Summer Tehqiq Research Program
Exploration of AI on Distribution Network Aggregation considering IBRs
Final Project Report

Group Members:

Batool Zehra Ladha

Manal Hasan

Rubab Shah

Supervisor:

Syed Muhammad Hur Rizvi

Dhanani School of Science & Engineering

Habib University

Summer Semester ’24

April 24, 2025

Contents

1 Introduction

1.1 Research Problem

How can ML be efficiently utilized to estimate operational conditions/parameters of modern power distribution networks focusing on load models, PV and BESS control modes, given the challenge of data scarcity and complex interactions of active components?

1.2 Rationale Research Gap

The landscape of power systems is undergoing significant transformation due to the integration of distributed generation sources, such as photovoltaic (PV) systems, generators, and battery energy storage systems (BESS). These technologies contribute to a more dynamic and bidirectional flow of electricity within the grid, fundamentally altering the traditional role of passive distribution networks. Usually, these networks are designed to transport electricity from centralized power plants to consumers with minimal interaction or feedback from the end-user. The introduction of bidirectional power flows, where energy can both be delivered to and drawn from the grid, creates new challenges in grid management.

This shift towards active distribution networks complicates the task of accurately estimating power system parameters. The effectiveness of grid management relies on precise parameter estimation to optimize performance and maintain stability. However, the current methodologies for estimating these parameters are constrained by a lack of comprehensive datasets. Often collected manually, existing datasets are limited in scope and fail to capture the full range of operational scenarios that include the complexities introduced by bidirectional power flows and interactions among smart devices, renewable energy sources, and the grid itself.

The challenge is further compounded by the fact that traditional estimation methods do not adequately account for the dynamic conditions presented by the integration of these technologies. As a result, there is a critical need for enhanced datasets that reflect the new dynamics of active distribution networks. This research aims to address this gap by leveraging modern techniques to automate the generation of extensive and diverse training datasets. By utilizing Python library and simulation tools such as OpenDSS, this project seeks to improve the accuracy and reliability of machine learning (ML) models used for estimating power system parameters by network aggregation.

The goal is to develop a more effective data-driven approach to better accommodate the complexities introduced by distributed generation sources and bidirectional power flows. This improved approach will contribute to more effective grid management and planning, ultimately enhancing the stability and efficiency of power systems in the face of ongoing technological advancements. Through this research, we aim to provide a solution that fills the current data gaps and supports the future evolution of grid management practices.

2 Literature Review

The evolution of power systems from the early to late twentieth century was largely characterized by a centralized structure, where electricity was generated at large-scale and then transmitted to major load centers. This approach was efficient for the technology and energy needs of the time. However, as we entered the 21st century, the energy landscape began to shift significantly due to a growing emphasis on renewable energy resources. This shift led to a fundamental transformation in the architecture of power systems, moving away from centralized generation towards a more distributed and decentralized model. This change was driven by the integration of various renewable energy sources, such as wind and solar power, which are inherently distributed.

The increasing adoption of renewable energy sources, particularly photovoltaic (PV) systems, has introduced new challenges and opportunities for the power grid. One significant change is the rise of distributed generation, where electricity is produced closer to where it is used rather than at centralized power plants. This shift has transformed grid management by necessitating new approaches to handle the diverse and complex nature of modern power loads. Intelligent loads like Electric Vehicles (EVs) and Battery Energy Storage Systems (BESS) further complicate this landscape. These smart loads not only consume electricity but also provide valuable services to the grid; EVs can adjust their charging based on grid demands, and BESS can store and release energy during peak periods. The integration of these technologies requires the grid to manage bidirectional energy flows and adapt to real-time changes in generation and consumption, highlighting the need for advanced management strategies to maintain stability and optimize efficiency. This evolving dynamic underscores the importance of innovative solutions to address both the challenges and opportunities presented by these advancements.

Alongside the physical changes in the grid's infrastructure, advancements in digital technology have significantly transformed the way power systems are monitored and controlled. The adoption of digital monitoring devices, including Phasor Measurement Units (PMUs) and smart meters, has played a crucial role in this evolution. These technologies facilitate the collection and analysis of real-time data, providing a detailed and timely picture of the grid's operational status. With PMUs, which measure the electrical waves on an electricity grid to determine the health of the system, and smart meters, which offer precise consumption data, grid operators can gain a more comprehensive understanding of power flow and system performance.

This real-time visibility allows for more accurate monitoring and quicker responses to any anomalies or fluctuations in the grid. By utilizing these tools, operators can enhance their ability to manage power system operations effectively, improving both efficiency and reliability. The ability to detect and address issues reduces the risk of outages and optimizes the overall performance of the power grid. Consequently, the integration of digital monitoring technologies has become a critical component in modernizing power systems and ensuring their reliability in the face of increasing complexity.

One of the major challenges brought about by the integration of renewable energy sources and smart loads is the increased uncertainty in power system operation. Traditional power systems, characterized by their predictable generation patterns and stable consumption levels, were relatively straightforward to manage using established models and control systems. However, the introduction of renewables such as photovoltaic (PV) systems, which are inherently variable and dependent on weather conditions, coupled with smart loads that can rapidly alter their consumption patterns, has significantly raised the complexity of grid management. This new level of unpredictability has made it more difficult for grid operators to maintain a reliable and stable power supply, requiring more advanced strategies and technologies to handle the fluctuating and dynamic nature of modern energy systems.

To address these challenges, power system loads are frequently modeled based on their dependency on voltage and frequency [1]. Among the various modeling approaches, two of the most commonly utilized are the ZIP load model and the exponential load model. The ZIP model, in particular, is especially valuable for representing loads as a combination of constant impedance (Z), constant current (I), and constant power (P) components [2]. This model allows for a better representation of how different types of loads behave under varying conditions.[2]

The parameters associated with these models, such as the coefficients for each load type in the ZIP model—namely, az (constant impedance), ai (constant current), and ap (constant power)—are essential for accurately forecasting load behavior. These coefficients help determine how loads will respond to changes in voltage and frequency, which is crucial for maintaining grid stability and reliability. This level of precision is increasingly important in modern power systems, where the characteristics of loads have evolved significantly. The rise of electronic devices and Electric Vehicles (EVs) has introduced new dynamics that affect how loads interact with the grid [1].

Furthermore, the IEEE 1547 standard [3] offers a thorough and detailed framework for managing distributed energy resources, with a particular focus on photovoltaic (PV) systems. This standard specifies various control modes, including Volt-VAR, Volt-WATT, Constant Power Factor (PF), Constant Reactive Power (Q), and WATT-VAR modes. Each of these control modes plays a critical role in maintaining grid stability and efficiency as the integration of PV systems becomes more prevalent.

The Volt-VAR mode adjusts reactive power to regulate voltage levels, while the Volt-WATT mode adjusts real power in response to voltage changes. The Constant Power Factor (PF) mode maintains a fixed power factor, and the Constant Reactive Power (Q) mode provides a constant level of reactive power. The WATT-VAR mode combines real and reactive power adjustments to manage voltage and power factor. These control modes are important for optimizing the performance of the grid, particularly as the share of distributed generation from PV systems grows.Despite the extensive guidelines and recommendations provided by IEEE 1547, there remains a significant gap in the literature concerning the real-time identification and implementation of these control modes and their associated parameters. Accurately understanding and applying these modes in practical, real-world scenarios is essential for enhancing the coordination between distributed resources and the conventional grid infrastructure. This knowledge is critical for optimizing grid performance and ensuring that the integration of distributed energy sources contributes positively to the overall stability and efficiency of the power system [3].

Research has revealed the complexities involved in managing the dynamic interactions between distributed energy resources (DERs), such as photovoltaic (PV) systems and battery energy storage systems (BESS), and the existing grid infrastructure. Traditional load models often rely on static assumptions that don't fully capture the diverse and dynamic behaviors of modern power systems. These models, which typically assume that loads behave consistently, may not account for the variability introduced by new technologies.

To address these issues, it is necessary to enhance these models by incorporating factors such as load elasticity and responsiveness to changes in voltage and frequency. Load elasticity refers to how much and how quickly a load can adjust its consumption based on fluctuations in grid conditions, while responsiveness involves how well loads react to changes in voltage and frequency. Including these factors in load models provides a more accurate representation of how loads behave under varying conditions, which is particularly important as the integration of DERs into the grid increases.

Improving load models with these considerations allows for a more realistic representation of how modern power systems operate and interact with distributed resources. This refinement is especially important for optimizing grid performance and ensuring stability as the proportion of DERs in the grid continues to grow. Effective modeling of these interactions helps in better planning and managing the grid, ultimately supporting the integration of DERs in a way that enhances overall system reliability and efficiency [4].

Moreover, as the integration of distributed energy resources (DERs) into the power grid increases, it introduces new variables and potential instabilities. This shift highlights the need for more dynamic and flexible load modeling techniques that can handle the variability and uncertainty brought about by DERs. Unlike traditional models, which often rely on static assumptions, these new approaches must adapt to the changing nature of modern power systems.To address these challenges, it is essential to use advanced data acquisition tools. Phasor Measurement Units (PMUs) and smart meters are particularly valuable in this regard.PMUs are designed to provide precise, time-synchronized measurements of critical electrical parameters. Specifically, they measure voltage and current phasors, which represent the magnitude and phase angle of sinusoidal waveforms, along with frequency and the rate of change of frequency (ROCOF). The sampling rates and GPS synchronization in PMUs ensure that data from different locations can be compared accurately in real time, providing a detailed snapshot of grid conditions. Smart meters collect detailed data on energy consumption and various electrical parameters. These meters measure active power (watts), reactive power (VARs), voltage, current, and power factor. Moreover, smart meters can record time-of-use data, enabling the analysis of consumption patterns and demand-side management. They are also capable of detecting anomalies, such as voltage sags or swells, and can communicate this information back to the utility for further analysis. By leveraging the real-time data provided by PMUs and smart meters, grid operators can develop more accurate load models. This detailed and time-sensitive information allows for better management of the grid, enabling quicker responses to fluctuations and more effective strategies for addressing the complexities associated with increased Distributed Energy Resources (DER) penetration. Ultimately, integrating these advanced data acquisition systems is essential for maintaining a stable and efficient power grid as it evolves with the growing presence of DERs. [4].

Ultimately, the shift towards a more distributed and decentralized power system has introduced both challenges and opportunities. As renewable energy sources and smart loads become more integrated into the grid, and as digital monitoring technologies advance, the complexity of power systems has increased significantly. While these changes present new difficulties, they also enhance the capabilities of the grid.To effectively manage this growing complexity, it is essential to develop more advanced models and strategies. These models must be capable of addressing the dynamic nature of modern power systems, ensuring that they remain reliable and efficient as they adapt to new conditions. By advancing our understanding and management of these evolving systems, we can support their continued operation and performance in a rapidly changing environment.

3 Methodology

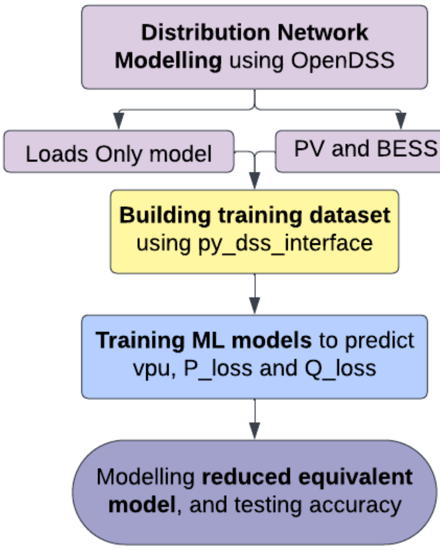


Figure 1: Summary of Methodology

3.1 Research Design

This research was designed to explore the potential of machine learning (ML) for power system parameter estimation, focusing on load models, and BESS and PV control modes, by simulating a reduced equivalent model by network aggregation. For this intent, the power distribution network as shown in *figure2*. is simulated on OpenDSS software.

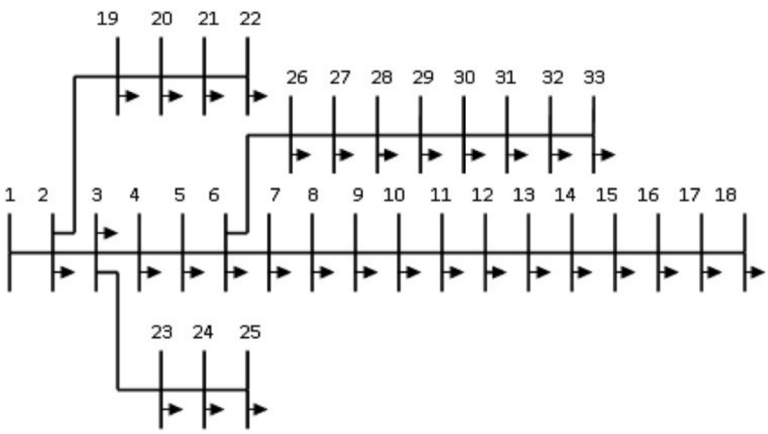


Figure 2: IEEE 33 Bus System with loads only

The distribution network was modeled in OpenDSS according to IEEE 1547 standards as a balanced three-phase system. This choice was made due to OpenDSS's capabilities in handling diverse power system configurations, renewable resources, and compatibility with Python for automation. This ensures the model accurately reflects real-world scenarios where energy can flow from generation to load (unidirectional) and from distributed generation back to the grid (bidirectional). The models included various operational scenarios, such as different loads, ZIPV Parameters, PV inverter characteristics, control modes, SOCs in BESS, and Generators such as Wind Turbine. The diversity of these scenarios was significant for representing the wide range of possible conditions in modern power distribution systems.

By replicating the distribution system, and solving the simulation, operational parameters such as power, current, and voltage are generated, which will then be used as features of a ML model to predict parameters to develop a reduced aggregated network model, as shown in *figure3*, for each branch in *figure2*. This research only deals with magnitudes of parameters rather than its values in polar coordinates, as they are easier to interpret and the angles for each phase difference can be ignored since the system under our consideration is balanced.

To prove the efficiency of the ML model deployed for the parameter estimation, Mean Squared Error (MSE) will be used as an error metric. Afterward, MSE will again be employed to validate the reduced simulated model parameters by comparing them with parameters from the original simulated model. With a low MSE score, the effectiveness and efficiency of ML in estimating the operational parameters can be deduced. This research design was appropriate for addressing the research problem because it provided a systematic way to generate and analyze large amounts of data, which is essential for training accurate ML models. Simulations ensured the data covered a range of real-world conditions, while the automation facilitated efficient data generation.

For detailed analysis, branches 6-18 are selected, the longest branch in the IEEE 33-bus system in *figure2*. This branch was chosen to provide a representative dataset covering a significant portion of the network. By re-modeling this branch, we aim to capture a wide range of operational conditions and parameters.

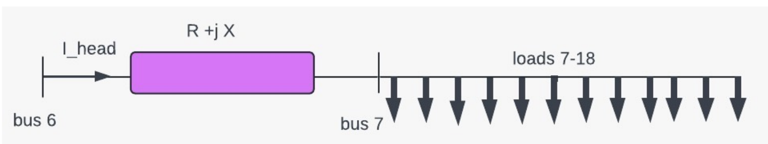


Figure 3: Reduced Equivalent of branch 6-18 by Network Aggregation

3.2 Data Generation & Collection

3.2.1 Data Generation in OpenDSS

Data for this study was generated using OpenDSS to simulate two modified IEEE 33-bus distribution systems, providing a comprehensive foundation for a simulation-based approach.
The load-only system was configured with a base frequency of 50 Hz and included various components such as transformers, lines, and loads. The network comprised 33 buses interconnected through lines and transformers. A new circuit named DEWAckt was initialized with a base voltage of 150 kV in a three-phase system, with the source bus designated as the main source. A step-down transformer (T1) was used to reduce the voltage from 150 kV to 12.66 kV, connecting to bus 1. The transformer's primary winding was connected to the source bus with a wye connection, while the secondary winding was connected to bus 1 with a delta connection. The transformer's parameters included a winding resistance of 0.05% and a reactance of 6%.
Lines with varying resistance and reactance values connected the buses, facilitating power flow between different parts of the network. Various loads were connected across the network at different bus locations, each characterized by its active power (kW), reactive power (kVAR), and voltage characteristics. The loads were modeled using the ZIPV model to simulate realistic behavior under different operating conditions. The control mode was set to static, and the simulation was run in snap mode.
The PV and BESS system was also configured with a base frequency of 50 Hz and included various components such as transformers, lines, loads, and distributed energy resources (DERs) like photovoltaic (PV) systems and battery energy storage systems (BESS). The configuration of the network, transformers, lines, and loads remained the same as in the load-only model. However, this network included several DERs strategically placed within the system:

- **Wind turbines** were connected at buses 7, 10, and 18 with power ratings of 150 kW, 60 kW, and 70 kW, respectively, each operating with a power factor of 0.85.
- **Battery Energy Storage Systems** were placed at buses 15 and 13, with a rated power of 50 kW, energy storage capacity of 1000 kWh, and maximum reactive power support of 600 kVar. These batteries are operated under Volt-VAR control mode to manage voltage levels within the network.
- A **PV System** was connected at bus 9 with a power rating of 500 kW and also operated under Volt-VAR control to maintain voltage stability.

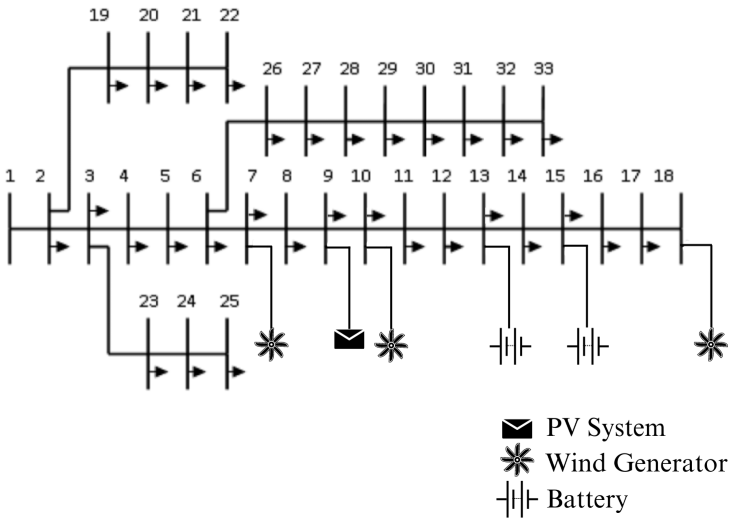


Figure 4: IEEE 33 Bus System with PV & BESS system

The voltage levels were established at 12.66 kV and 150 kV to facilitate per-unit calculations and ensure consistency with the IEEE 33-bus standard. The simulation parameters were configured to enable iterative control and ensure convergence during the simulations. The control mode was set to static, and simulations were conducted in daily mode with a one-hour time step over 12 hours. This setup enabled precise monitoring of the battery charging and discharging cycles. Data from these simulations were collected for further analysis.

3.2.2 Data Extraction

Extraction of data was automated by Python script using libraries; py_dss_interface, pathlib, numpy and pandas. Pathlib was used to access the modeled .dss files into our code, and the py_dss_interface enabled the control to OpenDSS by using its commands in Python.

For loads only model, the parameters extracted were ZIPV, active power (P_b in kW), reactive power (Q_b in kVar), and voltage per unit (vpu) for each load in the branch under consideration, the magnitude of active power loss (P_{loss}), the magnitude of reactive power loss (Q_{loss}) of each line section in the branch, and finally the magnitude of the current at the head (*I_head*) of the branch. P_{loss}, and Q_{loss} were calculated by extracting the power in the sending (bus1) and receiving (bus2) ends of a line section, then subtracting the bus2's power from bus1's power. Moreover, vpu is calculated by dividing the magnitude of the voltage at a load by base voltage. Whereas, the PV and BESS model, apart from having the same parameters as loads only model, also had several other parameters; stored efficiency, state (charging = 1, discharging = 0), and active power output in kW of the two batteries included in the model, and the irradiance of the PV for the entire branch.

The extracted features (ZIPV, P_b, Q_b, vpu, P_{loss}, Q_{loss}, *I_head*) are crucial for understanding the load dynamics, power flow, and losses in the branch. Moreover, stored efficiency, state, and kW of batteries, and irradiance of PV, are critical for capturing the interactions between DERs and the distribution network

3.2.3 Building Training Dataset & Data Management

To make a training dataset for loads only model that would later be utilized for ML, numpy was used to generate unique random magnitudes for ZIPV parameters, P_b (kW), and Q_b (kVar). Both P_b and Q_b are varied in the 50-150% range of the original parameter values. Whereas, ZIPV (general form [a_z, a_I, a_P, r_z, r_I, r_P, 0]) factors are varied making sure that the sum of a_z (active impedance), a_I (active current), a_P (active power) of each load equals to 1. The same goes for reactive factors of ZIPV r_z (reactive impedance), r_I (reactive current), and r_P (reactive power). These values were then overwritten with the original values (only virtually) to extract the modified values of vpu, P_{loss}, Q_{loss}, and *I_head* from the simulation that was affected due to changing *x_features*. Whereas for PV and BESS models, stored efficiency varied between 0-100, the state of the battery was either 0 or 1, kW varied from 0 to the size of the battery, and irradiance also changed between 0 and 1.
Both these datasets were first stored in an array of shape rank 3 tensors with a length of 5000 so that sufficient data is accessible for training and testing the ML model, and then by using pandas, dataframe was written in Excel sheets.

The general format of loads only model's dataset is: [[[$load_1$ parameters (ZIPV, kW, kVar, vpu, P_{loss}, Q_{loss})], [$load_2$ parameters (ZIPV, kW, kVar, vpu, P_{loss}, Q_{loss})], . . ., [$load_n$ parameters (ZIPV, kW, kVar, vpu, P_{loss}, Q_{loss}], [*I_head*]], . . .]

Similarly, general format of PV and BESS model is: [[[$battery_1$(efficiency%, state, kW), $battery_2$(efficiency%, state, kW)], [$load_1$ parameters (ZIPV, kW, kVar, vpu, P_{loss}, Q_{loss})], [$load_2$ parameters (ZIPV, kW, kVar, vpu, P_{loss}, Q_{loss})], . . ., [$load_n$ parameters (ZIPV, kW, kVar, vpu, P_{loss}, Q_{loss})], [*I_head*, irradiance]], . . .]

3.3 Data Analysis

3.3.1 Preparing Tensors For Analysis

In this research, the TensorFlow library in Python is utilized for machine learning purposes. Therefore, the initial step is to read the Excel sheet storing the dataset into tensors as numpy arrays due to the inhomogeneous shape of the dataset and split them into x and y features

(predicted). For training of the ML model, one for predicting vpu, P_{loss}, and Q_{loss}.
For loads only model, tensor for predicting vpu of each load, P_{loss} and Q_{loss} at each line section of the branch is:[[[$load_1$ parameters (ZIPV, kW, kVar], [$load_2$ parameters (ZIPV, kW, kVar], . . ., [$load_n$ parameters (ZIPV, kW, kVar]], . . .].

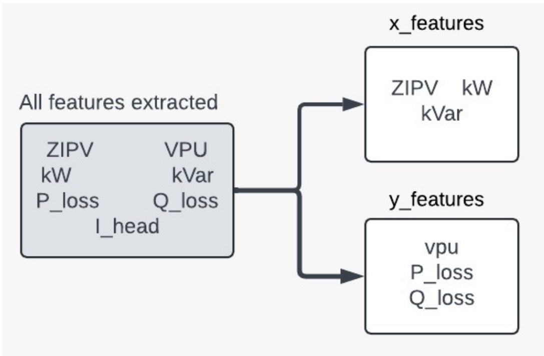


Figure 5: Parameters in the Loads-Only Model Tensor

The features selected for the loads-only model—ZIPV parameters, active power P_b, and reactive power Q_b—are essential for accurately predicting the per-unit voltage (vpu), active power loss P_{loss}, and reactive power loss Q_{loss} across the distribution network. ZIPV parameters reflect the load's response to voltage changes, directly impacting voltage stability and power quality. By including P_b and Q_b, the model captures the power flow dynamics that influence losses and voltage regulation. The aim is to predict these outputs to facilitate network aggregation, enabling the calculation of equivalent resistance (R) and reactance (X) from total P_{loss} and Q_{loss}, and to establish new power bases and ZIPV parameters necessary for simplified network modeling.
Similarly for For PV and BESS models, the tensor for predicting vpu of each load, P_{loss} and Q_{loss} is: [[$battery_2$(efficiency%, state, kW), $battery_2$(efficiency%, state, kW)], [$load_1$ parameters (ZIPV, kW, kVar], [$load_2$ parameters (ZIPV, kW, kVar], . . ., [$load_n$ parameters (ZIPV, kW, kVar]], . . .].

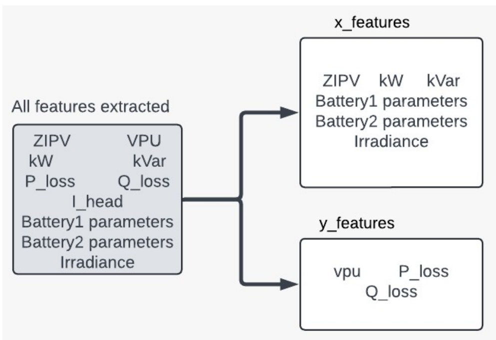


Figure 6: Parameters in the Load and PV BESS Model Tensor

For the PV and BESS models, additional features such as battery efficiency, state of charge, active power output, and irradiance were included to capture the operational characteristics of distributed energy resources. These factors are crucial for accurately predicting the impacts on vpu, P_{loss}, and Q_{loss}, given the variability introduced by PV generation and battery operation.
To split the data into training and validation (testing) datasets, sklearn's train_test_split library was imported. After splitting 80% of the data was set for training and the remaining 20% for the testing.

3.3.2 Machine Learning Model, parameter estimation, and prediction analysis

For the prediction of vpu, P_{loss} and Q_{loss} of both loads only, and PV and BESS models the neural network model employed is a sequential model built using the Keras library of Tensorflow. The architecture, as shown in *figure7* and *figure8*, is specifically designed for time-series prediction tasks and consists of the following layers:

1. **Input Layer**: The input layer is defined with a shape corresponding to the features of the training data, *x_train*, ensuring that the model accepts the input data correctly.

2. **LSTM Layers**: The model incorporates two Long Short-Term Memory (LSTM) layers, each with 64 units and configured to return sequences. This setup is effective in capturing temporal dependencies in the data. Using two LSTM layers allows the model to learn hierarchical temporal features, with each layer potentially capturing different levels of temporal patterns.

   - The first LSTM layer processes the input sequences and returns the full sequence to the next LSTM layer.
   - The second LSTM layer takes the sequence output from the first LSTM layer and also returns the full sequence for further processing.

3. **Time-Distributed Dense Layers**: Following the LSTM layers, the model includes two TimeDistributed dense layers:

   - The first TimeDistributed layer contains 32 neurons and uses the ReLU activation function to introduce non-linearity and capture complex patterns.
   - The second TimeDistributed layer adjusts its output to match the shape of the target data, y_train.

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 12, 64)	18,944
lstm_1 (LSTM)	(None, 12, 64)	33,024
time_distributed (TimeDistributed)	(None, 12, 32)	2,080
time_distributed_1 (TimeDistributed)	(None, 12, 3)	99

Total params: 54,147 (211.51 KB)
Trainable params: 54,147 (211.51 KB)

Figure 7: Loads only ML Model Summary

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 12, 64)	20,736
lstm_1 (LSTM)	(None, 12, 64)	33,024
time_distributed (TimeDistributed)	(None, 12, 32)	2,080
time_distributed_1 (TimeDistributed)	(None, 12, 3)	99

Total params: 55,939 (218.51 KB)
Trainable params: 55,939 (218.51 KB)
Non-trainable params: 0 (0.00 B)

Figure 8: PV & BESS ML Model Summary

The model is compiled with the Adam optimizer, using a learning rate of 0.001. The loss function selected is mean squared error (MSE), and MSE is also used as a performance metric to provide a clear measure of the model's prediction accuracy during training and validation. This configuration is chosen to minimize prediction errors and ensure efficient training convergence.

Moreover, to prevent over-fitting early stopping is employed, halting training when the validation loss stops improving. A patience of 10 epochs allows the model some leeway to improve before stopping. Restoring the best weights ensures the model retains the parameters with the best validation performance. Other hyper-parameters used are 50 epochs, which in our model provides a balance between sufficient training time and computational efficiency, allowing the model to converge without over-fitting, with a batch size of 32. All hyper-parameters, layers, and architecture are chosen to balance model complexity, training efficiency, and prediction accuracy after trial and error method.

For loads only model, the predicted vpu, P_{loss}, and Q_{loss} were used further to calculate overall active, reactive power loss, resistance (R), reactance (X), and also the new base powers P_b, Q_b, and new ZIPV parameters for the loads in the branch 6-18.
According to Joules Law it is known that power dissipated (P) from a circuit is $P = I^2R$, where

I and R are current and resistance, respectively.
To calculate the resistance R, we utilized the active power losses in the formula.

$$R = \frac{P_{\text{loss}}}{I^2} \tag{1}$$

For calculating the reactance X, we utilized the reactive power loss in the formula:

$$X = \frac{Q_{\text{loss}}}{I^2} \tag{2}$$

Since the entire branch 6-18 is reduced to lines 6-7, the line-to-line voltage at the receiving end of lines 6-7 in the reduced network model needs to be calculated using the predicted parameters. For the line-to-line voltage (V_{LL}) at the sending end (bus 6), we used the following formula:

$$V_{\text{LL}}(S) = V_{\text{base}} \times VPU_{\text{bus6}} \tag{3}$$

$$V_{\text{LL}}(S) = 12.66 \times VPU_{\text{bus6}} \tag{4}$$

The line-to-neutral voltage at the sending end, $V_{\text{LN}}(S)$ is given by:

$$V_{\text{LN}}(S) = \frac{V_{\text{LL}}(S)}{\sqrt{3}} \quad \text{since the model is a balanced 3-phase system.} \tag{5}$$

Now, the line-to-neutral voltage at the receiving end $V_{\text{LN}}(R)$ can be calculated using $V = IR$:

$$V_{\text{LN}}(R) = V_{\text{LN}}(S) - I(R + jX) \tag{6}$$

The per-unit voltage at the receiving end is then given by:

$$V_{\text{LN}}(R)pu = \frac{V_{\text{LN}}(R)}{V_{\text{base}}/\sqrt{3}} \tag{7}$$

Once the new per-unit V_{LN} is predicted, the new ZIPV parameters, along with the new active and reactive power bases (P_b and Q_b), needs to be predicted for each branch. Base power at the load can be calculated by:

$$P = P_b \left(a_z V^2 + a_i V + a_p\right) \tag{8}$$

Since we aim to achieve the reduced equivalent model:

$$P_b' \left(a_z' (V_{\text{LN}}(R)_{\text{pu}})^2 + a_i' (V_{\text{LN}}(R)_{\text{pu}}) + a_p\right) = P_b \left(a_z V^2 + a_i V + a_p\right) \tag{9}$$

To predict P_b' (predicted active power base), we use the fact that:

$$a_z' + a_i' + a_p \quad = 1 \tag{10}$$

From Equation (9), we can equate:

$$a_i' = \frac{P_b \times a_i \times V}{P_b' \times V_{\text{LN}}(R)} \tag{11}$$

$$a_z' = \frac{P_b \times a_z \times V^2}{P_b' \times (V_{\text{LN}}(R))^2} \tag{12}$$

$$a_p' = \frac{P_b \times a_p}{P_b'} \tag{13}$$

Substituting these into Equation (10):

$$P_b' = \frac{P_b \times a_z \times V^2}{(V_{\text{LN}}(R))^2} + \frac{P_b \times a_i \times V}{V_{\text{LN}}(R)} + (P_b \times a_p) \tag{14}$$

Similarly, the same approach can be applied to find the predicted reactive power base Q_b':

$$Q_b' = Q_b \times \left(\frac{a_z \times V^2}{(V_{\text{LN}}(R))^2} + \frac{r_i \times V}{V_{\text{LN}}(R)} + a_p \right) \tag{15}$$

Now the predicted P_b' will be substituted in Equations(11), (12), and (13) to find new active ZIPV factors, and Q_b' will be used to predict reactive ZIPV factors (r'_Z, r'_I, r'_P).

3.3.3 Reduced Equivalent Modeling and Validation

The predicted parameters, including resistance (R), reactance (X), and the new active P_b and reactive Q_b power components, along with the ZIPV model coefficients, were utilized to develop a reduced equivalent model. This model aimed to represent the network as an aggregate, using the estimated parameters to replicate the system's behavior. The primary objective of this approach was to assess the performance and accuracy of the reduced equivalent model by comparing it with the original, detailed loads-only model. By doing so, we sought to evaluate how well the simplified model could replicate the actual system dynamics and identify any discrepancies in performance metrics.
In the original model, we included all the lines and loads of the IEEE 33-bus system, covering every bus from 6-18 with detailed load data. This provided a thorough simulation of the network's behavior. For the reduced model, we simplified things by removing lines 7-8 till 17-18 and loads from buses 8 to 18. Instead of modeling these loads separately, we combined them all and represented them on bus 7. This remodeling was done to condense the entire branch in

between lines 6-7 with the predicted parameters of the entire branch, thus resulting in a reduced model equivalent to the original branch. The new load parameters—such as their active power (kW) and reactive power (kVA)—were predicted using a neural network machine learning approach.

The goal of this reduction was to see how well the simplified model could match the performance of the original, more detailed model. By combining loads into a single bus, we aimed to compare key metrics like power, voltage, and current between the two models. This comparison helped us understand how the simplifications affected accuracy and how closely the reduced model mirrored the behavior of the original system.

To further validate, the accuracy of the estimated parameters and how effective ML was in predicting them, total active loss, total reactive loss, current at the head of the branch in lines 6-7 *(I_head)*, voltages of each load in the reduced system were compared with the original load model using MSE as the error metric.

4 Result

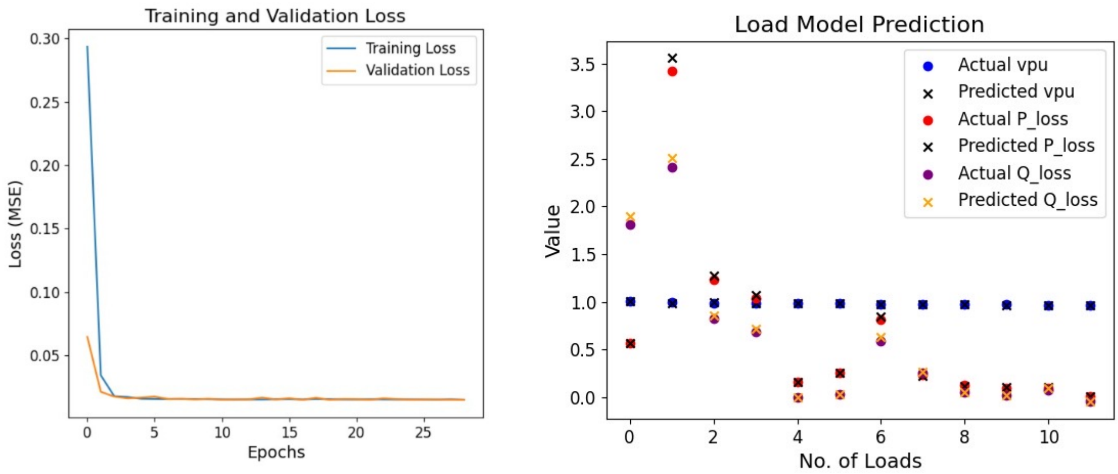
4.1 Research Results

The following findings are based on the longest branch depicted in *figure2*, which includes buses 6-18, and our efforts to develop a reduced equivalent model as shown in *figure3*. These findings pertain to the loads-only model:

1. **Accuracy of Machine Learning Model**
   The machine learning model was deployed to predict parameters such as voltage LN per unit (vpu) of each load in the branch, active power loss P_{loss}, and reactive power Q_{loss} across the line sections. The model achieved a mean squared error (MSE) of approximately 0.015, evidently shown in *figure9*.

2. **Overall Power Loss Using Predicted Values**
   Summing the predicted values of P_{loss} and Q_{loss}, the overall predicted P_{loss} was 8.3107, and the overall predicted Q_{loss} was 7.0382. In comparison, the manually calculated values from the OpenDSS simulation were 8.0719 for P_{loss} and 6.976 for Q_{loss}.

3. **Parameter Estimation Accuracy in Reduced and Original Model**
   To measure the discrepancies in parameter predictions, Mean Squared Error (MSE) was used as a metric to compare the values of reduced and original load simulations. The error calculated for the voltage line to neutral per unit for the entire 33-bus system was 0.04. MSE for total active and reactive power was 0.0008 and 0.00059 respectively. Whereas the *I_head's* error score was approximately 2.1009.



(a) Training vs. Validation Loss (b) Precision in Load Predictions

Figure 9: ML Accuracy for Load Model

Table1 shows the values from the circuit summary extracted from the original loads only and the reduced model.

	Original	Reduced
Max p.u Voltage	1.0488	1.0488
Min p.u Voltage	0.96825	0.99753
Total Active Power	3.5926 MW	3.58002 MW
Total Reactive Power	1.70019 Mvar	1.68095 Mvar
Total Active Losses	0.1192 MW (3.319%)	0.0966 MW (2.699%)
Total Reactive Losses	0.0782786 Mvar	0.0589674 Mvar

Table 1: Circuit Summary of Original and Reduced Load Models on OpenDSS

For PV and BESS simulation as well, the machine learning model was deployed to predict parameters such as voltage LN per unit (vpu) of each load in the branch, active power loss P_{loss}, and reactive power Q_{loss} across the line sections. The model achieved a mean squared error (MSE) of approximately 0.014, evidently shown in *figure10*

4.2 Discussion

The results from the analysis phase provided valuable insights into the effectiveness of the ML models for power system parameter estimation. Here are the interpretations of the findings from section 4.1:

1. **Accuracy of ML model**
   The accuracy of the machine learning (ML) model in predicting power network parameters for both loads only and PV BESS simulations demonstrates its effectiveness in cap-

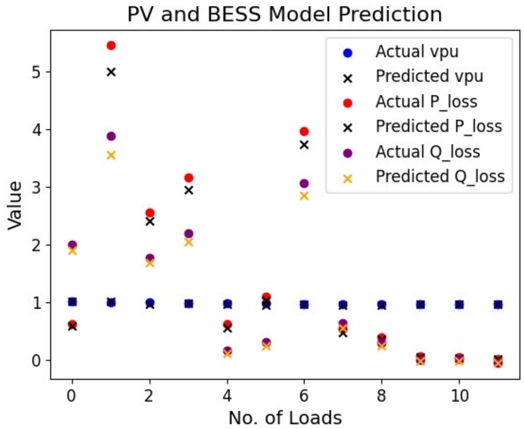


Figure 10: ML Accuracy in parameter prediction of PV BESS Model

turing complex interactions within the system. The low mean squared error (MSE) score of approximately 0.015 indicates that the ML model was proficient in learning and predicting voltage line to neutral per unit (vpu), active power loss P_{loss}, and reactive power loss Q_{loss} for each load in the studied branch. This efficiency underscores the model's ability to discern intricate relationships among the parameters, leading to highly accurate predictions. Its accuracy for both simulations further proved the credibility of the model developed in predicting these parameters for simulations with varying configurations.
Furthermore, comparing the overall predicted P_{loss} and Q_{loss} (8.3107 and 7.0382, respectively) with manually calculated values from OpenDSS simulations (8.0719 for P_{loss} and 6.9976 for Q_{loss}) reveals a small marginal difference. Each load's deviation from manual calculations was nearly 0.025, highlighting the ML model's capability to predict power losses across the branch with precision. This discrepancy also reflects the comprehensive nature of the dataset used for training, which effectively encapsulated the complex relations governing power network behavior.

2. **Error analysis of reduced equivalent simulation**
The findings from our research reveal significant insights into the parameter estimation accuracy between the reduced and original models. For the voltage line to be neutral per unit across the entire 33-bus system, an MSE of 0.04 was observed, indicating a relatively low level of discrepancy between the reduced and original simulations. This suggests that the reduced model retains a high degree of accuracy in voltage estimation, which is crucial for maintaining voltage stability in the network. Further analysis of the total active and reactive power reveals MSE values of 0.0008 and 0.00059, respectively. These extremely low error values underscore the precision of the reduced model in capturing the dynamics of active and reactive power within the system. Such accuracy is essential for effective load flow analysis and ensures that the reduced model can be reliably used for

power distribution planning and operational studies.

However, the *I_head's* error score, at approximately 2.1009, indicates a comparatively higher level of discrepancy. This higher error could be attributed to the complex interactions and non-linearities present in the head current calculations, which the reduced model may not fully capture. This highlights an area where further refinement of the model is necessary to improve its accuracy.

A comparison of parameter values of both models shown in *Table1*, shows how accurate the reduced equivalent model using network aggregation is, but those small discrepancies are justifiable since the difference could account for the lines that are now removed from the reduced network.

Overall, the low MSE values for voltage, active, and reactive power demonstrate that the reduced model performs well in most aspects, maintaining a high fidelity to the original model. This signifies the potential of the reduced model to serve as an efficient and reliable alternative for extensive simulations, enabling quicker analyses without substantial loss of accuracy.

5 Limitation or Key Challenges

1. **Learning New Software (OpenDSS)**: One of the challenges we faced was learning how to use OpenDSS. As it was new to us, we needed time to get familiar with its features and functionalities. This learning curve was steep and, at times, constrained our progress, but it was crucial for conducting our simulations effectively.

2. **Building training dataset**: Dealing with rank-3 tensors that had an inhomogeneous shape. Manipulating these tensors was complex because the data structure contained varying lengths and dimensions, which required careful handling to ensure consistency across all samples. Standard operations like flattening, slicing, indexing, and reshaping became more intricate.

3. **Choosing and Implementing Algorithms**: Determining the appropriate machine learning algorithms and implementing them with TensorFlow proved demanding. We had to make informed choices about which algorithms would be most effective for our research. This process required considerable learning and adjustment.

4. **Parameter calculation**: When calculating new ZIPV parameters for reduced model values initially did not satisfy the conditions we wanted due to the assumption that the aggregate model will have the same base power as the original model.

5. **Reducing Model Complexity and Validating Results**: A major challenge was simplifying the model while maintaining its accuracy. We carefully validated the results to ensure they were reliable. When inconsistencies occurred, we had to troubleshoot and find the source of the problem, which was a challenging aspect of the process.

6. **Scope of Simulation and Testing**: Our focus on specific branches of the IEEE 33-bus system was necessary for the project, but it also meant we couldn't explore every possible scenario. Deciding which branches to focus on, and then remodeling the entire analysis to concentrate on branches 6-18, was a challenge itself.

6 Role of Undergraduate Researchers

1. **Batool Zehra Ladha**:
   Batool is a rising Electrical Engineering Senior with an adequate background in power systems. Her role was to guide the computer science domain research members with power system modeling and simulation which was needed to ultimately generate the training data. Batool polished her skills in power systems theory, modeling, and analysis, particularly in the context of distribution networks using OpenDSS. Batool also assisted in the validation of models before and after the application of ML in the project.

2. **Manal Hasan**:
   Manal is a computer science rising junior. Her role in the project was to understand the distribution network aggregation problem and solve it using ML tools particularly neural networks. In this research Manal learned about the power system modelling using OpenDSS, solving a power systems problem using ML and generation of training data using python interface of OpenDSS.

3. **Rubab Shah**:
   Like Manal Rubab is also a computer science rising junior. She worked together with Manal in python-based generation of training data for the project using OpenDSS. Rubab developed good know how of python-based OpenDSS modeling and specification of training data requirements for application of ML to power systems problems.

7 Recommendation For Future Work

1. **Data Enhancement**: To improve the accuracy of ML models, it is recommended to expand the dataset with more diverse operational scenarios. This includes incorporating additional DERs and smart loads, as well as considering extreme weather conditions and

peak load situations. Enhanced datasets will enable more robust model training and validation.

2. **Advanced Load Modeling**: Future research should focus on refining load models by including factors like load elasticity and responsiveness to voltage and frequency changes. This will provide a more realistic representation of load behavior in modern power systems, enhancing the predictive accuracy of ML models.

3. **Integration of Real-Time Data**: The deployment of advanced data acquisition tools such as Phasor Measurement Units (PMUs) and smart meters should be expanded. Real-time data collection will facilitate dynamic analysis and allow for timely adjustments in grid management strategies.

4. **Implementation of IEEE 1547 Control Modes**: There is a need for further investigation into the real-time implementation of control modes specified in the IEEE 1547 standard, particularly for PV systems. Understanding these modes' practical application will enhance coordination between DERs and the traditional grid, optimizing performance and ensuring stability.

5. **Scalability and Application**: The methodologies developed in this study should be tested on larger and more complex distribution networks. Scaling the approach to different grid configurations will validate its applicability and effectiveness in diverse settings.

8 Future work in our consideration

For future work, we plan to significantly enhance the accuracy and reliability of our models by extending our analysis to additional branches of the IEEE 33-bus system. So far, we have concentrated on branches 6-18, but our goal is to broaden this focus to include other branches within the system. We are also aiming for a conference paper that will highlight the findings and methodologies from the work we've done up to this point. This paper will cover our current results and the improvements we've made to our models.

To address the higher discrepancy observed in the *I_head* parameter, future work should focus on refining the reduced model's ability to capture the complex interactions and nonlinearities associated with head current calculations. This could involve incorporating more detailed representations of network components or enhancing the model's approximation techniques.

Additionally, developing a machine learning (ML) model specifically designed to predict the current at the head of the branch could significantly improve accuracy. The model developed during the STRP was unsuccessful in predicting the *I_head*. By training the ML model on a comprehensive dataset that includes various operating conditions and configurations that affect

I_head, we can leverage its ability to identify patterns and relationships that traditional models may overlook. This approach would not only enhance the accuracy of *I_head* predictions but also provide a robust tool for dynamic and real-time analysis in power distribution systems.

9 Conclusion

This research highlights the transformative potential of artificial intelligence (AI) and machine learning (ML) in modernizing power distribution networks. By integrating distributed energy resources (DERs) such as photovoltaic (PV) systems and battery energy storage systems (BESS) with traditional load models within the IEEE 33-bus system, a robust framework for data generation and analysis is developed using OpenDSS and machine learning.

Our findings indicate that ML models can effectively estimate essential parameters, such as voltage, current, and power, across various operational scenarios. Our findings demonstrate that ML models are proficient in estimating crucial parameters, including voltage, current, and power, across various operational scenarios. The inclusion of ZIPV parameters significantly enhanced the models' ability to capture the complex dynamics of DERs, which is vital for grid stability and reliability. However, the research also targeted the challenges, including the limited availability of comprehensive data and the complexities of modeling bidirectional power flows, particularly with the increasing penetration of DERs. These challenges highlight the need for ongoing research and development to refine data collection methods and enhance model accuracy for predicting grid behavior under diverse conditions.

The advancements underscore ML's potential to revolutionize grid management and operational efficiency. Future research should address these challenges by improving data quality and developing advanced models to better navigate the complexities of modern power systems. This study provides valuable insights for advancing power system management and supports the shift towards more resilient and adaptable energy networks.

References

[1] S. M. H. Rizvi, S. K. Sadanandan, and A. K. Srivastava, "Real-time parameter tracking of power-electronics interfaced composite zip load model," *IEEE Transactions on Smart Grid*, vol. 13, no. 5, pp. 3891–3902, 2022.

[2] S. M. H. Rizvi, S. K. Sadanandan, and A. K. Srivastava, "Real-time zip load parameter tracking using sensitivity-based adaptive window and variable elimination with realistic synchrophasor data," *IEEE Transactions on Industry Applications*, vol. 57, no. 6, pp. 6525–6536, 2021.

[3] "Ieee standard for interconnection and interoperability of distributed energy resources with associated electric power systems interfaces," *IEEE Std 1547-2018 (Revision of IEEE Std 1547-2003)*, pp. 1–138, 2018.

[4] Z. K. Pecenak, H. V. Haghi, C. Li, M. J. Reno, V. R. Disfani, and J. Kleissl, "Aggregation of voltage-controlled devices during distribution network reduction," *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 33–42, 2021.