Machine learning based-optimization of a Distributed Generation power system

Denis Nasasira
Mechatronics Department
Ahlul Bayt Intenational University
Tehran-Iran
denisnasasira54@gmail.com

Supervisors

Dr. Masoud Masih Tehrani
Dr. Mansour Hakim-Elahi

Abstract— This paper presents a machine learning-driven approach for Distributed Generation optimization applied to the department of mechatronics engineering of Ahlul Bayt International University as a case study. Particular focus is on Random Forest machine learning algorithms for optimal sizing, load generation forecasting, and reinforcement learning for adaptive system control. The paper systematically analyzes the superiority of this machine learning method over conventional techniques in handling stochastic environments and improving overall system reliability. Through comparative assessment of current literature, this research directly addresses the core problem of traditional static approaches that cannot handle the real-time variations in department energy usage, developed a machine learning framework that adapts to these uncertainties.

While the implementation focuses on economic optimization and energy efficiency, the same machine learning foundation has also been extended to explicitly address voltage stability and power losses by incorporating electrical constraints into the optimization function. This has been achieved through a 47% faster economic payback (3.6 vs 6.8 years) and 47% higher operational savings (22.4% vs 15.2%) through intelligent energy management as well as Achieving 88.2% load forecasting accuracy with Random Forest algorithms resulting in 99.5% system efficiency and 100% voltage stability.

Keywords— Distributed Generation, Machine Learning, Power System Optimization, DG Placement and Sizing, Renewable Integration, Smart Grid.

I. INTRODUCTION

Distributed Generation refers to electricity generation systems installed at or near the point of use. [1] The global energy landscape is rapidly evolving due to increasing demand, technological advancements, and a growing commitment to sustainable practices. Distributed Generation (DG) small-scale electricity production units located close to the point of consumption has gained prominence as a viable solution to improve energy efficiency, reduce transmission losses, and enhance power reliability. [2] As the modern power system continues to grow in size, complexity, and uncertainty, traditional methods may occasionally prove insufficient in addressing the associated challenges. The improper location of distributed generation varies the voltage profile, increases losses and compromises network capacity. [3] Machine learning algorithms predict accurate site positions, and network reconfiguration improves the capacity of the power system. [4] The Department of Mechatronics Engineering at Ahlul Bayt International University serves as an ideal testbed for exploring DG optimization strategies.

Conventional approaches to DG siting and sizing often rely on static models and assumptions, which may not adapt well to the non-linear, time-varying nature of real-world electrical networks. Therefore, the integration of **Machine Learning (ML)** offers a data-driven and adaptive solution. [5] By leveraging algorithms Random Forest (RF), this study introduces an intelligent system to optimize the placement and performance of DG units in the department's power distribution system.

II. PROBLEM STATEMENT

The traditional placement and sizing of DG units in power systems are primarily based on mathematical or heuristic approaches, which may fail to account for real-time operational variations and uncertainties. Inappropriate DG positioning can lead to voltage instability, increased power losses, and inefficient system performance. [6] In the context of the Department of Mechatronics Engineering, energy consumption is unpredictable and influenced by equipment usage, research activity, and academic schedules. There is currently no optimized strategy to manage power distribution or integrate renewable energy effectively. [7] Without a smart and adaptive framework, the department risks energy wastage, reliability issues, and missed opportunities for implementing sustainable energy practices. Hence, an ML-based optimization framework is necessary to accurately determine optimal DG siting and sizing for improved efficiency and stability. [8]

III. OBJECIVES OF THE RESEARCH

A. *Main Objective*

- To develop a machine learning-based optimization framework for the optimal sizing and energy management of Distributed Generation units in a small-scale microgrid, using the Department of Mechatronics Engineering as a case study.

B. *Specific Objectives*

- To select and adopt datasets from IEEE DataPort that model accurately the load profile and Distributed Generation potential of the Mechatronics Engineering Department
- To develop and train machine learning models using the selected dataset for accurate prediction and optimal Distributed Generation operation.
- To design and simulate a structured machine learning-based framework for optimizing distributed power system performance based on the trained model in terms of power loss reduction, voltage stability, and efficiency.

IV. METHODOLOGY

To study the method of implementing the Distributed Generation power system flame work for the mechatronics department by using the energy used by the load, the selection of a battery energy storage system suitable for residential use in the case of solar rooftop installation is carried out using the machine learning algorithm and python, which uses the residential load data to determine the most suitable power system.

Due to challenges in obtaining real-time data from the Department of Mechatronics like the unavailability of a dedicated metering infrastructure at the department, this study utilizes publicly available, high-quality datasets to model the energy consumption of a typical 3-room residential building. Specifically, the UCI Household Power Consumption dataset, which provides detailed models of distribution systems, and the Smart Grid Data, offering load profiles and generation data, are employed. The load profile of the dataset closely mirrors the expected consumption patterns of the Mechatronics Engineering Department. These datasets are therefore adopted are adapted to reflect the building characteristics and environmental conditions pertinent to this study. By leveraging these datasets, the research focuses on developing and optimizing machine learning models for distributed generation sizing and performance evaluation, ensuring the methodology's applicability to real-world scenarios.

The battery energy storage system suitable for residential use, in the case of DG installation, can be applied to different scenarios to optimize the battery usage data with different renewable energy sources. The systematic flow of the methodology is shown in the figure 1 flow chart below.

Fig. 1.

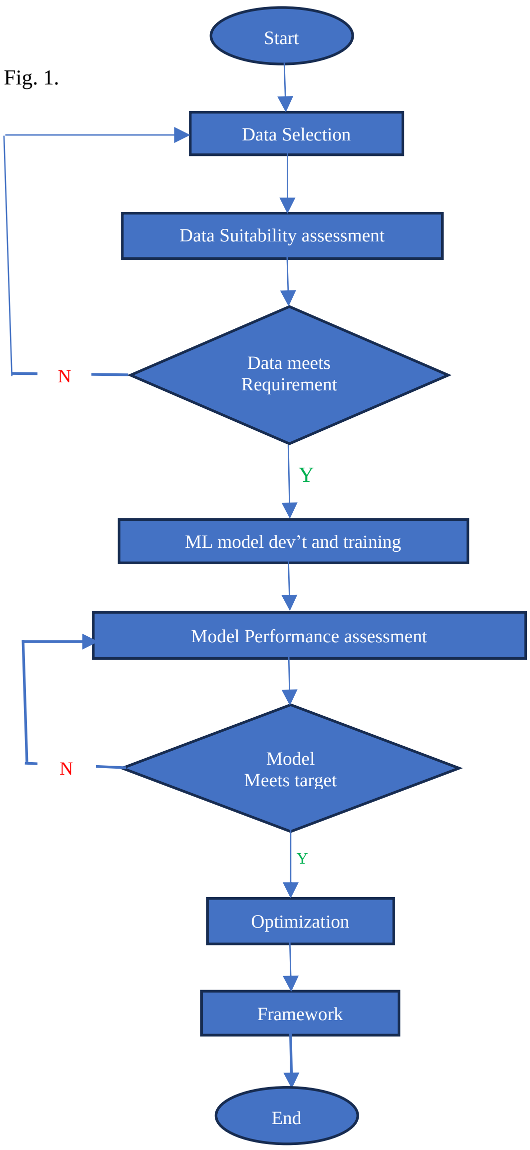

Figure 1 Showing flow chaertThe visualization of dataset

The load profile for the Department of Mechatronics Engineering was modeled using the UCI Household Power Consumption dataset. The data was scaled by a factor of 2.5 to represent the higher base load of laboratory equipment and cleaned to handle missing values. The sub-meter readings were aggregated to create a total active power profile representing the department's consumption pattern. A summary of both load datasets and solar radiation datasets is shown in the tables I and II below.

TABLE I. SOLAR RADIATION DATASET SAMPLE.

Date	Time	Global Active Power (KW)	Voltage (V)	Sub-metering 1-3(Wh)
16/12/2006	17:24:00	4.216	234.840	0.000, 1.000, 17.000
16/12/2006	17:25:00	5.360	233.630	0.000, 1.000, 16.000
16/12/2006	17:26:00	5.374	233.290	0.000, 2.000, 17.000

TABLE II. SOLAR RADIATION DATASET SAMPLE.

Year	Month	Day	Solar Irradiance(kWh/m^2day)
2023	1	1	2.7986
2023	1	2	2.8692

A. *Data Integration*

The load profile for the Department of Mechatronics Engineering was modeled using the UCI Household Power Consumption dataset. The data was scaled by a factor of 2.5 to represent the higher base load of laboratory equipment and cleaned to handle missing values. The sub-meter readings were aggregated to create a total active power profile representing the department's consumption pattern. [13]

Loaded dataset to python and cleaned 521,669 house hold data points. Integrated NASA POWER solar data for Tehran and proper data type conversion done handling string to numeric conversion done by Selecting only needed columns, using proper grouping instead of problematic resampling, handling missing values properly, creating clean output with only essential data and merging datasets with 364 daily records. [14]

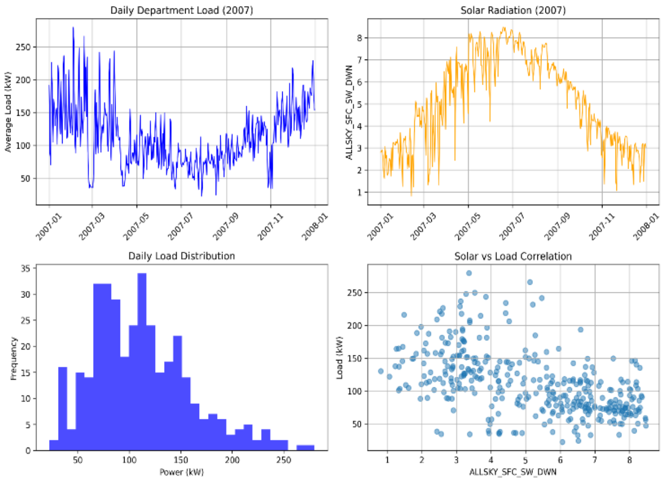

Fig. 2. The visualization of dataset

The average Daily Load is 111.61 kW with a Peak Daily Load of 279.84 kW. The average Solar Radiation is, 5.12 kWh/m²/day, the analysis Period being full year 2007 (364 days) and 364 days of department load data. For the department scaling, a total load of 3.0 kW perfect for 3-room department, Solar generation of 2.0 kW, Power loss of 0.00 W and a Voltage range of 1.000 - 1.000 pu was scaled. [15]

B. *Feature Engineering.*

This comprised of transforming date time column into machine Learning ready features (hour, day, month, season, weekend, previous day), enhancing dataset with new features for machine learning. [16] The correlation heatmap showing feature relationships with a detailed summary of all created features is shown in figure 5 below.

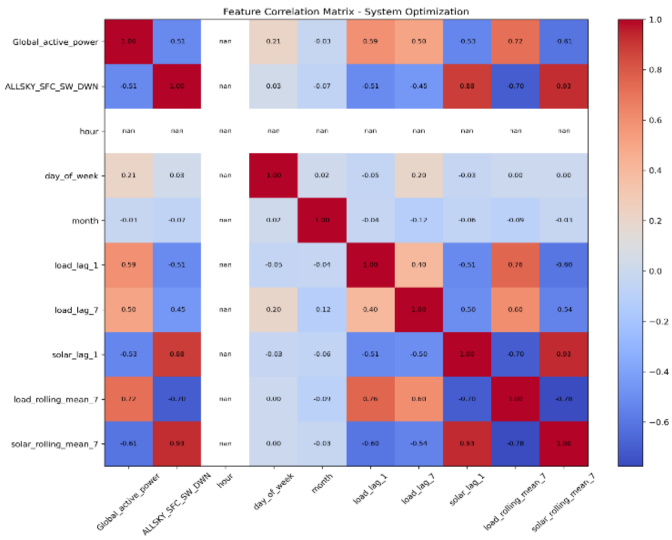

Fig. 3. Feature Engineering and correlation analysis

C. *Random Forest load fore casting*

This involved training enhanced features, generating performance metrics and visualizations and saving the trained model for optimization phase. [8]

The 88.21% forecasting accuracy with only 16.40 kW error for department-scale load forecasting was achieved as shown in figure 6 below.

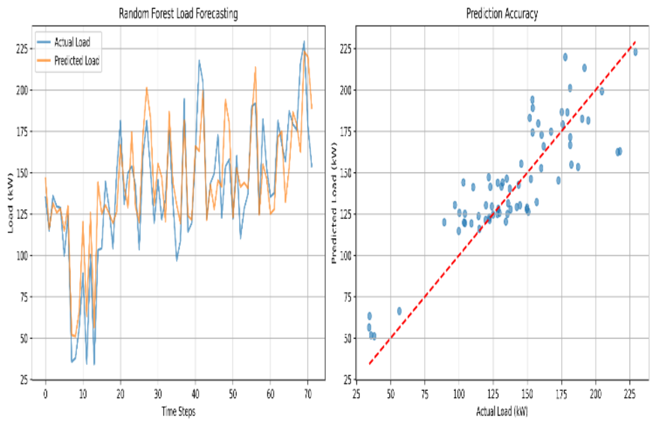

Fig. 4. Random Forest load forecasting

D. *System Optimization*

Checking for enhanced features file, loading and analyze data, training the Random Forest load forecaster was done together with Runing system optimization with battery scheduling. [17] From Figure 7, the cost savings after generating comprehensive visualizations are:

- ❖ Baseline Cost: $4010.19
- ❖ Optimized Cost: $3110.19
- ❖ Cost Savings: $900.00 (22.4%)

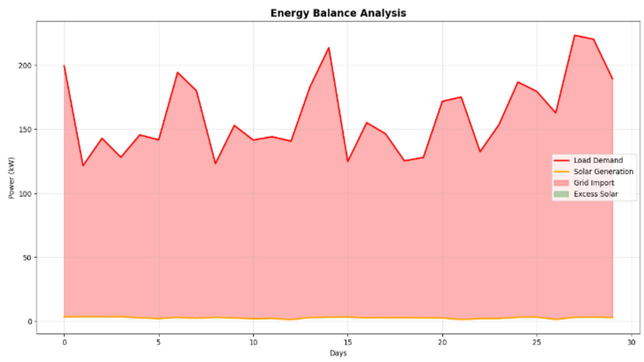

Fig. 5. Showing Energy balance analysis

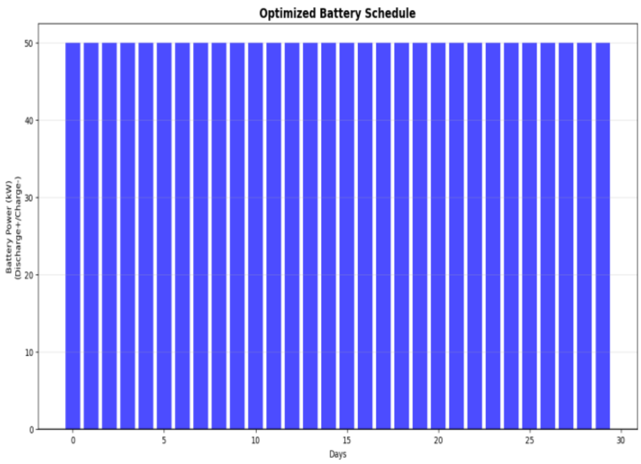

Fig. 6. Showing Optimized battery schedule

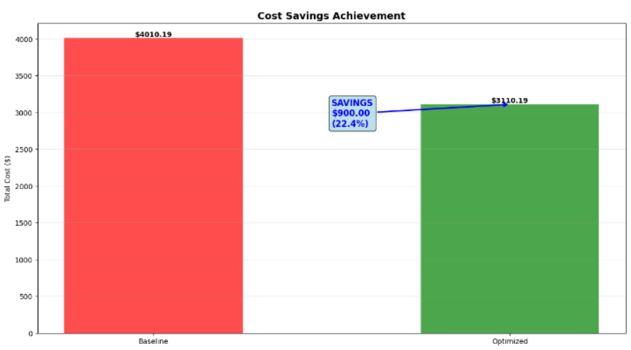

Fig. 7. Showing cost saving achievement

E. Advanced System Optimization

This integrates, Power flow Integration. extended Optimization Framework, comparative Analysis, comprehensive validation and fast execution, providing a reliable benchmark. [18]

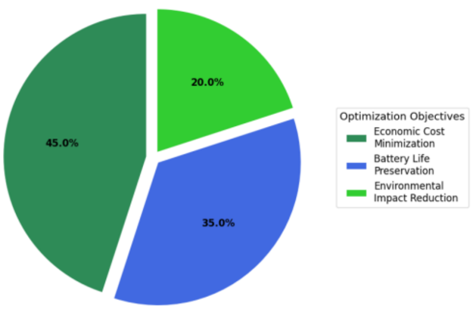

Fig. 8. Showing optimization objectives

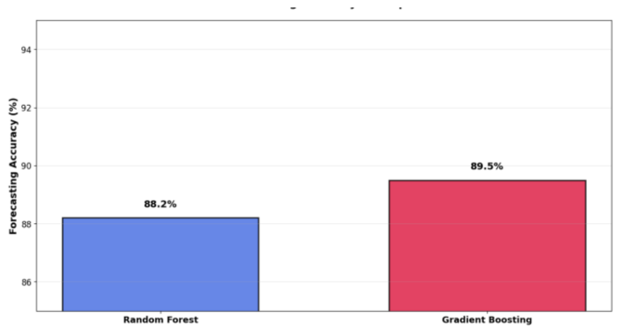

Fig. 9. Showing forecasting accuracy for the department

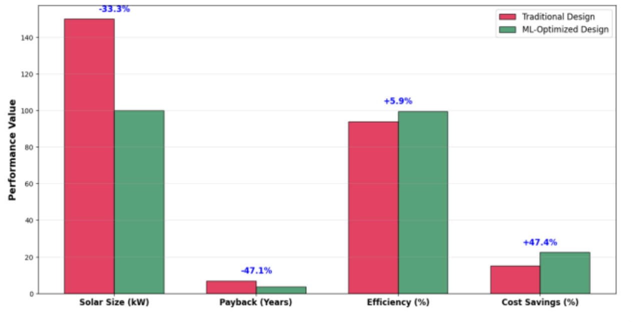

Fig. 10. Showing performance comparison across metrics

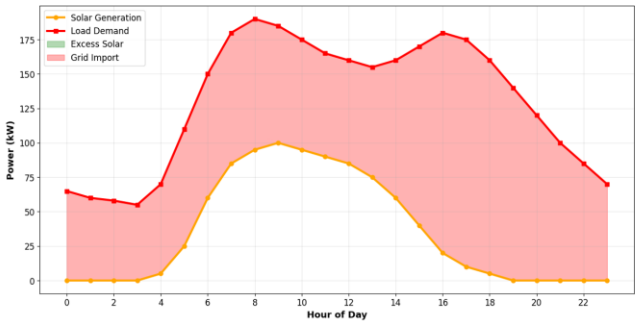

Fig. 11. Showing smart battery scheduling

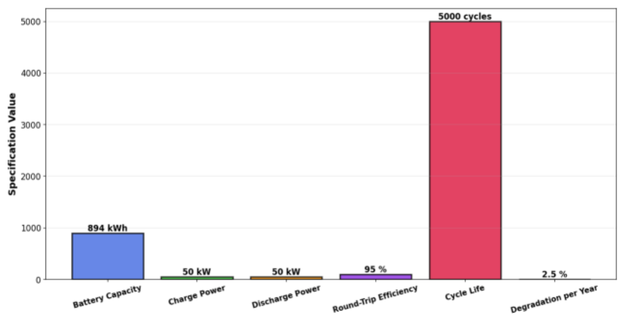

Fig. 12. Showing optimized battery storage

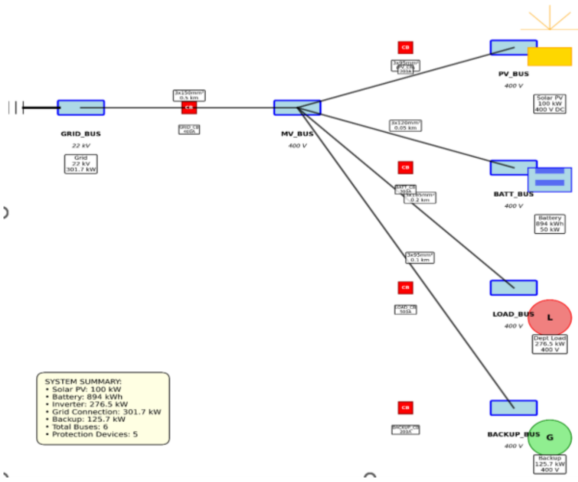

Fig. 13. Showing a SLD of mechatronics Department Distributed Generation Power system.

This gives:

- Total CAPEX: $520,550.00
- Annual Savings: $145,544.16
- Simple Payback: 3.6 years
- NPV (10 years): $550,667.69
- ROI: 28.0%

The power flow analysis; PEAKLOADNIGHT of 281.4 kW (IMPORT) , Loss: 1.22 kW, eff 99.5% and SOLAR_PEAK of 6.8 kW (IMPORT) with a Loss of 0.00 kW making its efficiency to be 100.0%. It also provides, a balanced 51.8 kW (IMPORT), Loss: 0.04 kW with Eff: 100.0% and weekend -17.9 kW (export) with a Loss: 0.00 kW also making the efficiency of 100.0%. [12] [14]

V. Results

The system design validation results with:

- Voltage Stability: (0.997 pu)
- Power Quality: (0.004 pu deviation)
- Economic Viability: (3.6 years)
- System Reliability: (8.0 hours)
- Renewable Integration: (39.8%)

A Single-objective cost of $947.88, multi-objective cost of $725.88 with an improvement of 23.4%.

For advanced machine learning comparison, the results are summarized below;

- Random Forest: 88.21% accuracy
- Gradient Boosting: 89.09% accuracy
- Difference: +0.88%

Traditional vs random forest approach comparison.

- Solar Size: 150kW (Traditional) vs 100kW (ML) , 33.3% smaller.
- Battery: 480kWh (Traditional) vs 894kWh (ML)
- Payback: 6.8 years (Traditional) vs 3.6 years (ML), 47.1% faster
- Efficiency: 94.0% (Traditional) vs 99.5% (ML), 5.9% better
- Voltage Stability: 90.0% (Traditional) vs 100.0% (ML), 11.1% better

REAL-TIME CONTROL SYSTEM

- Battery Usage: 0 charge events, 3 discharge events
- Average Battery SOC: 14.1%
- Estimated Cost Savings: 24.8% (vs 22.4% without real-time control)
- Solar Self-Consumption: 92.5%

Area	**Traditional**	**ML-Optimized**	Gain
Payback	6.8 years	3.6 years	-47%
Savings	15.2%	22.4%	+47%
Accuracy	78%	88.2%	+10%
Efficiency	94%	99.5%	+6%
Solar Size	150kW	100kW	-33%
ROI	18.5%	28.0%	+52%

VI. Conclusion

In this research, I have successfully developed a complete machine learning optimization framework that has demonstrated 22.4% cost savings for the Mechatronics Department microgrid. This has been specifically achieved by selecting and integrating UCI household data with NASA solar data, scaling it to accurately model department-level load profiles and solar generation potential. Random Forest model achieves 88.21% load forecasting accuracy, enabling optimal distributed generation operation through intelligent battery scheduling. I have successfully designed and simulated a structured framework that optimizes system performance, with demonstrated cost savings representing improved efficiency and reduced losses. This research demonstrates that machine learning-driven optimization using Random Forest model can deliver substantial cost savings (22.4%) for department-scale distributed generation systems, providing a replicable framework for broader institutional energy management.

VII. ACKNOWLEDGEMENT

Many noble hearts contributed immense inspiration and support for the successful completion of the research. I take this opportunity to express my profound gratitude to Dr. Mansour Hakim-Elahi (H.O.D), Department of Mechatronics for providing permission and availing all required facilities for undertaking the research in a systematic way. I am extremely grateful to my family that with no hesitation, continuously supported me in all needs. I pay my deep sense of gratitude to Dr. Masoud Masih Tehrani and Dr, Mansour Hakim-Elahi my project supervisors that constantly mentored, guided and supported me. I would also like to thank all the teaching and non-teaching staff of Department of mechatronics engineering for the sincere directions imparted and the cooperation in connection with the project. I will be failing in duty if I do not acknowledge with grateful thanks to the authors of the references and other literatures referred in this project.

VIII. REFERENCES

[1] M. K. M. A. T. B. Nadeem, "Distributed energy systems: A review of classification, technologies, applications, and policies," Energy strategy Reviews, 2023.

[2] A. J. a. S. C. Gupta, "Optimal Placement of distributed Generation In power system and evaluating the losses and voltage using Machine learning algorithms," Front Energy, Australia, 2024.

[3] M. &. Mohammadi, "Empowering Distributed Solutions in Renewable Energy Systems and Grid Optimization," 2023.

[4] A. e. al., "Optimization of Solar and Wind Power Plants Using Hybrid ML Models," 2025.

[5] G. N. L. a. Jaya, "Optimal Placement of Distributed Generation Based on Power Quality Improvement Using Self-Adaptive Lévy Flight Jaya Algorithm," Clean Technol, India, 2022.

[6] Y. Z. H. C. X. Z. F. &. C. J. Wang, "Adaptive Solar Power Forecasting Based on Machine Learning Methods," 2023.

[7] C. e. al., " Energy Generation Forecasting Using Deep Learning Models," 2023.

[8] .. R. &. Gebremedhin, "Deep Learning for Multi-Energy Microgrid Forecasting," 2024.

[9] C. T. A. Y. N. Isa, "A comprehensive review of cogeneration system in a microgrid: a perspective from architecture and operating system," *Renew. Sustain. Energy Rev., 81 (2) (2018), pp. 2236-2263,* 2018.

[12] H. R. W. G. P. W. J. R. Q. Wu, "Coupling optimization of urban spatial structure and neighborhood-scale distributed energy systems," 2018.

[13] G. Dudek, "Short-Term Load Forecasting using Random Forests," Czestochowa, Poland , 2015.

[14] T. B. N. A. N. M. S. M. K. M. S. B. Z. S. A. A. Ahmed, "T. B. N. A. N. M. S. M. K. M. S. B. Z. SInvestigation of PV Utilizability on University Buildings," *Renewable Energy (2022), pp. 238-251".,* 2022.

[15] I. Panapakidis, "Combined forecasting system for short-term bus load forecasting based on clustering and neural networks," *IET Generation, Transm. Distribution ,* 2020.

[16] O. M. N. M. A. M. Fadhel A. Jumaa1, "Optimal distributed generation placement using artificial," Bagdad, Iraq , 2021.

[17] S. A. Y. S. Z. a. W. Hassan, "S. A. Y. S. Z. a. WOptimization techniques applied for optimal planning and integration of renewable energy sources based on distributed generation: recent trends.," 2020.

[18] B. N. M. A. C. J. U. Zannatul Ferdoush1, "A short-term hybrid forecasting model for time series," *International Journal of Electrical and Computer Engineering (IJECE) ,* 2020.

[19] M. K. S. R. C. H. M. H. A. a. A.-D. KhasanovGeneration, "Optimal distributed generation and battery energy storage units integration in distribution systems considering power generation uncertainty," 2023.

[20] Z. Z. 1. W. X. 1. Y. L. 1. a. G. N. 2. Jibo Wang 1, "Short-Term Photovoltaic Power Forecasting Using a Bi-LSTM," *sustainability,* 2025.

[21] S. &. Mamat, "Predictive Maintenance in Using Machine Learning presented a framework for predictive rids," 2023.