IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 28, 2019, accepted July 23, 2019, date of publication August 23, 2019, date of current version September 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2937222

A New Hybrid Model for Short-Term Electricity Load Forecasting

MD. RASHEDUL HAQ[1] AND ZHEN NI[2], (Member, IEEE)
[1]Electrical Engineering and Computer Science Department, South Dakota State University, Brookings, SD 57007, USA
[2]Computer, Electrical Engineering, and Computer Science Department, Florida Atlantic University, Boca Raton, FL 33431, USA

Corresponding author: Zhen Ni (zhenni@fau.edu)

ABSTRACT Nowadays electricity load forecasting is important to further minimize the cost of day-ahead energy market. Load forecasting can help utility operators for the efficient management of a demand response program. Forecasting of electricity load demand with higher accuracy and efficiency can help utility operators to design reasonable operational planning of generation units. But solving the problem of load forecasting is a challenging task since electricity load is affected by previous history load, several exogenous external factors (i.e., weather variables, social variables, working day or holiday), time of day, and season of the year. To solve the problem of short-term load forecasting (STLF) and further improve the forecasting accuracy, in this paper we have proposed a novel hybrid STLF model with a new signal decomposition and correlation analysis technique. To this end, load demand time series is decomposed into some regular low frequency components using improved empirical mode decomposition (IEMD). To compensate for the information loss during signal decomposition, we have incorporated the effect of exogenous variables by performing correlation analysis using T-Copula. From the T-Copula analysis, peak load indicative binary variable is derived from value at risk (VaR) to improve the load forecasting accuracy during peak time. The data obtained from IEMD and T-Copula is applied to deep belief network for predicting the future load demand of specific time. The proposed data driven method is validated on real time data from the Australia and the United States of America. The performance of proposed load forecasting model is evaluated in terms of mean absolute percentage error (MAPE) & root mean square error (RMSE). Simulation results verify that, the proposed model provides a significant decrease in MAPE and RMSE values compared to traditional empirical mode decomposition based electricity load forecasting.

INDEX TERMS Short-term load forecasting, demand response, smart grid, improved empirical mode decomposition, T-Copula, peak load indicative variable, VaR, and deep belief network.

I. INTRODUCTION

Electricity load forecasting is essential for the utility provider to manage the demand response program efficiently in day ahead energy market. From the information of electricity load demand of consumers, utility providers can estimate how much electric energy is needed in the grid. The objective of the utility provider is to minimize the cost of energy production and purchasing [1]. In this scenario, a prior knowledge about the energy demand can help utility providers to make proper planning of generation units scheduling and amount of energy to be purchased [2]. The accurate electricity load forecasting has a significant role in power system, but any error in forecast incurs additional cost. According to Bunn and Farmer [3], [4], the cost of electric utility operator is saved by 10 million pounds due to 1% decrease of load forecasting errors. Power grid planning, investment and transaction are also based on accurate electricity load forecasting. Thus, accurate electricity load forecasting is prerequisite for making secure, reliable and economic operation of power system [5].

Over the last few years, researchers have proposed many models to forecast electricity load for varying time interval. Depending on the forecast time interval, those models can be classified into short-term, medium-term, and long-term load forecasting model [6]. Usually short-term load forecasting (STLF) models predict the half hourly or hourly load demand for next few hours to two weeks. The results

The associate editor coordinating the review of this article and approving it for publication was Bora Onat.

of STLF are used for short-term operational planning of the power system, e.g., generation scheduling. When load forecasting is intended for longer time interval i.e., from next few months to next few years, then medium-term and long-term load forecasting models are developed. Based on the model architecture, load forecasting models are primarily divided into two classes: traditional statistical models and advanced data driven models. Traditional statistical models are built using linear regression function where the problem of STLF is viewed as a time-series prediction problem [7]–[12]. The regression based models are effective for predicting stationary time series. However, load demand time series is non-stationary and shows nonlinear characteristics, thus advanced data driven models are proposed in recent times. To date, STLF problem has been investigated with different advanced data driven models. Advanced data driven models include: fuzzy logic based [13], artificial neural network (ANN) based [14]–[19] and exponential smoothing methods [20]. ANN based models are the most popular among advanced data driven models. Both statistical and advanced individual data driven models are commonly used for load forecasting. However, a single model is inadequate to represent inherent characteristics of electricity load demand because it depends on several factors including weather (i.e., temperature, humidity etc.,), time, and socio-economic constraints [21]. If we do not consider the exogenous variables (i.e., dry bulb temperature, wet bulb temperature, due point temperature and humidity), such models produce fair forecasting accuracy. Considering the effect of exogenous variables in load forecasting model incurs several challenges. First, the electricity load demand forecasting with high accuracy is a challenging task because those variables cause the data to be highly nonlinear and unpredictable. Second, artificial neural network based load forecasting model suffers from overfitting issue or falling into local minima. And, for traditional empirical decomposition based load forecasting model, end effect and envelop fitting limitation reduces the load forecasting accuracy.

In recent time, various hybrid models formed by integrating different models have been reported for improving the forecasting accuracy [22]–[33]. The reason is that, different models can capture the features of electricity load profiles. Motivating from the challenges as reported in [22]–[33], this paper has proposed a novel hybrid load forecasting model which includes a new signal decomposition technique and correlation analysis technique. To mitigate end effect and envelope fitting limitation associated with traditional empirical mode decomposition (EMD) [22], the improved empirical mode decomposition (IEMD) method proposed by Yang *et.al.* [35] and Yang and Yang [36] is investigated. Later on, to compensate for the information loss during signal decomposition, the effect of weather factors (i.e., exogenous variables) is incorporated by introducing T-Copula correlation analysis technique. Even though we have investigated IEMD to suppress the limitation of traditional EMD, good results can be obtained by just mirroring the extrema close to edges [22]. Interested reader can investigate end mirror extending in high frequency and least square polynomial in low frequency for IEMD. However, no complete solution is in sight currently. There is a room for suppressing the end effect. The main contributions of this paper are as follows:

- In this work, we have investigated a new method to suppress the end effect and envelope fitting limitation of traditional EMD. The end effect of EMD is reduced by employing linear interpolation to obtain the extreme points of observation interval boundary. Next, instead of cubic spline curve, nonuniform rational B-splines curve is used for suppressing envelope fitting limitation. This method can accurately measure the average value of instantaneous signal, which helps to realize the accurate signal decomposition.
- To compensate the information loss during signal decomposition, we have included the effect of exogenous variables to the load forecasting model. This is done by performing correlation analysis using T-Copula. T-Copula analysis can quantify the uncertainty between power load and external exogenous variables. Considering the importance of peak load forecasting accuracy, a tandem parameter VaR is computed from the fitted Copula model to determine the peak load indicative variable. For our research we have determined four peak load indicative variables for exogenous variables. Including those variables as input to load forecasting model improves load forecasting accuracy during peak time. Due to this incorporation, MAPE and RMSE value decreased by 9.59% and 7.57% respectively for case study #1, whereas for case study #2 these values are decreased by 8.64% and 7.83% respectively.
- The data obtained from IEMD and T-Copula helps for processing of more information. IEMD provides higher decomposition efficiency and peak load indicative variables help to improve the load forecasting accuracy during peak time. Therefore, we employ the DBN to process the data obtained from signal decomposition and correlation analysis. DBN is employed to overcome the deficiency of traditional neural network based models. The DBN learns to probabilistically reconstruct the input data and then detect feature patterns. Therefore, the proposed novel hybrid model consisting of IEMD, T-Copula, and DBN provides higher load forecasting accuracy with consideration of exogenous variables. Overall, MAPE and RMSE value decreased by 21.19% and 16.93% respectively for case study #1, whereas for case study #2 these values are decreased by 15.28% and 13.86% respectively.

This paper is organized as follows. The relevant work and problems associated with those STLF models are mentioned in Section II. The framework of the proposed hybrid model is presented in Section III and the design steps of proposed model for STLF is presented in IV. Simulation results and analysis are carried out in Section V. Finally the conclusion is presented in Section VI.

II. RELATED WORK OF SHORT TERM LOAD FORECASTING

If the load profile for a day is defined as $E_m(t) = [E_m(1), E_m(2), \ldots, E_m(N)]^T$, where $E_m(t)$ is the load profile on mth day and $t = 1, 2, 3, \ldots, N$ represents different time instances. The task of STLF model is to predict the load profile of future time instances i.e., $E_m(t+1)$ or $E_{m+1}(t)$. To avoid the notation complexity in later, we will use $E(t)$ as a load demand time series instead of load profile of a particular day $E_m(t)$.

Nowadays, the trend is to develop ensemble model or hybrid model for increasing the accuracy of load forecasting. Our focus is to develop a novel hybrid load forecasting model that follows ensemble strategy for making final prediction. The hybrid models are formed by integrating different models for improving the forecasting accuracy. The reason is that, different models can capture the features of electricity load profiles. In general, the hybrid models are classified into two main categories. For the first category model, electricity load is predicted separately by different models [22]–[27]. For the second category model, electricity load is decomposed into several components. Then each component is predicted by a suitable model [22], [28]–[33]. For example, in [26], [27], different models such as back propagation neural network, genetic algorithm back propagation neural network, wavelet neural network, radical basis function neural network, general regression neural network, support vector machine are used separately to predict the energy load demand. Then, multi-objective flower pollination algorithm is applied to optimize the weight of each model. The final prediction value is determined from weighted average. Although the performance of the first category model is better than single model, there is a problem in calculating the weight of each model, which leaves a riddle for determining the optimal weights. Therefore, the second category model has been proposed by many researchers. For the second category model, electricity load is decomposed into several low frequency components. Then each component is predicted by a suitable model and the final forecasting result is the sum of each components forecasting results. Li *et.al.* [28], used wavelet transform to decompose the original electricity load into several components. Then each component is predicted by extreme learning machine combined with partial least square regression. In [29], electricity load is decomposed by wavelet transform into some detailed sub series, and then each subseries is predicted by boundary network node model. Although wavelet transform can decompose the original electricity load into some low frequency components, it lacks the ability to extract the deep information as much as possible. To increase the efficiency of decomposition, in recent times EMD has been used by several researchers [22], [30]–[33]. In [33] X.H. Qiu et.al., used EMD to decompose the electricity load into several intrinsic mode functions and one residual function. X.H. Qiu et.al., have predicted the future load demand and reported a higher load forecasting accuracy. However, the end effect and envelope fitting limitation associated with EMD decreases the efficiency of signal decomposition which consequently decreases the load forecasting accuracy. Besides, X.H. Qiu et.al., also ignored the effect of exogenous variables. Therefore, there is a scope to improve the load forecasting accuracy of [33]. In our proposed novel hybrid load forecasting model, our objective is to improve the load forecasting accuracy by: (i) suppressing the end effect and envelope fitting limitation of traditional EMD, and (ii) incorporating the effect of exogenous variables into the load forecasting model.

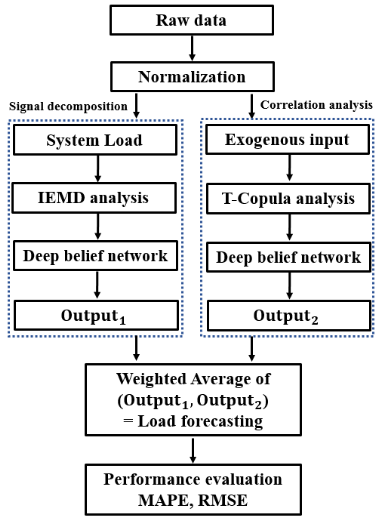

FIGURE 1. Framework of proposed hybrid STLF model.

III. FRAMEWORK OF THE PROPOSED METHOD

The framework of the proposed hybrid model for STLF is shown in Figure 1. The basic architecture of the proposed hybrid model consist of load demand time series decomposition and processing of exogenous input variables with the help of correlation analysis. Load demand time series and exogenous input variables are processed in parallel. Compared to [33], the application of IEMD will improve the signal decomposition efficiency and considering peak load indicative variable as input parameters will improve the load forecasting accuracy during peak load time. The binary peak load indicative variables for each of the exogenous inputs are determined from the VaR computed from the T-Copula correlation analysis.

The signal decomposition using IEMD will yield low frequency component called intrinsic mode functions e.g., ($IMF_1, IMF_2, IMF_3, \ldots, etc.,$) and a signal monotone function i.e., residual function. The steps of the load forecasting from signal decomposition is given below:

- 1st Step: In this step, IEMD is employed to decompose the electricity load demand time series into different sub-series with different frequencies i.e., ($IMF_1, IMF_2, IMF_3, \ldots, etc.,$), and a residual.
- 2nd Step: Each IMFs and residual are applied as DBN input, and the forecasting result for each of those are obtained.

- 3rd Step: The output obtained from each DBN are equally weighted and then aggregated to obtain $Output_1$.

When exogenous input variables are processed through T-Copula, the Gumbel-Hougaard Copula computes the upper tail dependence between energy load demand and the four exogenous input variables (e.g., dry bulb temperature, wet bulb temperature, dew point temperature, and humidity).

- 1st Step: First, we will start with computing upper tail dependence correlation parameter $\lambda^u = [\lambda_1, \lambda_2, \lambda_3, \lambda_4]$ and the tandem parameters i.e., VaR_1, VaR_2, VaR_3, VaR_4 for each of the variables. Then the peak load indicative variable for each of the exogenous variables are determined from VaR_1, VaR_2, VaR_3, VaR_4.
- 2nd Step: Each of the DBN models are pre trained with the load demand, correlation parameter and peak load indicative variable. The forecasting result obtained for each of the exogenous variables.
- 3rd Step: The output obtained from each DBN are equally weighted and then aggregated to obtain $Output_2$.

IV. DESIGN STEPS OF THE PROPOSED METHOD

A. LOAD DEMAND TIME SERIES SIGNAL DECOMPOSITION

There are several signal decomposition methods e.g., traditional wavelet transform, discrete wavelet transform, EMD. Compared to traditional wavelet transform EMD is highly preferable due to its applicability for non-stationary and non-linear time series. However, there are some problems (such as end effect, envelope fitting) that needs to be controlled in EMD [34]. IEMD is a modification of traditional EMD which is done by: (i) incorporating linear extrapolation to determine the end extremes so that the fitting envelope contain the given dataset, and (ii) employing nonuniform rational B-spline curve fitting envelope instead of cubic spline curve for processing complex signal [35], [36]. For clarification, first we presented the traditional EMD algorithm and it's issues as below [34]:

1) TRADITIONAL EMPIRICAL MODE DECOMPOSITION

EMD is an iterative shifting process which decomposes a signal into some regular low frequency components with different amplitude. The low frequency components include intrinsic mode functions (IMFs) and a residual function. The properties of the IMFs are given below:

(i) For each of the single IMF, the number of extrema and zero crossing throughout the whole length should be equal or differ by at most one.

(ii) At any data location, the mean value of the envelope defined by local extrema is zero.

In order to satisfy those two properties, the iterative shifting process for extracting IMFs from a given signal $E(t)$ is described below:

(i) Initially the local maxima ($E_{max}(t)$) and local minima ($E_{min}(t)$) of electricity load demand time series $E(t)$ are determined which are connected to construct upper and lower envelope with the help of cubic spline curve.

(ii) Then the difference between the mean of two envelopes and original load demand time series is determined. If the average of the upper and lower envelope is denoted as $g_1(t)$, and the difference between $E(t)$ & $g_1(t)$ is defined as $d_1(t)$ then,

$$d_1(t) = E(t) - g_1(t) \tag{1}$$

In order to be an IMF, the $d_1(t)$ must obey the properties of IMF as mentioned above. Whenever $d_1(t)$ satisfies the conditions of IMF, then it is selected as first IMF $I_1(t)$. Else, the above steps are iteratively repeated.

(iii) In the next step, the first IMF is subtracted from original electricity load demand time series to determine the residue $r_1(t)$,

$$r_1(t) = E(t) - I_1(t) \tag{2}$$

(iv) Now the residue $r_1(t)$ is considered as new data subject to the shifting process as described above. Repeat the above process until the residue time series $r_1(t)$ is a monotone function i.e., residue data is small enough so that there is no turning point.

(v) By using the EMD, the original electricity load can be expressed as follows:

$$E(t) = \sum_{i=1}^{N} I_i(t) + r_n(t) \tag{3}$$

Following this iterative shifting process, the data can be represented by IMFs and a residual function.

2) ISSUE'S WITH TRADITIONAL EMPIRICAL MODE DECOMPOSITION

Even though EMD decomposes complex time series more efficiently than other traditional decomposition techniques (e.g., wavelet transform or discrete wavelet transform), but EMD is associated with the following issues [34], [35]:

(i) End effect of traditional EMD will cause divergent phenomena for both ends of the data. The end extremes of signal cannot be determined to be a maximum or a minimum. It makes the envelope distorted and affects the EMD decomposition. For example, once the first decomposed component is faulty, the latter decomposition will show the same results distortion. Thus, the obtained IMFs are not appropriate enough [35]. On the other hand, serious end effect will appear in the Hilbert transform of IMF which will form a spectral leakage. To enable the Hilbert spectrum and to reflect the characteristics of the original signals, we must suppress this issue effectively.

(ii) Cubic spline fitting associated with traditional EMD will result in overshoot and undershoot phenomena. Thus the resulted envelope is not complete and consequently reflected into the extracted IMF.

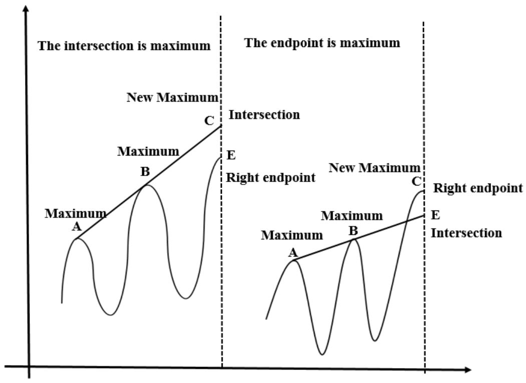

FIGURE 2. Determining the maxima of endpoint [35].

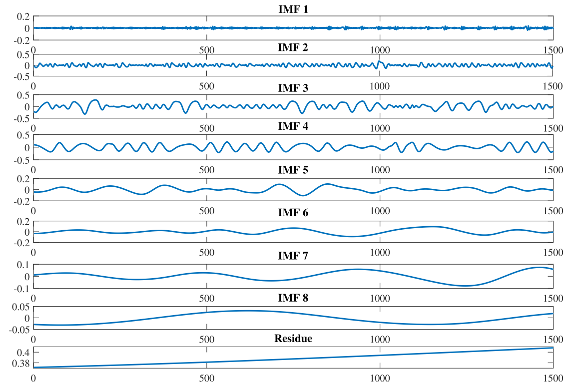

FIGURE 3. Signal decomposition using IEMD.

3) IMPROVED EMPIRICAL MODE DECOMPOSITION

To control the end effect and envelop fitting limitation of traditional EMD, we have investigated a recently proposed signal decomposition technique i.e., IEMD [35], [36]. By employing IEMD, we will suppress both end effect and envelope fitting limitations in the following way:

(i) Suppressing the end effect: In order to suppress the end effect and to achieve a real and effective decomposition, in this paper we have incorporated linear extrapolation to determine extreme ends of a signal so that the fitted envelope contain the given dataset. To make a complete envelope which will contain the entire signal data, we must make a deal with the endpoint of the signal. For more details, interested readers can look into [35].

The process by which this method determines the endpoints for upper envelope fitting is shown in Figure 2. Two maxima, A and B, are closest to an end. The straight line AB is linearly extended to the end point C. If point C is smaller than the endpoint value E of the signal, the point E is considered as a new maximum for the upper envelope fitting. Otherwise, if C is larger than the endpoint value E, point E is considered as a new maximum for the upper envelope intersection. Conversely we can determine the endpoints for lower envelope fitting.

(ii) Suppressing the envelope fitting limitation: The original EMD algorithm proposed by Huang had used cubic spline function to fit upper and lower envelope of the signal and then calculated the mean of the fitted upper & lower envelope. Because the power is low and easy to calculate, cubic spline curve fitting is simpler than others; however, the cubic spline fitting will cause the overshoot and undershoot phenomena, so that the envelope fitting deviates from the actual signal envelope and develop a incomplete envelope. In order to solve the overshoot and undershoot problem of cubic spline curve fitting, many researchers has proposed improvement method, such as high order spline function method, polynomial fitting, and piecewise power function interpolation method. These methods can solve the problem of the fitting overshoot or fitting undershoot based on their own characteristics [35].

In this paper, a nonuniform rational B-spline fitting method is used to fit the upper and lower envelope of signal, resulting in the mean envelope. We use the accumulative chord length parameterized algorithm to achieve BNURBS curve fitting. The same simulation signal uses nonuniform rational B-spline (NURBS) curve fitting envelope compared with fitting envelope by cubic spline function. After employing IEMD, the simulation result of signal decomposition is shown in Figure 3. With the decomposition results, it is obvious that IEMD algorithm can decompose the signal into different frequency components, and there is no mode mixing.

B. T-COPULA ANALYSIS

Preliminary research indicates that, there is a upper tail dependence between power load and exogenous input variables. In this research, the Gumbel-Hougaard copula model computes the upper-tail dependence between the power load and the four exogenous input variables. The classical bivariate Gumbel-Hougaard model can be defined as,

$$f(x_1(t), E(t)) = CP[f_{x_1}(x_1(t)), f_E(E(t))] \tag{4}$$

here $f_{x_1}(x_1(t))$ and $f_E(E(t))$ denotes the marginal cumulative distribution functions; x_1 represents one of the exogenous input variables, E denotes system load demand, $f(x_1, x_2)$ is the two dimensional joint distribution function; and $CP(x_1, E)$ is the Copula function. Now we need to determine the upper tail dependence parameter for each of the exogenous variables in the following way,

$$CP(x_1, E) = \exp\{-[(-\ln x_1)^{\alpha} + (-\ln E)^{\alpha}]^{1/\alpha}\} \tag{5}$$

The maximum likelihood method can be used to determine the copula model's parameter α. For the nonlinear relationship of system load demand and exogenous input variables, we adopt the Canonical Maximum Likelihood (CML) method that is implemented based on the empirical CDF of samples. The objective of the CML is expressed by:

$$\hat{\alpha} = \arg\min - \sum_{t=1}^{N} \ln f(x_1(t), E(t)) \tag{6}$$

here N indicates the number of exogenous input variables. Now upper-tail dependence parameter λ^1 of

TABLE 1. Pearson correlation matrix for correlation analysis between system load and input exogenous variables.

Correlation analysis system load and exogenous varibles					
	Dry bulb Temp.	Dew point Temp.	Wet bulb Temp.	Humidity	System Load
Dry bulb Temp.	1.00	0.64	0.89	-0.25	0.10
Dew point Temp.	0.64	1.00	0.91	0.56	-0.11
Wet bulb Temp.	0.89	0.91	1.00	0.20	-0.02
Humidity	-0.25	0.56	0.20	1.00	-0.27
System Load	0.10	-0.11	-0.02	-0.27	1.00

Gumbel-Hougard Copula is given by,

$$\lambda^1 = 2 - 2^{1/\alpha} \tag{7}$$

Following this method we can determine our desired copula parameter for each of the exogenous input variables. Due to the variety of fluctuations and spikes of power load data, an effective statistical estimation of the peak load is crucial. In this research work, a threshold parameter called VaR is introduced to determine the peak load indicative variable for each of the variables. The computed peak load indicative variables based on VaR helps to increase the load forecasting accuracy during peak load time. In our work, since exogenous input variables are stochastic and have impact on power load, we have determined the VaR from the following formula,

$$VaR_p^1 = CP^{-1}[f(x_1(t), E(t))] \tag{8}$$

here VaR_p^1 represents the pth upper percentile of bivariate distribution of exogenous input variable and system load. Hence, the binary value of peak load indicative variable is determined from the following formula,

$$M(x_1) = \begin{cases} 1, & \text{if } x_1(t) \geq VaR_p^1 \\ 0, & \text{if } x_1(t) < VaR_p^1 \end{cases} \tag{9}$$

here $M(x_1)$ represents the peak load indicative variable for one of the exogenous variables x_1 and the value of p is set as 0.95. For our research work, we will repeat this process for each of the exogenous input variables i.e., we need to do this calculation four times for four exogenous input variables.

In this research, the Gumbel-Hougaard Copula models fit the upper-tail dependence between system load versus exogenous weather variables. The default value of significance is set as 0.05 and the model parameters are estimated through maximum likelihood estimation. The upper-tail dependence parameter between the system load and the exogenous weather variables for case study #1 have been shown in Figure 4. The tail distribution of dry bulb temperature shows strong correlation between the points (0, 0) and (1, 1) in the tail. In other words, dry bulb temperature has great influence on system load. Other tail distribution also shows strong correlation. To get idea about positive or negative correlation of system load with exogenous variables, in this paper we have added the Pearson correlation matrix as in Table 1.

This correlation matrix gives the idea that, positive correlation of system load with dry bulb temperature and humidity.

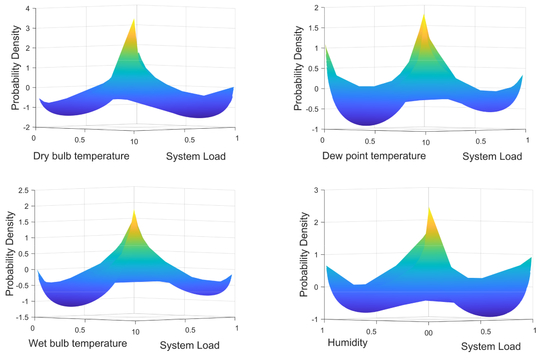

FIGURE 4. Correlation analysis of exogenous variables.

TABLE 2. Upper tail dependence parameter and VaR of system load with exogenous variables.

Exogenous variable	Upper-tail parameter		VaR	
	Case study #1	Case study #2	Case study #1	Case study #2
Dry bulb Temp.	3.12	3.58	98	94
Dew point Temp.	1.69	1.48	46	38
Wet bulb Temp	1.77	1.99	59	67
Humidity	2.34	2.78	83	89

Conversely, for other two exogenous variables there exists negative correlation. Then, the calculation of VaR is done at 95th percentage of significance level. The value of upper tail dependence parameter and VaR for both the case study is given in Table 2.

C. LEARNING WITH DEEP BELIEF NETWORK

A divide and conquer algorithm works by recursively breaking down a problem into two or more sub-problems of the same (or related) type, until these become simple enough to be solved directly. The solutions to the sub-problems are then combined to give a solution to the original problem. In the proposed method, the load demand data is decomposed into several IMFs and one residue by IEMD. During the correlation analysis we get upper tail dependence parameter, peak load indicative variable. The data from IMF, residue, upper tail dependence parameter, peak load indicative variable and system load are applied to DBN. A DBN has a unsupervised subpart which consists number of RBMs and a supervised part which is logistic regression layer i.e., ANN. Therefore learning with DBN is semi-supervised learning.

The DBN proposed by [38] provides a new way to train deep generative models, which is called layer-wise greedy pre-training algorithm. Figure 5 shows the architecture of a DBN. There is no inter-connection between units in each layer. An restricted Boltzmann machine (RBM) is a neural network which can learn the probability distribution over the input dataset. The DBN pre-training procedure treats each consecutive pair of layers in the MLP as a RBM [39] whose joint probability is defined as,

$$P_{h|v}(h|v) = \frac{1}{Z_{h,v}} * e^{(v^T Wh + v^T b + a^T h)} \tag{10}$$

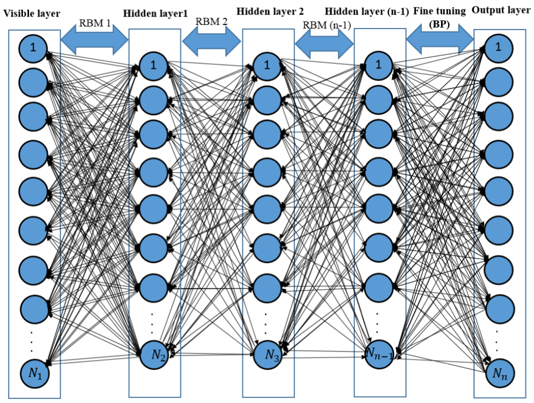

FIGURE 5. Deep belief network architecture.

here, h represents input applied to hidden layer, v represents output obtained from visible layer, W represents hidden neuron weights, and a represents activation. For each RBM there is pair of hidden layer and visible layer. For the BernoulliâĂ"Bernoulli RBM applied to binary v with a second bias vector b and normalization term $Z_{h,v}$, and

$$P_{h|v}(h|v) = \frac{1}{Z_{h,v}} * e^{(v^T Wh+(v-b)^T (v-b)+a^T h)} \tag{11}$$

for the GaussianâĂ"Bernoulli RBM applied to continuous variable v [40]. In both cases the conditional probability $P_{h|v}(h|v)$ has the same form as that in an MLP layer.

The objective function of an RBM is,

$$L(a, b, W) = \Sigma \log P_{h|v}(h|v) \tag{12}$$

The layer-wise pre-training method requires the DBN to be pre-trained following a stochastic gradient descent method on the objective function. The gradient method indicates that, the parameters (*e.g.*, a, b, W) are updated based on the gradients of the objective function (12). The gradients of the probability distribution function can be expressed in the following way,

$$\frac{\partial P_{h|v}(h|v)}{\partial W_{j,i}} = \langle v_i h_i \rangle_{P_{h|v}(h|v)} - \langle h_i v_i \rangle_{recon} \tag{13}$$

$$\frac{\partial P_{h|v}(h|v)}{\partial a_i} = \langle v_i \rangle_{P_{h|v}(h|v)} - \langle v_i \rangle_{recon} \tag{14}$$

$$\frac{\partial P_{h|v}(h|v)}{\partial b_i} = \langle h_i \rangle_{P_{h|v}(h|v)} - \langle h_i \rangle_{recon} \tag{15}$$

here $\langle h_i \rangle_{P_{h|v}(h|v)}$ is the expectation of the conditional distribution with respect to the input raw data; $\langle h_i v_i \rangle_{recon}$ is the expectation of the i^{th}-step reconstructed distribution. We can use contrastive divergence [41] to obtain the expectation of the reconstructed distribution through alternating Gibbs sampling. Later, we used the following updating formulas,

$$W_{i+1} = W_i + \eta(\langle v_i h_i \rangle_{P_{h|v}(h|v)} - \langle h_i v_i \rangle_{recon}) \tag{16}$$

$$a_{i+1} = a_i + \eta(\langle v_i \rangle_{P_{h|v}(h|v)} - \langle v_i \rangle_{recon}) \tag{17}$$

$$b_{i+1} = b_i + \eta(\langle h_i \rangle_{P_{h|v}(h|v)} - \langle h_i \rangle_{recon}) \tag{18}$$

To train multiple layers, one trains the first layer, freezes it, and uses the conditional expectation of the output as the input to the next layer and continues training next layers. Based on the layer wise pre-training approach, all the parameters of the DBN algorithm are initialized. Hinton and many others have found that initializing MLPs with pretrained parameters never hurts and often helps [38], [42]. Adjustment of these parameters in a supervised manner is conducted until the loss function the DBN reaches its minimum. Finally, back-propagation algorithm is applied for the fine-tuning process. All parameters are updated from the top to bottom which gives reduced forecasting errors.

Due to the influence from climate and social activities, the electricity load data shows three main nest cycles: daily, weekly and yearly. To identify cycles and patterns in load demand time series data, autocorrelation function (ACF) can be applied as a guidance for informative feature subset selection [43]. Suppose a time series data set is given as $E = E_t : t \in T$, where T is the index set. The lag k autocorrelation coefficient r_k can be computed by:

$$r_k = \frac{\sum_{t=k+1}^{n} (E_t - \overline{E})(E_{t-k} - \overline{E})}{\sum_{t=1}^{n} (E_t - \overline{E})^2} \tag{19}$$

where $\overline{E}$ is the mean value of all E in the given time series, r_k measures the linear correlation of the time series at times t and k.

V. SIMULATION RESULTS AND ANALYSIS

A. DESCRIPTION OF THE DATASET

The proposed hybrid load forecasting model is validated on the Australian Energy Market Operator (AEMO) data [44] and the dataset for one of urbanized regions of Houston, Texas, USA [46]. Specifically, the dataset include three main groups of measured variables: weather data (i.e., dry bulb temperature, wet bulb temperature, dew point temperature, and humidity), time categorical data (i.e., hour, month, day), social data (i.e., working day, weekend, holiday), and energy load demand for specific sampling time.

B. PERFORMANCE EVALUATION CRITERIA

The performance of the proposed load forecasting models are compared with respect to mean absolute percentage error (MAPE) and root mean square error (RMSE) [19], [34].

1) MAPE is defined as,

$$MAPE = (1/N) * \sum_{t=1}^{N} \frac{| E(t) - \hat{E}(t) |}{| E(t) |} * 100 \tag{20}$$

here $E(t)$ denotes actual load demand and $\hat{E}(t)$ denotes the forecasted load demand.

2) RMSE is defined as,

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} (E(t) - \hat{E}(t))^2} \tag{21}$$

The values of MAPE and RMSE provides the idea about forecasting accuracy. The less the values of MAPE and RMSE means higher forecasting accuracy.

C. LEARNING ENVIRONMENT SETUP

For signal decomposition based DBN, the day ahead load forecasting takes load demand of whole previous day E_{t-48} to E_{t-96} and the same day in the previous week E_{t-336} to E_{t-348}. Other than this decomposed signal, auto correlation between load demand and previous hour load demand (i.e., load demand at half hour before, and load demand at same day same time in previous week). Including the two bias input, the total number of inputs to DBN is 100. For the correlation analysis of exogenous variables, load demand of whole previous day E_{t-48} to E_{t-96} and the same day in the previous week E_{t-336} to E_{t-348} is applied as input to the DBN. In addition, upper tail dependence correlation parameter and peak load indicative variable is also applied as input to DBN. Similarly including the bias input, in this case also there are 100 inputs node to DBN. For both cases, the number of hidden layer is 3 and in each layer the number of hidden neuron is 30. These two parameters are determined from cross validation. Therefore, the structure of DBN is $100 - 30 - 30 - 30 - 1$. From the cross validation result we have determined the learning rate as 0.1 and number of iteration is 500. In order to perform the cross validation we have randomly selected 10% of raw data. For correlation analysis through T-Copula we have default threshold of significance as 0.05.

D. EXPERIMENTAL RESULTS

All the simulations are conducted using Matlab R2017b on a standard PC. The results are validated for two case study result. The dataset for two case study are: (i) AEMO data, Autsralia and (ii) Dataset of Houston, Texas, USA. For both of the case study, we have linearly scaled the dataset into [0, 1] using the following formula [33],

$$\bar{E_i} = \frac{E_{max} - E_i}{E_{max} - E_{min}} \tag{22}$$

(1) Case Study #1: In this case study we have collected dataset from AEMO [44] and [45]. The data collection date is from 1st January 2013 to 31st December 2013 with sampling time of half hour. We have divided the whole year dataset into four seasons: (i) January to March, (ii) April-June, (iii) July-September, and (iv) October-December. During the training time we have followed the simulation set up as given in learning environment set up subsection. Following this method for week ahead load forecasting we have considered the three week dataset of a month as training dataset and remaining week as the testing dataset. But here we assumed that we have the information of day-type i.e., the working day or holiday, time of day for avoiding uncertainty due to volatile nature of electricity load and exogenous input variables. The input dataset obtained from signal decomposition are eight IMFs and residual signal. With auto lag correlation, these decomposed signals are applied to DBN for load forecasting.

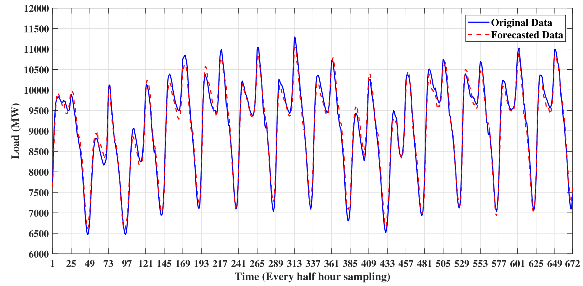

FIGURE 6. Two week-ahead load forecasting result.

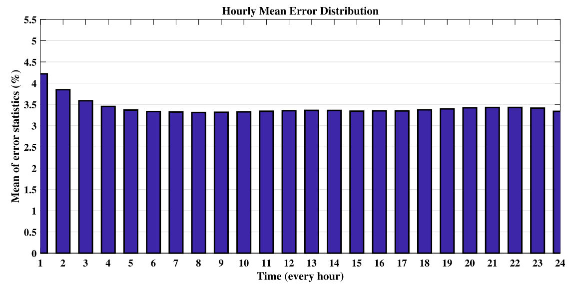

FIGURE 7. Hourly error distribution.

The input dataset obtained from correlation analysis includes upper tail dependence parameter, binary peak indicative variable and load demand dataset according to learning environment setup are applied to DBN for load forecasting due to exogenous variable. For a fair comparison, during each month of 2013 we have considered the three weeks load demand dataset as training dataset and remaining week is as testing dataset. This means our target is to forecast one-week load demand. For each of the season, we have considered two months dataset to evaluate the performance of load forecasting from our proposed method. Thus, within two months we have forecasted load demand for two weeks. The prediction result of the proposed model is compared with [33]. Notice that, here we have trained the DBN with batch of data of same time and similar day i.e., working day or weekend.

Now, we have forecasted load demand from signal decomposition and correlation analysis. For making final prediction i.e., forecasted load demand we have aggregated the results from signal decomposition and correlation analysis. We have considered the equal weighted average to determine the final forecasted load demand. The forecasted load demand from the proposed model is shown in Figure 6. The simulation result presented in Figure 6 is carried out in New South Wales (NSW), Australia during the month January- March 2013. Dataset of year 2013 is considered for comparison with [33]. We have presented the mean error distribution at every hour as shown in Figure 7. This error distribution is presented to show the load forecasting accuracy improvement during peak load time. From the mean error distribution result it is evident that, there is a improvement in load forecasting accuracy during peak time and this will help the utility operators to make proper generation scheduling and distribution maintenance planning. And for comparison with results pre-

TABLE 3. Load forecasting performance comparison: Case study 1, location five regions of Australia.

Month	Algorithm	NSW		TAS		QLD		VIC		SA	
		MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE	RMSE
Jan.-March	NN [34]	6.16	587.82	6.39	91.56	4.85	409.51	8.56	759.38	12.97	225.75
	DBN [34]	6.05	633.11	6.18	86.24	4.53	348.71	6.26	465.28	11.02	202.64
	EMD-DBN [34]	4.62	541.53	4.05	56.10	2.56	191.22	8.86	762.57	10.04	238.09
	Coupla-IEMD-DBN	**3.44**	**392.44**	**2.98**	**39.81**	**2.28**	**186.90**	**6.96**	**598.76**	**7.58**	**181.91**
April-June	NN [34]	6.64	743.44	8.27	123.11	5.65	399.51	9.06	569.34	13.76	201.87
	DBN [34]	6.44	699.17	7.32	102.71	5.27	369.54	6.44	426.41	11.21	182.81
	EMD-DBN [34]	3.22	377.63	5.80	85.13	2.93	243.68	4.35	321.59	6.76	125.31
	Coupla-IEMD-DBN	**2.41**	**314.03**	**4.34**	**66.42**	**2.39**	**201.32**	**3.69**	**276.51**	**5.14**	**98.72**
July-Sept.	NN [34]	7.64	732.24	8.70	161.28	5.38	372.26	8.39	728.06	14.19	381.68
	DBN [34]	5.17	480.69	6.48	119.53	5.11	357.87	7.85	546.61	11.43	223.44
	EMD-DBN [34]	3.08	322.04	4.93	73.91	2.08	142.84	3.83	285.45	9.60	192.74
	Coupla-IEMD-DBN	**2.42**	**271.21**	**3.84**	**58.32**	**1.68**	**118.36**	**3.41**	**228.89**	**7.54**	**158.08**
Oct.-Dec.	NN [34]	7.88	796.73	6.89	165.8	5.44	374.92	7.27	520.11	13.86	391.23
	DBN [34]	6.62	785.3	5.96	95.41	5.53	388.71	6.88	561.05	11.66	386.82
	EMD-DBN [34]	2.71	282.34	4.75	68.26	2.88	219.19	3.73	322.91	8.11	192.74
	Coupla-IEMD-DBN	**2.18**	**224.52**	**3.78**	**56.56**	**198.56**	**3.12**	**3.12**	**291.35**	**5.86**	**154.32**

sented in [33], we have done simulation for all of the regions of Australia as given in Table 3. As seen in Table 3, error in load forecasting results i.e., MAPE and RMSE values of the proposed model are lower than the other comparative models in [33]. The MAPE values of the proposed model are decreased by 21.19%, and the RMSE values decreased by 16.93% compared to [33]. The reason of performance improvement is due to : (i) IEMD signal decomposition, and (ii) T-Copula correlation analysis. IEMD improves the signal decomposition efficiency and T-Copula contributes to improve the load forecasting accuracy during peak time by computing peak load indicative variables from VaR.

(2) Case Study #2: For this case study we have collected the dataset from urbanized area of Houston, Texas, USA [46], [47]. The data collection date is from 1st January 2016 to 31st December 2016 with sampling time of one hour. We have divided the whole year dataset into four seasons: (i) January to March, (ii) April-June, (iii) July-September, and (iv) October-December. During the training time we have considered learning environment setup information. Following the similar procedure as mentioned for case study #1, for this case study we have again considered the three week dataset of a month as training dataset and remaining week as the testing dataset. The input dataset obtained from signal decomposition are eight IMFs and residual signal. With auto lag correlation, these decomposed signals are applied to DBN for load forecasting. The input dataset obtained from correlation analysis includes upper tail dependence parameter, binary peak indicative variable and load demand dataset. According to learning environment setup these inputs are applied to DBN for load forecasting due to exogenous variables.

The prediction result of the proposed model is compared with [37] and the results are presented in Table 4. As seen in Table 4, all MAPE and RMSE values of the proposed model are lower than the traditional EMD based STLF model. The MAPE values of the proposed model are decreased by 15.27%, and the RMSE values decreased by 13.86% compared to [37]. This significant decrease in MAPE and

TABLE 4. Load forecasting performance comparison: Case Study 2, Location Houston, Texas, USA.

		Location: Houston	
Month	Algorithm	MAPE	RMSE
Jan.-March	NN [[38]]	7.37	2521.19
	DBN [[38]]	6.99	2483.34
	Copula-DBN [[38]]	6.08	2263.61
	Coupla-IEMD-DBN	**4.11**	**2014.18**
April-June	NN [[38]]	8.16	1593.68
	DBN [[38]]	7.78	1479.97
	Copula-DBN [[38]]	6.63	1388.84
	Coupla-IEMD-DBN	**4.62**	**1325.30**
July-Sept.	NN [[38]]	7.19	2230.02
	DBN [[38]]	6.88	2020.01
	Copula-DBN [[38]]	6.21	1917.42
	Coupla-IEMD-DBN	**3.98**	**1940.68**
Oct.-Dec.	NN [[38]]	8.25	2213.67
	DBN [[38]]	7.99	2203.74
	Copula-DBN [[38]]	7.15	2110.45
	Coupla-IEMD-DBN	**5.46**	**1856.86**

RMSE values resulted from the combined effect of IEMD and T-Copula. These two method enables our proposed hybrid model for processing of more information.

VI. SUMMARY

This paper proposes a novel hybrid STLF model. First, load demand time series is decomposed by IEMD. Second, correlation analysis between system load and exogenous input variables are incorporated to increase the load forecasting accuracy during peak time. Third, the two components are predicted separately by the suitable model. Last, each component's forecasting results are added up to obtain the final forecasting results. Electricity load data from Australia and Texas electricity markets are used to validate the effectiveness of the proposed model. All case study results indicate that the proposed model improves the forecasting accuracy. Three facts emerge clearly from the results: (1) the linear and nonlinear component of electricity load can be extracted more accurately and effectively by the IEMD, (2) the peak load indicative variable computed from VaR through T-Copula model improves the load forecasting accuracy during peak time, (3) the DBN has a strong ability to fit the nonlinear component of the original electricity load. By using each model's advantage, the hybrid model can capture the different characteristics associated with electricity load. Therefore, the proposed model can provide a robust, stable and more accurate prediction results. Advanced models can be used to select suitable input variables for electricity load forecasting in the future. Besides, some other future influencing factors such as information of consumer related to incentive based demand response program, and uncertainty from distributed renewable energy integration can be added in the hybrid model as future research. This framework can be beneficial for practical short-term generation scheduling and operations for the grid network.

REFERENCES

[1] K. Chen, K. Chen, Q. Wang, Z. He, J. Hu, and J. He, "Short-term load forecasting with deep residual networks," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3943–3952, Jul. 2019.

[2] T. Roy, A. Das, and Z. Ni, "Optimization in load scheduling of a residential community using dynamic pricing," in *Proc. IEEE Power Energy Soc. Innov. Smart Grid technol. Conf. (ISGT)*, Apr. 2017, pp. 1–5.

[3] D. W. Bunn and E. D. Farmer, *Comparative Models for Electric Load Forecasting*, 1st ed. New York, NY, USA: Wiley, 1985.

[4] G. Černe, D. Dovzan, and I. Skrjanc, "Short-term load forecasting by separating daily profiles and using a single fuzzy model across the entire domain," *IEEE Trans. Ind. Electron.*, vol. 65, no. 9, pp. 7406–7415, Sep. 2018.

[5] B. Koo, M.-S. Kim, K. H. Kim, H.-T. Lee, J. H. Park, and C. Kim, "Short-term electric load forecasting using data mining technique," in *Proc. 7th Int. Conf. Intell. Syst. Control (ISCO)*, Jan. 2013, pp. 153–157.

[6] S. Fan and R. J. Hyndman, "Short-term load forecasting based on a semi-parametric additive model," *IEEE Trans. Power Syst.*, vol. 27, no. 1, pp. 134–141, Feb. 2012.

[7] C.-N. Ko and C.-M. Lee, "Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended Kalman filter," *Energy*, vol. 49, pp. 413–422, Jan. 2013.

[8] A. Al-Kandari, S. Soliman, and M. E. El-Hawary, "Fuzzy short-term electric load forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 26, no. 2, pp. 111–122, Feb. 2004.

[9] S.-J. Huang and K.-R. Shih, "Short-term load forecasting via ARMA model identification including non-Gaussian process considerations," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 673–679, May 2003.

[10] C.-M. Lee and C.-N. Ko, "Short-term load forecasting using lifting scheme and ARIMA models," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5902–5911, May 2011.

[11] C.-M. Huang, C.-J. Huang, and M.-L. Wang, "A particle swarm optimization to identifying the ARMAX model for short-term load forecasting," *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 1126–1133, May 2005.

[12] A. Kavousi-Fard, H. Samet, and F. Marzbani, "A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting," *Expert Syst. Appl.*, vol. 41, no. 13, pp. 6047–6056, 2014.

[13] H. J. Sadaei, R. Enayatifar, A. H. Abdullah, and A. Gani, "Short-term load forecasting using a hybrid model with a refined exponentially weighted fuzzy time series and an improved harmony search," *Int. J. Electr. Power Energy Syst.*, vol. 62, pp. 118–129, Nov. 2014.

[14] L. Hernández, C. Baladrón, J. M. Aguiar, B. Carro, A. Sánchez-Esguevillas, and J. Lloret, "Artificial neural networks for short-term load forecasting in microgrids environment," *Energy*, vol. 75, pp. 252–264, Oct. 2014.

[15] P. Li, Y. Li, Q. Xiong, Y. Chai, and Y. Zhang, "Application of a hybrid quantized Elman neural network in short-term load forecasting," *Int. J. Electr. Power Energy Syst.*, vol. 55, pp. 749–759, Feb. 2014.

[16] A. S. Khwaja, M. Naeem, A. Anpalagan, A. Venetsanopoulos, and B. Venkatesh, "Improved short-term load forecasting using bagged neural networks," *Electr. Power Syst. Res.*, vol. 125, pp. 109–115, Aug. 2015.

[17] C. Cecati, J. Kolbusz, P. Różycki, P. Siano, and B. M. Wilamowski, "A novel RBF training algorithm for short-term electric load forecasting and comparative studies," *IEEE Trans. Ind. Electron.*, vol. 62, no. 10, pp. 6519–6529, Oct. 2015.

[18] L. Wang, Y. Zeng, and T. Chen, "Back propagation neural network with adaptive differential evolution algorithm for time series forecasting," *Expert Syst. Appl.*, vol. 42, no. 2, pp. 855–863, 2015.

[19] R. Zhang, "Short-term load forecasting of Australian national electricity market by an ensemble model of extreme learning machine," *IET Gener., Transmiss. Distrib.*, vol. 7, no. 6, pp. 391–397, Apr. 2013.

[20] G. Sudheer and A. Suseelatha, "Short term load forecasting using wavelet transform combined with Holt-Winters and weighted nearest neighbor models," *Int. J. Electr. Power Energy Syst.*, vol. 64, pp. 340–346, Jan. 2015.

[21] Y. He, R. Liu, H. Li, S. Wang, and X. Lu, "Short-term power load probability density forecasting method using kernel-based support vector quantile regression and Copula theory," *Appl. Energy*, vol. 185, pp. 254–266, Jan. 2017.

[22] S. B. Kyung, Y.-S. Beak, D. H. Hong, and G. Jang, "Short term load forecasting for the holidays using fuzzy linear regression method," *IEEE Trans. Power Syst.*, vol. 1, no. 1, pp. 96–101, Feb. 2005.

[23] J. Che and J. Wang, "Short-term load forecasting using a kernel-based support vector regression combination model," *Appl. Energy*, vol. 132, pp. 602–609, Nov. 2014.

[24] P. Jiang, F. Liu, and Y. Song, "A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short-term load forecasting," *Energy*, vol. 119, pp. 694–709, Jan. 2017.

[25] F. Chahkoutahi and M. Khashei, "A seasonal direct optimal hybrid model of computational intelligence and soft computing techniques for electricity load forecasting," *Energy*, vol. 140, pp. 988–1004, Dec. 2017.

[26] L. Xiao, W. Shao, M. Yu, J. Ma, and C. Jin, "Research and application of a combined model based on multi-objective optimization for electrical load forecasting," *Energy*, vol. 119, pp. 1057–1074, Jan. 2017.

[27] L. Xiao, W. Shao, C. Wang, K. Zhang, and H. Lu, "Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting," *Energy*, vol. 180, pp. 213–233, Oct. 2017.

[28] S. Li, L. Goel, and P. Wang, "An ensemble approach for short-term load forecasting by extreme learbing machine," *Appl. Energy*, vol. 170, pp. 9–22, May 2016.

[29] M. Ghofrani, M. Ghayekhloo, A. Arabali, and A. Ghayekhloo, "A hybrid short-term load forecasting with a new input selection framework," *Energy*, vol. 170, pp. 86–777, Mar. 2017.

[30] G. F. Fan, S. Qing, H. Wang, W. C. Hong, and H. J. Li, "Support vector regression model based on empirical mode decompsition and auto regression for electric load forecasting," *Energies*, vol. 6, pp. 901–1887, Apr. 2016.

[31] T. Xiong, Y. Bao, and Z. Hu, "Interval forecasting of electricity demand: A novel bivariate EMD-based support vector regression modeling framework," *Int. J. Elect. Power Energy Syst.*, vol. 63, pp. 353–362, Dec. 2014.

[32] N. Liu, Q. Tang, J. Zhang, W. Fan, and J. Liu, "A hybrid forecasting model with parameter optimization for short-term load forecasting of micro-grids," *Appl. Energies*, vol. 129, pp. 45–336, Sep. 2014.

[33] X. Qiu, Y. Ren, P. N. Suganthan, and G. Amaratunga, "Empirical mode decomposition based ensemble deep learning for load demand time series forecasting," *Appl. Soft Comput.*, vol. 54, pp. 246–255, May 2017.

[34] N. E. Huang, Z. Shen, and S. R. Long, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Math. Phy. Eng. Sci*, vol. 454, pp. 95–903, Mar. 1998.

[35] X. Yang, G. Cheng, and H. Liu, "Improved empirical mode decomposition algorithm of processing complex signal for IoT application," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 10, 2015, Art. no. 862807.

[36] Z. Yang and L. Yang, "An improved empirical mode decomposition," in *Proc. 2nd Int. Congr. Image Signal Process.*, 2009, pp. 1–5.

[37] T. Ouyang, Y. He, H. Li, Z. Sun, and S. Baek, "A deep learning framework for short-term power load forecasting," 2017, *arXiv:1711.11519*. [Online]. Available: https://arxiv.org/abs/1711.11519

[38] G. Hinton, S. Osindero, and Y. W. Teh, "A Fast Learning Algorithm for Deep Belief Networks," *Neural Comput.*, vol. 18. no. 7, pp. 1527–1554, 2006.

[39] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[40] T. Yamashita, M. Tanaka, E. Yoshida, Y. Yamauchi, and H. Fujiyoshi, "To be Bernouli or to be Gaussian for a restricted Boltzman machine," in *Proc. 22nd Int. Conf. Pattern Recognit. Sci.*, 2014, pp. 1520–1525.

[41] G. E. Hinton, "Training products of experts by minimizing contarstive divergence," *Neural Comput.*, vol. 14, no. 8, pp. 1771–1800, 2002.

[42] G. E. Hinton, "A practical guide to training restricted Boltzman machine," *Neural Networks: Tricks of the Trade*, 2nd Ed., Berlin, Germany:Springer, 2012, pp. 599–619.

[43] 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.
[44] (2013). *AEMO, Australian Energy Market Operator 2013*. [Online]. Available: http://www.aemo.com.au/
[45] (2013). *Historical weather observations and statistics, Bureau of Meteorology Australian*. [Online]. Available: http://http://www.bom.gov.au/climate/data-services/
[46] (2016). *ERCOT Hourly Load Data*. [Online]. Available: http://www.ercot.com/gridinfo/load/load_hist
[47] (2013). *Weather Underground*. [Online]. Available: http://www.wunderground.com/history/daily/us/tx/houston/KHOU/date/2016-1-1

MD. RASHEDUL HAQ received the B.Sc. degree from the Department of Electrical and Electronic Engineering (EEE), Khulna University of University and Technology (KUET), Bangladesh, in 2010. He is currently pursuing the Master of Science degree in electrical engineering with South Dakota State University, Brookings, SD, USA.

He was served as a Lecturer and a Coordinator for EEE Department, University of Science and Technology Chittagong, from February 2011 to August 2017. His current research interests include power system data analytics, big data analytics, computational intelligence, machine learning, and power system resiliency. He has been actively involved in reviewing journal articles of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.

ZHEN NI (M'15) received the B.S. degree from the Department of Control Science and Engineering (renamed as College of Artificial Intelligence and Automation), Huazhong University of Science and Technology, Wuhan, China, in 2010, and the Ph.D. degree from the Department of Electrical, Computer, and Biomedical Engineering, The University of Rhode Island, Kingston, RI, USA, in 2015.

He was with the Department of Electrical Engineering and Computer Science, South Dakota State University, Brookings, SD, USA, from 2015 to 2019. He is currently an Assistant Professor with the Department of Computer, Electrical Engineering, and Computer Science, Florida Atlantic University, Boca Raton, FL, USA. His current research interests include computational intelligence, machine learning, smart grid, and cyber-physical systems. He received the prestigious IEEE Computational Intelligence Society Outstanding Ph.D. Dissertation Award, in 2020, International Neural Networks Society Aharon Katzir Young Investigator Award, in 2019, URI Excellence in Doctoral Research Award, in 2016, Chinese Government Award for Outstanding Students Abroad by Chinese Government, in 2014, and Second Prize of Graduate Student Poster Contest in IEEE Power and Energy Society General Meeting, in 2015. He has been actively involved in numerous conference and workshop organization committees in the society, including the General Co-Chair of the IEEE CIS Winter School, Washington, DC, USA, in 2016. He has been an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, since 2019, and the IEEE COMPUTATIONAL INTELLIGENCE MAGAZINE, since 2018. He was a Guest Editor of *IET Cyber-Physical Systems: Theory* and *Applications*, from 2017 to 2018.