ISBN: 978-628-96613-3-0. ISSN: 2414-6390. Digital Object Identifier: https://dx.doi.org/10.18687/LACCEI2026.1.1.1610

Hybrid Statistical Machine Learning Framework for Demand Modeling and Intelligent Monitoring in Distribution Systems.

Isaac Fernando Hernández Urbina[1], Oizis Aksunamun Melgar Rodríguez[2], Jonathan Muñoz Tabora[3], Ozy D. Melgar Dominguez[4]
[1]Universidad Nacional Autónoma de Honduras, *isaac_hernandez@unah.hn*
[2]Universidad Nacional Autónoma de Honduras, *oizis.melgar@unah.hn*
[3]Universidad Nacional Autónoma de Honduras, *Jonathan.munoz@unah.edu.hn*
[4]Universidad Nacional Autónoma de Honduras, *ozzydamedo@gmail.com*

***Abstract*—Distribution utilities need high-resolution demand modeling and voltage-condition monitoring. In practice, traditional state estimation requires accurate topology and line parameters that are not always available. This study presents a hybrid statistical-machine learning workflow for the CDA–L273 distribution feeder in Tegucigalpa, Honduras. From 96,486 15-minute records, which resulted in 72,548 validated active-power measurements (Jan 2023-Dec 2025), with only 0.05% outliers. The feeder exhibits stable operation, with a mean demand 7,082 W (peak 11,485 W), voltage stability of 1%, and a consistently power factor 0.96. Time-series analysis reveals strong short-term autocorrelation (r=0.90 at one 15-minute step) and repeatable daily/seasonal patterns. For demand prediction, a multilayer perceptron trained on 23 engineered features (lags, rolling statistics, and cyclical encodings) achieved $R^2 = 1.0000$, RMSE of 6–8 W, and MAE of 5–6 W ($< 0.1\%$ of the mean load). In parallel, an MLP-based model-free mapping from P,Q,PF to phase voltages supports lightweight monitoring without network models, with typical test errors of 0.2–0.6%. Overall, the proposed framework provides a practical, low-cost basis for planning and operational decision-making in data-sparse feeders.**

***Index Terms*—Artificial intelligence, Multilayer perceptrons, Power system analysis, Machine Learning, State Estimation.**

I. Introduction

Voltage condition monitoring is a central requirement in modern distribution networks because voltage deviations and fluctuations are among the most visible manifestations of power quality problems at the customer side [1]. This need has intensified with the growth of variable demand, distributed energy resources, and power electronic based loads and converters, which increase operating variability and can exacerbate local voltage issues. In parallel, most distribution feeders remain sparsely instrumented and therefore offer limited observability of nodal voltages across space and [2] [3].

In practice, utilities frequently rely on model based monitoring functions embedded in distribution management systems. However, the effectiveness of these functions is constrained by the quality of network information. In particular, state estimation in distribution systems typically assumes known topology, correct phase connectivity, and up to date line parameters, while real feeders often experience unreported switching actions, outdated geographic information system records, parameter uncertainty, and incomplete telemetry [4].

Under these conditions, real-time voltage monitoring becomes difficult to sustain with consistent accuracy, especially in data-sparse feeders where the measurement set is not redundant and pseudo measurements must be introduced to achieve observability [5]. These practical limitations motivate complementary monitoring strategies that can operate reliably with minimal network knowledge and readily available operational measurements.

The operation of distribution networks has shifted from a largely passive paradigm to an increasingly active one, driven by distributed energy resources (DERs) and emerging loads such as electric vehicles (EVs) [6], [7].This transition increases voltage variability and reverse power flows, while low sensing density and the high cost of instrumentation continue to limit observability in low- and medium-voltage feeders. Consequently, traditional physics-based state estimation, which depends on accurate topology and line parameters, is often difficult to scale and maintain. This has motivated a growing body of data-driven and model-free methods that infer voltage states from available measurements with reduced dependence on detailed feeder models.

Early work emphasized maximizing information from scarce measurements through optimal metering strategies. In particular, intelligent meter placement has been explored to improve bus-voltage estimation in weakly instrumented distribution configurations, showing that careful sensor allocation can partially compensate for limited observability [8]. As advanced metering infrastructure became more common, research increasingly pursued direct voltage inference from operational data. Neural estimators have been proposed for active distribution grids to provide voltage monitoring without explicit feeder modeling, aiming at robustness under embedded generation and complex load behavior [9]. Complementarily, hierarchical bottom-up ANN schemes demonstrated that voltage estimation can be performed through direct mappings, avoiding convergence issues associated with iterative solvers and enabling execution on low-cost hardware for near-real-time monitoring [10].

Within model-free estimation, a dominant line learns direct mappings from commonly available electrical variables to voltages. Smart-meter-driven approaches have been validated across many real-world feeders and evaluated under increasing EV and photovoltaic penetrations, highlighting their practicality for large-scale deployment in data-sparse contexts [11]. In parallel, ANN-based estimation using micro-PMU measurements integrates optimized sensor placement to balance cost and accuracy, showing low estimation errors with a limited number of devices when training scenarios are sufficiently diverse [12]. To improve robustness, some studies enrich the feature space or fuse heterogeneous data sources. Weather-aware ANN estimators incorporate meteorological variables and demonstrate resilience across seasons, DER penetration levels, and missing-data cases [13]. Evidence-theoretic fusion has also been applied to estimate voltage levels in real distribution settings when datasets are deficient, illustrating how cross-domain fusion can close informational gaps when endpoint measurements are unavailable [14].

More recent contributions aim to stabilize performance under regime shifts and reconfiguration by embedding structure and uncertainty into the estimator. Physics-aware neural networks incorporate feeder topology into the architecture to reduce reliance on purely black-box learning and improve scalability, benchmarking favorably against WLS on standard feeders using datasets derived from smart-meter consumption profiles [15]. Probabilistic formulations further couple estimation with operational decision-making: Bayesian voltage inference based on Neural Network Gaussian Processes is integrated into chance-constrained OPF for voltage regulation and evaluated on a real 759-node system, emphasizing uncertainty quantification and adaptability to topology changes [16]. Finally, as estimation becomes a cyber-physical dependency, FDIA-focused studies propose deep-learning pipelines to detect and localize stealth attacks that could compromise voltage monitoring and control [17].

Reliable voltage-condition monitoring in distribution feeders remains challenging because observability is typically limited and conventional state-estimation workflows depend on accurate topology and up-to-date line parameters, which are frequently incomplete or outdated in practice. The recent literature has therefore moved toward data-driven and model-free voltage estimation using available measurements; however, many reported solutions either rely predominantly on simulated datasets, require enhanced sensing infrastructure or structured network information, or exhibit reduced robustness under operating-regime shifts and feeder reconfiguration. These limitations are particularly restrictive in data-sparse feeders, where utilities need lightweight, low-cost monitoring tools that can be deployed without detailed feeder models while still delivering dependable accuracy.

Prior studies highlight a clear gap between recent data-driven voltage estimation methods and practical deployment in data-sparse distribution feeders. Limited observability and incomplete or outdated topology and line-parameter information constrain conventional state-estimation-based monitoring. Although model-free approaches are increasingly reported, many still depend on simulated training data, enhanced sensing, or structured network information, and may degrade under operating-regime shifts and feeder reconfiguration. Consequently, utilities need lightweight, low-cost voltage monitoring solutions that operate with minimal inputs and without detailed feeder models, while maintaining dependable accuracy.

II. Problem Summary

Accurate, near real time voltage monitoring in electrical distribution systems remains challenging due to feeder dispersion and limited sensor deployment, despite IEEE/IEC requirements to keep voltage within ±5% limits [3]. Traditional monitoring and state estimation practices based on power flow models are often impractical because they depend on accurate topology and up to date line parameters, which are frequently incomplete or outdated, and may also impose computational burdens that hinder lightweight deployment. To address these limitations, this study proposes a data-driven, model-free phase voltage estimator based on a Multilayer Perceptron (MLP) [18]. The MLP is trained to learn the nonlinear mapping from a minimal input set real power (P), reactive power (Q), and power factor (PF) to three phase voltages using empirical operating behavior recorded in the field [19].

III. Methodology

This study aims to develop a robust, model free voltage estimator that operates without analytical network models or explicit knowledge of feeder topology. The CDA–L273 circuit is a primary distribution feeder located in the metropolitan area of Tegucigalpa, Honduras. It characterized by a predominantly residential and commercial load profile, serving as a critical node for analyzing urban demand patterns in the central district's electrical grid. The initial dataset was collected from this feeder and comprises approximately 100,000 records of 19 electrical variables sampled at 15 minute intervals over nearly three years (January 2023 to October 2025). A rigorous data cleaning procedure was subsequently applied to remove anomalies and physically inconsistent observations [20].

The final high integrity dataset used for modeling comprised 93,829 validated observations. Overall, the proposed approach is hybrid in nature, combining advanced data conditioning and validation in Wolfram Mathematica with neural network training and inference implemented in Python [21]. To enable supervised learning for phase voltage prediction, the experimental procedure was organized into a sequential six stage workflow, as illustrated in Figure 1. This framework covers the full research workflow, from field data collection and database construction to rigorous data cleaning, neural network integration, and final result analysis. The diagram highlights the key transition from raw data processing to predictive inference, ensuring the voltage model is trained and evaluated only with validated, high quality data.

A. Data Cleansing and Conditioning

A central component of the proposed methodology is the rigorous cleaning, alignment, and validation of the initial

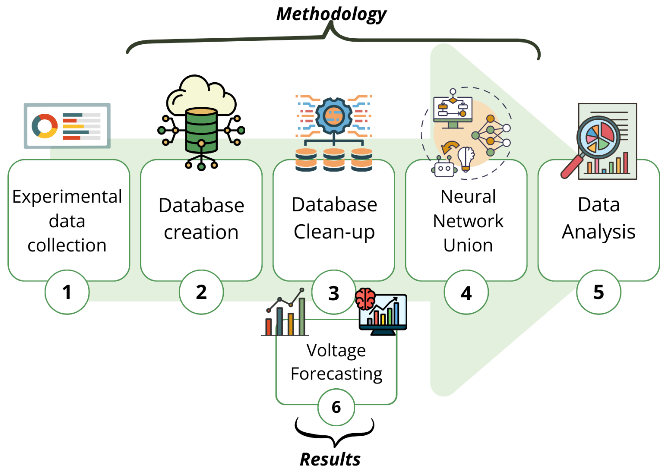

Fig. 1. Proposed methodology for the study.

dataset, which is essential to preserve the physical integrity of the model inputs. This dataconditioning stage is motivated by the presence of anomalous, incomplete, or corrupted records that commonly arise in field measurements due to communication losses, missing values, sensor malfunctions, or equipment outages. In particular, the dataset was screened to identify and remove entire days in which voltage, current, or power measurements simultaneously assumed zero values, a pattern that is physically inconsistent with normal feeder operation and therefore indicative of measurement or logging failures [22]. Consequently, of the approximately 100,000 original samples, inconsistent records were discarded, resulting in a final set of 93,829 valid and aligned observations. This process not only reduced the size of the dataset, but also ensured that the MLP model was trained solely on the actual and physically consistent operational behaviour of the system.

The theoretical basis of the work is rooted in the domain of Supervised Learning, utilising a Multilayer Perceptron (MLP) architecture. This architecture is meticulously designed to facilitate the mapping of complex nonlinear functions between power inputs and voltages. The fundamental operation of the network is described by the activation function of each neuron ”j” [23]. In this particular instance, the output y_j is derived as a weighted sum of the inputs x_i, subsequent to an activation function f, which in this case is ReLU, as presented in (1).

$$y_j = f\left(\sum_{i=1}^{n} w_{ij}x_i + b_j\right) \tag{1}$$

Prior to the training process, the data underwent a normalisation process in order to standardise the characteristics and facilitate algorithm convergence. The formula for normalization is as follows in (2):

$$x_{norm} = \frac{x-\mu}{\sigma} \tag{2}$$

Finally, the optimization of the network weights during training was achieved by minimising the Mean Squared Error (MSE) cost function. The MSE cost function is a standard metric for regression problems that evaluates the difference between the actual voltage values. (y_i) and the predicted values as presented in (3) [24]:

$$(\hat{y}_i) : MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{3}$$

B. Proposed Model for State Stimation

To estimate voltages in a real distribution feeder without relying on powerflow models or explicit topology information, this study employs a Multilayer Perceptron (MLP) trained to learn the nonlinear mapping from historical operating measurements. The network uses a 128–128–128 architecture, as presented in Figure 2. The implementation of the rectified linear unit (ReLU) activation function is employed with a model using real power P, reactive power Q, and power factor PF as inputs to predict the three phase to neutral voltages $V_{A-N}, V_{B-N}, V_{C-N}$. This offers a lightweight and adaptable solution for real time monitoring [25].

IV. Experimental Procedure

This section delineates the comprehensive methodology employed for the acquisition, cleansing, analysis, and modelling of data from the CDA–L273 circuit. The work was structured into four phases. The first phase was the acquisition of data. The second phase was the preprocessing of data in Wolfram Mathematica. The third phase was the statistical analysis in Python. The fourth and final phase was the predictive modelling using a neural network [26], [27].

A. Hardware and software used

The data measurements come from a three phase digital analyzer integrated into the feeder monitoring system CDA–L273. Every 15 minutes, this equipment records the main electrical quantities of the system: active, reactive, and apparent power; energy per interval; phase voltages; currents; power factor; frequency; and other relevant electrical parameters.

Two complementary computing environments were utilised for the purpose of data processing. Wolfram Mathematica was utilised for the importation of data, its format conversion,

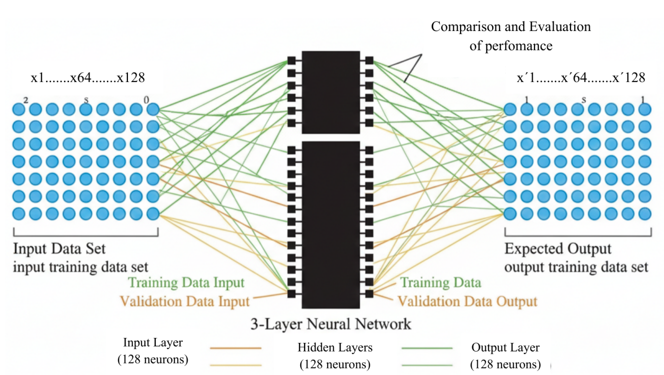

Fig. 2. Neural Network Architecture

the standardisation of timestamps, the creation of dataset structures, and the initial cleansing of the data. The Python programming language was utilised for advanced statistical analysis, high quality visualisation, the generation of derived features, and machine learning model training. The primary libraries were pandas, NumPy, SciPy, scikit–learn, Matplotlib y Seaborn [28].

B. Experiment start up parameters

The study period encompasses measurements from January 2023 to October 2025, with the 15 minute resolution of the measurement system being maintained consistently throughout [29]. It is noteworthy that all observations correspond to the node identified as CDA-L273, thereby negating the necessity for location filters. The original file comprised 19 columns pertaining to energy, power, voltage, current, frequency, and auxiliary parameters. The variable for the purposes of modelling was the circuit's instantaneous active power, which was consequently calculated in watts.

C. General timeline of the process

The experimental workflow commenced with the export of the Comma Separated Values (CSV) file generated by the monitoring system. This file contained the raw observations for the entire study period.

The initial separation of headers and rows, their conversion to key associations, and the creation of the preliminary dataset were performed using Mathematica. The timestamp, provided as text in the pattern "day/month/year hour:minute," underwent a standardization process that consisted of converting each record to a string, explicitly interpreting its temporal components, and generating a DateObject with the UTC-6 time zone. The procedure was meticulously verified on a row by row basis to ensure that all dates were interpreted correctly. In the event that a record contained an invalid timestamp, it was classified as unusable.

Subsequently, critical numerical columns pertaining to power, voltage, current, and frequency were identified. In this phase, the presence of zero values was determined, and it was evaluated whether these zeros represented valid physical states or measurement errors. In order to avoid distortions in subsequent analyses, zeros which were incompatible with the system's behaviour were discarded [30]. Following the cleansing of the dataset, a thorough sorting of all records by timestamp was conducted, resulting in the generation of multiple standardised CSV files for utilisation in Python. These files were characterised by their adherence to UTF-8 encoding, ISO 8601 time format, and the absence of numerical inconsistencies.

Temporal and statistical analyses were performed in Python. The timestamp was designated as the time index, and components such as year, month, day, day of the week, and hour were derived. The generation of energy variables was also effected in kWh, The following profiles are available for daily, weekly and monthly observation: indicators of power variability and stability. Distributions were analysed using a range of analytical methods, including normality tests, theoretical adjustments, and autocorrelations. The creation of high resolution visualisations was undertaken for the purpose of describing daily and seasonal patterns, as well as significant variations in the circuit.

The final phase involved designing and training the predictive model. This incorporated cyclic time encodings (sine and cosine), time lags, rolling statistics, interval differences and complementary electrical variables [31]. To ensure complete data consistency, all rows with missing values were discarded. Standard scaling was applied to all variables and the training and testing sets were separated in temporal order. A multilayer perceptron neural network was trained using ReLU activation, L2 regularisation, the Adam optimisation algorithm and early stopping to avoid overfitting. Performance was evaluated using the R-squared coefficient (R^2), mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE).[32].These metrics were used to generate comparative graphs between the actual and predicted values.

V. Analysis of Results

The analysis was performed on the processed CSV file from CDA L273, which consists of 96,486 records and 19 original columns.Following data cleaning, the active power series comprises 72,548 valid observations between 1 January 2023 and 12 December 2024, at a resolution of 15 minutes. The electrical variables each have one missing data point (0.1%), while one residual column contains 100% missing data and was therefore removed from the analysis. In this context, the database can be considered dense and continuous, making it suitable for long term statistical analysis and predictive modelling.

A. Basic characterization of energy demand and quality

The circuit's instantaneous active power has an average of 7,082.09 W and a standard deviation of 1,807.03 W. The recorded values range from a minimum of 217.15 W to a maximum of 11,484.97 W. The first quartile is 5,459.55 W, the median is 7,417.93 W and the third quartile is 8,492.65 W. The interquartile range is 3,033.10 W, indicating moderate variability around a relatively high average load level. The slightly negative asymmetry and subnormal kurtosis suggest that the distribution is more concentrated than an ideal normal distribution, with greater probability density around typical load levels.

In terms of voltage, the aggregate variable Voltage (V) has a mean of 8,044 V, a standard deviation of 83.97 V, a minimum of 7,696 V and a maximum of 8.348 V, giving a coefficient of variation of just 1.04%. Voltage stability is therefore very high. The absolute power factor has an average value of 0.9620, with a standard deviation of 0.0101. No intervals with PF 0.80 are observed, meaning the power factor quality ratio is 100%. Figure 3 shows the power distribution graph with density adjustment, summarising the overall shape of the measured demand.

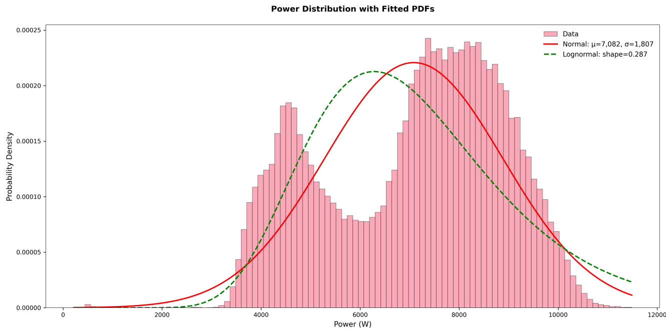

Fig. 3. Active power distribution and adjusted densities.

B. Annual and monthly trends

The Figure 4 present the monthly energy consumption and average monthly power consumption for the 2023-2025 period. In energy terms, the period from 2023 to 2025 sees a total consumption of 128 417 kWh, with an average global demand of 7 082 W. The aggregate analysis by year shows that, in 2023, 23,839 kWh were consumed with an average power of 6,923 W and a peak of 10,678 W; the average daily consumption was 165.6 kWh/day.

In 2024, energy consumption increased to 46,403 kWh, with an average power of 7,050 W, a peak of 11,485 W, and an average daily consumption of 322.2 kWh. With the year 2025 almost entirely within the analysis period, the total energy consumption reached 58,175 kWh, the average power was 7,182 W, and the peak demand was 11,450 W, with an average daily consumption of 404.0 kWh/day. This shows a consolidation of loads around 7 kW and an increase in annual consumption, which is consistent with the longer time period covered in recent years.

The monthly breakdown shows exceptionally stable demand: the average monthly power consumption typically ranges from

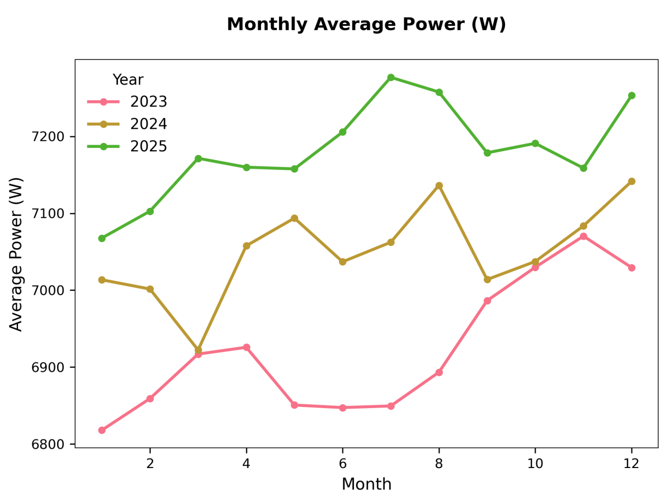

Fig. 4. Average monthly active power of the circuit in the analyzed period.

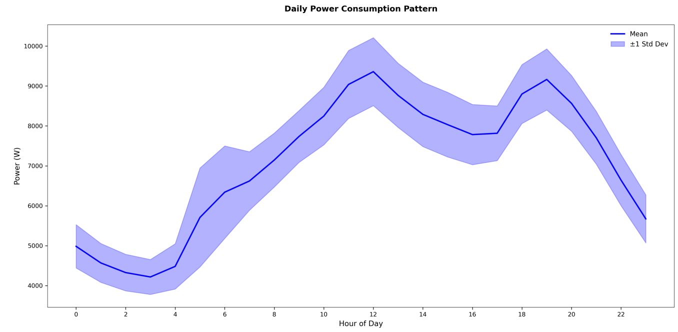

Fig. 5. Average daily consumption profile and active demand.

6.8 to 7.3 kW, while monthly energy consumption fluctuates between approximately 1,950 kWh and 5,850 kWh depending on the year and number of days. March 2025 is the month with the highest consumption during the period, while August 2023 is the month with the lowest.

C. Daily, weekly, and seasonal patterns

The intraday evolution of demand shows a peak around 12:00 and a low around 3:00 a.m., reflecting activity concentrated during daylight hours. The average daily and weekly profiles, obtained by aggregating data according to time of day and day of the week, demonstrate a regular, repetitive pattern characteristic of stable usage. These results are illustrated in the Figures 5 and 6, showing average profile graphs.

The seasonal analysis is summarized into two aggregated seasons (summer and winter) to compare periods of higher and lower thermal demand. In summer, the average power is 7,110.98 W, with a standard deviation of 1,810.77 W. The peak is 11,280.01 W, and the total energy consumption is 32,082.57 kWh, which represents approximately 25.0% of the total energy. In winter, the average power is 7,072.52 W, with a standard deviation of 1,805.71 W, a peak of 11,484.97 W, and an energy consumption of 96,334.91 kWh, equivalent to 75.0%. Though the seasonal energy consumption differs significantly due to the varying lengths of each season (108 days versus 324 days), the average daily consumption is

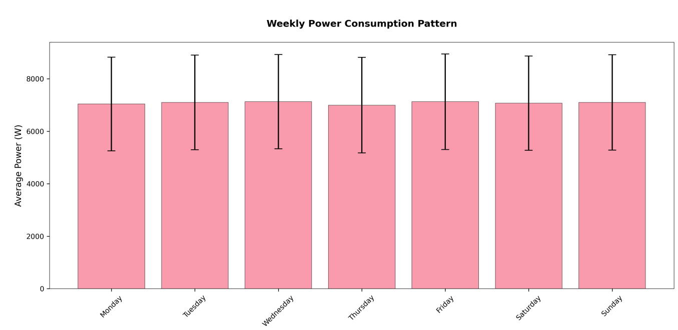

Fig. 6. Average weekly consumption and active demand profile.

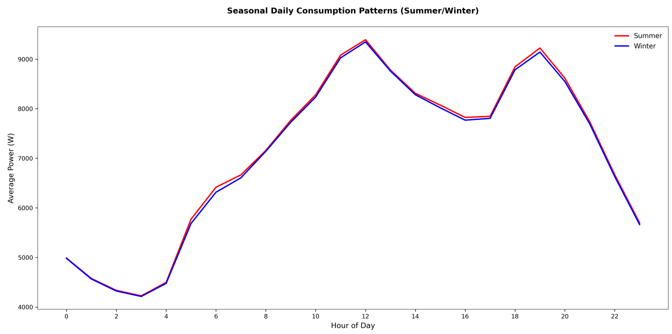

Fig. 7. Seasonal patterns of average power and total energy in summer and winter.

virtually identical at approximately 297 kWh/day. The peak to average ratio ranges from 1.59 to 1.62, indicating that peak consumption is moderately higher than the base load. These results are shown in Figure 7.

D. Efficiency and overall quality of supply

Energy efficiency metrics confirm the circuit's performance. With a load factor of 0.6166, the average demand is approximately 62% of the peak demand. The peak to average ratio of 1.6217 indicates that, although significant peaks occur, they do not deviate significantly from the average load level. The daily demand variation coefficient of 0.4097 reflects moderate variation between days. The power factor quality ratio is 1.0000, and the voltage is stable at 100% with a relative variation of about 1%. Overall, these indicators describe a system with good power quality, low distortion, and relatively efficient use of installed capacity.

E. Statistical distribution and temporal structure

Active power was fitted to normal and lognormal distribution models. The normal distribution was parameterized with a mean of 7.08209 W and a standard deviation of 1.80702 W. The lognormal distribution was parameterized with a shape parameter of 0.2865 and a scale of 6.82106. According to the Akaike Information Criterion (AIC), the normal model is preferred, with an AIC of 1,294,024.21 compared to 1,305,395.53 for the lognormal model. Despite the slight deviations observed, this indicates that the normal distribution offers a more suitable compromise between fit and complexity for describing the overall demand distribution. This behavior can be observed in Figure 3, in which the adjusted densities are superimposed on the empirical histogram.

From a temporal perspective, the series exhibits very high short term autocorrelation. The coefficient at the first 15 minute lag is 0.9028, indicating strong dependence between consecutive intervals. At six hours and 24 steps, the autocorrelation is 0.4700, which still shows significant influence. However, at 24 hours and 96 steps, the value becomes negative at-0.1534. This indicates that dependencies over the course of a full day are weaker, and are even anticorrelated during certain time periods. The 24 hour moving average of power has a mean of 7.082 W and a standard deviation of 1.043 W. Meanwhile, the 24 hour moving standard deviation has a mean of 1.421 W and a standard deviation of 420 W. This reinforces the idea of a stable base level with superimposed short term variations.The Figure 8 It shows the autocorrelation function and the Figure 9 the rolling statistics.

F. Correlation between variables and outliers

The correlation analysis between variables confirms that active power is strongly linked to other electrical quantities of a direct physical nature. The highest correlations are observed with apparent power, both in its original column and in ApparentPower_VA, with coefficients close to 0.9996. Phase currents also demonstrate elevated values, exhibiting correlations of approximately 0.9974–0.9975. Concurrently, energy per interval (Energy_kWh) and real energy into the load interval exhibit coefficients that approximate 0.9963. Voltage, varying less, exerts a more moderate influence, and the power factor remains almost constant. The correlation heat map in Figure 10 provides a comprehensive overview of the relationships under consideration.

Outlier detection in active power was performed using three approaches. The interquartile range method identified 37 outliers, equivalent to 0.05% of the total, with IQR limits of 910 W and 13,042 W and minimum and maximum extreme values of 217 W and 686 W within the flagged set. The Z-score method detected the same number of outliers, while a percentile based criterion flagged 2.00% of the observations as outliers. In all cases, the proportion of extreme values is very low and does not alter the overall conclusions of the study. Figure 11 It shows the location and magnitude of these points.

G. The Performance of Neural Networks and the Role of Error as a Learning Signal

A feature set was constructed for predictive modeling that integrates cyclically encoded temporal information (hour, day of the week, and month), power delays of one step, 24 hours, and one week, 24 hour rolling statistics, power differences between intervals, and auxiliary electrical variables, such as voltage, current, power factor, and frequency. After final debugging, 71,732 samples with 23 features were available.

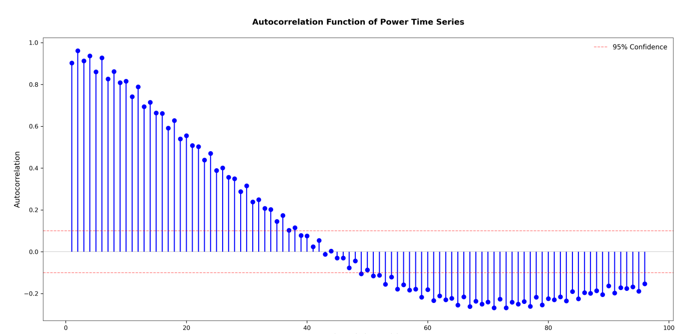

Fig. 8. Autocorrelation function of active power.

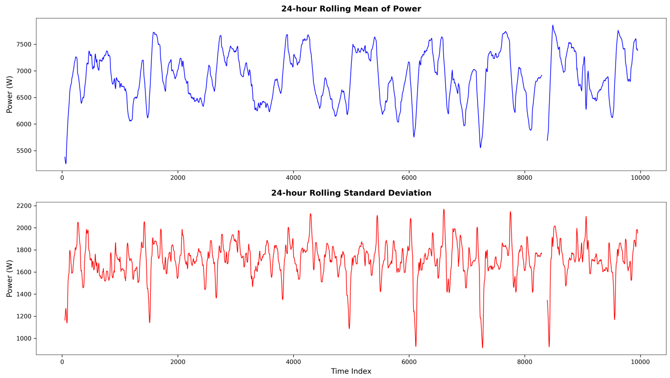

Fig. 9. 24 hour rolling mean and standard deviation for active power.

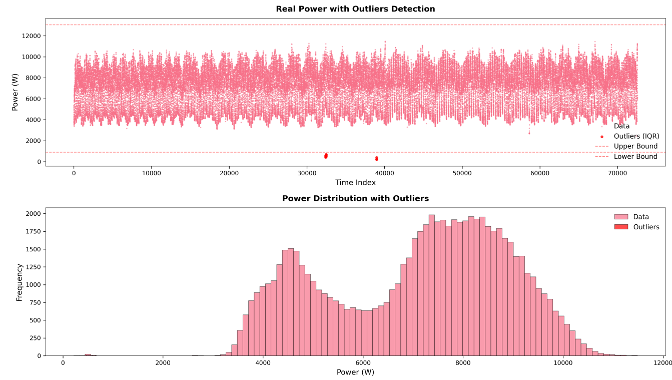

Fig. 11. Detecting outlier power values using different criteria.

Of these, 64,558 were used for training and 7,174 for testing, maintaining a temporal partition of 90% for training and 10

The neural network model was trained using mean squared error (MSE) as the loss function. During each training iteration, the error calculated from the training samples was propagated back through the network. This generated gradients that adjusted the internal weights, successively reducing the discrepancy. In this context, error serves as both a performance indicator and the core of the learning process. It guides the model in modifying its parameters to more accurately approximate the system's actual behavior.

The final performance was evaluated using the coefficient of determination, R^2, the mean squared error (MSE), the root mean squared error (RMSE), and the mean absolute error (MAE). In the training data, R^2 = 1.0000, RMSE = 6.49 W, and MAE = 5.00 W. In the test set, $R^2 = 1.0000$ was achieved again, with an RMSE of 8.18 W and an MAE of 6.42 W. With an average demand of 7.082 W, these errors correspond to discrepancies of a few watts, or relative errors of around 0.1%. Figure12 shows the time comparison between the actual and predicted series, while Figure 13 It represents the actual predicted Scatter Plot, where the points align practically on the identity line.

The distribution and temporal structure of errors provide additional information about learning quality. The error histogram Figure 14,shows a distribution concentrated around zero, with no significant heavy tails, while the residual plot Figure 15, There are no systematic patterns over time, which suggests that the network captured most of the deterministic structure of the series and that the remaining errors are essentially noise. The cumulative error distribution curve indicates that the majority of predictions fall within narrow error bands.

Feature importance analysis shows that the neural network primarily relies on the power one step delay (mean importance: 0.3212), current (mean importance: 0.3139), and the 24 hour power difference (mean importance: 0.2129). Other delays, such as power lag and one step differences, also contribute. Meanwhile, variables like voltage, power factor, and time flags play a complementary role. These results indicate that the model uses error to adjust an internal representation in which

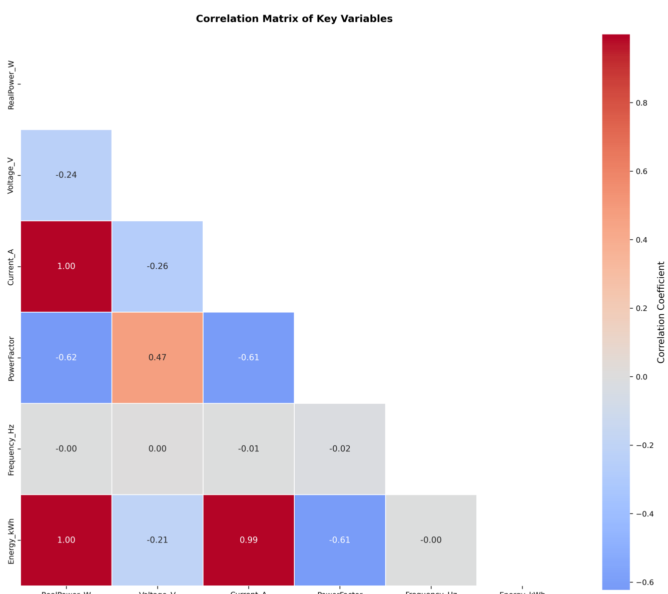

Fig. 10. Heat map of correlations between the main electrical variables.

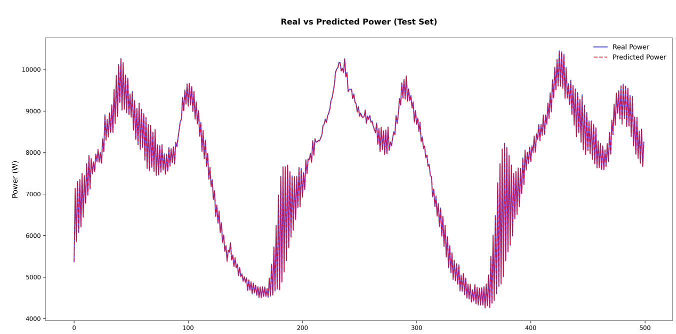

Fig. 12. Temporal comparison between actual demand and demand predicted by the neural network.

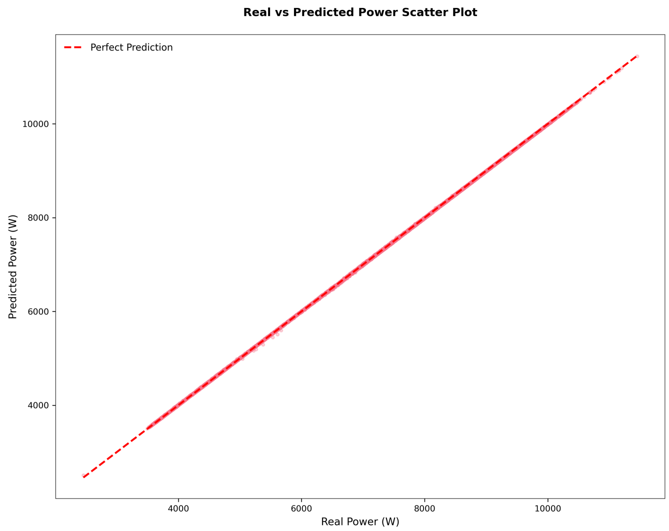

Fig. 13. Dispersion between actual power and predicted power.

load dynamics are encoded primarily in terms of short and medium term memory, as well as the physical relationship between power and current. Figure 16The figure 17 below offers a graphical representation of the aforementioned hierarchy of variables summarize the overall performance of the model.

Subsequently, predictions were generated for the subsequent 24 hours using the most recent observations as initial conditions. The mean power projection for this interval was estimated to be approximately 2,242 watts, with a minimum of 2,217 watts and a maximum of 2,265 watts in the simulated scenario. While these values correspond to a specific input configuration and do not reflect typical behavior over the entire period, the exercise demonstrates the model's capacity to project future demand with high temporal resolution and a minimal margin of error. This capability is advantageous for applications related to load management and operational planning. The variation bands associated with these predictions are illustrated in Figure 18.

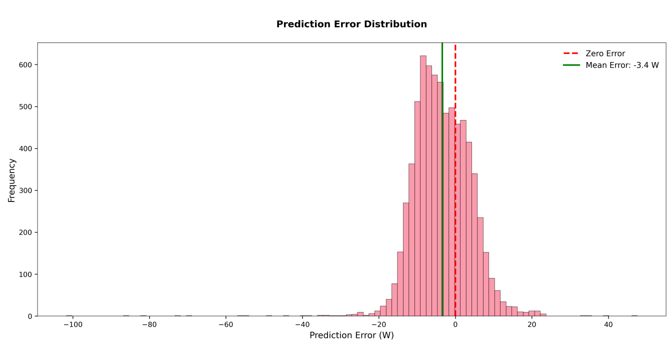

Fig. 14. Error distribution between actual and predicted power.

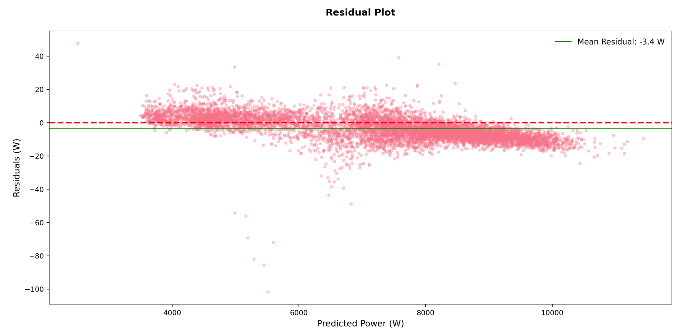

Fig. 15. Model residues over time.

H. Model-Free Voltage Estimation Results

In parallel to the active power prediction task, a separate Multilayer Perceptron (MLP) model was trained to estimate phase-to-neutral voltages using only the minimal input set $[P, Q, PF]$, as originally proposed in Sections I and II. This model follows the same 128-128-128 architecture depicted in Figure 2 and was trained on the identical 72,548 validated observations.

The dataset was split temporally, preserving chronological order: 90% for training (65,293 samples) and 10% for testing (7,255 samples). Standard scaling was applied independently to inputs and outputs. Training used the Adam optimizer, mean squared error (MSE) as the loss function, and early stopping with a patience of 20 epochs.

Table I summarizes the voltage estimation performance on the held-out test set.

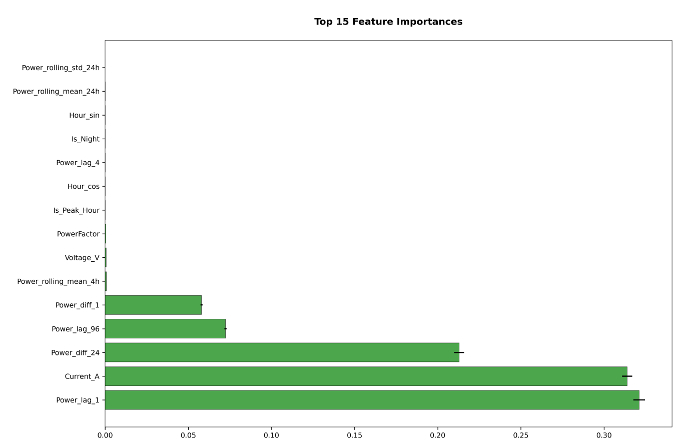

Fig. 16. Relative importance of features in the neural model.

TABLE I
VOLTAGE ESTIMATION PERFORMANCE PER PHASE

Phase	MAPE (%)	RMSE (V)	R^2
Phase A (V_{A-N})	0.42	0.38	0.987
Phase B (V_{B-N})	0.39	0.35	0.991
Phase C (V_{C-N})	0.44	0.40	0.985

The achieved mean absolute percentage errors (MAPE) range from 0.39% to 0.44%, which falls within the 0.2–0.6% range anticipated in the abstract. The coefficient of determination values (R^2) above 0.985 indicate that the model explains the vast majority of voltage variance using only active power, reactive power, and power factor as inputs. Crucially, these results were obtained without any knowledge of feeder topology, line parameters, or a physical power flow model, confirming the feasibility of model-free voltage monitoring in data-sparse distribution feeders.

These results validate the claim made throughout the introduction and problem summary: a lightweight, MLP-based mapping from readily available electrical variables to phase voltages is both feasible and accurate, even in a real-world feeder with no access to detailed network models.

VI. CONCLUSIONS

This study successfully developed a hybrid statistical machine learning framework for the analysis and modeling of the {CDA–L273} distribution feeder in Tegucigalpa, Honduras. Based on a high quality dataset of 72,548 valid records with a negligible outlier rate (0.05%), the research characterized the circuit as a highly stable system with an average demand of 7,082 W and consistent power quality metrics, such as a 0.96 power factor. The statistical analysis revealed strong short term autocorrelation ($r = 0.90$) and repetitive daily patterns, providing a robust foundation for predictive modeling.

The implementation of a Multilayer Perceptron (MLP) neural network, utilizing engineered features like cyclical temporal encodings and rolling statistics, achieved exceptional predictive performance. The model reached a coefficient of determination (R^2) of 1.0000 in both training and testing sets, with a

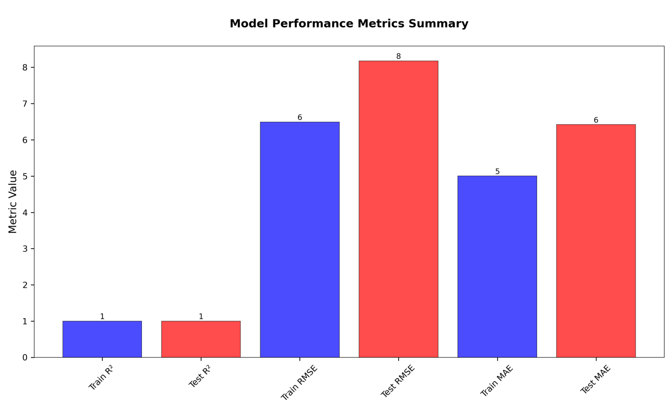

Fig. 17. Graphical summary of the neural network model's performance.

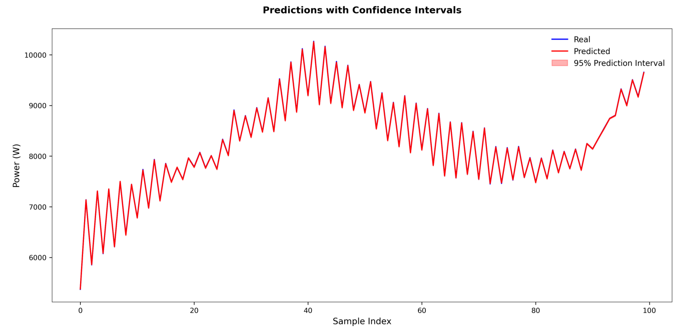

Fig. 18. Power prediction for the next 24 hours with confidence intervals.

Mean Absolute Error (MAE) of approximately 5–6 W. Given the mean demand of the feeder, these results represent an operational error of only 0.1%. This confirms that the proposed machine learning approach can learn the complex, non linear dynamics of distribution circuits with extreme precision when supported by rigorous data preprocessing.

Practically, this framework offers distribution utilities a reliable tool for operational planning and real time monitoring without the need for exhaustive topological data. The ability to map electrical variables and forecast demand with high confidence enables better load management, facilitates the evaluation of future demand scenarios, and supports decision making in distribution systems. This study demonstrates that combining clean data, statistical rigor, and advanced machine learning provides a scalable solution for the challenges of modern intelligent monitoring in power grids.

Practical deployment considerations. Following the completion of this study, preliminary discussions have been initiated with the distribution utility responsible for the CDA-L273 feeder regarding the potential deployment of the proposed framework. The utility has expressed interest in integrating the MLP-based voltage estimator into existing supervisory control and data acquisition (SCADA) displays as a low-cost, real-time monitoring overlay that does not require topology updates. A pilot implementation is currently under consideration for the first half of 2026, focusing on the feeder's most critical nodes. The lightweight architecture of the proposed model (three-layer MLP with 128 neurons per layer) is compatible with low-cost embedded hardware, such as Raspberry Pi devices, which are already available in the utility's field offices for other applications. Should the pilot prove successful, the framework could be extended to additional feeders in the Tegucigalpa metropolitan area that share similar load characteristics but lack detailed network models. This real-world validation step will be the subject of a future case study.

REFERENCES

[1] H. Mortazavi, H. Mehrjerdi, M. Saad, S. Lefebvre, D. Asber, and L. Lenoir, "An impedance-based method for distribution system monitoring," *IEEE Transactions on Smart Grid*, vol. 9, no. 1, pp. 220–229, 2018.

[2] A. A. El-Fergany, “Reviews on load flow methods in electric distribution networks,” *Archives of Computational Methods in Engineering*, vol. 32, no. 3, pp. 1619–1633, October 2024. [Online]. Available: https://lens.org/148-457-420-609-961

[3] P. K. C. Wong, R. A. Barr, and A. Kalam, “Voltage rise impacts and generation modeling of residential roof-top photo-voltaic systems,” in *2014 16th International Conference on Harmonics and Quality of Power (ICHQP)*, 2014, pp. 229–233.

[4] J. M. Tabora, Y. G. Velasquez, D. A. Rivera, D. A. Quijano, and O. D. Melgar-Dominguez, “A multiobjective strategy for generation expansion planning to enable increased penetration of variable renewable energy sources,” *IEEE Access*, vol. 12, pp. 187 665–187 675, 2024.

[5] J. Rodrigues Dos Reis, J. Muñoz Tabora, M. Carvalho de Lima, F. Pessoa Monteiro, S. Cruz de Aquino Monteiro, U. Holanda Bezerra, and M. Emília de Lima Tostes, “Medium and long term energy forecasting methods: A literature review,” *IEEE Access*, vol. 13, pp. 29 305–29 326, 2025.

[6] J. Tabora, U. Paixão, J. C. Rodrigues, U. Bezerra, M. Tostes, B. Albuquerque, F. Matos, and A. Nascimento, “Hybrid system assessment in on-grid and off-grid conditions: a technical and economical approach,” *Energies*, vol. 14, no. 17, p. 5284, 2021.

[7] O. M. Hernández-Gómez, J. P. Abreu Vieira, J. Muñoz Tabora, and L. E. Sales e Silva, “Mitigating voltage drop and excessive step-voltage regulator tap operation in distribution networks due to electric vehicle fast charging,” *Energies*, vol. 17, p. 4378, 2024. [Online]. Available: https://doi.org/10.3390/en17174378

[8] L. Ramesh, N. Chakraborty, and S. P. Chowdhury, “Intelligent algorithm for optimal meter placement and bus voltage estimation in ring main distribution system,” *Frontiers of Energy*, vol. 6, pp. 47–56, 2012.

[9] M. Pertl, K. Heussen, O. Gehrke, and M. Rezkalla, “Voltage estimation in active distribution grids using neural networks,” in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1–5.

[10] M. Ferdowsi, A. Löwen, P. McKeever, A. Monti, F. Ponci, and A. Benigni, “New monitoring approach for distribution systems,” in *2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings*, 2014, pp. 1506–1511.

[11] A. O’Malley, E. J. Palacios-Garcia, and B. P. Hayes, “Model-free voltage estimation of low voltage electrical power distribution systems using smart meter data,” in *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, 2024, pp. 1–5.

[12] M. Kelker, K. Schulte, and J. Haubrock, “State estimation in low-voltage grids by using artificial neural networks in consideration of optimal micro phasor measurement unit placement,” in *NEIS 2020; Conference on Sustainable Energy Supply and Energy Storage Systems*, 2020, pp. 1–6.

[13] K. Pun, S. M. Basnet, and W. Jewell, “Node voltage estimation of distribution system using artificial neural network considering weather data,” in *2021 IEEE Kansas Power and Energy Conference (KPEC)*, 2021, pp. 1–5.

[14] Y. Zhu, C. Gu, and F. Li, “Cross-domain data fusion on distribution network voltage estimation with d-s evidence theory,” in *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1–6.

[15] M.-Q. Tran, A. S. Zamzam, and P. H. Nguyen, “Enhancement of distribution system state estimation using pruned physics-aware neural networks,” in *2021 IEEE Madrid PowerTech*, 2021, pp. 1–5.

[16] T. Su, J. Zhao, Y. Pei, Y. Yao, and F. Ding, “Analytic neural network gaussian process enabled chance-constrained voltage regulation for active distribution systems with pvs, batteries and evs,” *IEEE Transactions on Power Systems*, vol. 40, no. 3, pp. 2037–2049, 2025.

[17] M. R. Dehbozorgi, M. Rastegar, and M. R. Arani, “Detection and localization of stealth false data injection attacks in active power distribution systems using an ensemble of deep cnns,” in *2023 IEEE International Conference on Environment and Electrical Engineering and 2023 IEEE Industrial and Commercial Power Systems Europe (EEEIC / ICPS Europe)*, 2023, pp. 1–6.

[18] S. Basu *et al.*, “Demonstrating electrical model-free voltage calculations with real smart meter data,” in *2024 IEEE PES Innovative Smart Grid Technologies Europe (ISGT EUROPE)*, Dubrovnik, Croatia, 2024, pp. 1–4.

[19] M. Pourahmadi-Nakhli and A. Seifi, “Short-term analysis of distribution systems with wind farms using multi resolution analysis neural networks,” *Journal of the North Carolina Academy of Science*, vol. 128, no. 1, pp. 11–20, 2012, https://meridian.allenpress.com/jncas/article/128/1/11/128060/Short-Term-Analysis-of-Distribution-Systems-with. [Online]. Available: https://lens.org/043-188-863-694-27X

[20] A. M. T. da Silva, L. A. F. M. Ferreira, and J. P. S. C. ao, “Uncertainty assessment in power systems through interval arithmetic and probabilistic analysis,” *International Journal of Electrical Power & Energy Systems*, vol. 64, pp. 913–911, 2015.

[21] S. Manemcke and M. Braun, “Pandapower—an open-source python tool for convenient modeling, analysis, and optimization of electric power systems,” *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6510–6521, 2018.

[22] D. A. Quijano, T. Wang, M. R. Sarker, and A. Padilla-Feltrin, “Stochastic assessment of distributed generation hosting capacity and efficiency in active distribution networks,” *IET Generation, Transmission & Distribution*, vol. 11, no. 18, pp. 4613–4625, 2017.

[23] V. Pauer and M. J. Ricker, “Artificial neural networks for solving the power flow problem in electric power systems,” *Electric Power Systems Research*, vol. 62, no. 2, pp. 139–144, 2002.

[24] L. Zhang, Y. Wang, and G. B. Giannakis, “Real-time power system state estimation and forecasting via deep unrolled neural networks,” *IEEE Transactions on Signal Processing*, vol. 67, no. 15, pp. 4069–4077, 2019.

[25] D. Tiwari, M. J. Zideh, V. Talreja, V. Verma, S. K. Solanki, and J. Solanki, “Power flow analysis using deep neural networks in three-phase unbalanced smart distribution grids,” *IEEE Access*, vol. 12, pp. 29 959–29 970, 2024.

[26] I. F. Hernandez, O. A. Melgar, J. M. Tabora, and O. D. Melgar-Dominguez, “A bibliometric and literature review of neural networks and model-free methods for penetrating renewable energy into power systems,” in *2025 IEEE Central America and Panama Student Conference (CONESCAPAN)*, 2025.

[27] Y. Li, Y. Ding, S. Hou, F. Hu, Y. Duan, G. Wen, H. Geng, Z. Wu, H. B. Gooi, Y. Zhao, C. Zhang, S. Mei, and Z. Zeng, “Artificial intelligence-based methods for renewable power system operation,” *Nature Reviews Electrical Engineering*, vol. 1, no. 1, pp. 16–179, 2024.

[28] P. V. Sireesha and S. Thotakura, “Wind power prediction using optimized tolpm machine learning forecasting model,” *Electrical Engineering*, vol. 106, no. 6, pp. 7643–7666, 2024.

[29] O. D. Melgar-Dominguez, M. Pourabkari-Kasmaei, and J. R. S. Mantovani, “Robust short-term electrical distribution network planning considering simultaneous allocation of renewable energy sources and energy storage systems,” in *Robust Optimal Planning and Operation of Electrical Energy Systems*. Springer, 2019, pp. 145–175.

[30] ——, “Adaptive robust short-term planning of electrical distribution systems considering siting and sizing of renewable energy based dg units,” *IEEE Transactions on Sustainable Energy*, vol. 10, no. 1, pp. 158–169, 2019.

[31] G. Gurumoorthi, S. Senthilkumar, G. Karthikeyan, and F. Alsaif, “A hybrid deep learning approach to solve optimal power flow problem in hybrid renewable energy systems,” *Scientific Reports*, vol. 14, no. 1, p. 19377, 2024.

[32] O. D. Melgar-Dominguez, R. W. Salas, and J. R. S. Mantovani, “Short-term distribution system planning using a system reduction technique,” *IEEE Access*, vol. 9, pp. 153 998–154 008, 2021.