Synthetic Voltage Data Generation, Modeling and Characterization of Power Quality Issues in AC Line Supply

Mr S Praveen[1]*, Dr. Ezhilarasan G[2],
Dr. Mahendra B M[3], Goutham B Bhandari[4], Vishwas R B[4], Naveen Raju S[4]
Dept. of ECE, Jain University, Bangalore[1], Dept. of EEE, Jain University, Bangalore[2], School of Computational and Data Sciences, Bangalore[3]
Dept. of ECE, RV College of Engineering, Bangalore[4]

ABSTRACT

AC power supply is critical for modern electrical systems, providing efficient energy transmission over long distances and powering a wide range of residential, commercial, and industrial applications. Its reliability and quality are essential for ensuring the optimal performance of connected devices and systems. Noise in real-time AC signals is generated due to factors like sudden load changes, switching operations in power systems, and external electromagnetic interference. Understanding and mitigating noise-induced deviations in AC signals is crucial for improving energy efficiency and system stability. This Work explores the challenges and solutions related to AC power quality, emphasizing the reconstruction of AC signals with embedded noise. Deviations from ideal sinusoidal waveforms, caused by noise components such as harmonic distortion, flicker, transients, and thermal noise, significantly impact energy efficiency and system reliability. Using data from a 230 V, 50 Hz AC supply, this study characterizes noise sources and employs machine learning models- Ridge Regression, K-Nearest Neighbors (KNN), and Random Forest for signal reconstruction. Random Forest achieved superior accuracy (R^2 = 0.98), effectively capturing linear and non-linear noise patterns. Preprocessing techniques like baseline normalization proved essential for enhancing model accuracy. Reconstructed signals demonstrated practical applications in smart grids, IoT devices, and diagnostic systems. The study sets a benchmark for signal reconstruction, offering insights for scalable, real-time solutions in power quality management and beyond. Research in this domain has explored the use of synthetic training datasets for grid stability analysis, which proved instrumental in improving the reliability of real-time power systems.

Keywords: AC signal reconstruction, power quality(PQ) , Machine learning(ML), K-Nearest Neighbors(KNN), Ridge Regression, Random Forest.

1. INTRODUCTION

In contemporary electrical and electronic systems, signal and power quality are pivotal to ensuring system dependability, efficiency, and performance. Signal quality pertains to the ability of an electrical signal, whether voltage or current, to retain its intended properties free from distortion or interference. With the increase in the addition of non-linear loads to electrical systems, it has become necessary to ensure that voltage degradation does not occur at the transmission or distribution stage. Although quality problems in DC power systems are frequently less complicated, they are nonetheless quite important. However, because of the sinusoidal shape of the voltage and current waveforms, AC power systems are more complicated and dominate contemporary electrical grids.Various practical problems associated with Power Quality namely Harmonic Distortions, Sags/Swells, Under/Over Voltages and Transients are major reasons for low quality power signals[3]. Various solutions exist to safeguard sensitive loads from the consequences of voltage disturbances. Among these solutions, power electronic devices serve as highly efficient and adaptable compensators, including series, parallel, and series-parallel compensators commonly referred to as custom power devices[1]. The growing use of electronic devices in Various applications has made the creation of dependable and effective voltage regulation systems necessary[2] along with that Maintaining high signal quality is also essential for accurate information transmission and the reliable operation of devices. Complementarily, power quality evaluates the suitability of an electrical power source to facilitate the smooth operation of connected systems. Poor power quality can result in system inefficiencies, equipment malfunctions, or even catastrophic failures.

Traditionally, probabilistic approach has been used for time varying signals in a power quality analysis assuming that the power line disturbance components vary too slowly to affect the accuracy of the analytical process[4]. Furthermore, a different investigations utilizes Wavelet Transform combined with optimized Artificial Neural Networks (ANN) [5] and Hilbert-Huang Transform along with feedforward neural networks[6] to classify power quality disturbances, indicating advancements in analytical techniques.there is a discussion on techniques for evaluating generative models, providing insights into assessing the performance of models used in power quality analysis and highlighting the relevance of generative approaches in this context [7].Addressing challenges related to signal and power quality requires sophisticated techniques that leverage advancements in signal processing, machine learning, and computational intelligence.

Synthetic data generation has also emerged as a promising solution to enhance the performance of machine learning models in small-signal stability assessments. Research in this domain has explored the use of synthetic training datasets for grid stability analysis, which proved instrumental in improving the reliability of real-time power systems [8][9]. Enhanced machine learning models have been utilized to assess voltage stability and improve predictive accuracy in complex power systems [10].Recent advances have expanded into the application of generative models for signal reconstruction and data augmentation. Diffusion models have been employed to generate synthetic data aimed at enhancing noise resilience in digital VLSI circuits [11]. Additionally, generative models have been shown to reconstruct high-fidelity signals with embedded noise, demonstrating the importance of model optimization for improving signal quality in data generation pipelines [12].

This paper builds upon these foundational works by focusing on synthetic voltage data generation and its utility in reconstructing AC signals affected by noise. The proposed methodologies, including Ridge Regression, K-Nearest Neighbors, and Random Forest algorithms, are evaluated for their performance in identifying and mitigating noise patterns in synthetic voltage data. These techniques aim to address challenges in smart grid applications, IoT devices, and low-power VLSI systems, where maintaining signal integrity is critical. By integrating machine learning techniques with robust noise characterization, this work seeks to set a benchmark for further innovation in power and signal quality enhancement.

2. METHODOLOGY

This section details the actual processes required to record data to assess the noise characteristics, preprocess the data, and employ machine learning models to construct the AC signal. Realisation of the AC signal waveform in the actual environment while keeping its inherent noise patterns is the key aim.

2.1 Data Logging

The V_{RMS} values of a standard 230 V, 50 Hz AC power supply are recorded using the Fluke 434 II Energy Analyzer, a high-accuracy equipment for power quality assessment. The analyzer acquired V_{RMS} measurements at 1-second intervals, yielding a dataset of 10,000 data points collected over a period of time. These measurements offered exceptional granularity, ensuring even slight fluctuations in the signal are noticed.

Analysis of the collected data indicated the presence of voltage fluctuations within a range of roughly ±5 V from the baseline. These oscillations are linked to noise created by elements like load changes, circuit disruptions, and external interference, making the signal non-ideal for practical applications without adequate modeling.

2.2 Signal Characterization

The recorded signal is analyzed to identify and quantify its noise components. Given that the noise is 3% compared to the original signal, this value is normalized to represent 100% of the overall noise for characterization purposes. The noise breakdown is as follows:

A. **Damping Noise (40%):** Caused by the dissipation of energy in resistive and capacitive components, resulting in a drop in oscillation amplitude over time.

B. **Flicker Noise (30%):** Low-frequency oscillations in voltage owing to load changes, occasionally noticeable as flickering in electrical equipment.

C. **Harmonic Distortion (15%):** Arising from non-linear loads, harmonic distortion provides additional frequencies deviating from the pure sinusoidal waveform.

D. **Transient Noise (10%):** Short-lived voltage spikes or dips generated by events like switching operations or electrical failures.

E. **Thermal Noise (5%):** Random high-frequency oscillations attributable to the thermal mobility of charge carriers in components.

This classification of noise is vital to exactly replicate its influence in the reconstructed signal, ensuring the final waveform matches the real-world ac signal scenario.

2.3 Data Preprocessing

The preprocessing procedure involved removing the noise component from the baseline voltage. The baseline is computed as the mean of all V_{RMS} values:

To extract the noise, each V_{RMS} reading is altered by removing this baseline:

$$N[i]=V_{RMS}[i]-V_{baseline} \qquad 1.1$$

This phase converted the dataset into a form where the noise is centered around zero, allowing machine learning models to focus exclusively on learning the noise patterns. By normalizing the signal in this manner, computational stability and model correctness are strengthened.

2.4 Signal Generation Using Machine Learning

The recovered noise is modeled using three machine learning algorithms: Ridge Regression, K-Nearest Neighbors (KNN), and Random Forest. The purpose is to forecast the noise levels for each time point, which, when added back to the baseline, rebuilt the original V_{RMS} signal. Each type had unique strengths:

A. **Ridge Regression**: A linear model with regularization to prevent overfitting. It sought to represent model the overall trend in noise using a simple linear equation:

$$N[i]=\alpha x+\beta N[i] \quad 1.2$$

While fast and interpretable, Ridge Regression struggled to capture the non-linear components of the noise

B. **K-Nearest Neighbors (KNN)**: This model assessed noise based on the average values of the 5 closest data points in the training set. By utilizing local patterns, KNN effectively simulated smaller-scale oscillations but needed considerable processing resources for big datasets.

C. **Random Forest Regressor**: This is a robust ensemble approach that integrated many decision trees. Each tree is trained on a subset of the data, and their outputs are averaged to create the final noise prediction. This method succeeded in capturing complicated non-linear interactions, making it the most successful model for this purpose.

The reconstructed signal is created by adding the baseline voltage and the expected noise:

$$V_{RMS}[i] = V_{baseline} + N[i] \quad 1.3$$

Random Forest exhibited the best accuracy, with an R^2 score of 0.98, indicating that the reconstructed signal closely matched the original waveform, including its noise characteristics.

3. IMPLEMENTATION

The implementation step required the combination of data preparation, machine learning models, and signal reconstruction into a cohesive workflow. This facilitated the continuous creation of a genuine AC signal.

3.1 Data Preprocessing

The raw V_{RMS} values are imported from the dataset, and the baseline voltage is computed as the mean of all observations. By deleting this baseline, the noise component is isolated, making the data more acceptable for machine learning systems. This preprocessing approach not only centered the noise around zero but also enhanced computation efficiency by decreasing the data structure.The following figure 1 depicts the input signal obtained.

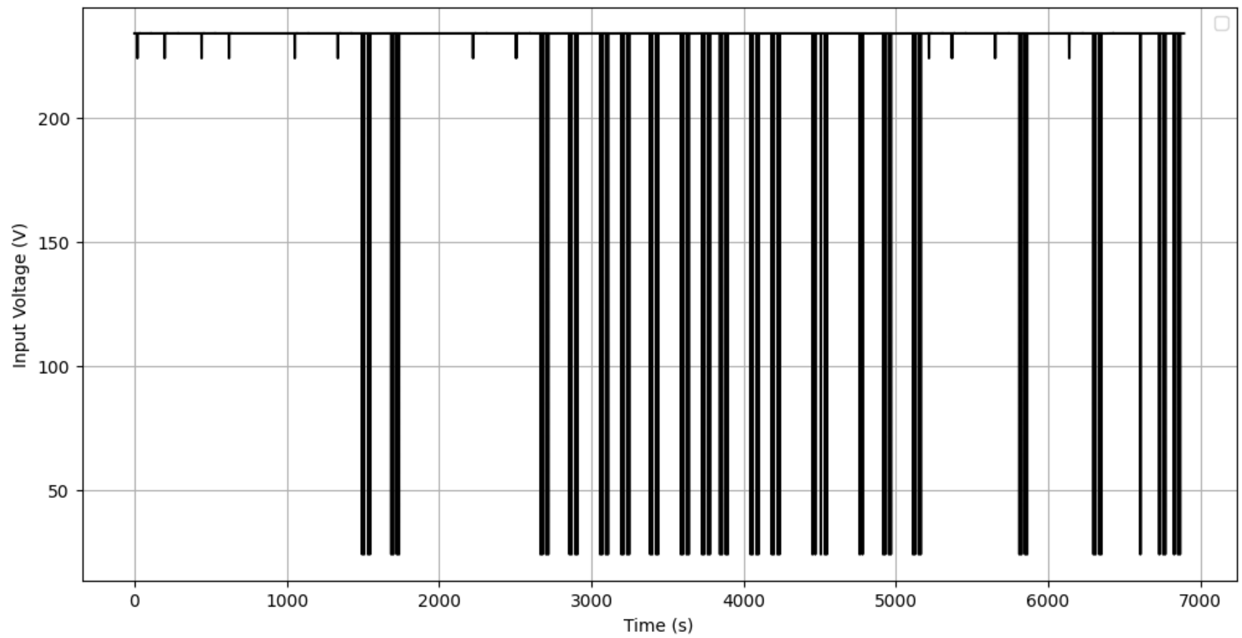

Figure 1: The input Voltage waveform

3.2 Model Training and Evaluation

As discussed in section 2.4 three models are considered for training and evaluation.

A. **Ridge Regression**:

Ridge Regression is a linear model that incorporates a regularization parameter to prevent overfitting. The model approximates the link between the noise and its time indices using a simple linear equation:

$$N[i]=\alpha x+\beta N[i] \quad 1.4$$

Here, α signifies the slope, and β the intercept. While computationally efficient and easy to grasp, this model struggled to capture non-linear noise patterns inherent in the data

B. **K-Nearest Neighbors (KNN)**:
KNN is a non-parametric approach that predicts the noise at a given time point by averaging the values of its k-nearest neighbors. With k=5k = 5k=5, the model efficiently detected local noise patterns, especially those with periodic characteristics. However, the computational cost of identifying nearest neighbors escalated with the dataset size, making KNN less practical for large-scale real-time applications.

C. **Random Forest Regressor**:
Random Forest is an ensemble learning strategy that builds several decision trees during training and averages their predictions to increase accuracy. Each tree is trained on a

random sample of the data, allowing the model to capture both linear and non-linear correlations in the noise. This resilience made Random Forest the most realistic model for reproducing the noise characteristics of the original signal.

The training step entailed fitting each model to the noisy data acquired from the training subset and assessing their performance using the testing subset
The R^2 score is chosen as the primary criterion to determine how well each model could predict the noise.

3.3 Signal Reconstruction

The estimated noise levels from each model are fed back to the baseline voltage to recreate the V_{RMS} signal. The Random Forest model is identified as the optimum answer thanks to its superior capabilities in maintaining both linear and non-linear noise patterns.

$$V_{RMS}[i]=V_{baseline}+N_{RF}[i] \qquad 1.5$$

This reconstructed signal is therefore available for real-time applications, where it may imitate real-world AC waveforms for low-power settings. The following figure 2 refers to the learning rate of various machine learning model used

The following figure 2 represent the reconstructed waveform with the original waveform which is completely aligned.

4. RESULT ANALYSIS

The results demonstrate that the suggested methodology is effective for reconstructing AC signals with embedded noise characteristics, with three machine learning models performing particularly well: Ridge Regression, K-Nearest Neighbors (KNN), and Random Forest. Each model is trained and evaluated using a dataset of over 10,000 real-world V_{RMS} measurements from a 230 V, 50 Hz AC power supply, which includes noise components like damping, flicker, harmonic distortion, transients, and thermal noise. Figure 3 depicts achieving ridge Regression of R^2 value of 0.75, demonstrating computational efficiency but struggling with nonlinear noise patterns. KNN, with an R^2 value of 0.86, efficiently captures localized oscillations but faces scalability issues. **Overall noise found in the signal is 3% which is categorized into various** noise components revealing a detailed characterization as shown in Table 1.

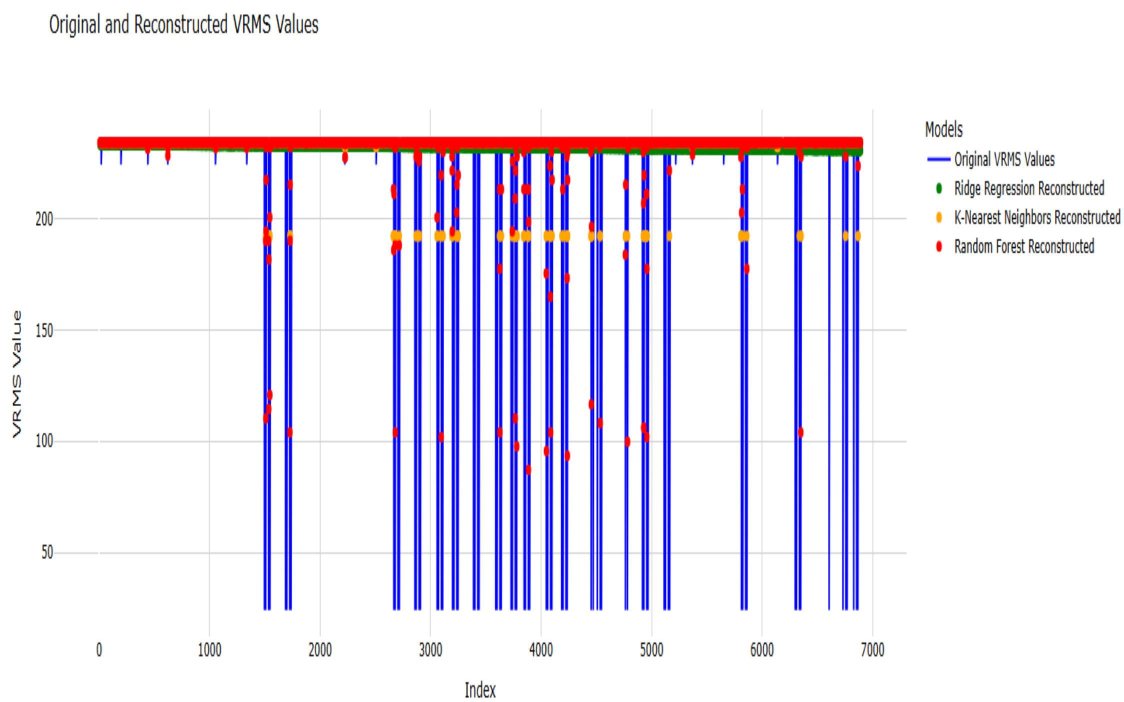

Figure 2. The original and reconstructed waveform

TABLE 1 : Categories of Noise with their respective values

Sl.no	Noise type	Amount of Noise present
1.	Damping Noise	40%
3.	Flicker Noise	30%
4.	Harmonic Distortion	15%
5.	Transient Noise	10%
6.	Thermal Noise	5%

The various types of noise include damping noise, which demonstrates gradual amplitude decay; flicker noise, which reflects load change-induced variations; harmonic distortion, which highlights frequency deviations; transient noise, which captures abrupt spikes; and thermal noise, which reconstructs random high-frequency patterns. Preprocessing approaches, such as baseline voltage normalization and noise isolation, are critical to the models' success, demonstrating the durability of the machine learning-driven methodology.

Figure 3: R^2 score of various models which includes KNN,Ridge regression,Random forest

TABLE 2 : MODEL PERFORMANCES (R^2 Scores)

Model	**R^2 (Noise Prediction)**	**R^2(Reconstructed VRMS)**
Ridge Regression	0.945	0.953
K- Nearest Neighbors	0.811	0.838
Random Forest	0.968	0.976

When compared to traditional signal processing methods, the machine learning-based methodology proves superior, as it adjusts dynamically to data-driven noise patterns rather than relying on predefined parameters. This adaptability leads to improved real-time noise modeling, accuracy, and scalability. The reconstructed signals closely match real-world waveforms, indicating practical applications across multiple domains. For instance, in smart grids, precise signal reconstruction improves diagnostics and performance monitoring. The method is appropriate for low-power devices in IoT and embedded systems due to its scalability and excellent input fidelity. Additionally, reconstructed signals serve as benchmarks for evaluating and optimizing electrical and electronic equipment. These findings emphasize the potential of combining powerful machine learning models with well-defined preprocessing pipelines for practical applications in current electrical systems, thereby establishing a high standard for signal reconstruction and power quality management.

Despite the optimistic findings, the study reveals limitations and opportunities for further improvement. While Random Forest outperforms the other models, its computing requirements may limit its usefulness in resource-constrained settings. To address this, lightweight neural network designs or hybrid models that combine Random Forest and deep learning can improve efficiency. Furthermore, incorporating other noise sources, such as electromagnetic interference and environmental perturbations, increases the methodology's usefulness. In a nutshell, the analysis demonstrates the impact of the novel approach on the advancement of power quality analysis and confirms its significance. The knowledge gathered not only validates the applicability of the methodology but also opens the door for further study and commercial applications aimed at ensuring the reliability and effectiveness of electrical systems.

5. CONCLUSION

This study offers a thorough investigation of AC power quality problems, emphasising their origins, consequences, and creative fixes via signal reconstruction. The study underlines the complexity brought about by departures from ideal sinusoidal waveforms and stresses how important it is to preserve signal and power quality for the effective operation of contemporary electrical systems. A thorough framework for comprehending the complex nature of power quality interruptions was provided by the careful characterisation of factors like flicker, transients, damping, harmonic distortions, and thermal noise. The technique used connects theoretical understanding with real-world application. The study showed how well these models could describe and recreate real-world noise patterns hidden in AC signals by utilising cutting-edge machine learning approaches like Ridge Regression, K-Nearest Neighbours (KNN), and Random Forest. With an astounding R2 score of 0.98—indicating an almost flawless alignment between the reconstructed and original signals—the Random Forest model was the most successful of them. This result emphasises how well ensemble approaches capture the linear and non-linear properties of noise components.

The study also emphasises how useful data preparation methods are for separating and normalising noise components so that machine learning models can only concentrate on the variability of the signal. In addition to enhancing computational stability, this preprocessing step guaranteed the accuracy of model predictions. This study offers a strong basis for enhancing the functionality and dependability of electrical systems by methodically tackling the problems of signal creation and noise reconstruction.In summary, by fusing cutting-edge analytics with workable solutions, this study significantly advances the fields of signal processing and power quality control. It creates a thorough framework for examining and reassembling AC signals, establishing a standard for further research in this area. In addition to advancing the state of the art, the approaches and results reported here open the door to more accurate, scalable, and efficient applications in electrical engineering and related domains.

6. FUTURE SCOPE

The research provides a robust foundation for advancing power quality analysis and AC signal reconstruction, opening avenues for further exploration. Optimizing machine learning models for scalability and real-time applications is critical for handling larger datasets in dynamic grids and industrial systems. Advanced architectures like CNNs and RNNs can enhance noise pattern predictions for complex, time-sensitive data. Hybrid approaches combining deep learning and ensemble techniques may balance computational efficiency with accuracy. Adaptive models using reinforcement learning could improve resilience to real-time noise fluctuations.

Practical implementation via FPGA-based systems can enable integration into industrial automation, smart grids, and portable tools. Expanding the methods to address electromagnetic interference and environmental disturbances will enhance applicability. Cross-disciplinary applications include telecommunications, medical electronics (e.g., ECG, EEG), and renewable energy systems for analyzing power quality. Signal reconstruction supports predictive maintenance, reducing downtime and boosting system reliability. Integration with smart grid technology allows real-time diagnostics, improved energy distribution, and reduced losses.

Future directions also include environmental impact assessments, energy efficiency evaluations, and participation in international standardization to ensure reconstruction accuracy and diagnostic reliability. These advancements will make AC signal reconstruction more scalable, accurate, and relevant, driving innovation in energy management and sustainable power systems globally.

7. REFERENCES

[1] S. Praveen, Dr. Ezhilarasan G, Jain University, Bangalore, " A Literature Review on Voltage Regulation to Overcome Power Quality Issues", Journal of Xi'an University of Architecture & Technology, Volume XVI, Issue 7, 2024, ISSN No : 1006-7930, 188-202. https://doi.org/10.37896/JXAT16.7/34411

[2] Mr S Praveen, Dr. Ezhilarasan G, Nandini Agrawal, Rakshita G Y, Riya Ketankumar Shah, Tanvi R, “Review on Design of Voltage Regulator to Overcome Power Quality Issues in Biomedical Applications”. – Journal of Xi'an University of Architecture & Technology, Volume XVI, Issue 12, 2024 Page No : 173-193. https://doi.org/10.37896/JXAT16.12/35313

[3] S. Govindarajan and S. S. Balakrishnan, "Analysis and interpretation of Power Quality issues with suitable corrective measures," *2013 International Conference on Green Computing, Communication and Conservation of Energy (ICGCE)*, Chennai, India, 2013, pp. 479-484, doi: 10.1109/ICGCE.2013.6823484.

[4] P. Kanirajan and V. Suresh Kumar, "Power quality disturbance detection and classification using wavelet and RBFNN," *Appl. Soft Comput. J.*, vol. 35, pp. 470-481, 2015.

[5] M. Ijaz, M. Shafiullah, and M. Abido, "Classification of power quality disturbances using Wavelet Transform and Optimized ANN," *2015 18th International Conference on Intelligent System Application to Power Systems (ISAP)*, 2015.

[6] S. Alshahrani, M. Abbod, and G. Taylor, "Detection and classification of power quality disturbances based on Hilbert-Huang transform and feed forward neural networks," *Proc. – 2016 51st Int. Univ. Power Eng. Conf. UPEC 2016*, vol. 2017-January, pp. 1-6, 2016.

[7] L. Theis, A. van den Oord, and M. Bethge, "A note on the evaluation of generative models," *International Conference on Learning Representations*, Apr. 2016.

[8] S. A. Dorado-Rojas, M. de Castro Fernandes, and L. Vanfretti, "Synthetic Training Data Generation for ML-based Small-Signal Stability Assessment," *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, Tempe, AZ, USA, 2020, pp. 1-7, doi: 10.1109/SmartGridComm47815.2020.9302991.

[9] M. Endres, M. Venugopal, and T. Tran, "Synthetic Data Generation: A Comparative Study," *Proceedings of the 26th International Database Engineered Applications Symposium*, 2022.

[10] O. Lennerhag and M. Bollen, *Power Quality*, Springer Handbooks, pp. 1171-1203, 2021

[11] P. Srivastava, P. Kumar, and Z. Abbas, "Enhancing ML model accuracy for Digital VLSI circuits using diffusion models: A study on synthetic data generation," *2024 IEEE International*

Symposium on Circuits and Systems (ISCAS), Singapore, Singapore, 2024, pp. 1-5, doi: 10.1109/ISCAS58744.2024.10558345.

[12] M. S. Sakib, M. R. Islam, S. M. S. H. Tanim, M. S. Alam, M. Shafiullah, and A. Ali, "Signal Processing-based Artificial Intelligence Approach for Power Quality Disturbance Identification," *2022 International Conference on Advancement in Electrical and Electronic Engineering (ICAEEE)*, Gazipur, Bangladesh, 2022, pp. 1-6, doi: 10.1109/ICAEEE54957.2022.9836389.

[13] H. -Y. Su and T. -Y. Liu, "Enhanced-Online-Random-Forest Model for Static Voltage Stability Assessment Using Wide Area Measurements," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6696-6704, Nov. 2018.

[14] D. Amuru, A. Zahra, H. V. Vudumula, P. K. Cherupally, S. R. Gurram, A. Ahmad, et al., "AI/ML algorithms and applications in VLSI design and technology," *Integration*, vol. 93, pp. 102048, 2023.

[15] M. V. Phanindra Kumar, A. Thayyilravi, R. Karthika, C. R. Manoj, and R. Kumar Soni, "An Approach to Translate Synthetic Lidar Point Cloud Data Similar to Real Lidar Data and Evaluate Using Deep Learning Techniques," *2023 IEEE 3rd Mysore Sub Section International Conference (MysuruCon)*, Hassan, India, 2023, pp. 1-6, doi: 10.1109/MysuruCon59703.2023.10396953.

[16] A. Munappy, J. Bosch, H. H. Olsson, A. Arpteg, and B. Brinne, "Data management challenges for deep learning," *2019 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, pp. 140-147, 2019.

[17] D. Viana, R. Teixeira, J. Baptista, and T. Pinto, "Synthetic Data Generation Models for Time Series: A Literature Review," *2024 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, Sydney, Australia, 2024, pp. 1-6, doi: 10.1109/ICECET61485.2024.10698494.

ACKNOWLEDGMENT

Dr. K U Vinayaka Assistant Professor, Electrical and Electronics Engineering has been very helpful in classification of power quality issues and always provided the guidance.