Applied Energy 269 (2020) 114915

Contents lists available at ScienceDirect

Applied Energy

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

Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid

Ghulam Hafeez[a,b,*], Khurram Saleem Alimgeer[a], Imran Khan[b]

[a] COMSATS University Islamabad, Islamabad 44000, Pakistan
[b] University of Engineering and Technology, Mardan 23200, Pakistan

HIGHLIGHTS

- A novel hybrid electric load forecasting model based on MMI-FCRBM-GWDO is developed for the decision making of a smart grid.
- To overcome the problem of curse of dimensionality a novel MMI technique is proposed for feature selection.
- The forecast accuracy and convergence rate of FCRBM is enhanced by GWDO.
- The proposed model is tested on hourly load data of USA power grids: FE, Dayton, and EKPC.

ARTICLE INFO

Keywords:
Deep learning
Electric load forecasting
Factored conditional restricted Boltzmann machine
Genetic wind driven optimization algorithm
Modified mutual information technique
PJM electricity market
Rectified linear unit
Smart grid

ABSTRACT

Accurate electric load forecasting is important due to its application in the decision making and operation of the power grid. However, the electric load profile is a complex signal due to the non-linear and stochastic behavior of consumers. Despite much research conducted in this area; still, accurate forecasting models are needed. In this article, a novel hybrid short-term electric load forecasting model is proposed. The proposed model is an integrated framework of data pre-processing and feature selection module, training and forecasting module, and an optimization module. The data pre-processing and feature selection module is based on modified mutual information (MMI) technique, which is an improved version of the mutual information technique, used to select abstractive features from historical data. The training and forecasting module is based on factored conditional restricted Boltzmann machine (FCRBM), which is a deep learning model, empowered via learning to forecast the future electric load. The optimization module is based on our proposed genetic wind-driven (GWDO) optimization algorithm, which is used to fine-tune the adjustable parameters of the model. The accuracy of the proposed framework is evaluated through historical hourly load data of three USA power grids, taken from publicly available PJM electricity market. The proposed model is validated by comparing it with four recent forecasting models like Bi-level, mutual information-based artificial neural network (MI-ANN), ANN-based accurate and fast converging (AFC-ANN), and long short-term memory (LSTM) in terms of accuracy and convergence rate.

1. Introduction

Smart grid (SG) emerged as a smart power system that has recently achieved a lot of popularity due to its importance in electric load forecasting [1]. A variety of novel research work has been conducted in the field of electric load forecasting, however, more accurate and robust electric load forecast models are still required. An accurate estimation of variation in future electric load is of great importance for both electric utility companies and consumers due to its application in the decision making and operation of the power grid [2]. However, the major obstacles in future electric load forecasting are the various influencing factors such as variable climate, temperature, humidity, occupancy patterns, calendar indicators, and social conventions. The valid mapping of these influencing factors and load variations is extremely cumbersome due to the stochastic and non-linear behavior of consumers. The emanation of advanced metering infrastructure (AMI), communication technologies, and sensing methods in the SG enable us to record, monitor and analyze the impact of these influencing factors on electric load forecasting [3]. In literature, both classical (time-series methods) and computational intelligence methods are applied for electrical load forecasting [4]. Both methods have their limitations. The former classical methods are blamed for their limited ability to handle

* Corresponding author at: COMSATS University Islamabad, Islamabad 44000, Pakistan.
E-mail addresses: ghulamhafeez393@gmail.com, ghulamhafeez@uetmardan.edu.pk (G. Hafeez).

https://doi.org/10.1016/j.apenergy.2020.114915
Received 3 June 2019; Received in revised form 27 March 2020; Accepted 31 March 2020

Nomenclature

AMI	Advanced metering infrastructure
ANN	Artificial neural network
AEMO	Australian energy market operator
ARMA	Aut-regressive and moving-average
BP	Back propagation
CS-SVM	Cuckoo search algorithm based SVM
CS	Cuckoo search algorithm
CRNN	Convolutional recurrent neural network
DRNN	Deep recurrent neural network
DR	Demand response
DWT	Discrete wavelet transform
ESSVR	Ensemble subsampled support vector regression
FFI	Fruit-fly immune algorithm
GWDO	Genetic wind driven optimization
x	Fine tuning
F_t	Forecasted load values
g	Gravitational constant
$F_{pr+1}(i)$	Global best solution according to fitness function
E	Historical electric load data
$\circ$	Hadamard product
E_i	Input discrete random discrete variable
$p(E_i, E_j^t)$	Joint probability of two discrete variables
LSTM	Long short-term memory
LM	Levenberg-Marquardt
MILP	Mixed integer linear programming
MAPE	Mean absolute percentage error
$MI(E_i, E_j^t)$	Mutual information between the two variables
NFIS	Neuro-fuzzy inference system
PSO	Particle swarm optimization
RRMSE	Relative relative root mean error
RTP	Real time pricing
RMSE	Root mean square error
RBF-ELM	Radial basis function-ELM
$y^T w^y$	Style factored
SVM	Support vector machine
E_k^n	Second target discrete random variable, which is average value
SRSVR	Seasonal recurrent support vector regression
S_z	Supplementary variable, which is second target variable
E_j^t	Target discrete random variable
RT	Universal gas constant
$x_{new}^{n+1}(i)$	Updated position
v_i	Velocity
AHLM	Adaptive hybrid learning model
AFC-ANN	Accurate fast converging-ANN
ARIMA	Auto-regressive integrated moving average
BPNN	BP neural network
$\bar{b}$	Bias
w^v, w^y, w^h	Corresponding layer weights
CABCA	Chaotic artificial bee colony algorithm
CRBM	Conditional restricted Boltzmann machine
$\hat{a}$	Dynamic bias of visible layer
$\hat{b}$	Dynamic bias of hidden layer
ELM	Extreme learning machine
FCM	Fuzzy C means
FCRBM	Factored conditional restricted Boltzmann machine
$FF(x_{new}(i))$	Fitness function of position
$FF(v(i))$	Fitness function of velocity
α	Frictional coefficient
GA	Genetic algorithm
HEMS	Home energy management system
$h^T w^h$	Hidden factored
$I_t r$	Irrelevancy threshold
$H(E, E^t)$	Joint entropy of two discrete variables
LSSVM	Least squares SVM
$F_{pr}(i)$	Local best solution according to fitness
MEDEA	Modified EDE
MMI	Modified mutual information technique
MI-ANN	Mutual information artificial neural network
NREL	National renewable energy laboratory
τ	Number of hours under consideration
X_{new}	Position
RBM	Restricted Boltzmann machine
R_t	Redundancy threshold
RNN	Reglet neural network
ReLU	Rectified linear unit
SG	Smart grid
SSA	Singular spectrum analysis
SSVR	Subsampled support vector regression
$\mathscr{X}$	Selected candidate inputs
T_t	Target load values
E_l^m	Third target discrete random variable, which is moving average
$v_{n+1}(i)$	Updated velocity
$v^T w^v$	Visible factored
WNN	Wavelet neural network

non-linear data. On the other hand, computational intelligence methods are criticized for problems like handcrafted features, limited learning capacity, impotent learning, inaccurate appraisal, and insufficient guiding significance. Although, there are some existing machine learning models applied for electric load forecasting, which partially resolve the aforementioned problems and have improved performance due to the use of ingenious design [5]. A suitable mechanism is required to solve the aforementioned problems because low forecast accuracy results in huge economic loss. One percent increase in the forecast error will cause a 10 million increase in the overall utility cost. Therefore, electric utility companies are trying to develop a fast, accurate, robust, and simple short-term electric load forecasting model. Moreover, accurate forecasting can also be beneficial for the detection of potential faults and reliable grid operation.

Over the last two decades, numerous load forecasting models have been developed due to its application in the decision making of the power grid. Boroojeni et al. proposed a generalized method to model off-line data that have different seasonal cycles (e.g., daily, weekly, quarterly, and annually). Both seasonal and non-seasonal load cycles are modeled individually with the help of auto-regressive and moving-average (ARMA) components [6]. Li et al. investigated ensemble subsampled support vector regression (ESSVR) for forecasting and estimation of load [7]. A deep belief network restricted Boltzmann machine (RBM) is used for electric load forecasting. The network reduced the forecast error with affordable execution time [8]. Hong et al. forecasts the electric load of Southeast China with the help of the hybrid model based on seasonal recurrent support vector regression (SRSVR) model and chaotic artificial bee colony algorithm (CABCA). The performance of the model is validated by comparing to the auto-regressive integrated moving average (ARIMA) model [9]. Though, these references provide a good study for future electric load forecasting. However, the load of a microgrid is more volatile, containing high frequency, and sharp variation as compared to a load of the power grid. Moreover, the previous literature focuses on feature engineering and traditional methods like decision tree (DT), ARIMA, and ANN. Although, the DT confronts overfitting problems, means DT outperforms in training and worst in forecasting, and ANN face limited generalization capability and it is hard to control its convergence rate. Also, these learning models are not suitable for large data because their performance is compromised as the size of data increases. Moreover, in addition to feature engineering, there is also a need to integrate the optimization module with forecaster for outstanding performance.

In this regard, a novel hybrid model for short-term electric load forecasting is proposed. The main contributions of this paper are demonstrated as follows:

1. A novel hybrid forecast model composed of modified mutual information (MMI), factored conditional RBM (FCRBM), and genetic wind driven optimization (GWDO) techniques is proposed for short-term electric load forecasting. The aforementioned techniques are integrated in a coordinated modular framework to construct the proposed hybrid model.
2. Based on the existing mutual information (MI) technique [10], a new MMI technique for feature selection is proposed. The proposed MMI technique works systematically on both linear and non-linear load time series data and rank the candidate inputs according to their information value to select key features, and discard irrelevant and redundant features to overcome the problem of curse of dimensionality.
3. Auxiliary variables are proposed for our MMI feature selection technique based on four joint discrete variables.
4. A deep learning technique FCRBM is adapted, which is empowered via learning to forecast the day and week-ahead electric load.
5. A GWDO algorithm is proposed, which is a hybrid of genetic algorithm (GA) and wind-driven optimization (WDO) algorithm. The proposed algorithm has a global powerful search capability and fast convergence rate.
6. The adjustable parameters of both data pre-processing and feature selection module, and the training and forecasting module are fine-tuned by our proposed GWDO algorithm. The purpose is to optimize the performance of the proposed model.
7. The proposed model is tested on historical hourly load data of three USA power grids: FE, Dayton, and EKPC. Results utilizing the proposed model have proven more accurate when compared to the benchmark models like Bi-level, MI-artificial neural network (MI-ANN), accurate fast converging-ANN (AFC-ANN), and long short-term memory (LSTM) in terms of accuracy and convergence rate.

The remaining sections of this paper are arranged in the following manner: related work is presented in Section 2. The proposed system model is demonstrated in Section 3. Simulation results are discussed in Section 4. At the end, the paper is concluded in Section 5. The acronyms and symbols used in this work are listed in **NOMENCLATURE**.

2. Related work

Short-term electric load forecasting normally covers the hours to week prediction horizon and is crucial in the decision making of the power grid especially at large scales, where countries and groups of countries have common power systems such as the European Union. In literature, both statical models and machine learning models are commonly used for short-term load forecasting. To well understand the state of the art short-term load forecasting models, these models are classified into two categories: single models that do not use feature engineering and optimization techniques with forecasters; hybrid models that use an integrated framework of feature engineering, forecasting, and optimization techniques. The detail discussion is as follows:

2.1. Single short-term load forecasting models

The main assumption for these single models is that only the forecaster model is used to forecast the future electric load. Authors proposed distributed methods in [11] to forecast the future load using weather information. The power system is divided into two subnetworks according to weather variations. Moreover, separate forecasting models, i.e., ARIMA and grey are established for both subnetworks. The adapted models are evaluated by comparing with the traditional models using two performance metrics, i.e., relative root mean square error (RRMSE) and mean absolute percentage error (MAPE).

A deep recurrent neural network (DRNN) based model is proposed to forecast the household load [12]. This method overcomes the problems of overfitting created by classical deep learning methods. The results show that DRNN outperforms the existing methods like ARIMA, SVR, and convolutional RNN (CRNN) by 19.5%, 13.1%, and 6.5%, respectively, in terms of RMSE. In [13], long short-term memory RNN (LSTM-RNN) based framework is proposed to forecast the future residential load. The accuracy of the proposed framework is enhanced by embedding appliance consumption sequences in the training data. The proposed framework is validated on the real-world data. However, the authors focus only on accuracy while the convergence rate and computational complexity are ignored. A demand response (DR) scheme based on real-time pricing (RTP) is proposed in [14] for industrial facilities. The scheme adopted ANN for forecasting the future prices for global time horizon optimization. The energy cost minimization is facilitated by price forecasting and is formulated by mixed-integer linear programming (MILP). The proposed framework performance analysis is carried out by the practical case study of steel powder manufacturing. Simulation results illustrate that hourly ahead DR is better than a day ahead DR, with an improved ability to satisfy industrial demand with reducing cost while satisfying targets. Authors in [15] presented a probabilistic forecasting model to forecast solar power, electrical energy consumption, and netload across the seasonal variations and scalability. Dynamic Gaussian process and Quantile regression models are employed on the data of metropolitan area Sydney, Australia for probabilistic forecasting. Simulation results depict that the proposed model outperforms in all three scenarios of forecasting: solar power generation, electricity consumption, and netload. Authors in [16], investigated the recency effect of electricity load forecasting using preceding hours load and temperature. The aim is to determine lagged hourly temperature and daily moving average temperature to enhance forecast accuracy. The data used for network training and validation is of global energy competition 2012. The recency effect is investigated in three scenarios: aggregated level geographic hierarchy, bottom level geographic hierarchy, and individual level hours of the day. However, accuracy is enhanced at the cost of model complexity. In [17], proposed a long-term forecasting model to improve the relative forecast accuracy of electric utility resource integrated planning. The analysis was conducted on twelve Western US electric utility in the mid-2000s for both peak and normal energy consumption. Though single models are robust and fast converging, however, their accuracy is still low and not up to the required level.

2.2. Hybrid short-term load forecasting models

In hybrid models, the feature engineering and optimization modules are integrated with the forecaster module to improve the forecast accuracy and are suitable for the situation where accuracy is of prime importance. In [18], the authors presented an intelligent model to forecasts the load on distributed generation (DG) and examine the power supply structure. First, the support vector machine (SVM) and fruit-fly immune (FFI) algorithm are used to predict DG load. Second, a combined neural network and a polynomial regression model are used for power supply structure analysis about hourly load and weather factors. Finally, the impact of DG on the regional power system structure is analyzed in terms of load reduction on the main electric grid station. This combined intelligent model has a low-performance error and strong generalization. However, higher accuracy is achieved at the cost of slow convergence rate and high computational complexity.

Authors in [19], proposed an IoT-based deep learning system to forecast the future load with high precision. Moreover, the proposed method also qualitatively analyzed influencing factors such as variable climates, temperature, humidity, and social conventions that have a great impact on the forecast. However, the transfer of a huge amount of data on existing communication infrastructure is challenging.

In [20], the adaptive hybrid learning model (AHLM) is proposed to forecast the intensity of solar irradiation. The linear and dynamic behavior of data are captured by time-varying and multiple layer linear models. Also, a hybrid model of backpropagation (BP), GA, and neural network is used to learn the non-linear behavior of the data. The proposed AHLM learn linear, temporal, and non-linear behavior from the off-line data and predict the intensity of the solar with more precision. The proposed model outperforms for both short- and long-term forecast horizons.

To optimally harvest the potential of solar energy, forecasting of solar power is indispensable. Thus, the least absolute shrinkage and selection operator model is proposed for forecasting solar energy generation [21]. The proposed model is trained using historical weather data aiming not only to reduce prediction error but also to reveal the weather variables' significance in model training for forecasting. An algorithm is developed based on a single index, least absolute shrinkage, and selection operator models that maximize Kendall 's coefficient to estimate forecasting model coefficients. The goal of this algorithm is to ignore less important variables and increase the sparsity of the coefficient vector. With the proposed model, either prediction accuracy is improved or tradeoff between accuracy and complexity is achieved. However, accuracy is improved at the cost of more high system complexity.

For short-term load prediction, a hybrid model is proposed in [22]. This model is based on improved empirical mode decomposition, ARIMA, and wavelet neural network (WNN) optimized by the FFI optimization algorithm. For performance demonstration of the proposed model electric load data of Australian and New York electricity market are used. Simulation results show that the proposed model prediction is more accurate as compared to the existing models.

In [23], a deep learning-based electric load prediction model is proposed to forecast the future load. The proposed model extracts abstracted features using stacked denoising auto-encoders technique. With these abstracted features, the SVR model is trained to forecast the future load. The proposed model is evaluated by comparing it with plain SVR and ANN in terms of accuracy improvement.

The ANN model is used to forecast the hourly energy consumption of buildings in the Sugimoto Campus of Osaka City University, Japan [24]. The presented model is trained with Levenberg-Marquardt (LM) and BP algorithms. The six parameters are given as input such as dry bulb, humidity, temperature, global hourly irradiance, previous hourly, and weekly energy consumption. The accuracy of the proposed model is evaluated in terms of correlation coefficient and RMSE. Simulation results illustrate that RMSE is largest in the science and technology area of the university campus as compared to the humanities college area and old liberal arts area.

A novel type of hybrid system based on artificial intelligence is discussed in [25] to forecast 24 h load profile of the Polish grid station. The proposed hybrid system was tested on the off-line data of Poland and a few other countries. The MAPE varies from 1.08% to 2.26% in this scenario depending on the country. In the paper [26], an ensemble model based on empirical mode decomposition algorithm and deep learning is proposed for load forecasting. The proposed model is tested and validated on the electrical energy consumption datasets of the Australian energy market operator (AEMO). Moreover, the electric energy consumption data is decomposed into intrinsic mode functions (IMF) and the proposed model was used to model each of the IMF to improve the forecast accuracy. An autocorrelation function is for selecting input parameters and least squares SVM (LSSVM) is for forecasting is discussed in [27]. The main contribution of the paper is to provide a fully automated machine learning model without human intervention to forecast the future load. A hybrid incremental learning approach is proposed in [28], that is composed of discrete WT (DWT), empirical mode decomposition, and random vector functional link network (RVFLN), is discussed for short-term load forecasting. To evaluate the proposed model, the AEMO electricity load data is used. Simulation results depict that the proposed system is effective as compared to eight benchmark prediction methods.

In the paper [29], an extreme learning machine (ELM) model based on a mixed kernel for future load forecasting is discussed. The half-hour resolution electric load data is used to validate the proposed model. This the electric load data of the state of New South Wales, Victoria and Queensland in Australia. Simulation results illustrate that our proposed method is better as compared to the existing three methods like radial basis function-ELM (RBF-ELM), UKF-ELM, and mixed-ELM in terms of accuracy.

In [30], the authors proposed a hybrid of ELM and new switching delayed particle swarm optimization (PSO) algorithm for short-term load forecasting. The weights and biases are optimized with new switching delayed algorithm. Tanh function is used as an activation function because it has a better generalization problem and avoids the unnecessary hidden nodes and overtraining problem. Experimental results show that the proposed model outperforms the RBF neural network. The proposed model is successfully applied for short-term load forecasting in the power system.

A novel hybrid model, which is a combination of singular spectrum analysis (SSA), SVM, and cuckoo search (CS) algorithm, is proposed in [31] to forecast the future load. The historical data is pre-processed with the help of SSA. The pre-processed data is fed to the SVM model to forecast the future load and performance is optimized with the help of the CS algorithm. The performance of the proposed model is evaluated in terms of accuracy by comparing it with SVM, CS-SVM (CS-SVM), SSA-SVM (SSA-SVM), ARIMA, and BP neural network (BPNN).

In [32], the clustering-based hybrid model is proposed to predict the hourly electricity demand of hotel buildings. The operating buildings are non-stationary because of irregular electric temporal features. The on-line modified predictor model is proposed. The model is a combination of SVR and wavelet decomposition algorithm, which takes extracted training samples as input by fuzzy C means (FCM). The proposed model has improved accuracy as compared to the traditional models.

A deep neural network model is adopted for short-term load and probability density forecasting in [33]. The proposed model is evaluated on electricity consumption case studies of three Chinese cities for the year 2014. The simulation results demonstrate that: (i) deep learning-based model has better forecast accuracy relative to random forest and gradient boosting model, (ii) temperature, weather, and other environmental variables have a significant impact on electricity consumption, and iii) the probability density forecasting method can provide a high-quality prediction.

In [34], a hybrid forecast model is proposed, which is a combination of feature extraction technique and two-stage forecast engine. The two-stage forecast engine using Ridgelet neural network (RNN) and Elman neural network to provide accurate predictions. The optimization algorithm is applied to optimally select the control parameters for the forecast engine.

Authors in [35] proposed a short-term load forecasting model based on SSVR. The main objective is to improve relative forecast accuracy and efficiency. The relative forecast accuracy and efficiency are improved by giving the output of the forecast module to optimization module for fine-tuning of parameters. However, the forecast accuracy is improved at the cost of computational complexity.

A hybrid model of GA and non-linear AR with an exogenous neural network is proposed for short and medium-term forecasting in [36]. To fine-tune input parameters for the proposed model statistical and pattern recognition-based schemes are employed. The GA is used for selection weights and biases for the training of the neural network. The proposed model is validated by comparing it with the existing models such as average with exogenous inputs and regression tree models.

In [37], data-analytic based framework is proposed to forecast solar energy. The proposed framework is developed and validated on eight years (2005–2012) large dataset of a golden site of USA with a one-minute resolution taken from the national renewable energy laboratory (NREL). The uniqueness of this method is that data preprocessing is performed using integrated serial time-domain analysis coupled with multivariate filtering.

A short-term load forecasting framework based on dynamic mode decomposition is proposed in [38]. The proposed model improves prediction accuracy with the help of dynamic mode decomposition and extreme value constraint method.

Authors in [39] proposed robust short-term load forecasting framework with automatic data cleansing methods for distribution feeders load forecasting. A day-ahead building level load forecasting model based on deep learning is proposed [40]. The proposed deep learning model is validated by comparing with traditional models in terms of accuracy. An integrated framework of VMD, LSTM, and Bayesian optimization algorithm is proposed in [41]. The purpose of this model is outperforms exiting models in terms of both accuracy and stability. A hybrid model of modified multi-objective cuckoo search algorithm (CSA) and GNRR is proposed in [42]. The proposed model is tested on Australia energy market operator (AEMO) real-time load data in comparison to the existing models in terms of forecast accuracy.

Authors proposed a hybrid approach to forecasting the electricity production from solar panel-based microgrid in [43]. The hybrid model is based on GA, PSO, and neuro-fuzzy inference systems (NFIS). The proposed model is tested on real-time power generation data obtained from gold wind microgrid found in Beijing. The forecasting models

Table 1
Recent and relevant work brief summary in terms of strategies, objectives, repository, limitations, and critical remarks.

Strategies	Objectives	Repository	Limitations	Remarks
Weather information based electric load forecasting of a bulk power system [10]	Forecast accuracy improvement for effective performance of bulk power system	Fujian Province bulk power system China	This model is suitable and quite effective only for bulk power system	The performance of the model can be improved for both bulk and distributed power system by incorporating exogenous parameters
Household forecasting using DRNN [11]	To improve the comfort of the users by the reliable provision of electricity	Ireland commission for energy regulation	The complexity of the model is increased	The household load forecasting is possible only by sharing consumers consumption pattern to the commission of energy regulation, which is a threat to the consumers
LSTM-RNN based residential load forecasting [12],	Accuracy improvement to facilitate the residential consumers	Canadian residential load data	The proposed model improved only the meter level forecast accuracy	The accuracy is improved while the convergence rate is compromised
A big data approach for electric load forecasting [15]	Forecast accuracy improvement for scalable models	Global energy forecasting competition 2012	The model has a complex structure and slow convergence rate	The forecast accuracy is improved at the expense of high complexity
Intelligent model for forecasting based on SVM and FFI algorithm [17]	DG forecasting and regional power supply structure analysis	Certain area data in Northeast of Chine	The model is suitable for the short horizon of forecasting	The high accuracy and better generalization is obtained at cost of model complexity
IoT-based electric load forecasting [18]	Improvement of the accuracy and capability for effective power system operation	Electric load record of the urban area in south China	The framework has large complexity	The complex framework has direct on the impact the convergence rate
A deep model with stacked denoising auto-encoders for day-ahead load forecasting [22]	Forecast accuracy improvement	California electric load data	The system model performance is compromised with the decrease in the datasize	The accuracy is notably improved with large datasize however, the convergence rate is compromised
ANN-based prediction model [23]	To reduce the RMSE and improve forecast accuracy	Real data of Sugimoto Campus of Osaka City University Japan	Objectives are achieved at the cost of high convergence rate	The convergence rate is decreased due to the sigmoidal function and model complexity, which deny its application in real life
Intelligent hybrid model based load forecast [24]	Day-ahead electric load forecasting	Historical load data of Poland	To efficiently manage the generation of the Polish grid	The reliability of the Polish grid is improved at the cost of high modeling complexity

must have the ability to learn non-linear behavior of the consumers forms historical data most efficiently to forecast the future electric load. In this regard, ANN is one of the machine learning techniques mostly used to forecast future electric load due to easy and flexible implementation [44]. However, the performance of ANN is highly dependent on adjustable tuning parameters such as learning rate, number of layers, and number of neurons in the layers. The learning algorithms for training neural networks such as gradient descent, multivariate AR, and BP algorithm may suffer from premature convergence and overfitting [45]. To cure the aforementioned problems, hybrid forecast strategies in literature have been proposed. However, hybrid forecast strategies have improved modeling capabilities as compared to non-hybrid methods. Still, there is a problem of slow convergence and high execution time due to many adjustable parameters. In [46], the authors have used a Bi-level strategy, which is based on ANN and differential evolutionary algorithm (DEA) for electric load forecasting. An AFC-ANN and modified enhanced differential evolutionary algorithm (MEDEA) [47] based strategy is proposed to forecast the future load [10].

However, these strategies are highly dependent on the modular's knowledge and experience. Moreover, the performance of the aforesaid strategies is satisfactory for small data size and their performance is compromised as the size of data increases. There is no mechanism proposed to handle the large data (big data) and in real life, the data size is increasing dramatically. The proposed model has better performance as compared to existing ANN and linear regression-based models. The related work is comprehensively summarized in Table 1.

Three conclusions can be drawn from the above mentioned recent and relevant work: (i) there is no universal forecast model, which is perfect in all perspective, though, some models are better for some objectives and some conditions, (ii) there is a problem of overfitting, means a model outperforms in training and worst in forecasting, and (iii) there is a trade-off between forecast accuracy and convergence rate, when forecast accuracy is increased convergence rate will be compromised and vice versa. In this regard, a novel hybrid forecast model is proposed, which is an integrated framework of three modules: (i) MMM based data pre-processing and feature selection module, (ii) FCRBM based training and forecasting module, and (iii) GWDO based optimization module. The proposed model aims to perform high-quality electric load forecasting ranging from the day ahead to a week ahead of time horizon with comparatively high convergence speed for the decision making of SG.

3. Proposed system model

In this study , a novel hybrid model based on MMI technique, deep learning (FCRBM) model, and GWDO algorithm is proposed for short-term electric load forecasting, as shown in Fig. 1. This work is the extension of our earlier conference paper [48]. The earlier work is only for day-ahead load forecasting while the current is for both day and week ahead load forecasting with the novel concept of scalability. The proposed model is an integrated framework of three modules as illustrated in Fig. 1: (i) MMI based data pre-processing and feature selection module (ii) FCRBM based training and forecasting module, and (iii) our proposed GWDO algorithm-based optimization module. The entire implementation process of the proposed model is illustrated in in Fig. 2.

Before performing electric load forecasting, it is indispensable to identify the factors, which influence the load behavior. These influencing parameters include weather factors (humidity, temperature, and dew point), occupancy patterns, and calendar indicators. However, it is not feasible to apply all aforementioned candidate inputs to FCRBM based training and forecasting module. Moreover, the candidate inputs include ineffective features that complicate and degrade the performance of the model. Thus, the candidate inputs are first feed into the data pre-processing phase. . Then, pre-processed data is feed to the MMI based feature selection phase. The output of data pre-processing and features selection module is given as an input to training and forecasting module based on FCRBM. The output of this module is feed into optimization module based on GWDO, which is a novel contribution of this study. The optimization module first calculates error between the real and forecasted value. Then, it minimizes the error in order to make accurate predictions. The detailed demonstration of the proposed system model is as follows:

3.1. Data pre-processing and feature selection module

Let, E is the historical hourly load data of USA power grids taken from publicly available PJM electricity market, which is represented in the matrix form. This hourly electric load data is feed into the data pre-processing and feature selection module.

$$E = \begin{bmatrix} E(1,1) & E(2,1) & E(3,1) & E(4,1) & \dots & E(x,1) \\ E(1,2) & E(2,2) & E(3,2) & E(4,2) & \dots & E(x,2) \\ E(1,3) & E(2,3) & E(3,3) & E(4,3) & \dots & E(x,3) \\ E(1,4) & E(2,4) & E(3,4) & E(4,4) & \dots & E(x,4) \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ E(1,y) & E(2,y) & E(3,y) & E(4,y) & \dots & E(x,y) \end{bmatrix} \tag{1}$$

where $E(1, 1)$ is the electric load of first day first hour, $E(2, 1)$ is the electric load of second day first hour, such that $E(x, y)$ is the electric load of x^{th} day and y^{th} hour. The data is of four years having 1460 days and each day has 24 hours. The dimension of the data set is 1460×24.

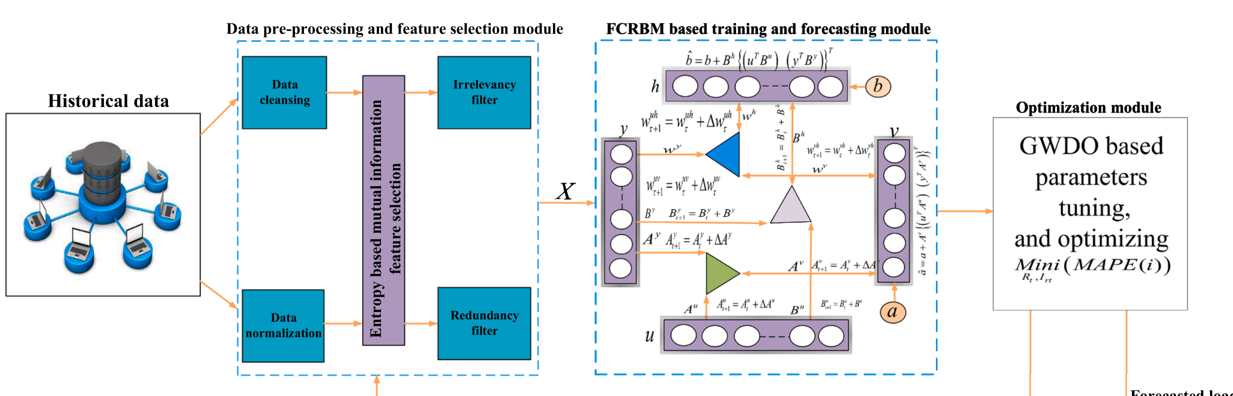


Fig. 1. Schematic diagram and main procedure of the FCRBM based proposed system model for hour and week ahead electric load prediction with hour resolution. Single arrowhead denotes one-way data flow and double arrowhead denotes two-way data flow.

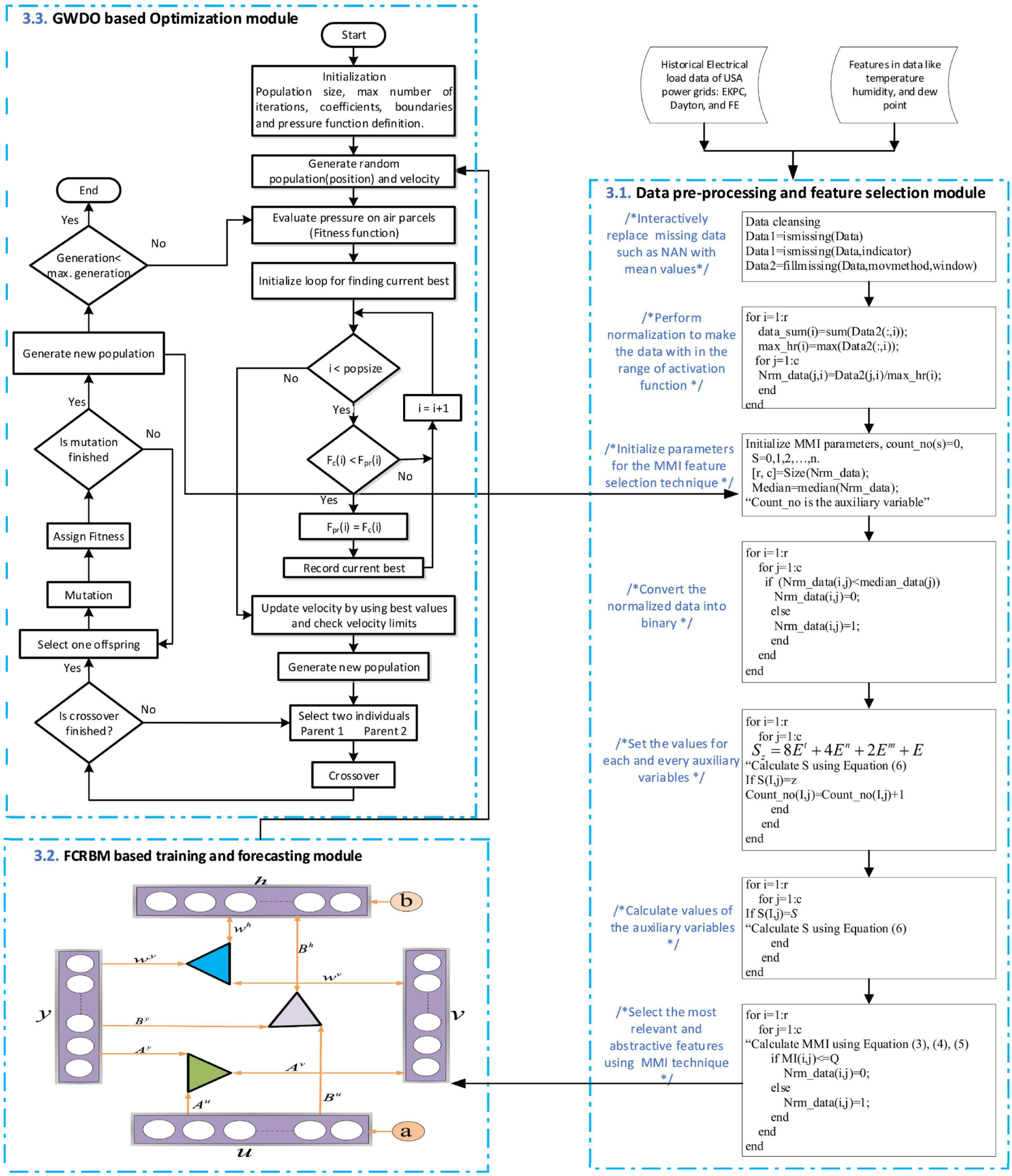


Fig. 2. Implementation process of the proposed system model an integrated framework of three modules: (i) 3.1 MMI based data pre-processing and feature selection module, (ii) 3.2 FCRBM based training and forecasting model, and (iii) 3.3 GWDO based optimization module.

Thus, the total number of data samples is 35040. The rows show the number of hours and columns show the number of days. The value of x is linked with the tuning of FCRBM training, larger the value of x implies fine-tuning and vice versa. There is a performance tradeoff between fine-tuning and convergence rate. This input data is first passed through the data cleansing phase, where defective and missing values are replaced by the average value of preceding days' electric load data. The cleansed data is passed through the normalization phase because the data have outliers and the weight matrix is extremely small, to make the overall weighted sum within the limits of the activation function. This data pre-processing is depicted in Fig. 2. In machine learning, feature selection is a process to select abstracted features and filter out unimportant features. The feature selection process has significant importance because it helps to avoid the curse of dimensionality and highly contributes to accuracy. In this regard, the entropy-based MI technique is a feature selection technique, which is used in a variety of taxonomy problems such as image processing, cancer categorization, image recognition, and data mining. The MI features selection technique is developed and used by [45,10] for features selection. In this work, the MMI feature selection technique is developed by devising modifications in MI technique subjected to accuracy and convergence rate. The cleansed and normalized data is passed through the MMI based feature selection phase to rank the inputs according to their information values. The ranked inputs are filtered using the irrelevancy and redundancy filters to remove irrelevant and redundant information. The subset of selected features contains best and more relevant information which highly contributes to the accuracy. First, the existing MI features selection technique is discussed. Then, the MMI feature selection technique will be discussed.

MI is a measure between two (possibly multi-dimensional) random variables E and E_t, that quantifies the amount of information obtained about one random variable, through the other random variable. The information found commonly in two random variables is of importance in our work, which is defined as mutual information between the two variables. The mathematical description is as follows:

$$H(E, E^t) = -\sum_i \sum_j p(E_i, E_j^t)\log_2(p(E_i, E_j^t)) \qquad \forall\ i, j \in \{1,\ 2\}, \tag{2}$$

where $p(E_i, E_j^t)$ is the joint probability of two random variables, E_i is the input random variables, and E_j^t is the target value. In feature selection, the information which is common among both variables are indispensable, which is formulated as in [45]:

$$MI(E, E^t) = \sum_i \sum_j p(E_i, E_j^t)\log_2\left(\frac{p(E_i, E_j^t)}{p(E_i)p(E_j^t)}\right), \tag{3}$$

where $MI(E_i, E_j^t)$ is used to find the mutual information between the two variables E_i and E_j^t. In this case, the candidate inputs are ranked by MI technique between input and the target value. From entropy-based MI technique, the following three conclusions can be drawn:

- If $MI(E_i, E_j^t) = 0$, it indicates that the discrete random variables E_i and E_j^t are irrelevant.
- If $MI(E_i, E_j^t)$ has some large value, it indicates that discrete random variables E_i and E_j^t are highly relevant.
- If $MI(E_i, E_j^t)$ has small value, it indicates that discrete variables E_i and E_j^t are lightly related.

In [45], among the training data samples last value of every hour of the day is chosen as the target value. The target value or last sample is very close to next day with respect to time and seems logical, however, it may cause serious forecast errors due to ignorance of average behavior while forecasting. In [10], the authors have used average value in addition to the target value because both average and target values are of equal importance. The Eq. (3) is modified for three variables as follows:

$$MI(E, E^t, E^n) = \sum_i \sum_j \sum_k p(E_i, E_j^t, E_k^n) \times \log_2\left(\frac{p(E_i, E_j^t, E_k^n)}{p(E_i)p(E_j^t)p(E_k^n)}\right), \tag{4}$$

where E_k^n is the average value, which indicates the second target. However, the average value will be very low, if some values in the selected features are very small. The addition of average with other two parameters is not enough because it may cause serious prediction problems due to ignorance of moving average. Thus, the Eq. (3) is modified for four variables as follows:

$$MI(E, E^t, E^m, E^n) = \sum_i \sum_j \sum_k \sum_l p(E_i, E_j^t, E_k^n, E_l^m) \times \log_2\left(\frac{p(E_i, E_j^t, E_k^n, E_l^m)}{p(E_i)p(E_j^t)p(E_k^n)p(E_l^m)}\right), \tag{5}$$

where the third target value E_l^m is moving average. The Eq. (5) is expanded for binary encoded information as in Eq. (6) as:

$$\begin{aligned}
&MI(E_i, E^t, E_k^n, E_l^m)= \\
&\quad p(E_i = 0, E_j^t = 0, E_k^n = 0, E_l^m = 0) \times \log_2\left(\frac{p(E_i = 0, E_j^t = 0, E_k^n = 0, E_l^m = 0)}{p(E_i = 0)p(E_j^t = 0)p(E_k^n = 0)p(E_l^m = 0)}\right) \\
&\quad + p(E_i = 0, E_j^t = 0, E_k^n = 0, E_l^m = 1) \times \log_2\left(\frac{p(E_i = 0, E_j^t = 0, E_k^n = 0, E_l^m = 1)}{p(E_i = 0)p(E_j^t = 0)p(E_k^n = 0)p(E_l^m = 1)}\right) \\
&\quad + p(E_i = 0, E_j^t = 0, E_k^n = 1, E_l^m = 0) \times \log_2\left(\frac{p(E_i = 0, E_j^t = 0, E_k^n = 1, E_l^m = 0)}{p(E_i = 0)p(E_j^t = 0)p(E_k^n = 1)p(E_l^m = 0)}\right) \\
&\quad + p(E_i = 0, E_j^t = 0, E_k^n = 1, E_l^m = 1) \times \log_2\left(\frac{p(E_i = 0, E_j^t = 0, E_k^n = 1, E_l^m = 1)}{p(E_i = 0)p(E_j^t = 0)p(E_k^n = 1)p(E_l^m = 1)}\right) \\
&\quad + p(E_i = 0, E_j^t = 1, E_k^n = 0, E_l^m = 0) \times \log_2\left(\frac{p(E_i = 0, E_j^t = 1, E_k^n = 0, E_l^m = 0)}{p(E_i = 0)p(E_j^t = 1)p(E_k^n = 0)p(E_l^m = 0)}\right) \\
&\quad + p(E_i = 0, E_j^t = 1, E_k^n = 0, E_l^m = 1) \times \log_2\left(\frac{p(E_i = 0, E_j^t = 1, E_k^n = 0, E_l^m = 1)}{p(E_i = 0)p(E_j^t = 1)p(E_k^n = 0)p(E_l^m = 1)}\right) \\
&\quad + p(E_i = 0, E_j^t = 1, E_k^n = 1, E_l^m = 0) \times \log_2\left(\frac{p(E_i = 0, E_j^t = 1, E_k^n = 1, E_l^m = 0)}{p(E_i = 0)p(E_j^t = 1)p(E_k^n = 1)p(E_l^m = 0)}\right) \\
&\quad + p(E_i = 0, E_j^t = 1, E_k^n = 1, E_l^m = 1) \times \log_2\left(\frac{p(E_i = 0, E_j^t = 1, E_k^n = 1, E_l^m = 1)}{p(E_i = 0)p(E_j^t = 1)p(E_k^n = 1)p(E_l^m = 1)}\right) \\
&\quad + p(E_i = 1, E_j^t = 0, E_k^n = 0, E_l^m = 0) \times \log_2\left(\frac{p(E_i = 1, E_j^t = 0, E_k^n = 0, E_l^m = 0)}{p(E_i = 1)p(E_j^t = 0)p(E_k^n = 0)p(E_l^m = 0)}\right) \\
&\quad + p(E_i = 1, E_j^t = 0, E_k^n = 0, E_l^m = 1) \times \log_2\left(\frac{p(E_i = 1, E_j^t = 0, E_k^n = 0, E_l^m = 1)}{p(E_i = 1)p(E_j^t = 0)p(E_k^n = 0)p(E_l^m = 1)}\right) \\
&\quad + p(E_i = 1, E_j^t = 0, E_k^n = 1, E_l^m = 1) \times \log_2\left(\frac{p(E_i = 1, E_j^t = 0, E_k^n = 1, E_l^m = 1)}{p(E_i = 1)p(E_j^t = 0)p(E_k^n = 1)p(E_l^m = 1)}\right) \\
&\quad + p(E_i = 1, E_j^t = 1, E_k^n = 0, E_l^m = 0) \times \log_2\left(\frac{p(E_i = 1, E_j^t = 1, E_k^n = 0, E_l^m = 0)}{p(E_i = 1)p(E_j^t = 1)p(E_k^n = 0)p(E_l^m = 0)}\right) \\
&\quad + p(E_i = 1, E_j^t = 1, E_k^n = 0, E_l^m = 1) \times \log_2\left(\frac{p(E_i = 1, E_j^t = 1, E_k^n = 0, E_l^m = 1)}{p(E_i = 1)p(E_j^t = 1)p(E_k^n = 0)p(E_l^m = 1)}\right) \\
&\quad + p(E_i = 1, E_j^t = 1, E_k^n = 1, E_l^m = 0) \times \log_2\left(\frac{p(E_i = 1, E_j^t = 1, E_k^n = 1, E_l^m = 0)}{p(E_i = 1)p(E_j^t = 1)p(E_k^n = 1)p(E_l^m = 0)}\right) \\
&\quad + p(E_i = 1, E_j^t = 1, E_k^n = 1, E_l^m = 1) \times \log_2\left(\frac{p(E_i = 1, E_j^t = 1, E_k^n = 1, E_l^m = 1)}{p(E_i = 1)p(E_j^t = 1)p(E_k^n = 1)p(E_l^m = 1)}\right)
\end{aligned} \tag{6}$$

A supplementary S_z variable is defined in Eq. (10) to find the joint and individual probabilities, such that:

$$S_z = 8E^t + 4E^n + 2E^m + E \quad \forall \quad E, E^t, E^n, E^m \in \{0, 1\}, \tag{7}$$

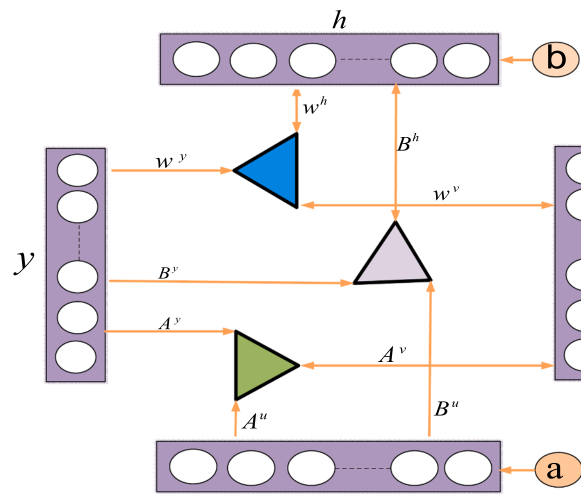


Fig. 3. Architecture of factored conditional restricted Boltzmann machine, where u is the history input layer, h is the hidden layer, y is the style layer, and v is the visible output layer.

where $S_z \in \{0, 1, 2, 3, ...,15\}$. S_z counts the number of zeros, ones, twos, threes, and finally the number of fifteens. From the aforesaid discussion, the joint and individual probabilities can be find using Eq. (8) subjected to Eq. (6) as:

$$
\begin{aligned}
p(E=0) &= \frac{S_{0z}+S_{2z}+S_{4z}+S_{6z}+S_{8z}+S_{10z}+S_{12z}+S_{14z}}{L}\\
p(E=1) &= \frac{S_{1z}+S_{3z}+S_{5z}+S_{7z}+S_{9z}+S_{11z}+S_{13z}+S_{15z}}{L}\\
p(E^n=0) &= \frac{S_{0z}+S_{1z}+S_{4z}+S_{5z}+S_{8z}+S_{9z}+S_{12z}+S_{13z}}{L}\\
p(E^n=1) &= \frac{S_{2z}+S_{3z}+S_{6z}+S_{7z}+S_{10z}+S_{11z}+S_{14z}+S_{15z}}{L}\\
p(E^m=0) &= \frac{S_{0z}+S_{1z}+S_{2z}+S_{3z}+S_{8z}+S_{9z}+S_{10z}+S_{11z}}{L}\\
p(E^m=1) &= \frac{S_{4z}+S_{5z}+S_{6z}+S_{7z}+S_{12z}+S_{13z}+S_{14z}+S_{15z}}{L}\\
p(E^t=0) &= \frac{S_{0z}+S_{1z}+S_{2z}+S_{3z}+S_{4z}+S_{5z}+S_{6z}+S_{7z}}{L}\\
p(E^t=1) &= \frac{S_{8z}+S_{9z}+S_{10z}+S_{11z}+S_{12z}+S_{13z}+S_{14z}+S_{15z}}{L}
\end{aligned}
\tag{8}
$$

The Eqs. (5)–(8) are the MMI technique equations, which are used to find the mutual information between the four variables such as E_i, E^t, E^n and, E^m. The candidate inputs are ranked based on this mutual information to remove irrelevant and redundant information. The MMI feature selection technique provides twofold benefits: (a) selection of suitable and relevant features minimizes the forecast error, and (b) selection of the subset of features improves the convergence rate. The step by step procedure of overall data pre-processing and feature selection module is illustrated in Fig. 2.

Before feeding them to the training and forecasting module, the selected features are split into training, testing, and validation data samples for training and validation of the FCRBM. The selected subset of key features is given as an input to the training and forecasting module based on FCRBM.

3.2. FCRBM based training and forecasting module

This module aims to devise a framework that is enabled via learning to forecast day and week ahead electric load. In this regard, in literature, a wide variety of short-term load forecasting models such as dynamic regression, transfer function, and AR heteroscedastic, have been proposed. Although, these models are capable of linear predictions and the behavior of load is non-linear and stochastic. To solve the aforesaid problems, the authors used novel strategies for short-term load forecasting in [49,50] based on ANN. These forecasting strategies are capable to handle the non-linear behavior of electric load and forecast the future load. However, the performance of these strategies is compromised with the increase in data size. The deep learning models such as RBM, conditional RBM (CRBM), and FCRBM have better performance at the large datasets. These models have deep layers layout, which can capture the highly abstracted features. Thus, FCRBM from the pool of deep learning models is selected to forecast the future electric load because it provides high-quality forecasting. The training and forecasting module is based on FCRBM, which is an indispensable part of our proposed hybrid forecasting model. At first, the architecture of the FCRBM model is determined. The model has four layers along with neurons, i.e., hidden layers, visible layer, style layer, and history layer as depicted in [51]. As discussed earlier, the FCRBM network must be trained and enabled via learning to forecast the future electric load. Generally, learning is of 3 types, i.e., supervised, unsupervised, and reinforced. Since in our scenario, we use historical load data, thus we use supervised learning. In literature, many supervised learning activation functions exist such as sigmoidal, tangent hyperbolic, rectified linear unit (ReLU), and SoftMax. However, we chose the ReLU activation function as shown in Eq. (9) because it has faster convergence and also overcomes the problem of vanishing gradient.

$$
\begin{aligned}
&f(X, b) = \max(0,\ \beta(X, b))\\
&\Delta f(X, b)\begin{cases}1 & \text{if } \beta(X, b) \geqslant 0\\ 0 & \text{otherwise,}\end{cases}
\end{aligned}
\tag{9}
$$

where X is the selected candidate inputs, b indicates bias value, and β is for steepness control of activation function. The FCRBM architecture is described in detail in the following subsection.

3.2.1. Architecture of the FCRBM

FCRBM is an extension of the CRBM introduced by Taylor and Hinton in [52]. In FCRBM [53], they add the concept of factor and styles to mimic multiple human actions (as shown in Fig. 3). Its contrastive divergence does not suffer from the issue of vanishing gradient as in back-propagation. It has a rich, distributed hidden state which permits the simple and exact inference that helps in preserving the temporal information present in the electrict load time series data [54]. We proposed a new way to adopt deep learning technique, i.e., FCRBM for short-term load forecasting, where the successive layers take the output from preceding trained layers and improve the forecast accuracy.

The FCRBM comprises of four layers as shown in Fig. 3: visible layer v, history layer u, hidden layer h, and style layer y. The visible and history layers are real-valued while the hidden layer is binary. These layers are significant for the proper operation of FCRBM. The visible layer is responsible for encoding the present time series data to forecast the future value, while the history layer will encode historical time series data. Hidden layer is responsible for the discovery of significant features that are required for analysis. The different styles and parameters, which are essential for forecasting, are embedded into the style layer. The last style layer represents multiple parameters that are important for load forecasting. The relation and interaction between the layers, weights, and factors are expressed by an error function as:

$$
E(v, u, h; w) = -v^T\hat{a} - h^T\hat{b} - \sum\{(v^Tw^v)\circ(y^Tw^y)\circ(h^Tw^h)\}
\tag{10}
$$

where E is the error function, v^Tw^v is the visible factored, y^Tw^y is the style factored, and h^Tw^h is the hidden factored. It $\circ$ is a hadamard product used for element wise multiplication. The $\hat{a}$ and $\hat{b}$ elements represent dynamic biases associated with visible and hidden layers, respectively, which are defined as follows:

$$
\begin{aligned}
\hat{a} &= a + A^v\{(u^TA^u)\circ(y^TA^y)\}^T\\
\hat{b} &= b + B^h\{(u^TB^u)\circ(y^TB^y)\}^T
\end{aligned}
\tag{11}
$$

where w^v, w^y, w^h are weights of the corresponding layers and $A^v, A^u, A^y, B^h, B^u, B^y$ are the connections of the corresponding layers to factors, these are also known as model free parameters. The connections and weights are the parameters that must be fine tuned and trained for accurate performance of deep learning technique FCRBM.

3.2.2. Conditional probability

In case of FCRBM, conditional probability determines probability distribution of one layer conditioned over all the remaining layers. In first case, we define probability distribution of hidden layer conditioned over all the remaining layers $p(h|v, u, y)$. There is no intra-layer connection between the neurons of the same layer, but inter-layer connection between the neurons of different layers. The conditional probability of hidden layer can be calculated as:

$$p(h|v, u, y) = \text{ReLU}[\hat{b} + w^h\{(v^T w^v)\circ(y^T w^y)\}] \tag{12}$$

where ReLU is defined as in Eq. (9).

For all inputs, probability of hidden layer neurons is evaluated using ReLU activation function.

In second case, we determine the probability of the visible layer i.e., $p(v|h, u, y)$ conditioned over remaining layers. The conditional probability of visible is defined as:

$$p(v|h, u, y) = \text{ReLU}[\hat{a} + w^v\{(h^T w^h)\circ(y^T w^y)\}] \tag{13}$$

Finally, we define the joint probability distribution of visible and hidden layer neurons conditioned on history layer, style layer, and model parameters $p(v, h|u, y, ...)$. The restriction is that there is no intra-layer connection between the neurons while there is only inter-layer connection between the neurons of different layers. The joint probability is calculated as:

$$\begin{aligned} p(v, h|u, y, ...) &= \text{ReLU}([\hat{b} + w^h\{(v^T w^v)\circ(y^T w^y)\}]\times \\ &\quad [\hat{a} + w^v\{(h^T w^h)\circ(y^T w^y)\}]) \end{aligned} \tag{14}$$

Eq. (14) represents the joint probability distribution of visible and hidden layer neurons.

3.2.3. FCRBM weights and biases learning rules

We adopt stochastic gradient decent for learning and updating rules to overcome the problem of vanishing gradient. Moreover, the stochastic gradient decent converges faster and avoids overfitting on large datasets as compared to batch gradient decent and mini-batch gradient decent algorithms [55]. The gradient of the weights for each layer is calculated as:

$$\begin{aligned} \Delta w_t^h &= -\eta\tfrac{\partial E}{\partial w^h} \\ \Delta w_t^v &= -\eta\tfrac{\partial E}{\partial w^v} \\ \Delta w_t^y &= -\eta\tfrac{\partial E}{\partial w^y} \end{aligned} \tag{15}$$

For each layer the gradient of connections are calculated as follows:

$$\begin{aligned} \Delta A_t^u &= -\eta\tfrac{\partial E}{\partial A^u} \\ \Delta A_t^v &= -\eta\tfrac{\partial E}{\partial A^v} \\ \Delta A_t^y &= -\eta\tfrac{\partial E}{\partial A^y} \end{aligned} \tag{16}$$

$$\begin{aligned} \Delta B_t^u &= -\eta\tfrac{\partial E}{\partial B^u} \\ \Delta B_t^h &= -\eta\tfrac{\partial E}{\partial B^h} \\ \Delta B_t^y &= -\eta\tfrac{\partial E}{\partial B^y} \end{aligned} \tag{17}$$

The gradient of dynamic biases are as follow:

$$\begin{aligned} \Delta\hat{a} &= -\eta\tfrac{\partial E}{\partial v} \\ \Delta\hat{b} &= -\eta\tfrac{\partial E}{\partial h} \end{aligned} \tag{18}$$

The weights of corresponding layers are updated as:

$$\begin{aligned} w_{t+1}^h &= w_t^h + \Delta w_t^h \\ w_{t+1}^y &= w_t^y + \Delta w_t^y \\ w_{t+1}^v &= w_t^v + \Delta w_t^v \end{aligned} \tag{19}$$

The connections are updated as follows:

$$\begin{aligned} A_{t+1}^u &= A_t^u + \Delta A_t^u \\ A_{t+1}^v &= A_t^v + \Delta A_t^v \\ A_{t+1}^y &= A_t^y + \Delta A_t^y \end{aligned} \tag{20}$$

$$\begin{aligned} B_{t+1}^u &= B_t^u + \Delta B_t^u \\ B_{t+1}^h &= B_t^h + \Delta B^h t \\ B_{t+1}^y &= B_t^y + \Delta B^y t \end{aligned} \tag{21}$$

The dynamic biases are updated as follows:

$$\begin{aligned} \hat{a}_{t+1} &= \hat{a}_t + \Delta\hat{a}_t \\ \hat{b}_{t+1} &= \hat{b}_t + \Delta\hat{b}_t \end{aligned} \tag{22}$$

where Eq. (19) is weight update equation for each layer, Eqs. (20)–(22) are dynamic biases update equations.

The training and learning procedure iterate for the number of epochs to enable the network for forecasting. The FCRBM is enabled via training and learning to forecast the future electric load. Moreover, the performance metric, MAPE, is considered as validation error, which is formulated as follows:

$$MAPE = \left(\frac{1}{\tau}\sum_{t=1}^{\tau}\frac{|T_t - F_t|}{|T_t|}\right)\times 100, \tag{23}$$

where T_t represents actual load values, F_t indicates forecasted load values, and τ represents number of hours under consideration. Further details of the FCRBM working and learning activation function can be found in [56]. The output of this module is fed into the GWDO based optimization module to further improve forecast accuracy with an affordable convergence rate.

3.3. GWDO algorithm based optimization module

The objective of this module is to minimize the forecast error with an affordable convergence rate. The authors used DEA [45] and MEDEA [10] with forecaster module to optimize the performance of the model. Both algorithms have a slow convergence rate and low precision [57]. Furthermore, the aforementioned algorithms are trapped in local optimum [57]. To remedy the aforementioned problems, the GWDO algorithm is proposed, which is a hybrid of WDO and GA algorithms [58]. The step by step procedure is given in Fig. 4 and their parameters are listed in Table 2. The proposed algorithm takes benefit from the features of both algorithms (GA and WDO). The GA enables the diversity of the population and WDO has faster convergence. The GWDO based module receives the forecasted load with some error that is minimum as per the ability of FCRBM. This forecasting error can be minimized with the proposed GWDO optimization technique. The sole objective of GWDO based optimization module is to fine-tune the adjustable parameters of the model to improve forecast accuracy with an affordable convergence rate. In other words, the optimization module is integrated with the FCRBM based forecaster to minimize error and improve the forecast accuracy. Thus, error minimization (MAPE) becomes the objective function of the optimization module, which is mathematically modeled as:

$$\underset{R_t, I_{rt}}{Mini}\, MAPE(j) \quad \forall\ j \in \{1, 2, 3, ...\tau\}, \tag{24}$$

where R_t and I_{rt} are the thresholds of redundancy and irrelevancy, respectively. The GWDO based optimization module feed the optimized values of the thresholds to MMI based feature selection module to select key features from the given data. The integration of the optimization module to the forecasting model increase the execution time, which

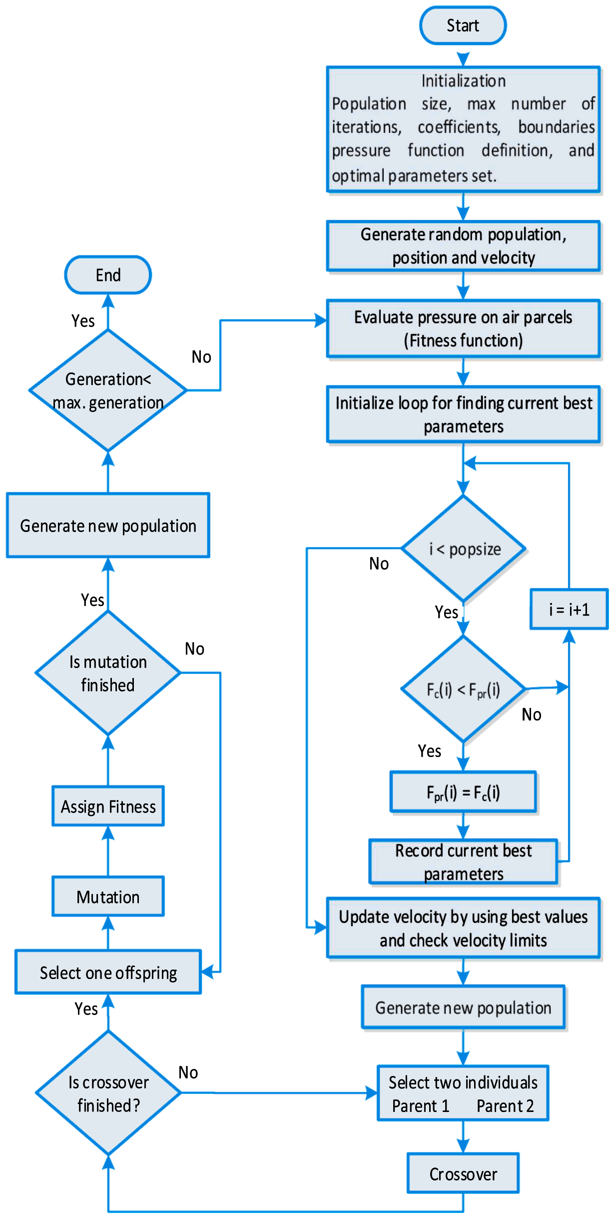


Fig. 4. Our proposed GWDO algorithm used in the optimization module to optimize error performance.

Table 2
Parameters used in simulations.

Parameters	Values
Population size	24
Number of decision variables	2
Number of iterations	100
RT	3
g	0.2
α	0.4
dimMin	−5
dimMax	5
vmax	0.3
vmin	−0.3
crossoverrate	0.9
mutation rate	0.1
Learning rate	0.0001
Weight decay	0.0002
Momentum	0.5

disturbs the convergence rate because of the tradeoff between execution time and convergence rate. The integration of the optimization module is favorable for applications where forecast accuracy is of primary importance and vice versa. Our proposed GWDO algorithm among the heuristic algorithms is preferred due to the following reasons: (i) it avoids premature convergence and (ii) it has faster convergence. The GWDO algorithm randomly produce population, i.e., the position 25 and velocity matrix 26 as in [58]:

$$\begin{cases} x_{new} = 1 \ \ if \ \ rand(1) \leqslant sig(j, i) \\ x_{new} = 0 \ if \ \ rand(1) > sig(j, i) \end{cases} \tag{25}$$

$$v_i = vmax \times 2 \times (rand(populationsize, \ \ n) - 0.5) \tag{26}$$

The fitness functions for velocity and position are defined as in Eqs. (27) and (28) because the position vector and velocity vector will now be updated by comparing random number ($rand(.) \in [0, 1]$ with fitness function ($FF(.) \in [0, 1]$) as shown in Eq. (29).

$$FF(v(i)) = \frac{MAPE(x_{new}(i))}{MAPE(v(i)) + MAPE(x_{new}(i))} \tag{27}$$

$$FF(x_{new}(i)) = \frac{MAPE(v(i))}{MAPE(x_{new}(i)) + MAPE(v(i))} \tag{28}$$

If random number is less than fitness function, then load value will be update because our objective function is minimization problem.

$$F_{pr}(i) = \begin{cases} v_n(i) & if \ \ rand(i) \leqslant FF(v(i)) \\ x_{new}^n(i) & if \ \ rand(i) \leqslant FF(x_{new}(i)) \end{cases} \tag{29}$$

Now there is question mark, why load update has influence on random value. We cure this problem by eliminating the load update influence on random number, now the comparison is between fitness function of the candidate input to the previous one as shown in Eq. (30). Thus, the selected load update value will have high quality of accuracy.

$$F_{pr+1}(i) = \begin{cases} v_{n+1}(i) & \frac{v_n(i)}{v_n(i_{\max})} \leqslant FF(v(i)) \\ x_{new}^{n+1}(i) & \frac{x_{new}^n(i)}{x_{new}^n(i_{\max})} \leqslant FF(x_{new}(i)) \end{cases} \tag{30}$$

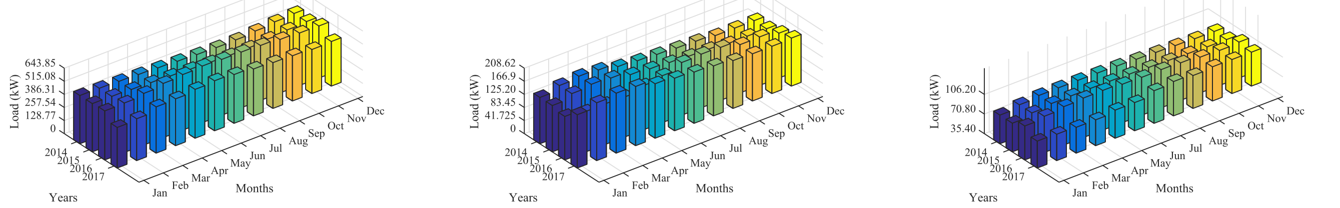
With the integration of GWDO algorithm based optimization module, the accuracy is improved while the convergence rate is compromised because there is a trade-off between accuracy and convergence rate. However, the proposed short-term load forecasting model outperforms the existing models like MI-ANN [45], Bi-level [46], and AFC-ANN [10] in terms of accuracy. It is because ANN-based models have a shallow layout and their performance is degraded with the increase in datasize. The FCRBM has improved performance with the large datasize due to its deep layers' layout.

4. Simulation results and discussions

For performance evaluation of the proposed short-term load forecasting model, simulations are conducted in MATLAB. In simulations, the proposed model is compared with existing short-term load forecasting models like AFC-ANN [10], MI-ANN [45], Bi-level [46], and LSTM. The aforementioned models are selected as benchmark models due to their closer architectural similarities with the proposed model. Two performance metrics i.e., accuracy and convergence rate are used for performance evaluation. Accuracy is defined as accuracy = 100-MAPE and is measured in percentage (%). The execution time is defined as the time spent by the forecasting strategy during execution and is measured in seconds. The detailed demonstration is as follows.

4.1. Description of the benchmark dataset

Historical hourly electric load data is taken from publicly available PJM electricity market [59] for performance evaluation of the proposed



(a) FE grid (b) Dayton grid (c) EKPC grid

Fig. 5. Historical hourly electric load data of four years ranging from 2014 to 2017 of FE, Dayton, and EKPC power grids of the USA with month and year indexes.

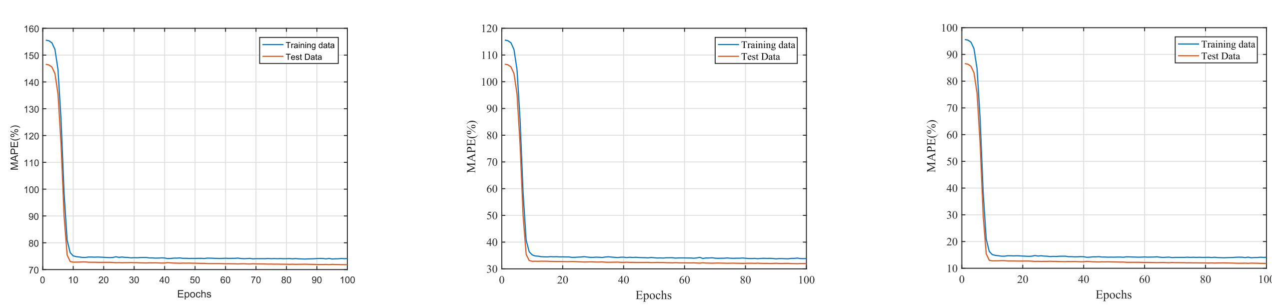
model. Different variables (humidity, temperature, dew point, and hour of the day) that are typical for the training forecasting model are used. The four years (2014–2017) historical hourly load data of three USA power grids is depicted in Fig. 5. The data consists of humidity, temperature, and load values. It is obvious from Fig. 5 that the FE power grid is serving the highly-dense area and have the highest load profile while the Dayton power grid has less energy consumption with a lower load profile as compared to FE power grid and more energy consumption as compared to EKPC power grid. The dataset is passed through a data pre-processing and feature selection module where abstracted features from the given dataset are extracted. The subset (abstracted features) of data is divided into training, and testing data samples. The three years of data are used to train the network and one year of data is used to test the network. The training data samples are from 2014–2016, consisting of the input vector, the aforementioned variables, and the target measured load profile. The testing data samples are of one-year 2017, which is used for testing purpose. The validation data samples are constructed from training data samples to exploit validation error for proper parameter selection.

4.2. Evaluation of learning curve

A learning curve is a graphical representation that compares the performance of models on training and testing data samples across a varying number of epochs. The learning curve enables us to verify whether the chosen model is learning or memorizing the data. When there are high variance and bias, the learning curve is bad, and the model is memorizing not learning. Due to high bias, the training and testing error rate is high and the convergence rate is fast. In contrast, the high variance occurs, when the gap between training and testing errors is large. In both cases, the model is not good and leads to poor generalization. Overfitting occurs when the test error at a certain point starts to increase and training error decrease. This shows that the model is memorizing but not learning. Thus, such a model leads to bad generalization. The overfitting problem is prevented using the dropout method and early stopping [60]. However, in the case of FCRBM, it is observed that the testing error gradually decreases as the training error does for FE, Dayton, and EKPC power grids of USA as illustrated in Fig. 6. Thus, the FCRBM model resolved the problem of overfitting. Moreover, the gap between training error and testing error is small and there is no bias and variance as clearly depicted in Fig. 6 for FE, Dayton, and EKPC power grids of USA.

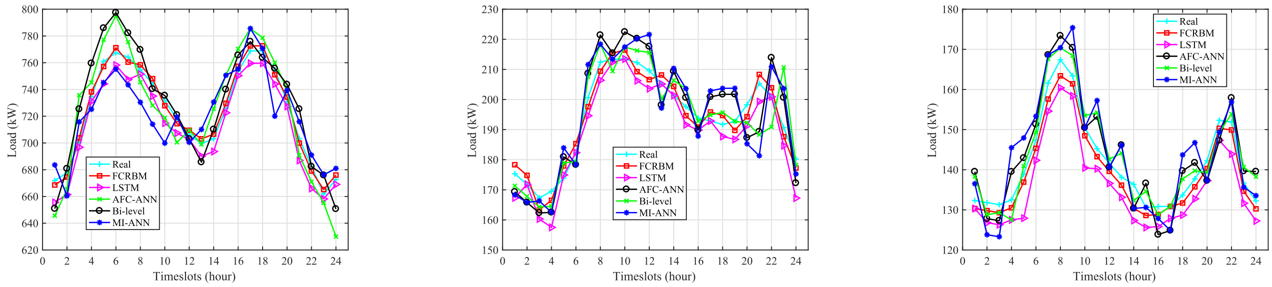
4.3. Evaluation of actual and forecasted load for day ahead time horizon with hour resolution

The day ahead forecasted electric load profile with hour resolution of the proposed model and benchmark models like LSTM, MIANN, Bi-level, and AFC-ANN for three USA power girds (EKPC, FE, and Dayton) is depicted in Fig. 7. It is obvious from the graphical illustration that all prediction models (our proposed FCRBM based model, and four benchmark models like Bi-level, MI-ANN, AFC-ANN, and LSTM) are capable to capture non-linear behavior of load from historical data and based on the captured behavior forecast future electric load. It is also clear that models like Bi-level, MI-ANN, AFC-ANN, and LSTM use the sigmoidal activation function, Levenberg-Marquardt, and multi-variate AR algorithms for network training. In contrast, the adapted FCRBM network is trained using ReLU due to having small execution time. It is verified from the Fig. 7 that the proposed model closely follows the target curve as compared to the benchmark models like AFC-ANN, Bi-level, MI-ANN, and LSTM for all three USA power grids: FE, Dayton, and EKPC. The day ahead with hour resolution forecasted load numerical observations in terms of MAPE for EKPC, FE, and Dayton, power grids are listed in Tables 3–5, respectively. The day ahead with hour resolution forecasted load based on FCRBM based model, and benchmark models like LSTM, MI-ANN, Bi-level, and AFC-ANN numerical results are listed in Table 3. The MAPE error of the proposed FRCBM based model is 0.4920%, the Bi-level model is 2.5186%, the MI-ANN model is 4.3371%, the AFC-ANN model is 2.4741%, and LSTM model is 2.7582%. The MAPE of the proposed model is lower as compared to benchmark models, lower MAPE results in better accuracy. The forecasted load of AFC-ANN is better than Bi-level, and Bi-level is better than MI-ANN in terms of accuracy. The reason for this comparatively better performance is that AFC-ANN model used MEDEA for optimization and the Bi-level model used DEA for optimization, which improves the forecast accuracy by minimizing the error. Although, this



(a) FE grid (b) Dayton grid (c) EKPC grid

Fig. 6. The deep learning model FCRBM learning evaluation on FE, Dayton, and EKPC power grids hourly load data ofthe USA.



(a) FE grid (b) Dayton grid (c) EKPC grid

Fig. 7. Day ahead with hour resolution evaluation of actual and predicted load in terms of forecast accuracy of the proposed FCRBM based model and benchmark models like LSTM, MI-ANN, AFC-ANN, and Bi-level on FE, Dayton, and EKPC power grids hourly load data of the USA.

improved accuracy is obtained by paying a cost of increased execution time. The proposed FCRBM based model outperforms Bi-level, MI-ANN, and AFC-ANN models due to the integration of MMI based feature selector and GWDO based optimization module in the framework. Numerical observations of the proposed model and benchmark models in terms of MAPE for the Dayton power grid are in Table 4. The MAPE of the proposed model, Bi-level, MI-ANN, AFC-ANN, LSTM is 0.4998%, 3.0220%, 4.4988%, 2.7633%, and 2.5432%, respectively. Furthermore, the future forecasted load based on FCRBM is more accurate as compared to the benchmark models like Bi-level, MI-ANN, AFC-ANN, and LSTM due to the use of MMI technique, deep neural network, and GWDO algorithm. The numerical results of the proposed model and benchmark models in terms of MAPE for the EKPC power grid are listed in Table 5. The MAPE of the proposed model is 0.4525%, Bi-level is 2.4202%, MI-ANN is 4.3280%, AFC-ANN is 2.7530%, and LSTM is 3.1234%. Thus, the proposed model is better as compared to the benchmark models like Bi-level, MI-ANN, AFC-ANN, and LSTM in terms of forecast accuracy. From the results and discussion, we conclude that the proposed deep learning model FCRBM has superior performance as compared to benchmark models. The average numerical results in terms of MAPE of FCRBM based model for FE power grid is 0.4920%, for Dayton power grid is 0.4998%, and for EKPC power grid 0.4525%, which is less as compared to the benchmark models.

4.4. Evaluation of proposed and benchmark models in terms of convergence rate

Performance evaluation of the proposed deep learning model FCRBM and benchmark models like LSTM, MI-ANN, Bi-level, and AFC-ANN in terms of convergence rate for three USA power grids (EKPC, FE, and Dayton) is depicted in Fig. 8. There is a trade-off between forecast accuracy and convergence rate. The accuracy of the Bi-level strategy has improved as compared to the MI-ANN model. This improved accuracy is achieved at the cost of more execution time because DEA based optimization module is integrated with a Bi-level strategy. It is obvious from the Fig. 8a, b, and c that the execution time is increased

Table 3
Evaluation of actual and predicted load (P.load) in terms of MAPE of the proposed FCRBM based model and benchmark models like LSTM, MI-ANN, AFC-ANN, and Bi-level on FE power grid hourly load data of USA.

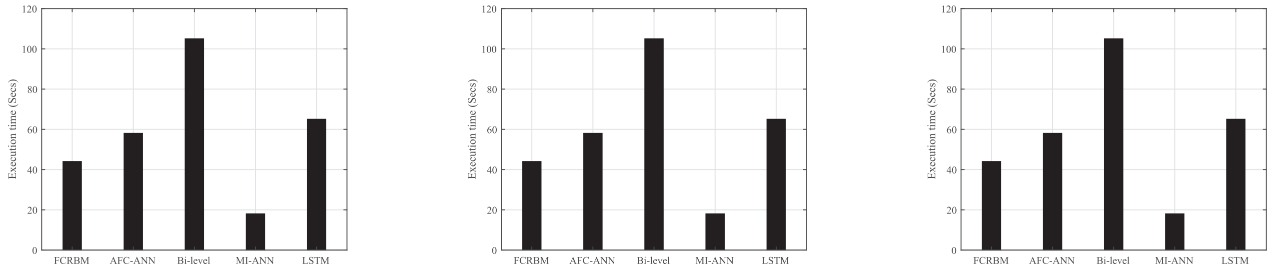
Hours	Target (kW)	Proposed and benchmark electric load forecasting models									
		FCRBM		AFC-ANN		Bi-level		MI-ANN		LSTM	
		P.load (kW)	MAPE (%)	P.load (kW)	MAPE (%)	P.load (kW)	MAPE (%)	P.load (kW)	MAPE (%)	P.load (kW)	MAPE (%)
00.00	671.8923	668.5829	0.4925	683.5829	1.7400	645.5829	3.9157	650.8923	3.1255	651.8030	3.3370
01.00	677.7923	674.4538	0.4926	660.4538	2.5581	665.4538	1.8204	630.7923	6.9343	661.4538	2.9028
02.00	700.3192	703.7687	0.4926	715.7687	4.6806	635.7687	9.2173	725.3192	3.5698	720.4538	2.2302
03.00	734.5654	738.1835	0.4926	725.1835	2.0696	745.1835	1.4455	759.5654	3.4034	726.1835	2.4302
04.00	760.9115	757.1637	0.4924	743.1637	1.5923	777.1637	2.1359	740.9115	2.6284	776.1637	2.9528
05.00	767.4346	771.2146	0.4925	755.2146	4.4182	795.2146	3.6199	797.4346	3.9091	793.2146	3.3252
06.00	754.7077	758.4250	0.4915	730.4250	4.0749	745.4250	0.8851	750.7077	2.3554	752.7077	2.8520
07.00	744.3962	748.0627	0.4923	714.0627	4.3205	755.0627	1.2300	740.3962	0.5300	720.0627	3.3250
08.00	731.4692	727.8664	0.4925	699.8664	1.7461	718.8664	1.4329	735.4692	0.5373	716.8664	1.9325
09.00	717.9577	714.4214	0.4926	705.4214	0.7822	700.4214	1.7229	720.9577	0.5468	701.4214	1.8530
10.00	706.0231	709.5006	0.4926	700.5006	1.4930	730.5006	2.4425	760.0231	6.4179	609.2220	2.6538
11.00	699.6500	703.0961	0.4925	710.0961	3.9058	699.0961	0.0792	685.6500	2.0010	687.6500	1.9825
12.00	703.1462	706.6095	0.4925	730.6095	2.8131	725.6095	3.1947	707.6213	0.9955	690.6500	2.2350
13.00	726.0346	729.6107	0.4926	705.6107	2.2272	750.6107	3.3850	710.1462	1.9283	715.1462	2.2102
14.00	753.6077	757.3196	0.4925	755.3196	1.1835	700.3196	7.0711	740.0346	1.5923	735.0346	2.2152
15.00	768.8000	772.5867	0.4925	785.5867	3.6695	785.5867	2.1835	765.6077	1.2435	787.5867	2.3250
16.00	768.8538	772.6408	0.4925	740.6408	4.5999	778.6408	1.9500	720.8000	6.6503	780.6408	2.2523
17.00	754.7423	751.0248	0.4925	720.0248	1.1768	740.0248	0.1325	763.8538	2.1325	745.0248	1.9525
18.00	730.7462	734.3454	0.4926	739.3454	1.5246	690.9239	1.3136	755.7423	3.7790	742.3454	2.2980
19.00	703.3885	699.9239	0.4926	715.9239	2.1544	670.9967	1.7721	743.7462	8.8145	715.3885	3.3528
20.00	682.3577	678.9967	0.4925	760.9967	1.2358	655.1795	1.6650	765.3885	0.1153	695.3577	1.9982
21.00	661.9192	665.1795	0.4926	676.1795	2.1540	630.0057	1.0182	682.3577	2.1151	680.1795	3.3850
22.00	672.6923	676.0057	0.4926	681.0057	1.2822	630.1417	6.3456	675.9192	0.4797	682.6923	2.5650
23.00	676.6923	680.1750	0.4926	686.5057	1.4502	633.1530	5.8778	680.6923	1.1893	696.6923	3.1252
Avg.			0.4920		2.4741		2.9186		4.3371		2.7582

Table 4
Evaluation of actual and predicted load (P.load) in terms of MAPE of the proposed FCRBM based model and benchmark models like LSTM, MI-ANN, AFC-ANN, and Bi-level on Dayton power grid hourly load data of USA.

Hours	Target (kW)	Proposed and benchmark electric load forecasting models									
		FCRBM		AFC-ANN		Bi-level		MI-ANN		LSTM	
		P.load (kW)	MAPE (%)	P.load (kW)	MAPE (%)	P.load (kW)	MAPE (%)	P.load (kW)	MAPE (%)	P.load (kW)	MAPE (%)
00.00	175.3224	178.5829	0.7112	168.5829	3.9927	171.5829	2.2816	169.8923	3.4223	188.5829	2.2345
01.00	171.7990	174.8321	0.7463	165.4538	3.4925	167.4538	2.3284	165.7923	3.4925	185.8321	2.1342
02.00	167.3052	174. 8754	0.3909	166.7687	0.5977	164.7687	1.7932	162.3192	2.9886	145.2520	3.3250
03.00	169.5134	163.5234	0.7699	162.1835	4.1298	178.1835	2.9498	162.5654	2.1298	155.3825	3.0325
04.00	173.9342	166.3543	0.2999	183.1637	5.7497	179.1637	2.8748	180.9115	3.0248	155.3250	2.9980
05.00	182.3425	177.3824	0.6453	178.2146	2.1938	207.2146	4.6453	198.4346	2.1938	160.4346	3.1028
06.00	200.6273	185.9854	0.4955	211.4250	2.4834	218.4250	3.4894	208.7077	3.9879	204.6273	1.5328
07.00	212.3932	197.9234	0.4125	218.0627	2.8249	209.0627	2.8249	221.3962	4.2374	199.9234	1.3125
08.00	213.4532	209.4321	0.9372	213.8664	0.0999	217.8664	1.8743	215.4692	0.9372	229.4321	3.5325
09.00	212.4507	215.9234	0.4054	217.4214	1.8739	216.4214	1.8739	222.9577	4.2163	200.4507	2.3829
10.00	209.2531	216.8374	0.4134	220.5006	3.7691	215.5006	1.8845	220.0231	3.7691	218.2031	3.0029
11.00	205.1459	209.9321	0.4313	221.0961	5.7250	200.0961	2.8625	217.6500	3.8167	200.1459	1.6320
12.00	201.1462	206.5612	0.4624	197.6095	2.8997	206.6095	2.4373	198.6213	3.4122	188.1462	3.6740
13.00	197.3546	208.6578	0.4899	210.6107	4.4697	203.6107	2.4832	209.1462	3.9731	180.3260	3.9982
14.00	193.6077	204.3196	0.5182	203.3196	3.0363	192.3196	3.0363	200.0346	1.5182	200.3852	2.2352
15.00	192.8345	194.9134	0.5477	187.5867	3.0954	194.5867	0.5159	189.6077	2.0636	202.6332	2.1235
16.00	191.8538	190.7534	0.5556	202.6408	4.1853	195.6408	1.0371	200.8000	4.1482	205.3219	3.3780
17.00	192.7023	195.8034	0.5649	203.0248	6.2597	192.0248	0.0019	201.8538	5.2164	202.9876	3.6872
18.00	198.7462	194.3480	0.0173	203.3454	5.7070	192.9239	3.0260	201.7423	4.6694	182.3854	2.8876
19.00	205.2805	208.4567	0.4612	185.9239	6.5564	188.9967	8.2803	187.7462	5.5477	195.3290	2.2327
20.00	200.3075	203.6512	0.4933	181.9967	5.6898	190.1795	4.9775	189.3885	7.7932	192.3075	1.6288
21.00	200.9032	187.9234	0.5735	210.1795	4.9775	210.0057	6.2301	213.3577	6.4708	191.8765	1.9923
22.00	190.6533	187.5915	0.6644	203.2057	6.8187	210.1417	10.4902	200.9192	5.2451	201.7135	2.3589
23.00	180.2476	177.7435	0.6644	175.5057	2.7740	178.1530	1.1096	170.6923	3.4383	188.6534	1.9789
Avg.			0.49980		2.7633		3.0220		4.4988		2.5432

Table 5
Evaluation of actual and predicted load (P.load) in terms of MAPE of the proposed FCRBM based model and benchmark models like LSTM, MI-ANN, AFC-ANN, and Bi-level on EKPC power grid hourly load data of USA.

Hours	Target (kW)	Proposed and existing forecast models									
		FCRBM		AFC-ANN		Bi-level		MI-ANN		LSTM	
		P.load (kW)	MAPE (%)	P.load (kW)	MAPE (%)	P.load (kW)	MAPE (%)	P.load (kW)	MAPE (%)	P.load (kW)	MAPE (%)
00.00	132.8923	130.5829	0.5115	136.5019	4.5345	138.9234	3.1741	139.8923	5.4414	145.8923	3.3520
01.00	131.7923	129.4538	0.5175	123.3458	2.2763	128.4832	6.0700	127.7923	3.0350	146.8923	3.2523
02.00	131.3192	129.7687	0.5232	123.6685	1.5232	129.8723	6.0929	127.3192	3.0464	120.3192	2.9268
03.00	132.5654	130.1835	0.5094	145.1345	3.7736	127.1358	9.8113	139.5654	5.2830	122.4328	2.0280
04.00	138.9115	136.1637	0.4396	147.1370	1.4396	140.1378	6.4784	142.9115	2.8793	125.9115	2.2025
05.00	147.4346	145.2146	0.3575	153.1106	1.3575	149.4634	4.0724	151.4346	2.7149	135.2823	2.1345
06.00	161.2115	157.4474	0.4751	168.4130	3.7127	167.4013	4.3315	168.2115	4.3315	144.2115	3.3542
07.00	167.7077	163.4250	0.3895	170.2327	1.7922	170.2723	1.7922	173.7077	3.5843	157.7077	3.0119
08.00	163.3962	161.0627	0.2239	175.6234	3.0598	168.8602	7.3435	170.3962	4.2837	153.3962	2.9812
09.00	150.4692	148.8664	0.3293	150.4214	1.9939	153.1423	0.7231	150.4000	0.0692	162.6750	3.1234
10.00	145.9577	143.4214	0.3769	157.0023	6.1961	154.5602	8.2614	153.9577	5.5076	136.7750	2.2105
11.00	141.0231	139.5006	0.4124	140.3261	0.7062	142.6101	0.7062	140.0231	0.7062	130.3210	3.2650
12.00	138.6500	136.0961	0.4477	146.9501	4.3432	144.1295	5.7910	146.6500	5.7910	128.3210	2.2860
13.00	136.1462	130.6095	0.4003	130.6107	2.9335	132.6071	4.4003	130.1462	2.4003	125.6540	3.1032
14.00	130.0346	128.6107	0.5313	130.3602	3.0626	134.3106	0.3421	136.0346	4.5939	115.2430	3.1028
15.00	130.6077	128.3196	0.5286	127.5671	1.5286	128.5027	2.2930	123.6077	5.3503	121.7710	2.1025
16.00	130.8000	130.5867	0.2133	124.6418	9.2307	130.4082	4.5853	124.8903	4.5853	119.8351	2.2540
17.00	133.8538	131.6408	0.4959	143.3485	2.9917	137.1348	7.4793	139.8038	4.4876	121.7652	3.2501
18.00	137.7423	135.0248	0.4519	146.3541	1.4519	139.3924	6.5338	141.0233	2.9039	122.3542	3.2340
19.00	142.7462	140.3454	0.4057	137.9339	2.1085	139.2267	3.5142	137.7053	3.5142	120.6856	3.2523
20.00	152.3885	150.9239	0.3131	149.9557	1.9697	149.1534	1.9697	147.3885	3.2828	140.3456	3.1236
21.00	151.3577	149.9967	0.3166	156.9520	1.3166	153.2357	3.2916	157.3577	3.9499	135.7654	3.3675
22.00	136.9192	134.1795	0.4636	135.1257	2.9271	140.1017	0.7318	139.9192	2.1953	121.9512	2.9925
23.00	132.6923	130.0057	0.5123	133.5231	4.5369	138.1530	0.9830	139.6053	5.5199	115.8976	3.0025
Avg.			0.4525		2.4202		2.7530		4.3280		3.1234



(a) FE grid (b) Dayton grid (c) EKPC grid

Fig. 8. Comparative evaluation of the proposed FCRBM based model and benchmark models like LSTM, MI-ANN, AFC-ANN, and Bi-level in terms of convergence rate on FE, Dayton, and EKPC power grids hourly load data of USA.

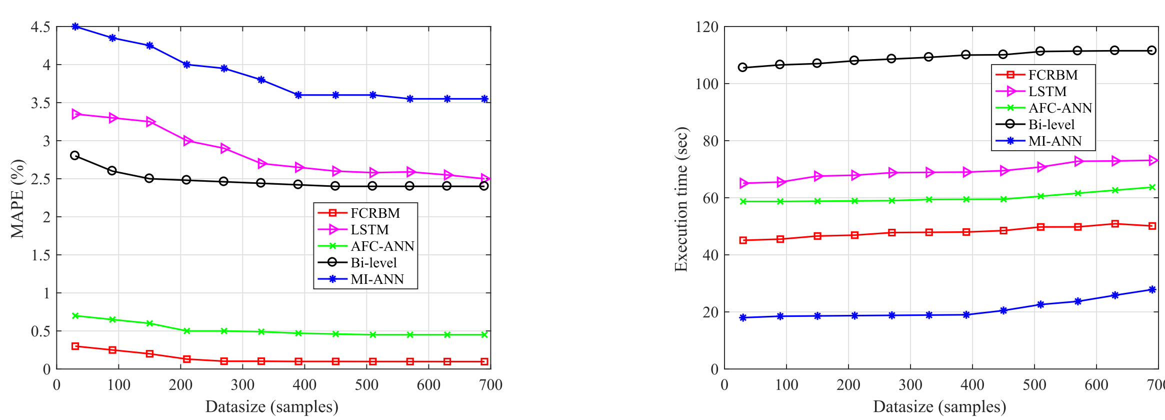
from 16.52s to 102s as the optimization module is integrated with the forecasting module. The proposed model has reduced execution time due to the following reasons: (i) highly abstractive features are given as an input to training and forecasting module which reduces the network training time, (ii) it replaces the sigmoidal activation function by ReLU, and (iii) it uses GWDO instead of MEDEA algorithm due to relatively faster convergence rate. The proposed short-term load forecasting model decreased the execution time from 102s to 43s due to aforesaid modifications in the existing models like LSTM, AFC-ANN and Bi-level. In contrast, the MI-ANN has excellent performance in terms of convergence rate as compared to the other models like FCRBM, LSTM, AFC-ANN, and Bi-level because no optimizer is integrated with MI-ANN model. This behavior is clearly depicted in Fig. 8.

4.5. Scalability analysis

The scalability analysis enables us to identify whether the proposed model is scalable or its suitable for the said scenarios. The Eq. (1), is manipulated for input samples, features, weights, and bias, and correspondingly forecasted results are analyzed to the scalability of the model. For example, the weights of the input samples are increased but the number of input samples remains constant and the proposed forecast strategy is not affected. On the other hand, if the number of input samples (size of the data), influencing factors, and forecast-horizon are increased, that affect the convergence rate and accuracy. The impact of these factors on the accuracy and convergence rate is illustrated in Fig. 9 in terms of error (Fig. 9)a and execution time (Fig. 9b) for the proposed and benchmark models. The accuracy analysis of the general trend in terms of error performance is shown in Fig. 9a. The forecast accuracy is improving as the number of data samples is increasing from 0 to 720 and tends toward stability as the number of samples is further increased. The results are obvious because the value of x in the Eq. (1) is tightly linked with the training of FCRBM. During the training process, a large value of x means fine-tuning and results in improved forecast accuracy. Similarly, Fig. 9b illustrates number of samples vs execution time. It is verified from Fig. 9b that as the number of samples increases the execution time increases and vice versa. The proposed model has relatively high scalability as compared to the benchmark models due to the use of MMI technique for features selection , FCRBM for forecasting, and GWDO algorithm for optimization.

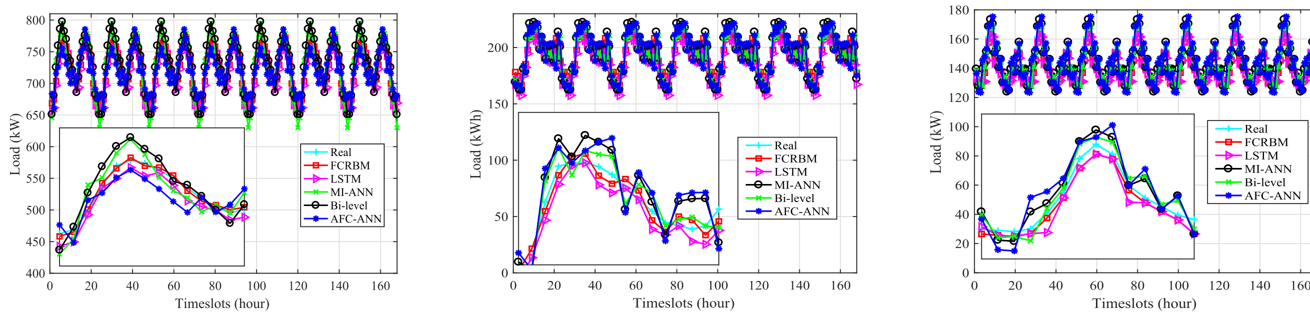
4.6. Evaluation of actual and forecasted load forthe week ahead time horizon with the hour resolution

The weekly forecasted load profile based on the proposed model and benchmark models for three USA power grids like EKPC, FE, and Dayton is illustrated in Fig. 10. The FE power grid predicted load profile for a week ahead with hour resolution is illustrated in Fig. 10a. In this figure, some portion is zoomed to clearly show the behavior of the proposed and benchmark models. It is obvious from the figure that the proposed FCRBM based prediction closely follow the target load as compared to LSTM, MI-ANN, Bi-Level, and AFC-ANN models. The hourly and daily MAPE values for the proposed and benchmark models are listed in Table 3. It is noticed that FCRBM has the best accuracy as compared to existing models. The week ahead forecasted load profile with hour resolution based on our proposed model FCRBM, and benchmark models like LSTM, MI-ANN, Bi-level, and AFC-ANN for Dayton power grid is depicted in Fig. 10b. It is seen that FCRBM based predicted load closely follow the target load as compared to the benchmark models. The proposed model has better accuracy due to the use of deep layer layout and GWDO based optimization module. Similarly, in the case of EKPC power grid's, the proposed deep learning model FCRBM weekly predicted load



(a) Error performance (b) Convergence rate

Fig. 9. Scalability evaluation of the proposed model and benchmark models like MI-ANN, LSTM, AFC-ANN, and Bi-level in terms: (a) error performance and (b) convergence rate.



(a) FE grid (b) Dayton grid (c) EKPC grid

Fig. 10. Performance evaluation of actual and predicted load for the week ahead with hour resolution of the proposed model and benchmark models like MI-ANN, LSTM, AFC- ANN, and Bi-level in terms of accuracy on FE, Dayton, and EKPC power grids hourly load data of USA.

profile with hour resolution is more accurate as compared to the benchmark models as depicted in Fig. 10c. Overall, it is observed that the proposed short-term load forecasting model has a more accurate prediction as compared to the benchmark models.

4.7. Cumulative distribution function of error

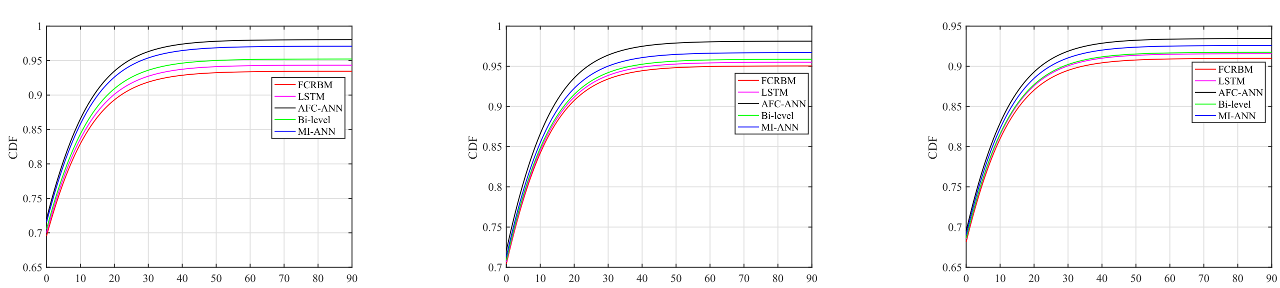
The cumulative distribution function (CDF) of error for three power grids: FE, Dayton, and EKPC is illustrated in Fig. 11 for the proposed and benchmark models like AFC-ANN [10], MI-ANN [45], Bi-level [46], and LSTM. The proposed short-term load forecasting model is better in all three scenarios in terms of error CDF, i.e., FE, Dayton, and EKPC as compared to the benchmark models. It is not surprising because the FCRBM has more computational power as compared to the AFC-ANN, MI-ANN, Bi-level, LSTM models. Roughly, when the error is under 4%, the Bi-level has better performance than MI-ANN, and worst performance than AFC-ANN; although FCRBM predicts more reliably even if there is more uncertainty because their deep layers structure can capture the highly abstracted features. Thus, the proposed model would be a better choice for the utility in the decision making of the SG because its performance is improving with the increase in datasize.

5. Conclusion

Accurate electric load forecasting is indispensable due to its application in decision making and operation of the power grid. With accurate electric load forecasting, operators are capable to develop an optimal market plan to enhance the economic benefits of energy management. Therefore, it is a significant goal for scholars and industry to develop a forecasting model, which provides robust, stable, and accurate load forecasting. However, the performance of individual prediction models is not satisfactory due to inherent limitations. In contrast, hybrid models fully utilize the advantages of individual models and have improved performance. In this paper, a short-term load forecasting model is proposed. The proposed model is an integrated framework of three modules: (i) MMI based data pre-processing and feature selection module, (ii) deep learning technique FCRBM basedtraining and forecasting module, and (iii) GWDO algorithm based optimization module. The proposed model is tested on historical hourly electric load data of three USA power grids: FE, EKPC, and Dayton. The proposed model is validated by comparing it to the four benchmark models like MI-ANN, Bi-Level, AFC-ANN, and LSTM in terms of accuracy and convergence rate. Based on results, performance evaluation, and discussion the following conclusions are made can be drawn. First, the proposed MMI technique improves the forecast accuracy by selecting desired features from data and then feed these desired features into into the training module based on FCRBM to reduce the training time. Secondly, the adapted deep learning model FCRBM is trained and enabled via learning to forecast day and week-ahead electric load with hour resolution. Thirdly, the proposed GWDO algorithm is used in the optimization module to fine-tune the adjustable parameters of the model. Finally, the proposed model is validated by comparing it with four benchmark models like AFC-ANN, Bi-level, MI-ANN, and LSTM models. In short, the proposed electric load forecasting model outperforms MI-ANN by 31.2%, Bi-level by 17.3%, and AFC-ANN by 4.7% in terms of forecast accuracy. Furthermore, the average execution time of the proposed model is 52s, the AFC-ANN model is 58s, the Bi-level model is 102s, MI-ANN model is 16.5s, and LSTM model is 63s. Thus, it is concluded that the proposed model outperforms benchmark models in terms of both forecast accuracy and convergence rate.

CRediT authorship contribution statement

Ghulam Hafeez: Conceptualization, Data curation, Methodology, Writing - original draft, Resources, Software, Writing - review & editing. **Khurram Saleem Alimgeer:** Supervision, Project administration, Visualization, Investigation, Writing - review & editing. **Imran Khan:** Formal analysis, Funding acquisition, Writing - review & editing.



(a) FE grid (b) Dayton grid (c) EKPC grid

Fig. 11. CDF of daily MAPE evaluation of the proposed model and benchmark models like Bi-level, AFC-ANN, MI-ANN, and LSTM on FE, Dayton, and EKPC power grids hourly load data of the USA.

Declaration of Competing Interest

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

References

[1] Javaid Nadeem, Hafeez Ghulam, Iqbal Sohail, Alrajeh Nabil, Alabed Mohamad Souheil, Guizani Mohsen. Energy efficient integration of renewable energy sources in the smart grid for demand side management. IEEE Access 2018;6:77077–96. https://doi.org/10.1109/ACCESS.2018.2866461.
[2] Xiao Liye, Shao Wei, Wang Chen, Zhang Kequan, Haiyan Lu. Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting. Appl Energy 2016;180:213–33. https://doi.org/10.1016/j.apenergy.2016.07.113.
[3] Alahakoon Damminda, Xinghuo Yu. Smart electricity meter data intelligence for future energy systems: A survey. IEEE Trans Industr Inf 2016;12(1):425–36. https://doi.org/10.1109/TII.2015.2414355.
[4] Hernandez Luis, Baladron Carlos, Aguiar Javier M, Carro Belén, Sanchez-Esguevillas Antonio J, Lloret Jamie, Massana Joaquim. A survey on electric power demand forecasting: future trends in smart grids, microgrids and smart buildings. IEEE Commun Surveys Tut 2014;16(3):1460–95. https://doi.org/10.1109/SURV.2014.032014.00094.
[5] Rahman Aowabin, Srikumar Vivek, Smith Amanda D. Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. Appl Energy 2018;212:372–85. https://doi.org/10.1016/j.apenergy.2017.12.051.
[6] Boroojeni Kianoosh G,Hadi Amini M., Bahrami Shahab, Iyengar SS, Sarwat Arif I, Karabasoglu Orkun. A novel multi-time-scale modeling for electric power demand forecasting: from short-term to medium-term horizon. Electric Power Syst Res 2017; 142:58–73. doi:10.1016/j.epsr.2016.08.031.
[7] Li Yanying, Che Jinxing, Yang Youlong. Subsampled support vector regression ensemble for short term electric load forecasting. Energy 2018;164:160–70. https://doi.org/10.1016/j.energy.2018.08.169.
[8] Dedinec Aleksandra, Filiposka Sonja, Dedinec Aleksandar, Kocarev Ljupco. Deep belief network based electricity load forecasting: an analysis of Macedonian case. Energy 2016;115:1688–700. https://doi.org/10.1016/j.energy.2016.07.090.
[9] Hong Wei-Chiang. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. Energy 2011;36(9):5568–78. https://doi.org/10.1016/j.energy.2011.07.015.
[10] Ahmad Ashfaq, Javaid Nadeem, Guizani Mohsen, Alrajeh Nabil, Ali Khan Zahoor. An accurate and fast converging short-term load forecasting model for industrial applications in a smart grid. IEEE Trans Industr Inf 2016;13(5):2587–96. https://doi.org/10.1109/TII.2016.2638322.
[11] Liu Dunnan, Zeng Long, Li Canbing, Ma Kunlong, Chen Yujiao, Cao Yijia. A distributed short-term load forecasting method based on local weather information. IEEE Syst. J. 2018;12(1):208–15. https://doi.org/10.1109/JSYST.2016.2594208.
[12] Shi Heng, Minghao Xu, Li Ran. Deep learning for household load forecasting—a novel pooling deep RNN. IEEE Trans Smart Grid 2018;9(5):5271–80. https://doi.org/10.1109/TSG.2017.2686012.
[13] Kong Weicong, Dong Zhao Yang, Hill David J, Luo Fengji, Xu Yan. Short-term residential load forecasting based on resident behaviour learning. IEEE Trans Power Syst 2018;33(1):1087–8. https://doi.org/10.1109/TPWRS.2017.2688178.
[14] Huang Xuefei, Hong Seung Ho, Li Yuting. Hour-ahead price based energy management scheme for industrial facilities. IEEE Trans Industr Inf 2017;13(6):2886–98. https://doi.org/10.1109/TII.2017.2711648.
[15] van der Meer DW, Munkhammar J, Widén J. Probabilistic forecasting of solar power, electricity consumption and net load: investigating the effect of seasons, aggregation and penetration on prediction intervals. Sol Energy 2018; 171:397–413. doi:10.1016/j.solener.2018.06.103.
[16] Wang Pu, Liu Bidong, Hong Tao. Electric load forecasting with recency effect: a big data approach. Int J Forecast 2016;32(3):585–97. https://doi.org/10.1016/j.ijforecast.2015.09.006.
[17] Carvallo, Pablo Juan, Larsen Peter H, Sanstad Alan H, Goldman Charles A. Long term load forecasting accuracy in electric utility integrated resource planning. Energy Policy 2018; 119:410–422. doi:10.1016/j.enpol.2018.04.060.
[18] Xu Xiaomin, Niu Dongxiao, Wang Qiong, Wang Peng, Dash Wu Desheng. Intelligent forecasting model for regional power grid with distributed generation. IEEE Syst J 2017;11(3):1836–45. https://doi.org/10.1109/JSYST.2015.2438315.
[19] Li Liangzhi, Ota Kaoru, Dong Mianxiong. When weather matters: IoT-based electrical load forecasting for smart grid. IEEE Commun Mag 2017;55(10):46–51. https://doi.org/10.1109/MCOM.2017.1700168.
[20] Wang Yu, Shen Yinxing, Mao Shiwen, Cao Guanqun, Nelms Robert M. Adaptive learning hybrid model for solar intensity forecasting. IEEE Trans Industr Inf 2018;14(4):1635–45. https://doi.org/10.1109/TII.2017.2789289.
[21] Tang Ningkai, Mao Shiwen, Wang Yu, Nelms RM. Solar power generation forecasting with a LASSO-based approach. IEEE Internet Things J 2018;5(2):1090–9. https://doi.org/10.1109/JIOT.2018.2812155.
[22] Zhang Jinliang, Wei Yi-Ming, Li Dezhi, Tan Zhongfu, Zhou Jianhua. Short term electricity load forecasting using a hybrid model. Energy 2018;158:774–81. https://doi.org/10.1016/j.energy.2018.06.012.
[23] Tong Chao, Li Jun, Lang Chao, Kong Fanxin, Niu Jianwei, Rodrigues Joel JPC. An efficient deep model for day-ahead electricity load forecasting with stacked denoising auto-encoders. J Parallel Distrib Comput 2018; 117:267–73. doi:10.1016/j.jpdc.2017.06.007.
[24] Yuan Jihui, Farnham Craig, Azuma Chikako, Emura Kazuo. Predictive artificial neural network models to forecast the seasonal hourly electricity consumption for a University Campus. Sustain Cities Soc 2018;42:82–92. https://doi.org/10.1016/j.scs.2018.06.019.
[25] Brodowski Stanisław, Bielecki Andrzej, Filocha Maciej. A hybrid system for forecasting 24-h power load profile for Polish electric grid. Appl Soft Comput 2017; 58:527–39. doi:10.1016/j.asoc.2017.04.053.
[26] Qiu Xueheng, Ren Ye, Suganthan Ponnuthurai Nagaratnam, Amaratunga Gehan AJ. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. Appl Soft Comput 2017; 54:246–55. doi:10.1016/j.asoc.2017.01.015.
[27] Yang Ailing, Li Weide, Yang Xuan. Short-term electricity load forecasting based on feature selection and Least Squares Support Vector Machines. Knowl-Based Syst 2019;163:159–73. https://doi.org/10.1016/j.knosys.2018.08.027.
[28] Qiu Xueheng, Suganthan Ponnuthurai Nagaratnam, Amaratunga Gehan AJ. Ensemble incremental learning Random Vector Functional Link network for short-term electric load forecasting. Knowl-Based Syst 2018; 145:182–96. doi:10.1016/j.knosys.2018.01.015.
[29] Chen Yanhua, Kloft Marius, Yang Yi, Li Caihong, Li Lian. Mixed kernel based extreme learning machine for electric load forecasting. Neurocomputing 2018; 312:90–106. doi:10.1016/j.neucom.2018.05.068.
[30] Zeng Nianyin, Zhang Hong, Liu Weibo, Liang Jinling, Alsaadi Fuad E. A switching delayed PSO optimized extreme learning machine for short-term load forecasting. Neurocomputing 2017;240:175–82. https://doi.org/10.1016/j.neucom.2017.01.090.
[31] Zhang Xiaobo, Wang Jianzhou, Zhang Kequan. Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm. Electric Power Syst Res 2017;146:270–85. https://doi.org/10.1016/j.epsr.2017.01.035.
[32] Chen Yibo, Tan Hongwei, Berardi Umberto. Day-ahead prediction of hourly electric demand in non-stationary operated commercial buildings: a clustering-based hybrid approach. Energy Build 2017;148:228–37. https://doi.org/10.1016/j.enbuild.2017.05.003.
[33] Guo Zhifeng, Zhou Kaile, Zhang Xiaoling, Yang Shanlin. A deep learning model for short-term power load and probability density forecasting. Energy 2018;160:1186–200. https://doi.org/10.1016/j.energy.2018.07.090.
[34] Ghadimi Noradin, Akbarimajd Adel, Shayeghi Hossein, Abedinia Oveis. Two stage forecast engine with feature selection technique and improved meta-heuristic algorithm for electricity load forecasting. Energy 2018;161:130–42. https://doi.org/10.1016/j.energy.2018.07.088.
[35] Li Yanying, Che Jinxing, Yang Youlong. Subsampled support vector regression ensemble for short term electric load forecasting. Energy 2018;164:160–70. https://doi.org/10.1016/j.energy.2018.08.169.
[36] Jawad Muhammad, Ali Sahibzada M, Khan Bilal, Mehmood Chaudry A, Farid Umar, Ullah Zahid, et al. Genetic algorithm-based non-linear auto-regressive with exogenous inputs neural network short-term and medium-term uncertainty modelling and prediction for electrical load and wind speed. J Eng 2018;2018(8):721–9. https://doi.org/10.1049/joe.2017.0873.
[37] Manjili Yashar Sahraei, Vega Rolando, Jamshidi Mo M. Data-analytic-based adaptive solar energy forecasting framework. IEEE Syst J 2018;12(1):285–96. https://doi.org/10.1109/JSYST.2017.2769483.
[38] Kong Xiangyu, Li Chuang, Wang Chengshan, Zhang Yusen, Zhang Jian. Short-term electrical load forecasting based on error correction using dynamic mode decomposition. Appl Energy 2020;261:114368https://doi.org/10.1016/j.apenergy.2019.114368.
[39] Huyghues-Beaufond Nathalie, Tindemans Simon, Falugi Paola, Sun Mingyang, Strbac Goran. Robust and automatic data cleansing method for short-term load forecasting of distribution feeders. Appl Energy 2020;261:114405https://doi.org/10.1016/j.apenergy.2019.114405.
[40] Cai Mengmeng, Pipattanasomporn Manisa, Rahman Saifur. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. Appl Energy 2019;236:1078–88. https://doi.org/10.1016/j.apenergy.2018.12.042.
[41] He F, Zhou J, Feng ZK, Liu G, Yang Y. A hybrid short-term load forecasting model based on variational mode decomposition and long short-term memory networks considering relevant factors with Bayesian optimization algorithm. Appl Energy 2019;237:103–16. https://doi.org/10.1016/j.apenergy.2019.01.055.
[42] Wu Zhuochun, Zhao Xiaochen, Ma Yuqing, Zhao Xinyan. A hybrid model based on modified multi-objective cuckoo search algorithm for short-term load forecasting. Appl Energy 2019;237:896–909. https://doi.org/10.1016/j.apenergy.2019.01.046.
[43] Semero Yordanos Kassa, Zhang Jianhua, Zheng Dehua. PV power forecasting using an integrated GA-PSO-ANFIS approach and Gaussian process regression based feature selection strategy. CSEE J Power Energy Syst 2018;4(2):210–8. https://doi.org/10.17775/CSEEJPES.2016.01920.
[44] Mocanu Elena, Mocanu Decebal Constantin, Nguyen Phuong H, Liotta Antonio, Webber Michael E, Gibescu Madeleine, Slootweg Johannes G. On-line building energy optimization using deep reinforcement learning. IEEE Trans Smart Grid 2018;10(4):3698–708. https://doi.org/10.1109/TSG.2018.2834219.
[45] Amjady Nima, Keynia Farshid, Zareipour Hamidreza. Short-term load forecast of microgrids by a new bilevel prediction strategy. IEEE Trans Smart Grid 2010;1(3):286–94. https://doi.org/10.1109/TSG.2010.2078842.
[46] Amjady Nima, Keynia Farshid. Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm. IEEE Trans Power Syst 2008;24(1):306–18. https://doi.org/10.1109/TPWRS.2008.

2006997.
[47] Hafeez Ghulam, Islam Noor, Ali Ammar, Ahmad Salman, Usman Alimgeer Muhammad, Saleem Khurram. A modular framework for optimal load scheduling under price-based demand response scheme in smart grid. Processes 2019;7(8):499. https://doi.org/10.3390/pr7080499.
[48] Hafeez Ghulam, Javaid Nadeem, Riaz Muhammad, Ali Ammar, Umar Khalid, Iqbal Zafar. Day Ahead Electric Load Forecasting by an Intelligent Hybrid Model Based on Deep Learning for Smart Grid. In: Conference on complex, intelligent, and software intensive systems. Cham: Springer, 2019. p. 36–49. doi:10.1007/978-3-030-22354-0_4.
[49] Abedinia Oveis, Amjady Nima, Zareipour Hamidreza. A new feature selection technique for load and price forecast of electrical power systems. IEEE Trans Power Syst 2016;32(1):62–74. https://doi.org/10.1109/TPWRS.2016.2556620.
[50] Khwaja AS, Naeem M, Anpalagan A, Venetsanopoulos A, Venkatesh B. Improved short-term load forecasting using bagged neural networks. Electric Power Syst Res 2015;125:109–15. https://doi.org/10.1016/j.epsr.2015.03.027.
[51] Hafeez Ghulam, Javaid Nadeem, Ullah Safeer, Iqbal Zafar, Khan Mahnoor, Ur Rehman Aziz. Short term load forecasting based on deep learning for smart grid applications. In: International conference on innovative mobile and internet services in ubiquitous computing. Cham: Springer; 2018. p. 276–88. doi:10.1007/978-3-319-93554-6_25.
[52] Taylor GW, Hinton GE, Roweis ST. Two distributed-state models for generating high-dimensional time series. J Mach Learn Res 2011;12:1025–68.
[53] Mocanu Decebal Constantin, Bou Ammar Haitham, Lowet Dietwig, Driessens Kurt, Liotta Antonio, Weiss Gerhard, Tuyls Karl. Factored four way conditional restricted boltzmann machines for activity recognition. Pattern Recogn Lett 2015; 66:100–8. doi:10.1016/j.patrec.2015.01.013.
[54] Mocanu Elena, Nguyen Phuong H, Gibescu Madeleine, Larsen Emil Mahler, Pinson Pierre. Demand forecasting at low aggregation levels using factored conditional restricted boltzmann machine. In: 2016 Power Systems Computation Conference (PSCC). IEEE; 2016. p. 1–7. doi:10.1109/PSCC.2016.7540994.
[55] Introduction to various types of gradient descent. https://machinelearningmastery.com/gentle-introduction-mini-batch-gradient-descent-configure-batch-size/ [accessed 13 Aug 2019].
[56] Mocanu Elena, Nguyen Phuong H, Gibescu Madeleine, Kling Wil L. Deep learning for estimating building energy consumption. Sustain Energy Grids Networks 2016;6:91–9. https://doi.org/10.1016/j.segan.2016.02.005.
[57] Bao Zongfan, Zhou Yongquan, Li Liangliang, Ma Mingzhi. A hybrid global optimization algorithm based on wind driven optimization and differential evolution. Math Probl Eng 2015;2015:608–20. https://doi.org/10.1155/2015/389630.
[58] Ghulam Hafeez, Javaid Nadeem, Iqbal Sohail, Khan Farman. Optimal residential load scheduling under utility and rooftop photovoltaic units. Energies 2018;11(3):611. https://doi.org/10.3390/en11030611.
[59] PJM electricity market. Available online: https://www.pjm.com/ [accessed 17 Feb 2019].
[60] Srivastava Nitish, Hinton Geoffrey, Krizhevsky Alex, Sutskever Ilya, Salakhutdinov Ruslan. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15(1):1929–58. https://doi.org/10.5555/2627435.2670313.

Ghulam Hafeez has completed his B.Sc. in Electrical Engineering from University of Engineering and Technology Peshawar, Pakistan, and MS in Electrical Engineering from COMSATS University Islamabad, Islamabad, Pakistan. He is pursuing towards Ph.D. from the same University. Ghulam Hafeez is lifetime charted engineer from Pakistan Engineering Council. Ghulam Hafeez is working as a lecturer in the Department of Electrical Engineering, University of Engineering and Technology, Mardan. He has authored or co-authored over 15 peer-reviewed research papers in reputed international journals and conferences. His research interests include optimization, planning, energy management, and machine learning applications in smart/micro grids, etc.

Khurram Saleem Alimgeer did his bachelor's degree in IT in 2002 and completed his MS in Telecommunications (Gold Medal) in 2006. He did his PhD in Electrical Engineering with specialization in Antenna design in 2014. Currently, he is Assistant Professor at COMSATS University Islamabad, and working as researcher in RF-Lab, COMSATS University Islamabad, Pakistan. He is serving as editor and reviewer of few reputed journals since 2008. He has published more than 80 research papers at reputed journals and conferences in the fields of Antenna Design, wave propagation, mathematical modeling, Wireless Communications, Image Processing, and Energy Management in the Smart/Micro Grid.

Imran Khan received the B.Sc. degree in Electrical Engineering from N.W.F.P. University of Engineering and Technology, Peshawar, Pakistan in 2003 and M.Sc. degree in telecommunication engineering from the Asian Institute of Technology, Thailand, in 2007. He did Ph.D. degree at the Telecommunications FOS, School of Engineering and Technology, Asian Institute of Technology, Thailand, in 2010. Currently he is working as professor in Electrical Engineering Department, University of Engineering Technology, Mardan. His research interests include performance analysis of Wireless Communication Systems, OFDM, OFDMA, MIMO, Cooperative Networks, Cognitive Radio Systems, and Energy Management in the Smart Grid.