2017 16th IEEE International Conference on Machine Learning and Applications

A Review of Deep Learning Methods Applied on Load Forecasting

Abdulaziz Almalaq, George Edwards
Department of Electrical and Computer Engineering
University of Denver
Denver, CO 80210
a.almalaq@hotmail.com, george.edwards@du.edu

***Abstract*—The utility industry has invested widely in smart grid (SG) over the past decade. They considered it the future electrical grid while the information and electricity are delivered in two-way flow. SG has many Artificial Intelligence (AI) applications such as Artificial Neural Network (ANN), Machine Learning (ML) and Deep Learning (DL). Recently, DL has been a hot topic for AI applications in many fields such as time series load forecasting. This paper introduces the common algorithms of DL in the literature applied to load forecasting problems in the SG and power systems. The intention of this survey is to explore the different applications of DL that are used in the power systems and smart grid load forecasting. In addition, it compares the accuracy results RMSE and MAE for the reviewed applications and shows the use of convolutional neural network CNN with k-means algorithm had a great percentage of reduction in terms of RMSE.**

***Keywords*-Artificial neural networks, Forecasting, Learning (artificial intelligence), Machine learning, Smart grids.**

I. Introduction

Deep learning (DL) is a type of machine learning that has deeper inner hidden layers cascaded into the network. Its goal is to make machines like computers think and understand as human thinks by mimicking the grid of the human brain connection. Artificial Intelligence (AI) has been investigated in many industries for automation processes such as automated labor, picture and audio detection, decision maker in critical fields and scientific research assistants. Machine learning (ML) algorithms are the essential algorithms of AI which extract patterns from raw data to make subjective decisions. One of the simplest ML algorithms is logistic regression that can recommend the use of caesarean delivery decision for patients [1]. In addition to simple ML algorithms, nave Bayes is a tool for splitting spam emails from legitimate email [1]. ML algorithms are subdivided into supervised learning and unsupervised learning. A supervised learning algorithm is applied to a dataset that has features and each of those features associated with a label [1]. However, a unsupervised learning algorithm is applied to a dataset which has many features in order to learn useful properties from the structure of the dataset [1]. The dataset is always split into training and testing sub-datasets in the learning algorithms. The training dataset helps the model to learn from the raw dataset and the test dataset helps to validate the output of the model. A limitation of ML algorithms is the insufficient learning models for high dimensional datasets. Another approach of AI is representation learning which is more developed than ML. ML is widely used in load forecasting on difference papers and different applications. Artificial Neural Network (ANN) is the standard neural network with one input layer, one hidden layer and one output layer. ANN is used for load forecasting in [2] [3] [4] [5]. A review of using neural network for short term load forecasting is presented in [6] for collection of papers (published between 1991 and 1999). Another approach of ML that is used for load forecasting is Support Vector Machine (SVM) in [7] [8] [9]. The architecture of DL is more complex than other neural network models because it has more layers and computations [1]. Essentially, DL builds sophisticated networks that are applicable to analysis results in many disciplines such as computer vision, audio processing, natural language processing (NLP), bioinformatics, finance, power systems, smart grids and more. DL is a complex computational model which is designed in multiple hidden layers and represents the data in different abstractions by using features as inputs to the DL model. DL algorithms are used for different tasks of learning especially for unsupervised learning. Load forecasting is one of the applications that benefited from DL algorithms in many papers in the literature. This paper shows an overview of the DL algorithms and methods that are used in smart grid load forecasting. Table I shows a collection of reviewed papers for DL methods used in the load forecasting papers (published between 2014 and 2016). The following parts of this paper are a survey of state of the art DL methods a

Table I
SHOWS A REVIEW OF DL METHODS USED IN THE LOAD FORECASTING PAPERS.

DL method	Number of Papers	Methods combined with
Autoencoder	2	Autoencoder and LSTM
RNN	1	Deep RNN
LSTM	2	Sequence to Sequence
CNN	1	k-mean Clustering
RBN	1	-
DBN	1	DBN and SVR
DBM	1	Predictive DBM

0-7695-6321-X/17/31.00 ©2017 IEEE
DOI 10.1109/ICMLA.2017.0-110

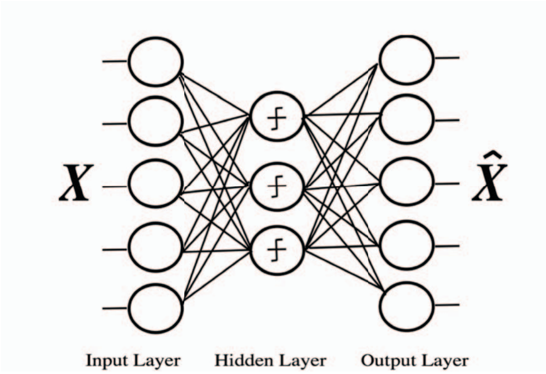


Figure 1. Shows the architecture of the Autoencoder learning algorithm.

load forecasting. In section II introduction to common DL methods and their computational graphs that are applied on smart grid load forecasting. Section III will provide a brief overview of relevant DL architectures of state of the art load forecasting problems in the SG. Section IV is the conclusion of the survey.

II. Different Deep Learning Methods

A. Autoencoder

Autoencoder is one of the feed forward neural network that is used to copy input neurons to output neurons by passing through a hidden layer or multiple hidden $\boldsymbol{h}$ layers as in stacked Autoencoders [1]. The main parts of Autoencoder network architecture are based on encoder function $\boldsymbol{h} = f(x)$ and a decoder for reconstruction $\widehat{\boldsymbol{x}} = g(\boldsymbol{h})$. Therefore, the reconstructed output is $\widehat{\boldsymbol{x}} = g(f(x))$ which copies the data input. The mathematical representation of Autoencoder is described as follows:

$$\widehat{\boldsymbol{x}} = g(\boldsymbol{Wx}+\boldsymbol{b}) \tag{1}$$

where $\boldsymbol{x}$ is the input, $\boldsymbol{W}$ is the weights, $\boldsymbol{b}$ is the bias and g is the activation function which could be sigmoid or a rectified linear function. Figure 1 shows the simple architecture of Autoencoder for three layers input, hidden and output layer. This learning algorithm is usually used for dimensionality reduction, feature learning or corrupted data reconstruction. The Autoencoder models that are used for these kinds of problems are known as Undercomplete Autoencoder, Sparse Autoencoder and Denoising Autoencoder, respectively [1].

B. Recurrent Neural Network (RNN)

RNN network is basically sequential data neural network processor because it has internal memory to update the state of each neuron in the network with the previous input. RNN are usually trained with Backpropagation. However, it has a weakness and it usually fails because of vanishing gradient descent [10]. Deep RNN stacks more layers into the architecture of RNN that provides more significant benefits [8].

The computational graph of RNN has N sequential $\boldsymbol{x}^i$ input and sequential $\boldsymbol{y}^i$ outputs when sequence time is denoted from t to τ [11]. In the following mathematical representation of the RNN model computational graph, the number of layer is denoted as l from 1 to N . The activation function is denoted correspondingly to l as h_l:

$$a_1(t) = b_1 + W_1 \times h_1^{(t-1)} + U_1 \times x^{(t)} \tag{2}$$

$$a_l(t) = b_l + W_l \times h_l^{(t-1)} + U_l \times h_l^{(t)} \tag{3}$$

$$y(t) = b_N + W_N \times h_N^{(t-1)} + U_N \times h_N^{(t)} \tag{4}$$

where $x(t)$ is the input and $y(t)$ is the predicted output.

C. Long Short Term Memory (LSTM)

The RNN model fails on vanishing gradient decent as mentioned above, so LSTM is designed to compensate for its failure. LSTM is designed to provides a longer-term memory. In the model of LSTM, there are internal self-loops that are used for storing information. There are five important elements in the computational graph of the LSTM: 1) input gate, 2) forget gate, 3) output gate, 4) cell, and 5) state output. Figure 2 shows the LSTM computational model in the cell layer and these are combined with other cells to form the RNN model. The gates operations such as reading, writing and erasing are performed in the cell memory state [10]. The following equations show the mathematical representation of the LSTM model.

$$x_i = \sigma(x_t W_{x_i n} + h_{(t-1)} W_{x_i m} + b_i) \tag{5}$$

$$x_f = \sigma(x_t W_{x_f n} + h_{(t-1)} W_{x_f m} + b_f) \tag{6}$$

$$x_o = \sigma(x_t W_{x_o n} + h_{(t-1)} W_{x_o m} + b_o) \tag{7}$$

$$U = \tanh(x_t W_{Un} + h_{(t-1)} W_{Um} + b_U) \tag{8}$$

$$C_t = x_f \times C_{t-1} + x_i \times U \tag{9}$$

$$h_t = x_o \times \tanh(U) \tag{10}$$

where $x_i$is the input of the input gate, x_f is the input of the forget gate, x_o is the input of the output gate, U is the update signal, C_t is the state value at the time t and h_t is the output of the LSTM cell. The memory state can be modified by the decision of the input gate using sigmoid function with on/off state. If the value of the input gate is very small and close to zero, there will be no change in the state cell memory C_t . Stacked LSTM can be represented in multi LSTM layers in the network model.

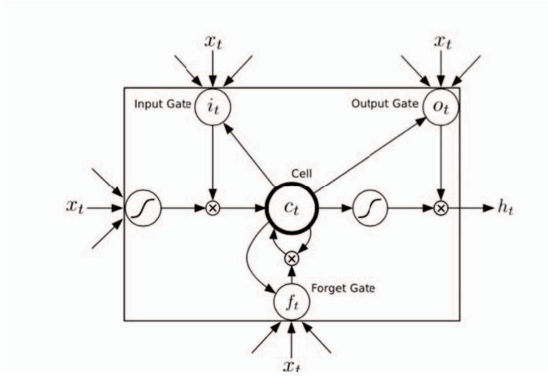


Figure 2. Shows LSTM cell and architecture graph.

D. Convolutional Neural Network (CNN)

CNN is a feedforward neural network which mimics the human neurons in its structure. This type of deep learning CNN has been applied widely in the visual and audio processing, video recognition and natural language processing. CNN are usually used for processing grid data topology [1]. The construction of the time series data is 1D grid at a time interval and image data construction is 2D grid of pixel [1]. The convolution mathematical operation is employed in at least one of the CNN layers [1]. The operation of convolution in signal processing is described generally in the following equation:

$$s(t) = (x * w)(t) = \int x(a)w(t-a)da \tag{11}$$

where x is called input, w is the kernel filter and s the output which is called feature map for the continuous time t. The discrete time convolution operation is descried as follows:

$$s(t) = (x * w)(t) = \sum_{a=-\infty}^{\infty} x(a)w(t-a)da \tag{12}$$

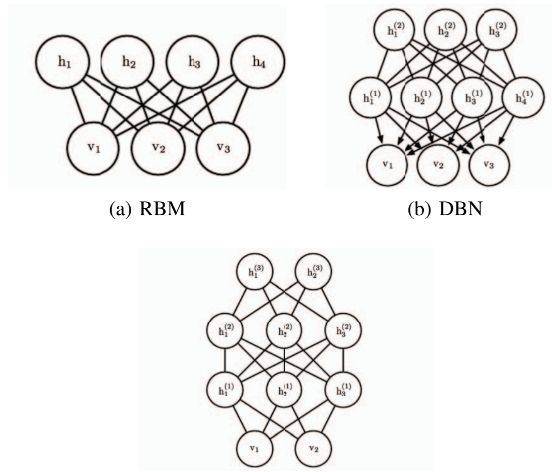
For the two-dimensional image I and a 2D kernel filter K, convolution is also called cross correlation below [1]:

$$s(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i+m, j+n)K(m,n) \tag{13}$$

Typically, CNN consists of three main layers that build the architecture of the network. The first layer has several parallel convolutions layers, the second layer is the detector stage which has rectified linear activation and the third stage is the pooling function layer [1].

E. Restricted Boltzman Machine (RBM)

RBM is one of the most famous Deep probabilistic models which are undirected probabilistic graphical models [1]. RBM has two main layers where the first contains visible inputs and the second layer contains hidden variables. Usually, RBM is stacked in order to make it deeper by building one on top of the other. Figure 3a shows simple RBM for the undirected probabilistic graphical model for two layers.



(a) RBM (b) DBN

(c) DBM

Figure 3. Shows different architecture of a. RBM with undirected connections between visible inputs and hidden variables. b. DBN with directed connections toward visible inputs and the others undirected connections for the hidden layers. c. DBM with undirected connections for one visible inputs layer and multi hidden layers.

F. Deep Belief Network (DBN)

DBN architecture has several layers of hidden units known as stacked RBM with multiple hidden layers that are trained using the Backpropagation algorithm [12]. Basically, the connection units in the DBN architecture is between each unit in a layer with each unit in the next layer, however, there are no intra connection for each layer units [1]. Figure 3b shows DBN for three layers configuration, two hidden and one visible layer where the top two layers are undirected, but, the connection between all other layers should be directed pointing towards the data layer [1]. Generally, a DBN is RBM with more hidden layers, however, RBM usually have only one hidden layer.

G. Deep Boltzman Machine (DBM)

DBM neural network is like RBM architecture but with more hidden variables and layers than RBM. Also, DBM is unlike to DBN because DBM architecture has entirely undirected connections between variable within all layers, such a visible layer and multi hidden layers [1]. Figure 3c shows the architecture of DBM neural network for three hidden layers and one visible layer.

III. Review on Load Forecasting

Smart grid vision is to build a robust, reliable, efficient grid and minimizing the cost of production. The objective of the smart grid is gaining the advancement of the dramatically increase of the usage of technology and communication by investing in bidirectional flow of power and data. Smart grid effectiveness relies on three major roles: 1) load forecasting, 2) dynamic pricing, and 3) demand side management that can help maintain and manages the grid effectively. Specifically, load forecasting is one of the most important factor in planning and operating the power system. One of the most important aspect in the economy, as well as, developed industries is load forecasting for the power systems. Electrical Load forecasting can be influenced by many factors that can change the patterns of load consumptions. There are different types of load forecasting papers that applies DL in the power systems to Short-term forecasting (STLF), Mid-term forecasting (MTLF) and Long-term forecasting (LTLF). Generally, load forecasting can be influenced by many factors such as time, weather and the type of customer especially for STLF problems [13]. However, MTLF and LTLF forecasting are functions of historical data for power consumption plus weather, number of customers, number of appliances and demographic data [13]. STLF mostly have been studied in the literature for a time interval ranging from an hour to couple of weeks. STLF is used in different applications such as real time control, energy transfer scheduling, economy dispatch and demand response [13]. MTLF can be used for the range of interval from a month to five years in order to plan for near future power plants and show the dynamics of the power system within the time interval. LTLF has been used for an average interval of 5 to 20 years. Usually, LTLF is used to plan the generation power plants by the size and the type in order to satisfies the future requirement and cost efficient [14]. In the following paragraphs, provides an overview of how DL methods that are described previously are used to solve a problem of load forecasting.

The Autoencoder DL algorithm is used for STLF of the electricity price [5]. The author of this study proposed Stacked Denoising Autoencoders for short term forecasting. The forecasting approaches are designed for two models online forecasting and one day ahead forecasting. Stacked Denoising Autoencoder showed effective forecasting results especially for day ahead forecasting [15]. The results are compared with state of art forecasting methods such as classical Neural network(NN), Support Vector Machine (SVM), multivariate adaptive regression splines (MARS) and least absolute shrinkage and selection operator (Lasso) [15]. Another study has investigated Autoencoder and Long short term memory for forecasting renewable energy power plants [16]. The study combined the Autoencoder and LSTM for forecasting and compared the results to the state of art approaches such Artificial Neural Network, LSTM and DBN [16]. The method is applied for 21 solar power plants when proposed method showed decreasing in the average of root mean square error (RMSE) for training and testing results. In most of the papers RMSE and mean absolute error (MAE) are used to evaluate the performance of the proposed method in comparison with other methods that are in the literature. RMSE and MAE are used usually to determine the error of the forecasting accuracy. Performance equations of DL methods used were as follows:

$$RMSE(y,x) = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y-x)^2} \tag{14}$$

$$MAE(y,x) = \frac{1}{N}\sum_{n=1}^{N}|y-x| \tag{15}$$

Where x is the measured input time series, y is the predicted output time series and N are the number of samples of the time series. The study of Autoencoder and LSTM claimed great performance of the proposed method with RMSE 0.0713 and followed by the DBN with RMSE 0.0714.

A recent study for household load forecasting has used a novel approach of Pooling Deep Recurrent Neural Network (PDRNN) [11]. The authors claimed that adding more layers to the neural network improves the forecasting performance [11]. This study has used a pool of inputs for a group of customers to increase the data diversity and volume [11]. The data of this study is collected from 920 smart metered customers in Ireland. The authors also used GPU in their hardware in order to accelerate the computational time by parallelizing the models. The results were compared to the literature load forecasting techniques such as ARIMA, SVR and DRNN [11]. This novel approach showed improvement results in RMSE when compared with; ARIMA by 19.5%, SVR by 13.1% and DRNN by 6.5% [11].

One of the common methods of deep learning which is been used in short term load forecasting is Long Short Term Memory algorithm (LSTM). This method is used for two approaches in building energy load such as standard LSTM and LSTM-based sequence to sequence (S2S) [10]. They are implemented on consumption dataset of residential load that is trained and tested with one hour and one minute resolution [10]. In this study, the standard LSTM performed successfully in one hour resolution, however, failed in one minute resolution [10]. For the second approach, both dataset resolutions performed well compared to results to the literature [10]. Another study investigated LSTM for residential load for short term forecasting has studied the usage of appliances for individual load forecasting [17]. The researchers showed their algorithm outperforms the state of art approaches of residential load forecasting [17].

Convolutional Neural Network (CNN) with k-mean clustering was applied in short term forecasting [18]. The K-mean algorithm is used on a large dataset to create clusters for training the CNN [18]. The study selected the data from

August 2014 for summer data and from December 2014 for winter data to apply the method [18]. The comparison results in this study shows great improvement in terms of RMSE for CNN with k-means 0.2194 in summer experiment and 0.2399 in winter experiment. Neural network was applied to both experiment which resulted RMSE 0.2379 in summer and 0.2839 in winter. The author's also applied CNN only on the datasets in order to compare the results with the proposed method. CNN only resulted in lower accuracy in terms of RMSE 0.2502 in summer and RMSE 0.2614 in winter. The study concluded that CNN outperforms other methods with help of clustering techniques [18].

A study that has demonstrated great performance using DL in costumer's load electricity forecasting used Restricted Boltzmann Machine (RBM) method [12]. The training process in the study was driven by two cases pre-training process and Rectified Linear Unit without pre-training. DL approach is structured by using a heuristic method to determine the number of hidden neurons that are used in each hidden layer [12]. There were four hidden layers and 150 hidden neurons. A sigmoid function is used for RBM pre-training. A linear function is used for the prediction layer which is the output layer [12]. DL forecasting results are compared with other known methods in the literature for forecasting such sallow neural network (SNN), double seasonal Holt-Winters (DSHW) and autoregressive integrated moving average (ARIMA) [12]. The results are compared with the previous work in the literature to verify the performance of the proposed approach of DL short term load forecasting. In addition, the results are verified by mean absolute percentage error (MAPE) and relative root mean square error (RRMS) [12]. This approach reduced (MAPE) and (RRMS) by 17% and by 22% in comparison with (SSN) and by 9% and by 29% in comparison with (DSHW) [11].

Deep belief network (DBN) algorithm was proposed in a study for load short forecasting [19]. In this study, the authors aimed to improve the performance of DBN with load forecasting by using ensemble methods and emerge Support Vector Regression (SVR) [19]. The authors used three electricity load demand datasets and three regression datasets to apply their proposed method. The results show that the ensemble deep learning by combining DBN with SVR outperformed SVR, Feedforward NN, DBN and ensemble NN [19]. The prediction results for load demand of South Australia for the proposed method with RMSE 30.598, however, RMSE for SVR is 44.674 and RMSE for Feedforward NN is 38.8585.

A study that proposed Predictive Deep Boltzmann Machine (PDBM) for predicting wind speed was applied to wind energy [20]. The authors used raw wind speed data to forecast short term and long term wind speed. The algorithm was used to predict wind speed one hour ahead and one day ahead. The proposed method showed improved performance in comparison with autoregressive (AR), adaptive neuro-fuzzy inference system (ANFIS), and support vector machine (SVM) for support vector regression (SVR) [20]. The study was made for short term and long term forecasting. Short term forecasting is designed to forecast from 10 minutes ahead to 2 hours ahead. The result of 10 minutes ahead in terms of mean square error (MSE) is 0.2951 for the proposed method as correspond to the MSE for SVR is 0.6340. Long term forecasting is designed to forecast from one day ahead to seven days ahead. The proposed algorithm for one day ahead forecasting resulted MSE 1.2926 which is an improvement over SVR that has MSE 1.3678.

Table II
SHOWS PERCENTAGE OF REDUCTION FOR RMSE IN COMPARISON BETWEEN PROPOSED METHOD AND BENCHMARK STATE OF ART METHOD.

Benchmark method	Proposed Method	RMSE % of reduction
MLP	Autoencoder and LSTM	5.51%
ARIMA	PDRNN	19.2%
NN	CNN	7.7%
CNN	CNN with k-means	12.3%
FNN	DBN and SVR	21.2%
SVR	PDBN	2.85%

The following table summarize most of the reviewed papers in terms of the comparison between the proposed method in each paper and common benchmark method in the state of art. Table II shows RMSE percentage of reduction in comparison with the other method. Table II, the reduction of RMSE in the method of ensemble DBN and SVR is remarkable in comparison with FNN. In addition, the method of PDRNN shows great reduction in comparison to ARIMA. Then, CNN with k-means showed also great percentage of reduction in RMSE.

IV. CONCLUSION

Artificial Intelligence techniques were developed in the last decade for the automation processes and intelligent decisions. Supervised and unsupervised machine learning, representation learning and deep learning algorithms have been applied to many applications in the AI problems for different areas such computer vision, neuroscience, biomedical engineering and power systems. This paper has reviewed the famous deep learning methods that applied to the SG load forecasting. Most of these learning algorithms have successful approached the forecasting analysis and outperformed state of the art forecasting problems of machine learning and neural networks. There are couple of problems due to forecasting analysis regarding the type of load, time, weather, seasons, costumers behaviors and vacations. For example, in residential load consumption forecasting for individuals varies based on the nature of the appliances usage.

Neural network techniques are used widely in the problem of load forecasting. The challenges of neural network for power systems such as training time, upgrading NN training and integration of technologies [14]. This paper has focused on the review for state of the art DL methods in different studies applied to the SG load forecasting. DL showed significant outperformance on many applications of unsupervised learning. Stacked Denoising Autoencoder algorithm is applied for STLF electricity price. It showed great performance fore day ahead forecasting. In addition, Autoencoder and LSTM algorithms were combined to provided great RMSE results in forecasting renewable energy power plants. Pooling Deep RNN showed outperformance other forecasting methods such as ARIMA, SVR on load forecasting of smart meters in Ireland. LSTM sequence to sequence applied to building energy forecasting performed well in the one hour resolution. CNN with k-mean clustering shows great results on RMSE for STLF in summer and winter forecasting. RBM is also used for customer's load forecasting that resulted great performance in comparision with ARIMA, DSHW and SSN. By using ensemble learning and combining DBN and SVR most of the best performances in load forecasting. Lastly, Predictive DBM is used for wind speed forecasting for the smart grid. It showed great performance in comparison with AR and SVR in terms of RMSE. From table II, three combined methods from the reviewed state of the art DL methods for load forecasting showed great percentage of reduction in terms of RMSE. Ensemble DBN with SVR was the highest reduction in comparison with other methods.

There are many papers that are not covered in this review. Therefore, there will be more review in future work on the other applications of SG using DL methods.

REFERENCES

[1] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[2] M. Hayati and Y. Shirvany, "Artificial neural network approach for short term load forecasting for illam region," *World Academy of Science, Engineering and Technology*, vol. 28, pp. 280–284, 2007.

[3] N. Kandil, R. Wamkeue, M. Saad, and S. Georges, "An efficient approach for short term load forecasting using artificial neural networks," *International Journal of Electrical Power & Energy Systems*, vol. 28, no. 8, pp. 525–530, 2006.

[4] D. C. Park, M. El-Sharkawi, R. Marks, L. Atlas, and M. Damborg, "Electric load forecasting using an artificial neural network," *IEEE transactions on Power Systems*, vol. 6, no. 2, pp. 442–449, 1991.

[5] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks:: The state of the art," *International journal of forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[6] H. S. Hippert, C. E. Pedreira, and R. C. Souza, "Neural networks for short-term load forecasting: A review and evaluation," *IEEE Transactions on power systems*, vol. 16, no. 1, pp. 44–55, 2001.

[7] B.-J. Chen, M.-W. Chang *et al.*, "Load forecasting using support vector machines: A study on eunite competition 2001," *IEEE transactions on power systems*, vol. 19, no. 4, pp. 1821–1830, 2004.

[8] P.-F. Pai and W.-C. Hong, "Support vector machines with simulated annealing algorithms in electricity load forecasting," *Energy Conversion and Management*, vol. 46, no. 17, pp. 2669–2688, 2005.

[9] Z. Zhu, Y. Sun, and H. Li, "Hybrid of emd and svms for short-term load forecasting," in *Control and Automation, 2007. ICCA 2007. IEEE International Conference on*. IEEE, 2007, pp. 1044–1047.

[10] D. L. Marino, K. Amarasinghe, and M. Manic, "Building energy load forecasting using deep neural networks," in *Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE*. IEEE, 2016, pp. 7046–7051.

[11] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting–a novel pooling deep rnn," *IEEE Transactions on Smart Grid*, 2017.

[12] S. Ryu, J. Noh, and H. Kim, "Deep neural network based demand side short term load forecasting," *Energies*, vol. 10, no. 1, p. 3, 2016.

[13] S. Khatoon, Ibraheem, A. Kr, and P. Singh, "Effects of various factors on electric load forecasting,," in *International Conference (PIICON), in 2014 6th IEEEPower India*, 2014.

[14] M. T. Haque and A. Kashtiban, "Application of neural networks in power systems; a review," *Power*, vol. 2005, 2000.

[15] L. Wang, Z. Zhang, and J. Chen, "Short-term electricity price forecasting with stacked denoising autoencoders," *IEEE Transactions on Power Systems*, 2016.

[16] A. Gensler, J. Henze, B. Sick, and N. Raabe, "Deep learning for solar power forecasting?an approach using autoencoder and lstm neural networks," in *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference on*. IEEE, 2016, pp. 002 858–002 865.

[17] W. Kong, Z. Y. Dong, D. J. Hill, F. Luo, and Y. Xu, "Short-term residential load forecasting based on resident behaviour learning," *IEEE Transactions on Power Systems*, 2017.

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

[19] X. Qiu, L. Zhang, Y. Ren, P. N. Suganthan, and G. Amaratunga, "Ensemble deep learning for regression and time series forecasting," in *Computational Intelligence in Ensemble Learning (CIEL), 2014 IEEE Symposium on*. IEEE, 2014, pp. 1–6.

[20] C.-Y. Zhang, C. P. Chen, M. Gan, and L. Chen, "Predictive deep boltzmann machine for multiperiod wind speed forecasting," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1416–1425, 2015.