2020 IEEE Electric Power and Energy Conference (EPEC)
2020 IEEE Electrical and Power Engineering Conference (EPEC), Edmonton, Canada

Electric Power Load Forecasting Based on Multivariate LSTM Neural Network Using Bayesian Optimization

Mohammad Munem*, T. M. Rubaith Bashar†, Mehedi Hasan Roni‡, Munem Shahriar§, Tasnim Binte Shawkat¶ and Habibur Rahaman‖
Department of Electrical & Computer Engineering
Rajshahi University of Engineering & Technology, Bangladesh*†‡§¶
Memorial University of Newfoundland, Canada‖
*munem.mh@gmail.com**, *rubaithbashar@gmail.com*†, *rony110465@gmail.com*‡, *munem140059@gmail.com*§,
m.shawkat.bd@ieee.org¶ and *mhrahaman@mun.ca*‖

2020 IEEE Electric Power and Energy Conference (EPEC) | 978-1-7281-6489-2/20/$31.00 ©2020 IEEE | DOI: 10.1109/EPEC48502.2020.9320123

***Abstract*—With rapid growth and development around the world, electricity consumption is increasing day by day. As the production and consumption of electricity is simultaneous, an electric power load forecasting technique with higher accuracy can play a pivotal role in a stable and effective power supply system. In this paper, a multivariate Bayesian optimization based Long short-term memory (LSTM) neural network is proposed to forecast the residential electric power load for the upcoming hour. Bayesian optimization algorithm is conducted to select the best-fitted hyperparameter values since deep learning networks are associated with different hyperparameters which play a vital role in the performance of a network architecture. Our proposed Bayesian optimized LSTM neural network has obtained almost perfect prediction performance and it surpasses the other established model such as convolutional neural network (CNN), artificial neural network (ANN) and support vector machine (SVM) where mean absolute error (MAE), root mean squared error (RMSE) and mean squared error (MSE) are found 0.39, 0.54 and 0.29 respectively for the individual household power consumption dataset.**

***Index Terms*—Electric power load, Deep learning , Long short-term memory, Bayesian optimization**

I. Introduction

In recent times, the consumption of electricity has been increased a lot around the world with the massive growth of population. It also impacts the social and economic growth directly and no house can be imagined without electricity. Electricity consumption is divided into three areas *i.e.* industrial, commercial and residential areas. Among them more than 65% of the total production is used by residential areas [1]. However, a large portion of the production gets dissipated because of inefficient management systems. Therefore, to minimize power blackout, an efficient management system is necessary. Electrical load prediction plays one of the major roles in the power system to improve the efficiency of the whole power system [2]. Moreover, the electrical energy has to be consumed at the same time it is produced, because of the internal nature and characteristics of electricity. Therefore, to manage a stable electrical power system, an effective electrical power load forecasting model is essential.

Electric power load varies with time and it is associated with multiple features [3]. Normally several features are fed into a model to predict electric power consumption [4]. Achieving better electrical energy consumption predictions using classical models are very difficult as there is a subject of uncertainty and irregular seasonal trends [5]. Electrical load forecasting is mainly divided into three types but some researchers have also divided into four types based on time interval [6]. The three categories basis on time intervals are:

1) Long term load forecasting (1 year to 10 years ahead)
2) Medium term load forecasting (1 month to 1 year ahead)
3) Short term load forecasting (1 minute to 1 week ahead)

An LSTM neural network model where the important hyperparameters are tuned by Bayesian optimization algorithm to predict hourly electric power load for the next hour is proposed by us.

II. Related Works

Many of the research on prediction of electrical energy consumption have been able to construct modestly accurate and efficient model architecture which can be divided into three types of model based on the solving category. These model categories are determined on type of solving techniques such as statistical, machine-learning based and deep learning based intrinsically focus on different parts of the prediction problem. Mainly machine learning models have performed the most effectively as K. P. Amber presented a multi linear regression model to forecast daily energy consumption through extracting five unique features and eliminating unnecessary features [7]. Y. Chen conducted a SVM based prediction of electrical load for residential buildings [8]. S. A. Bogomolov produced a load predicting model consisting of a random forest modeling technique for 1 week ahead consumption forecasting

[9]. H. Chen developed a power load prediction method using ANN for short term house-hold usage of power [10]. K. Amarasinghe attempted to predict house-hold load using the same dataset on a CNN model [11].

Daniel L. Marino constructed a standard LSTM model and a sequence to sequence LSTM model to predict the load capacity of a single house hold for a 1 hour period [12]. Adding to this RUNHAI JIAO et al. formulated a multi sequence LSTM technique to predict the electric load demand for non-residential areas [13]. Of all the neural networks LSTM is most powerful in recognising patterns within sequential temporal data. Hence W. Kong et al. devised an LSTM approach to predict the electric load demand of both residence and non-residential ie. industrial areas for a short period of time using clustering techniques [14]. They mostly concentrated on noisiness of the data and importance of the features to make a highly noise free dataset that is efficient.

Uniquely in temporal datasets LSTM model retain the previous information well due to the fact that it utilizes different cell structures to preserve the past data sequence. LSTM includes various parameters that influence different aspect of the model architecture. The predicted outputs massively depend on these parameter values which are tedious and difficult to be calculated by a human. Thus a computational algorithm is required which is filled in by Bayesian optimization algorithm. In literature, an LSTM model for short term power consumption forecasting where different important hyperparameters are picked by Bayesian optimization algorithm to find out more accurate performance with lower loss is proposed.

III. Data Description

The dataset used in this study is provided by Hebrail and Berard (https://archive.ics.uci.edu/ml/datasets/) in UCI Machine Learning Repository which contains 2,075,259 measurements for a residential house situated in France for the year of 2006 to 2010. The dataset holds minutely averaged measurements with 25,979 missing values which are removed in data preprocessing. New features such as day of the year, hour of the day and day of the week were added as these

TABLE I: Input Feature Range & Average Values

Input Feature	Range	Average
Global active power (kW)	0.12-6.5	1.09
Global reactive power (kW)	0-0.77	0.12
Voltage (V)	225.83-251.90	240
Global intensity (A)	0.5-28.38	4.6
Sub metering 1 (Wh)	0-48.37	1.12
Sub metering 2 (Wh)	0-46.43	1.29
Sub metering 3 (Wh)	0-21.55	6.45
Day of the year	1-365	–
Hour of the day	00:00-23:00	–
Day of the week	1-7	–

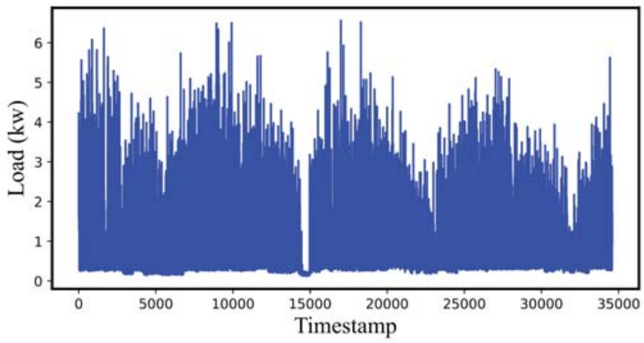

Fig. 1: Electric power load for hourly time resolution

new features play a vital role in predicting electric load due to human interactions in specific times. The whole dataset is resampled into hourly data with 10 features: day of the year, hour of the day, day of the week, global active power (GAP), global reactive power (GRP), voltage, global intensity (GI), sub metering 1, sub metering 2 and sub metering 3. TABLE I depicts the input features with ranges and average values. Sub metering 1, sub metering 2 and sub metering 3 readings are corresponding to electrical load of different equipment.

In this study, feature standardization is done as different input features have different ranges and the measuring units are also different which cause massive problems in different deep learning models. The input feature standardization is conducted using (1) equation:

$$Z = \frac{x - \mu}{\sigma} \tag{1}$$

Where μ denotes input feature mean , σ denotes input feature standard deviation and x denotes input feature value and Z denotes standardized value.

LSTM cells require the data to be sequentially fed to the hidden layers meaning the data has to be sequenced to a 4D array of size (1424,24,1,10). Here, 5-k cross validation method is used to divide the dataset into training and test or validation set. Fig. 1 represents electrical load with hourly time resolution for the entire dataset.

IV. Performance Evaluation Metrics

To compare the performance of different deep learning models on forecasting electrical load, different performance evaluation metrics are available. In this study, we have used mean squared error (MSE), mean absolute error (MAE) and root mean squared error (RMSE).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \tag{3}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{4}$$

Here, n is the total number of test observations, $\hat{y}_i$ denotes predicted value and y_i denotes the actual true value of the test data.

V. Methodology

In this section the overview of our proposed model architecture is stated. The main components of the proposed model are described through the long short term memory cell and Bayesian optimization algorithm. LSTM cells act like the model memory cells that memorize input sequence characteristics. Several combined LSTM cells can internalise a huge amount of sequential data. Bayesian optimization algorithm tunes the LSTM cell hyperparameter based on the Bayes theorem. In the subsection below the LSTM cell and the Bayesian optimization algorithms are further elaborated.

A. Long short-term Memory (LSTM)

LSTM is adequately employed deep learning neural network which retains information better than any other deep neural networks due its memory cell configuration. This type of configuration allows LSTM to evaluate past data efficiently on the other hand SVM performs analogous to linear regression technique with having disregard for retention of the past data. ANN model are feed forward models that propagate the information while having temporal information being lost fairly quickly. CNN, most prominent for image processing, handles data in dimensional fashion to extract the specific dimensional feature constituting good spatial feature extraction which is helpful for spatial data where as temporal data requires careful retention of past data . Thus LSTM is the most suitable candidate for the aforementioned task.

LSTM neural network architecture combines several LSTM cells to create a stable memory sequence which eliminate the problem of vanishing gradient. LSTM cells are composed of several state boxes acquiring inputs sequentially through time. For every time step the LSTM cell takes an input vector and the output of which is calculated as below:

$$h_t = f_w(h_{t-1}, x_t) \tag{5}$$

Where the x_t term represents the input vector, h_t and h_{t-1} represents state vector at time t and $(t-1)$ and f_w represents the non linear activation function having w as the weight parameters. Fig. 2 exhibits the common LSTM cell representation.

The mathematical expression of the LSTM cell displayed in Fig. 2 is stated in equation (6) to (13).

$$F(t) = \sigma\left(W_f \cdot [H_{t-1},\ X_t] + b_f\right) \tag{6}$$

$$I(t) = \sigma\left(W_i \cdot [H_{t-1},\ X_t] + b_i\right) \tag{7}$$

$$\widetilde{C}(t) = tanh\left(W_c \cdot [H_{t-1},\ X_t] + b_c\right) \tag{8}$$

$$C(t) = f_i * (C_{t-1} + I_t * \widetilde{C}_t)) \tag{9}$$

$$O(t) = \sigma\left(W_o \cdot [H_{t-1},\ X_t] + b_o\right) \tag{10}$$

$$H(t) = O_t * tanh(C_t)) \tag{11}$$

$$sigmoid\,(x) = \frac{1}{1+e^{-x}} \tag{12}$$

$$tanh\,(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{13}$$

Where X_t represents the sequential input, b_f, b_i, b_c and b_o denotes the bias weights, W_f, W_i, W_c and W_o depicts the input weights, t is the latest time step and $t-1$ is the previous time step; H denotes the output and C signifies the cell state.

B. Bayesian Optimization

Deep learning models like LSTM have several hyperparameters that are deeply depended on the dataset itself. Humanly ensuring a perfect fit for these hyperparameter is more like an art then an exact science. Deep learning models have several hyperparameter tuning techniques such as grid search, random search [15] etc. Bayesian optimization is a powerful strategy to find the extrema of a given objective function which estimates the objective function as a form of a Gaussian process and interprets it as a proxy function [16]. Bayesian optimization is most effective when the closed-form expression of the given objective function is not known but some observation can be derived from the objective function. In our proposed model Bayesian optimization is applied to find the optimum hyperparameters to locate the minima of the test or validation loss.

Here, P is the hyperparameter search space, where the model parameters are comprised like the number of hidden layer is denoted as N_a, dropout rate as N_b, batch size as N_c etc. Thus the objective function F can be expressed as

$$F : P\ (N_a, N_b, N_c, ..., N_n) \subset \mathbb{R}^n \to \mathbb{R} \tag{14}$$

To find the optimal model hyperparameter arrangement the search space can be defined as $p^* \in P$ such that

$$p^* = \arg\min_{p \in \mathbb{R}}\ F \tag{15}$$

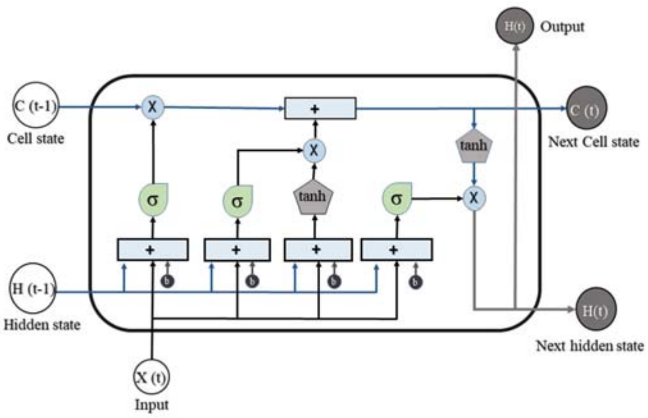

Fig. 2: LSTM cell architecture

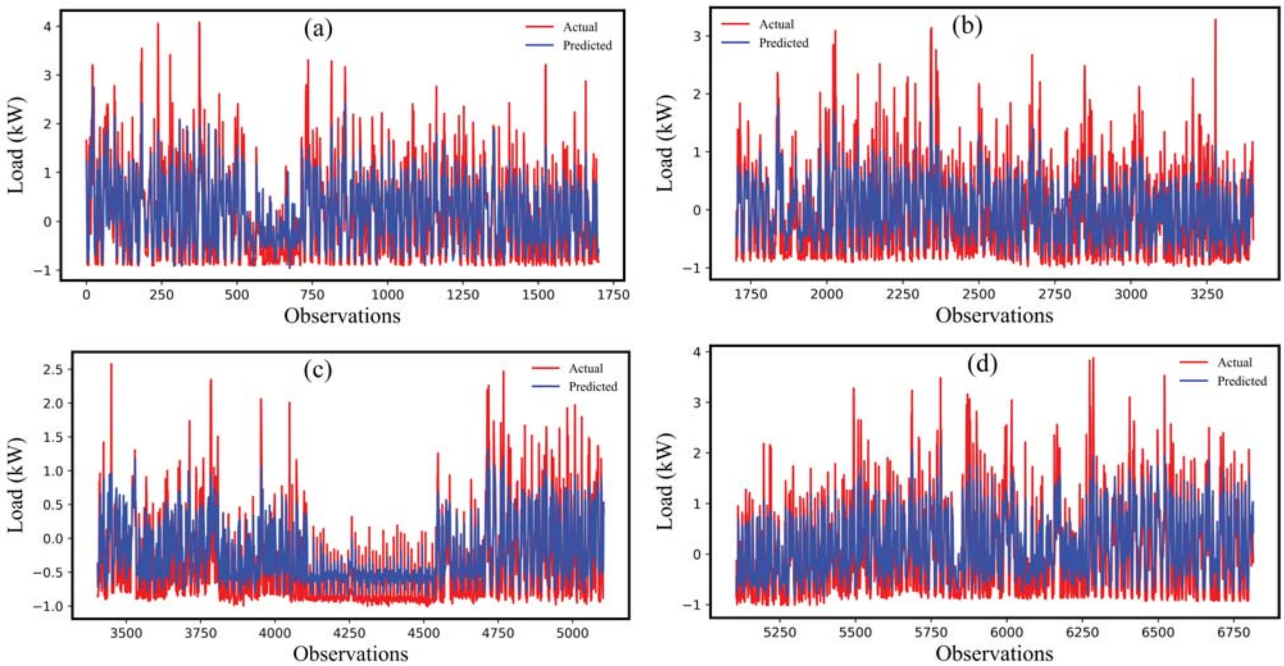

Fig. 3: Actual verses predicted electric power load (a)Partial result A (b)Partial result B (c)Partial result C (d)Partial result D

Here, the observations of the objective function can be expressed as $D_{1:n} = (p_{1:n}, F_{p_{1:n}})$ which helps Bayesian optimization to create a probabilistic model within $F(p)$ allowing the exploitation of the model to find out the next position in P in the sample.

Bayesian optimization computes the posterior distribution of the objective function with the help of Bayes' theorem where the next hyperparameter combination is chosen from this distribution. It uses the information of the previous sampling point to detect the shape of the objective function and finding the hyperparameters that maximizes the target output.

In our paper the model hyperparameters produce that objective function and the target output is to maximize the negative of the validation loss. Bayesian optimization is accomplished to select important LSTM hyperparamters values. TABLE II represents the hyperparamters and their search space.

TABLE II: LSTM Model Hyperparameter Search Space

Hyperparameters	Search Space
Number of LSTM cells	4-512
Activation function	ReLU, Linear, Sigmoid, Tanh, ELU
Optimization method	SGD, Adam, Nadam, Adamax, Adadelta, Adagrad, RMPSprop
Neurons in hidden layer	4-512
Dropout rate	0-0.8
Batch size	4-128

VI. Result

In this section the performance of the proposed model is evaluated and later it is compared with the performance of other existing deep learning models. All experiments are executed in Python with the help of *Scikit-Learn*, *Keras* and *Tensorflow* libraries.

The plotted RMSE graph of training set and validation set is a good measurement that determines the learning capability of a model. Fig. 5 depicts RMSE of training and validation set for 60 epochs. The training and validation RMSE loss reduce rapidly at first, then converge on a line on the horizontal axis which indicates that the model is learning from the training data. The small vertical gap between RMSE of the two datasets after 25 epoch proves that no over-fitting or under-fitting has occurred.

As shown in Fig. 4, Bayesian optimization is run for 60 calls to find out the best suitable hyperparameter values in

TABLE III: LSTM Model Hyperparameters Values Selected by Bayesian Optimization Algorithm

Hyperparameters	Selected Hyperparameters Values
Number of LSTM cells	360
Activation function	ReLU
Optimization method	Adamax
Neurons in hidden layer	102
Dropout rate	0.19
Batch size	57

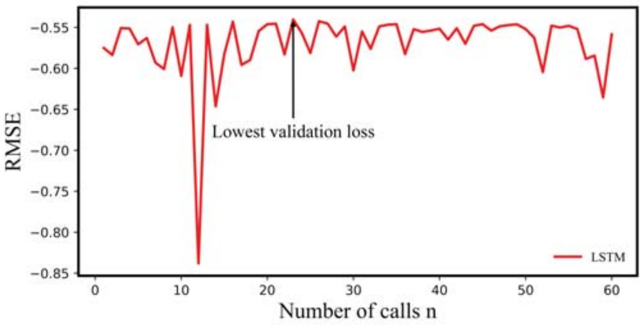

Fig. 4: The optimum configuration for LSTM model by Bayesian optimization

terms of RMSE. The best performance is found at 23 call and no extra effective improvement is found after that. TABLE III. demonstrates the selected hyperparameter values by applying Bayesian optimization algorithm on the 5-k cross validation set with 60 calls where the lowest RMSE is 0.54.

Fig 3 represents actual hourly load and predicted hourly load by our proposed model. It is justified from Fig 3 that our proposed model is able to capture the trend of electric power load of the particular house and shows great stability with MAE, RMSE and MSE of 0.39, 0.54 and 0.29 respectively.

In, TABLE IV MAE, MSE and RMSE of our proposed model is compared with other established deep learning models like CNN, ANN and SVM, where all important hyperparameters of these models are selected by Bayesian optimization algorithm accomplished with 60 calls. It is indicated from TABLE IV that, though our proposed model and CNN perform tremendously well, our proposed Bayesian optimization based LSTM model surpasses the other three established model in terms of MAE, MSE and RMSE. In MAE ranking the lowest to the highest corresponds to: LSTM (0.39), CNN (0.43), ANN (0.52) and SVM (0.68). Therefore, it is verified that though CNN and ANN model perform efficiently, our proposed Bayesian optimized LSTM neural network is the leading performer in terms of MAE, MSE and RMSE. Compared to previous works with the same dataset our

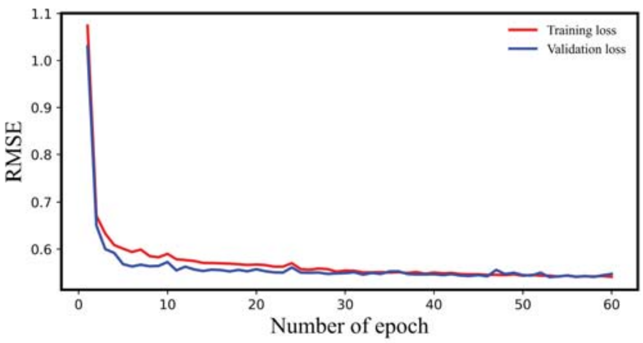

Fig. 5: Performance of LSTM model

TABLE IV: Model Performance Comparison of Deep Learning Methods in Terms of Error Metrics

Method	RMSE	MSE	MAE
SVM	0.92	0.85	0.68
ANN	0.66	0.44	0.52
CNN	0.58	0.34	0.43
LSTM	**0.54**	**0.29**	**0.39**

model produced overwhelmingly adequate results [12]. The compared model has one LSTM hidden layer with several units of LSTM cells where as our proposed model has one hidden layer with Bayes optimized LSTM hyperparameters to reduces the error to a significant amount. TABLE V compares the RMSE with other developed models with the same dataset.

VII. Conclusion

The proposed Bayesian optimization algorithm based multivariate LSTM neural network delivers robust and effective prediction of household electrical load for the next hour. The proposed model showed stable performance and ability to catch the irregular trends of electric power load for the individual residence. The introduced model is compared with other existing models and it surpasses these models regarding MAE, RMSE and MSE which indicates the usefulness and excellence of the introduced model. Bayesian optimization helps to reduce the initial manual validation loss by tuning the hyperparameters to an optimum level. The result of this paper shows that Bayesian optimization based multivariate LSTM neural network model offers the best performance to predict electric power load of the upcoming hour compared to the other established competitive benchmarks. Though the introduced model performs conspicuously, it can be improved by adding feature selection technique with the model.

In future we will increase the number of hidden layers to establish a hybrid model by combining with additional benchmark neural networks that further improves our current predictions.

TABLE V: Model Performance Comparison with other LSTM based models in Terms of Error Metrics

Model	RMSE (Test dataset)
LSTM (Laer:1, Unit:5)	0.640
LSTM (Laer:1, Unit:20)	0.657
LSTM (Laer:1, Unit:50)	0.686
LSTM (Laer:1, Unit:100)	0.729
Proposed LSTM	**0.54**

2020 IEEE Electrical and Power Engineering Conference (EPEC), Edmonton, Canada

References

[1] X. Masip-Bruin, E. Marín-Tordera, A. Jukan, and G.-J. Ren, "Managing resources continuity from the edge to the cloud: Architecture and performance," *Future Gener. Comput. Syst.*, vol. 79, pp. 777–785, 2018.

[2] S. Ruzic, A. Vučković, and N. Nikolić, "Weather sensitive method for short term load forecasting in electric power utility of serbia," *Power Systems, IEEE Transactions on*, vol. 18, pp. 1581 – 1586, 12 2003.

[3] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, "A review on time series forecasting techniques for building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 74, pp. 902 – 924, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1364032117303155

[4] N. Arghira, L. Hawarah, S. Ploix, and M. Jacomino, "Prediction of appliances energy use in smart homes," *Energy*, vol. 48, no. 1, pp. 128 – 134, 2012, 6th Dubrovnik Conference on Sustainable Development of Energy Water and Environmental Systems, SDEWES 2011. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544212002903

[5] G. Mentel, J. Brożyna, and B. Szetela, "Influence of double seasonality on economic forecasts on the example of energy demand," *Journal of International Studies*, vol. 9, 12 2016.

[6] S. Rahman and O. Hazim, "A generalized knowledge-based short-term load-forecasting technique," *IEEE Transactions on Power Systems*, vol. 8, no. 2, pp. 508–514, 1993.

[7] K. Amber, M. Aslam, and S. Hussain, "Electricity consumption forecasting models for administration buildings of the uk higher education sector," *Energy and Buildings*, vol. 90, pp. 127 – 136, 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0378778815000110

[8] Y. Chen, P. Xu, Y. Chu, W. Li, Y. Wu, L. Ni, Y. Bao, and K. Wang, "Short-term electrical load forecasting using the support vector regression (svr) model to calculate the demand response baseline for office buildings," *Applied Energy*, vol. 195, pp. 659–670, 06 2017.

[9] A. Bogomolov, B. Lepri, R. Larcher, F. Antonelli, F. Pianesi, and A. Pentland, "Energy consumption prediction using people dynamics derived from cellular network data," *EPJ Data Science*, vol. 5, 03 2016.

[10] Hong Chen, C. A. Canizares, and A. Singh, "Ann-based short-term load forecasting in electricity markets," in *2001 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.01CH37194)*, vol. 2, 2001, pp. 411–415 vol.2.

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

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

[13] R. Jiao, T. Zhang, Y. Jiang, and H. He, "Short-term non-residential load forecasting based on multiple sequences lstm recurrent neural network," *IEEE Access*, vol. 6, pp. 59 438–59 448, 2018.

[14] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on lstm recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851, 2019.

[15] J. S. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyper-parameter optimization," in *Advances in Neural Information Processing Systems 24*, J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2011, pp. 2546–2554. [Online]. Available: http://papers.nips.cc/paper/4443-algorithms-for-hyper-parameter-optimization.pdf

[16] M. Aoki, "On some convergence questions in bayesian optimization problems," *IEEE Transactions on Automatic Control*, vol. 10, no. 2, pp. 180–182, 1965.