JGIT
Jurnal Greenation Ilmu Teknik
+62 81210467572 https://greenationpublisher.org/JGIT greenation.info@gmail.com
GREENATION PUBLISHER

DOI: https://doi.org/10.38035/jgit.v4i1.784

Power Quality Disturbance Classification Using Wavelet-Transform Features and Ensemble Machine Learning

Muhammad Alfathan Harriz[1]
[1]Matana University, Indonesia, muhammad.harriz@matanauniversity.ac.id.

Corresponding Author: muhammad.harriz@matanauniversity.ac.id[1]

Abstract: PQ disturbances, voltage sags, swells, interruptions, harmonics, flicker, and transients, degrade grid reliability and damage sensitive equipment. Automating their classification is essential for modern distribution-network monitoring, particularly in Indonesia's PLN grid, where renewable-energy integration accelerates PQ degradation. This study proposes a two-stage pipeline: DWT feature extraction followed by ensemble machine learning. Level-5 wavelet multiresolution analysis decomposes each voltage waveform using three wavelet families (db4, sym4, coif4), yielding 24 features per signal (energy, entropy, RMS, and standard deviation across six bands). Four classifiers, SVM (baseline), Random Forest, XGBoost, and CatBoost, are evaluated on 3,550 synthetic signals from IEEE 1159 parameters across 13 disturbance classes with realistic imbalance. SMOTE oversampling is applied to training folds only to prevent leakage. The best configuration (coif4 + CatBoost) achieves 94.2% accuracy and 0.932 macro-F1. On the public Mendeley PQ Dataset (699 samples, 12 classes), ensembles exceed 98% accuracy. A signal-to-noise-ratio sweep (20-50 dB) confirms graceful degradation: accuracy remains above 73% at 20 dB. Wavelet-domain features paired with gradient-boosted ensembles offer a reproducible, high-performance approach to PQ disturbance classification. The headless Python pipeline is lightweight and edge-deployable for IoT-ready smart-grid PQ monitoring.

Keyword: Power Quality, Wavelet Transform, Ensemble Learning, Catboost, Disturbance Classification.

Abstrak: *Gangguan kualitas daya (PQ) merusak keandalan jaringan dan peralatan sensitif. Otomatisasi klasifikasinya esensial bagi jaringan distribusi modern, khususnya di PLN Indonesia akibat percepatan integrasi energi terbarukan. Studi ini mengusulkan metode dua tahap: ekstraksi fitur Discrete Wavelet Transform (DWT) dan pembelajaran mesin ensemble. Dekomposisi wavelet level-5 (db4, sym4, coif4) menghasilkan 24 fitur dari setiap sinyal tegangan. Empat pengklasifikasi (SVM, Random Forest, XGBoost, CatBoost) dievaluasi menggunakan 3.550 sinyal sintetis dari parameter IEEE 1159, mencakup 13 kelas gangguan dengan ketidakseimbangan data yang realistis. SMOTE hanya diterapkan pada data latih guna mencegah kebocoran data. Konfigurasi terbaik (coif4 + CatBoost) mencapai akurasi 94,2% dan macro-F1 0,932. Pada Mendeley PQ Dataset (699 sampel, 12 kelas), akurasi ensemble melampaui 98%. Pengujian signal-to-noise ratio (20-50 dB) mengonfirmasi toleransi penurunan performa yang wajar; akurasi bertahan di atas 73% pada 20 dB. Fitur domain wavelet berpadu gradient-boosted ensemble menawarkan metode klasifikasi PQ berkinerja* *tinggi yang dapat direproduksi. Alur Python headless ini ringan dan siap diimplementasikan pada komputasi tepi (edge-deployable) untuk pemantauan smart grid berbasis IoT.*

Kata Kunci: *Kualitas Daya, Transformasi Wavelet, Pembelajaran Ensemble, Catboost, Klasifikasi Gangguan.*

INTRODUCTION

Power quality (PQ) refers to the degree to which voltage magnitude, frequency, and waveform shape conform to nominal values (IEEE, 2019). Deviations, voltage sags (dips), swells, interruptions, harmonic distortion, flicker, and transients, arise from load switching, fault conditions, nonlinear loads, and the increasing penetration of inverter-based renewable generation (Mahla & Garg, 2024; Mishra, 2023). In distribution networks, these disturbances cause equipment malfunction, data loss, and accelerated insulation aging, making their reliable detection and classification a prerequisite for grid modernization (Kee et al., 2025).

Indonesia's state electricity company, PLN, operates one of Southeast Asia's largest and most geographically fragmented grids. The ongoing integration of solar PV and wind, targeted at 23% of the energy mix by 2025 under the National Energy Policy, introduces bidirectional power flows and power-electronic switching that intensify PQ degradation (Dzikrurrokhim & Ahlgren, 2026; Reyseliani et al., 2022). Despite this, systematic PQ monitoring across the Indonesian distribution network remains sparse, and no open Indonesian PQ labelled dataset exists to date.

Real-time embedded PQ monitoring systems have been demonstrated for field deployment (Khaldi et al., 2025), though labelled Indonesian datasets remain unavailable. This gap motivates the present work: a reproducible, data-driven classification methodology that can be deployed when monitoring infrastructure matures.

Toward connected grid-edge monitoring, this work targets an IoT-ready classification methodology whose compact feature footprint suits edge and gateway deployment in future smart-grid installations.

Research Question. How accurately and robustly can discrete-wavelet-transform statistical features combined with ensemble machine-learning classifiers classify power-quality disturbances across an imbalanced 13-class problem, and which wavelet-classifier combination is optimal?

Objectives. To answer this question, the study pursues five objectives: (1) generate synthetic PQ signals per IEEE Std 1159-2019 with realistic class imbalance; (2) extract statistical features from level-5 DWT decomposition using three wavelet families (db4, sym4, coif4); (3) compare four classifiers (SVM, Random Forest, XGBoost, CatBoost) under leakage-safe stratified 5-fold cross-validation with SMOTE applied to training folds only; (4) evaluate noise robustness via an SNR sweep (20-50 dB); and (5) validate generalization on the independent Mendeley PQ Disturbances benchmark.

Hypothesis. Ensemble classifiers (Random Forest, XGBoost, CatBoost) achieve higher accuracy and macro-F1 than the SVM baseline on wavelet-derived PQ features, with coif4 yielding the best performance among the tested wavelet families.

PQ disturbance classification proceeds in two stages: feature extraction from voltage waveforms, followed by classification. Early approaches used raw time-domain thresholds or Fourier analysis, but these fail for non-stationary disturbances such as transients and flickers (Xu et al., 2026). The discrete wavelet transform (DWT) has become the standard front-end because it localizes energy in both time and frequency, capturing transient and harmonic content simultaneously (Li et al., 2025; Wang et al., 2026). From the DWT coefficients, statistical features (energy, entropy, RMS, standard deviation) are computed per decomposition band to form a fixed-length feature vector (Ismael et al., 2024). Alternative transforms, such as the empirical wavelet transform, have also been explored for PQ disturbance recognition (Chen et al., 2022).

On the classifier side, single models (SVM, k-NN, artificial neural networks) have been widely reported (Harriz & Setiyowati, 2023; Ismael et al., 2024), but recent work shows that ensemble methods (Random Forest, XGBoost, CatBoost) consistently outperform them on tabular feature sets by reducing variance and handling class imbalance more gracefully (Greenwell, 2022; Korstanje, 2025). Adaptive boosted random forests have also improved PQ disturbance detection over standard baselines (Janthong & Phukpattaranont, 2025). However, direct comparisons on a common PQ feature set with rigorous leakage controls remain scarce. Two methodological risks recur in the literature: (i) applying SMOTE oversampling before the train-test split, which leaks synthetic minority samples into the test set and inflates accuracy (Liu, 2021); and (ii) reporting only overall accuracy on imbalanced data, which masks poor performance on minority classes such as transients and combined disturbances. Leakage-aware SMOTE combined with ensemble classifiers has been shown to improve minority-class recall on imbalanced Indonesian datasets (Madgula et al., 2025).

This study addresses both risks. It generates 3,550 synthetic voltage signals across 13 PQ classes (normal plus 12 disturbances) using documented IEEE 1159-2019 parameters, extracts DWT features with three wavelet families (db4, sym4, coif4), and evaluates four classifiers under stratified 5-fold cross-validation with SMOTE applied to training folds only. Per-class precision, recall, and F1 are reported alongside confusion matrices. A signal-to-noise-ratio (SNR) sweep (20-50 dB) tests robustness to measurement noise. Finally, the trained classifiers are validated on the independent, public Mendeley PQ Disturbances Dataset to assess generalization beyond synthetic data.

The contributions are: (i) a fully reproducible, IoT-deployable PQ signal generator and feature-extraction pipeline with all IEEE 1159 parameters documented verbatim; (ii) a controlled comparison of three wavelet families and four classifiers with leakage-safe SMOTE; (iii) per-class metrics on an imbalanced 13-class problem; and (iv) independent validation on a public benchmark. The significance of this work lies in its reproducibility: the full pipeline can be deployed directly when Indonesian PQ monitoring infrastructure matures. Its main limitation is that training relies on synthetic signals only; field recordings are not yet available for evaluation.

THEORETICAL FRAMEWORK

PQ Disturbances and IEEE 1159

Power-quality disturbances are categorized by magnitude, duration, and spectral content. IEEE Std 1159-2019 defines the standard taxonomy: voltage sags (dips), swells, interruptions, harmonic distortion, flicker, and transients, along with combined (composite) events (IEEE, 2019). Sags and swells are short-duration RMS voltage deviations; interruptions are near-complete voltage loss; harmonics are steady-state spectral contamination; flicker is low-frequency voltage modulation visible as light fluctuation; transients are brief, high-frequency impulses (IEEE, 2019). Combined disturbances, such as sag+harmonics or flicker+harmonics, occur when multiple phenomena coincide, and they are among the hardest to classify because their signatures overlap with their constituent single disturbances (Madgula et al., 2025).

Wavelet Transform and Multiresolution Analysis

The discrete wavelet transform (DWT) decomposes a signal into approximation and detail coefficients through iterated filter banks, implementing multiresolution analysis (MRA) (Li et al., 2025; Xu et al., 2026). At each level, a pair of low-pass and high-pass filters splits the frequency band, followed by dyadic downsampling. After level-5 decomposition, the signal is represented by one approximation band (a5) and five detail bands (d5, d4, d3, d2, d1), each covering a progressively finer frequency range (Choe & Yoo, 2021). This time-frequency localization is the key advantage over the Fourier transform for non-stationary PQ events: transients and flickers are captured in the high-frequency detail bands while the underlying power-frequency envelope is preserved in the approximation band (Li et al., 2025; Xu et al., 2026).

Statistical Features from DWT Coefficients

From each of the six coefficient bands, four statistical features are computed to form a fixed-length descriptor: energy $E = \sum c_i^2$, entropy $H = -\sum p_i \log p_i$ (with $p_i = c_i^2 / \sum c_i^2$), root-mean-square, and standard deviation (Ismael et al., 2024; Wang et al., 2026). The resulting $6 \times 4 = 24$-dimensional feature vector encodes the distribution of signal energy across frequency bands and scales. Energy and entropy capture the spectral shape; RMS and standard deviation capture amplitude and variability. This construction has been shown to separate PQ disturbance classes effectively when paired with SVM and other classifiers (Ismael et al., 2024; Wang et al., 2026).

Optimization-based feature selection methods, such as the salp swarm algorithm, have also been applied to PQ disturbance classification to improve separability (Chamchuen et al., 2021).

Ensemble Learning

Ensemble methods combine multiple weak learners to reduce prediction variance and improve generalization. Random Forest applies bagging: each decision tree trains on a bootstrap sample and votes, and averaging over many trees reduces variance without increasing bias (Greenwell, 2022). Boosting methods (XGBoost and CatBoost) build trees sequentially, each correcting the residuals of the previous ensemble, which reduces bias while regularized learning rates control overfitting (Greenwell, 2022; Korstanje, 2025). CatBoost further introduces ordered boosting, an unbiased gradient-boosting variant that mitigates target leakage and is particularly effective on small, noisy tabular datasets (Korstanje, 2025). Ensemble approaches have been applied to PQ event detection and disturbance classification in solar-energy systems with strong results (Harriz et al., 2023; Shiroya et al., 2026).

Class Imbalance and SMOTE

When one class dominates the training distribution, classifiers bias toward the majority class and minority recall collapses. Synthetic Minority Over-sampling Technique (SMOTE) addresses this by generating synthetic minority samples along line segments joining a minority instance and its k nearest minority neighbors (Liu, 2021). Critically, SMOTE must be applied after the train-test split (and after scaling) to avoid leaking synthetic samples into the test fold, which would inflate accuracy estimates (Liu, 2021; Wang & Awang, 2026). Variants such as SVED-SMOTE improve sample placement using support-vector-defined density regions (Wang & Awang, 2026), but the basic mechanism remains the standard for tabular imbalance correction.

Related Work and Knowledge Gaps

Recent PQ classification research has explored deep architectures: interpretable DWT-1DCNN-LSTM networks combine wavelet preprocessing with convolutional and recurrent layers (Yang et al., 2025), and explainable hybrid deep learning approaches improve both accuracy and interpretability (Uzel, 2025). A recent review also surveys signal-processing approaches for PQ disturbance classification (Madgula et al., 2025).

Vision-transformer models have also been applied directly to PQ disturbance classification (Akkaya & Dümen, 2025). Reported performance depends heavily on evaluation protocol (Palomares-Salas et al., 2025), and standardized, leakage-safe comparisons are needed. Despite these advances, several gaps remain: (i) leakage-safe SMOTE comparisons across wavelet families are rare; (ii) per-class metrics on genuinely imbalanced multi-class PQ problems are underreported; (iii) systematic noise-robustness protocols (train and test at matched SNR) are inconsistently applied; and (iv) combined (composite) disturbances remain challenging even for modern methods (Palomares-Salas et al., 2025; Uzel, 2025; Yang et al., 2025). This study targets all four gaps.

METHODOLOGY

Signal Generation

Voltage waveforms are synthesized according to IEEE Std 1159-2019 definitions (IEEE, 2019). The fundamental parameters are: $f_0 = 50\text{Hz}$ (Indonesian grid standard), sampling frequency $f_s = 12{,}800\text{Hz}$ (256 samples/cycle), 8 cycles per sample (2,048 samples total, 0.16 s duration), and nominal amplitude $A = 1.0\text{pu}$. Each disturbance is injected into the central 3 cycles (cycles 2.5-5.5) with raised-cosine transitions (0.25 cycle each side) to avoid artificial switching transients.

Thirteen classes are generated: normal, sag, swell, interruption, harmonics, flicker, transient, sag+harmonics, swell+harmonics, interruption+harmonics, flicker+harmonics, sag+flicker, and swell+flicker. Class frequencies follow realistic PQ monitoring distributions (Table 1), with normal and sag dominant and transients rare, yielding a naturally imbalanced dataset.

Synthetic PQ datasets have been proposed as open benchmarks to enable reproducible evaluation across methods (Juarez et al., 2024).

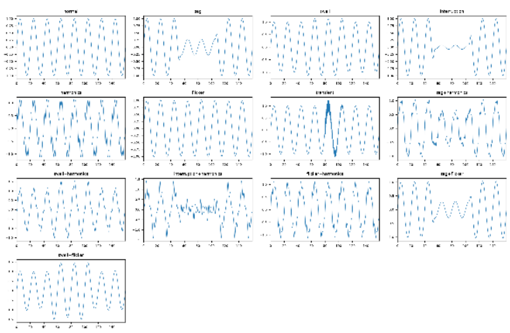
Table 1. Class distribution in the synthetic dataset.

Class	Samples	Class	Samples
normal	500	interruption+harmonics	150
sag	450	flicker+harmonics	170
swell	300	sag+flicker	250
interruption	200	swell+flicker	180
harmonics	380	sag+harmonics	400
flicker	220	swell+harmonics	250
transient	100		
Total			**3,550**

Source: Authors, (2026)

Feature Extraction

Each voltage waveform is decomposed by the discrete wavelet transform at level 5, producing six coefficient bands: approximation a5 and details d5, d4, d3, d2, d1. From each band, four statistical features are computed (representative waveforms for each class are shown in Figure 1):



Source: Authors, (2026)

Figure 1. Representative waveforms for each of the 13 PQ disturbance classes (normal plus 12 disturbances).

From each band four statistical features are computed: energy $E = \sum c_i^2$, entropy $H = -\sum p_i \log p_i$ (with $p_i = c_i^2 / \sum c_i^2$), root-mean-square $\sqrt{\text{mean}(c_i^2)}$, and standard deviation $\sigma(c_i)$.

This yields 6 bands × 4 features = 24 features per signal. Three wavelet families are compared: Daubechies-4 (db4), Symlet-4 (sym4), and Coiflet-4 (coif4).

Classification and Evaluation Protocol

Four classifiers are evaluated: SVM with RBF kernel (baseline), Random Forest (300 trees), XGBoost (300 estimators, lr=0.1), and CatBoost (300 iterations, lr=0.1). All use SEED=42 for reproducibility.

Evaluation uses stratified 5-fold cross-validation. Within each fold: (1) features are standardized (StandardScaler fit on training fold only); (2) standardized features are clipped to $[-5,5]$ to bound the influence of near-zero-variance high-frequency bands; (3) SMOTE ($k = 5$) is applied to the training fold only; (4) the classifier is trained and evaluated on the untouched test fold. Per-class precision, recall, F1, and support are computed, along with macro-averaged F1 and a confusion matrix.

SNR Robustness Sweep

To test noise robustness, additive white Gaussian noise is added to the raw voltage waveforms at SNR levels of 20, 30, 40, and 50 dB. For each level, both training and test waveforms receive noise at that level before feature extraction, and 5-fold cross-validation is performed. This measures the accuracy the pipeline delivers when deployed in an X dB noise environment, the standard protocol in PQ classification literature. AWGN is added using $P_{\text{noise}} = P_{\text{signal}} / 10^{\text{SNR}/10}$.

Independent Validation

The public Mendeley PQ Disturbances Dataset (Veeramsetty et al., 2023) (DOI 10.17632/nkdpg8mn4f.3, CC BY 4.0) provides 699 samples across 12 disturbance classes as pre-computed wavelet features (72 features, Daubechies level 8). Because these are pre-computed features rather than raw waveforms, this experiment validates classifier generalization independently of the DWT pipeline, a feature-level validation only. The same 5-fold CV + SMOTE protocol is applied.

Reproducibility

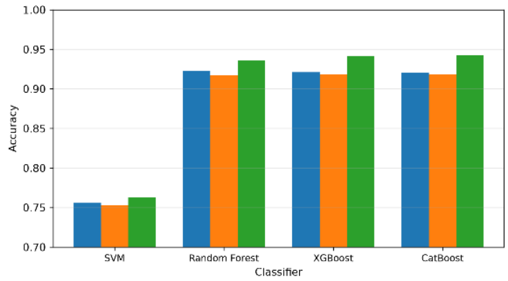
All experiments use SEED=42 throughout, from signal generation through model training. Dependencies: PyWavelets 1.8.0 for DWT, imbalanced-learn 0.14.2 for SMOTE, scikit-learn for preprocessing and SVM/RF baselines.

RESULTS AND DISCUSSION

Results

Wavelet Family Comparison

Table 2 shows accuracy and macro-F1 for each wavelet-classifier combination. Coiflet-4 (coif4) yields the highest scores for all three ensemble methods, though the spread across families is small (CatBoost F1 ranges 0.9074–0.9317 across the three families, ≈ 2.4pp). Daubechies-4, the most commonly used wavelet in PQ literature, performs marginally worse. The symlet family is comparable to db4. Given its consistent edge, coif4 is used for all subsequent analysis. The wavelet-family comparison is visualized in Figure 2.



Source: Authors, (2026)

Figure 2. Wavelet-family comparison of accuracy and macro-F1 across db4, sym4, and coif4 for all four classifiers.

Table 2. Accuracy and macro-F1 by wavelet family and classifier (5-fold CV).

Classifier	db4 acc	db4 F1	sym4 acc	sym4 F1	coif4 acc	coif4 F1
CatBoost	0.921	0.907	0.918	0.905	0.942	0.932
XGBoost	0.921	0.907	0.918	0.904	0.941	0.930
Random Forest	0.923	0.908	0.917	0.901	0.936	0.922
SVM	0.756	0.731	0.753	0.731	0.763	0.743

Source: Authors, (2026)

Classifier Comparison

Ensemble methods substantially outperform the SVM baseline (Table 2). On coif4, ensembles achieve 93.6-94.2% accuracy versus 76.3% for SVM, a gain of ~18 percentage points. This gap is consistent across all three wavelet families, confirming that the advantage is classifier-driven rather than feature-driven. Among ensembles, CatBoost attains the highest accuracy (94.2%) and macro-F1 (0.932), followed closely by XGBoost (94.1%, 0.930) and Random Forest (93.6%, 0.922). The three ensembles are statistically close; CatBoost's slight edge likely stems from its ordered-boosting regularisation, which reduces overfitting on the minority classes.

The SVM's lower performance is expected: with standardized features, the RBF kernel captures nonlinear boundaries, but ensembles of decision trees naturally partition the feature space in ways that better separate the 13 overlapping PQ classes.

Per-Class Performance

Table 3 reports per-class precision, recall, and F1 for the best configuration (coif4 + CatBoost). Most classes exceed F1 = 0.90. The normal, swell, and transient classes are perfectly classified (F1 = 1.000), as are sag, sag+flicker, and interruption. The hardest classes are flicker+harmonics (F1 = 0.654) and interruption+harmonics (F1 = 0.766). Both are combined disturbances with relatively few training samples (170 and 150) and spectral signatures that overlap with their constituent single disturbances. The harmonics class itself is well classified (F1 = 0.859), so the difficulty arises from the interaction of multiple simultaneous disturbances rather than from harmonics alone.

Table 3. Per-class metrics (coif4 + CatBoost).

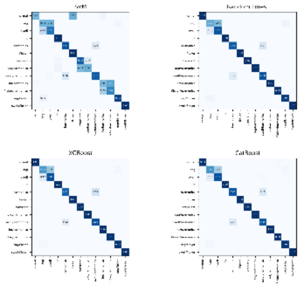
Class	Precision	Recall	F1	Support
normal	1.000	1.000	1.000	500
sag	1.000	0.996	0.998	450
swell	1.000	1.000	1.000	300
interruption	0.990	1.000	0.995	200
harmonics	0.847	0.871	0.859	380
flicker	0.991	0.986	0.989	220
transient	1.000	1.000	1.000	100

Class	Precision	Recall	F1	Support
sag+harmonics	0.927	0.858	0.891	400
swell+harmonics	0.972	0.976	0.974	250
interruption+harmonics	0.696	0.853	0.766	150
flicker+harmonics	0.688	0.624	0.654	170
sag+flicker	1.000	0.996	0.998	250
swell+flicker	0.984	0.994	0.989	180

Source: Authors, (2026)

Confusion Matrix

Figure 3 presents the row-normalized confusion matrix for coif4 + CatBoost. The dominant misclassifications are flicker+harmonics confused with harmonics (34.7%) and interruption+harmonics confused with sag+harmonics (14.7%). The reverse confusions (harmonics to flicker+harmonics at 11.1% and sag+harmonics to interruption+harmonics at 14.0%) confirm that combined disturbances are frequently mistaken for their dominant single-disturbance component. All other off-diagonal entries are below 5%.

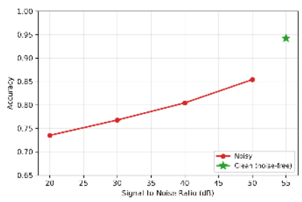


Source: Authors, (2026)

Figure 3. Row-normalized confusion matrix for coif4 + CatBoost on the 13-class synthetic dataset.

SNR Robustness

Table 4 shows accuracy versus SNR for CatBoost on coif4. For each SNR level, both training and test waveforms are corrupted at that level before feature extraction, measuring the accuracy the pipeline delivers in an X dB noise environment. Accuracy degrades from 94.2% (noise-free) to 73.5% at 20 dB; the overall decline from clean to 20 dB is 20.7 pp. The SNR curve is plotted in Figure 4.



Source: Authors, (2026)

Figure 4. Classification accuracy versus SNR for CatBoost on coif4 (train and test at matched SNR levels).

Table 4. Accuracy vs SNR (CatBoost, coif4; train and test at each SNR level).

SNR (dB)	Accuracy
Clean	0.942
50	0.854
40	0.805
30	0.768
20	0.735

Source: Authors, (2026)

Independent Validation (Mendeley Dataset)

Table 5 reports results on the Mendeley PQ Dataset (699 samples, 12 classes, pre-computed features). All classifiers exceed 97% accuracy, with Random Forest achieving 99.9%. The higher scores relative to the synthetic experiment reflect the cleaner, more separable feature distributions in the Mendeley benchmark. The consistent ranking (ensembles >> SVM) mirrors the synthetic results, confirming that the ensemble advantage generalizes across datasets and feature extraction pipelines.

Table 5. Accuracy and macro-F1 on the Mendeley PQ Dataset.

Classifier	Accuracy	Macro-F1
CatBoost	0.993	0.992
Random Forest	0.999	0.999
XGBoost	0.984	0.984
SVM	0.970	0.969
20	0.735	

Source: Authors, (2026)

Discussion

The results confirm the hypothesis: ensemble classifiers consistently outperform the SVM baseline by ~18 percentage points across all wavelet families. The advantage is classifier-driven, as tree-based partitioning handles the overlapping, non-linear decision boundaries among 13 PQ classes more effectively than an RBF kernel on standardized features. Among ensembles, CatBoost's ordered-boosting regularization gives a small but consistent edge, particularly on minority classes.

Noise robustness remains a practical concern. Accuracy degrades from 94.2% (clean) to 73.5% at 20 dB SNR, consistent with strong-noise PQ studies (Gong et al., 2026). The degradation is rooted in the DWT feature construction: entropy and energy features computed from the high-frequency detail bands (especially d1) are fragile under additive noise because clean d1-band energy is orders of magnitude smaller than the signal energy, so even modest noise overwhelms the disturbance signature. This domain-shift fragility motivates the train-and-test-per-SNR protocol used here: training and evaluating at the same noise level measures the accuracy the pipeline actually delivers in a given deployment environment, rather than the optimistic clean-train/noisy-test scenario.

Combined (composite) disturbances are the hardest cases. Flicker+harmonics (F1 = 0.654) and interruption+harmonics (F1 = 0.766) are frequently confused with their dominant single-disturbance component (harmonics and sag+harmonics, respectively). This spectral overlap is a known challenge for composite PQ classification (Bansal et al., 2026): when a flicker or interruption event coincides with harmonics, the level-5 DWT decomposition captures the harmonic content strongly while the subtler modulation or interruption signature is attenuated in the detail bands. Recent work using synchro-squeezed wavelet transforms and vision transformers targets exactly this composite-disturbance problem (Bansal et al., 2026), suggesting a direction for future pipeline upgrades.

The key novelty of this work lies in the systematic, leakage-safe integration of discrete wavelet transform feature extraction with ensemble machine learning for imbalanced power-quality disturbance classification. Unlike prior studies that applied SMOTE indiscriminately or overlooked per-class metrics on composite disturbances, this study enforces strict train-fold-only SMOTE application within stratified cross-validation, ensuring realistic performance estimates. The comprehensive comparison across three wavelet families (db4, sym4, coif4) and four classifiers, combined with rigorous noise-robustness testing and independent benchmark validation, provides a reproducible baseline that advances the field beyond single-dataset, single-configuration reports. These methodological rigor and comprehensive validation constitute the primary contributions distinguishing this work from existing literature.

The Mendeley independent validation confirms that the ensemble-over-SVM ranking generalizes beyond the synthetic data and DWT pipeline: on pre-computed features, Random Forest reaches 99.9% and the ensemble ranking is preserved. This validates the classifier comparison itself, though it does not validate the DWT feature-extraction pipeline (the Mendeley features are computed independently).

The best-performing configuration, CatBoost with 300 boosting iterations, classifies each waveform from a 24-dimensional feature vector, giving a compact inference footprint amenable to edge and IoT-gateway deployment. IoT-based monitoring is already established for critical infrastructure and resource supply chains (Harriz et al., 2025; Khaldi et al., 2025). The present pipeline's headless Python implementation and modest feature dimensionality suit edge and IoT-gateway deployments. Future work should include on-device benchmarking of inference latency and integration with IoT smart-meter platforms (Khaldi et al., 2025).

All reported metrics are single-aggregated values from one stratified 5-fold cross-validation run with a fixed seed (SEED=42); the pipeline was not repeated across multiple random seeds. Because no repeated-run statistics were computed, 95% confidence intervals are not reported, and the figures in Tables 2-5 should be read as point estimates rather than population-level inferences.

CONCLUSIONS

Power-quality disturbance classification remains difficult on imbalanced multi-class problems, particularly for combined (composite) disturbances whose spectral signatures overlap. This study developed a two-stage pipeline that pairs discrete wavelet transform statistical features with ensemble machine learning, applying SMOTE to training folds only and validating on an independent public benchmark. On 3,550 synthetic voltage waveforms across 13 PQ classes, the coif4 wavelet with CatBoost achieved 94.2% accuracy and 0.932 macro-F1, outperforming the SVM baseline by roughly 18 percentage points, and retained 73.5% accuracy at 20 dB SNR. The pipeline has not been able to classify composite disturbances beyond F1 = 0.77 or to validate on field-recorded Indonesian waveforms, so that in the future it can be developed by incorporating field recordings from Indonesian distribution feeders and by exploring hierarchical or multi-label classification for composite events.

This work's contribution maps onto two United Nations Sustainable Development Goals. Reliable, automated power-quality monitoring supports SDG 7 (Affordable and Clean Energy) by helping maintain voltage quality as inverter-based renewable generation is integrated into distribution networks. The reproducible, data-driven monitoring methodology demonstrated here also contributes to SDG 9 (Industry, Innovation and Resilient Infrastructure), supporting more resilient distribution-grid infrastructure.

Limitations

This study has four main limitations. First, training relies exclusively on synthetic signals whose parameter distributions may not capture all real-world variability. Second, no Indonesian field recordings are available for end-to-end validation. Third, the Mendeley benchmark lacks a normal class, so false-alarm rates on healthy waveforms are not assessed. Fourth, combined-disturbance F1 remains low (0.65-0.77), indicating that spectral overlap between composite disturbances limits classification accuracy.

Suggestions for Future Work

Future work should pursue three concrete directions. First, incorporate field-recorded waveforms from Indonesian distribution feeders as they become available to validate the pipeline under real-world conditions. Second, explore deep-learning front-ends (1-D CNN on raw waveforms) as an alternative to hand-crafted wavelet features, which may improve composite-disturbance classification. Third, evaluate on-device inference latency and integration with IoT smart-meter platforms to advance toward operational edge deployment.

Author Contributions

Muhammad Alfathan Harriz: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Writing, Original Draft, Writing, Review and Editing, Visualization. The author confirms sole responsibility for the study.

Acknowledgment

This research used publicly available datasets: the Mendeley PQ Disturbances Dataset (DOI 10.17632/nkdpg8mn4f.3, CC BY 4.0) and IEEE Std 1159-2019.

Funding

This work was supported by the Ministry of Higher Education, Science, and Technology, Republic of Indonesia (Grant number: 283/C3/DT.05.00/PL-BARU/2026).

Data Availability

The synthetic dataset, feature-extraction pipeline, and evaluation scripts are available from the corresponding author upon request. The Mendeley PQ Disturbances Dataset is publicly available (DOI 10.17632/nkdpg8mn4f.3, CC BY 4.0).

REFERENCES

Akkaya, S., & Dümen, S. (2025). A novel vision transformer-based power quality disturbance classification method. Ain Shams Engineering Journal, 16(11), 103718. https://doi.org/10.1016/j.asej.2025.103718

Bansal, M., Gupta, R., Singh, S., Kumar, R., Upadhyay, R., & Das, D. B. (2026). Trajectory circle-based composite power quality disturbance classification using synchro squeezed wavelet transform and vision transformer. Electric Power Systems Research, 252, 112460. https://doi.org/10.1016/j.epsr.2025.112460

Chamchuen, S., Siritaratiwat, A., Fuangfoo, P., Suthisopapan, P., & Khunkitti, P. (2021). Adaptive salp swarm algorithm as optimal feature selection for power quality disturbance classification. Applied Sciences, 11(12), 5670. https://doi.org/10.3390/app11125670

Chen, S., Li, Z., Pan, G., & Xu, F. (2022). Power quality disturbance recognition using empirical wavelet transform and feature selection. Electronics, 11(2), 174. https://doi.org/10.3390/electronics11020174

Choe, S., & Yoo, J. (2021). Wavelet packet transform modulus-based feature detection of stochastic power quality disturbance signals. Applied Sciences, 11(6), 2825. https://doi.org/10.3390/app11062825

Dzikrurrokhim, M. R., & Ahlgren, E. O. (2026). Target-based energy transitions in the electricity systems of Southeast Asia: A comparative study of Indonesia, the Philippines, and Vietnam. Renewable and Sustainable Energy Transition, 10, 100152. https://doi.org/10.1016/j.rset.2026.100152

Gong, F., Diao, R., Cai, C., & Han, J. (2026). POA-optimized 1D-CNN with channel attention for power quality disturbance classification under strong noise conditions. Energies, 19(15), 3514. https://doi.org/10.3390/en19153514

Greenwell, B. M. (2022). Gradient boosting machines. In Tree-based methods for statistical learning in R. Chapman and Hall/CRC. https://doi.org/10.1201/9781003089032-8

Harriz, M. A., Akbariani, N. V., Setiyowati, H., & Santoso, H. (2023). Enhancing the efficiency of Jakarta's mass rapid transit system with XGBoost algorithm for passenger prediction. Jambura Journal of Informatics, 5(1), 1-6. https://doi.org/10.37905/jji.v5i1.18814

Harriz, M. A., & Setiyowati, H. (2023). Komparasi algoritma decision tree dan KNN dalam mengklasifikasikan daerah berdasarkan produksi listrik. JIKO (Jurnal Informatika dan Komputer), 7(2), 167. https://doi.org/10.26798/jiko.v7i2.787

Harriz, M. A., Setiyowati, H., & Akbariani, N. V. (2025). A literature review on IoT applications in fishery cold chains: Integrating blue economy principles for sustainable food security. Journal of Food Technology Research, 12(4), 380-393. https://doi.org/10.18488/jftr.v12i4.4631

IEEE. (2019). IEEE recommended practice for monitoring electric power quality (IEEE Std 1159-2019). https://doi.org/10.1109/IEEESTD.2019.8796486

Ismael, M. R., Abd, H. J., & Homod, R. Z. (2024). An algorithm for the classification of power quality disturbance signals using a tunable-Q-factor wavelet transform and ensemble learning methodology. Electrical Engineering, 106, 5253-5265. https://doi.org/10.1007/s00202-024-02294-y

Janthong, S., & Phukpattaranont, P. (2025). Power quality disturbance detection using improved grasshopper optimization and adaptive boosted random forest. Electrical Engineering, 107, 14265-14300. https://doi.org/10.1007/s00202-025-03260-y

Juarez, M., Zamora, A., Ortiz-Bejar, J., Silva, J., Cerda, J., & Paternina, M. (2024). PQ-SyDa: Power quality synthetic disturbances dataset. In Proceedings of the IEEE PES Generation, Transmission and Distribution Latin America Conference and Industrial Exposition (GTDLA) (pp. 1-6). IEEE. https://doi.org/10.1109/gtdla61236.2024.10913752

Kee, O. K. H., Kee, K.-K., Yong, C. Y., Rashidi, R., & Lo, T. H. (2025). Development of a data-driven energy monitoring system for power consumption and power quality monitoring. Journal of Renewable Energy and Smart Grid Technology, 20, 163-175. https://doi.org/10.69650/rast.2025.263660

Khaldi, B., Dekhandji, F., & Recioui, A. (2025). Low-cost IoT-based smart meter for real-time power quality monitoring and disturbance detection using embedded 1D CNN. Journal of Energy Systems, 9(4), 365-379. https://doi.org/10.30521/jes.1718242

Korstanje, J. (2025). Gradient boosting with XGBoost, LightGBM, and CatBoost. In Advanced forecasting with Python. Apress. https://doi.org/10.1007/979-8-8688-2028-1_16

Li, S., Zhu, X., & Zhou, D. (2025). Power quality disturbance signal denoising and detection based on improved DBO-VMD combined with wavelet thresholding. Electric Power Systems Research*, 238, 111193. https://doi.org/10.1016/j.epsr.2024.111193

Liu, J. (2021). Importance-SMOTE: A synthetic minority oversampling method for noisy imbalanced data. Soft Computing, 26(3), 1141-1163. https://doi.org/10.1007/s00500-021-06532-4

Madgula, S., Veeramsetty, V., & Durgam, R. (2025). Signal processing approaches for power quality disturbance classification: A comprehensive review. Results in Engineering, 25, 104569. https://doi.org/10.1016/j.rineng.2025.104569

Mahla, R., & Garg, M. M. (2024). Power quality disturbances and detection methods: A brief review. In Proceedings of the IEEE 11th Power India International Conference (PIICON) (pp. 1-6). IEEE. https://doi.org/10.1109/piicon63519.2024.10995132

Mishra, M. K. (2023). Power quality in power distribution systems: Concepts and applications. CRC Press. https://doi.org/10.1201/9781032617305

Palomares-Salas, J. C., Aguado-González, S., & Sierra-Fernández, J. M. (2025). Robustness of machine learning and deep learning models for power quality disturbance classification: A cross-platform analysis. Applied Sciences, 15(19), 10602. https://doi.org/10.3390/app151910602

Reyseliani, N., Hidayatno, A., & Purwanto, W. W. (2022). Implication of the Paris agreement target on Indonesia electricity sector transition to 2050 using TIMES model. Energy Policy, 169, 113184. https://doi.org/10.1016/j.enpol.2022.113184

Shiroya, N. J., Dey, M., & Rana, S. P. (2026). Ensemble learning for event detection and disturbance classification in power quality data from solar energy systems. Next Energy, 11, 100556. https://doi.org/10.1016/j.nxener.2026.100556

Uzel, H. (2025). An explainable hybrid deep learning approach for power quality disturbance classification. Bulletin of the Polish Academy of Sciences: Technical Sciences, Article 157326. https://doi.org/10.24425/bpasts.2026.157326

Veeramsetty, V., Gundapu, R. K., Aitha, D., & Aluri, N. (2023). PQ disturbances dataset [Data set]. Mendeley Data. https://doi.org/10.17632/nkdpg8mn4f.3

Wang, J., & Awang, N. (2026). SVED-SMOTE: Support vector enhanced density SMOTE oversampling technique for imbalanced data classification. International Journal of Machine Learning and Cybernetics, 17. https://doi.org/10.1007/s13042-026-03240-z

Wang, T., Zhuo, J., Hou, Y., Lu, Z., & Li, Y. (2026). Power quality disturbance classification via a time-frequency feature-fused transformer model with cross-attention mechanism. Electric Power Systems Research, 251, 112330. https://doi.org/10.1016/j.epsr.2025.112330

Xu, W., Wang, R., Zhang, Y., Wang, J., Wang, Z., Wu, X., Li, W., Li, X., Zhang, M., & Sun, L. (2026). A power quality disturbance classification method using a hybrid transformer and discrete wavelet transform model. Electric Power Systems Research, 253, 112547. https://doi.org/10.1016/j.epsr.2025.112547

Yang, S., Shan, T., & Yang, X. (2025). Interpretable DWT-1DCNN-LSTM network for power quality disturbance classification. Energies, 18(2), 231. https://doi.org/10.3390/en18020231