Journal of Artificial Intelligence and Capsule Networks (ISSN: 2582-2012)
www.irojournals.com/aicn/

# Design and Implementation of Short-Term Load Forecasting using LightGBM

**Surendar S.[1], Midhun Keshava Abimanyu M.E.[2], Sweetha M.[3], Syamala M.[4], Manikandan V.[5]**

[1-4]Students,[5]Professor, Department of EEE, Coimbatore Institute of Technology, Anna University, Coimbatore, India

**Email:** [1]surendar6347@gmail.com, [2]midhunkeshavaabimanyu@gmail.com, [3]sweetham2003@gmail.com, [4]syamalamanick20@gmail.com, [5]manikantan@cit.edu.in

## Abstract

In the modern world, energy efficiency and smart systems are more necessary than ever before, making it imperative to possess a system capable of monitoring and forecasting power consumption in real-time. This research presents an advanced Data Acquisition System (DAS) combining hardware sensors, cloud computing, and machine learning to deliver accurate power monitoring and forecasting. The system employs sensors connected to an Arduino and ESP8266, which wirelessly transmit voltage and current information to Firebase for processing and storage. A machine learning algorithm implemented in Python subsequently forecasts power demand, with the outcomes displayed on a user-friendly web dashboard developed with Flask. This dashboard updates dynamically, displaying real-time power information and visualizing predictions every five minutes. Through the implementation of REST APIs, the system ensures smooth and efficient data transmission without requiring local storage. This research provides a real-world and cost-efficient solution that can integrate energy management systems with intelligent, real-time knowledge for better decision-making.

**Keywords:** Real-Time Power Monitoring, Load Forecasting, Machine Learning, Energy Management, Cloud Database.


*DOI: https://doi.org/10.36548/jaicn.2025.1.004*
*Received: 26.02.2025, received in revised form: 25.03.2025, accepted: 10.04.2025, published: 21.04.2025*

## 1. Introduction

Several studies have contributed to advancements in load forecasting, influencing this research. Machine learning models like LightGBM and regression-based approaches have demonstrated effectiveness for short-term load forecasting in industrial and substation environments [1-2], aligning with this study's methodology. The role of predictive analytics in smart grids has been highlighted [3, 4], emphasizing the need for real-time forecasting solutions. Furthermore, multi-model forecasting approaches have been introduced [5], supporting the hybrid methodologies used for accuracy enhancement. Research on data acquisition systems [6, 7] has reinforced the necessity of cloud-based monitoring, similar to the research on Firebase-integrated DAS, while secure IoT-based data collection has been emphasized for ensuring reliability in real-time applications [8]. The explorations of time-series forecasting techniques enables to gain knowledge on the model selection [9, 10]. Finally, optimization strategies in smart grids have been discussed [11], aligning with the objective of scalable and efficient energy forecasting. From these studies, this research integrates a cloud-connected DAS with a LightGBM-based forecasting model to enhance real-time power monitoring and decision-making. The research demonstrates a cost-effective and scalable solution for energy monitoring and forecasting. With its smart design and real-time capabilities, it holds great potential for real-world applications and promotes better energy management.

### 1.1 Background of the Problem

Fast population growth, urbanization, and increased use of high-powered electrical equipment force the rising energy needs worldwide . The International Energy Agency (IEA) forecasts that world electricity consumption will increase roughly 5 percent yearly next ten years. More than 60% of nations as of 2023 noted grid inefficiencies and power imbalances caused by uncertain demand profiles. This growing requirement together with the movement to more sustainable energy sources emphasize the urgent nature of smart energy management systems. Modern energy grids now depend on power forecasting to solve these difficulties. Balancing supply and demand, incorporating renewable energy, and avoiding energy deficits or grid interruptions depend critically on accurate forecasting. Paired with machine learning and cloud technologies, live data acquisition systems offer sophisticated tools to monitor, forecast, and dynamically manage energy use. The deficiencies of conventional energy systems, namely, grid inefficiencies and the lack of usable information, emphasizes the need to embrace more intelligent, data-driven approaches. Predictive analytics and machine learning models can achieve optimized load balancing, lower running costs, and improved grid resilience. This combination not only enables more sustainable energy distribution but also enhances the overall dependability and efficiency of power networks in light of changing energy needs

### 1.2 Proposed Solution

Solution combines IoT-based data capture with machine learning to offer real-time energy monitoring and forecasting. The configuration involves voltage and current sensors attached to an Arduino UNO, which sends real-time power data to Firebase through an ESP8266 module. A Flask web dashboard offers dynamic updates and forecasting outputs, allowing users to track energy consumption in real time. Using a machine learning algorithm for demand prediction, the system provides timely and accurate power forecasts. The integration of hardware, cloud storage, and predictive analytics provides a scalable and cost-efficient means of smart energy management, addressing the increasing need for efficient and data-driven power solutions.

## 2. Data Acquisition System

### 2.1 Hardware Selection and Configuration

Selecting the right hardware, especially for reliable forecasting data, was important in creating this energy forecasting tool. The machine learning model's forecasting capacity is analytically dependent on sensor data. The accuracy of these prediction rests entirely on the exceptionally thorough quality of all this data. We selected these products mainly due to their outstanding performance, meaningful cost-effectiveness, strong connectivity and smooth compatibility. This new strategy uses a cloud connected Data Acquisition System (DAS) to acquire and smoothly transfer many real-time data points to a cloud database. Effective data retention and faster data retrieval, improve energy management and prediction accuracy for dynamic forecasting.

Sensors: ZMPT101B, and ACS712-20A was chosen for capturing the voltage and the current readings respectively. These sensors allow the power to be calculated in real-time.

Microcontroller: An Arduino microcontroller was then incorporated into the setup for the processing of sensor signals and data communication. It was highly compatible with other peripheral devices in support of real time functioning.

Communication Module: The ESP8266 Wi-Fi interface module was linked to bridge the hardware set with Firebase cloud storage for the reliable and secure transmission of data wirelessly

## 2.2 Block diagram

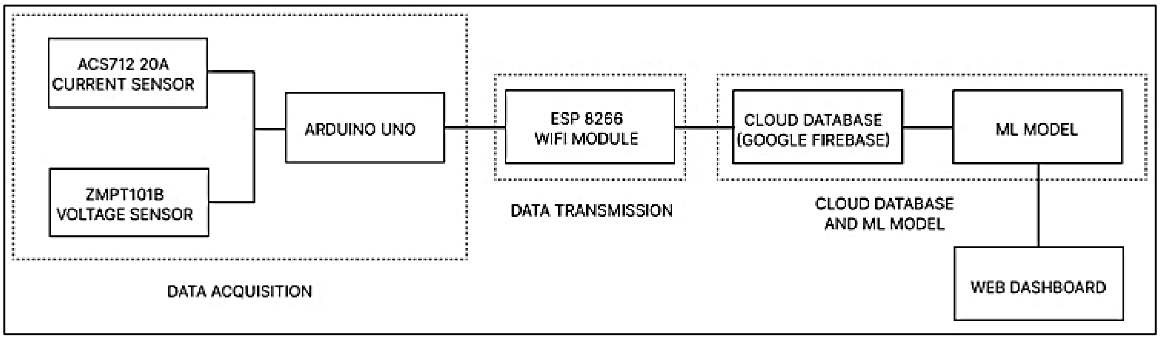


**Figure 1.** Block Diagram

The Figure 1 shows the block diagram of the real-time data acquisition and forecasting system for power monitoring and prediction. The scalability and efficiency of the system are offered through sensor technology, microcontrollers, cloud computing and machine learning [3,7]. The system connects an Arduino Uno microcontroller with ZMPT101B voltage sensors and ACS712-20A current sensors. To calculate the power consumption, these sensors record live electric parameters including voltage and current.

**Arduino UNO:** This will be the core microcontroller. It will receive the analog signal from the sensor, process and format data.

**Data Transmission:** The ESP8266 Wi-Fi module transmits the acquired data from the Arduino to a cloud database in Google Firebase, ensuring reliable and wireless communication.

**Cloud Database:** Firebase functions as a central storage system for live sensor data acquisition.

1) This will include cloud real-time storage, through which Google Firebase will be, managing all the sensor readings with secured and scaled storages and ensuring continuous availability of data to further process.

2) Machine Learning Model: A machine learning model for predicting power demand is directly given this data. It is a short-term load forecasting based on historical data as well as real-time information.

3) Web Dashboard: The interactive web dashboard shows the predicted as well as real-time data. The dashboard offers an easy-to-use interface for forecasting present power consumption patterns and forward projections. The overall flow is shown in Figure 2.

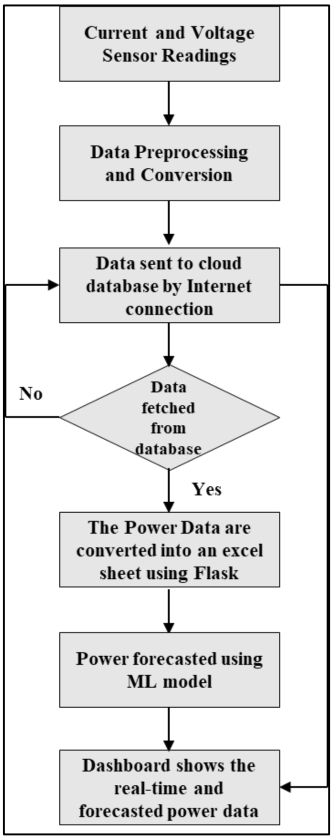


**Figure 2.** Flowchart

## 3. Machine Learning Model

Because of the outstanding efficiency and accuracy , the Light Gradient Boosting Machine (LightGBM) was chosen in this research for forecasting the power [1,11]. A gradient boosting structure, LightGBM uses tree-based learning techniques to improve forecasting ability. LightGBM uses a histogram-based approach that discretizes continuous features into bins, therefore greatly reducing computing overhead and memory needed. This makes it good in handling the enormous amounts of live data obtained from the DAS connected to the cloud. LightGBM improves general energy management and decision-making processes by handling high-dimensional data effectively and offering quick training times, therefore the forecasting model can produce dependable and timely predictions. The overall model can be expressed as an additive series of functions,

$$\hat{y_i} = \sum_{t=1}^{T} f_t\,(x_i) \tag{1}$$

A key aspect of LightGBM is its method for finding the optimal split. The gain from splitting a node into two child nodes is calculated as,

$$\left(Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} + \frac{(G_L + G_R)^2}{H_L + H_R + \lambda}\right] - \Upsilon\right) \tag{2}$$

The algorithm leverages gradient-based one-sided sampling (GOSS) and exclusive feature bundling (EFB) methods to optimize the training process. GOSS ensures that training prioritizes data samples with significant gradient values, capturing important information while reducing computational overhead. EFB efficiently groups mutually exclusive features to minimize feature dimensions, making LightGBM highly scalable for large datasets.

To prevent overfitting and improve generalization capabilities, LightGBM incorporates regularization methods such as L1 and L2 penalties. Furthermore, it supports parallel and distributed training, making it well-suited for real-time applications requiring low latency and high throughput.

This study selected LightGBM because of its superior performance in managing the time-series power consumption data gathered from the Godishala 33/11kV substation. The model accurately predicted power by effectively leveraging feature dependencies and temporal patterns, as illustrated in Fig. 3. Its high computational efficiency enabled seamless execution of the training and prediction stages, ensuring timely updates to the web dashboard for real-time observation and decision-making.

### 3.1 Training and Testing

The dataset obtained from the Godishala, Telangana, 33/11 kV substation, was used to train and validate the machine learning model [2]. The dataset includes power usage numbers gathered from January 1, 2023, through December 31, 2023, thereby guaranteeing a year of representative data for a wide range of seasonal and operational changes. Hourly data gathering caught important changes in energy need daily. The data was preprocessed for training to extract pertinent features, handle missing data, and make sure data is normalized. Roughly 75%

ISSN: 2582-2012 

of the dataset was split for training the LightGBM model to learn power consumption patterns; the rest 25% was used for model testing of accuracy and generalization abilities. This detailed and wide dataset allowed the model to forecast power demand with great accuracy and efficiently change, therefore supporting a stable and trustworthy energy forecast system [10,5]. Sample data from the database is listed in Table 1.

**Table 1.** Sample Data from Dataset

| DATE | TIME | POWER(kW) |
|---|---|---|
| 1/1/2021 | 01-00 | 1967 |
| 1/1/2021 | 02-00 | 1967 |
| 1/1/2021 | 03-00 | 1967 |
| 1/1/2021 | 04-00 | 2443 |
| 1/1/2021 | 05-00 | 2756 |
| 1/1/2021 | 06-00 | 3203 |
| 1/1/2021 | 07-00 | 4169 |
| 1/1/2021 | 08-00 | 5103 |
| 1/1/2021 | 09-00 | 5214 |
| 1/1/2021 | 10-00 | 4857 |

To enhance the accuracy of short-term load prediction, the current study employed the Light Gradient Boosting Machine (LightGBM) algorithm due to its scalability and efficiency when dealing with large data. The load data were initially divided into accurate time periods with distinct patterns of energy usage throughout the day. Time-filtered data was extracted for every period, and Min-Max scaling was applied to normalize it in order to keep similar feature values. Temporal features such as hour of day, day of week, and month were intended to incorporate cyclical and seasonal trends into the feature space.

A baseline model of LightGBM was initially trained on the datasets acquired to compare baseline performance. For further increasing model accuracy, a hyperparameter tuning step was conducted using Grid Search Cross-Validation (CV=3) over a provided parameter grid, including parameters n_estimators, learning_rate, max_depth, num_leaves, reg_alpha, and reg_lambda. The highest-performing estimator from the grid search was retained as the ultimate model for each time interval. This module-based approach facilitated the adaptation of every interval-based model to separate consumption behaviors and hence increased forecasting precision as well as corresponding the hybrid technique explored in the given studies [1-2].

A comparison of the test power measurements and the prediction produced by the LightGBM model is shown in Figure.3. The graph graphically depicts the model's predictive performance and its capacity to very closely follow the actual power usage patterns. Small variances between the test and predicted values suggest intrinsically difficult forecasting problems in dynamical power systems. But the predicted curve with the test data emphasizes how well the model learns temporal patterns and feature dependencies [9]. The small error margins shown in the comparison indicate the model's robustness and establish it as a dependable instrument for real-time prediction and energy management choices. This evaluation demonstrates that LightGBM model is appropriate for accurate and fast load prediction across different timeframes.

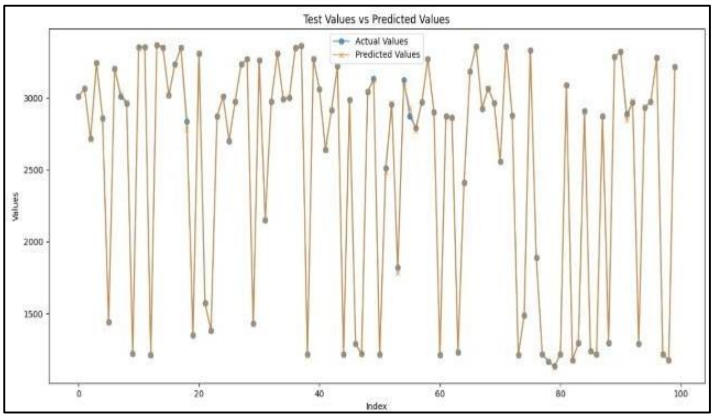


**Figure 3.** Comparison of Test values and Forecasted Values

The error distribution of the power values created by the LightGBM model is shown in Figure.4. The graph offers some understanding of the variance between the forecasted and actual power values, therefore assisting in evaluating the model's forecast accuracy and dependability. A focused error distribution near zero shows that the forecasts strongly correspond with the actual values, therefore testifying to the model's sensitivity. The distribution's symmetrical and tight spread indicates little variance in forecast errors, hence steady and consistent performance. Any small irregularities apparent in the graph call attention to singular instances of deviation that might be caused by unexpected power demand oscillation. On the whole, the distribution highlights how well the LightGBM model reduces prediction errors and improves forecasting accuracy

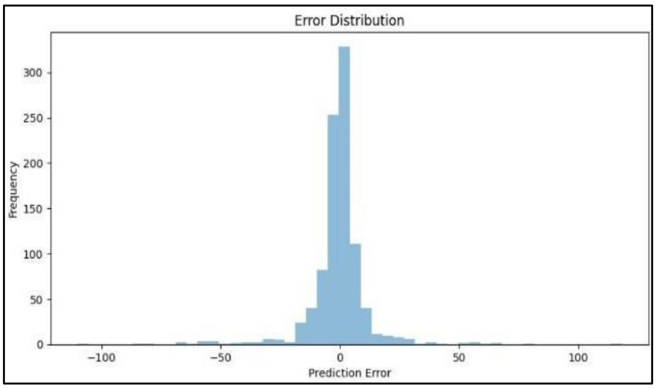


**Figure 4.** Error Distribution

## 3.2 Forecasted Result Overview

The forecasting results demonstrate the model's ability to accurately predict power consumption trends over time. The predictions closely align with the observed test values, emphasizing the effectiveness of the LightGBM model in handling time-series data. Performance metrics such as accuracy and RMSE emphasize the reliability of the forecasted values, ensuring real-time updates for monitoring and decision-making. Some results are presented in Table 2, providing a comprehensive view of the model's performance.

**Table 2.** Comparison of Actual Power and Forecasted Power values

| Actual Power (kW) | Forecasted Power(kW) |
|---|---|
| 2949.82 | 2954.89 |
| 2947.38 | 2952.04 |
| 2947.00 | 2953.18 |
| 2973.06 | 2972.92 |
| 2970.86 | 2964.18 |
| 2966.62 | 2966.22 |
| 2961.35 | 2960.43 |
| 2956.93 | 2956.21 |
| 2954.19 | 2950.28 |
| 2953.22 | 2885.61 |

### 3.3 Evaluation

The 99.99 percent achieved accuracy shows the outstanding forecasting power of the LightGBM model with very little variance from true values. A Root Mean Squared Error (RMSE) of 0.092 helps to confirm the accuracy of the model by giving a mean of prediction errors, refer Table 3.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y_i} - y_i)^2}{n}} \tag{3}$$

The model is effective at reducing mistakes and preserving good prediction stability throughout the data set; hence this small RMSE value points to that. The findings emphasize the accuracy and precision of the applied power load estimation system.

**Table 3.** Results of evaluating parameters

| S.NO | Parameter | Value |
|---|---|---|
| 1. | Accuracy | 99% |
| 2. | RMSE | 0.092 |

## 4. Machine Learning Model Implementation

For facilitating real-time forecasting and user access, the trained LightGBM model was made operational using a web framework based on Flask in Python. Flask was used owing to its flexibility and lightness, as it was easy to integrate with machine learning pipelines and support RESTful APIs. The deployment layout consisted of setting up a REST API endpoint via which live or user-provided data could be accessed in JSON form. The model returned predicted power values, which were formatted and returned in the form of a JSON response. This deployment pipeline ensured that the ML system could respond to requests in low latency, allowing it to be integrated with the real-time data acquisition system (DAS) and dashboard interface. This combination of backend intelligence and frontend visualization enabled the forecasted data to be displayed dynamically, hence enabling real-time decision-making for smart grid and energy management applications.

## 5. Web Dashboard with ML model and Cloud Database

The development of a modernized web dashboard with reference to [6], aims to provide a sophisticated monitoring and forecasting tool for energy management by integrating machine learning models with cloud-based data acquisition systems. This approach addresses the growing need for intelligent and predictive monitoring solutions in energy systems, enabling a comprehensive platform for stakeholders to make informed decisions.

The DAS (Data Acquisition System) is equipped with ZMPT voltage and ACS712-20A current sensors connected to an Arduino, which processes the sensor data. The processed data is transmitted wirelessly to Firebase using the ESP8266 module. Firebase serves as the cloud database, enabling secure and centralized storage for real-time data access [8].

A REST API framework is implemented using Flask to retrieve power data from Firebase at regular intervals. The retrieved data is dynamically converted into an Excel-compatible format, ensuring efficient data storage and preprocessing. This preprocessing step is essential for maintaining data integrity and compatibility with the forecasting model.

Every 5 minutes, the structured data from Firebase is fed into the LightGBM machine learning model. The forecasting model's high computational efficiency ensures that predictions are generated promptly without compromising performance.

The web dashboard is designed to provide dynamic updates by displaying live readings of voltage, current, power, every 3 seconds. Graphical plots offer a clear visualization of forecasted power usage trends, allowing users to assess consumption patterns and predict future requirements. The forecasted results are displayed along the live data for easy comparison, with a focus on maintaining accuracy and clarity. The REST API facilitates seamless communication between the backend (data processing and forecasting) and the web dashboard. This lightweight and efficient communication framework ensures that both live and forecasted data are accurately transmitted to the dashboard in real-time, minimizing latency and maximizing system responsiveness.


**Figure 5.** Web Dashboard

The web dashboard's interactive and user-friendly interface (Figure.5) empowers stakeholders to monitor power consumption trends, evaluate prediction accuracy, and access error metrics for better decision-making. The system's cloud-based architecture, coupled with ML-driven forecasting and REST API integration, ensures scalability and futures cope for expanding data requirements.

By seamlessly integrating hardware components, cloud storage, and advanced machine learning algorithms, this web dashboard provides a state-of-the-art solution for real-time power monitoring and predictive analytics. It demonstrates the power of modern technologies in delivering intelligent, scalable, and efficient energy management systems that meet contemporary energy challenges.

## 6. Conclusion

This research presents a comprehensive solution for real-time energy monitoring and forecasting using a cloud-integrated data acquisition system (DAS) combined with machine learning models. By utilizing modern technologies such as Firebase cloud storage, REST APIs, and the LightGBM forecasting algorithm, the system efficiently handles dynamic power data, enabling accurate predictions and real-time visualization through a user-friendly web dashboard. The inclusion of hardware components for live data acquisition ensures continuous monitoring and data accuracy, crucial for effective energy management. The selection of the LightGBM model , enabled a high computational efficiency, robust handling of feature dependencies, and superior prediction accuracy for time- series data. With a forecasting accuracy of 99 % and a minimal root mean square error (RMSE) of 0.092, the model demonstrated its capability to provide reliable insights for stakeholders.This methodology not only modernizes power consumption monitoring but also supports data-driven decision-making through timely forecasts and error analysis. The scalable architecture and cloud-based approach make the system adaptable for future expansion, ensuring that it can accommodate growing energy requirements and data volumes. Overall, the solution serves as a vital step towards intelligent, efficient, and automated energy management systems, contributing to more sustainable and optimized power distribution practices.

## References

[1] Wang, Yuanyuan, Jun Chen, Xiaoqiao Chen, Xiangjun Zeng, Yang Kong, Shanfeng Sun, Yongsheng Guo, and Ying Liu. "Short-term load forecasting for industrial customers based on TCN-LightGBM." IEEE Transactions on Power Systems 36, no. 3 (2020): 1984-1997.

[2] Veeramsetty, Venkataramana, Arjun Mohnot, Gaurav Singal, and Surender Reddy Salkuti. "Short term active power load prediction on a 33/11 kv substation using regression models." Energies 14, no. 11 (2021): 2981.

[3] Shete, Monika G., and P. M. Joshi. "Practical implementation of load forecasting in smart grid: operational domain feature." In 2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1171-1176. IEEE, 2019.

[4] Saini, Vikash Kumar, Rajesh Kumar, Ameena S. Al-Sumaiti, and Ehsan Heydarian-Forushani. "Learning based short term wind speed forecasting models for smart grid applications: An extensive review and case study." Electric Power Systems Research 222 (2023): 109502.

[5] Zeng, Pan, Min Jin, and Md Fazla Elahe. "Short-term power load forecasting based on cross multi-model and second decision mechanism." IEEE Access 8 (2020): 184061-184072.

[6] Tanjim, Meer Shadman Shafkat, Ashrafun Nushra Oishi, Ali Azam Sojib, Md Bashir Ahmmad, Md Shaiful Islam, and Suman Chandra Das. "Web-Based Small Scaled Portable Data Acquisition and Control Unit for Demand Side Load." In 2020 IEEE Region 10 Symposium (TENSYMP), pp. 564-567. IEEE, 2020.

[7] Lee, Youngiae, Jinhong Kim, and Kiltaek Lim. "Development of data acquisition system for collecting and processing data." In 2020 IEEE International Conference on Consumer Electronics (ICCE), pp. 1-2. IEEE, 2020.

[8] Z. Guan, J. Li, L. Wu, Y. Zhang, J. Wu and X. Du, "Achieving Efficient and Secure Data Acquisition for Cloud-Supported Internet of Things in Smart Grid," in IEEE Internet of Things Journal, vol. 4, no. 6, pp. 1934-1944, Dec. 2017, doi: 10.1109/JIOT.2017.2690522.

[9] Chemetova, S., P. J. Santos, and M. Ventim-Neves. "Load peak forecasting in different load patterns situations." In 2016 10th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG), pp. 148-151. IEEE, 2016.

[10] Kong, Weicong, Zhao Yang Dong, Youwei Jia, David J. Hill, Yan Xu, and Yuan Zhang. "Short-term residential load forecasting based on LSTM recurrent neural network." IEEE transactions on smart grid 10, no. 1 (2017): 841-851.

[11] Skolthanarat, Siriya, Udom Lewlomphaisarl, and Kanokvate Tungpimolrut. "Short-term load forecasting algorithm and optimization in smart grid operations and planning." In 2014 IEEE Conference on Technologies for Sustainability (SusTech), pp. 165-171. IEEE, 2014.