Deep Learning Based Load Forecasting with Decomposition and Feature Selection Techniques

Siva Sankari Subbiah[1] and Senthil Kumar P[2]*

[1]Department of Computer Science and Engineering, Sreenivasa Institute of Technology and Management Studies, Chittoor 517 127, Andhra Pradesh, India

[2]School of Information Technology & Engineering, Vellore Institute of Technology, Vellore 632 014, Tamil Nadu, India

The forecasting of short term electricity load plays a vital role in power system. It is essential for the power system's reliable, secure, and cost-effective functioning. This paper contributes significantly for enhancing the accuracy of short term electricity load forecasting. It presents a hybrid forecasting model called Gated Recurrent Unit with Ensemble Empirical Mode Decomposition and Boruta feature selection (EBGRU). It is a hybrid model that addresses the non-stationary, non-linearity and noisy issues of the time series input by using Ensemble Empirical Mode Decomposition (EEMD). It also addresses overfitting and curse of dimensionality issues of load forecasting by identifying the pertinent features using Boruta wrapper feature selection. It effectively handles the uncertainty and temporal dependency characteristics of load and forecasts the future load using deep learning based Gated Recurrent Unit (GRU). The proposed EBGRU model is experimented by using European and Australian Electricity load datasets. The temperature has high correlation with load demand. In this study, both load and temperature features are considered for the accurate short term load forecasting. The experimental outcome demonstrates that the proposed EBGRU model outperforms other deep learning models such as RNN, LSTM, GRU, RNN with EEMD and Boruta (EBRNN) and LSTM with EEMD and Boruta (EBLSTM).

Keywords: Boruta feature selection, Electricity load prediction, Ensemble empirical mode decomposition, Gated recurrent unit, Recurrent neural network

Introduction

The electric power industry is the important industry contributing rigorously in the development of states. The short term load forecasting (STLF) is indispensable for directing the day to day activities of the society and industries. It assists the power system in making decisions regarding power generation planning, fuel purchase scheduling, resource allocation, maintenance schedule, and preparation of the electricity load dispatch schedule. The power system cannot generate large quantity of electricity in advance and store it for future use. The cost of storing the electricity is more compared to the cost of generating the electricity. So, it is necessary for the power system to forecast the future load demand in advance.[1] The power system can make the resources to be available for generating the electricity on demand economically without storing it. STLF is also very useful for the energy policy makers and power system managers for making proper decision. The

high quality, uninterrupted and stable electric energy provided by the power system relies on the accurate STLF.[2]

As load demand depends on different unstable factors like meteorological conditions and electricity price, it adds volatile, non-linearity and non-stationary characteristics to the load.[3,4] The accurate forecasting becomes the challenging task in power systems. The development of an efficient forecasting model is essential to overcome these challenges. Many researchers have devised numerous approaches to increase the accuracy of short term load forecasting over the last few decades. They are divided into two categories: traditional and intelligent methods. The traditional methods like ARIMA,[5] Kalman filtering[6], regression analysis[7], and exponential smoothing[8] are not a satisfactory methods for STLF due to an inability of handling the non-linear and multivariate data. The intelligent methods such as artificial neural network, support vector regression, expert system, random forest, and fuzzy logic have the ability to handle the non-linear and multivariate inputs, however they cannot guarantee the accurate

————————
*Author for Correspondence
E-mail: senbe@rediffmail.com

forecasting in all types of problems.[9] The performance of these models are affected due to the volatile nature, the increase of the volume of data, the number of layers, and the long term dependencies among the data.[10,11] The neural network suffers from the slow convergence and also the local minimum issues. In order to overcome these issues and improve the performance of forecasting, the combination of various methods can be used.[1,12]

This paper introduces the EBGRU hybrid model as a combination of three methods namely, feature selection, decomposition and deep neural network. The load data has a strong non-linear nature and it consists of noisy data. The decomposition methods can handle these non-linear and noisy issues by decomposing the time series load and temperature inputs into smaller components. Feature selection is an important dimensionality reduction technique that selects useful characteristics from the input to decrease the curse of dimensionality. It is usually done in filter or wrapper feature selection. The relevancy of the features is chosen in filter feature selection based on their statistical dependence on the target feature. The wrapper feature selection evaluates every possible combination of features from the original list and chooses the best subset depending on the learning model. Hence it is called as model based feature selection. It guarantees an accurate result than the filter feature selection.

The deep learning is the complex artificial neural network that has an ability to find the sequence dependency from the time series data. The uncertainty and the non-linearity issues are effectively handled by the deep neural network. The real time load is the complex and time series in nature. So, the deep learning based gated recurrent unit is utilized in this paper for short term load forecasting. The proposed EBGRU model supports well the current working environment by enhancing the accuracy of the short term load forecasting at a great extent. It is useful for the reliable and an economic functioning of the power sectors, power utilities companies, power manufacturers and society. EBGRU helps power sector in making decisions on operational planning and regional energy exchange. It helps electricity power utility companies in making decisions related to schedule, generate, integrate, control and dispatch of an electricity. It also helps the electricity power manufacturers in making decisions related to load increment/decrement, preparing maintenance schedule and planning energy storage optimizations. Most

importantly, it protects the society from the shortage of electricity and helps to continue the regular schedules smoothly.

Yu et al.[10] designed a hybrid model for improving the performance of STLF using ensemble empirical mode decomposition. The random volatility behavior of the load series was effectively handled by employing denoising technique. The author utilized the decomposition for denoising purpose. The decomposed subseries are modeled using back propagation neural network (BPNN). Finally, the forecasting outcome from each subseries was combined and formed the final forecast result. The BPNN with EEMD model improved the accuracy of load forecasting. Zheng et al.[12] developed a STLF model in which the selection of similar day (SD) was done using XGBoost and k-means clustering. Then, the time series load is decomposed into sub time series for reducing the unnecessary interactions in the load series. The long term dependencies are effectively dealt with LSTM and the load was forecasted. The result showed that, the LSTM model with similar day selection and EMD combination achieved better forecasting performance than LSTM with SD, Auto Regressive Integrated Moving Average (ARIMA) with SD, LSTM with EMD and Support Vector Regression (SVR) with SD.

Gao et al.[1] developed a hybrid load forecasting model. The model forecasted the short term electricity load using GRU along with empirical mode decomposition and the Pearson correlation coefficient feature selection. The decomposition and feature selection concepts were utilized for improving the performance of GRU. The performance of this model is experimented using three different datasets and compared against random forest, random forest with EMD, SVR and SVR with EMD. The GRU with EMD proved an improved performance than others. Qiu et al.[4] introduced an efficient load demand forecasting model. The author utilized the Deep Belief Network (DBN) for forecasting load demand. The extracted Intrinsic Mode Functions (IMF) from decomposition process was modeled by using EMD and DBN with two Restricted Boltzmann Machine (RBM). As a result, DBN with EMD achieved better results.

In load demand forecasting, the EMD technique has been widely utilized for denoising purpose. Each sub-series comprise a portion of the actual load demand series. It makes the sub series to become considerably simpler than the original series and allows it for accurate forecasting. Fan et al.[13] developed the hybridized

model for electricity load forecasting using SVR with differential EMD. The model guaranteed the better interpretability and good accuracy of forecasting. In order to estimate the electricity load demand for a certain date and season, Bedi and Toshniwal[14] merged the EMD approach with the LSTM model. It results better performance than the single model. As stated above, the decomposition algorithm and prediction model of the hybridized are primarily different. However the establishment procedure is nearly identical.

For short term load forecasting, many studies used hybrid models. The methods discussed in literature performed the decomposition on the univariate load data and also performed the forecasting directly from the decomposed components. It may add complexity, introduce random errors and also increases the workload.[15] Using ensemble empirical mode decomposition, the EBGRU model efficiently addresses the noise, non-linearity, and non-stationary aspects of the load data in current study. The temperature has high correlation with the electricity load. So, in addition to load, the temperature data also considered for improving the performance of STLF.

As the size of the input grows larger, so does the complexity of the forecasting model. The dimension of the input should be reduced for reducing the forecasting model complexity. The feature selection is one of the solutions for reducing the dimension of input data. The learning algorithm for obtaining the optimal subset of features is taken into account by the wrapper feature selection. It aids in projecting performance improvement. In this paper, the curse of dimensionality is reduced by reducing the dimensionality of the decomposed component using the Boruta wrapper feature selection. Subsequently, the temporal dependency and uncertainty of the time series load is efficiently analyzed and the over fitting is reduced by using the deep learning based gated recurrent unit. In earlier research, the forecasting models were constructed for each sub series directly. But, in this paper the feature selection is applied on each decomposed sub series to find the significantly correlated features. Consequently, the GRU forecasting model was constructed using that selected features to enhance the forecasting performance. The main contribution of the paper is as follows,

- Multivariate Input: The load data is more sensitive to the temperature. So, the history of load and temperature data is considered as the input features for improving STLF. The previous

24 hours load and temperature features are utilized as input for forecasting the next hour load.
- Decomposition: The multivariate input features are decomposed into different load and temperature components (IMFs and Residue) using EEMD.
- Feature Selection: The dimension of each IMFs and residue is reduced by using the Boruta wrapper feature selection.
- Deep Learning: The selected features from each IMFs and residue are given as input to deep recurrent neural network based GRU network for forecasting. The final load forecast result is constructed by combining the forecast outcome from each IMFs and residue.
- Performance Evaluation: The performance of the proposed EBGRU model is compared against RNN, LSTM, GRU, EBRNN and EBLSTM in terms of RMSE, MAE and MAPE.

The nomenclature used in this paper is given in Table 1.

The rest of the paper is organized as follows. The significance of EEMD and Boruta feature selection are presented in the following section. In the next section, the deep learning based RNN, LSTM, and GRU models are discussed. Followed by, the architecture and algorithm of the proposed EBGRU model is enlightened. The experimental findings of EBGRU are evaluated in terms

Table 1 — Nomenclature

AEMO	Australian energy market operator
ANN	Artificial Neural Network
AR	Auto regression
ARIMA	Auto regressive integrated moving average
AU	AEMO dataset (Site1)
DBM	Deep belief network
DWT	Discrete wavelet transform
EBGRU	Gated recurrent unit with EEMD and boruta
EBLSTM	Long short term memory with EEMD and boruta
EBRNN	Recurrent neural network with EEMD and boruta
EEMD	Ensemble empirical mode decomposition
EMD	Empirical mode decomposition
FT	Fourier transform
GRU	Gated recurrent unit
IMF	Intrinsic mode decomposition
LSTM	Long short term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
RBM	Restricted Boltzmann machine
RMSE	Root mean square error
RNN	Recurrent neural network
SD	Similar day selection
STLF	Short term load forecasting
SVR	Support vector regression
SW	Switzerland dataset

of error measures in the next section. Finally, the present study ends with conclusions.

Ensemble Empirical Mode Decomposition and Feature Selection

The electricity load demand is non-linear, noisy and time series in nature. The accurate STLF can be achieved by effectively handling these characteristics. Many researchers made a deep analysis and designed the models with the combination of signal decomposition technique as an important pre-processing method to enhance the performance of STLF.[16] The Fourier Transform (FT) was utilized for pre-processing the non-linear data. It concentrates the frequency resolution but fails to consider the time stamp.[14,17] The Discrete Wavelet Transform (DWT) has been utilized as an alternate to FT for STLF. It considers the frequency components along with the time stamp. But, it has the critical decimation issues. Next, the self-adaptive Empirical Mode Decomposition (EMD) was introduced for transforming the high frequency components in the load series into low frequency components based on the local characteristics of the data instead of using the basis function as other decomposition methods. However, there is a presence of different frequencies in a single IMF. As a result, EMD has a mode mixing issue. Hence, the random volatility nature of the load data introduces the noise.[6,18] The traditional EMD is not sufficient for obtaining the accurate STLF. The denoising process should be incorporated with EMD to solve the mode mixing problem.[16,19] The method called EEMD has the noise assisted data analysis capacity that can be utilized for improving the short term load forecasting.

Ensemble Empirical Mode Decomposition (EEMD)

Wu and Hung *et al.*[20] presented the EEMD as a self-adaptive signal decomposition approach. It decomposes any series into a number of IMFs and residue. The conditions to be satisfied by each IMF are as follows. First, the difference between the number of zero crossings and the number of extrema must be zero or one. Second, at every moment in time, the mean value of the envelope defined by local minima and the envelope formed by local maxima is zero.[21,22] Let us consider the original signal x(t). To tackle the mode mixing problem in EEMD, white noise is mixed with original signal across the entire time-frequency space. As shown below, the original time series signal is decomposed into a number of IMFs and residue.

$$x(t) = \sum_{j=1}^{n} c_j(t) + r(t) \qquad \dots (1)$$

where, r(t) denotes the residue and c_j(t) denotes IMF. Due to signal intermittence, the single IMF may have disparate frequencies. It leads to the mode mixing problem. This problem can be handled well by introducing the EEMD. To tackle the mode mixing problem in EEMD, white noise is mixed with the original signal across the entire time-frequency space.[23] The following is the EEMD's working procedure.

First, the random noise signal is added to the original signal. After that, the decomposition is performed to form the IMFs and residue. Repeat the steps by adding the random white noises at each iteration. Then, find the iteration that generates the minimum number of IMFs. Finally, compute the final decomposition result by taking the ensemble mean of IMFs.[18] The original signal is decomposed by using EEMD as follows,

$$x(t) = x(t) + n_p(t) \qquad \dots (2)$$

$$x_p(t) = \sum_{p=1}^{N_p} c_{j,p}(t) + r_{n_p}(t) \qquad \dots (3)$$

where, n_p(t) denotes the white noise added at the pth trail, x(t) denotes input signal, x_p(t) denotes the input signal with added noise at the pth iteration, $c_{i,p}$ denotes the jth IMF of the pth iteration and r_{n_p} denotes the residue at the pth iteration. The ensemble mean of the IMFs is calculated as follows,

$$c_j = \left(\sum_{p=1}^{N} c_{j,p} \right) \Big/ N \qquad \dots (4)$$

where, 'N' denotes the number of iteration.

Boruta Feature Selection

Feature selection is a key strategy for reducing dimensionality. It reduces the dimension of the dataset by identifying the relevant features. There are two types of feature selection namely, filter method and wrapper method. The filter method works faster as compared to wrapper method. It uses the statistical characteristics of the features as feature evaluator. It assigns a numerical value to each feature based on the statistical characteristics like correlation between it and the target feature. So, the feature with least important are identified as irrelevant and are removed from the input features list. On the other hand, the learning algorithm is utilized for identifying the best subset of features in wrapper feature selection.[24] It checks all the possible combination of features to find

the optimal features. It reduces over fitting and curse of dimensionality issues by identifying an optimal subset of features.[25] The process of finding all the strongly relevant and all the weakly relevant features is not an easy task.[26] The wrapper feature selection provides the solution for all relevant problems. The Boruta feature selection is used in this paper to locate all of the important features for increasing STLF performance.

The Boruta feature selection is the model-based (wrapper) feature selection. It is an enhanced version of the random forest regression.[27] It provides the best solution for finding the all relevant problems. The random correlations are formed by the random fluctuation in time series data and the fixed size of samples while using random forest regression. Hence, this issue increases with the decreasing size of data. So, the Boruta feature selection creates the random features by performing the permutations of the samples for each feature and adds it with original dataset. Then, it finds the importance of both original and random features. After that, it selects the features that have the relevance (z-score) more than the maximal relevance of its random features as important features. It repeats these processes until all features are selected or rejected.

Deep Learning Models

The deep neural network is the complex ANN that has an ability to overcome the limitations of artificial neural network by handling the data of any size and any number of layers easily. It analyses the input by stacking numerous hidden layers between the input and output layers. The DNN has the capability to analyze the temporal dependency in the input complex time series data and keep those information for future process. It works well with time series data and successfully uncovers hidden temporal patterns. As the load demand has time series nature, the deep neural network is suitable for processing and forecasting the load. The deep learning methods have been utilized in variety of applications such as load forecasting, heart disease prediction, wind speed prediction, toxicity prediction and crop yield prediction. It is also suitable for processing big datasets with high dimensions.[11] The deep learning methods like GRU, LSTM and RNN have the capability of handling the temporal dependency. So, they are suitable for analyzing the time series load data. The discussion about these methods is given in the following section.

Recurrent Neural Network

A recurrent neural network is a type of ANN with a recurrence connection in its network topology. It captures the sequential information from the input and shares it with different time steps by looping back to itself. So, it is well suited to the time series forecasting. The parameter sharing capability leads to the reduction of computational time. Despite the fact that it enhances forecasting accuracy, it has two significant flaws: vanishing and exploding gradient problems. Hence, it can remember the past information for a shorter period only.[28]

Long Short Term Memory

A more advanced version of the RNN is the LSTM. It overcomes the gradient descent and short term memory issues of RNN by introducing memory in each cell. Each memory consists of three different gates namely, input, forgot and output gates.[29] It tunes the information flow through the network by employing the gates. The input gate finds the new information from the current input and makes use of it for updating the current cell state. The required information is taken from the cell state by the output gate and adds it to the output. The forget gate discards the unwanted information that cannot be reused. In each cell, it retains the required information and discards the unwanted information from the past computation. It remembers the long sequences of temporal dependency information in the cell itself and utilizes it for further computation.[11] In recent years, many researchers utilized LSTM for time series forecasting and attained better results.

Gated Recurrent Unit

The GRU is an extended version of the LSTM that addresses the vanishing and exploding gradient problems of RNN. It can also successfully learn the long-term dependencies that occur in time series data.[30] It is similar to LSTM but, it has fewer parameters than the LSTM. Hence, it produces better results than the LSTM for the same problem.[31] The GRU utilizes only the reset and update gates for regulating the information flow through the neural network.[28] The architecture of gated recurrent unit is depicted in Fig.1.

The update gate decides what information from the past needs to keep in the network and what information the network should forget. The update gate feeds the sigmoid function, the previous concealed state information as well as the current input. It finds update state at time 't' as follows,

$$u_t = \sigma(W_u x_t + W_u h_{t-1} + b_u) \qquad \dots (5)$$

where, 'u_t' represents update state at time 't' and 'x_t' represents the current input. The 'h_{t-1}' denotes the
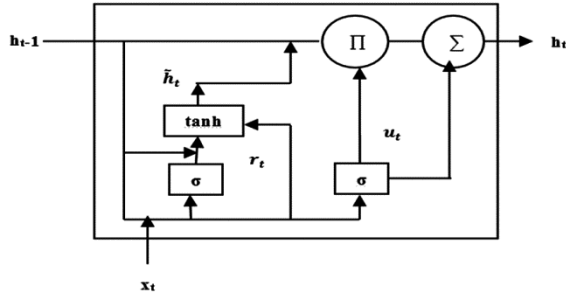
Fig. 1 — Architecture of Gated Recurrent Unit

hidden state information at the time period 't−1'. The 'W$_u$' represents the weight matrix of the update gate and 'b$_u$' denotes the bias weight matrix of the update gate during sigmoid function.

The update gate forgets the previous state information when the outcome of the sigmoid function is closer to 1. It remembers the previous hidden state information when the outcome is 0. The reset gate also utilizes the sigmoid function for resetting the hidden state. It feeds the current state and previous hidden state information as input to the sigmoid function and finds whether the current state have any new information for resetting or it has only the previous information for resetting.[32,33] When the reset gate produces the value closer to 0, it throws away the previous state information and when the value is closer to 1, it resets the state by the new information as follows,

$$r_t = \sigma(W_r x_t + W_r h_{t-1} + b_r) \qquad \text{... (6)}$$

The hidden state, h$_t$, is updated as follows,

$$\tilde{h}_t = tanh(W.[r_t * h_{t-1}] + W.x_t) \qquad \text{... (7)}$$

$$h_t = (1 - u_t) * h_{t-1} + u_t * \tilde{h}_t \qquad \text{... (8)}$$

where, $\tilde{h}_t$ denotes candidate memory cell.

Proposed Model
EBGRU Architecture

The proposed deep learning based EBGRU model is a hybrid model. The architecture of EBGRU consists of four phases namely decomposition, feature selection, deep learning and performance evaluation. The EBGRU model's architecture is depicted in Fig. 2. In first phase, decomposition phase, the load (L) and temperature (T) series are decomposed sequentially into a number of subseries by using EEMD. However, the non-stationary and non-linearity nature of load and temperature adds complexity to the forecasting process, it cannot be ignored while performing the load forecasting. The divide and conquer technique of EEMD guarantees the reduction of model complexity by dividing the input time series into a number of sub time-series. It also removes high-frequency components from other frequency modules and makes it possible to minimize the impact of measurement noise. In the second phase, feature
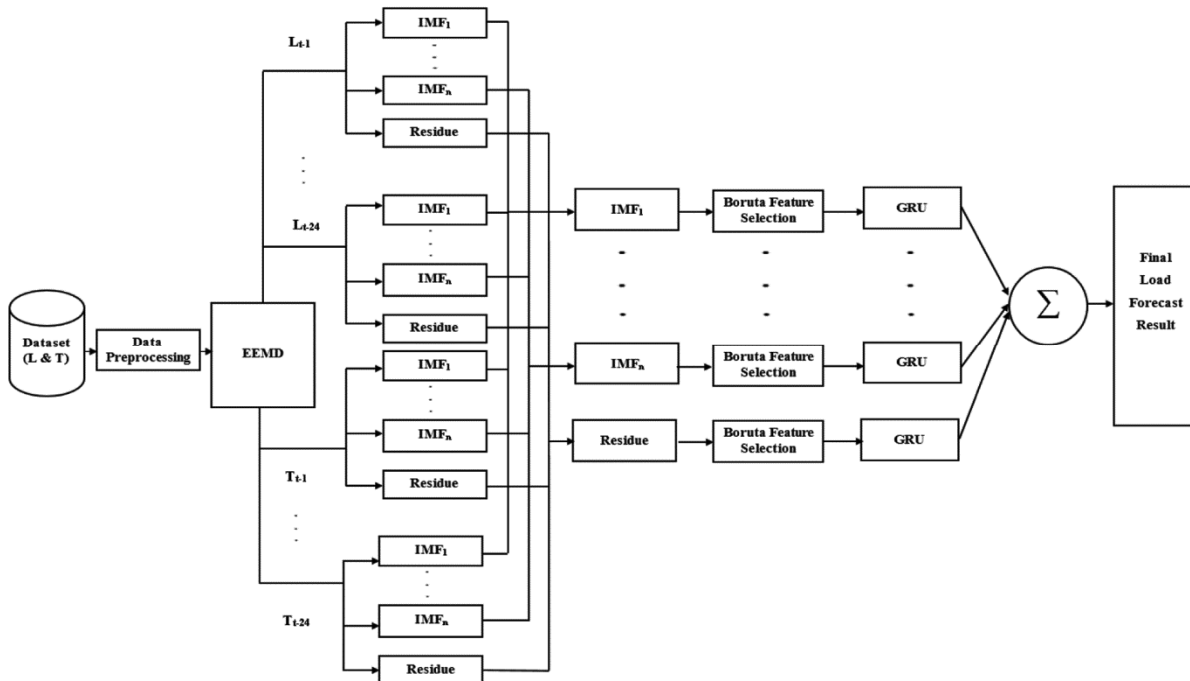

Fig. 2 — Architecture of EBGRU model

selection phase, the dimension of the input for the forecasting process is reduced by finding relevant features using accurate wrapper. Using Boruta the ideal subset of features is selected from load and temperature data of each IMF and residue.

In third phase, deep learning phase, the selected features are given as input to GRU for forecasting. Gated recurrent unit is slightly less complex, faster and consumes less memory than LSTM. It controls the flow of information without requiring the use of a memory unit. In fourth phase, performance evaluation phase, the dataset with previous day 24 hours load and temperature is considered as the input and the next hour load is predicted as an output. In terms of root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE), the proposed EBGRU model is compared against RNN, LSTM, GRU, EBRNN, and EBLSTM.

EBGRU Algorithm

The proposed EBGRU model's functioning procedure is as follows. The EBGRU model cleans the input load (L) and temperature (T) time series by removing the incompleteness. It replaces the missing values using the mean value of the features. Then, it transforms the load and temperature to the new range (0 to 1) by normalizing it by using the min-max normalization. Consequently, it performs the decomposition, feature selection, forecasting and performance evaluation. First, it decomposes the input series using ensemble empirical mode decomposition. Prior to decomposition, it adds the white noise to the original signal and generates the noise added original time series for the input time series as given in Eq. (2). After that it decomposes the noise added original signal into the series of IMFs and residue as given in Eq. (3). It repeats the decomposition process for 'K' number of trials and 'A' number of features where A = 1, 2, 3, …, 49. After that it combines the corresponding IMFs and residue of the related features and forms the final IMFs and residue. Subsequently, it identifies the optimal subset of features by using Boruta feature selection for improving the forecasting performance. It creates the random features by permuting the original features. Then, it adds the random features to the original features and builds the random forest regression using the extended dataset. Next, it selects the original features as the relevant features if its z-score is higher than the maximum z-score of its random features. EBGRU repeats the process of finding optimal subset of features for each IMFs and residue. Consequently, it

forecasts the load for each IMFs and residue using gated recurrent unit (GRU). Then, it combines the forecast output of each IMFs and residue to construct the final load forecast output. Finally, it tests the performance of EBGRU against RNN, LSTM, GRU, EBRNN and EBLSTM in terms of error measures. It tests the generality of the proposed EBGRU forecasting model by using two different datasets of hourly load and temperature data collected from AEMO and European country Switzerland.

Results and Discussion

Dataset

The experiment is conducted using the load and temperature data recorded at an hourly basis from the European and Australian countries. The Switzerland (SW) load and temperature from January 2008 to February 2012 and AEMO (AU) load and temperature from January 2004 to February 2005 are used for forecasting. The data from 1st January 2008 to 31st December 2010 is utilized as the training dataset, the data from 1st January 2011 to 31st December 2011 is used as the validation dataset, and the data from 1st January 2012 to 31st February 2012 is used as the testing dataset for the SW dataset. The data from 1st January 2004 to 31st December 2004 is used as the training dataset, 1st January 2005 to 31st January 2005 is utilized as validation dataset, and 1st February 2005 to 28th February 2005 is used as the testing dataset for the AU dataset.

Performance Evaluation

To show the superiority of the proposed EBGRU model, the performance is evaluated by using RMSE, MAE and MAPE. Let 'Actual$_t$' and 'Forecast$_t$' be the actual and forecast load for the period 't'. Let 'N' be the number of samples.[34,35] The RMSE, MAPE and MAE of the forecasted load are determined as follows,

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(Actual_t - \text{Forecast}_t)^2} \qquad \dots (9)$$

$$\text{MAPE} = \frac{1}{N}\sum_{t=1}^{N}\left|\left(\frac{Actual_t - Forecast_t}{Actual_t}\right)*100\right| \qquad \dots (10)$$

$$\text{MAE} = \frac{1}{N}\sum_{t=1}^{N}|Actual_t - Forecast_t| \qquad \dots (11)$$

Case Study – I: Results obtained from SW dataset

The results of the experiment conducted using SW dataset are as follows. The previous day load data has

strong positive correlation and previous day meteorological temperature data has strong negative correlation with the current load. So, the previous day load and temperature series for the 24 hours load and temperature are taken as input and it is decomposed into 6 IMFs and residue. Then, the IMFs of the previous day load and temperature for 24 hours is formed by combining the corresponding hour IMFs and residue. After that, the Boruta feature selection is applied at each IMFs and residue to select the relevant features from the previous day 24 hours load and temperature features related to the next hour load. The features extracted are L_{t-24}, L_{t-23}, L_{t-20}, L_{t-22}, L_{t-1}, L_{t-2}, L_{t-19}, T_{t-13}, L_{t-21}, L_{t-4}, T_{t-5}. Then, these features are given as input to the GRU for forecasting.

GRU is designed as network with three hidden layers of 20 units in each. The number of epochs tested varies between 10 and 200. Finally, the number of epochs is set to 140. The training, validation and

testing are performed by using the training, validation and testing datasets respectively. The validation is performed by setting MAE as the loss function and Adam as the optimizer. The load requirement is high during the high winter season (January and February) in Switzerland. So, the winter season load demand is forecasted. The comparison of the training and testing losses for SW dataset is shown in Fig. 3. The comparison of forecast load generated by using RNN, LSTM, GRU, EBRNN, EBLSTM and EBGRU against actual load for SW dataset is shown in Fig. 4 and Fig. 5.

The training and testing loss shown in Fig. 3 depicts the improvement of the EBRNN, EBLSTM and EBGRU performance in each iteration. The testing loss reduces gradually as the number of iteration increases. The loss started to flat from 20 iterations. Hence, the overfitting of the EBGRU is reduced by reducing the testing loss. So, the EBGRU is designed by setting the number of epochs as 140. The forecasted load of
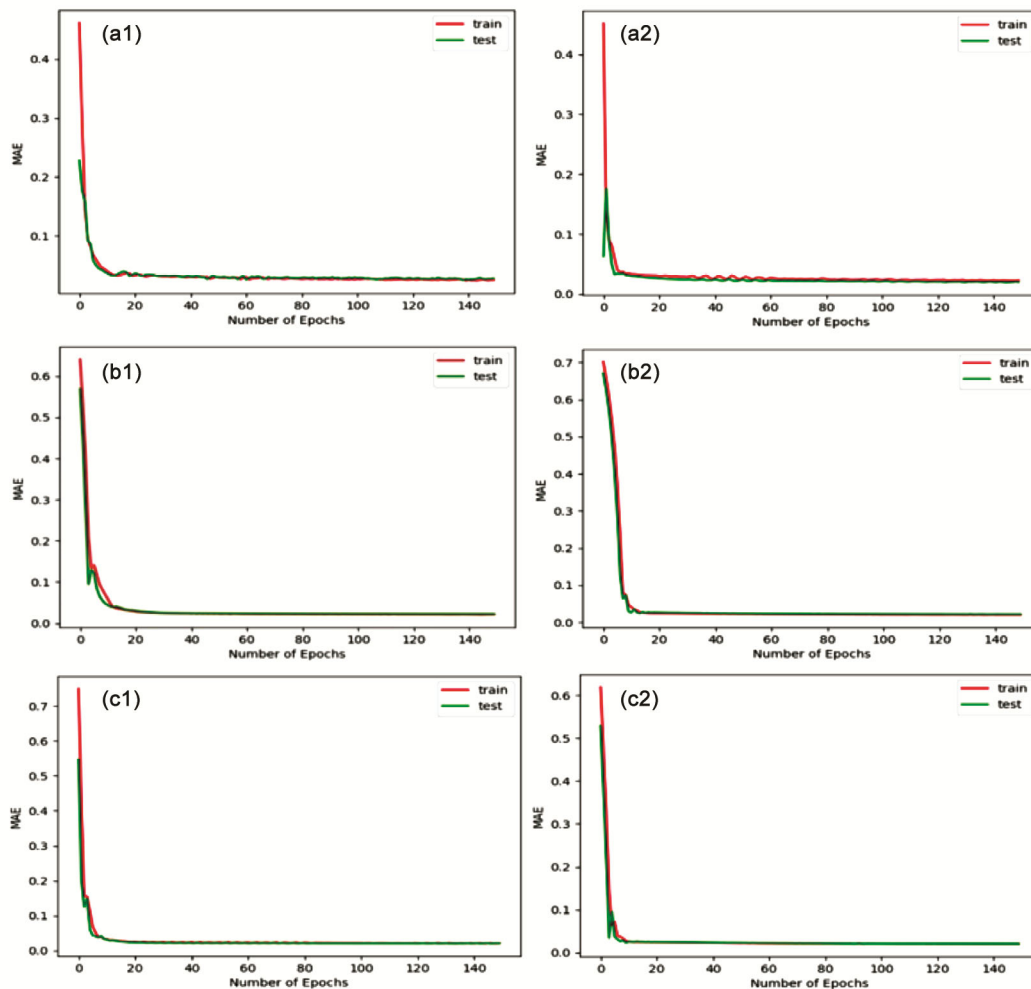

Fig. 3 — Comparison of training and testing losses for SW dataset: (a1) RNN (a2) EBRNN (b1) LSTM (b2) EBLSTM (c1) GRU (c2) EBGRU
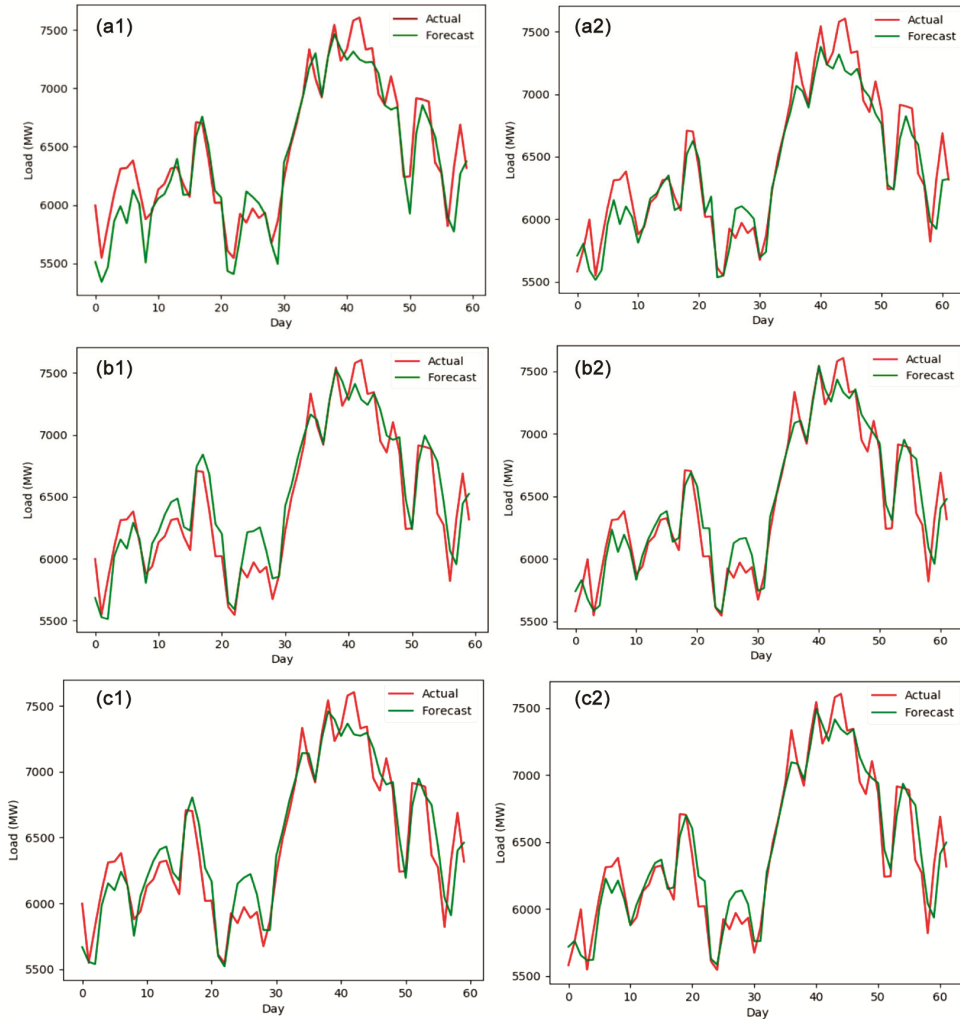
Fig. 4 — Comparison of actual and forecast load for SW dataset: (a1) RNN (a2) EBRNN (b1) LSTM (b2) EBLSTM (c1) GRU (c2) EBGRU
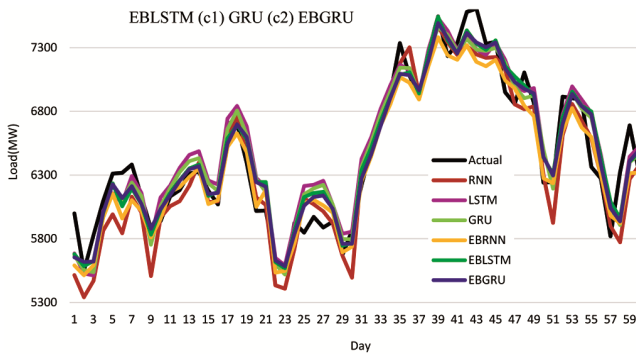

Fig. 5 — Sample of actual and forecast load: SW dataset

Table 2 — Comparison of load forecasting results of SW

Deep learning methodologies	RMSE	MAE	MAPE
RNN	209.684	160.068	2.512
LSTM	184.141	149.595	2.356
GRU	173.694	141.753	2.220
EBRNN	170.504	130.182	2.002
EBLSTM	160.815	125.417	1.977
EBGRU	154.901	120.784	1.895

EBRNN, EBLSTM and EBGRU is almost as similar as the actual load. It is shown in Fig. 4. However, the EBGRU forecast load is almost closer to the actual load compared to EBRNN and EBLSTM. The sample of the EBGRU forecast load with actual and other models is shown in Fig. 5. The forecasting with EEMD decomposition and Boruta feature selection produces less error compared to forecasting without EEMD and Boruta feature selection as shown in Fig. 4 and Fig. 5. In addition to that, the proposed GRU with EEMD and Boruta feature selection shows the superiority by producing more accurate forecasting results compared to RNN, EBRNN, LSTM, EBLSTM and GRU.

The forecasting outcomes for the SW dataset are compared using the performance indicators such as RMSE, MAPE, and MAE in Table 2. The EBRNN,

EBLSTM and EBGRU produces less RMSE of 39.18, 23.326 and 18.793, MAE of 29.886, 24.178 and 20.969, MAPE of 0.51, 0.379 and 0.325 compared to RNN, LSTM and GRU respectively. Hence, among all the EBGRU achieves better forecasting results by producing less RMSE of 15.603, MAE of 9.398 and MAPE of 0.107 with EBRNN. It is also producing less RMSE of 5.914, MAE of 4.633 and MAPE of 0.082 with EBLSTM.

Case Study – II : Results Obtained from AU Dataset

AEMO load and temperature data are also used to test the proposed model. It consists of daily load and temperature data recorded at every hour at AEMO. The previous day 24 hours load and temperature are more congruous with the next hour load. The input load and temperature series is decomposed into IMFs and residue using EEMD. The decomposition on AU dataset creates 6 IMFs and one residue for each feature. Then, all the corresponding IMFs and the residue of the related

features are combined together to form the final 6 IMFs and residue. The Boruta wrapper is then used to extract all relevant features from the IMFs and residue. It selects the L_{t-24}, L_{t-2}, L_{t-5}, L_{t-1}, T_{t-1}, L_{t-14}, L_{t-23}, L_{t-3}, L_{t-4}, L_{t-15}, L_{t-7}, L_{t-16}, L_{t-12} features as relevant subset of features to the load at time 't'. Then, the selected features are given as input to the GRU. The gated recurrent unit is constructed as follows.

The number of hidden layers in the GRU network is set to three with each layer has a total of 20 units. The number of epochs is set as 1200 epochs, the loss function as MAE and the optimizer as Adam. The GRU network is trained and validated by using the training and validation datasets. The load demand is high in the summer at the AEMO. So, the load for the summer (February) is forecasted by using the GRU. The comparison of the training loss and the testing loss during validation of AU dataset is shown in Fig. 6. The
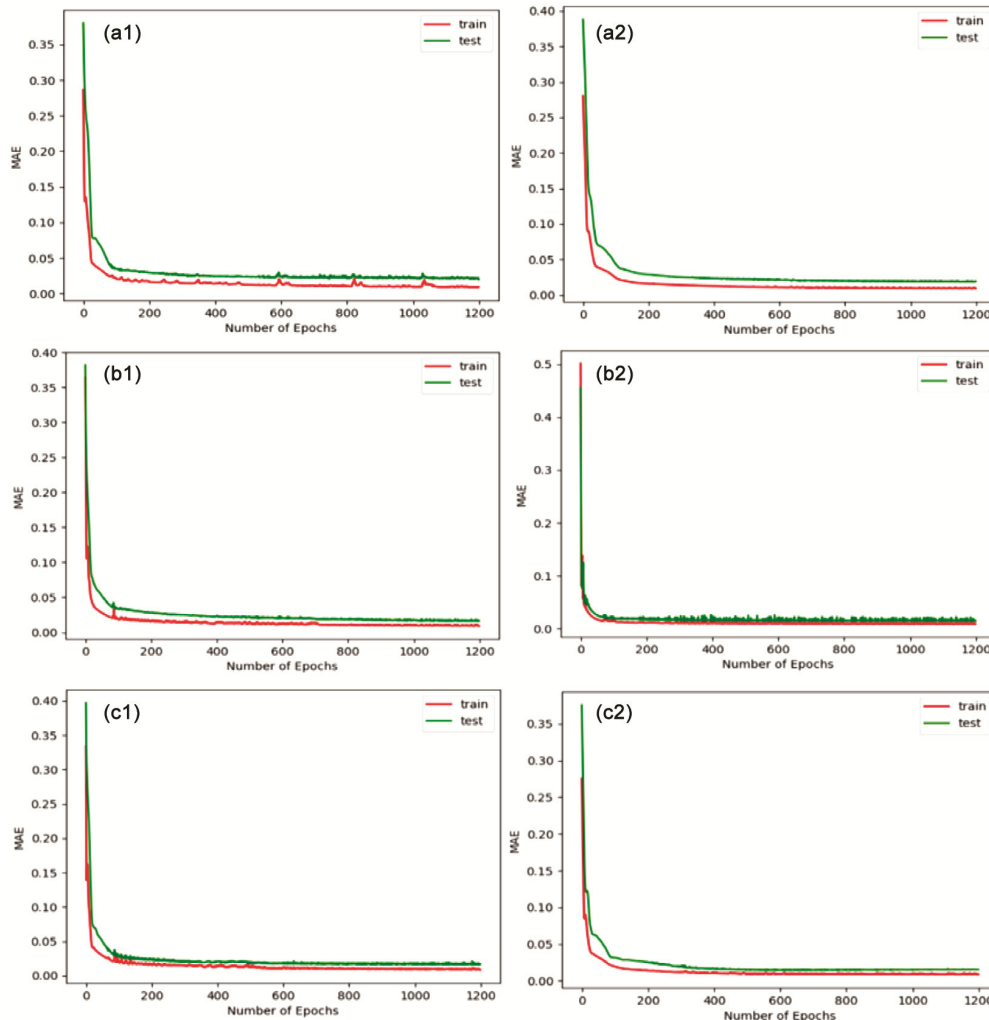

Fig. 6 — Comparison of training and testing losses for AU dataset: (a1) RNN (a2) EBRNN (b1) LSTM (b2) EBLSTM (c1) GRU (c2) EBGRU
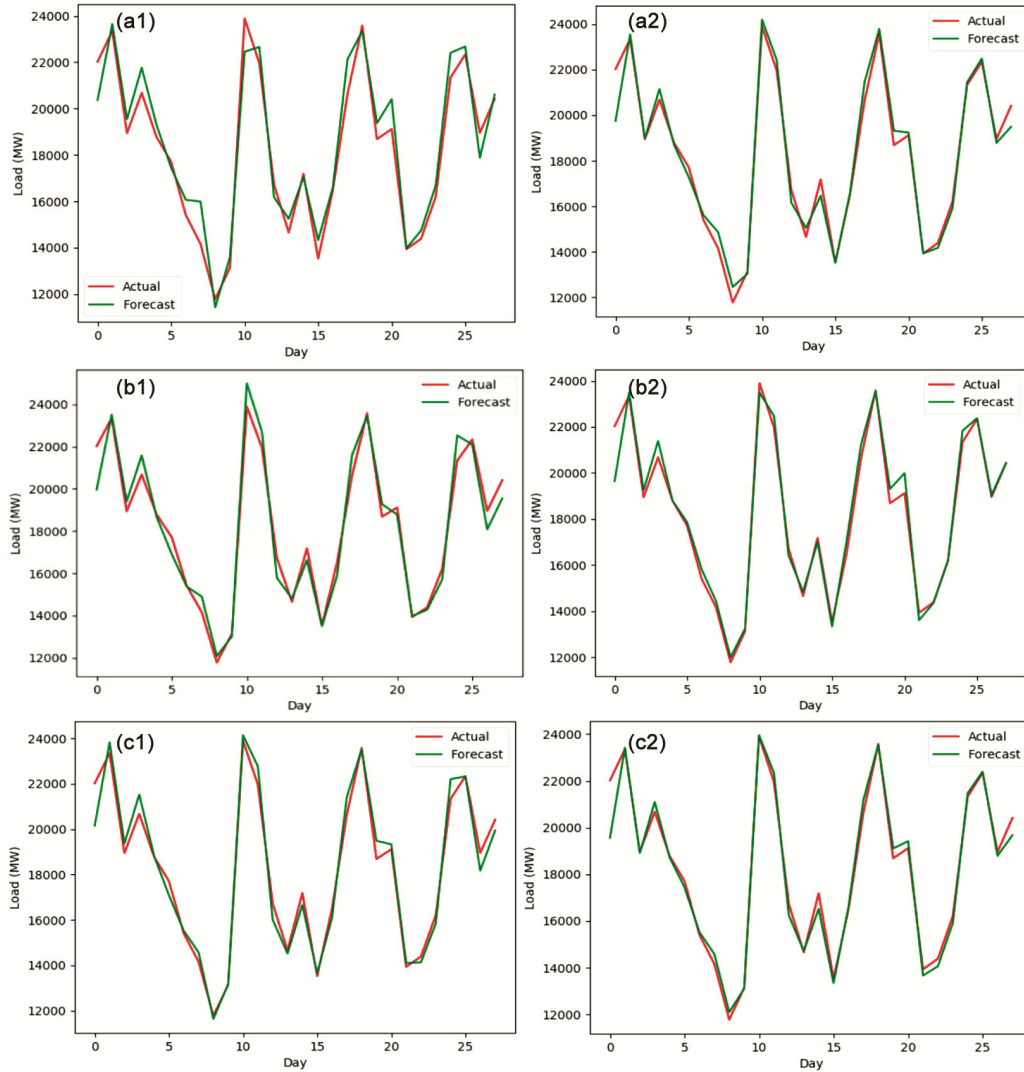
Fig. 7 — Comparison of actual and forecast load for AU dataset: (a1) RNN (a2) EBRNN (b1) LSTM (b2) EBLSTM (c1) GRU (c2) EBGRU

comparison of actual and forecast load predicted by RNN, LSTM, GRU, EBRNN, EBLSTM, and EBGRU are shown in Figs 7 & 8.

The significant reduction of loss in each iteration is shown in Fig. 6. The loss reduces with increasing number of iterations. The loss reduces gradually from 200th iteration and maintains the lowest value at the iteration 1200. So, the EBGRU is designed by setting the number of iterations as 1200. The forecast load of EBRNN, EBLSTM and EBGRU is almost closer to the actual load. The graph in Fig. 7 demonstrates this. Hence, compared to EBRNN and EBLSTM, the forecast load of EBGRU is more similar to the actual load. The sample of the comparison of forecast load generated by EBRNN, EBLSTM and EBGRU against the actual load is shown in Fig. 8. The load forecasted
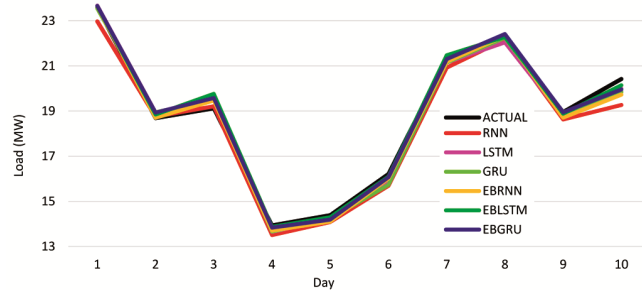

Fig. 8 — Sample of actual and forecast load of AU dataset

by the RNN, LSTM, GRU with EEMD-Boruta feature selection produces better results than the RNN, LSTM, GRU without EEMD-Boruta feature selection. It is shown in Fig. 7 and Fig. 8. Hence, the proposed EBGRU outperforms other models. The performance

Table 3 — Comparison of load forecasting results of AU

Deep learning methodologies	RMSE	MAE	MAPE
RNN	0.689	0.419	2.255
LSTM	0.590	0.347	1.866
GRU	0.602	0.335	1.791
EBRNN	0.634	0.325	1.699
EBLSTM	0.565	0.271	1.410
EBGRU	0.592	0.268	1.377

comparison of load forecast results obtained by using RNN, LSTM, GRU, EBRNN, EBLSTM and EBGRU is given in Table 3.

EBRNN, EBLSTM and EBGRU produces less RMSE of 0.055, 0.025 and 0.01, MAE of 0.093, 0,075 and 0.067, MAPE of 0.556, 0.456 and 0.414 compared to RNN, LSTM and GRU respectively. Hence, among all the EBGRU achieves better forecasting results by producing less MAPE of 0.878, 0.489, 0.414, 0.32 and 0.033 compared to RNN, LSTM, GRU, EBRNN and EBLSTM.

Conclusions

The accurate short term load forecasting is a critical task for making energy policy in the electricity system. The proper planning, scheduling and dispatching of load demand is also important for maintaining the stability and reliability of power system economically. In this paper, the EBGRU model is proposed as a hybrid model for improving the accuracy of the STLF. The highly correlated temperature data is also considered along with the history of load data for STLF. The time series load and temperature data is decomposed into IMFs and residue for denoising purpose. The relevant features are extracted effectively by using the Boruta feature selection for reducing the dimensionality. For tackling the uncertainty and temporal dependency of load, the forecasting is done using a GRU deep neural network. In terms of RMSE, MAE, and MAPE, the proposed EBGRU model is compared to the RNN, LSTM, GRU, EBRNN, and EBLSTM models. The EBRNN, EBLSTM and EBGRU produces less RMSE of 0.055, 0.025 and 0.01, MAE of 0.093, 0,075 and 0.067, MAPE of 0.556, 0.456 and 0.414 compared to RNN, LSTM and GRU respectively. The result shows that the EBGRU model shows its generality and proves the superiority by outperforming other models. The proposed EBGRU model can be used for various time series applications like weather forecasting, financial forecasting and wind speed forecasting in the future. In particular, the proposed approach can be used to

improve forecasting performance by combining time series data with large data of high dimension.

References

1   Gao X, Li X, Zhao B, Ji W, Jing X & He Y, Short-term electricity load forecasting model based on EMD-GRU with feature selection, *Energies*, **12(6)** (2016) 1140.

2   Senthil K P, A Review of Soft Computing Techniques in Short-Term Load Forecasting, *Int J Appl Eng Res*, **12(18)** (2017), 7202–7206.

3   Zhang Z & Hong W C, Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm, *Nonlinear Dyn*, **98(2)** (2019) 1107–1136.

4   Qiu X, Ren Y, Suganthan P N & Amaratunga G A, Empirical mode decomposition based ensemble deep learning for load demand time series forecasting, *Appl Soft Comput*, **54** (2017) 246–255.

5   Barak S & Sadegh S S, Forecasting energy consumption using ensemble ARIMA–ANFIS hybrid algorithm, *Int J Electr Energy syst*, **82** (2016) 92–104.

6   Takeda H, Tamura Y & Sato S, Using the ensemble Kalman filter for electricity load forecasting and analysis, *Energy*, **104** (2016) 184–198.

7   Vu D H, Muttaqi K M & Agalgaonkar A P, A variance inflation factor and backward elimination based robust regression model for forecasting monthly electricity demand using climatic variables, *Appl Energy*, **140** (2015) 385–94.

8   Dong Z, Yang D, Reindl T & Walsh W M, Short-term solar irradiance forecasting using exponential smoothing state space model, *Energy*, **55** (2013) 1104–1113.

9   Matsuo Y & Oyama T, Forecasting daily electric load by applying artificial neural network with fourier transformation and principal component analysis technique, *J Oper Res Soc China*, **8(4)** (2020) 655–667.

10 Yu Y L, Li W, Sheng D R & Chen J H, A hybrid short-term load forecasting method based on improved ensemble empirical mode decomposition and back propagation neural network, *J Zhejiang Univ-Sc A*, **17(2)** (2016) 101–114.

11 Subbiah S S & Chinnappan J, An improved short term load forecasting with ranker based feature selection technique, *J Intell Fuzzy Syst*, **39(5)** (2020) 6783–6800.

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

13 Fan G F, Peng L L, Hong W C & Sun F, Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression, *Neurocomputing*, **173** (2016) 958–970.

14 Bedi J & Toshniwal D, Empirical mode decomposition based deep learning for electricity demand forecasting, *IEEE Access*, **6** (2018) 49144–49156.

15 Zhang Z, Hong W C & Li J, Electric load forecasting by hybrid self-recurrent support vector regression model with variational mode decomposition and improved cuckoo search algorithm, *IEEE Access*, **8** (2020) 14642–14658.

16 Gaci S, A new ensemble empirical mode decomposition (EEMD) denoising method for seismic signals, *Energy Procedia*, **97** (2016) 84–91.

17 Prasad R, Deo R C, Li Y & Maraseni T, Weekly soil moisture forecasting with multivariate sequential, ensemble

empirical mode decomposition and Boruta-random forest hybridizer algorithm approach, *Catena*, **177** (2019) 149–166.

18 Liu T, Luo Z, Huang J & Yan S, A comparative study of four kinds of adaptive decomposition algorithms and their applications, *Sensors*, **18(7)** (2018) 2120.

19 Zhang J, Wei Y M, Li D, Tan Z & Zhou J, Short term electricity load forecasting using a hybrid model, *Energy*, **158** (2018) 774–781.

20 Wu Z & Huang N E, Ensemble empirical mode decomposition: a noise-assisted data analysis method, *Adv Adapt Data Anal*, **1(01)** (2009) 1–41.

21 Raj N & Brown J, An EEMD-BiLSTM algorithm integrated with boruta random forest optimiser for significant wave height forecasting along coastal areas of Queensland, Australia, *Remote Sens*, **13(8)** (2021) 1456.

22 Zhang C, Wei H, Zhao J, Liu T, Zhu T & Zhang K, Short-term wind speed forecasting using empirical mode decomposition and feature selection, *Renew Energy*, **96** (2016) 727–37.

23 Wu T Y & Chung Y L, Misalignment diagnosis of rotating machinery through vibration analysis via the hybrid EEMD and EMD approach, *Smart Mater Struct*, **18(9)** (2009) 095004.

24 Ahmed A A M, Deo R C, Ghahramani A, Raj N, Feng Q, Yin Z & Yang L, LSTM integrated with Boruta-random forest optimiser for soil moisture estimation under RCP4.5 and RCP8.5 global warming scenarios, *Stoch Environ Res Risk Assess*, **35** (2021) 1–31.

25 Subbiah S S & Chinnappan J, Opportunities and challenges of feature selection methods for high dimensional data: A review, *Ingénierie des Systèmesd'Information*, **26(1)** (2021) 67–77.

26 Senthil K P & Lopez D, A review on feature selection methods for high dimensional data, *Int J Eng Technol*, **8(2)** (2016) 669–672.

27 Szul T, Tabor S & Pancerz K, Application of the BORUTA algorithm to input data selection for a model based on rough set theory (RST) to prediction energy consumption for building heating, *Energies*, **14(10)** (2021) 2779.

28 Din G M, MautheA U & Marnerides A K, Appliance-level short-term load forecasting using deep neural networks in *Proc Int Conf Comput Netw Communi (ICNC)*, 2018 53–57.

29 Paramasivan S K, Deep learning based recurrent neural networks to enhance the performance of wind energy forecasting: A review, *Revue d'Intelligence Artificielle*, **35(1)** (2021) 1–10.

30 Gao B, Huang X, Shi J, Tai Y & Xiao R, Predicting day-ahead solar irradiance through gated recurrent unit using weather forecasting data, *J Renew Sustain Energy*, **11(4)** (2019) 043705.

31 Subbiah S S & Chinnappan J, A review of short term load forecasting using deep learning, *Int J Emerg Technol* **11(2)** (2020) 378–84.

32 Zheng J, Chen X, Yu K, Gan L, Wang Y & Wang K, Short-term power load forecasting of residential community based on GRU neural network, *Int Conf Power Syst Technol*, (2018) 4862–4868.

33 Subbiah S S & Chinnappan J, A Review of bio-inspired computational intelligence algorithms in electricity load forecasting, in *Smart Buildings Digitalization*, **1st edn**, vol I, edited by O V G Swathika, K Karthikeyan & S K Padmanaban (CRC Press) 2022, 169–192.

34 Subbiah S S & Chinnappan J, Short-term load forecasting using random forest with entropy-based feature selection, *Artificial Intelligence and Technologies, Lecture Notes in Electrical Engineering* (Springer, Singapore) 2022, 73–80, December.

35 Senthil K P, Improved prediction of wind speed using machine learning, *EAI Endorsed Trans Energy Web*, **6(23)** (2019) 1–7.