Big Data Analytics for Load Forecasting in Smart Grids: A Survey

Sana Mujeeb[1], Nadeem Javaid[1,*], Sakeena Javaid[1], Asma Rafique[2], Manzoor Ilahi[1]
[1]Department of Computer Science, COMSATS University Islamabad, Islamabad 44000, Pakistan
{sana.mujeeb.22, nadeemjavaidqau, sakeenajavaid}@gmail.com, tamimy@comsats.edu.pk
[2]Department of Computer Engineering, School of Sciences, Karabuk University, Turkey
asmamcs@gmail.com
[*]Correspondence: nadeemjavaidqau@gmail.com; www.njavaid.com

***Abstract*—Recently big data analytics are gaining popularity in the energy management systems (EMS). The EMS are responsible for controlling, optimization and managing the energy market operations. Energy consumption forecasting plays a key role in EMS and helps in generation planning, management and energy conversation. A large amount of data is being collected by the smart meters on daily basis. Big data analytics can help in achieving insights for smart energy management. Several prediction methods are proposed for energy consumption forecasting. This study explores the state-of-the-art forecasting methods. The studied forecasting methods are classified into two major categories: (i) univariate (time series) forecasting models and (ii) multivariate forecasting models. The strengths and limitations of studied methods are discussed. Comparative anlysis of these methods is also done in this survey. Furthermore, the forecasting techniques are reviewed from the aspects of big data and conventional data. Based on this survey, the gaps in the existing research are identified and future directions are described.**

***Index Terms*—Big Data, Data Analytics, Load Forecasting, Artificial intelligent Forecasters, Deep Learning.**

I. Introduction

The modernization of power systems has brought a revolution in the electricity generation and distribution sectors in recent years. With the introduction of smart grid, the communication technology is integrated with conventional electricity meters, known as smart meters. These smart meters measure electricity consumption (and other measurements) at every small time intervals and communicate to energy suppliers, resulting in generation of very huge amount of data. Due to availability of the huge amount of data, many innovative programs are implemented like real-time pricing, off peak time usage lesser tariffs, etc. In near future, all the conventional energy meters will be replaced by smart meters. It is estimated that, more than 800 million smart meters will be deployed world wide till 2020. Power utilities receive a deluge of data after the deployment of smart meters. This data is termed as energy big data. Big data have a few major characteristics referred as 4 V's.

- *Volume:* The major characteristic that makes data *big* is its huge volume. Tera bytes and exabytes of smart meter measurements are recorded daily.
- *Velocity:* The frequency of recorded data is very high. Smart meter measurements are recorded in very small time intervals. It is a continuous streaming process.
- *Variety:* The data can be in different structures, e.g., sensors data, smart meters data and communication modules data are different. Both structured and unstructured data is captured. Unstructured data is standardized to make it meaningful and useful.
- *Veracity:* The trustworthiness and authenticity of data is referred as veracity. The recorded data may have noisy or false readings. The false readings can be due to the malfunctioning of sensors.

TABLE I: List of abbreviations

Abbreviation	Full Form
ABC	Artificial Bee Colony
AEMO	Australia Electricity Market Operators
ANN	Artificial Neural Networks
ARIMA	Auto Regressive Integrated Moving Average
CNN	Convolution Neural Networks
CART	Classification and Regression Tree
DNN	Deep Neural Networks
DSM	Demand Side Management
DT	Decision Tree
DE	Differential Evaluation
GA	Genetic Algorithm
ISO NECA	Independent System Operator New England Control Area
KNN	K Nearest Neighbor
LSSVM	Least Square Support Vector Machine
LSTM	Long Short Term Memory
MAPE	Mean Absolute Percentage Error
NYISO	New York Independent System Operator
PJM	Pennsylvania-New Jersey-Maryland (Interconnection)
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SAE	Stacked Auto Encoders
STLF	Short Term Load Forecast
SVM	Support Vector Machine

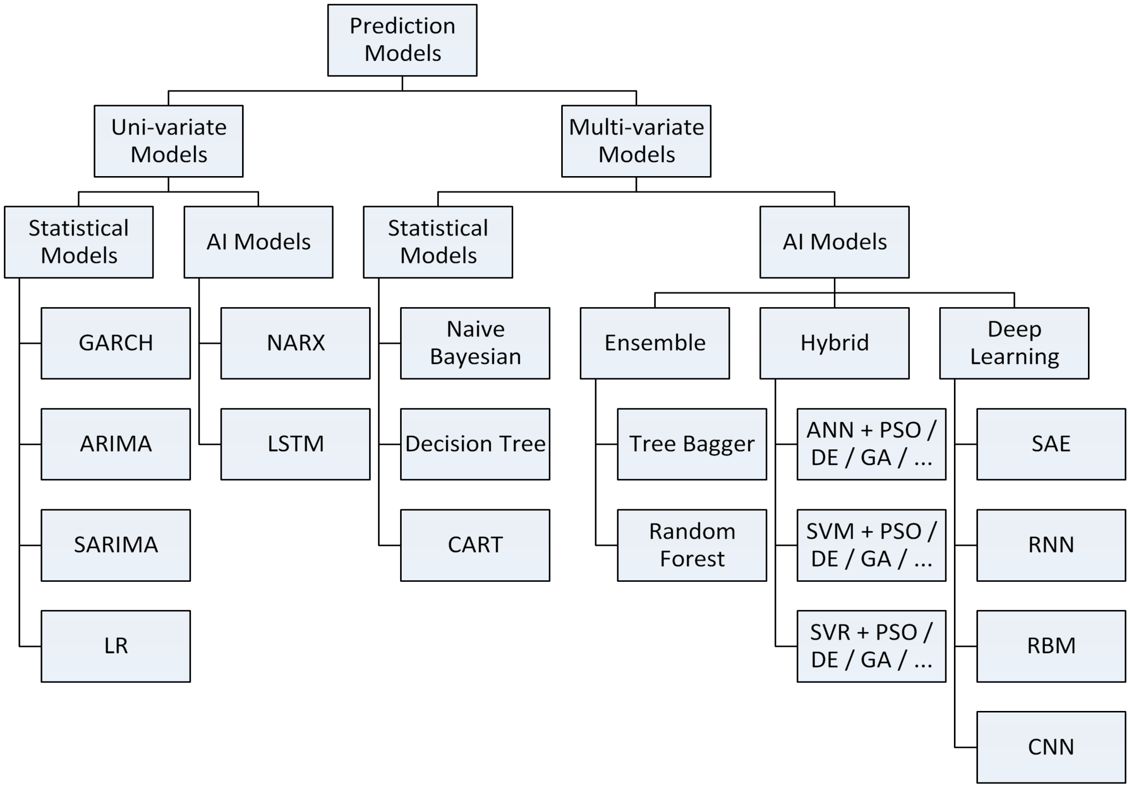

Fig. 1: Classification of prediction models.

TABLE II: List of symbols

Symbol	Description
b	SVM bias
c	cost penalty
η	insensitive loss function parameter
ρ	SVM marginal plane
σ	SVM kernel function parameter
w_i	ANN weights
W_{hx}	RNN weights

Besides, the 4V's of big data, the energy big data exhibits a few more characteristics: (i) data as an energy: big data analytics should cause energy saving, (ii) data as an exchange: energy big data should be exchanged and integrated with other sources big data to identify its value, (iii) data as an empathy: data analytics can help in improvement of service quality of energy utilities [1]. Approximately 220 million smart meter measurements are recorded daily, in a large sized smart grid.

In order to avoid the failure of electricity distribution networks, the suppliers rely on generation and demand balancing. For balancing demand generation and filling the demand response gap, the utilities have to estimate the energy demand patterns of different consumers. The demand pattern is not always even, therefore electricity load estimation is a very difficult task. Several prediction methods are proposed for energy load forecasting. Classic statistical methods to modern computationally intelligent prediction techniques are proposed for electricity load prediction. This work surveys the state-of-the-art load forecasting models from the literature of past four years. The focus of this survey is on the univariate and multivariate prediction models. The major contribution of this work is the comparative analysis of prediction methods with respect to their input, i.e., conventional traditional data and big data. Energy big data is also explained. The existing load forecasting surveys mostly focus on traditional data forecasting techniques [2]-[5]. The existing surveys and reviews, discuss only one or two forecasting horizons (short-term, medium-term). Whereas, all the forecasting horizons, i.e., short-term, medium-term and long-term are discussed in this study. An analysis is presented on electricity load forecasting with big data approaches [6]-[16] and conventional data [17]-[36].

List of abbreviations used in this article is given in Table 1 and list of symbols is shown in Table 2. A comparison of traditional and big data analysis is presented in Table 3. Rest of the paper is organized as: Section 2 is comparison of forecasting models, Section 3 is critical analysis and section 4 is conclusion.

II. Comparison of Forecasting Models

In this section, the forecasting models are categorized as: uni-variate (time series) models and multi-variate models. Brief explanation of sub categories of these models, is also

TABLE III: Comparison of traditional data and big data analysis.

Feature	Traditional Data	Big Data
Size	Limited size	Very huge (terabytes, exabytes)
Sources	Power utilities production data only	All the influential factors, e.g., population, weather, economic conditions, government policies, customer behavior patterns, etc.
Algorithms	Classical, statistical, Machine learning, AI	Feature extraction, correlation analysis, dimension reduction, deep learning, parallel processing algorithms
Accuracy	High for short term predictions, degrades with noisy data	Accurately model noisy and data, risk of falling in local optimum
Usage / benefits	Can impact decisions in the present, i.e., short-term decision making, used for analysing current situations and short-term forecasting, online monitoring, fault detection (instant response to the situation)	Helps in: long-term decision making, budgeting, investment, policy making, assets allocation, maintenance planning, recruitment strategies etc.

given. The classification hierarchy of prediction models is shown in Fig. 1. Moreover, a comparative analysis of the discussed models is given at the end of this section.

4.1 Load Forecasting based on Time series Models

Electricity consumption recorded at successive equally spaced time intervals is known as electricity consumption time series. Time series forecasters predict the future values based on previously observed values. Following are few popular time series prediction models implemented for forecasting energy consumption.

4.1.1 Autoregressive Integrated Moving Average

ARIMA is the most popular method for time series forecasting. First introduced by Jinkens *et al.*, ARIMA [37] is also known as Jenkins-Box approach. It can calculate the probability of a future value lying in a specified range of values. ARIMA is combination of Auto-Regression (AR) and Moving Average (ML). AR process means that the current value of the series depends on the previous values of same series. ML is a process which assumes that the current deviation of a value from the mean of series depends on the previous deviation. ARIMA is donated as $ARIMA(p, q, d)$, where p is the number of autoregressive terms, q is the number of non-seasonal differences and q is the number of lagged forecast errors (from the prediction equation). Three basic steps of ARIMA are: model identification, parameter estimation and model verification (shown in Fig. 2). For establishing the forecasting equation of ARIMA, the base are the following equations [34]:

$$For \quad d = 0 : y_t = Y_t \tag{1}$$
$$For \quad d = 1 : y_t = Y_t - Y_{t-1} \tag{2}$$
$$For \quad d = 2 : y_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \tag{3}$$

Where y is the d^{th} difference of Y. From the above equations, the generalized equation of ARIMA forecaster can be written as follows:

$$\hat{y}_t = \varepsilon + \phi_1 y_{t-1} + ... + \phi_p y_{t-p} - \theta_1 e_{t-1} - ... - \theta_q e_{t-q} \tag{4}$$

Where, ε is error term, ϕ is the parameter of the auto regressive part and θ is the moving average parameter.

Fig. 2: Steps of ARIMA prediction model.

4.1.2 Artificial Neural Network

ANN is network of interconnected small computational units called neurons, inspired by biological neurons. Equation of multi layer perceptron (shown in Fig. 3) neural networks is given below:

$$y(x_1, \ldots, x_n) = f(w_0 + w_1 x_1 + \ldots + w_n x_n) \tag{5}$$

Where, x_i are the inputs, $f()$ is input to output mapping function, w_i are the weights and w_0 is the bias. The function is given by following equation:

$$f(v) = \frac{1}{1 + e^{-v}} \tag{6}$$

The output activation function can be written as following that is a simple binary discrimination (zero-centered) sigmoid:

$$f(v) = \frac{1 - e^{-v}}{1 + e^{-v}} \tag{7}$$

ANN models can be used for prediction of both time series and multivariate inputs. Some of the popular time series ANN prediction models are Elman network [12], ELM [22],[34], NARX and LSTM [17].

4.1.2.1 Non-linear Autoregressive Network with Exogenous Variable

NARX ia a non-linear and autoregressive recurrent neural network (RNN). It has a feedback architecture, in which output layer is connected to the hidden layers of the network. It is different from back propagation ANN (shown in Fig. 3), as its feed back connection encloses several hidden layers, and not the input layer. NARX also utilizes the memory ability by using the past predicted values or actual observations. It models a nonlinear function by recurrence from the past values of the time series. This recurrence relation is used to predict the new values in time series. The input to the network is the past lagged values of the same time series. For example,

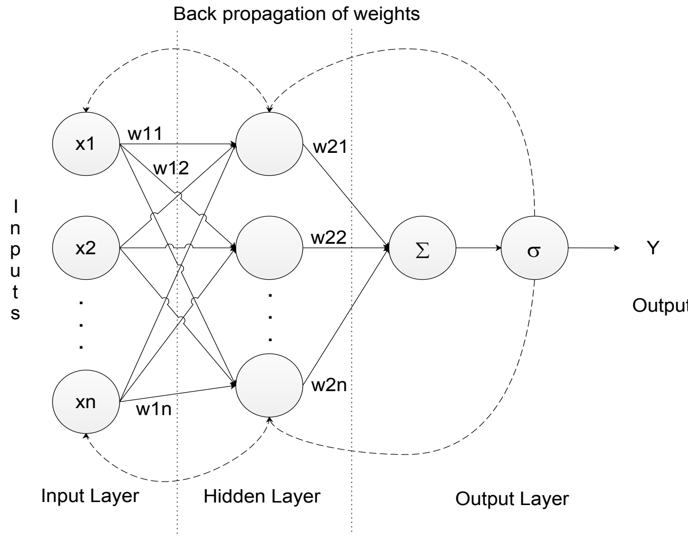

Fig. 3: Simplest ANN: multilayer perceptorn with back propagation of weights.

to predict a future value y_t, the inputs of the network are $(y_{t-1}, y_{t-2}, ..., y_{t-p})$. While the training of network the past predicted values are also used as an input. NARX can be defined by the following equation:

$$\hat{y}_{t+1} = F(y_t, y_{t-1} - ... - y_{t-n}, x_{t+1}, x_t, ..., x_{t-n}) + \varepsilon_t \quad (8)$$

Where, $\hat{y}_{t+1}$ is output of network at time t, that is the one step ahead predicted value of future time, $t+1$. F(.) is the non-linear mapping function of the network (e.g., polynomial, sigmoid, etc.), $y_t, y_{t-1}, ...$ are the true past observations also called the desired outputs, $x_{t+1}, x_t, ...$ are the network inputs that are the lagged values of the time series, n is the number of delays and ε_t is the error term. NARX network is shown in Fig. 4.

4.1.2.2 Long Short Term Memory

LSTM is a deep learning method that is variant of RNN. It is first introduced by Hochreiter *et al.* in 1997 [38]. The basic purpose of proposing LSTM was to avoid the problem of vanishing gradient (using gradient descent algorithm), that occurs while training of back propagation neural network (BPNN) (shown in Fig. 3).. In LSTM every neuron of hidden layer is a memory cell, that contains a self-connected recurrent edge. This edge has a weight of 1, which makes the gradient pass across may steps without exploding or vanishing [17].

4.1.3 Comparative Analysis of Time Series Forecasting Models

ARIMA is better suited to short-term forecasting, on the other hand, ANN models perform better at long-term forecasting. ANNs can detect the underlying patterns of the data with the help of hidden layer nodes, therefore, they can model non-stationary time series [12],[22],[34]. A major benefit of neural network is their ability to flexibly create a nonlinear mapping between input and output data. They can capture the nonlinearity of the time series very well.

4.2 Load Forecasting based on Multivariate Models

Multivariate models take multiple inputs. These inputs are the factors that influence the electricity consumption, also called exogenous variables. These variables can be weather parameters (temperature, humidity, cloud cover, wind speed, etc.), calendar variables (hour of the day, day of the week, etc.), fuel price etc. Multivariate forecasting methods are categorized into three main categories, i.e., ensemble, hybrid and deep learning models. Brief description of these categories and the papers implemented these methods for electricity load forecasting, is given in this section.

4.2.1 Load Forecasting based on Ensemble Models

Ensemble methods are the prediction models that combine different learners in order to achieve better performance. Ensemble models are supervised learning techniques. Multiple weak learning methods are combined to establish a strong and accurate model. Ensemble method is a combination of multiple models, that helps to improve the generalization errors which might not be handled by a single modeling approach (shown in Fig. 5).

Let us assume, there are three prediction models: A, B and C and their prediction accuracy is 88%, 83%, 76% respectively. Suppose, A and C are highly correlated and model B is not correlated with both A and C. In such a scenario, combining models A and C will not reduce the prediction error, however combining model B with model A or model C would improve the accuracy. Every prediction method is assigned a certain weight. These weight are assigned by the standard techniques. Following are some weight assigning techniques:

- *Collinearity calculation:* Calculate the collinearity of all models which each other in order to decide the base models. Exclude the highly correlated models so that the final model is generalized enough to generate less generalization error.

- *Weight assignment by ANN:* Neural Networks can be used to determine the appropriate weights for the prediction models.
- *Weight assignment by Bayesian:* Weights are assigned by calculating the posterior probability of all the models. One of the two techniques can be used: (i) Bayesian model averaging that is an in-sample technique, (ii) Predictive likelihood scaling that is an out-of-bag technique.
- *Equal weight assignment:* Assign equal weights to all the models. This is the simplest method and often performs well as compared to the complex methods. However, it is unable to rank the models based on their performance. Other approaches include bagging, boosting of input samples, learner's forward selection, etc.

4.2.1.1 Random Forest

Random forest (RF) is one of the most popular ensemble learning model. From a large sized data, samples are drawn with replacement that are subsets of data's features. Random samples are taken from the data to establish decision trees (DT). Several DT are made with these randomly drawn data samples, that makes a random forest. DT can be made using any tree generation algorithm, e.g., ID3, CART (Classification And Regression Tree) or c4.5, etc. The parameters of RF algorithm are number of trees and decision tree related parameters like split criteria. For example, 100 trees are generated from a data. A test sample is given for prediction, every tree generates a response to the test sample, that makes 100 predictions for a test sample. A weighted average of these responses is the final predicted value of the random forest. There are many trees in the forest made with different data samples, therefore, the prediction model is highly generalized with no possibility of overfitting.

In paper [24], authors have predicted short-term electricity load of a university campus building using random forest. A two staged models is proposed for load prediction. In the first stage, the electricity consumption patterns are considered using the moving average method. In the second stage, RF is trained with the optimal hyper parameter, i.e., number of trees, split criteria of decision tree, minimum split, etc. The optimal parameters are selected by trial and error method. The model is trained on five years hourly load data. The trained model is verified by modified Time Series Cross Validation (TSCV). The performance of a prediction method degrades if the difference between the training time and prediction time, is very large. This problem arises when training data is much larger as compared to the test data. To overcome this problem, TSCV is applied for one step ahead forecast (point forecast). This proposed model outperforms SVR and ANN in terms of MAPE and RMSE. Results prove the effectiveness of proposed method for short-term load forecasting.

4.2.2 Load Forecasting based on Hybrid Models

Hybrid forecasting methods are combination of data smoothing, regression and other techniques. Hybrid approaches combine the strengths of two or more methods while mitigating their individual weaknesses. Generally, a meta heuristic optimization algorithm is combined with forecasting method, to fine tune the hyper parameters of the forecaster. To train an accurate model on the training data, the hyper parameters of model must be chosen according to the data. Default hyper parameters do not guarantee good training for every input data.

4.2.2.1 Hybrid Support Vector Machine

SVM is a really efficient prediction method. Due to its computational simplicity and accuracy, it is one of the most used methods for prediction. SVM was originally proposed by Vapnik *et al.* in 1995 [39]. SVM create an optimal hyper plane (exactly in the middle) to divide training examples into their respective classes. SVM has three main hyper parameters that are: cost penalty c, insensitive loss function parameter η and sigma kernel parameter σ. SVM predictor can be written in the form of following equation:

$$g(x) = \text{sign}\Big(\sum_i y_i \alpha_i K(x_i, x) + b\Big) \tag{9}$$

$$= \text{sign}\Big(\sum_{i:y_i=1} \alpha_i K(x_i, x) - \sum_{j:y_j=-1} \alpha_j K(x_j, x) + b\Big) \tag{10}$$

$$= \text{sign}\Big(h_+(x) - h_-(x) + b\Big). \tag{11}$$

For a two class problem, the following discriminant can be used:

$$s(x) = \text{sign}[p(x|1) - p(x|-1)], \tag{12}$$

by assuming equal class priors $p(1) = p(-1)$. Suppose, the class conditional densities use Parzen estimates:

$$p(x|1) - p(x|-1) = \frac{\sum_i \beta_i y_i K(x, x_i)}{2\sum_i \beta_i}, \tag{13}$$

Where,

$$\beta_i \geq 0, \tag{14}$$

$$\sum_i \beta_i y_i = 0, \tag{15}$$

Essentially we are picking weights or a distribution of the examples while remaining consistent with the equal class priors assumption.

Now the margin of an example under this discriminant is

$$m_i = y_i s(x_i) = y_i[p(x_i|1) - p(x_i|-1)], \tag{16}$$

that is a measure of correctness of the classified examples. In other words, large and positive margins correspond to confident and correct classifications.

In [33], the authors optimize the hyper parameters of least square SVM (LSSVM), by using modified ABC optimization algorithm. The hybrid model outperform several prediction models. In [35], the author utilize hybrid SVR for prediction of electricity load. The hyper parameters of SVR are tuned using modified firefly optimization algorithm. Firefly algorithm (FA)

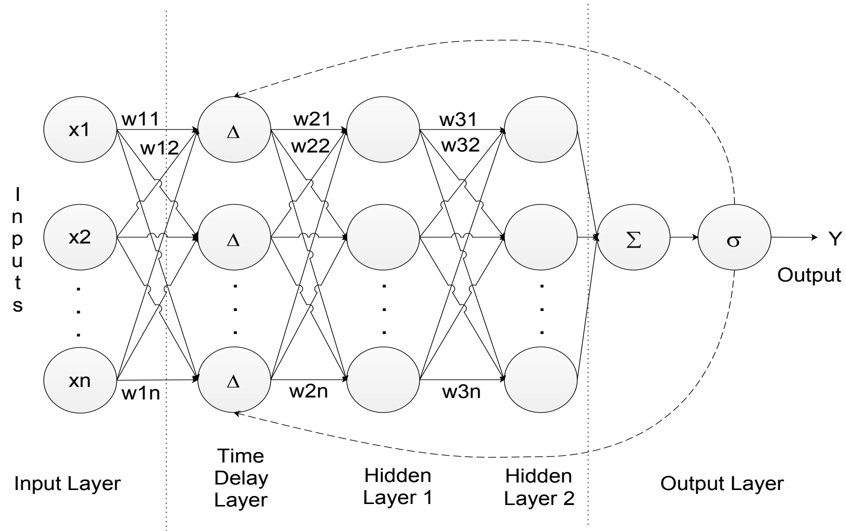

Fig. 4: NARX network.

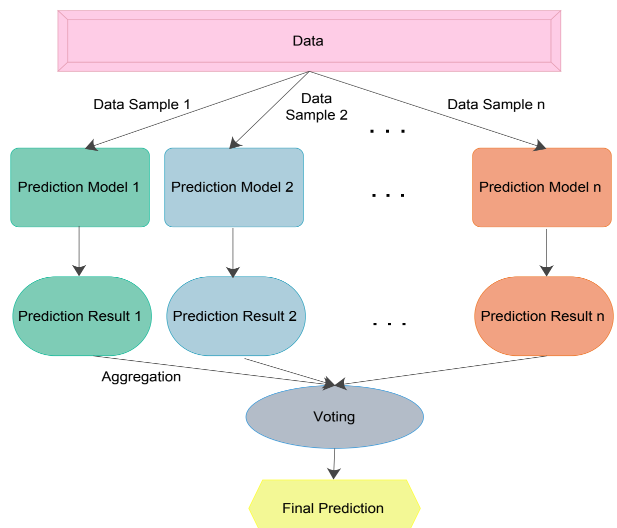

Fig. 5: General representation of ensemble models.

is a nature inspired meta heuristic optimization approach, that is based on flashing behavior of fireflies. The original FA has a possibility of trapping into local optimum. To overcome this issue, two modifications were suggested by the authors. Firstly, improving the population diversity by the aid of two mutations and three cross over operations. Secondly, encouraging the total firefly population to move toward the best promising local or global individual. The SVR model is optimized using enhanced FA. The prediction results proves the effectiveness of this hybrid model. It outperforms several prediction methods, i.e., ANN, ARMA, PSO-SVR, GA-SVR, FA-SVR, etc.

4.2.2.2 Hybrid ANN

The performance of ANN depends on how well the model is fit on the training data. The hyper parameters of ANN are number of neurons, number of hidden layers, learning rate, momentum and bias. A hybrid ANN prediction model is proposed in [18]. The hyper parameters of ANN are optimized using genetic algorithm. The results prove the efficiency and good accuracy of proposed model as compared to other models.

4.2.3 Load Forecasting based on DNN Models

DNN are variants of ANN, that has deep structure with number of hidden layers cascaded into the network. Automatic feature learning capability of DNN allows the network to learn the non-linear complex function, and create mapping from input to output without requirement of hand crafted features [13],[17].

4.2.3.1 Stacked Autoencoder

Autoencoder is a feed forward neural network, that is a unsupervised learning method. As the name suggests, autoencoders encodes the inputs by using an encoder function $y = f(x)$. The encoded values are reconstructed on the output layer by passing through a decoder function $x\prime = g(x)$. The reconstructed out can be written as, $x\prime = g(f(x))$. Basically, the inputs are copied to output layer by passing through hidden layers. The purpose of using autoencoders is the dimensionality reduction of input data. In stacked autoencoder, multiple encoding layers are stacked together as hidden layers of the network as shown in the Fig. 6. The equation of autoencoders is:

$$x\prime = g(wx + b) \tag{17}$$

Where, $x\prime$ are the reconstructed inputs, $g(.)$ is the encoding function, w are the weights and b is the bias.

4.2.3.2 Restricted Boltzman Machine

Visible units are conditionally independent on hidden units and vice versa. For a RBM, energy function can be calculated using following equation:

$$Energy(v, h) = -b'h - c'v - h'Wv.$$

Where b, c are offsets or biases and W comprises the weights connecting units The joint probability of (v,h)

$$P(v,h) = \frac{1}{Z} e^{-Energy(v,h)}$$

Where Z is the normalization term.

- Given initial $v^{(0)}$, we sample $h^{(0)} \sim sigm(Wv^{(0)} + c)$
- Then it can be sampled $v^{(1)} \sim sigm(W' h^{(0)} + b)$
- After t steps, its obtain $(h^{(t)}, v^{(t)})$
- As $t \rightarrow \infty$, sample $(h^{(t)}, v^{(t)})$ are guaranteed to be accurate sample of $P(v,h)$

4.2.3.4 Convolution Neural Network

CNN is a feed forward ANN, that perform mathematical operation convolution on input data. Generally, CNN has three basic layers that are used to build the network. These layers are convolution, rectified linear unit (ReLU) and pooling layer.

TABLE IV: Comparison of existing methods for load prediction.

Inputs	Platform	Duration	Forecast Horizon	Region	Prediction Method	Features	Limitations
Historic energy consumption, demand	Daily, hourly and 15 minutes energy consumption of entertainment venues	2012-2014	Medium term, month ahead	Ontario, Canada	Artificial Neural Networks, SVR [6]	Suitable for big data processing	High time complexity
Historic load, weather data	Hourly load of 1.2 million consumers (residential, commercial, industrial and municipal) of real distribution system	2012	Short term, day and week ahead	Not mentioned	Hierarchical clustering (Bottom up), Classification and regression tree (CART) [7]	Computationally simple	Unable to capture high nonlinearity
Temperature, humidity	Global Energy Forecasting Competition 2012, hourly load and temperature	2004-2007	Short term, Day and week ahead	21 zones of USA	Recency effect [8]	Good performance on big data	High complexity
Historical traffic, weather data	Hourly traffic and weather data observed on a national route from Goyang to Paju, total 20.12 million EVs	2014-2015	Short term, day ahead	Traffic Monitoring System (TMS) of the Ministry of Land, Infrastructure and Transport (MOLIT), South Korea	Decision Tree [9]	Simple	Unable to capture high nonlinearity
Historic load	Every second load of three houses of Smart dataset	May-July 2012	Short term, day and week ahead	Umass Trace online Repository	Adaptive Neuro Fuzzy Inference System (ANFIS) [10]	Good accuracy, simple	Hard to choose suitable kernel method
Historic consumption	15 minutes consumption of Budweiser Gardens event venue, total 43,680 measurements	January-March 2014	Short term, day and week ahead	Ontario, Canada	SVR [11]	Simple and fast	Accuracy degrades with extremely nonlinear data
Historic consumption, weather parameters, social and economical variables of smart city	North-eastern China smart city dataset	2006-2015	Short-term, medium-term	China	Modified Elman Network [12]	Efficiently capture nonlinearity, good accuracy, high convergence rate	High computational and space complexity
Historic load	1.4 million hourly electricity load records	2012-2014	Short term, day and week ahead	Not mentioned	K means, CNN [13]	High accuracy	High complexity
Historic load, electricity parameters	Individual household electric power consumption dataset	2006-2010	Short-term	Not mentioned	CNN [14]	High accuracy, models big data well	High complexity
Historic appliance consumption	(i) Domestic Appliance Level Electricity dataset, (ii)Time series data of power consumption, (iii) Synthetic dataset	2012-2015	Short term	(i) UK-Dale, (ii) Southern England, (iii) Canada	Bayesian network [15]	Efficiently learns data patterns and relationships in data, mitigate missing data, avoid overfitting	High complexity
Weather variables	Historic temperature, humidity and load data	2014-2016	Short-term	Not mentioned	MLR [16]	Simple and fast	Unable to deal with highly non-stationary data
System load, day ahead demand, weather data, hourly consumption	Hourly weather, consumption data of New England	2003–2016	Short-term, day and week ahead	ISO NE CA, New England, USA	Empirical mode decomposition, LSTM [17]	High accuracy, ability of accurately predict long-term load	High complexity
Historic load	Half hourly consumption data of three states	2006-2009	Short-term	New South Wales, State of Victoria, Queensland, Australia	BPNN, RBFNN, GRNN, genetic algorithm optimized back propagation neural network (GABPNN), cuckoo search algorithm [18]	Higher accuracy, outperforms compared optimized ANN models	Possibility of stuck in local optimum
Historic load	5 min ahead forecasting, Australian electricity load data	2006-2007	Short term, hour ahead	Australia	MI, ReliefF, ANN, LR [19]	Trained model on highly correlated inputs, high accuracy	High complexity
Calendar variables, weather variables, lagged loads	15 minute electricity load of "Smart Metering Customer Behavior Trial" from 5000 homes of Irish Social Science Data Archive (ISSDA)	2009-2010	Ireland, New York	Very Short-term, 15 minutes and hour ahead	ANN [20]	Robustness to noisy data, automatic feature engineering	Computationally expensive, requires large training data
Historical load	15 minutes load of individual household meter data	2010-2012	Short-term	Taipei, Taiwan	Decision tree, BPNN [21]	Robustness to noisy data, high accuracy	High complexity, vanishing gradient problem leading to overfitting
Temperature, date type	30 minutes load from Smart meter data of Irish households from the Irish Social Science Data Archive (ISSD), 3000 households	2009-2010	Short term	Ireland	K-mean, Online Sequential ELM [22]	Fast in learning	Difficult to select appropriate kernel function
Temperature, annual holidays, maximum daily electrical loads	EUNITE, a historical electricity load dataset	1997-1998	Short term	Middle region of the Delta in Egypt	Hybrid KN3B predictors, KNN and NB classifier [23]	High accuracy	Computationally expensive
Historic load, weather variables	Hourly load, temperature, humidity	2013-2015	Short term, day and week ahead	Not mentioned	Multi-variable linear regression (MLR) [24]	Simple	Unable to model highly nonlinear data well
Historical temperature and power load data	Hartcourt North Building of National Penghu University of Science and Technology	January-May 2015, September-October 2015	Short-term	Taiwan	Multipoint fuzzy prediction (MPFP) [25]	High accuracy	High complexity
Historic load	Real-time hourly load data (in MWHrs.) of NSW State	April-October 2011	Short term, day and week ahead	Australia	RBFNN [26]	High accuracy	High complexity
Outdoor temperature, relative humidity, supply and return chilled water temperature, flow rate of the chilled water	One-year building operational data from campus building in the Hong Kong Polytechnic University	2015	Short-term	Hong Kong	Decision tree model, association rule mining [27]	Simple	Accuracy degrades on noisy, missing data
Historic load	EMS's electricity information collection system data	Not mentioned	Short and medium-term	Not mentioned	Coordination optimization model [28]	High accuracy	High complexity
Weather data, electricity consumption	15-minute intervals consumption data of 5000 households from project with Electric Power Board (EPB) of Chattanooga	2011-2013	Short term, day and week ahead	Chattanooga, Tennessee, U.S.	Sparse coding, ridge regression [29]	High accuracy	High complexity
Historic price, meteorological attributes	Hourly consumption of HVAC system of a five-star hotel in Hangzhou City	Not mentioned	Short term, day ahead	State Grid Corporation of China Hangzhou, China	SVR [30]	Simple	Difficult to select appropriate kernel function
Historic load data	10 minutes load of Belsito Prisciano feeder Azienda Comunale Energia e Ambiente (ACEA) power grid, 10,490 km of Rome city	2009-2011	Short term, 10 minutes and day ahead	Rome, Italy	Echo State Network [31]	High accuracy	Trained network is a black-box, cannot be understood
Indoor and outdoor temperature, humidity, solar radiation, calendar attributes, consumption	Consumption and weather of a university of Girona's office building	2013-2014	Short term, day and week ahead	Not mentioned	ANN, SVR, MLR [32]	Regression models simpler and faster than ANN, however less accurate	ANN: high complexity, LR:unable to capture high nonlinearity in data
Historical load and price	Hourly price and load of NYISO, PJM and New South Wales	2010, 2014	Short term, day ahead	NYISO, PJM, NSW AEMO energy markets	QOABC-LSSVM [33]	High accuracy	High complexity, possibility of overfitting
Historic load	Load Diagrams Dataset	2011-2014	Short-term	Portugal	ELM [34]	High accuracy	High complexity
Historic load	Hourly consumption of 5 cities	2007-2010	Shoer-term	FARS electric power company	Firefly-SVR [35]	High accuracy	High complexity

In the convolution layer a convolution filter is applied to extract features from input data [13]. The convolution operation can be defined by following equation:

$$y(t) = (x * w)(t) = \int x(a)w(t-a)da \tag{18}$$

Where, x is the input, w is the kernel filter and y is the output, that is feature map of input at time t.

4.2.4 Comparative Analysis of Multivariate Forecast Models

This section provides a brief overview of strengths and limitations of prediction models discusses above. Comparative analysis of these models is also given here. The basic limitation of RF is that prediction by large number of trees make the model very complex in terms of computation and time. Therefore, this model will be ineffective for the real-time predictions. RF are fast to train, however the prediction process of trained model is a time consuming process. The scenarios where running time is important, other prediction approaches are preferable.

DNN produce good forecasting results in presence of enough data, big model and high computation. DNN have a significant advantage over other predictors, that it don't require feature engineering (a computationally expensive process). It is highly adaptive models towards the new problems. The major limitation of DNN is that, it require a large amount of data for training a good model. The training of DNN is a very expensive in terms of time and space. The complex most DNN models are trained for weeks with hundreds of special machines containing GPUs (Graphics Processing Unit). Selection of suitable training method and hyper parameters is a difficult task (as no standard theories are present). However, DNNs are the most suitable prediction methods for big data as it has a great computational power [12],[13]. The conventional prediction models cannot handle huge volume and complexity present in big data. DNN manages memory by training models on mini batches of training data. It make partitions of data and train parallel on multiple processor cores. The basic features of discussed prediction methods are shown in Table 4.

III. Critical Analysis

The comprehensive survey of recent load forecasting methods lead us to the following findings. These finding can help in improving comprehension of load forecasting.

- *Critical Comment* 1*:* Modifying of optimization algorithm to converge fast may led to fall in the local optimum and unstable solution.
- *Critical Comment* 2*:* DNN are computationally expensive. In process of selecting optimal network parameters, the number of neurons in hidden layers and number of layers, increase should be in very small successive steps. Because both time and space complexity increase with increase in number of layers or neurons.
- *Critical Comment* 3*:* The optimization of a predictor's hyper parameters for a certain test dataset may lead to over fitting on that specific dataset [10],[18],[23],[35]. This optimized model is not guaranteed to perform well on the unseen data. Therefore, degree of optimization of any algorithm is a matter of special care.
- *Critical Comment* 4*:* For establishing any prediction model, enough data must be fed as model input, as the load data contains seasonality. Enough data that cover the whole seasonality pattern should be input for development of stable and generalized prediction model.
- *Critical Comment* 5*:* The study of relevant literature of load forecasting reveals that the forecasting of long-term energy load is very rare. There is a lot of research scope in the field of long-term energy forecasting as this area is still very immature.
- *Critical Comment* 6*:* Big data is not considered in most of the analysis performed through load forecast [16]-[36]. Analysis of big data can unveil the un-precedent insights useful for market operation planning and management.

IV. Conclusion

This work is expected to serve as an initial guide for those novice researchers, who are interested in the area of energy consumption prediction. Particularly, energy big data is focused in this study. Following conclusions are drawn from this study:

1) Most of the research work is on short or medium-term load forecasting. Long-term term load forecasting is an area that still needs to be explored in detail.
2) There is no universal technique for electricity load prediction and the choice of prediction models depends on the scenario and forecast horizons.
3) It is concluded that multivariate prediction models are suitable for large dataset, whereas, univariate predictors perform well on small datasets.
4) Overall, deep learning prediction methods outperform all the classic and machine learning prediction methods in terms of accuracy. As well as, their high computational power makes them the most suitable choice for big data prediction and analytics, where other machine learning methods cannot perform very well. Furthermore, DNN has proved to be an effective method for long-term forecasting.
5) Energy big data analytics is an emerging field. There is a lot of research scope for novice researchers in this area. The unprecedented insights drawn from big data can be beneficial for energy utilities in: improving service quality, maximizing profit, detecting and preventing energy thefts and many other ways.

References

[1] Zhou, K., Fu, C. and Yang, S., 2016. Big data driven smart energy management: From big data to big insights. *Renewable and Sustainable Energy Reviews*, 56, pp.215-225.

[2] Fallah, S.N., Deo, R.C., Shojafar, M., Conti, M. and Shamshirband, S., 2018. Computational Intelligence Approaches for Energy Load Forecasting in Smart Energy Management Grids: State of the Art, Future Challenges, and Research Directions. *Energies*, 11(3), p.596.

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

[4] Martinez-Alvarez, F., Troncoso, A., Asencio-Cortes, G. and Riquelme, J.C., 2015. A survey on data mining techniques applied to electricity-related time series forecasting. *Energies*, 8(11), pp.13162-13193.

[5] Amasyali, K. and El-Gohary, N.M., 2018. A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81, pp.1192-1205.

[6] Grolinger, K., L'Heureux, A., Capretz, M.A. and Seewald, L., 2016. Energy forecasting for event venues: big data and prediction accuracy. *Energy and Buildings*, 112, pp.222-233.

[7] Zhang, P., Wu, X., Wang, X. and Bi, S., 2015. Short-term load forecasting based on big data technologies. *CSEE Journal of Power and Energy Systems*, 1(3), pp.59-67.

[8] Wang, P., Liu, B. and Hong, T., 2016. Electric load forecasting with recency effect: A big data approach. *International Journal of Forecasting*, 32(3), pp.585-597.

[9] Arias, M.B. and Bae, S., 2016. Electric vehicle charging demand forecasting model based on big data technologies. *Applied energy*, 183, pp.327-339.

[10] Sulaiman, S.M., Jeyanthy, P.A. and Devaraj, D., 2016, October. Big data analytics of smart meter data using Adaptive Neuro Fuzzy Inference System (ANFIS). *International Conference on Emerging Technological Trends (ICETT)*, pp.1-5.

[11] Grolinger, K., Capretz, M.A. and Seewald, L., 2016, June. Energy consumption prediction with big data: Balancing prediction accuracy and computational resources. *2016 IEEE International on Congress Big Data*, pp.157-164.

[12] Wei, Z., Li, X., Li, X., Hu, Q., Zhang, H. and Cui, P., 2017, August. Medium-and long-term electric power demand forecasting based on the big data of smart city. *Journal of Physics: Conference Series*, 887(1), pp.012025-012033.

[13] Dong, X., Qian, L. and Huang, L., 2017, February. Short-term load forecasting in smart grid: A combined CNN and K-means clustering approach. *IEEE International Conference on Big Data and Smart Computing (BigComp) 2017*, pp.119-125.

[14] Amarasinghe, K., Marino, D.L. and Manic, M., 2017, June. Deep neural networks for energy load forecasting. *IEEE 26th International Symposium on Industrial Electronics (ISIE) 2017*, pp.1483-1488.

[15] Singh, S. and Yassine, A., 2018. Big data mining of energy time series for behavioral analytics and energy consumption forecasting. *Energies*, 11(2), p.452.

[16] Saber, A.Y. and Alam, A.R., 2017, November. Short term load forecasting using multiple linear regression for big data. *IEEE Symposium Series on Computational Intelligence (SSCI) 2017* pp.1-6.

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

[18] Xiao, L., Wang, J., Hou, R. and Wu, J., 2015. A combined model based on data pre-analysis and weight coefficients optimization for electrical load forecasting. *Energy*, 82, pp.524-549.

[19] Koprinska, I., Rana, M. and Agelidis, V.G., 2015. Correlation and instance based feature selection for electricity load forecasting. *Knowledge-Based Systems*, 82, pp.29-40.

[20] Quilumba, F.L., Lee, W.J., Huang, H., Wang, D.Y. and Szabados, R.L., 2015. Using Smart Meter Data to Improve the Accuracy of Intraday Load Forecasting Considering Customer Behavior Similarities. *IEEE Transaction on Smart Grid*, 6(2), pp.911-918.

[21] Hsiao, Y.H., 2015. Household Electricity Demand Forecast Based on Context Information and User Daily Schedule Analysis From Meter Data. *IEEE Transaction on Industrial Informatics*, 11(1), pp.33-43.

[22] Li, Y., Guo, P. and Li, X., 2016. Short-term load forecasting based on the analysis of user electricity behavior. *Algorithms*, 9(4), p.80.

[23] Saleh, A.I., Rabie, A.H. and Abo-Al-Ez, K.M., 2016. A data mining based load forecasting strategy for smart electrical grids. *Advanced Engineering Informatics*, 30(3), pp.422-448.

[24] Moon, J., Kim, K.H., Kim, Y. and Hwang, E., 2018, January. A Short-Term Electric Load Forecasting Scheme Using 2-Stage Predictive Analytics. *IEEE International Conference on Big Data and Smart Computing (BigComp) 2018*, (pp. 219-226). IEEE.

[25] Chang, H.H., Chiu, W.Y. and Hsieh, T.Y., 2016. Multipoint fuzzy prediction for load forecasting in green buildings. *International Conference on Control Robotics Society*, pp.562-567.

[26] Lu, Y., Zhang, T., Zeng, Z. and Loo, J., 2016, December. An improved RBF neural network for short-term load forecast in smart grids. *IEEE International Conference on Communication Systems (ICCS) 2016* pp.1-6).

[27] Xiao, F., Wang, S. and Fan, C., 2017, May. Mining Big Building Operational Data for Building Cooling Load Prediction and Energy Efficiency Improvement. *IEEE International Conference on Smart Computing (SMARTCOMP) 2017* pp.1-3.

[28] Fu, Y., Sun, D., Wang, Y., Feng, L. and Zhao, W., 2017, October. Multi-level load forecasting system based on power grid planning platform with integrated information. *IEEE Chinese Automation Congress (CAC) 2017* pp.933-938.

[29] Yu, C.N., Mirowski, P. and Ho, T.K., 2017. A sparse coding approach to household electricity demand forecasting in smart grids. *IEEE Transactions on Smart Grid*, 8(2), pp.738-748.

[30] Chen, Y., Tan, H. and Song, X., 2017. Day-ahead Forecasting of Non-stationary Electric Power Demand in Commercial Buildings: Hybrid Support Vector Regression Based. *Energy Procedia*, 105, pp.2101-2106.

[31] Bianchi, F.M., De Santis, E., Rizzi, A. and Sadeghian, A., 2015. Short-term electric load forecasting using echo state networks and PCA decomposition. *IEEE Access*, 3, pp.1931-1943.

[32] Massana, J., Pous, C., Burgas, L., Melendez, J. and Colomer, J., 2015. Short-term load forecasting in a non-residential building contrasting models and attributes. *Energy and Buildings*, 92, pp.322-330.

[33] Shayeghi, H., Ghasemi, A., Moradzadeh, M. and Nooshyar, M., 2015. Simultaneous day-ahead forecasting of electricity price and load in smart grids. *Energy Conversion and Management*, 95, pp.371-384.

[34] Ertugrul, O.F., 2016. Forecasting electricity load by a novel recurrent extreme learning machines approach. International Journal of *Electrical Power & Energy Systems*, 78, pp.429-435.

[35] Kavousi-Fard, A., Samet, H. and Marzbani, F., 2014. A new hybrid modified firefly algorithm and support vector regression model for accurate short term load forecasting. Expert systems with applications, 41(13), pp.6047-6056.

[36] Barak, S. and Sadegh, S.S., 2016. Forecasting energy consumption using ensemble ARIMA-ANFIS hybrid algorithm. *International Journal of Electrical Power & Energy Systems*, 82, pp.92-104.

[37] Box, G.; Jenkins, G. Time Series Analysis: Forecasting and Control; John Wiley and Sons: Hoboken, NJ, USA, 2008.

[38] Sepp Hochreiter and Jurgen Schmidhuber. Long short-term memory. Neural Computation, 9(8):1735-1780, 1997.

[39] Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine learning*, 20(3), pp.273-297.