Energy Proceedings
Vol 45, 2024

ISSN 2004-2965

# Multivariate Regression for Electricity Load Forecasting in Power Systems

E.R.Joy[1], R.C.Bansal [1,2*] , C. Ghenai[3,4], E.Gryazina[5], R.Kumar[6], Sujil.A[7],

1 Department of Electrical Engineering, University of Sharjah, United Arab Emirates

2 Department of Electrical, Electronic and Computer Engineering, University of Pretoria, Soth Africa

3 Sustainable Energy and Renewable Energy Engineering Department, University of Sharjah, University of Sharjah

4 Renewable Energy and Energy Efficiency Research Group, Sustainable Energy and Power System Research Centre, Research Institute of Science and Engineering, University of Sharjah, United Arab Emirates

5 Centre for Energy Science and Technology, Skoltech Institute of Science and Technology, Russia

6 Department of Electrical Engineering, Malaviya National Institute of Technology, Jaipur, India

7 Department of Electrical Engineering, Rajarambapu Institute of Technology, Maharashtra, India

(*Corresponding Author: rbansal@sharjah.ac.ae)

**ABSTRACT**

The development of smart grids in power system necessitates the need for forecasting the electricity load for the safe and economic functioning of electricity markets. A case study has been carried out considering a city's electricity load data using Multivariate Regression model. An input database of the model is generated taking into account of peak and off-peak hours based on maximum and minimum load data obtained from the utility operator. The characteristics of the electricity load over the whole year have been primarily analyzed to obtain a better intuition on the load behavior. In this context, the information in the form of temperature, days, different time duration i.e., peak and off-peak hours and past load data have been given as input to the regression model. The accuracy of the method has been evaluated using Root Mean Square Error (RMSE). The results of the adapted model have been compared with Neural Network, Ensemble and Kernel methods.

**Keywords:** electricity load, load forecasting, machine learning, multivariate regression, Sharjah, SEWA

## 1. INTRODUCTION

Sharjah is one of the populous cities in UAE and consumes the most electricity in the Emirate of Sharjah. The power systems in Sharjah have largely expanded to meet the escalating demand of electricity in the city [1], [2]. Sharjah Electricity, Water and Gas Authority (SEWA) provides electricity to the city with good quality power at reasonable rates [3]. Over the past few years, SEWA has experienced continuous increase in energy demand in Sharjah City [3]. However, the yearly electricity load pattern in Sharjah changes based on temperature variations, seasonal effects, peak/off-peak hours in a day etc. The electricity supplied to a city must be generated as per the demand in the city [4], [5]. Both over production and under production of electricity would cause financial loss to the operators [6]. One solution to overcome this problem is forecasting the load, which helps in predicting the amount of power required to meet the demand in the city [7], [8]. Weather parameters, peak/off-peak hours in a day, weekdays/weekend days, customer types, historical load data etc. are the prominent influencing factors that can affect the load pattern in the city. Accurate load forecasting would help the operators' devising policies for load management, scheduling generating units, predicting future demands, contract evaluations, tariff adjustments and prepares well for electricity market in advance [9], [10].

Several studies exist in literature for load forecasting based on physical models, statistical models and artificial intelligence models [11]-[15]. Among which artificial intelligence models have become more popular due to its capability to perform complex tasks. In [13], a deep learning algorithm based on long short-term memory

# This is a paper for 15th International Conference on Applied Energy (ICAE2023), Dec. 3-7, 2023, Doha, Qatar.

(LSTM) is used to forecast the actual load of a power plant in China. The numerical results obtained in this work have shown that the LSTM method has produced the prediction results with higher accuracy. Another study reported in [14], has proposed a deep belief network (DBN) to forecast the load of multiple low-voltage users. The verification results have shown a good accuracy when compared to LSTM model. A hybrid forecasting model based on discrete wavelet transform (DWT) and support vector regression (SVM) has been proposed in [15], to forecast the day ahead peak load of Greece power system. The load series in this work is first decomposed into low frequency and high frequency series and then SVR is applied to forecast the data. The method has shown a robust performance. Although many forecasting techniques based on intelligent techniques and hybrid methods exist in literature, these methods are very complex and it requires more training data to learn effectively.

The study presented in this article aims to develop a multivariate regression model for forecasting the electricity load considering the variables like temperature, days, peak/off-peak hours in a day and past load data as its features. Root Mean Square Error (RMSE) is used to validate the accuracy of the technique. The accuracy of the adapted method has been compared with neural networks, kernel and ensemble model.

The rest of the paper is organized in five different sections. Section II describes various components affecting the electricity load of Sharjah city. The methodology of multivariate regression technique, analytical expressions, evaluation process and error metrics are detailed in Section III. The details of data, performance evaluation and analysis are presented in Section IV. Finally, Section V summarizes the Conclusions of the study.

## 2. COMPONENTS AFFECTING ELECTRICITY LOAD

This Section has discussed various possible components that can affect the electricity load in a city over the year. In this study, electricity load of Sharjah city has been considered. The section has analyzed the characteristics of electricity load data of Sharjah City of the year 2021 [1]. The test data have been obtained from SEWA's statistical book published in SEWA website [1].

Figure 1, shows the Sharjah city's maximum and minimum load across the year. As shown in the Figure 1, the peak load has substantially increased between the months May to October due to hot weather conditions in UAE. UAE has a hot and desert climate. Seasonal change in Emirate is the well-known component and it largely affects the electricity load profile. The temperature in UAE is hot from the first week of April to the end of October. This period is extremely hot, and the electricity usage is more. The temperature is normal after October until the start of April. In Figure 1, the maximum and minimum load data has increased during summer and it has decreased during the winter.

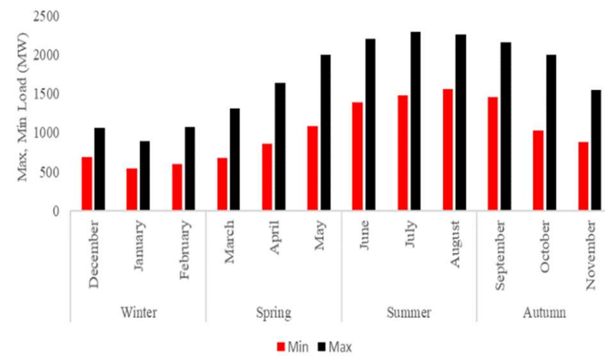


Fig. 1. Sharjah City's minimum and maximum loads in the year 2021

Figure 2 shows the electricity load variation from the month of December to November. As stated earlier, the load has become peak during summer. The test data shown in Figure 2 has considered 8 days in a month. For each day, the load data at certain time interval during peak and off-peak hours have been considered for analysis. The load variation is less than 1000MW at the start of the graph and then it increases around 2300MW during the peak summer.

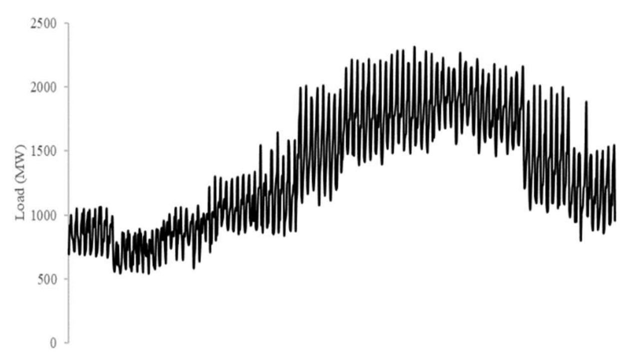


Fig. 2. Electricity load variation in a year

Figure 3 shows the relation between the load and temperature across the year. The temperature in UAE varies from 57°F to 106°F [16]. The temperature in summer varies between 86°F to 106°F. During this time, the climate is very hot outside. So, most of the consumers use electrical and electronic appliances inside home. This is a common load pattern in UAE as cooling demands are apparently larger in summer than those of other seasons. This phenomenon influences the load behavior.

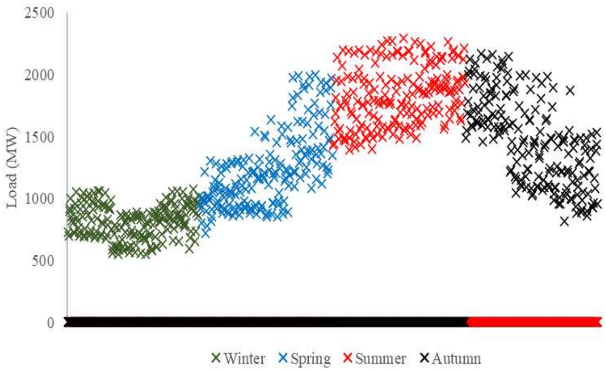


Fig. 3. Relation between load and temperature across the year

Apart from this, the repetitive events like peak hours, week days, weekends, holidays etc. can also affect the load data. Another key effect is the daily electricity demand in peak and off-peak hours in a day. An example of a different daily load patterns in the month of December is shown in Figure 4.

In this figure, during off-peak hours between 24.00hrs to 4.00hrs, the load consumption is less, whereas during peak hours between 19.00hrs to 22.00hrs the load consumption has gone to maximum.

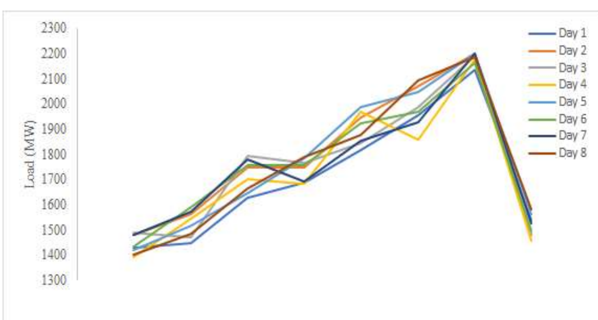


Fig. 4. Different loads patterns of 8 days in December month

From the above analysis, it has been observed the forecasting study must include important variables like temperature, days, hours in the form of quantitative variables. This study has considered 3 independent variables such as temperature, days and hours and one dependent variable load.

## 3. MULTIVARIATE REGRESSION IN LOAD FORECASTING

### *3.1 Analytical Expressions*

Based on the earlier analysis, the electricity load forecasting for Sharjah city load, should accommodate the important features for better prediction. The study has used multivariate regressions, which is a traditional method of forecasting for obtaining superior performance. The study has included multiple features like temperature, days, peak/off-peak hours and electricity load. The model has been represented using $h_\theta(x)$ as shown in (1).

$$h_\theta(x) = \theta_0 x_0^{(i)} + \theta_1 x_1^{(i)} + \theta_2 x_2^{(i)} + \theta_3 x_2^{(i)} \quad (1)$$

For convenience $x_o(i)$ is taken as 1, $x_1(i)$, $x_2(i)$, $x_3(i)$ denotes multiple features like temperature, days and hours, i=1...m is the training set. $\theta_0$, $\theta_1$, $\theta_2$, $\theta_3$ are the parameters of the equation. The parameter $\theta$ has been solved analytically, where $\theta$ is a n+1 dimensional vector $\theta \in \Re^{n+1}$. The parameter equation is given by (2) using mean squared error by taking the mean of the squares of the average difference between the predicted value and the actual value.

$$J(\theta_0, \theta_1, \dots \theta_n) = \frac{1}{2m} \sum_{n=1}^{m} ( h_\theta(x^{(i)}) - y^{(i)})^2 \quad (2)$$

Where, n is the number of features, $x^{(i)}$ is the index of the feature of $i^{th}$ data set. $x_i^{(i)}$ is the value of feature j in $i^{th}$ data set, $y^{(i)}$ is the predicted output and m is the total number of training set. The value of $\theta$ can be minimized by solving iteratively. $\theta_0, \theta_1, \dots\dots \theta_n$, can be solved for every j.

### *3.2 Performance Indicator*

In order to evaluate the accuracy of the predicted data set with different data distribution, the root mean square error (RMSE) is used as an error metrics to measure the predictive error. The RMSE is a measure of an absolute error and has the same unit as the predicted error. The RMSE is calculated at the end of each learning task, which is given in (3).

$$\text{RMSE} = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \left( y^{(i)} - \widehat{y^{(i)}} \right)^2} \quad (3)$$

where $y^{(i)}$ and $\widehat{y^{(i)}}$ are the actual and predicted power, m is the number of samples in the training set. The RMSE values shows how well the regression learning model learns the current task.

### *3.3 Evaluation Process*

The hypothesis $h_\theta(x)$ given in (1) is used to forecast the load variation in the work. The variation of load depends on multiple factors like temperature, days, and peak/off-peak hours. The feature vectors are n+1 dimensional vector and the parameters are another n+1 dimensional vector. The step-by-step evaluation process of multivariate regression has been illustrated in the flowchart shown in Figure 5. The procedure is briefly discussed in steps.

- Specify the parameters. Independent (temperature, days and hours) and dependent variables (load).
- Build the hypothesis $h_\theta(x)$ and cost function J ($\theta_0$), for every $x^{(i)}$ and $y^{(i)}$.
- Determine the optimal parameter $\theta$ values, such that J($\theta_0$) is minimum and $x^{(i)}$ is close the value of $y^{(i)}$.
- Solve the value of $\theta$ iteratively, for every j.
- Evaluate the accuracy of the error value using RMSE equation given in (3).

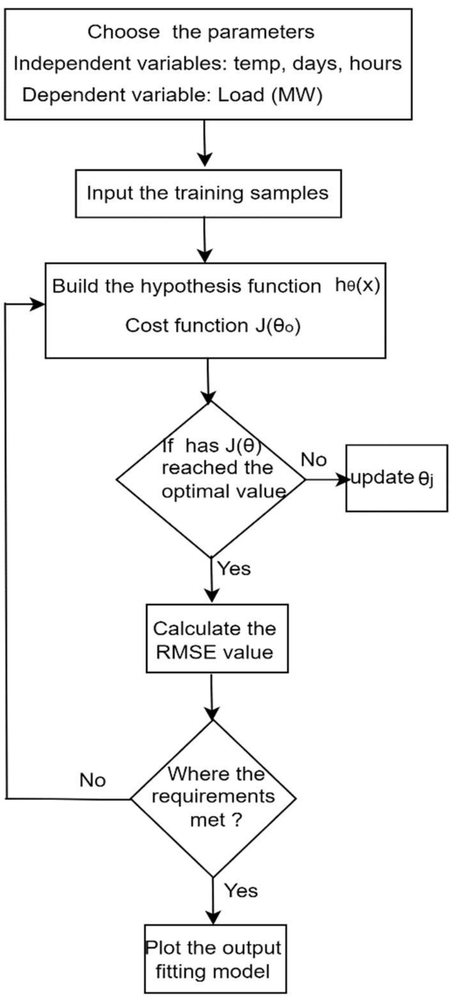


Fig. 5 Evaluation process flow chart

## 4. COMPONENTS AFFECTING ELECTRICITY LOAD

To comprehensively illustrate the Multivariate Regression technique, electricity load data of Sharjah City is used for analysis in this paper. Electricity load data of 12 months have been considered based on the information obtained from SEWA's annual statistical report [1]. The training data is formed in accordance with the maximum and minimum load data published in SEWA's annual statistical book, which is given in Table 1. The tested data set have chosen a total of 96 days load data in a year and each month 8 days are considered. The data set has 8 different peak and off-peak hours of a day. A total of 768 data set have been chosen from a pool of data available in a year. The training dataset have been examined based on the seasons starting from December.

The data have been partitioned as winter season (Dec. to Feb.), spring season (Mar.-May.), summer season (Jun.- Aug.) and autumn season (Sep.-Nov.). Extensive numerical cases with different seasons have been investigated based on the considered load data. The data has also been experimented with different benchmark models like Neural network, Ensemble and Kernel.

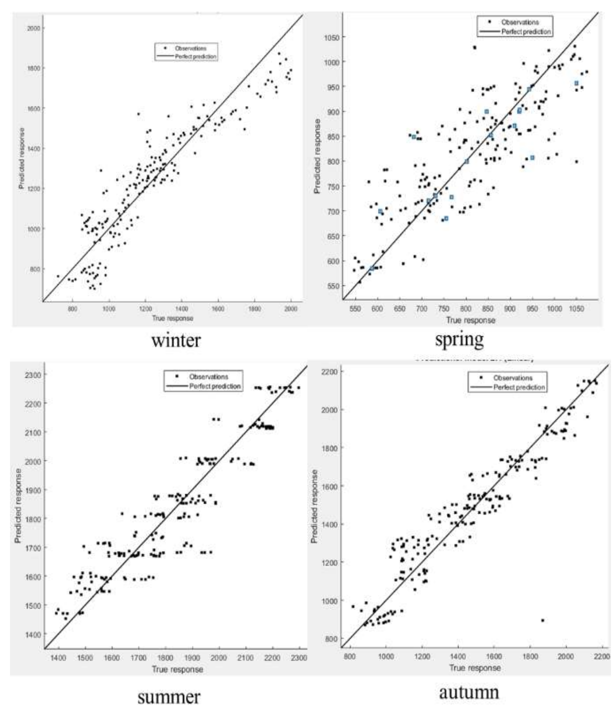


Fig. 6 Predicted value versus true response for four seasons

Figure 6 shows the variation of observation point and the perfect prediction for four different seasons. The line shown in the figure is the perfect prediction and the data points are observation points. A perfect prediction of any learning algorithm means the observation point should

Table I
Sharjah City's Maximum Loads in MW [1]

| *Seasons* | *Winter* | | | *Spring* | | | *Summer* | | | *Autumn* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***Months*** | *Dec* | *Jan* | *Feb* | *Mar* | *Apr* | *May* | *Jun* | *July* | *Aug* | *Sep* | *Oct* | *Nov* |
| ***Min*** | *691* | *543* | *597* | *671* | *853* | *1079* | *1392* | *1483* | *1561* | *1462* | *1025* | *883* |
| ***Max*** | *1062* | *889* | *1073* | *1313* | *1633* | *2000* | *2201* | *2297* | *2266* | *2159* | *1996* | *1548* |

lie very close to the prediction line. From the plot it has been observed several observation points lies very close to the prediction line. The distance from any point to the line should be less, that is the error of that particular observation point. The accuracy of the model is calculated by finding the RMSE value of the predicted data set.

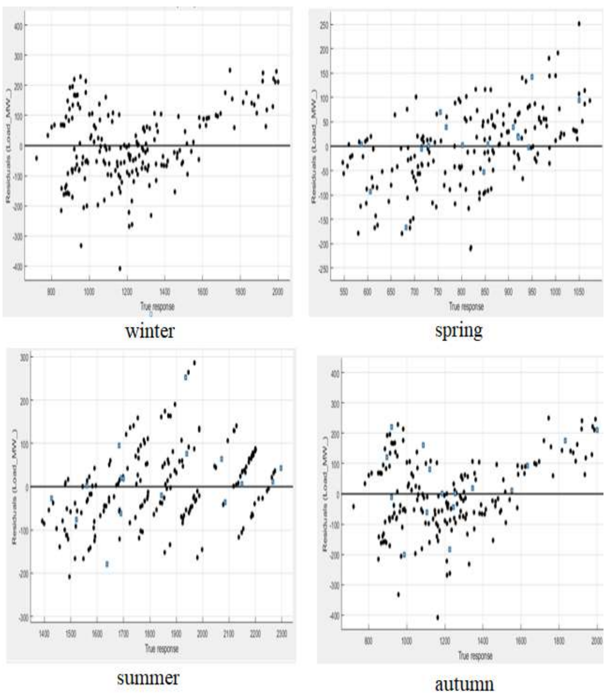


Fig. 7 Residuals versus true response for four seasons

Table II
Comparison of RMSE values

| | *RMSE* | | | |
|---|---|---|---|---|
| Models | Winter | Spring | Summer | Autumn |
| Multivariate Regressions | 77.954 | 118.44 | 85.735 | 118.87 |
| Neural Network | 84.458 | 74.198 | 94.85 | 109.03 |
| Ensemble | 80.81 | 144.52 | 118.13 | 159.53 |
| Kernel | 134.11 | 170.63 | 157.69 | 193.23 |

Figure 7 depicts the residual vs perfect prediction plot for regression models of four seasons. Most of the observation points are seen close to the prediction line. The points lying exactly on the line indicates the accurate prediction. In Figure 7, positive observation points on the y-axis indicates the prediction was too low and the negative observation points indicates, the prediction was too high.

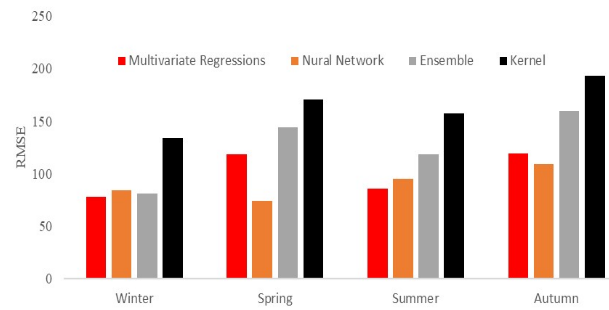


Fig. 8 RMSE comparison of different models

Furthermore, the RMSE value of the multivariate regression model are compared with the benchmark models such as Neural network, Ensemble and Kernel, which are given in Table II. The model used for comparison are Neural network, Ensemble and Kernel. The multivariate regression model has provided RMSE value of 77.954 for Winter, 118.44 for Spring, 85.735 for Summer and 118.87 for Autumn. It is clear that regression model has lower error in winter and summer, while neural network model has performed well for spring and autumn. Table II also indicates Ensemble and Kernel model have relatively higher forecasting errors than the other two models.

Figure 8 shows the RMSE comparison of 4 different models as presented in Table II. The bar chart shows the performance comparison of the machine learning models. Multivariate regression model shows better performance for winter and summer seasons, while for spring and autumn the neural network model has shown a good performance.

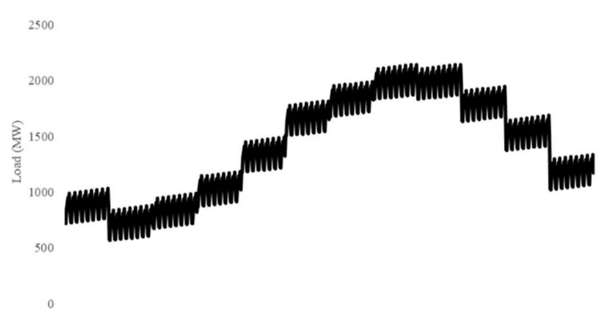


Fig. 9 Load variation predicted for another 1 year

Figure 9 shows the load predicted for next 1 year using Multivariate regression model. The figure has exhibited similar load pattern behavior when compared to the actual load behavior shown in Figure 2.

## 5. CONCLUSIONS

The paper has provided a forecasting study on Sharjah city's electricity load data using Multivariate Regression model. Various components influencing the electricity load including temperature, peak/off-peak hours, week days etc. has been analyzed to get a better insight on the change in performance of the load behavior. The mathematical formulations of multivariate regression model have been explained with the flowchart. The performance of the model is compared with the other state of the art models like Neural Networks, Ensemble and Kernel. The accuracy of the models has been evaluated using the RMSE value. The obtained results indicate multivariate regression model has provided a good accuracy when compared to Ensemble and Kernel methods. The prediction performance of the load can be further, if more number of features and datasets are used in the analysis improved in the future work.

## REFERENCE


[1] S. AlSuwaidi (2023, August 4) The Annual Statistical Book 2021. Retrieved from https://www.sewa.gov.ae/uploads/statspdfar/SEWAStatBook2021.pdf

[2] A. H. Siddique, H. K. Al Bloushi and L. A. Lamont, Distributed Generation and smart power grid UAE vision for 2030, *3rd IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe)*, Berlin, Germany, pp. 1-6, 2012.

[3] SEWA Marketing and branding Team 2020 (2023, August 4). Retrieved from https://www.sewa.gov.ae/uploads/pdfs/SEWAVISION2020.pdf

[4] R.C. Bansal and T.S. Bhatti, Small Signal Analysis of Isolated Hybrid Power Systems: Reactive Power and Frequency Control Analysis, *Alpha Science International*, Oxford, U.K.2008.

[5] A. Mnatsakanyan, A. H. AlMazrooqi, E. B. Muruaga and P. A. Banda, Electricity Market Structure with Individual Pricing Mechanism, *9th IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*, Charlotte, NC, USA, 2018, pp. 1-6, 2018.

[6] R. C. Bansal, Automatic reactive power control of autonomous hybrid power systems, *PhD thesis, IIT Delhi*, India, 2003.

[7] B. Neupane, K. S. Perera, Z. Aung and W. L. Woon, Artificial Neural Network-based electricity price forecasting for smart grid deployment, *International Conference on Computer Systems and Industrial Informatics*, Sharjah, United Arab Emirates, 2012, pp. 1-6, 2012.

[8] D. Masri, H. Zeineldin and W. L. Woon, Electricity price and demand forecasting under smart grid environment, *IEEE 15th International Conference on Environment and Electrical Engineering (EEEIC)*, Rome, Italy, 2015, pp. 1956-1960, 2015.

[9] D. Kumar, H.D. Mathur, S. Bhanot, and R.C. Bansal, Forecasting of Solar and Wind Power using LSTM RNN for load Frequency Control in Isolated Microgrid, *Int. Journal of Modelling and Simulation*, vol.41, no.4, 2021, 311-323.

[10] R.C. Bansal and J.C. Pandey, Load Forecasting using Artificial Intelligence Techniques: a Literature Survey, *Int. Journal of Computer Applications in Technology*, vol.22, no.2/3, pp.109-119, 2005.

[11] Y. Li, X. He and W. Zhang, Applications of adaptive CKF algorithm in short-term load forecasting of smart grid, *Proc. 33rd Chinese Control Conf.*, Nanjing, China, pp. 8145-8149, 2014.

[12] K.B. Song, Y.S. Baek, D. H. Hong and G. Jang, Short-term load forecasting for the holidays using fuzzy linear regression method, *IEEE Trans. Power Systems*, vol. 20, no. 1, pp. 96-101, Feb. 2005.

[13] C. Cui, M. He, F. Di, Y. Lu, Y. Dai and F. Lv, Research on Power Load Forecasting Method Based on LSTM Model, *IEEE 5th Information Technology and Mechatronics Engineering Conf. (ITOEC)*, Chongqing, China, pp. 1657-1660, 2020.

[14] Q. Cheng *et al.*, Short-Term Load Forecasting Method for Low Voltage Users Based on Deep Belief Neural Network, *IEEE Sustainable Power and Energy Conference (iSPEC)*, Chengdu, China, pp. 2564-2569, 2020.

[15] I. P. Panapakidis, G. C. Christoforidis, N. Asimopoulos and A. S. Dagoumas, combining wavelet transform and support vector regression model for day-ahead peak load forecasting in the Greek power system, *IEEE International Conf. Environment and Electrical Engineering, (EEEIC / I&CPS Europe)*, Milan, Italy, pp. 1-6, 2017.

[16] W. Spark (2023, August 4), The Climate and Average Weather Year-Round, Retrieved from https://weatherspark.com/y/105470/Average-Weather-in-Dubai-United-Arab-Emirates-Year-Round,