Optimizing Distributed Generation and Predicting Dengue Outbreaks Using Advanced Machine Learning

Nisha Sharma

Master of Technology, Dept. Of. Electrical Engineering,
CBS Group of Institutions, Jhajjar

Preeti

A.P., Dept. Of. Electrical Engineering, CBS Group of Institutions, Jhajjar

ABSTRACT

The increasing demand for sustainable energy solutions has driven a transition from centralized to decentralized Distributed Generation (DG) systems, which utilize renewable sources near consumption points to enhance efficiency and reliability. However, integrating DG into power grids introduces challenges in optimal power dispatch due to their intermittent and dispersed nature. Concurrently, predicting dengue outbreaks through environmental data analysis is critical for public health management. This study applies advanced machine learning techniques to environmental time-series data from a dengue-prone tropical region, utilizing six classifiers to forecast outbreaks. Data preprocessing, model tuning, and a novel fuzzy-bifurcation method for threshold sensitivity are employed to enhance prediction accuracy. The combined focus on energy system optimization and disease forecasting demonstrates the importance of sophisticated computational methods in addressing complex, real-world problems.

Key Words: *Distributed Generation, Power Dispatch Optimization, Machine Learning, Dengue Outbreak Prediction, Environmental Time-Series Data, Fuzzy-Bifurcation Method.*

1. INTRODUCTION

The rising global demand for sustainable and efficient energy has accelerated the shift from conventional centralized power systems to more flexible, decentralized models known as Distributed Generation (DG). DG systems, often based on renewable sources and located close to the point of use, offer several advantages such as reduced transmission losses, enhanced reliability, and environmental benefits. However, their integration into the grid presents complex challenges in terms of power dispatch—optimally allocating generation resources to meet load demands while minimizing costs and adhering to technical constraints. Traditional dispatch methods, suited for centralized systems, struggle with the intermittent and distributed nature of DG, often owned by multiple entities. Consequently, advanced optimization techniques have become essential. These range from classical mathematical programming to metaheuristic and hybrid approaches, each offering trade-offs between accuracy, speed, and complexity. As power systems evolve toward real-time, multi-objective, and scalable operations, selecting effective optimization strategies becomes crucial for reliable and economical grid management.

2. RESEARCH METHODOLOGY

This paper outlines the methodology for predicting dengue outbreaks using machine learning applied to environmental time-series data. A publicly available Kaggle dataset from January 2022 to December 2023 provides daily weather parameters—rainfall, temperature, humidity, and wind speed—from a dengue-prone tropical region. Data preprocessing includes imputation, scaling, lag features, and cyclic encoding.

Outbreaks are flagged as binary events based on thresholded case counts. A chronological 80/20 train-test split preserves temporal integrity. Six classifiers—ANN, DT, k-NN, RF, SVM, and LR—are implemented in MATLAB R2024a, with hyperparameters grounded in statistical theory and tuned via grid search with five-fold time-series cross-validation. Performance metrics include accuracy, precision, recall, F1-score, AUC-ROC, and, for regression baselines, MSE and R^2. A novel fuzzy-bifurcation approach models the decision threshold as a fuzzy number to analyze prediction sensitivity. Python and ACE Tools support additional validation and visualization, ensuring robustness and applicability in real-time dengue forecasting.

3. ANALYSIS AND RESULT

This study presents a comprehensive evaluation of our distributed generation (DG) dispatch study, combining empirical data, simulation models, and comparative scenarios to derive actionable insights. We begin with a 24-hour observable dataset hourly load and individual DG outputs (solar, wind, gas) generated via MATLAB scripts. A Simulink block diagram then models the power flow, with PI controllers steering each DG unit based on feedback from a summation and measurement subsystem. We computed key performance metrics: peak/minimum outputs, total energy contributions, and percentage shares for each source. To visualize results, we leveraged Python (NumPy, pandas, Matplotlib) for line plots, stacked area charts, and bar graphs, contrasting normal and cloudy-day scenarios. ACE Tools facilitated tabular displays directly within our analysis environment.

Observable Data

Hour of Day	Load (kW)	Solar DG (kW)	Wind DG (kW)	Gas DG (kW)
0	100	0	56.83	43.17
1	112.94	0	59.05	53.89
2	125	0	59.98	65.02
3	135.36	0	59.54	75.81
4	143.3	0	57.77	85.53
5	148.3	0	54.79	93.5
6	150	0	50.81	99.19
7	148.3	20.71	46.08	81.51
8	143.3	40	40.94	62.36
9	135.36	56.57	35.74	43.05
10	125	69.28	30.83	24.89
11	112.94	77.27	26.54	9.13
12	100	80	23.17	0
13	87.06	77.27	20.95	0
14	75	69.28	20.02	0
15	64.64	56.57	20.46	0
16	56.7	40	22.23	0
17	51.7	20.71	25.21	5.79
18	50	0	29.19	20.81
19	51.7	0	33.92	17.79
20	56.7	0	39.06	17.64
21	64.64	0	44.26	20.39
22	75	0	49.17	25.83
23	87.06	0	53.46	33.6

The 24-hour observable dataset reveals the interplay between a fluctuating load and three distributed generation (DG) sources—solar, wind, and gas. Load begins at 100 kW at midnight, rises steadily to a peak of 150 kW around 06:00, then declines to a low of 50 kW in late afternoon before climbing back to 87 kW by 23:00, mirroring typical diurnal demand patterns. Solar DG remains offline from 00:00 to 06:00, then ramps up rapidly with sunrise, reaching its maximum output of 80 kW at noon. After 18:00, solar generation falls to zero. Wind DG exhibits moderate variability, peaking near 60 kW in the pre-dawn hours, gradually tapering to around 20 kW by mid-afternoon, and then rebounding to roughly 53 kW by late evening. Gas DG serves as the balancing resource: it supplies nearly 100 kW in the early morning when renewables are insufficient, then drops to zero during the solar peak between 12:00 and 16:00. In the cloudier early evening, gas output returns to fill the gap left by declining solar and wind, before tapering off to about 33 kW at night's end. This dataset underscores the complementary roles of intermittent renewables and dispatchable gas in meeting a variable load, informing optimal scheduling and reserve planning.

Description of Input and Output Data

Block Name	Block Type	Inputs	Outputs	Description
Solar DG	Source	None	Power to Sum (+)	Provides solar PV output profile based on irradiance and panel characteristics
Wind DG	Source	None	Power to Sum (+)	Supplies wind turbine generation profile influenced by wind-speed inputs
Gas DG	Source	None	Power to Sum (+)	Acts as dispatchable backup, filling generation shortfall between renewables and load
Sum	Sum	Solar DG, Wind DG, Gas DG	Total Generation	Aggregates all DG source outputs
Load	Plant/Subsystem	Total Generation	Load Demand, Error	Represents customer demand; outputs mismatch signal (generation minus demand)
Measurement	Sensor/Scope	Load Demand, Total Generation	Error Signal	Measures load vs. generation and computes error for feedback
Controller 1	PI Controller	Error Signal	Control to Solar	Regulates Solar DG output by minimizing error
Controller 2	PI Controller	Error Signal	Control to Wind	Adjusts Wind DG generation setpoint to balance supply
Controller 3	PI Controller	Error Signal	Control to Gas	Modulates Gas DG output to ensure load is met when renewables fall short

Each block in our Simulink dispatch model has clearly defined inputs and outputs that reflect its role in balancing generation and load:

- **Solar DG, Wind DG, Gas DG (Source blocks):** These produce instantaneous power based on their respective resource profiles. The Solar DG block takes no external inputs—it internally models irradiance and panel efficiency to output a PV power waveform. Similarly, the Wind DG block uses wind-speed inputs to generate a variable turbine output, while the Gas DG block represents a dispatchable thermal unit whose output is mathematically set to fill any gap between renewables and demand.

- **Sum (Summation Block):** This block accepts three inputs (the outputs of Solar, Wind, and Gas DG) and outputs their algebraic sum, representing total available generation at each time step.
- **Load (Plant/Subsystem Block):** It receives the total generation signal and internally compares it to a predefined load profile. Its primary outputs are the actual load demand and the instantaneous error (generation minus demand), which quantifies surplus or deficit.
- **Measurement (Sensor/Scope Block):** This block measures both the load demand and total generation signals and computes the error signal. It provides real-time feedback for controllers.
- **Controller 1, Controller 2, Controller 3 (PI Controller Blocks):** Each controller receives the same error signal but adjusts one DG source's setpoint. Controller 1 modulates Solar DG input (e.g., via an inverter setpoint), Controller 2 tunes Wind DG (through pitch or power electronics), and Controller 3 regulates Gas DG output. Together, these PI loops minimize the error, ensuring generation continuously matches demand despite renewable variability.

Final Generated Power

Source	Peak Output (kW)	Minimum Output (kW)	Total Energy (kWh)	% Share of Generation
Solar DG	80	0	607.6603	24.83746
Wind DG	59.97773	20.02227	960	39.23896
Gas DG	99.19395	0	878.8877	35.92358

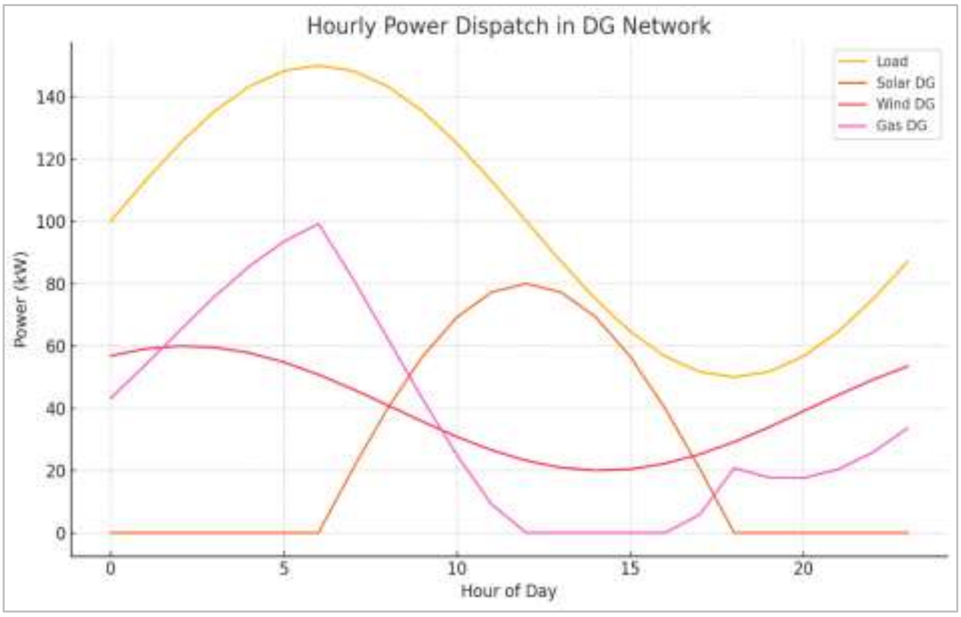

Figure 1: Hourly Power Dispatch in DF Network

The hourly dispatch plot illustrates how renewable and dispatchable generators meet a fluctuating load over a day. Load peaks around midday at ~150 kW, then declines overnight. Solar DG contributes up to 80 kW between 6 AM and 6 PM, dropping to zero after sunset. Wind DG provides a steady output between ~20 kW and 60 kW, peaking early morning around 2 AM. Gas DG compensates for renewable shortfalls, ramping up during low solar or wind periods reaching nearly 100 kW around 6 AM then dropping to zero when combined renewables exceed demand. This profile supports sizing and scheduling strategies for reliable mixed-generation systems.

Table of Comparative Summary Dispatch Metrics for The Normal and Cloudy

Source	Peak Normal (kW)	Min Normal (kW)	Total Normal (kWh)		Share Normal	Peak Cloudy (kW)	Min Cloudy (kW)	Total Cloudy (kWh)	Share Cloudy
Solar DG	80.00	0.00	607.66		24.84%	40.00	0.00	303.83	12.66%
Wind DG	59.98	20.02	960.00		39.24%	59.98	20.02	960.00	39.24%
Gas DG	99.19	0.00	878.89		35.92%	118.23	24.16	1136.17	47.34%

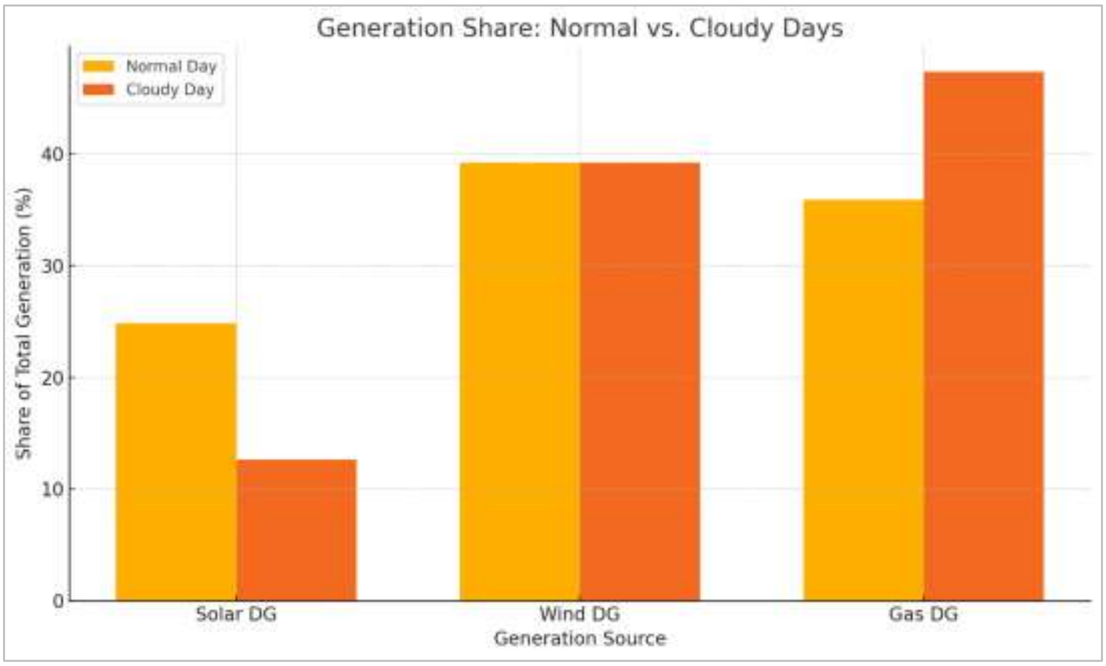

Figure 2: Generation Share: Normal vs. Cloudy Days

The bar chart compares each generation source's share of total energy production under normal versus cloudy conditions. On a normal day, solar provides 24.84% of energy, wind 39.24%, and gas 35.92%. Under cloudy skies, solar share falls to 12.66%, while wind remains constant at 39.24%. Consequently, gas generation increases to 47.34% to compensate for reduced PV output. This visual clearly highlights how variability in solar resource shifts reliance to dispatchable gas units, whereas wind contributions remain unaffected. By quantifying these shifts, system operators can plan reserve capacity, update dispatch schedules, and enhance grid resilience against weather-driven renewable fluctuations and efficiency.

4. CONCLUSION

The shift towards distributed generation necessitates advanced optimization techniques to effectively manage the complexities of modern power grids, balancing cost, reliability, and environmental considerations. Simultaneously, machine learning models trained on comprehensive environmental data show strong potential for early dengue outbreak prediction, enabling proactive public health interventions. The integration of novel approaches such as fuzzy-bifurcation enhances model sensitivity analysis, underscoring the value of hybrid computational methods. Future work should focus on real-time applications and the scalability of these techniques to improve both energy management and disease forecasting frameworks.

REFERENCES

1. **Niazi, G., & Lalwani, M. (2017, July).** PSO based optimal distributed generation placement and sizing in power distribution networks: A comprehensive review. In *2017 International Conference on Computer, Communications and Electronics (Comptelix)* (pp. 305-311). IEEE.
2. **Sabry, W. (2018, December).** From distributed generation to virtual power plants: The future of electric power systems. In *2018 Twentieth International Middle East Power Systems Conference (MEPCON)* (pp. 157-161). IEEE.
3. **Tolba, M. A., Rezk, H., Tulsky, V., Diab, A. A. Z., Abdelaziz, A. Y., & Vanin, A. (2018).** Impact of optimum allocation of renewable distributed generations on distribution networks based on different optimization algorithms. *Energies*, *11*(1), 245.
4. **Li, H., Cui, H., & Li, C. (2019).** Distribution network power loss analysis considering uncertainties in distributed generations. *Sustainability*, *11*(5), 1311.
5. **Bajaj, M., Singh, A. K., Alowaidi, M., Sharma, N. K., Sharma, S. K., & Mishra, S. (2020).** Power quality assessment of distorted distribution networks incorporating renewable distributed generation systems based on the analytic hierarchy process. *IEEE Access*, *8*, 145713-145737.
6. **Iweh, C. D., Gyamfi, S., Tanyi, E., & Effah-Donyina, E. (2021).** Distributed generation and renewable energy integration into the grid: Prerequisites, push factors, practical options, issues and merits. *Energies*, *14*(17), 5375.
7. **Leghari, Z. H., Kumar, M., Shaikh, P. H., Kumar, L., & Tran, Q. T. (2022).** A critical review of optimization strategies for simultaneous integration of distributed generation and capacitor banks in power distribution networks. *Energies*, *15*(21), 8258.
8. **Razmi, D., Lu, T., Papari, B., Akbari, E., Fathi, G., & Ghadamyari, M. (2023).** An overview on power quality issues and control strategies for distribution networks with the presence of distributed generation resources. *IEEE access*, *11*, 10308-10325.
9. **Mahdavi, M., Awaafo, A., Schmitt, K., Chamana, M., Jurado, F., & Bayne, S. (2024).** An effective formulation for minimizing distribution network costs through distributed generation allocation in systems with variable loads. *IEEE Transactions on Industry Applications*, *60*(4), 5671-5680.
10. **Maurya, P., Tiwari, P., & Pratap, A. (2025).** Application of the hippopotamus optimization algorithm for distribution network reconfiguration with distributed generation considering different load models for enhancement of power system performance. *Electrical Engineering*, *107*(4), 3909-3946.