon_Peak_Load_Pred_Res. (f) Rainy_Wed_Peak_Load_Pred_Res. (g) Rainy_Fri_Peak_Load_Pred_Res. (h) Rainy_Sun_Peak_Load_Pred_Res. (i) Winter_Mon_Peak_Load_Pred_Res. (j) Winter_Wed_Peak_Load_Pred_Res. (k) Winter_Fri_Peak_Load_Pred_Res. (l) Winter_Sun_Peak_Load_Pred_Res.

estimate future demand. Table 2 and 3 show the comparison of average and peak demand prediction results of *RNN*, *LSTM*, *(EMD+RNN)* and proposed EMD based LSTM *(EMD+LSTM)* models in terms of RMSE (both normalized and Un-normalized). From the comparison of prediction results in Table 2 and 3, it is apparent that EMD based LSTM model (proposed approach) does accurate average and peak demand forecasting than other regression models. Moreover, the prediction error using proposed *EMD+LSTM* approach varies from 5 to 8%. Figure 7 and 8 show the plots of prediction results of EMD based LSTM model for the selected days in all three seasons. The proposed EMD based LSTM approach has several advantages:

- The proposed approach can be used to accurately forecast the electricity demand for a time_interval (specified by the user) and its succeeding time_interval of a day.
- The approach can actively support large input and output mapping.
- The window-based approach used in this study provides support for dynamic learning and can be used to efficiently handle the long-term historical dependencies in the electricity time series data.

V. CONCLUSION

Electricity demand estimation and management has a significant impact on various social and economic policies of a nation. In the past decades, various physical models, statistical tools, and AI (Artificial Intelligence) based methods were developed to accurately estimate electricity demand. Out of these methods, the AI-based techniques were found to be most promising while capturing nonlinear variations of data. However, these techniques are not capable of handling historical data dependencies. In this study, a hybrid EMD based deep learning approach is proposed to predict electricity demand by taking care of long-term data dependencies. The approach uses window technique of LSTM network model to optimally forecast demand for a user-specified time interval and its next time interval. The average and peak demand prediction results of the proposed (EMD+LSTM) approach are compared with RNN, EMD based RNN (EMD+RNN) and LSTM models in terms of RMSE and absolute percentage error. Based on the comparison of prediction results, the following inferences can be drawn:

- The proposed EMD based deep learning approach *(EMD+LSTM)* outperforms other regression models for electricity demand time series forecasting.
- The proposed deep learning approach *(EMD+LSTM)* can be utilized to deal efficiently with nonlinear features of electricity time series data.
- The proposed approach can provide support for dynamic learning by continually shifting the input and output windows during model building and testing phases.

REFERENCES

[1] W. He, "Load forecasting via deep neural networks," *Procedia Comput. Sci.*, vol. 122, pp. 308–314, Dec. 2017.

[2] V. Saxena and P. C. Bhattacharya, "Inequalities in LPG and electricity consumption in India: The role of caste, tribe, and religion," *Energy Sustain. Develop.*, vol. 42, pp. 44–53, Feb. 2018.

[3] I. Koprinska, M. Rana, and V. G. Agelidis, "Correlation and instance based feature selection for electricity load forecasting," *Knowl.-Based Syst.*, vol. 82, pp. 29–40, Jul. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0950705115000714

[4] L. Friedrich and A. Afshari, "Short-term forecasting of the Abu Dhabi electricity load using multiple weather variables," *Energy Procedia*, vol. 75, pp. 3014–3026, Aug. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1876610215013843

[5] W.-C. Hong, Y. Dong, W. Y. Zhang, L.-Y. Chen, and B. K. Panigrahi, "Cyclic electric load forecasting by seasonal SVR with chaotic genetic algorithm," *Int. J. Elect. Power Energy Syst.*, vol. 44, no. 1, pp. 604–614, 2013.

[6] Y. H. Chen, W.-C. Hong, W. Shen, and N. N. Huang, "Electric load forecasting based on a least squares support vector machine with fuzzy time series and global harmony search algorithm," *Energies*, vol. 9, no. 2, p. 70, 2016.

[7] G. Dudek, "Short-term load forecasting using random forests," in *Intelligent Systems*. Cham, Switzerland: Springer, 2015, pp. 821–828.

[8] A. S. Ahmad *et al.*, "A review on applications of ANN and SVM for building electrical energy consumption forecasting," *Renew. Sustain. Energy Rev.*, vol. 33, pp. 102–109, May 2014.

[9] 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*, pp. 1–8, Sep. 2017.

[10] S. Sood, M. Kumar, R. B. Pachori, and U. R. Acharya, "Application of empirical mode decomposition–based features for analysis of normal and cad heart rate signals," *J. Mech. Med. Biol.*, vol. 16, no. 1, p. 1640002, 2016.

[11] H. Wang, L. Liu, S. Dong, Z. Qian, and H. Wei, "A novel work zone short-term vehicle-type specific traffic speed prediction model through the hybrid EMD–ARIMA framework," *Transportmetrica B, Transport Dyn.*, vol. 4, no. 3, pp. 159–186, 2016.

[12] F. Cannarile *et al.*, "An unsupervised clustering method for assessing the degradation state of cutting tools used in the packaging industry," in *Proc. Eur. Saf. Rel. Conf. (ESREL)*, 2017, pp. 138–139.

[13] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. Hoboken, NJ, USA: Wiley, 2015.

[14] H.-X. Zhao and F. Magouls, "A review on the prediction of building energy consumption," *Renew. Sustain. Energy Rev.*, vol. 16, no. 6, pp. 3586–3592, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032112001438

[15] A. Foucquier, S. Robert, F. Suard, L. Stéphan, and A. Jay, "State of the art in building modelling and energy performances prediction: A review," *Renew. Sustain. Energy Rev.*, vol. 23, pp. 272–288, Jul. 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032113001536

[16] M. Parti and C. Parti, "The total and appliance-specific conditional demand for electricity in the household sector," *Bell J. Econ.*, vol. 11, no. 1, pp. 309–321, 1980.

[17] M. Aydinalp-Koksal and V. I. Ugursal, "Comparison of neural network, conditional demand analysis, and engineering approaches for modeling end-use energy consumption in the residential sector," *Appl. Energy*, vol. 85, no. 4, pp. 271–296, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S030626190600136X

[18] A. Aranda, G. Ferreira, M. D. Mainar-Toledo, S. Scarpellini, and E. L. Sastresa, "Multiple regression models to predict the annual energy consumption in the Spanish banking sector," *Energy Buildings*, vol. 49, pp. 380–387, Jun. 2012.

[19] S. A. Kalogirou, "Artificial neural networks in energy applications in buildings," *Int. J. Low-Carbon Technol.*, vol. 1, no. 3, pp. 201–216, 2006.

[20] G. Escrivá-Escrivá, C. Álvarez-Bel, C. Roldán-Blay, and M. Alcázar-Ortega, "New artificial neural network prediction method for electrical consumption forecasting based on building end-uses," *Energy Buildings*, vol. 43, no. 11, pp. 3112–3119, 2011.

[21] M. A. R. Biswas, M. D. Robinson, and N. Fumo, "Prediction of residential building energy consumption: A neural network approach," *Energy*, vol. 117, pp. 84–92, Dec. 2016.

[22] M. S. Al-Musaylh, R. C. Deo, J. F. Adamowski, and Y. Li, "Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in queensland, Australia," *Adv. Eng. Informat.*, vol. 35, pp. 1–16, Jan. 2018.

[23] M. E. Günay, "Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: Case of Turkey," *Energy Policy*, vol. 90, pp. 92–101, Mar. 2016.

[24] A. Rahman, V. Srikumar, and A. D. Smith, "Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks," *Appl. Energy*, vol. 212, pp. 372–385, Feb. 2018.

[25] Y. Yaslan and B. Bican, "Empirical mode decomposition based denoising method with support vector regression for time series prediction: A case study for electricity load forecasting," *Measurement*, vol. 103, pp. 52–61, Jun. 2017.

[26] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*. Hoboken, NJ, USA: Wiley, 2001.

[27] K. Bandara, C. Bergmeir, and S. Smyl. (2017). "Forecasting across time series databases using long short-term memory networks on groups of similar series." [Online]. Available: https://arxiv.org/abs/1710.03222

[28] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[29] *Jakob Aungiers*. Accessed: Jan. 15, 2018. [Online]. Available: http://www.jakob-aungiers.com/articles/a/LSTM-Neural-Network-for-Time-Series-Prediction

[30] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[31] G. Rilling, P. Flandrin, and P. Goncalves, "On empirical mode decomposition and its algorithms," in *Proc. IEEE Workshop Nonlinear Signal Image Process. (EURASIP)*, vol. 3, Jun. 2003, pp. 8–11.

[32] *ENVIS Centre of Chandigarh State of Environment*. Accessed: Jan. 15, 2018. [Online]. Available: http://chenvis.nic.in

[33] S. García, J. Luengo, and F. Herrera, *Data Preprocessing in Data Mining*. Berlin, Germany: Springer, 2016.

[34] N. M. Nasrabadi, "Pattern recognition and machine learning," *J. Electron. Imag.*, vol. 16, no. 4, p. 049901, 2007.

[35] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. London, U.K.: Pearson Education, 2015.
[36] A. K. Jain, M. N. Murty, and P. J. Flynn, ''Data clustering: A review,'' *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, Sep. 1999.
[37] P. Senin, ''Dynamic time warping algorithm review,'' Dept. Inf. Comput. Sci., Univ. Hawai'i Mānoa, Honolulu, HI, USA, Tech. Rep., vol. 855, 2008, pp. 1–23.
[38] E. Keogh, L. Wei, X. Xi, M. Vlachos, S.-H. Lee, and P. Protopapas, ''Lb_keogh supports exact indexing of shapes under rotation invariance with arbitrary representations and distance measures,'' in *Proc. 32nd Int. Conf. Very Large Data Bases*, 2006, pp. 882–893.
[39] M. Yan, ''Methods of determining the number of clusters in a data set and a new clustering criterion,'' Ph.D. dissertation, Virginia Tech, Blacksburg, VA, USA, 2005.
[40] *Robert Gove Block*. Accessed: May 22, 2017. [Online]. Available: https://bl.ocks.org/rpgove/0060ff3b656618e9136b

JATIN BEDI received the bachelor's degree from K.U. Kurukshetra in 2013, and the master's degree from DCSA, K.U. Kurukshetra, in 2015. He is currently pursuing the Ph.D. degree with IIT Roorkee, Roorkee, India. His main research areas are data mining, big data and time series data. He was a recipient of the MHRD Fellowship for Doctoral Research. He has presented his work in several reputed conferences, including SIAM SDM, and MLDM, USA.

DURGA TOSHNIWAL (M'11) received the bachelor's degree in engineering and the M.Tech. degree from the National Institute of Technology, Kurukshetra, and the Doctor of Philosophy from IIT Roorkee, Roorkee, India. She is currently an Associate Professor with the Department of Computer Science and Engineering, IIT Roorkee. She is also with IBM Research India. She has published her research work in several international journals and conferences. Her research interests include data mining and databases, mining time series, data streams and social media data, privacy preserving DM, soft computing in DM, and machine learning. She is a fellow of the ACM. She has received various awards and honors. She has attended, chaired sessions, and presented her work in several reputed international conferences in USA, U.K., Australia, and Europe.