Artificial Intelligence Research and Development
T. Alsinet et al. (Eds.)

doi:10.3233/FAIA240442

Comparative Analysis of Electricity Demand Forecasting at Substation Level

Alex SEGURA [a,1], Jérémy LANCEREAU [b], Joaquim MELÉNDEZ [a] and Carolina M.R. DO CARMO [a]
[a] *Universitat de Girona (UdG), Spain*
[b] *École Nationale Supérieure de l'Électronique et de ses Applications (ENSEA), France*

Abstract.
Growing electrical demand on the electric system along with the rising use of renewable energy sources is highlighting the importance of energy flexibility management on the electric grid. The Electric System Operators at both transmission (TSO) and distribution level (DSO) are responsible to ensure the security of supply and efficiency of the grid under strict balancing conditions (demand equals supply at every time instant). Acting on both generation and demand to maintain this equilibrium considering the technical constraints of the grid is known as flexibility management and it requires accurate generation and demand forecasting to predict possible critical events and react accordingly. The objective of this paper is to analyze the performance of different forecasting methods for predicting demand at the substation level. Substation level data is the result of aggregating the consumption and generation data of multiple points on the grid. Results show that current state of the art algorithms, such as deep learning models, perform better than simpler methods, such as random forests, specially when datasets do not present clearly repetitive profiles. Deep learning models manage to reduce forecasting error by 16% on average compared to random forest models on next day load forecasting, whereas the forecasting error reduction on next hour load forecasting is 5%.

Keywords.
Deep Learning, Machine Learning, Power Systems, Renewable Energy Sources, Smart Grids, Time-series

1. Introduction

The investment in Renewable Energy Sources (RESs) has helped decarbonize the energy system in Europe [1]. However, the installation of these RESs along the distribution grid, combined with increasing electricity demand as well as electrification of the heating and transport sectors [2], has reached a point where the distributed RESs are exerting stress on transmission and distribution networks. This fact has highlighted the importance of not only generation management but also demand side management for a proper operation of electrical grids [3]. Energy flexibility is the ability to manage variations in demand or generation, which is essential in ensuring the stability and efficiency of the grid [4]. The Distribution System Operator (DSO) requires reliable electricity demand fore-

[1] Corresponding Author: Alex Segura, alex.segura@udg.edu

casts in order to, first, properly determine the occurrence of critical events (i.e. congestion or voltage variations) and, second, manage this flexibility (changes on generation or demand) to maintain the grid operating under safe conditions to guarantee the supply.

Forecasting demand, generation, and flexibility (capability to change demand, or generation) is essential for the DSO to operate and manage electric power networks [5]. Energy forecast is also necessary to cope with risk management and to prevent potential congestion problems on the grid. To achieve this, several artificial intelligence (AI) methods are used [6]. They can be divided into three main categories which are: statistical models [7,8], machine learning (ML) models [9,10,11] and deep learning (DL) models [12,13,14].

1.1. Literature review

With the development of information technology, AI methods have replaced mathematical models due to their abilities of learning features and processing data. Consequently, numerous studies and reviews on AI forecasting techniques have been conducted in the literature [15].

In [16], a comparison between conventional models and more recent AI models is made. Conventional models encompass time-series models (derived from auto-regressive and moving average models), regression models (linear, non-linear, logistic models), and gray models. For more recent ML and DL models, they consider Artificial Neural Network (ANN), Support Vector Regression (SVR), and Random Forest (RF). The conclusion drawn is that for daily and hourly energy consumption forecasting, more recent ML and DL-based models are more accurate than conventional models.

A ML-prediction oriented review [17] exposes the superiority of eXtreme Gradient Boosting (XGB) methods compared to Multiple Linear Regression (MLR), ELastic Net (ELN), RF, Gradient Boosting Machines (GBM), or SVR. Additionally, the performance and accuracy of Deep Neural Network (DNN) models are also praised in this research work.

DL (considered as a subgroup of ML methods) has proven its ability to forecast time-series. In particular, these models can be used to highlight inherent abstract characteristics and invariant structures in the data. [18] provides a review of different DL methods such as Auto-encoder, DNN, Convolution Neural Network (CNN), and Recurrent Neural Network (RNN). The article highlights the interest of these models for feature extraction in demand forecasting, achieving good results with CNN for short-term forecasting.

Other articles are more focused on refining the results of a single method. This is the case of [19], which aims to optimize the performance of CNN, widely used in time-series forecasting. The conclusion drawn is that smoothed CNN (a combination of an exponential smoothing with CNN) outperforms other tested methods.

Other methods studied in the literature include Long Short-Term Memory (LSTM) methods. [20] demonstrates the potential of using LSTM, attention-based LSTM, and Seq2Seq LSTM for multi-step prediction. Another major area of research in the field of electricity consumption prediction concerns the creation of multiple models.

1.1.1. Short-term forecasting

Short-term usually includes forecasts up to seventy-two hours ahead, while medium and long-term start from one-week, months to year(s). Several studies on day-ahead electricity forecasting point to the great capabilities of ML models for short, medium and long-term electricity forecasting [21]. Most research is conducted on short-term forecasting. It is more precise and accurate than the medium and long-term [21].

There is, however, strengths and limitations to each particular machine learning algorithm. A RF model is an example of ensemble learning while CNN and LSTM fall under the DL approach. Both DL methods are relatively not easy to use and take long to train. CNN requires large amounts of training set in the building field to yield efficient forecasting performance, while the LSTM is prone to overfitting during the learning process and has an extremely large number of parameters in the network. Other gradient-boosting algorithms besides RF were also tested, such as CatBoost and LightGBM (Light Gradient-Boosting Machine). Despite the small differences between algorithms, RF has been maintained as reference for comparison with DL in this work.

In this paper, performance of prediction algorithms for forecasting electricity demand at the substation level is assessed. Substations refer to the location of power transformers in the grid. Power transformers together with lines and cables are expensive physical devices which replacement requires planning and large investments. Thus, flexibility management is a convenient alternative to avoid occasional events and deferring investment. Data has been obtained as the aggregation of consumption and generation data from customers downstream of these transformers resulting in time-series records with hourly frequency. The paper evaluates the performance of several machine learning algorithm (i.e. RF, CNN and LSTM) in two different scenarios. Data has been gathered by new instruments deployed in the grid with the objective of flexibility management. Thus, main interest is to identify suitable forecasting strategies capable to lever flexibility management.

The rest of the paper is organized as follows: Section 2 reviews the data and the methodology used. In section 3 we present and discuss the results of the machine learning algorithms. Finally, the relevant conclusions and insights are provided in section 4.

2. Materials and methods

2.1. Data description

Data has been gathered in substations of a German grid (Table 1) and it was collected in the period of five years (from January 2018 to December 2022) with one hour sampling rate. This data was collected during FEVER, an European Research & Innovation project. FEVER's objective is to promote optimal management of the power grids in the future energy system based on renewable sources. The dataset is not publicly available.

2.2. Forecasting algorithms

The study covers the comparison between a ML algorithm, RF, and DL algorithms, CNN and LSTM, when predicting energy consumption in an electrical grid.

The ML algorithm, RF, can be used to solve both classification and regression assignments. In this work, a regression assignment (e.g. forecasting substation energy consumption), the final output of a RF model is the mean of the outputs from all trees. The selection of hyperparameters and the input features are important factors to increase forecasting performance. Selection process for both was performed in this study.

Table 1. Description of substations consumption in kWh

Substation	mean	std	min	max	1st Qu.	2nd Qu.	3rd Qu.
1	35.640	20.454	0.000	100.0410	19.045	26.469	52.778
2	75.200	20.3084	0.000	159.8430	59.253	74.895	89.750
3	33.269	11.330	0.000	97.5440	24.305	31.758	39.950
4	18.462	8.209	0.000	62.832	11.730	16.838	23.632
5	35.317	20.121	0.000	121.086	18.796	29.554	48.455
6	13.163	7.596	0.000	50.326	7.069	10.345	18.297
7	34.622	14.550	0.000	98.100	22.66	31.523	44.605
8	3.690	3.290	0.000	24.613	0.000	3.247	5.606

With respect to the DL methods, we focus our study on the use of two DL methods, namely: CNN and LSTM. These methods are known for their ability to extract characteristics and learn patterns, and they were considered ideal candidates for forecasting electricity consumption.

CNN is a popular technique for forecasting time-series. CNNs are initially known for their efficiency on image processing tasks. Thus, in the same way as for an image, we consider time-series as a 1-dimension input vector that can be read and processed by the CNN model. CNNs have the ability to filter noise and extract local features from temporal data, such as seasonal patterns or trends. They also generally enable faster training compared to other models using MLP, RNN, or Attention modules, by reducing the number of parameters [18].

RNNs process a time-series step by step, maintaining an internal state from one time step to the next. The main idea of RNNs is to keep a memory of the past in hidden layers. In recurrent models, the previous state of a hidden layer influence the current state which is not the case in feed-forward neural networks. To avoid vanishing gradient and to enable long-term dependencies, we use a special RNN technology: LSTM. Due to their architecture composed of forget gates and input gates, LSTM are especially well suited for time-series forecasting. However, these models generally require a large amount of data in order to not overfit and also extract relevant features.

Table 2 shows the architectures of the DL models that were used to obtain the results presented in the next section.

Table 2. Deep learning model settings

Task	Model	Specific Layers	Prediction Layers
Real time	CNN	64 filters, kernel size = 2	Dense 8 units ; Dense 1 unit
Real time	LSTM	128 LSTM units	Dense 100 units ; Dense 1 unit
Non-real time	CNN	256 filters, kernel size=2 512 filters, kernel size=2	Dense 200 units ; Dense 24 unit
Non-real time	LSTM	1024 LSTM units	Dense 400 units ; Dense 24 unit

Lastly, the results are evaluated based on Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). MAE is generally used for regression problems and measures errors between paired observations expressing the same phenomenon while MAPE is used due to its very intuitive interpretation in terms of relative error. To find optimum parameters of the NNs, the optimization algorithm used was Adaptive Moment Estimation (ADAM) as in [22].

2.3. Feature selection

To improve the performance of algorithms, it is essential to extract relevant characteristics. We used temporal features such as hour of the day, since the frequency of data is hourly, and day of the week. However, it was found that the standard approach with discrete values (a value from 1 to 24 for the hours of the day) is not optimal for training models. For this reason, a sine/cosine transformation to these characteristics was applied to enhance the performance of the forecasting models [23]. We then obtain two periodic and continuous functions for the hour of the day and also two functions for the day of the week.

To further improve the results of the models we extract characteristics related to the seasonality and modes of the data. To do this, three functions were used. The first is obtained by averaging consumption over the last 7 days at the time we wish to predict. This highlights the recent trend at a specific time of day. Then, the second characteristic corresponds to the trend in consumption over the medium term. The value of this characteristic on day n at time t will be equal to the difference between the consumption value at time t on day $n\text{-}1$ and the consumption value at time $t\text{-}1$ on day $n\text{-}1$. This allows the model to see more clearly the trend in consumption on the previous day at the same time. The last characteristic corresponds to the very short-term trend, and is an average of the last five consumption values. To select these characteristics, we test their importance by carrying out a mixture test. To do so, we calculate the accuracy of our model with a characteristic on a test sample. Then, we recalculate this accuracy after mixing the values of this characteristic. By taking the difference between the two Mean Squared Error (MSE) values, we obtain an importance value for the characteristic. The more negative the value, the more the model has been degraded by this mixture, which means that the feature is important. A value close to 0 indicates that the model does not use this characteristic and that it is therefore useless. Finally, a positive value indicates that the model has obtained better results with the mixed characteristic, this feature should therefore be removed. This is how the three features were selected.

2.4. Comparative analysis settings

Aggregated consumption data at the substation level with an hourly frequency is analyzed in this paper. There are two very different but clear scenarios:

- *Non real-time scenario day ahead*: This is the case we worked with in the mentioned FEVER project. We received substation consumption data on a daily basis excluding data of the previous day due to pilot's constraints. This non real-time forecasting included prediction values for the whole entire next day.
- *Real-time scenario hour ahead*: This is the ideal case where we have substation consumption data on a real-time basis and we forecast the next hour using the past week of consumption.

We are analyzing the impact on performance on the forecast models between these two scenarios using RF, CNN and LSTM models.

3. Results and comparative analysis

3.1. Analysis of substation level data

Figure 1 shows consumption for substation 1 grouped by day of the week. In this substation we appreciate distinct consumption patterns depending on the day of the week. However, in other substations this pattern is not so clear (Figure 2).

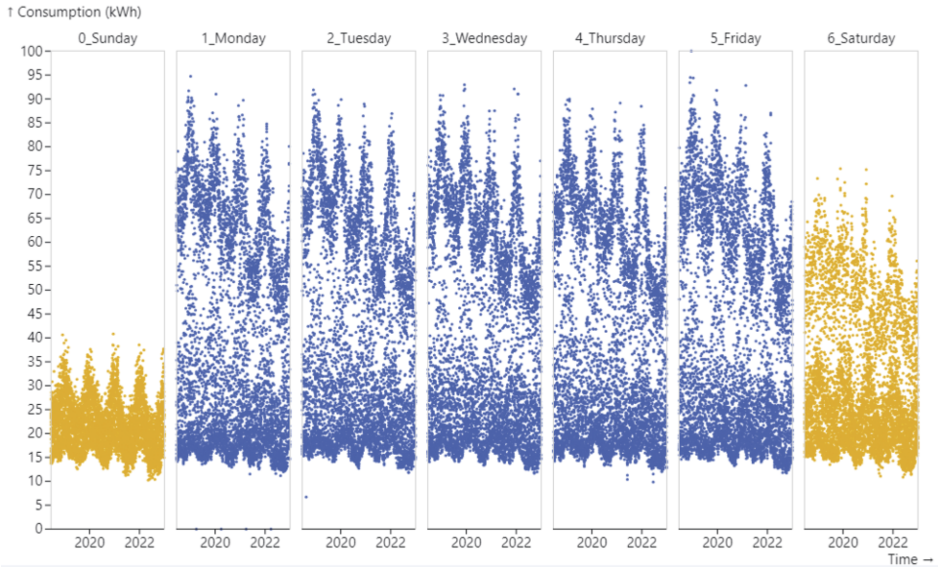

Figure 1. Substation 1: Consumption by day of the week

Furthermore, it is interesting to see how consumption patterns change each month throughout the year (Figure 3). We see how the consumption is lower during the warmer months of the year, this could be attributed to a lower use of Heating, Ventilation and Air Conditioning (HVAC) systems.

3.2. Results

The results obtained with the different forecasting methods are grouped in Table 3. We can see the accuracy values of the RF, CNN and LSTM models according to the MAE and MAPE for our two case studies: real-time and non-real-time. The values in bold represent the most accurate method for each substation.

It can be inferred that in real-time forecasting, the consumption observed immediately before the forecast will express the future consumption accurately. Thus, real-time scenario is going to have more accurate forecasting than non real-time forecasting.

Results obtained with the three methods are fairly similar in terms of MAE and MAPE in both scenarios. Of special consideration is substation 5 showing higher forecasting errors than the rest. This can be due to the higher unpredictability of substation 5 when compared to other substations.

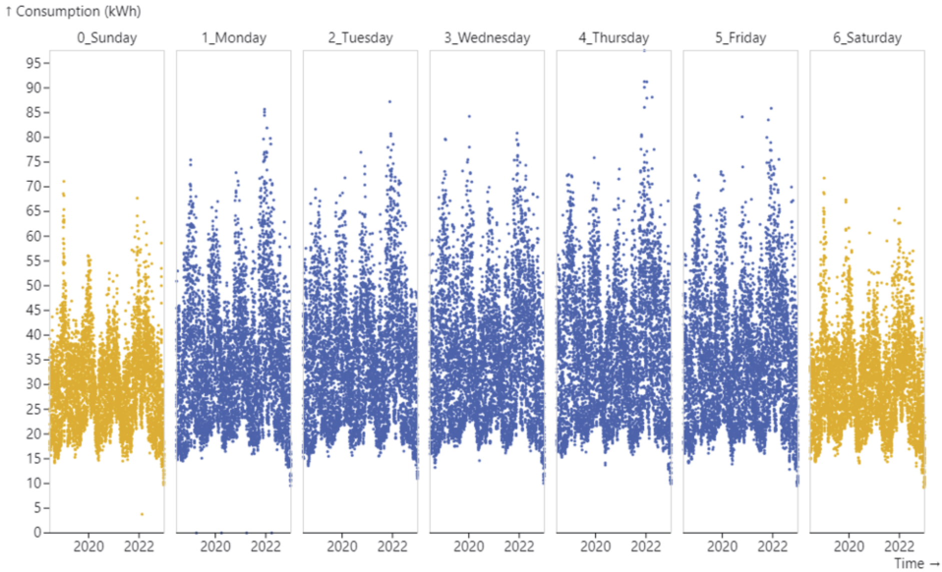

Figure 2. Substation 3: Consumption by day of the week

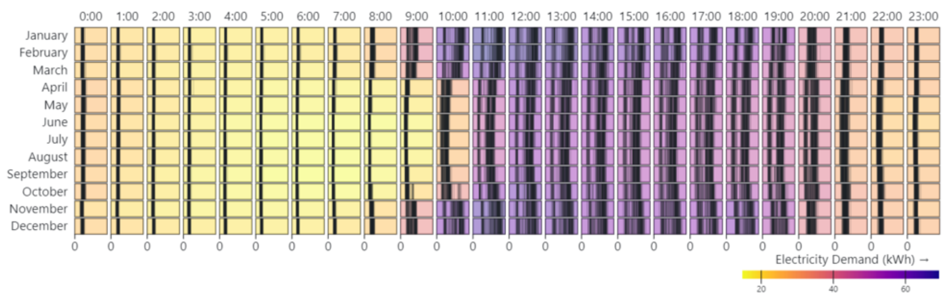

Figure 3. Substation 1: Hourly consumption by month

For the real-time scenario, the results are fairly similar between the three different methods. However, for all the substations except substation 5, the model using CNNs is more accurate than RF and LSTM. Substation 5 is the only substation obtaining better results with LSTM model, which is more complex and uses more parameters.

For non real-time prediction, there is an improvement in results with the use of DL, on all the nodes. DL models outperform traditional ML models especially on tasks that involve complex patterns and relationships. Thus, better results obtained in non real-time scenario by DL models can be attributed to this fact.

If we take substation 2, for example, with a MAE of 3.499 in the non real-time scenario, the model based on CNNs provides an improvement of over 20% compared with the RF model (MAE = 4.436). In the non real-time scenario it is harder to make accurate predictions, since we cannot use data from the previous day we want to forecast. This fact has increased the differences in results between the models in said scenario.

CNNs appear to be the most accurate model for substations 1, 2, 3, 4 and 7 in non real-time scenario. However, for substations 5, 6 and 8, LSTMs outperformed the other

Table 3. Comparative analysis between algorithms in both scenarios

Substation	Algorithm	Real-time		Non real-time	
		MAE	MAPE	MAE	MAPE
1	Random Forest	3.930	0.054	5.683	0.079
	CNN	**3.796**	**0.051**	**5.090**	**0.070**
	LSTM	3.829	0.052	5.372	0.074
2	Random Forest	2.316	0.076	4.436	0.150
	CNN	**2.200**	0.075	**3.499**	**0.123**
	LSTM	2.280	0.075	3.769	0.133
3	Random Forest	2.576	0.081	4.912	0.159
	CNN	**2.419**	**0.072**	**4.178**	**0.127**
	LSTM	2.55	0.077	4.549	0.142
4	Random Forest	2.116	0.112	3.812	0.210
	CNN	**2.110**	**0.111**	**3.240**	**0.184**
	LSTM	2.164	0.114	3.668	0.203
5	Random Forest	8.692	0.305	9.484	0.340
	CNN	8.460	0.290	9.523	0.340
	LSTM	**8.220**	**0.270**	**9.039**	**0.309**
6	Random Forest	1.524	0.117	2.375	0.190
	CNN	**1.520**	**0.116**	2.304	0.187
	LSTM	1.530	0.117	**2.216**	**0.181**
7	Random Forest	3.314	0.115	6.651	0.253
	CNN	**2.990**	**0.099**	**5.093**	**0.183**
	LSTM	3.130	0.108	5.478	0.200
8	Random Forest	0.807	0.168	1.182	0.263
	CNN	0.807	0.168	1.171	0.257
	LSTM	0.809	0.170	**1.127**	**0.254**

methods. When we compare the MAPE errors, we can observe that the models based on LSTMs seem to be better suited to the most difficult substations to predict. This is the case for substations 5 and 8, which have a large error on average that the rest of substations. CNN models seem to be better suited to substations that are simpler to predict. Finally, RF does not perform as well as the DL algorithms but it competes in the subsets of data with the least error.

3.3. Discussion

As we have seen in the literature review, DL methods have a number of drawbacks such as having many parameters and a high complexity. However, despite their drawbacks, they have the ability to capture the complicated interactions of substations load profiles.

Other studies using ML and DL algorithms, for example, [14] compared the performance of DL methods and SVM with respect to forecasting the energy consumption in institutional buildings. The experimental results indicate that the forecasting accuracy of DL methods is better than SVM. In our case study, results also indicate that the use of DL gives better forecasting accuracy compared to simpler ML methods.

In [24], the best performing method is not so clear but ANN has the least performance in terms of prediction accuracy, while the SVM method has a steady behaviour with low accuracy deviations. In our study, RF manages to achieve similar results to the DL methods in some particular substations, specially in the real-time scenario.

4. Conclusions

Accurate forecasting methods will be essential as DSOs need to ensure the security of grid supply in scenarios with higher RESs integration and electrification of heating and transport sectors. This paper presents a comparative analysis of electrical demand forecasting at DSO substation level between different machine learning methods. Two forecasting scenarios are studied depending on the availability of data: real-time hour ahead , which forecasts the next hour with the entire past week of consumption and non real-time day ahead, which forecasts the entire next day with the past week of consumption data excluding the previous day. Besides using consumption data, we propose using other features such as day of the week or month of the year, and using holidays information such as country festivities or local holidays.

The performance of the forecasts is extremely related to the characteristics of our data. However, DL methods achieve better results compared to simpler ML methods such as RF. Additionally, LSTM performs better on substations with higher MAE and MAPE values. Using the LSTM algorithm we managed to improve by approximately 3% the results of the forecast in the worst performing substation.

Real-time forecasting is between 3% and 8% more accurate than non real-time forecasting, depending on substation. DSOs ability to manage the grid is extremely dependant on the forecasting performance. Therefore, it is clear that DSOs ability to manage the electrical grid would improve with the transition from non real-time to real-time.

In addition, we believe that by adding extra information to the prediction models, such as outside temperature, could increase the overall accuracy achieved so far.

Acknowledgments

The FEVER project - Flexible Energy Production, Demand and Storage-based Virtual Power Plants for Electricity Markets and Resilient DSO Operation - is acknowledged by contributing with the data used in this work. FEVER was funded by the European Union (grant agreement N°864537).

References

[1] Commission E, Eurostat. Sustainable development in the European Union – Monitoring report on progress towards the SDGs in an EU context – 2023 edition. Publications Office of the European Union; 2022.

[2] Bongers T, Kellermann J, Franz M, Moser A. Impact of curtailment of renewable energy sources on high voltage network expansion planning. 2016:1-8.

[3] Guchhait R, Sarkar B. Increasing Growth of Renewable Energy: A State of Art. Energies. 2023;16(6). Available from: https://www.mdpi.com/1996-1073/16/6/2665.

[4] Hussain S, Lai C, Eicker U. Flexibility: Literature review on concepts, modeling, and provision method in smart grid. Sustainable Energy, Grids and Networks. 2023;35:101113. Available from: https://www.sciencedirect.com/science/article/pii/S2352467723001212.

[5] Villar J, Bessa R, Matos M. Flexibility products and markets: Literature review. Electric Power Systems Research. 2018;154:329-40. Available from: https://www.sciencedirect.com/science/article/pii/S0378779617303723.

[6] Khalil M, McGough AS, Pourmirza Z, Pazhoohesh M, Walker S. Machine Learning, Deep Learning and Statistical Analysis for forecasting building energy consumption — A systematic review. Engineering Applications of Artificial Intelligence. 2022;115:105287. Available from: https://www.sciencedirect.com/science/article/pii/S0952197622003372.

[7] Divina F, García Torres M, Goméz Vela FA, Vázquez Noguera JL. A Comparative Study of Time Series Forecasting Methods for Short Term Electric Energy Consumption Prediction in Smart Buildings. Energies. 2019;12(10). Available from: https://www.mdpi.com/1996-1073/12/10/1934.

[8] Massana J, Pous C, Burgas L, Melendez J, Colomer J. Short-term load forecasting in a non-residential building contrasting models and attributes. Energy and Buildings. 2015;92:322-30. Available from: https://www.sciencedirect.com/science/article/pii/S0378778815001024.

[9] Huang Y, Yuan Y, Chen H, Wang J, Guo Y, Ahmad T. A novel energy demand prediction strategy for residential buildings based on ensemble learning. Energy Procedia. 2019;158:3411-6. Innovative Solutions for Energy Transitions. Available from: https://www.sciencedirect.com/science/article/pii/S1876610219309804.

[10] Divina F, García Torres M, Goméz Vela FA, Vázquez Noguera JL. A Comparative Study of Time Series Forecasting Methods for Short Term Electric Energy Consumption Prediction in Smart Buildings. Energies. 2019;12(10). Available from: https://www.mdpi.com/1996-1073/12/10/1934.

[11] Jiang B, Cheng Z, Hao Q, Ma N. A Building Energy Consumption Prediction Method Based on Random Forest and ARMA. 2018:3550-5.

[12] Sülo I, Keskin SR, Dogan G, Brown T. Energy Efficient Smart Buildings: LSTM Neural Networks for Time Series Prediction. 2019:18-22.

[13] Marino DL, Amarasinghe K, Manic M. Building energy load forecasting using Deep Neural Networks. 2016:7046-51.

[14] Mlangeni S, Ezugwu AE, Chiroma H. Deep Learning Model for Forecasting Institutional Building Energy Consumption. 2020:1-8.

[15] Barja-Martinez S, Aragüés-Peñalba M, Íngrid Munné-Collado, Lloret-Gallego P, Bullich-Massagué E, Villafafila-Robles R. Artificial intelligence techniques for enabling Big Data services in distribution networks: A review. Renewable and Sustainable Energy Reviews. 2021;150:111459. Available from: https://www.sciencedirect.com/science/article/pii/S1364032121007413.

[16] Wei N, Li C, Peng X, Zeng F, Lu X. Conventional models and artificial intelligence-based models for energy consumption forecasting: A review. Journal of Petroleum Science and Engineering. 2019;181:106187. Available from: https://www.sciencedirect.com/science/article/pii/S092041051930600X.

[17] Fan C, Xiao L, Zhao Y. A short-term building cooling load prediction method using deep learning algorithms. Applied Energy. 2017 06;195:222-33.

[18] A review of deep learning for renewable energy forecasting. Energy Conversion and Management. 2019;198:111799. Available from: https://www.sciencedirect.com/science/article/pii/S0196890419307812.

[19] Wibawa A, Putra Utama AB, Elmunsyah H, Pujianto U, Dwiyanto F, Hernandez L. Time-series analysis with smoothed Convolutional Neural Network. Journal of Big Data. 2022 04;9:44.

[20] Lindemann B, Müller T, Vietz H, Jazdi N, Weyrich M. A survey on long short-term memory networks for time series prediction. Procedia CIRP. 2021 05;99:650-5.

[21] Ahmad T, Zhang H, Yan B. A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. Sustainable Cities and Society. 2020;55:102052. Available from: https://www.sciencedirect.com/science/article/pii/S2210670720300391.

[22] Sezer OB, Gudelek MU, Ozbayoglu AM. Financial time series forecasting with deep learning : A systematic literature review: 2005–2019. Applied Soft Computing. 2020;90:106181. Available from: https://www.sciencedirect.com/science/article/pii/S1568494620301216.

[23] Chakraborty D, Elzarka H. Advanced machine learning techniques for building performance simulation: a comparative analysis. Journal of Building Performance Simulation. 2018 07;12:1-15.

[24] Zeng A, Liu S, Yu Y. Comparative study of data driven methods in building electricity use prediction. Energy and Buildings. 2019;194:289-300. Available from: https://www.sciencedirect.com/science/article/pii/S0378778818303682.