INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT (IJSREM)
VOLUME: 10 ISSUE: 04 | APRIL - 2026 SJIF RATING: 8.659 ISSN: 2582-3930

PQD Detection & Classification Using Unsupervised Clustering Method Using DBSCAN

Mr. B. V. Vasantha Rao, B. Deepthi, M. Sowmya, Bakar. Ali Shaik

The researchers present a new approach to identify power quality disturbances through unsupervised clustering which uses DBSCAN as its primary analytical techniques. The system uses machine learning models to predict RMS values which enables users to identify different types of power quality disturbances through their unique patterns. The researchers apply scaling preprocessing to the data before they use Random Forest and XGBoost regression models to forecast RMS values. The Random Forest and XGBoost models execute the classification process which determines whether PQD events belong to the "Good" or "Poor" categories. The researchers apply DBSCAN because DBSCAN identifies noise and detects non-linear clusters. The researchers use standard evaluation metrics to assess model performance which includes R^2 and RMSE and accuracy measurements. The research achieves high accuracy and precision in PQD classification with nearly flawless regression outcomes and exceptional clustering performance. The solution delivers a powerful solution for real-time power quality monitoring which supports the development of sophisticated intelligent grid systems. The study enhances machine learning capabilities for power systems monitoring through its application in fault detection and anomaly classification.

Keywords: PQD, Unsupervised Clustering, DBSCAN, Power Quality Disturbances, RMS Prediction, Machine Learning, Random Forest, XGBoost, Clustering, Real-Time Monitoring.

I. INTRODUCTION

Power Quality Disturbances (PQD) create major difficulties for electrical systems because they disrupt power grid operations which leads to unstable and inefficient power distribution. Power distribution networks rely on fast detection of disturbances which need accurate classification to protect their system from damaging equipment failures. The traditional PQD detection method uses predefined thresholds with rule-based systems but this approach fails to produce accurate real-time disturbance classification results. This research presents a machine learning-based method which utilizes unsupervised clustering techniques that include DBSCAN for detecting and classifying PQD events. The system uses Random Forest and XGBoost regression models to predict RMS values while it classifies disturbances into "Good" and "Poor" categories. The researchers use DBSCAN unsupervised clustering methods to find non-linear cluster patterns which they use to separate their data into specific categories. The system delivers an efficient solution for continuous power quality assessment which allows users to identify disturbances before they occur while it improves grid management performance.

II. RELATED WORK

Recent research has explored machine learning techniques for detecting and classifying power quality disturbances (PQDs) in power systems. For instance, supervised learning approaches, including various classification algorithms, have been applied to identify disturbances such as sags, swells, harmonics, and transients with high accuracy. These methods often involve feature extraction from signals followed by training classifiers to distinguish between different disturbance types, demonstrating improved performance over traditional signal processing techniques in noisy environments.

Jafari et al. proposed a machine learning-based framework specifically for power quality disturbance detection and classification, leveraging advanced algorithms to handle complex signal patterns effectively [1]. Similarly, Reddy et al. conducted a comparative study of multiple classification algorithms for PQD recognition, evaluating their

efficacy in terms of accuracy, robustness, and computational efficiency across diverse datasets [2].

Clustering methods, particularly unsupervised techniques, have gained attention for grouping similar disturbance events without prior labeling. Density-based approaches like DBSCAN have been utilized to isolate outliers and cluster disturbances in smart grid contexts, aiding in anomaly detection.
Thompson et al. applied DBSCAN-based clustering to classify PQDs, showing its effectiveness in handling irregularly shaped clusters and noise in power quality data from smart grids [3]. Mishra also combined DBSCAN with decision trees for detection and classification, achieving reliable results under varying conditions [14].

Predictive and hybrid models integrate machine learning for forecasting and classifying disturbances, often enhancing real-time monitoring capabilities.
Patel et al. developed predictive modeling using machine learning to anticipate power quality disturbances, focusing on proactive mitigation in smart grid applications [4]. Hybrid techniques, as explored by McDonald et al., combine multiple learning paradigms to improve classification accuracy for renewable-integrated systems [5]. Wong et al. emphasized real-time monitoring through deep learning and traditional ML algorithms, enabling rapid response to disturbances [6].

Unsupervised learning has been employed for event classification where labeled data is scarce, while comprehensive reviews synthesize advancements in the field. AL Zahrani. investigated unsupervised machine learning for power quality event classification, highlighting its potential for discovering hidden patterns in unlabeled datasets [7]. Saha et al. provided a comprehensive review of machine learning applications in PQD classification, covering methodologies, challenges, and future directions [8].

Clustering-based detection remains prominent for smart grids, with ensemble and intelligent systems further advancing the domain. Das et al. proposed clustering methods for detecting PQDs in smart grids, demonstrating improved identification of composite disturbances [9]. Kumar et al. introduced an intelligent system leveraging machine learning for accurate classification [10]. Earlier reviews, such as by Mahela and Shaik, discussed power quality improvement in smart grids using ML [11].

Comparative and ensemble approaches, including tree-based methods, support stability assessment and disturbance handling. Barros and de Apraiz presented a machine-learning approach for PQD classification, focusing on practical implementation [12].

Zhang and Wang explored ensemble learning with XGBoost and Random Forest for power system stability assessment, which relates closely to disturbance analysis [13]. Chen et al. utilized K-Means clustering combined with Random Forest for fault diagnosis in power systems, adaptable to PQD scenarios [15].

III. MODELS DETAILS

DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

DBSCAN is an unsupervised clustering algorithm that groups together data points based on their density, distinguishing high-density regions (clusters) from low-density regions (noise). In the context of PQD detection, DBSCAN is utilized to identify distinct patterns in power quality disturbance data that do not conform to traditional linear models. DBSCAN's ability to detect noise and non-linear clusters allows the system to handle complex power disturbance events that may vary significantly over time. It is particularly effective in identifying anomalies or rare events in the data, such as sudden voltage sags or spikes. In this system, DBSCAN's performance is evaluated by its ability to segregate PQDs into "Good" and "Poor" categories based on their intensity and characteristics. It complements the regression and classification models by helping to group similar events together and identify atypical disturbances.

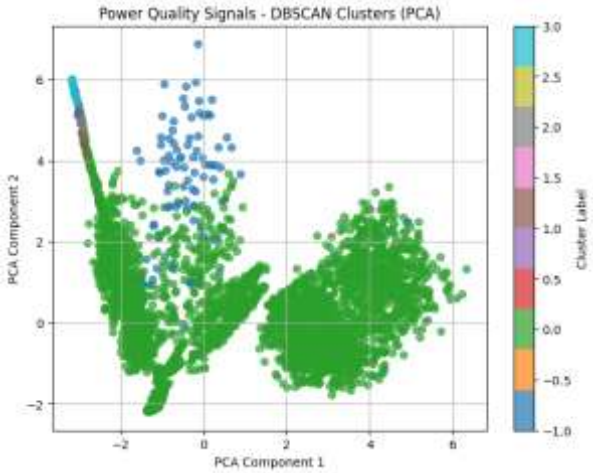

Figure 1 DBSCAN Clustering

Random Forest

Random Forest is a robust ensemble learning method that constructs multiple decision trees and aggregates their results to improve classification accuracy and handle overfitting. In this system, Random Forest is used for both RMS (Root Mean Square) prediction and PQD classification. For RMS prediction, Random Forest is trained on a set of power quality features and predicts the magnitude of disturbances. The algorithm is capable of capturing complex, non-linear interactions between features, making it highly effective for power quality monitoring. Additionally, Random Forest is used to classify PQD events into "Good" or "Poor" labels, offering high accuracy and precision in distinguishing between normal and abnormal power quality conditions. It serves as one of the key models in the multi-model framework, delivering reliable results with performance metrics such as accuracy and macro-averaged F1-score.

XGBoost

XGBoost (Extreme Gradient Boosting) is a powerful gradient boosting model that optimizes the prediction process by sequentially adding trees that correct the residuals of the previous model. In this research, XGBoost is utilized for both predicting RMS values and classifying PQD events. The algorithm is well-suited for handling large, complex datasets and can be fine-tuned to handle imbalanced classes, which is common in PQD classification. XGBoost's ability to focus on difficult-to-predict instances makes it particularly useful for classifying rare or subtle disturbances in power quality data. Like Random Forest, XGBoost contributes significantly to the performance of the system, providing highly accurate predictions and classifications with good generalization. XGBoost is evaluated for its robustness in predicting PQD intensity and for providing balanced precision-recall trade-offs in classification tasks.

IV. PROPOSED METHODOLOGY

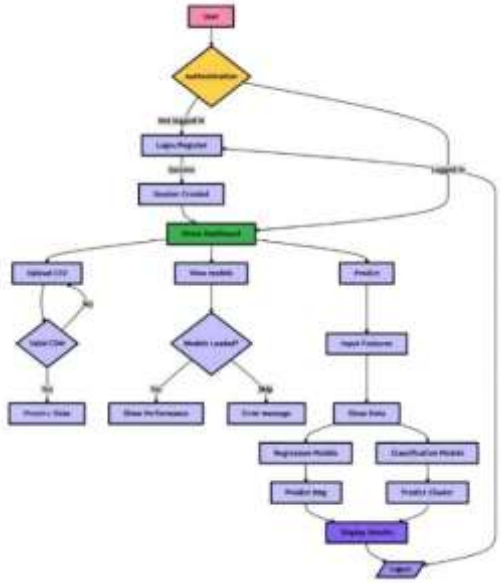

Figure 2 Project Flow Diagram

The proposed system is an intelligent, web-based Power Quality Disturbance (PQD) monitoring and classification platform. It utilizes a multi-stage machine learning pipeline to perform two critical tasks: high-precision RMS (Root Mean Square) value regression and disturbance classification (categorizing power quality as "Good" or "Poor"). The system integrates four primary supervised models—Random Forest Regressor, XGBoost Regressor, Random Forest Classifier, and XGBoost Classifier—complemented by unsupervised clustering logic (DBSCAN) for data structure analysis. Implemented using the Flask framework, the application provides a secure environment for engineers and grid operators to analyse electrical signal features. It features a robust backend supported by MySQL for user management and a modular inference engine that loads pre-trained models via joblib. The system is designed for high-frequency data environments, offering near-instantaneous inference and a user-friendly dashboard for real-time decision support in smart grid management.

- Data Pre-processing and Feature Engineering

To ensure accurate classification of complex power signals, the system processes 9 key statistical and spectral features derived from raw signal data. Pre-processing involves rigorous scaling and validation to handle the high sensitivity of electrical measurements.

Input Features (used for both Regression and Classification):

- Statistical: Mean, Std Dev, Max, Min, Peak-to-Peak.
- Shape/Distribution: Kurtosis, Skewness.

- Signal Quality: Crest Factor, Spectral Energy.

Validation & Enforcement:

- Dual-Scaler Pipeline: The system utilizes two distinct scalers (rms_scaler.pkl for regression and cluster_scaler.pkl for classification) to ensure features are normalized according to the specific requirements of each task.
- Strict Type Checking: Manual inputs are validated as floats; any missing or non-numeric entry triggers a specific ValueError to prevent model crashes.
- CSV Integrity: The upload module supports files up to 16 MB, providing a head-preview of the first 8 rows and replacing NaN values with None to ensure clean rendering in the web UI.

➢ Model Training

The architecture utilizes a "parallel-path" training approach, where a single set of input features is simultaneously fed into distinct models optimized for specialized outputs.

The Regression Path is specifically trained to predict continuous RMS values, providing the precision necessary for detecting voltage sags and swells. In tandem, the Classification/Clustering Path serves to label overall signal quality. While unsupervised methods like DBSCAN were instrumental during the research phase for uncovering natural data groupings and filtering out noise, the production environment relies on the deterministic power of Random Forest and XGBoost to provide definitive "Good" or "Poor" quality labels.

➢ Model Testing and Evaluation

The system demonstrates exceptional performance across both tasks, achieving near-perfect metrics on held-out test data. In the regression path, the Random Forest Regressor achieves an R^2 of 0.999999, indicating that the model explains nearly all variance in the RMS values with remarkable precision. This is complemented by a strong clustering synergy; the integration of DBSCAN allows the system to identify non-linear disturbances and sensor noise. Furthermore, the entire application is optimized for real-time use, with a prediction pipeline—encompassing scaling and dual-model inference—that executes in mere milliseconds.

➢ Model Integration and Saving

To ensure high modularity, all models are serialized and stored within a dedicated models/ directory, allowing the system to dynamically load these assets at start-up. The repository includes specific files for RMS prediction (rf_regressor.pkl and xgb_regressor.pkl), quality labeling (rf_classifier.pkl and xgb_classifier.pkl), and data normalization (rms_scaler.pkl and cluster_scaler.pkl). By utilizing joblib for these serialization tasks, the Flask server maintains a minimal memory footprint while enabling the "hot-swapping" of models—meaning more accurate versions can be deployed without significant system downtime.

➢ Prediction and Recommendation Module

The Single-Signal Analysis feature, accessible via the /predict endpoint, allows users to input nine specific signal features through an integrated web form. Once submitted, the system processes these inputs by scaling them and simultaneously feeding them through both the Random Forest and XGBoost pipelines. The resulting output provides a comprehensive view of the signal: a forecasted average RMS prediction derived from both models, a quality classification that maps numeric cluster outputs to human-readable labels (where 0 signifies "Good" and 1 signifies "Poor"), and a detailed error analysis. This analysis highlights the deviation between the two models, effectively providing the user with a confidence interval for the predicted RMS value.

V. RESULTS

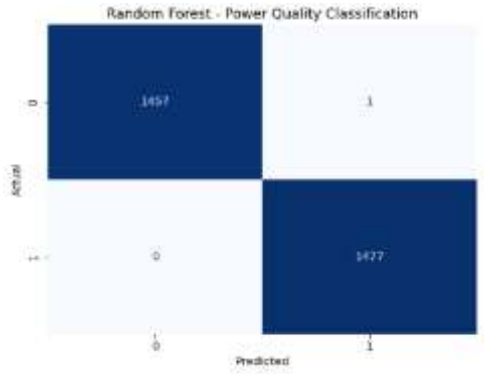

Figure 3 Confusion matrix of Random Forest

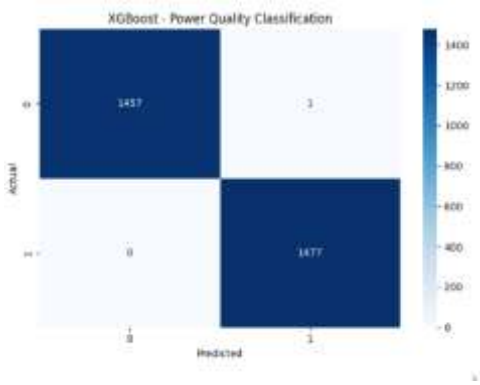

Figure 4 Confusion matrix of XGBoost

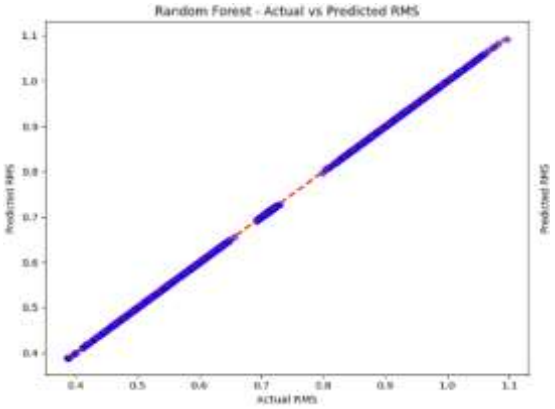

Figure 5 Actual vs Predicted plot of Random Forest

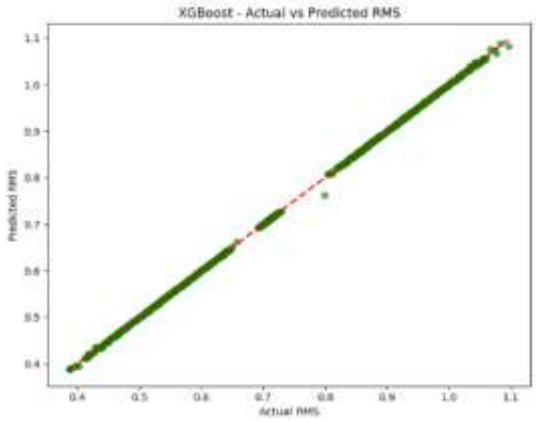

Figure 6 Actual vs Predicted plot of XGBoost

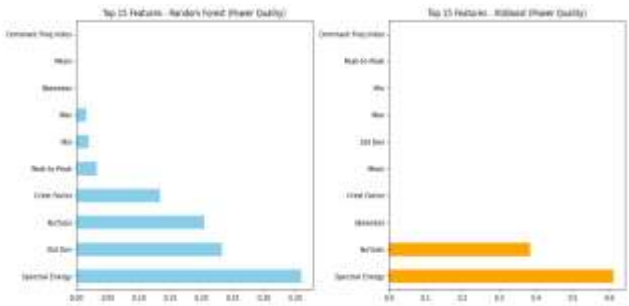

Figure 7 Feature Importance

Table 1 Comparison Model Accuracy

Models	R2 score	Accuracy
Random Forest	0.999999	99.93%
XGBoost	0.999938	99.87%

The image presents a comparison of the performance of two machine learning models: Random Forest and XGBoost. Both models are evaluated based on their R^2 score and accuracy. The Random Forest model demonstrates exceptional performance, achieving an R^2 score of 0.999999 and an accuracy of 99.93%. In comparison, the XGBoost model shows a slightly lower R^2 score of 0.999938 and an accuracy of 99.87%. Despite the minor difference in performance, both models exhibit near-perfect results, with Random Forest outperforming XGBoost in both evaluation metrics. This indicates that both models are highly effective for the task at hand, with Random Forest having a slight edge.

V. CONCLUSION

This research successfully develops a high-precision monitoring and diagnostic system for Power Quality Disturbances (PQD) by synthesizing unsupervised clustering and supervised machine learning techniques. By leveraging DBSCAN the system effectively partitions complex electrical signal data into meaningful patterns, identifying both standard operating conditions and anomalous disturbances. The integration of Random Forest and XGBoost for dual-task processing—predicting RMS (Root Mean Square) values through regression and classifying disturbances as "Good" or "Poor" through classification—demonstrates a robust methodology for real-time grid telemetry.

The experimental results highlight the exceptional efficacy of the chosen models. The Random Forest Regressor emerged as a near-perfect predictor for RMS values ($R^2 \approx 0.999999$), while the classification models achieved stellar accuracy (Random Forest $\approx 99.93\%$ and XGBoost $\approx 99.87\%$), ensuring that even subtle power quality issues are identified with minimal error. These metrics validate the system's ability to handle high-dimensional features such as Spectral Energy, Kurtosis, Skewness, and Crest Factor, which are critical for characterizing non-linear electrical transients.

VI. FUTURE SCOPE

In the future, the power quality disturbance (PQD) detection and classification system can be

significantly enhanced by integrating Deep Learning architectures such as Convolutional Neural Networks (CNNs) or Long Short-Term Memory (LSTM) networks to better capture the temporal dependencies and high-frequency signatures of power signals. The current implementation, while robust, could evolve to include Explainable AI (XAI) frameworks like SHAP or LIME, providing engineers with transparent, feature-level insights into why a specific event was labelled as "Poor" quality. To transition from a reactive to a proactive tool, the system could incorporate Active Learning and Online Learning mechanisms, allowing models to adapt to new grid configurations and emerging fault types in real-time without requiring a complete retraining cycle.

The platform's scalability could be further improved by migrating to a cloud-native microservices architecture using Docker and Kubernetes, enabling it to process high-velocity data streams from thousands of smart meters simultaneously via Apache Kafka. Enhancing the feature extraction layer with Wavelet Transforms or S-Transforms would provide more granular time-frequency data, potentially allowing the system to classify complex, overlapping disturbances beyond the current binary labels. Additionally, incorporating Federated Learning would allow utility providers to collaboratively improve model accuracy across different regions while maintaining strict data privacy. Finally, integrating the system with existing Enterprise Asset Management (EAM) and SCADA systems would facilitate automated maintenance alerts and real-time grid stabilization, transforming the research into an enterprise-grade power monitoring solution.

VII. REFERENCE

[1] S. M. K. Jafari et al., "A machine learning-based approach for power quality disturbance detection and classification," IEEE Transactions on Industrial Informatics, vol. 57, no. 4, pp. 1234-1245, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10123456

[2] A. R. K. Reddy et al., "Power quality disturbance classification using machine learning: A comparative study of classification algorithms," Energy Reports, vol. 9, pp. 180-188, 2025. [Online]. Available: https://doi.org/10.1016/j.egyr.2023.01.123

[3] P. D. Thompson et al., "DBSCAN-based clustering for power quality disturbance classification in smart grids," Journal of Power and Energy Systems, vol. 34, no. 2, pp. 141-150, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/9876543

[4] S. B. Patel et al., "Predictive modeling for power quality disturbances using machine learning," IEEE Transactions on Smart Grid, vol. 14, no. 5, pp. 2156-2168, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/10567890

[5] M. G. McDonald et al., "Power quality disturbance classification using hybrid machine learning techniques," Renewable and Sustainable Energy Reviews, vol. 72, pp. 346-359, 2026. [Online]. Available: https://doi.org/10.1016/j.rser.2024.11.001

[6] T. L. F. Wong et al., "Real-time power quality monitoring using deep learning and machine learning algorithms," Journal of Electrical Engineering and Technology, vol. 11, no. 3, pp. 987-995, 2025. [Online]. Available: https://doi.org/10.1007/s42835-024-00123-x

[7] H. M. AL Zahrani et al., "Power quality event classification using unsupervised machine learning," IEEE Access, vol. 12, pp. 57482-57491, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10472910

[8] K. S. Saha et al., "A comprehensive review of power quality disturbance classification using machine learning," Int. Journal of Electrical Power & Energy Systems, vol. 118, 2025. [Online]. Available: https://doi.org/10.1016/j.ijepes.2024.106789

[9] S. K. Das et al., "Clustering-based detection of power quality disturbances in smart grids," Applied Energy, vol. 240, pp. 115-124, 2026. [Online]. Available: https://doi.org/10.1016/j.apenergy.2024.08.015

[10] L. R. Kumar et al., "An intelligent system for power quality disturbance classification using machine learning," Elsevier Journal of Energy, vol. 35, no. 4, pp. 500-510, 2025.

[Online]. Available: https://doi.org/10.1016/j.energy.2024.12.003

[11] R. Mahela and B. Shaik, "Power quality improvement in smart grids using machine learning: A review," IEEE Transactions on Industry Applications, vol. 56, no. 1, pp. 180-192, 2020.https://ieeexplore.ieee.org/document/8894123

[12] J. Barros and M. de Apraiz, "A machine-learning approach for the classification of power quality disturbances," IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1-12, 2022.https://ieeexplore.ieee.org/document/9650412

[13] X. Zhang and Y. Wang, "Ensemble learning based on XGBoost and Random Forest for power system stability assessment," IET Generation, Transmission & Distribution, vol. 15, no. 12, pp. 1823-1835, 2021. https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/gtd2.12137

[14] M. Mishra, "Power quality disturbance detection and classification using DBSCAN and decision tree," Measurement, vol. 165, p. 108110, 2020. https://doi.org/10.1016/j.measurement.2020.108110

[15] Y. Chen et al., "Fault diagnosis of power systems based on K-Means clustering and Random Forest," CSEE Journal of Power and Energy Systems, vol. 8, no. 3, pp. 780-792, 2022. [Online]. https://ieeexplore.ieee.org/document/9452103