



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 2, February 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379

 9940 572 462

 6381 907 438

 ijircce@gmail.com

 www.ijircce.com

Electricity Consumption Forecasting through Machine Learning: A Research Investigation

Mrs.Pragathi M¹, Mrs.Khudaija Nazhath M R²

Assistant Professor, Akash Institute of Engineering and Technology Devanahalli Bengaluru, India

Assistant Professor, Akash Institute of Engineering and Technology Devanahalli Bengaluru, India

ABSTRACT: Electricity is crucial for the economy and an integral component of contemporary living. Electricity is utilised by a wide range of individuals and enterprises for many purposes, including everyday need. An effective forecasting model is crucial for optimising the power system's operations and planning due to the uncertainties surrounding consumers' electricity use. Enhancing the precision and reliability of electricity usage projections can enhance power planning and reduce daily maintenance costs. The primary objective of this project is to facilitate the planning of future power consumption and load to achieve equilibrium between power usage and generation. This will aid in decreasing operational expenses and minimising resource wastage. This project aims to develop a machine learning platform model capable of forecasting energy consumption. The two methods suggested for the prediction model are Linear Regression and k-Nearest Neighbour. The data will undergo analysis and pre-processing before being utilised for training and testing models.

KEYWORDS: Electricity, Power Usage, Linear Regression, and k-Nearest Neighbour.

I. INTRODUCTION

Electricity is a remarkable and crucial human invention. Despite being a ubiquitous energy source globally, humanity existed for centuries without it. At night, the world was quite black, illuminated only by a sparse scattering of candles. Although individuals have historically survived without it, the human species would certainly struggle to prosper without it. The creation of power enabled expansion and progress. Everything changed when the concept of generating energy and using it to provide life to the planet was introduced. Electricity not only powers our houses and facilitates daily activities like cooking and cleaning, but it also significantly benefits several industries, particularly the technology sector. Electricity is a cost-effective energy source that can be easily converted into many forms. It can be efficiently and rapidly distributed to numerous locations. This type of energy is easily manageable and monitorable. Therefore, it is the most widely used energy source. Electricity is utilised for various reasons such as lighting, heating, cooling, refrigeration, medical equipment, appliances, electronics, computers, public transit systems, and more. Today, it is difficult to envision a life devoid of power. Energy conservation refers to the reduction of energy usage. This can be achieved by reducing energy consumption or by utilising energy-efficient devices. Conserving energy is essential for sustainable development. Conserving energy is crucial for various reasons. In addition to reducing our reliance on finite energy sources such as fossil fuels, it also enables cost savings on energy expenses such as electricity bills and other related costs. Conserving energy can help reduce the pace of urban expansion in regions where natural resources such as oil or lithium are being extracted. Conserving energy and improving its efficiency reduces the emission of greenhouse gases into the Earth's atmosphere. Consumers value energy conservation for its cost-saving benefits and environmental protection. Saving energy leads to cost savings, resulting in reduced living expenses. It aids in saving money and safeguarding the environment by reducing resource use and carbon dioxide emissions.

II. ELECTRICITY CONSUMPTION

Energy consumption is a crucial component of energy systems. Following the energy crisis of the 1970s, individuals began to consider their energy consumption. The global energy consumption is increasing rapidly. Each country attempts to minimise energy consumption in several sectors, including buildings, farms, industry, and cars. Energy is

derived from three distinct sources: fossil fuels, renewable resources, and nuclear energy. Monitoring the utilisation of each energy type in various locations requires significant effort. By doing this, we can calculate the energy consumption in various locations and develop customised programmes accordingly. In 2020, India's energy consumption per capita was approximately 0.6 toe (Tonne of Oil Equivalent), which is roughly half of the energy consumption per capita in the rest of Asia. Each individual consumed 940 kWh of electricity in 2020, around one-third of the regional average for Asia. In 2020, the total energy consumption decreased by 5.6% to 885 Mtoe because to the Covid-19 crisis, following a rapid increase of 4% per year from 2010 to 2019.

Coal accounted for 44% of the country's energy in 2020. 24% comes from oil and 6% comes from biomass, totaling 30%. Natural gas accounts for 6 percent, while hydroelectric, nuclear, solar, and wind power collectively account for the remaining 4 percent.

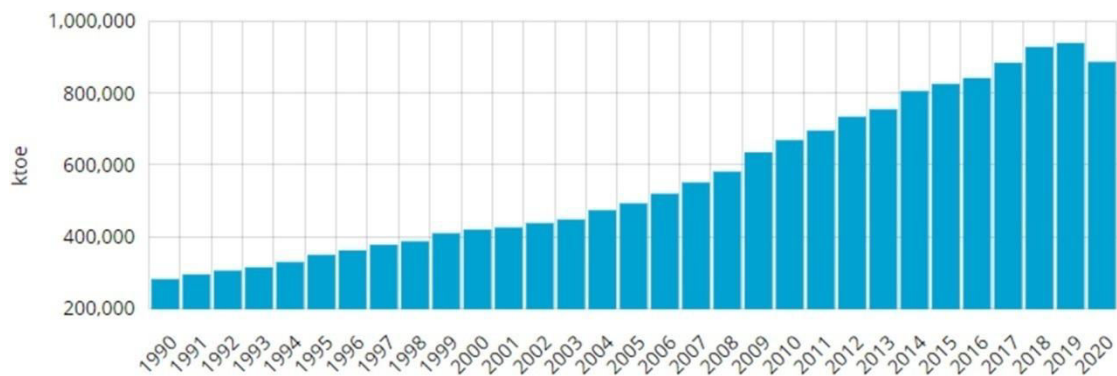


Figure 1.2.1: Total Consumption of Electricity in India over last 20 years

The figure 1.2.1 describes the consumption of Electricity in India over last 20 years

The data used in this project is collected from KPTCL Hadadi Road, Davanagere spanning over the years 2017 to 2022.

III. LITERATURE REVIEW

Energy forecasting can be classified into three main categories: short-term, medium-term, and long-term. Long-term forecasting is difficult for academics, especially when working with one-minute or five-minute incremental data. The study provided focuses on forecasting energy levels for short, medium, and long-term durations. Energy load forecasting can be achieved by physics-based models or statistical/machine learning techniques. Physics-based models rely on engineering principles that incorporate a complex interplay of material, structural, and geometric characteristics. Statistical models utilise machine learning algorithms to examine historical energy consumption data and provide mathematical representations of the relationship between the data and factors that impact energy usage.

This section emphasises statistical methods as our work aligns with that classification. Traditional machine learning methods such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVM) have been used to forecast energy use. Jetcheva et al. [17] presented an artificial neural network (ANN) model to forecast day-ahead energy use at the building level, utilising an ensemble approach to select model parameters. Artificial Neural Networks (ANNs) have been extensively researched for general load forecasting over short, medium, and long forecasting timeframes in numerous academic studies [18] [19].

Naji et al. [20] employed an Extreme Learning Machine technique to predict building energy usage using information on building material thickness and thermal insulation capacity. Several research have proposed using artificial neural network (ANN) and support vector machine (SVM) models to forecast energy use and assessed their effectiveness. Convolutional Neural Networks (CNN) have been utilised in load forecasting and have shown better performance than SVM models, achieving results comparable to Artificial Neural Networks (ANN) and other sophisticated deep learning techniques. Mocanu et al.'s work shown that the newly developed stochastic models, Factored Conditional Restricted

Boltzmann Machine, outperformed ANN, SVM, and traditional RNNs in short-term prediction. The work emphasises S2S GRU and LSTM models, which differ from earlier research on load forecasting. S2S models offer a more reliable analysis for time series problems by conveying their underlying hidden state through a directed graph over a sequence. Sequence-to-sequence (S2S) models outperform traditional artificial neural networks (ANNs), support vector machines (SVMs), and convolutional neural networks (CNNs) in maintaining the integrity of sequential input. LSTM and GRU recurrent neural networks are becoming increasingly popular for time series regression applications. Malhotra et al. (2015) used traditional Long Short-Term Memory (LSTM) networks to detect irregularities in electricity consumption. Bouktif et al. [26] used a standard LSTM model along with a genetic algorithm to predict short to medium term aggregate load. Both studies employed LSTM models for load forecasting, focusing on traditional LSTM prediction rather than S2S prediction, which is the main focus of our research. In typical LSTM models, the Encoder's outputs are used as predictions, while S2S models combine both the Encoder and Decoder to use the sequential outputs from the Decoder as predictions. Our work closely resembles the research carried out by Marino et al. [15]. They used classic LSTM and LSTM-based S2S models to forecast energy consumption at the household level utilising datasets with one-hour and one-minute intervals. The S2S model exhibited robust performance on both datasets and produced outcomes comparable to prior deep learning techniques [16]. We used S2S models in our work, however we varied in approach, sample creation, and the duration of the prediction sequence. The study does not include the method of using the previous Decoder output as the next input for their LSTM-S2S model. This paper introduces a novel approach for building-level load forecasting by combining sample creation with the S2S algorithm.

Energy forecasting can be classified into three main categories: short-term, medium-term, and long-term. Long-term forecasting is difficult for academics, especially when working with one-minute or five-minute incremental data. The study provided focuses on forecasting energy levels for short, medium, and long-term durations.

Energy load forecasting can be achieved by physics-based models or statistical/machine learning techniques. Physics-based models rely on engineering principles that incorporate a complex interplay of material, structural, and geometric characteristics. Statistical models utilise machine learning algorithms to examine historical energy consumption data and establish mathematical representations of the relationship between the data and factors that impact energy usage. This section emphasises statistical methods as our work aligns with that classification.

Traditional machine learning methods such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVM) have been used to forecast energy use. Jetcheva et al. [17] presented an artificial neural network (ANN) model to forecast day-ahead energy use in buildings. They employed an ensemble method to select model parameters. Artificial neural networks (ANNs) have been extensively researched for general load forecasting over short, medium, and long forecasting timeframes in numerous academic studies [18] [19]. Naji et al. [20] employed an Extreme Learning Machine technique to predict building energy usage using information on building material thickness and thermal insulation capacity. Several research have proposed using artificial neural network (ANN) and support vector machine (SVM) models to forecast energy use and assessed their effectiveness. Convolutional Neural Networks (CNN) have been utilised in load forecasting and have shown better performance than SVM models, achieving results comparable to Artificial Neural Networks (ANN) and other sophisticated deep learning methods. Mocanu et al.'s work shown that the newly developed stochastic models, Factored Conditional Restricted Boltzmann Machine, outperformed ANN, SVM, and traditional RNNs in short-term prediction. The work emphasises S2S GRU and LSTM models, which differ from earlier research on load forecasting.

S2S models offer a more reliable analysis for time series problems by conveying their underlying hidden state through a directed graph over a sequence. Sequence-to-sequence (S2S) models outperform traditional artificial neural networks (ANNs), support vector machines (SVMs), and convolutional neural networks (CNNs) in maintaining the integrity of sequential input.

LSTM and GRU recurrent neural networks are becoming increasingly popular for time series regression challenges. Malhotra et al. [25] used traditional LSTM networks to detect anomalies in electricity demand. Bouktif et al. [26] used a standard LSTM model along with a genetic algorithm to predict short to medium term aggregate load. The two studies used LSTM models for load forecasting, focusing on traditional LSTM prediction rather than S2S prediction, which is the main focus of our research.

In typical LSTM models, the Encoder's outputs are used as predictions, while S2S models combine both the Encoder and Decoder to use the sequential outputs from the Decoder as predictions.

Our work closely resembles the research carried out by Marino et al. [15]. They used classic LSTM and LSTM-based S2S models to forecast energy consumption at the household level utilising datasets with one-hour and one-minute intervals. The S2S model showed high performance on both datasets and produced results comparable to earlier deep learning techniques [16]. We used S2S models in our work, however we varied in approach, sample creation, and the duration of the prediction sequence. The study does not include the method of using the previous Decoder output as the next input in their LSTM-S2S model.

This paper introduces a novel approach for building-level load forecasting by combining sample creation with the S2S algorithm.

Literature Survey Summary

An analysis of energy consumption data was conducted to assess the current energy usage of the university campus. Next, the following events unfold: The air conditioning system on a college campus operates mostly according to the cooling and heating seasons and the outdoor air temperature. Cooling is active from mid-June to mid-September with a predetermined temperature of 27°C. Heating is operational from mid-November to mid-March, with the temperature maintained at 18°C. The Electric Heat Pump (EHP) technology is gradually supplanting traditional methods of cooling and heating.

A critical view was expressed on the utilisation of new and renewable energy due to its lack of profitability. Electricity's affordability in comparison to other energy sources results in its high specific gravity when used for educational purposes. Electricity consumption has been steadily increasing over an extended period. Electricity consumption is higher in the heating season compared to the cooling season. During vacations, electricity consumption decreases, whereas it peaks in March and September when schools are in session.

An investigation was conducted to examine the relationship between the number of members or the total gross floor area and the electricity consumption. Therefore, it became a mutual linear proportionate relationship. The average electricity consumption per unit area of the university campus is approximately 71.34 kWh/m² per year.

Existing System

They obtained power and data with the assistance of a remote monitoring device. Analysed the data to assess its predictive accuracy for monthly electricity expenses.

Electricity consumption was predicted using a convolutional neural network and a support vector machine approach. A support vector machine was utilised to forecast power consumption with an accuracy of 12%, known as the relative prediction error (RPE). The machine-learning models improved the RPE of electricity by 54% compared to the previous method using multiple linear regression.

Further investigation revealed that the support vector machine overestimated electricity use by 4% during the year (mean percentage error, MPE). From March to October, the actual electricity use exceeded the initial prediction by only 1%.

Problem Statement

The power system is experiencing increased demand and loads, particularly from non-linear loads, which are exceeding its design capacity and complicating the operation of the network. This compromises the system's safety due to insufficient regulation of high power currents. This may result in significant equipment loss and financial expenses. This project aims to develop a novel method for analysing and forecasting electricity use in a specific area over a defined period using historical data.

Proposed Objectives

This analysis is organized as follows:

- Get a public dataset of how much electricity is used in the area and clean it up by dealing with missing values and outliers.
- To see and understand how people use electricity over time.
- To create a training and testing model using Machine Learning Algorithms, Linear Regression, and k-Nearest Neighbor.
- To figure out what can be learned from how well the algorithms work.

Gaps Identified

- Using combinations of occupant characteristics instead of each one alone gives new information about how the occupant affects energy demand. Studying occupant traits instead of actual behaviour data lets you work with bigger data sets.
- Using a big set of data, the average real and theoretical energy use of different building types and features for each energy label was compared.

Methodology Presented

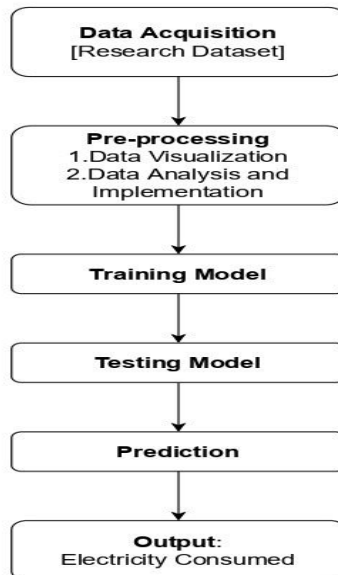


Figure: Methodology
The Figure describes the Methodology

Detailed Description of The Methodology

1. Dataset Acquisition:

Acquisition of datasets is the process of getting the datasets needed for the proposed project. Here, we are collecting data from the public that predicts how much electricity a certain area or building will use.

2. Pre-Processing:

During pre-processing, the data is displayed so that patterns, bad data, and outliers can be found and used to learn more about the dataset. Data pre-processing is needed to clean the data and make it fit for a machine learning model. It also makes a machine learning model more accurate and efficient.

- Dealing with null and missing values: We'll figure out the average of a column or row that's missing a value and put that number in its place. This strategy works well for features that have numbers, like age. Here, we'll use the Scikit-learn library, which has many libraries for building machine learning models.
- You can also deal with missing values by getting rid of the rows or columns that have null values. If more than half of the rows in a column are empty, the whole column can be dropped.
- Getting rid of "outliers," which are samples that are very different from the rest of the data. Putting the features or data points on a graph is the easiest way to find an outlier. One of the best and easiest ways to figure out what the overall data means and what the outliers are is to look at the data visually. Most of the time, scatter plots and box plots are the best ways to find outliers.
- Feature selection: The main goal of feature selection is to get rid of useless or duplicate data from the model.
- The entire dataset is divided into training data and testing data.

3. Training of the Model:

A training model is a set of data used to teach an ML algorithm how to do something. It is made up of a set of sample output data and the sets of input data that affect that output. The training model is used to run the input data through the algorithm to see if the output from the algorithm matches the output from the sample. The model is changed based on what this correlation tells us. Training a model is as easy as using labelled examples to find good values for all the weights and the bias. We will teach the model using data that has already been cleaned up.

4. Testing:

In machine learning, model testing is the process of judging how well a model works on a testing set after it has been fully trained. The testing set, which is made up of a set of testing samples, should be kept separate from both the training and validation sets. However, it should follow the same probability distribution as the training set. We will test the model with test data that has already been cleaned up and evaluate the model.

5. Calculation and Prediction:

The electricity use for each mode (Active, Sleep, and Off) is calculated separately, and then all three are added together with the power mode to get the total electricity use by mode. The UEC, or unit electricity consumption, is used to figure out how much electricity a device uses each year. The average consumption is then found, and the

electrical load is used to predict consumption over a period of time.

IV. EXPERIMENTAL RESULT AND DISCUSSION

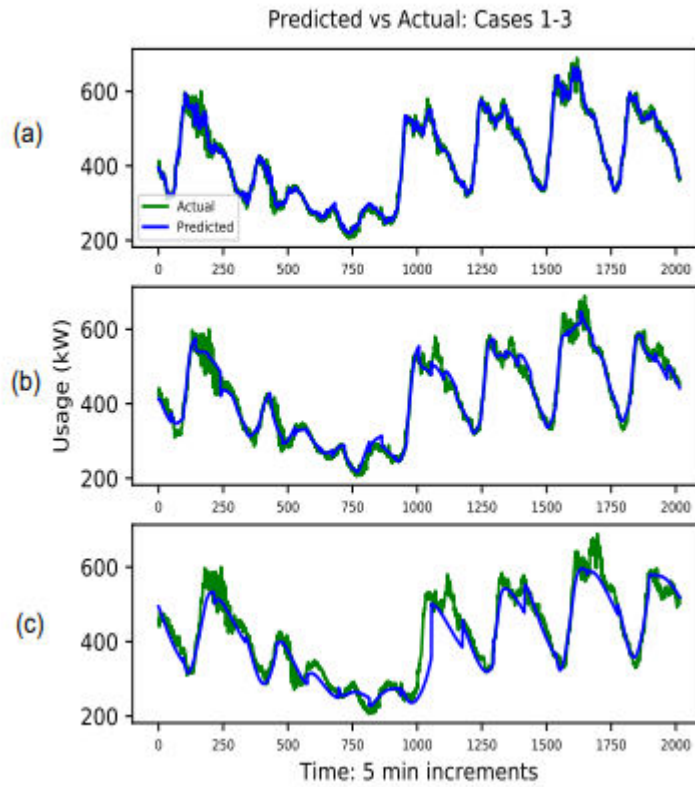


Figure . Predicted usage compared to actual, for the three cases: Hence, (a) represents case 1, (b) and (c) represent cases 2 and 3 respectively.

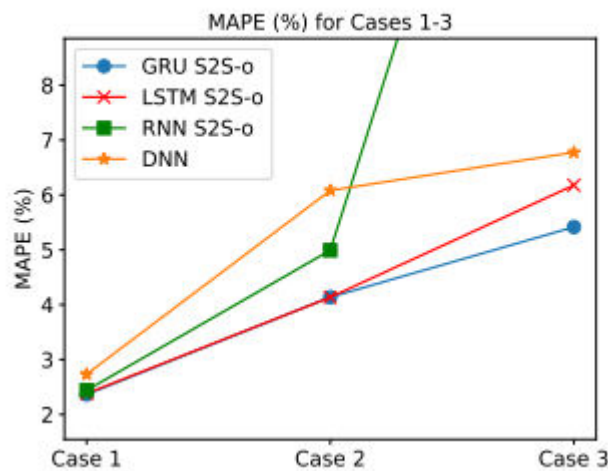


Figure . MAPE (%) for the three cases and for each model. confirmed by Chung et al. [31] and Jozefowicz et al. [32].

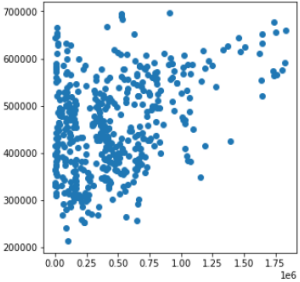
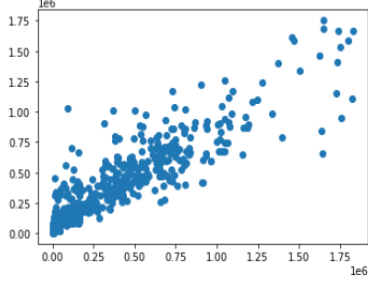
LINEAR REGRESSION ALGORITHM	K-NEAREST NEIGHBOUR ALGORITHM
<pre data-bbox="212 562 708 618"> IN [30]: PLT.SCATTER(Y_TEST, PREDICTIONS) Out[30]: <MATPLOTLIB.COLLECTIONS.PATHCOLLECTION AT 0x27FEEB1DE50> </pre> 	<pre data-bbox="818 562 1347 618"> IN [35]: PLT.SCATTER(Y_TEST, Y_PRED) Out[35]: <MATPLOTLIB.COLLECTIONS.PATHCOLLECTION AT 0x27FEF667F40> </pre> 

Figure: Comparison of Accuracy
The figure shows Comparison of Accuracy.

Based on how well each model did in different tests, the K-Nearest Neighbor Algorithm did the best and is good for predicting how much electricity will be used.

V. CONCLUSION

During the summer and winter months, people have the highest demand for power service. Belavanur, Shamnur, and Saraswathi are the areas that consume the most electricity, whereas Tolahunse is the area that consumes the least amount of electricity. With the assistance of the K-Nearest Neighbour algorithm, the prediction model performs with a higher degree of precision. In order to make the model even more precise and dependable, it could be beneficial to take into account the number of installations in a certain region as well as bank losses. Another way in which this could be improved is by obtaining data that is more trustworthy and up to date in order to make the model more accurate.

REFERENCES

[1] H. Ritchie and M. Roser, "Urbanization," 2018, Our World inData, Available: <https://ourworldindata.org/urbanization>. [Accessed: Jan. 2019].

[2] U.S. Energy Information Administration, "Use of Energy in the United States Explained," 2017, Available:<https://www.eia.gov/energyexplained>. [Accessed: Jan. 2019].

[3] H. C. Akdag and T. Beldek, "Waste Management in GreenBuilding Operations Using GSCM," International Journal of Supply Chain Management, Vol. 6 (3), pp. 174–180, 2017.

[4] Independent Electricity System Operator, "Globaladjustment and Peak Demand Factor," 2019, Available:<http://www.ieso.ca/Sector-Participants/Settlements/GlobalAdjustment-and-Peak-Demand-Factor>. [Accessed: Jan. 2019]

[5] J. Schmidhuber, "Deep learning in neural networks: an overview," Neural Networks, Vol. 61, pp. 85–117, 2015.

[6] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning representations by back-propagating errors," Nature, Vol. 323 (6088), pp. 533–536, 1986.

- [7] A. Graves, S. Fernandez, F. Gomez and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," In Proc. of the International Conference on Machine Learning (ICML), pp.369–376, 2006.
- [8] I. Sutskever, O. Vinyals and Q. V. Le, "Sequence to sequence learning with neural networks," In Proc. of Advances in Neural Information Processing Systems (NIPS), pp. 3104–3112, 2014.
- [9] G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition," IEEE Transaction on Audio, Speech, and Language Processing, Vol. 20 (1), pp. 30–42, 2012.
- [10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," In Proc. of Advances in Neural Information Processing Systems (NIPS), pp. 1097–1105, 2012.
- [11] P. J. Werbos, "Backpropagation through time: what it does and how to do it," Proceedings of the IEEE, Vol. 78 (10), pp.1550–1560, 1990.
- [12] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," In Proc. of the International Conference on Machine Learning (ICML), pp. 1310–1318, 2012.
- [13] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, Vol. 9 (8), pp. 1735–1780, 1997.
- [14] K. Cho et al., "Learning phrase representations using RNN encoder–decoder for statistical machine translation," In Proc. of the Empirical Methods in Natural Language Processing (EMNLP), pp. 1724–1734, 2014.
- [15] D. L. Marino, K. Amarasinghe and M. Manic, "Building energy load forecasting using deep neural networks," In Proc. of the IEEE Industrial Electronics Society (IECON), pp. 7046–7051, 2016.
- [16] E. Mocanu, P. H. Nguyen, M. Gibescu and W. L. Kling, "Deep learning for estimating building energy consumption," Sustainable Energy, Grids and Networks, Vol. 6 (1), pp. 91–99, 2016.
- [17] J. G. Jetcheva, M. Majidpour and W. Chen, "Neural network model ensembles for building-level electricity load forecasts," Energy and Buildings, Vol. 84 (1), pp. 214–223, 2014.
- [18] S. Ferlito et al., "Predictive models for building's energy consumption: an artificial neural network (ANN) approach," In Proc. of the Italian Association of Sensors and Microsystems (AISEM), pp. 1–4, 2015.
- [19] Y. T. Chae, R. Horesh, Y. Hwang, and Y. M. Lee, "Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings," Energy and Buildings, Vol. 111 (1), pp. 184–194, 2016.
- [20] S. Naji et al., "Estimating building energy consumption using extreme learning machine method," Energy, Vol. 97 (1), pp. 506–516, 2016.
- [21] D. B. Araya, K. Grolinger, H.F. ElYamany, M. Capretz and G. Bitsuamlak, "An ensemble learning framework for anomaly detection in building energy consumption," Energy and Buildings, Vol. 144 (1), pp. 191–206, 2017.
- [22] S. Seyedzadeh, F. P. Rahimian, I. Glesk and M. Roper, "Machine learning for estimation of building energy consumption and performance: a review," Visualization in Engineering, Vol. 6 (1), pp. 5–25, 2018.
- [23] N. L. Tasfi, W. A. Higashino, K. Grolinger and M. A. Capretz, "Deep neural networks with confidence sampling for electrical anomaly detection," In Proc. of the IEEE International Conference on Smart Data, pp. 1038–1045, 2017.
- [24] K. Amarasinghe, D. L. Marino and M. Manic, "Deep neural networks for energy load forecasting," In Proc. of the IEEE International Symposium on Industrial Electronics (ISIE), pp. 1483–1488, 2017.
- [25] P. Malhotra, L. Vig, G. Shroff and P. Agarwal, "Long short term memory networks for anomaly detection in time series," In Proc. of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN), pp. 89–95, 2015.
- [26] S. Bouktif, A. Fiaz, A. Ouni and M. A. Serhani, "Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: comparison with machine learning approaches," Energies, Vol. 11 (7), pp. 1636–1656, 2018.
- [27] Weather Underground, "Historical Weather," 2018, Available: <https://www.wunderground.com/history>. [Accessed: May 2018]
- [28] D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv preprint arXiv:1412.6980, 2014.
- [29] E. Hoffer, I. Hubara and D. Soudry, "Train longer, generalize better: closing the generalization gap in large batch training of neural networks," In Proc. of Advances in Neural Information Processing Systems (NIPS), pp. 1731–1741, 2017.
- [30] A. Paszke, S. Gross, S. Chintala, and G. Chanan, "PyTorch: from research to production," 2016. Available: <https://pytorch.org>. [Accessed: May, 2018]
- [31] J. Chung, C. Gulcehre, K. Cho and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," arXiv preprint arXiv:1412.3555, 2014.
- [32] R. Jozefowicz, W. Zaremba, and I. Sutskever, "An empirical exploration of recurrent network architectures," In Proc. of the International Conference on Machine Learning (ICML), pp. 2342–2350, 2015



INNO  **SPACE**
SJIF Scientific Journal Impact Factor
Impact Factor: 8.379



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 **9940 572 462**  **6381 907 438**  **ijircce@gmail.com**



www.ijircce.com

Scan to save the contact details