# Analysis of Energy Data Forecasting Performance in Low-Voltage DC Microgrids

Daniel-Cătălin Mitroi
*Automation and Industrial Informatics*
*University Politehnica of Bucharest*
Bucharest, Romania
daniel.mitroi1806@upb.ro

Cristina Nichiforov
*Automation and Industrial Informatics*
*University Politehnica of Bucharest*
Bucharest, Romania
cristina.nichiforov@upb.ro

Iulia Stamatescu
*Automation and Industrial Informatics*
*University Politehnica of Bucharest*
Bucharest, Romania
iulia.stamatescu@upb.ro

Grigore Stamatescu
*Automation and Industrial Informatics*
*University Politehnica of Bucharest*
Bucharest, Romania
grigore.stamatescu@upb.ro

***Abstract*—Low voltage Direct Current (LVDC) systems represent an emerging alternative for efficient usage and local energy distribution. Such systems are able to integrate renewable energy sources, energy storage and DC-native and AC loads using a common DC bus and multiple voltage levels. We present an end-to-end implementation of a prediction framework for LVDC microgrids loads in a residential scenario using a data-driven approach. Given the limited availability of public DC microgrid datasets, a data augmentation pipeline is first proposed and deployed using open source software libraries. Results provide a comparative analysis between the SARIMA and Prophet load forecasting (regression) models using standardised evaluation metrics.**

***Index Terms*—dc microgrids, load forecasting, data augmentation, sarima, prophet.**

## I. Introduction

The deployment of LVDC residential and industrial microgrids is currently emerging as a practical solution for efficient and secure energy usage [1]. Artificial intelligence (AI) and machine learning (ML) techniques applied to DC energy systems can improve system resilience and real-time control towards efficient and reliable operation [2]. Forecasting in DC microgrids poses different challenges compared to AC systems. The protection mechanisms in DC grids require faster and more precise responses, making real-time forecasting critical. Moreover, consumption patterns are different, transmission losses are lower, and the lack of standardized datasets for DC systems makes the forecasting task even more complex for developing accurate prediction models.

In this context the main contribution of the article is considered to be an applied artificial intelligence approach to mitigate public dataset scarcity for DC microgrids using data augmentation together with comparative evaluation of time series load forecasting models for such task. The data augmentation framework was specifically designed for DC microgrids, making the synthetic data more useful and closer to real-word conditions.

The rest of the paper is structured as follows. Section II briefly discusses related work with regard to deep learning methods applied for DC microgrid load forecasting. In Section III we present the main elements of a DC microgrid architecture, together with the methodology used for data augmentation and load forecasting and associated dataset. Results are presented in comparative manner in Section IV using standardised metrics. Section V concludes the paper with outlook on future work.

## II. Related work

This section presents previous research focused on creating accurate models for load forecasting in DC microgrids. Load forecasting estimates the amount of electricity needed to satisfy future demand and is thus a key component in efficient energy management [3].

The paper [4] presents a recurrent neural network architecture, the LSTM method, for load forecasting. A hardware-in-the-loop approach is used, which integrates a 255W solar array (combining polycrystalline and monocrystalline panels), a 12V lead-acid battery bank, and IoT enabled sensors (LDR, DHT11, voltage/current sensors) with a NODEMCU ESP8266 for real-time data acquisition. The prediction model realized to forecast energy consumption in DC microgrid powered by solar panels and batteries achieved an accuracy of about 95%. However, this performance can be affected by features such as weather data and others.

The autors in [5] presented an LSTM-based forecasting model for predicting load demand and power generation in DC microgrids, while simultaneously addressing the problem of optimal storage system management (ESS) through consensus-based distributed control. The proposed solution integrates a distributed extended Kalman Filter (DEKF) algorithm for efficient training of LSTM models. This approach achieved a voltage deviation below 0.1V and increased the microgrid's autonomous operation duration by up to 30% under renewable energy source (RES) outage conditions. Building upon these prior approaches, we next describe the specific DC microgrid configuration used in our study, followed by the proposed methodology for data augmentation and forecasting.

## III. DC Microgrid

A DC microgrid is a small-scale electricity grid that operates mainly with direct current (DC) and functions as a localized and integrated energy system, capable of operating both connected to the national grid and independently (island mode).

At the foundation of a DC microgrid's arhitecture is the DC common bus, which interconnects generation sources, storage systems and energy consumers. Bassically, it manages the microgrid's energy. Through this concept, DC microgrids significantly reduce losses from inefficient conversions (AC-DC) that occur in traditional AC grids, as they optimize the energy flow.

The structure of DC microgrid is illustrated in Figure 1. Solar panels, wind turbines, small hydro power plants, biomass & biogas and fuel cells are connected to the DC bus through the converter. Solar panels are connected to the DC bus through a DC-DC converter. Wind turbines are connected to the DC bus through an AC-DC converter. Small hydropower plant are connected to the DC bus through an AC-DC converter. Biomass and biogas are connected to the DC bus through an AC-DC converter. Fuel cells are connected to the DC bus through an DC-DC converter. In addition to renewable and conventional sources of power generation, the DC bus is also connected to the main grid. The connection to the main AC grid is realized through bidirectional DC-AC converters, which allow both the import of energy from the grid when local generation is insufficient and the export of surplus energy to the main grid.

Energy storage systems and electric vehicles are connected to the DC bus through bidirectional DC-DC converters, facilitating energy transfer in both directions: from the DC bus to the storage systems and electric vehicles and from the storage systems and electric vehicles to the DC bus.

DC microgrids enable a multi-level approach to operating voltages. Within this architecture, in addition to the main bus, we also have a sub-bus at lower voltages, such as 48V, specifically designed for particular applications such as telecommunications or electronic devices.

DC microgrids offer numerous advantages, including increased energy efficiency, simplified integration of renewable energy sources, enhanced reliability and reduced maintenance costs. The spread and evolution of DC microgrids makes efficient energy management more and more relevant. This involves load forecasting (i.e. estimating how much electricity will be needed to supply future demand) to optimize system performance [4].

Over time, electricity load forecasting has been approached in a variety of ways, including statistical modeling, artificial intelligence and machine learning methodologies. In this paper we implement and compare the SARIMA [6] and Prophet

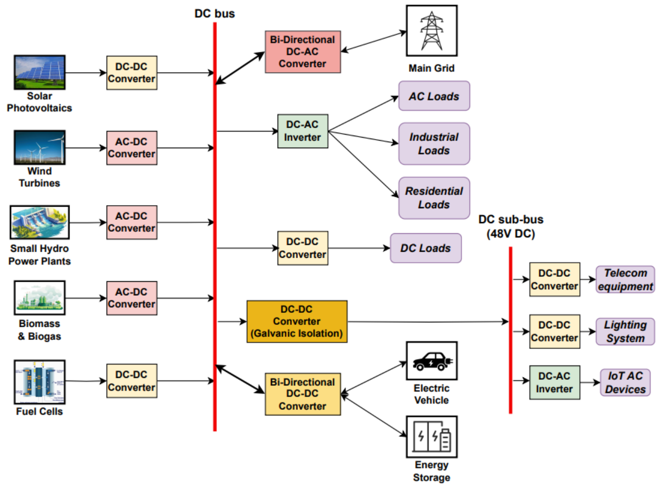


Fig. 1. Structure of a DC microgrid

time-series prediction algorithms for forecasting energy consumption. These algorithms have been implemented on a synthetically augmented and generated dataset, since currently large datasets are not readily available at suitable quality levels.

We chose as a starting point a dataset monitoring a residential DC microgrid system for 5 households, each equipped with photovoltaic panels and individual batteries. The dataset tracks the interaction between consumption, solar production and energy storage over a full day, measured at 15-minute intervals. The specific DC microgrid architecture described by the dataset is shown in Figure 2. In order to scale, extend, and enhance the available samples, synthetic data generation produces realistic power consumption time series that maintain the key features of real-world data. This method makes it possible to get beyond restrictions on the historical data that is currently accessible and improves the forecasting model resilience.

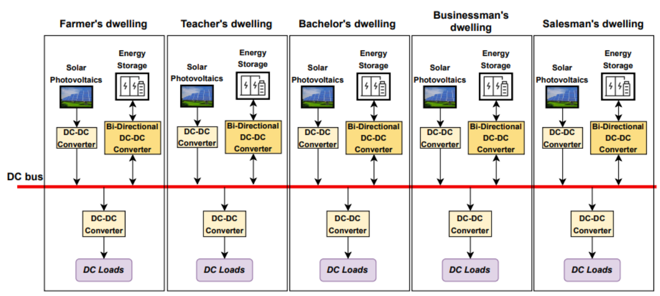


Fig. 2. DC microgrid architecture for the dataset

### A. Methodology

Synthetic data generation creates realistic power consumption time series that preserve the essential characteristics of real-world data while providing the ability to scale, extend, and augment the available samples. This approach allows to overcome limitations in available historical data and enhance the robustness of forecasting models.

The key features of the process are: the extraction of base patterns, multiscale temporal modeling, and consideration of domain-specific constraints. The following paragraphs describe each step.

- Extraction of Base
  - The system extracts daily consumption patterns from existing samples
  - If limited data is available, patterns are looped to create full day cycles
  - Core temporal features of the original data are preserved
- Multi-scale Temporal Modeling
  - Implements daily patterns with realistic morning and evening peaks
  - Creates weekly patterns distinguishing weekdays (higher consumption) from weekends (approximately 15% reduced consumption)
  - Models seasonal effects including yearly temperature variations and quarterly transitions
- Considering Domain-Specific Constraints
  - Enforces power system physics such as PV plant output being zero at night
  - Maintains proper power balance between consumption, generation, and battery storage
  - Creates realistic correlations between system components

The data generator implements different augmentation techniques that can be applied selectively (Fig. 3): noise addition, amplitude scaling, seasonal effects and frequency variations.

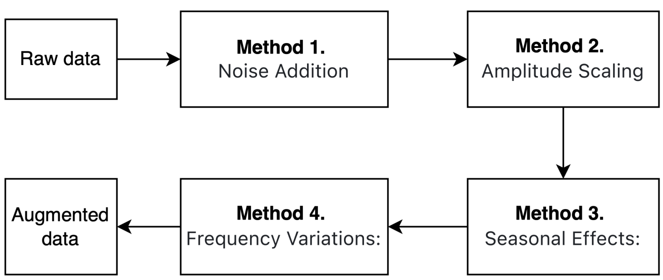


Fig. 3. Data augmentation framework

In order to provide new data and enable the model to learn different variations and enhance generalisation performance, noise injection entails introducing random noise to the original data. The *Noise Addition Method* creates new data by supplementing the old data with a predetermined amount of random noise. In this work, we employed time-varying Gaussian noise with a dynamic scale ranging from 3% to 7% of the mean power value. The noise intensity follows a sinusoidal pattern aligned with the time of day, simulating higher variability during peak consumption hours and lower variability during off-peak periods. This approach creates more realistic augmented data by reflecting the natural fluctuations in measurement precision that occur throughout the day.

*The Amplitude Scaling Method* applies global consumption level shifts using a ±10% variation. The method simulates day-to-day variations in overall energy demand maintaining the relative proportions of peaks and troughs.

*The Seasonal Effects Method* implements annual cycles using sinusoidal modulation. The method simulates seasonal variations in PV generation (higher in summer) and creates quarterly variations in consumption patterns.

*The Frequency Variations Method* introduces cyclical components at different frequencies. The method simulates appliance cycles, HVAC operation, and other periodic behaviors in order to create realistic sub-daily variations in consumption patterns.

The ARIMA model is a mathematical model used for time series forecasting, based on auto regression (AR), integration (I) and differencing and moving average (MA). It first checks stationarity and seasonality, then identifies the AR, MA parameters. Thus the differencing process first takes place to convert non-stationary data into stationary data and then the ARIMA (auto-regressive integrated moving average) model is generated.

The SARIMA (Seasonal auto regressive integrated moving average) model, also called Box-Jenkins, is similar to the ARIMA model, except that it is used when the time series exhibit seasonality. The general form of a SARIMA model is denoted as follows:

$$\mathrm{SARIMA}(p,d,q)\times(P,D,Q)_s \tag{1}$$

The model parameters p,d and q represent the non-seasonal part (p, d and q represent the nonstationary AR order without seasonal differencing, and the nonstationary MA order, respectively), and P,D,Q and S the seasonal part (P , D , Q and S correspond to the seasonal AR order, seasonal differencing, seasonal MA order and seasonal pattern repetition time interval, respectively. [7] Mathematically, it can be expressed in terms of a composite model as in equation 2:

$$\Phi_P(L^s)\,\phi_p(L)\,(1-L)^d\,(1-L^s)^D\,y_t = \Theta_Q(L^s)\,\theta_q(L)\,\varepsilon_t \tag{2}$$

where:

- $L$ is the lag operator
- $\phi_p(L)$, $\Phi_P(L^s)$ - autoregressive polynomials (non-seasonal and seasonal)
- $\theta_q(L)$, $\Theta_Q(L^s)$ - moving average polynomials (non-seasonal and seasonal)
- $(1-L)^d$ - non-seasonal differencing component
- $(1-L^s)^D$ - seasonal differencing component
- $\varepsilon_t$ - white noise

Prophet is a forecasting model developed by Facebook and is available in Python and R [8]. It decomposes the dataset into 3 main components, i.e. trend, seasonality and holidays. It can be represented as in equation 3:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t \tag{3}$$

where the model parameters g(t), s(t), h(t) are interval linear curves with automatic change point detection. The seasonality component is modeled using a Fourier series which provides a flexible model that captures varying periodic effects [9]:

$$s(t) = \sum_{n=1}^{N} \left[ a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right] \quad (4)$$

where p represents the regular periodicities.

## IV. Results

### A. Metrics evaluation

Once the forecasting approach and augmented dataset generation methodology are established, we evaluate the performance of the SARIMA and prophet models using standard metrics. These two methods were chosen because they have the ability to model seasonality, an important feature in residential DC microgrid consumption patterns. While many AI-based methods exist, they require large volumes of training data and extensive computational resources. In contrast, SARIMA and Prophet can be trained effectively on limited datasets, offering a solid balance between interpretability, computational efficiency, and seasonal modeling capability.

Each subplot of Figure 4 represents a version of the data in which specific augmentation techniques were applied to increase the availability of the initial training set. The first subset represents the original data: the x-axis represents time (with hourly intervals on January 1st) and the y-axis represents power consumption in kilowatts (kW). The second subplot repesents the data after the Noise Addition Method was implemented. In this method random noise is added to the original signal in order to simulate measurement errors or sensor fluctuations. The third graph represents data after the Amplitude Scaling method was implemented while the fourth represents the input data after the Seasonal Effects method. The last plot represents the data after the Frequency Variations Method.

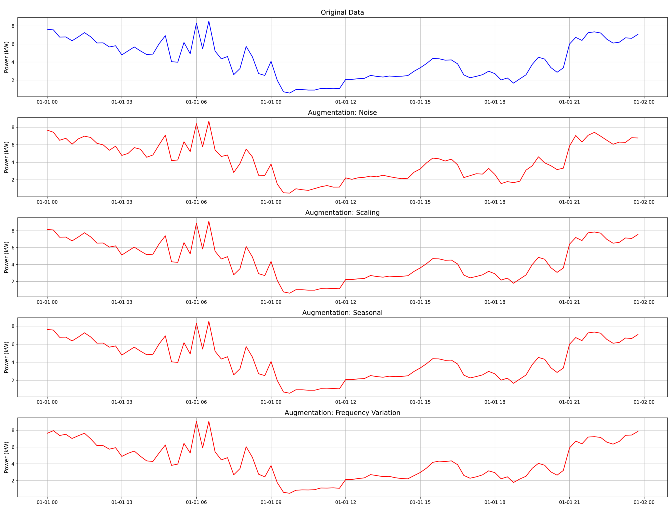


Fig. 4. Data augmentation results

Using the raw data for one single day has led to the generation of a one month augmented dataset, with the results presented in Figure 5 being obtained. The power (blue) curve is influenced by both PV generation (green) and battery charging (red). During the day, solar generation helps offset load, while the battery may charge when excess solar is available. At night, the absence of PV production and ongoing battery behavior suggests reliance on stored energy or external sources. On this data the two proposed energy forecasting models: SARIMA and Prophet, were implemented.

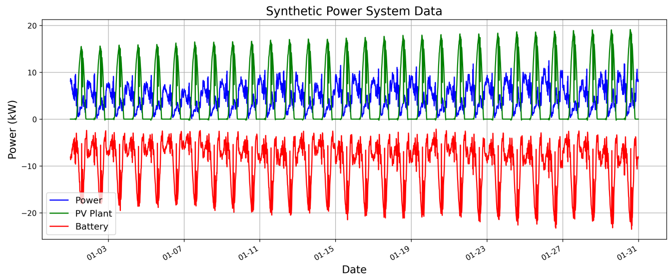


Fig. 5. Synthetic data visualization

Figure 6 and Figure 7 reflect the performance of the SARIMA model — specifically SARIMA (1,0,1) × (0,1,1,96) and the Prophet model in forecasting power consumption over the course of a single day (January 30th). The black line represents the actual power usage values (in kilowatts), while the blue line corresponds to the predicted values generated by the SARIMA or the Prophet model. Surrounding the blue line is a shaded blue region, which illustrates the 95% confidence interval, giving a visual indication of the uncertainty in the predictions.

The SARIMA model performs notably well, capturing the general trends, including both the peaks and troughs in the power usage. The alignment between the predicted and actual values suggests that the model is accurately tracking the temporal patterns in the data. The goodness-of-fit metrics shown in the upper-left corner provide a quantitative measure of the model's performance: Root Mean Square Error (RMSE) is 0.7905, Mean Absolute Error (MAE) is 0.5832, and the coefficient of determination ($R^2$) is 0.9230 — indicating that over 92% of the variance in the actual data is explained by the model.

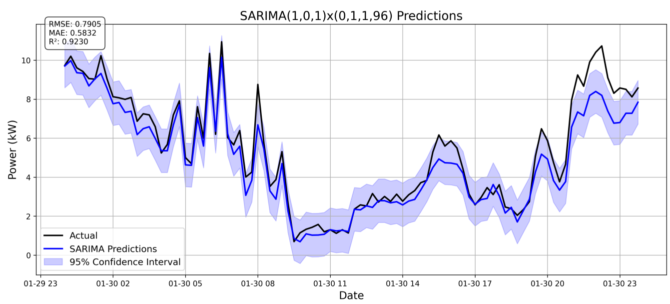


Fig. 6. SARIMA predictions

The Prophet model performs fails in capturing the peaks and troughs in the power usage. The goodness-of-fit metrics shown in the upper-left corner provide a quantitative measure of the model's performance: Root Mean Square Error (RMSE) is 0.8886, Mean Absolute Error (MAE) is 0.6253, and the coefficient of determination (R²) is 0.9027 — indicating that over 90% of the variance in the actual data is explained by the model.

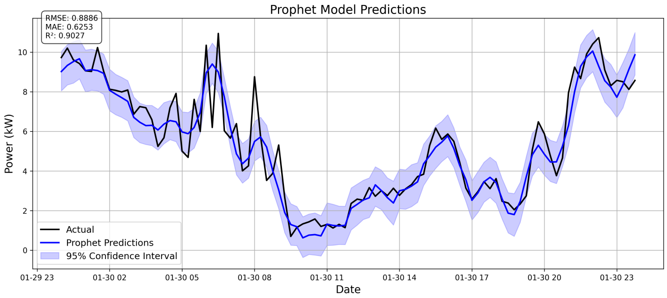


Fig. 7. Prophet analysis

The coefficient of determination evaluates the quality of the actual predictions of regression models, including in the context of time series prediction. $R^2$ is the square of the correlation between the actual and predicted variable. It ranges from 0 to 1, where 0 indicates that the variables are uncorrelated and 1 that the model explains all predicted values [10].

$$R^2 = 1 - \frac{\sum_1^n (E_{\text{act}} - E_{\text{pre}})^2}{\sum_1^n (E_{\text{act}} - \overline{E}_{\text{act}})^2} \tag{5}$$

The Mean Absolute Error (MAE) is a metric used to measure the average of absolute differences between predicted and actual values. It has range $(0, +\infty)$, where a smaller value indicates a more accurate prediction model [11].

$$\text{MAE} = \frac{1}{n} \sum_1^n |E_{\text{pre}} - E_{\text{act}}| \tag{6}$$

Root Mean Squared Error (RMSE) measures the vertical distance between predicted and actual values following the regression line. Compared to the MAE, the RMSE is more relevant for observing large errors because of the way it is calculated [11]:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_1^n (E_{\text{pre}} - E_{\text{act}})^2} \tag{7}$$

What this study shows, beyond the comparison itself, is that with the right kind of synthetic data, even classical statistical models such as SARIMA can outperform more complex models like Prophet in forecasting short-term load profiles in DC microgrids. This highlights the importance of data quality and structure, even more than model complexity, in achieving reliable predictions. Table I, Figure 8 and Figure **??** depict the a comparison summary of the power presdiction models. Based on the evaluation metrics, the SARIMA model performed best overall with the lowest RMSE of 0.7905, MAE = 0.5832 and $R^2$ = 0.9230.

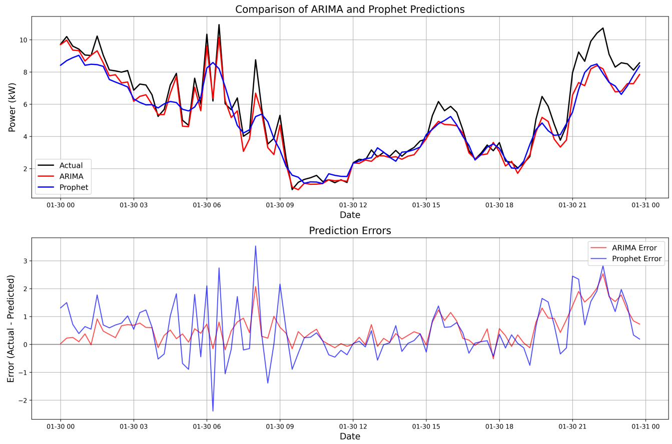


Fig. 8. Model comparison

Figure 8 highlights the SARIMA model's effectiveness in capturing seasonality and trends, especially in a setting where patterns repeat cyclically (like energy consumption in a smart grid or residential setting). However, it is important to note that while the synthetic dataset was carefully designed to reflect realistic consumption patterns, there may still be differences compared to real-world data. Factors such as sudden unexpected load spikes, faults, or irregular user behavior could lead to differences in actual implementation. Overall, this visualization demonstrates that the SARIMA model is a strong candidate for time series forecasting in power systems, particularly when seasonality is a key feature of the data.

TABLE I
MODEL PERFORMANCE METRICS

| **Model** | ***RMSE*** | ***MAE*** | $R^2$ |
|---|---|---|---|
| SARIMA | 0.7905 | 0.5832 | 0.9230 |
| Prophet | 0.8886 | 0.6253 | 0.9027 |

Finally, Figure 9 illustrates the standardised metrics comparison form RMSE, MAE and $R^2$ in a graphical manner in order to observe the relative performance of the two implemented methods.

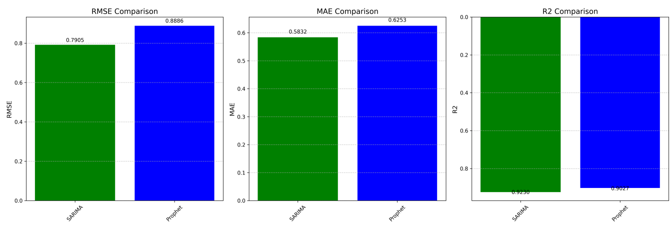


Fig. 9. Performance comparison

## V. Conclusion

The article discussed a viable approach to online load forecasting in LVDC microgrids using both conventional (SARIMA) and data-driven deep learning methods (Prophet). The results analysis has shown that both methods can be suitable for robust regression tasks by combining data augmentation, model parametrisation and domain expertise. Deployment of models on embedded hardware such as smart meters is considered feasible through embedded ML techniques.

Ongoing future work is focused on extending the presented load forecasting framework with multiple models for time series data analysis, as well as enhancing data quality through extended dataset collection and validation.

## References


[1] G. Stamatescu, R. Plamanescu, and M. Albu, "Online embedded analytics for energy time series pre-processing in lvdc microgrids," in *7th International Conference on DC Microgrids (ICDCM)*, 2025.

[2] A. N. Akpolat, M. R. Habibi, E. Dursun, A. E. Kuzucuoğlu, Y. Yang, T. Dragičević, and F. Blaabjerg, "Sensorless control of dc microgrid based on artificial intelligence," *IEEE Transactions on Energy Conversion*, vol. 36, no. 3, pp. 2319–2329, 2021.

[3] N. Shabir, O. Husev, H. N. Hokmabad, D. Kamran, M. Jawad, and J. Martins, "Multi-step ahead short-term residential dc load forecasting: A comparative study of ngboost-based algorithms," in *7th International Conference on DC Microgrids (ICDCM)*, 2025.

[4] P. T. B. A. C. C. D. V. Pranav, M. Ashwin and N. Venugopal, "Smart management of dc microgrid using forecasting," in *ICT Analysis and Applications Proceedings of ICT4SD 2023*, 2023, pp. 271–279.

[5] S. A. Alavi, K. Mehran, V. Vahidinasab, and J. P. S. Catalão, "Forecast-based consensus control for dc microgrids using distributed long short-term memory deep learning models," in *IEEE Transactions on Smart Grid ( Volume: 12, Issue: 5, September 2021)*, 2021, pp. 3718 – 3730.

[6] C. Nichiforov, I. Stamatescu, I. Făgărăşan, and G. Stamatescu, "Energy consumption forecasting using arima and neural network models," in *2017 5th International Symposium on Electrical and Electronics Engineering (ISEEE)*, 2017, pp. 1–4.

[7] K. Samal, K. Babu, S. Das, and A. Acharya, "Time series based air pollution forecasting using sarima and prophet model," in *ITCC 2019: Proceedings of the 2019 International Conference on Information Technology and Computer Communications*, 08 2019, pp. 80–85.

[8] Facebook, "Automatic forecasting procedure," accessed: 2025-05-09.

[9] ——, "Prophet: forecasting at scale [online]," accessed: 2025-05-09.

[10] A. Jierula, S. Wang, T.-M. OH, and P. Wang, "Study on accuracy metrics for evaluating the predictions of damage locations in deep piles using artificial neural networks with acoustic emission data," in *Applied Sciences*, vol. 11, no. 5, 2021, p. 2314.

[11] D.-C. Mitroi, I. Stamatescu, N. Arghira, and I. Făgărăşan, "Lithium-ion battery soc estimation: A neural network approach," in *2024 28th International Conference on System Theory, Control and Computing (ICSTCC)*, 2024, pp. 432–437.