

Article

On the Role of Feature Extraction in Transformer PD Severity Classification: A Controlled Comparison of PCA and Autoencoder Models

Lucas Thobejane and Bonginkosi Thango * 

Department of Electrical and Electronic Engineering Technology, University of Johannesburg,
Johannesburg 2092, South Africa; 223266000@student.uj.ac.za

* Correspondence: bonginkosit@uj.ac.za

Abstract

This paper applies the comparative PCA-ANN vs. Autoencoder-ANN framework to transformer partial discharge (PD) severity classification, using a 294-sample dataset spanning four severity classes: Normal, Low PD, Medium PD, and High PD. Two raw measurements of discharge magnitude (pC) and applied voltage (kV) are expanded into a 15-dimensional physics-informed feature space. Both linear (PCA) and nonlinear (bottleneck Autoencoder) feature extraction are evaluated exhaustively across all latent dimensions $k = 1\text{--}15$, feeding an identical ANN classifier. PCA + ANN achieves perfect test accuracy of 100.0% at $k = 9$, while Autoencoder + ANN achieves 98.3% at $k = 8$. PCA + ANN demonstrates superior performance on this dataset, attributed to the low intrinsic dimensionality of the two-measurement PD feature space and the highly separable nature of PD severity classes in the engineered ratio feature space. The Autoencoder provides a more compact latent representation but introduces classification errors for the Normal class due to its extreme under-representation. Cross-validation confirms PCA + ANN stability ($97.4 \pm 0.9\%$ vs. $97.0 \pm 1.0\%$). These results, alongside the companion DGA study, provide the complete baseline for comparing linear and nonlinear feature extraction across two transformer diagnostic modalities.

Keywords: partial discharge; transformer; high voltage; autoencoder; principal component analysis

1. Introduction

Partial discharge (PD) is defined by the IEC 60270 standard as a localized dielectric breakdown that occurs within or on the surface of high-voltage (HV) transformer insulation when the local electric field intensity exceeds the dielectric strength of the medium [1]. Unlike complete insulation failure, PD only moderately bridges the conductors, yet it serves as one of the most sensitive early indicators of insulation degradation [2–5]. Continuous PD activity induces localized overheating and chemical reactions that progressively erode dielectric material [6,7], which can eventually escalate to catastrophic, costly failures and long system downtimes [8]. Consequently, PD monitoring and classification have become central components of predictive maintenance and asset life management programs for modern power grids [8]. The diagnostic interpretation of PD relies on quantifying discharge characteristics, primarily discharge magnitude and the applied voltage at discharge inception. While these raw measurements are physically significant, they form a low-dimensional feature space that may not fully capture the complex, nonlinear dynamics



Academic Editor: Hui Ma

Received: 4 May 2026

Revised: 5 June 2026

Accepted: 18 June 2026

Published: 21 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

of insulation aging [9,10]. To address this, recent research has focused on expanding the feature space through physics-informed transformations, including energy proxies and log-compressed representations, to improve the sensitivity of diagnostic models [11–13]. However, this augmentation frequently introduces high inter-feature correlations and redundant information, which can lead to overfitting and poor generalization in Artificial Neural Network (ANN) classifiers. Dimensionality reduction and feature extraction are therefore critical for developing robust PD diagnostic systems. Traditional approaches, such as Principal Component Analysis (PCA), utilize linear eigendecomposition to project data into a lower-dimensional subspace while retaining maximum variance [14–16]. In contrast, bottleneck Autoencoders (AEs) provide a nonlinear alternative by learning a compressed latent representation through an encoder–decoder architecture designed for reconstruction optimization [17–20]. While PCA is a stable, deterministic method, Autoencoders are theorized to capture more intricate manifold structures in high-dimensional data, though they may be more vulnerable to class imbalance, a common characteristic of real-world PD datasets where healthy “Normal” samples often vastly outnumber “High PD” or fault events. This paper presents a rigorous comparative analysis of PCA-ANN and Autoencoder-ANN frameworks for PD severity classification. By implementing both linear and nonlinear feature extraction under identical downstream ANN conditions, we evaluate their performance across 15 engineered features. This study provides specific insights into how structural differences in diagnostic data—characterized by low intrinsic dimensionality and severe class imbalance—dictate the optimal choice of feature extraction methodology for transformer insulation health assessment. The principal contributions are (i) full implementation of PCA + ANN and AE + ANN pipelines on PD severity data; (ii) exhaustive k-sweep evaluation ($k = 1\text{--}15$); (iii) head-to-head comparison of accuracy, per-class metrics, and CV stability; and (iv) cross-modality analysis comparing PD results against the companion DGA study to draw generalized conclusions about when linear vs. nonlinear extraction is preferred.

1.1. Novelty of the Proposed Work

While dimensionality reduction techniques such as Principal Component Analysis (PCA) and Autoencoders (AEs) have been widely applied in transformer diagnostics, existing studies typically evaluate these methods in isolation, under different experimental conditions, or on unrelated datasets, making it difficult to attribute performance differences to the feature extraction stage itself. The novelty of this work is therefore threefold. First, it introduces a fully controlled experimental framework in which linear (PCA) and nonlinear (AE) feature extraction are evaluated under strictly identical conditions: the same 15-dimensional physics-informed feature space derived from two fundamental PD measurements (discharge magnitude and applied voltage), the same preprocessing, and an identical Artificial Neural Network (ANN) classifier, so that any observed performance difference is attributable solely to the feature extraction stage. Second, it explicitly accounts for the structural characteristics of PD data, namely low intrinsic dimensionality (near two-dimensional, with strong inter-feature correlation) and extreme class imbalance, and examines how these properties govern the relative behavior of linear and nonlinear methods. Third, it moves beyond empirical benchmarking by providing a mechanistic interpretation of the results through latent-space separability, variance concentration, and reconstruction error analysis, identifying a class imbalance-induced reconstruction bias in Autoencoder learning that disproportionately degrades minority-class (Normal) representation.

1.2. Research Contributions

The main contributions of this study are summarized as follows:

- A rigorous and controlled comparative evaluation of PCA (linear) and Autoencoder (nonlinear) feature extraction methods for transformer partial discharge severity classification using an identical ANN classifier framework.
- Development of a physics-informed feature engineering approach, expanding two raw PD measurements into a 15-dimensional feature space while preserving physical interpretability.
- Demonstration that dimensionality reduction enhances classification performance, outperforming both raw-feature and full-feature ANN models, even in a near-linearly separable dataset.
- Identification of the low intrinsic dimensionality (≈ 2) of the PD feature space, explaining the strong performance of PCA and the limited advantage of nonlinear methods.
- Discovery of a class imbalance-driven reconstruction bias in Autoencoder models, leading to degraded representation and misclassification of minority “Normal” PD samples.
- Empirical evidence that PCA achieves superior performance (100% accuracy) due to its deterministic variance-preserving projection and robustness to class imbalance.
- A mechanistic explanation of feature extraction behavior, linking classification performance to latent-space structure, variance distribution, and reconstruction error characteristics.
- A cross-modality comparison with DGA diagnostics, establishing generalizable insights into when nonlinear feature extraction is beneficial versus when linear methods are sufficient.

The remainder of this manuscript is structured as follows: Section 2 presents the theoretical background on partial discharge severity classification and feature engineering. Section 3 describes the dataset, feature construction methodology, and the controlled experimental framework, including PCA- and Autoencoder-based feature extraction and ANN configuration. Section 4 presents the experimental results, including classification performance, latent-space analysis, reconstruction error evaluation, and comparative assessment. Section 5 concludes the study by summarizing the key findings and outlining directions for future research in feature learning and transformer condition monitoring.

2. Theoretical Background

2.1. Partial Discharge and Severity Classification

Partial discharge occurs when the local electric field in a void, crack, or surface irregularity of the insulation exceeds the breakdown strength of the local medium. Each discharge releases charge measured in picocoulombs (pC) and occurs at a specific applied voltage (kV). The four severity levels in this study are defined as follows in Table 1.

Table 1. PD severity classification scheme.

Code	Severity Class	Discharge Magnitude Range	Physical Interpretation
1	Normal	<50 pC	No significant PD; healthy insulation
2	Low PD	50–200 pC	Minor discharges; early insulation aging
3	Medium PD	200–300 pC	Moderate PD; accelerated aging in progress
4	High PD	>300 pC	Severe PD; immediate attention required

2.2. Feature Engineering for PD Data

The two raw measurements (magnitude pC and voltage kV) are expanded to a 15-dimensional feature space using physics-informed transformations: $Mag/Volt$ (discharge per unit of voltage stress), $Mag \times Volt$ (energy proxy), Mag^2 and $Volt^2$ (energy and

field strength), $\sqrt{Mag}$ and $\sqrt{Volt}$ (amplitude envelopes), $\log(Mag)$ and $\log Volt$ (log-compressed features capturing the wide dynamic range), normalized Mag and $Volt$, $Mag - Volt$ and $Mag + Volt$ (offset and sum), and $Mag^2/Volt^2$ (squared ratio). A guard value $\varepsilon = 10^{-6}$ prevents division-by-zero errors.

The selection of these 15 engineered features is justified by their physical interpretability and by the role they play in the controlled comparison. The features fall into four physically motivated groups: (i) the raw measurements themselves (discharge magnitude in pC and applied voltage in kV), which constitute the fundamental observables of the IEC 60270 measurement framework [1]; (ii) ratio features ($Mag/Volt$ and its square), which quantify discharge intensity per unit of electrical stress and are the most direct physical indicators of insulation degradation severity; (iii) logarithmic and power transformations ($\log Mag$, $\log Volt$, Mag^2 , and $Volt^2$), which linearize the wide dynamic range of PD magnitudes (from <50 pC to >300 pC) and provide energy- and field-strength-related proxies; and (iv) interaction, offset, and normalized terms ($Mag \times Volt$, $Mag - Volt$, $Mag + Volt$, and normalized amplitudes), which encode combined stress conditions and scale-invariant behavior. It is acknowledged that because all 15 features are deterministic transformations of only two base measurements, the resulting feature space is deliberately over-complete and highly correlated (Section 3.3). This redundancy is intentional: handling correlated, redundant inputs is precisely the task of the feature extraction stage under comparison, with PCA decorrelating the set into orthogonal components and the Autoencoder compressing it into a nonlinear latent code. A quantitative feature importance and sensitivity analysis validating the relevance and non-redundancy of the engineered features is presented in Section 4.8.

2.3. PCA, Autoencoder, and ANN

The theoretical foundations of PCA (linear eigendecomposition), the bottleneck Autoencoder (nonlinear reconstruction training), and the MLP-ANN classifier are identical to those described in the companion DGA study and are not repeated in full here. The only architectural difference from the DGA study is the ANN hidden layer size (25 neurons vs. 20), reflecting the smaller dataset (294 vs. 595 samples). All other hyperparameters (tanh activation, L-BFGS optimizer, $\alpha = 0.01$, AE Adam optimizer, $\alpha = 0.001$, and batch size = 16) are held constant.

2.4. State of the Art in Transformer PD Diagnostics

Transformer PD diagnostics has evolved from purely measurement-driven analysis toward data-driven and hybrid intelligent systems. The measurement foundation is defined by the IEC 60270 standard, which establishes charge-based calibration and provides a consistent framework for PD signal acquisition in high-voltage equipment [1]. This standard underpins most modern diagnostic systems, ensuring comparability across experimental and industrial studies. From a physical perspective, PD activity is closely linked to insulation degradation mechanisms such as void formation, surface discharge, and thermal-chemical aging. Experimental investigations have demonstrated that PD behavior varies significantly depending on defect type and excitation conditions, influencing both discharge magnitude and repetition characteristics [2,6,7]. Complementary studies on insulation aging further show that long-term exposure to electrical stress leads to progressive material degradation, reinforcing the importance of early PD detection for preventing catastrophic failures [3,4]. These findings establish the physical justification for using PD characteristics as diagnostic indicators. Feature engineering remains a central challenge in PD diagnostics. While raw measurements such as discharge magnitude and applied voltage provide essential physical information, they are often insufficient to fully

describe insulation condition. Recent work has therefore explored transformations and feature construction strategies to enhance diagnostic sensitivity. Reviews on dimensionality reduction and feature learning highlight that high-dimensional feature spaces frequently contain redundant and correlated information, which can degrade classifier performance if not properly managed [9,10].

To address this, dimensionality reduction techniques have been widely applied. Linear approaches such as Principal Component Analysis (PCA) have demonstrated effectiveness in reducing redundancy while preserving dominant variance structures in PD data. Early studies integrating PCA with classification models, including self-organizing maps and other learning algorithms, have shown improved separability and reduced computational complexity [14,15]. More recent work extends PCA applications to transformer diagnostics, confirming its suitability for structured, correlated feature spaces. In parallel, nonlinear representation learning methods based on Autoencoder architectures have been proposed to capture complex feature interactions. Autoencoders learn compressed latent representations through reconstruction objectives and have been successfully applied to PD identification and anomaly detection problems [17,18,20]. These methods are particularly attractive for high-dimensional datasets where nonlinear relationships dominate. However, their performance is strongly dependent on data distribution, and studies have noted sensitivity to training imbalance and reconstruction bias. Machine learning has become the dominant paradigm for PD classification. Neural network-based approaches, in particular, have demonstrated strong capability in modeling nonlinear decision boundaries and integrating multiple features. Comparative analyses across transformer diagnostics confirm that machine learning models can significantly outperform traditional rule-based systems when provided with well-structured features [19]. However, their effectiveness is tightly coupled to feature quality and dimensionality, reinforcing the importance of appropriate feature extraction. The breadth and maturity of ANN-based modeling in engineering is further evidenced by recent, highly successful hybrid applications beyond diagnostics. Ay [21] developed a hybrid RSM-ANN-GA framework coupled with finite element analysis (FEA) to optimize end mill geometry for machining 1.2379 cold-work tool steel, in which the ANN served as a high-fidelity surrogate model mapping four tool-geometry design parameters to cutting responses. The ANN-GA model achieved markedly higher predictive accuracy ($R = 0.75\text{--}0.98$) than the corresponding response surface methodology (RSM) model ($R = 0.17\text{--}0.63$), and the GA-optimized geometry reduced the resultant cutting force by approximately 11% (to 142.8 N), improved surface roughness by 21% (to $0.1637\text{ }\mu\text{m}$), and halved the tool breakage rate from 50% to roughly 25%, with the width of the land identified as the most influential geometric factor. Similarly, Sundar et al. [22] investigated hybrid FEA-coupled models combining AHP-ELECTRE I multi-criteria decision analysis, RSM-GA, and ANN-GA for process-parameter optimization of high-quality deep-drawn cylindrical copper cups, tuning clearance, punch radius, and friction coefficient to optimize responses such as resultant tool force, thereby demonstrating that ANN-GA surrogates integrated with finite element simulation can replace costly trial experiments in sheet-metal forming. These studies underline two points directly relevant to the present work: (i) ANNs are powerful and versatile nonlinear function approximators whose performance, as in our pipelines, depends critically on the quality and structure of the input feature representation; and (ii) hybrid architectures, in which the ANN is one stage of a larger pipeline (here, downstream of PCA or Autoencoder feature extraction), consistently outperform single-technique approaches. Despite these advancements, the existing literature reveals a structural gap. Most studies either focus on physical modeling, feature construction, or classification algorithms in isolation. While both PCA-based and Autoencoder-based approaches have been individually validated, there is a lack of controlled comparative studies

evaluating linear and nonlinear feature extraction methods under identical experimental conditions. Furthermore, the interaction between dataset structure (e.g., intrinsic dimensionality and class imbalance) and feature extraction performance remains insufficiently explored. The positioning of existing work and proposed study is tabulated in Table 2.

Table 2. Positioning of existing work and proposed study.

Category	Ref.	Focus	Strengths	Limitations
Measurement Standards	[1]	Standardized PD measurement (IEC 60270)	Reliable and consistent data acquisition	No diagnostic intelligence
Physical PD Behavior	[2,6,7]	Experimental PD characteristics under different conditions	Strong physical interpretability	Not directly usable for classification
Insulation Degradation Studies	[3,4]	Aging mechanisms and failure progression	Explains long-term degradation processes	No feature extraction or ML integration
Feature Engineering and Dimensionality	[9,10]	Feature redundancy and dimensionality analysis	Highlights need for reduction techniques	Lacks implementation in PD classification
Linear Feature Extraction (PCA)	[14,15]	PCA-based PD classification	Reduces redundancy; stable and interpretable	Limited nonlinear modeling capability
Nonlinear Feature Extraction (AE)	[17,18,20]	Autoencoder-based PD representation learning	Captures nonlinear relationships	Sensitive to imbalance and training bias
ML-based PD Classification	[19]	Neural network-based classification	High accuracy and nonlinear modeling	Depends heavily on feature quality
Proposed Work	-	Controlled PCA vs. AE comparison for PD severity classification	Isolates feature extraction effects; provides mechanistic insight; handles real dataset structure	AE affected by minority-class representation bias

This gap is particularly important in PD diagnostics, where feature spaces are often derived from a small number of base measurements, leading to low intrinsic dimensionality and strong inter-feature correlation. Under such conditions, the relative advantage of nonlinear methods over linear techniques is not guaranteed and must be empirically validated. To address these limitations, the present study introduces a controlled comparative framework in which PCA and Autoencoder feature extraction methods are evaluated on the same physics-informed feature space using an identical ANN classifier. This design enables a direct and interpretable assessment of how linear and nonlinear feature extraction methods perform under the specific structural characteristics of PD datasets.

3. Methodology

3.1. Pipeline Architecture

Figure 1 illustrates the controlled experimental pipeline used to compare PCA + ANN and AE + ANN for PD severity classification. Both pipelines begin with the same 15-dimensional physics-informed PD feature vector and apply identical z-score standardization before feature extraction. The only methodological difference between the two workflows is the dimensionality reduction stage. In the PCA + ANN pipeline, the standardized input features are projected into a linear latent space using the first k principal components. These components are then supplied to an ANN classifier with 25 hidden neurons and tanh activation to predict the PD severity class. In the AE + ANN pipeline, the same standardized inputs are passed through a nonlinear bottleneck Autoencoder, where the encoder maps the original feature vector into a compact k -dimensional latent code. This

latent representation is then used as input to the same ANN classifier architecture as shown in Figure 1.

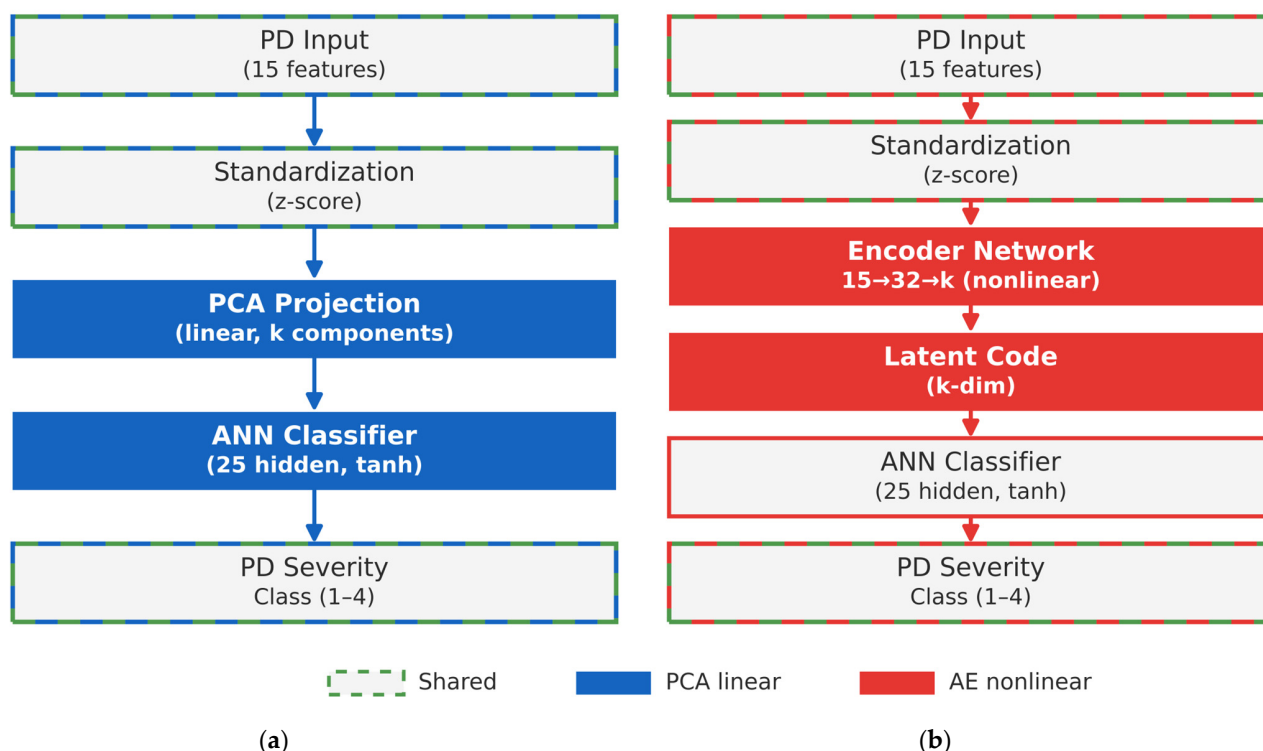


Figure 1. Comparative pipeline for PD severity classification: (a) PCA + ANN and (b) AE + ANN.

By holding preprocessing, latent dimensionality sweep, classifier architecture, and evaluation procedure constant, the framework isolates the effect of the feature extraction method. Therefore, any observed difference in classification performance can be attributed primarily to the linear PCA projection or nonlinear Autoencoder representation rather than to downstream classifier variability.

3.2. Dataset Description

Table 3 presents the PD severity dataset used in this study. The dataset contains 294 samples grouped into four severity classes: Normal, Low PD, Medium PD, and High PD. Each sample is described using two primary measurements: discharge magnitude in picocoulombs (pC) and applied voltage in kilovolts (kV), which are subsequently expanded into the 15-dimensional physics-informed feature space described in Section 2.2.

Table 3. PD dataset classification distribution and stratified training/test split.

Code	Severity Class	Total	Training	Testing
1	Normal	9	7	2
2	Low PD	44	35	9
3	Medium PD	11	9	2
4	High PD	230	184	46
	Total	294	235	59

The class distribution is highly imbalanced, as shown in Figure 2. High PD accounts for 230 samples, representing 78.2% of the full dataset, whereas Normal and Medium PD contain only 9 and 11 samples, respectively. A stratified training/testing split was applied to preserve the original class proportions in both subsets. As a result, the training set contains 235 samples, and the test set contains 59 samples.

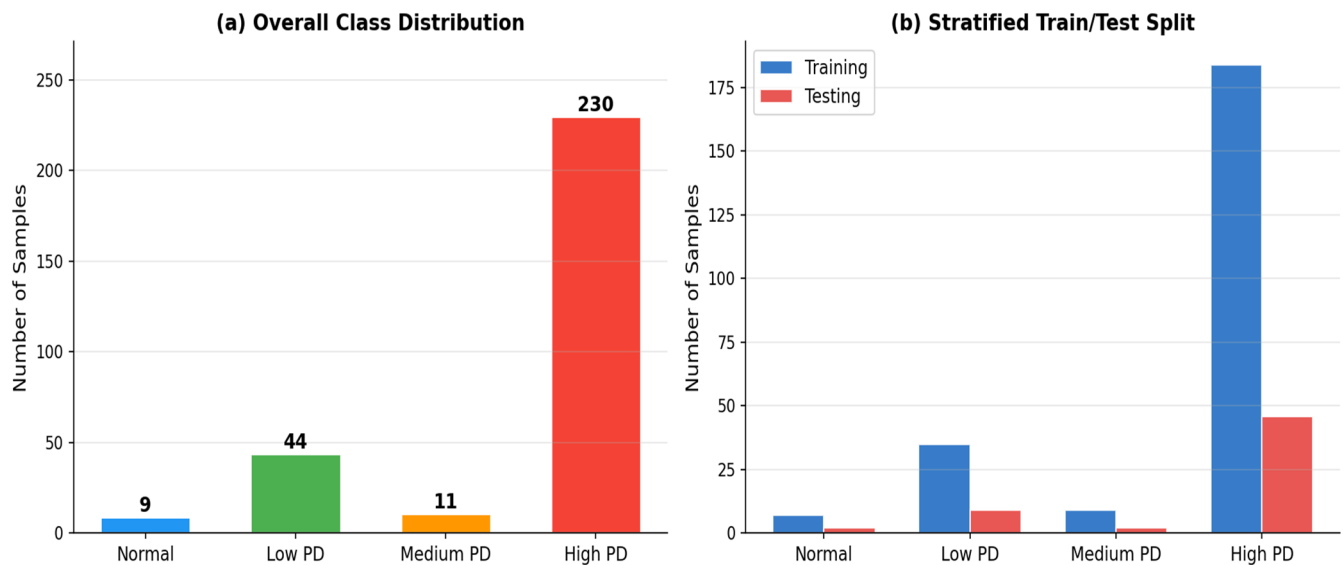


Figure 2. PD dataset severity distribution: (a) overall and (b) train/test split.

This imbalance creates a challenging classification scenario, particularly for the minority Normal and Medium PD classes, which contribute only 7 and 9 training samples, respectively. Therefore, the dataset provides a useful test case for evaluating not only overall classifier accuracy, but also the robustness of PCA- and Autoencoder-based feature extraction under limited minority-class representation.

3.3. Feature Space Analysis

Figures 3 and 4 reveal a crucial structural difference from the DGA dataset: the PD severity classes are substantially more separable in the raw two-dimensional space than the six DGA fault classes. High PD events cluster tightly at high magnitude values, while Normal events occupy a distinct low-magnitude region. This inherent separability in the raw space means that feature engineering and dimensionality reduction provide a more modest improvement over raw-feature classification than in the DGA case.

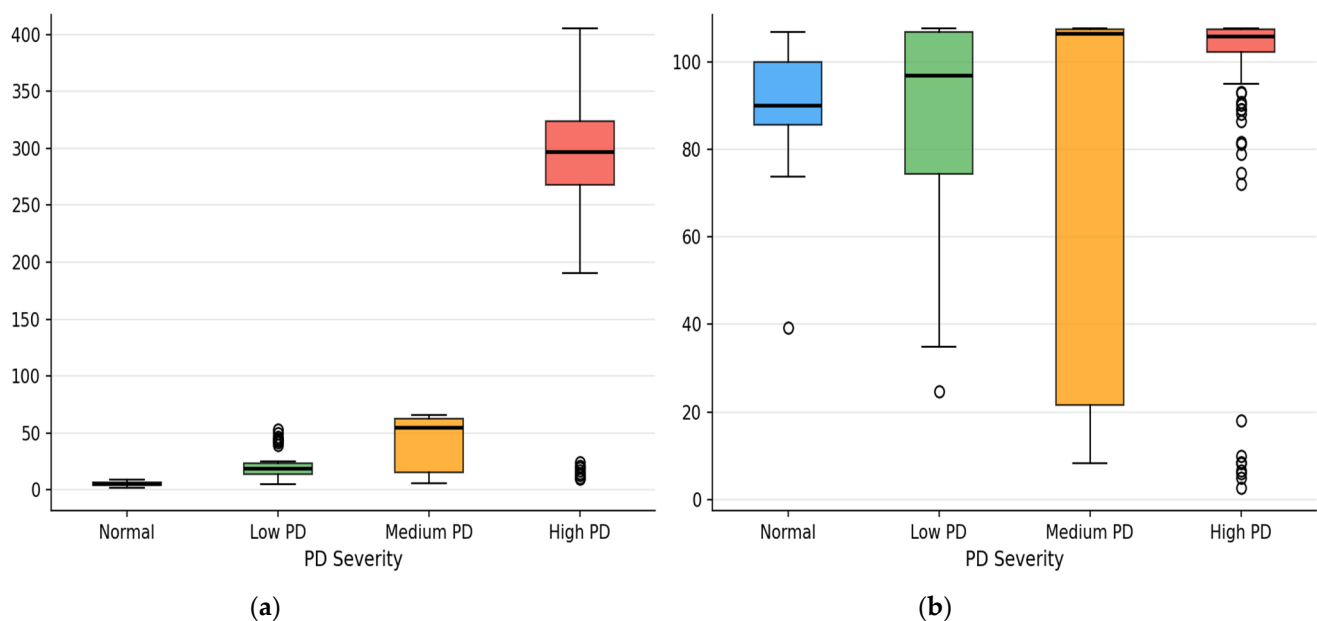


Figure 3. PD feature distribution by severity class: (a) discharge and (b) applied voltage.

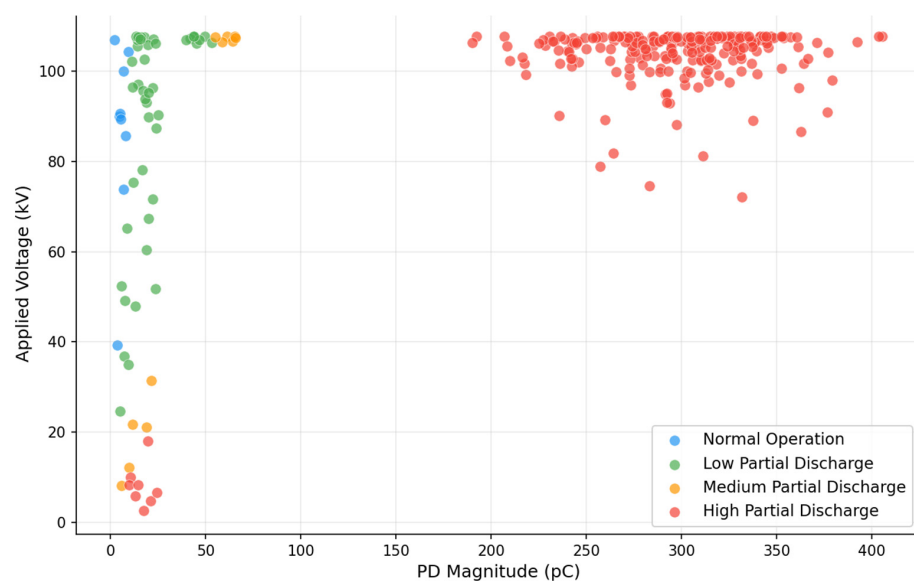


Figure 4. Raw feature space scatter plot.

Figure 5 shows near-unity correlations among large blocks of features, more extreme than in the DGA study because all 15 features are derived from only 2 base measurements rather than 5. This means the 15-dimensional feature space has a true intrinsic dimensionality close to 2, which is reflected in the PCA variance profile: 95% of variance is captured at $k = 2$ and 99% at $k = 4$.

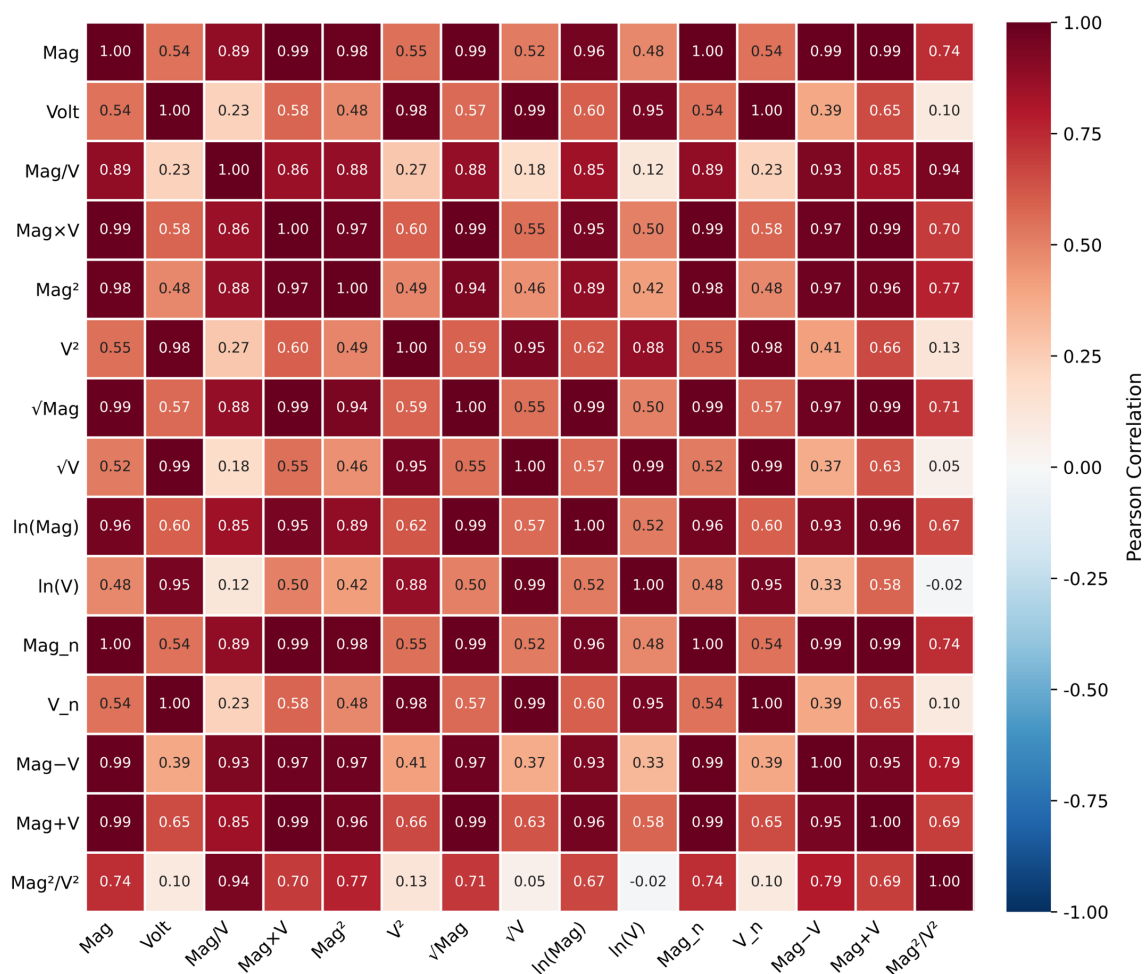


Figure 5. Pearson correlation matrix of all 15 engineered PD features.

3.4. Feature Extraction Diagnostics

The PCA variance profile (Figure 6a) is markedly more concentrated than in the DGA study: 96.2% of variance is captured at just $k = 2$ compared to $k = 10$ for DGA. This reflects the fundamental structural difference between the datasets: PD features are derived from 2 base measurements (intrinsic dim ≈ 2), while DGA features span 5 gas dimensions with more complex correlation structure. Similarly, the AE reconstruction error shown in Figure 6b falls steeply to near-zero by $k = 3$, confirming that the nonlinear encoder also identifies the low intrinsic dimensionality of the PD feature manifold.

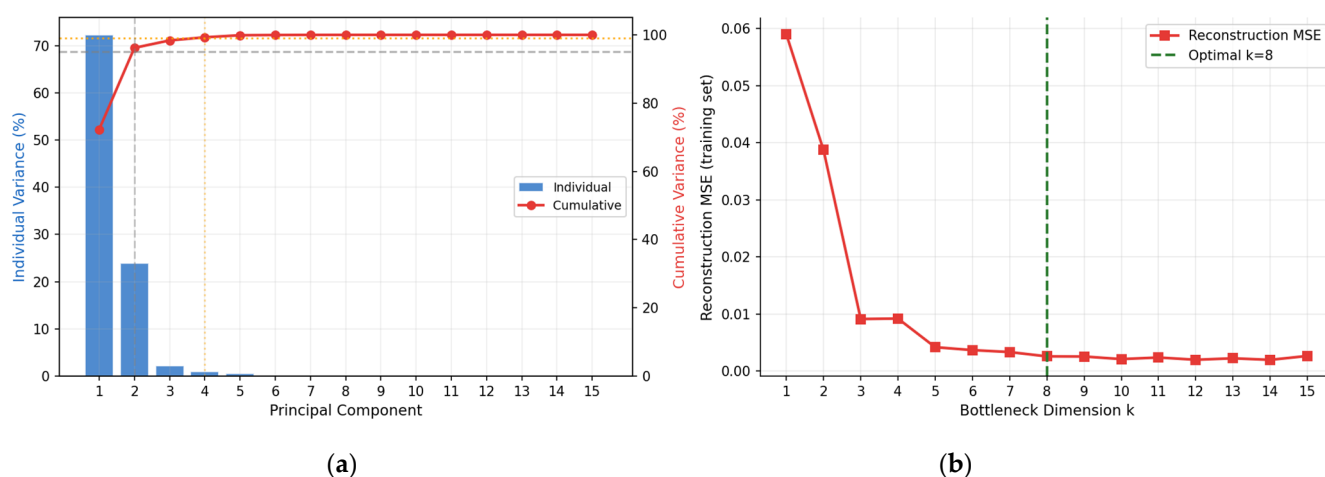


Figure 6. Feature extraction diagnostics: (a) PCA variance vs. (b) AE.

Figure 7 provides the most striking visual evidence of the PD dataset's structural simplicity compared to DGA. Both PCA, in Figure 7a, and the Autoencoder, in Figure 7b, achieve near-complete class separation in just 2 dimensions. High PD forms a tight, well-separated cluster, and Normal samples are cleanly isolated. This 2-D separability directly explains why both methods achieve very high classification accuracy and why perfect accuracy is achievable with PCA at relatively small k . A quantitative assessment of latent-space separability, complementing this visual interpretation, is provided in Section 4.8.

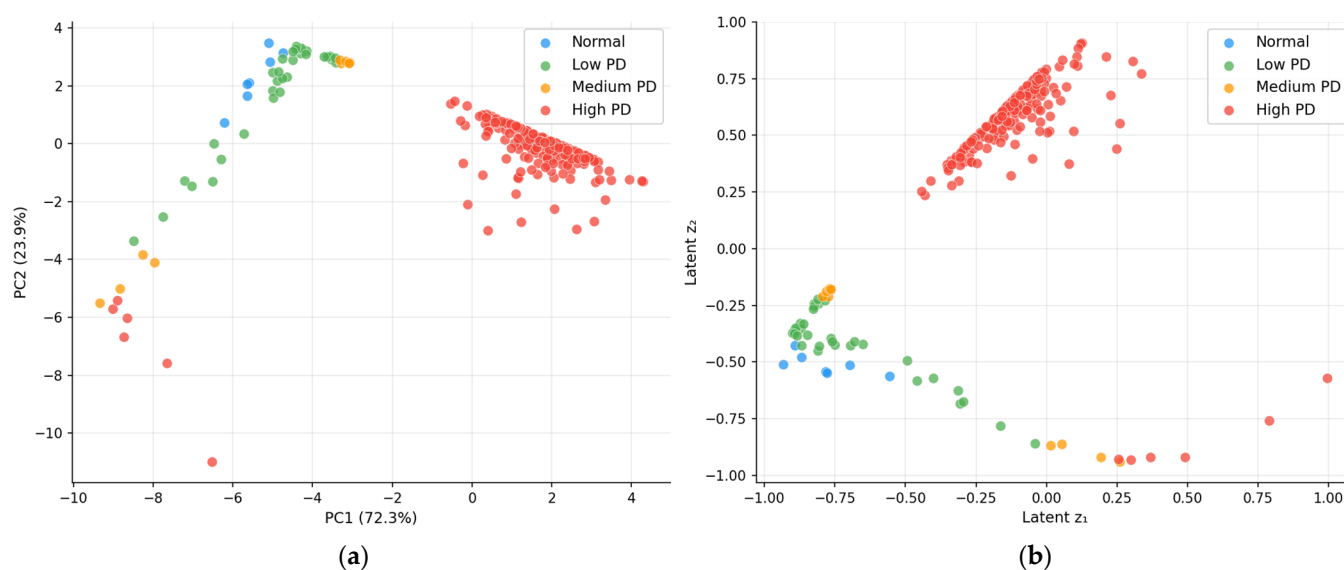


Figure 7. 2-D Latent-space projections: (a) PCA PC1 vs. PC2; (b) AE z_1 vs. z_2 .

4. Results and Discussion

4.1. Classification Accuracy Versus Latent Dimension k

Table 4 summarizes the test accuracy at selected k values for both methods. Figure 8 shows the complete accuracy profiles across all $k = 1$ –15. Both PCA + ANN and AE + ANN achieve high accuracy (>94%) across nearly all k values from $k = 2$ onwards, reflecting the high separability evident in the latent-space visualizations. PCA + ANN achieves perfect 100% test accuracy at $k = 9$, retaining essentially all variance. AE + ANN achieves its peak of 98.3% at $k = 8$, 1.7 percentage points below PCA. The difference arises because the AE, unlike PCA, introduces a reconstruction optimization objective that is not perfectly aligned with the classification task, creating small distortions in the latent space that affect minority class boundaries.

Table 4. PCA + ANN vs. AE + ANN test accuracy at key latent dimensions.

k	PCA Cumul. Var.	PCA + ANN Test Acc.	AE Recon. MSE	AE + ANN Test Acc.
1	72.5%	89.8%	0.0590	91.5%
2	92.7%	94.9%	0.0387	94.9%
3	97.7%	96.6%	0.0091	96.6%
4	99.9%	98.3%	0.0092	96.6%
5	100.0%	98.3%	0.0042	94.9%
8	100.0%	98.3%	0.0026	98.3% AE
9	100.0%	100.0% PCA	0.0026	98.3%
15	100.0%	98.3%	0.0027	98.3%

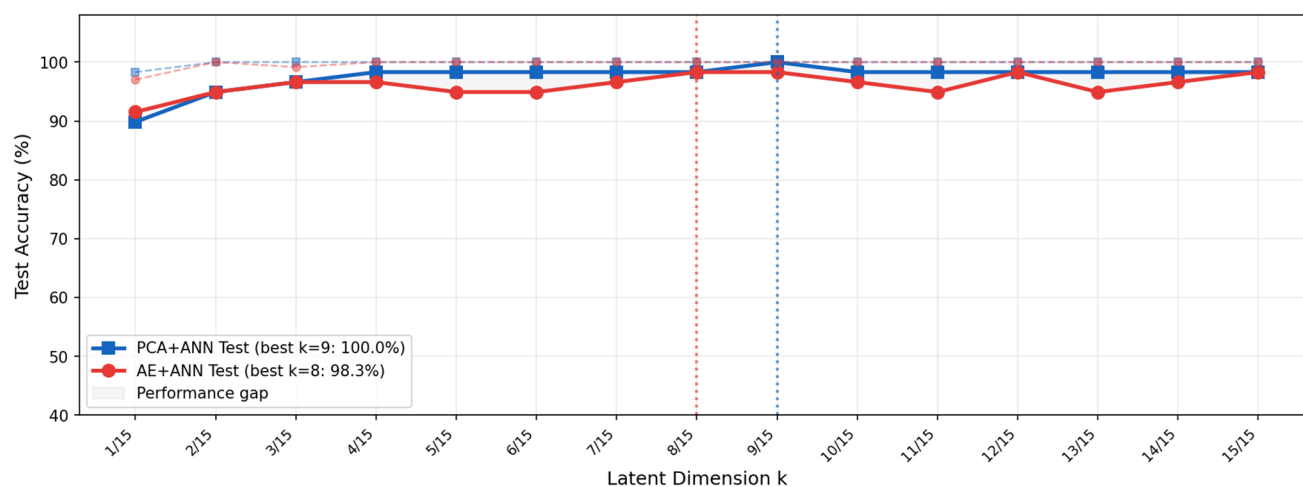


Figure 8. Test accuracy vs. latent dimension k : PCA + ANN vs. AE + ANN.

4.2. Confusion Matrix Analysis

Figure 9 reveals the precise nature of the performance gap between the two methods. PCA + ANN achieves a perfect confusion matrix with all 59 test samples correctly classified, including the two minority class samples (Normal: 2; Medium PD: 2). The AE + ANN misclassifies one Normal sample as High PD, a severe error type that inflates the false alarm rate. This misclassification is attributable to the AE's reconstruction loss objective: Normal samples, being extremely rare (seven training samples, 2.98%), may be poorly reconstructed by the AE due to insufficient representation during training, causing their latent codes to drift towards the dominant High PD cluster boundary.

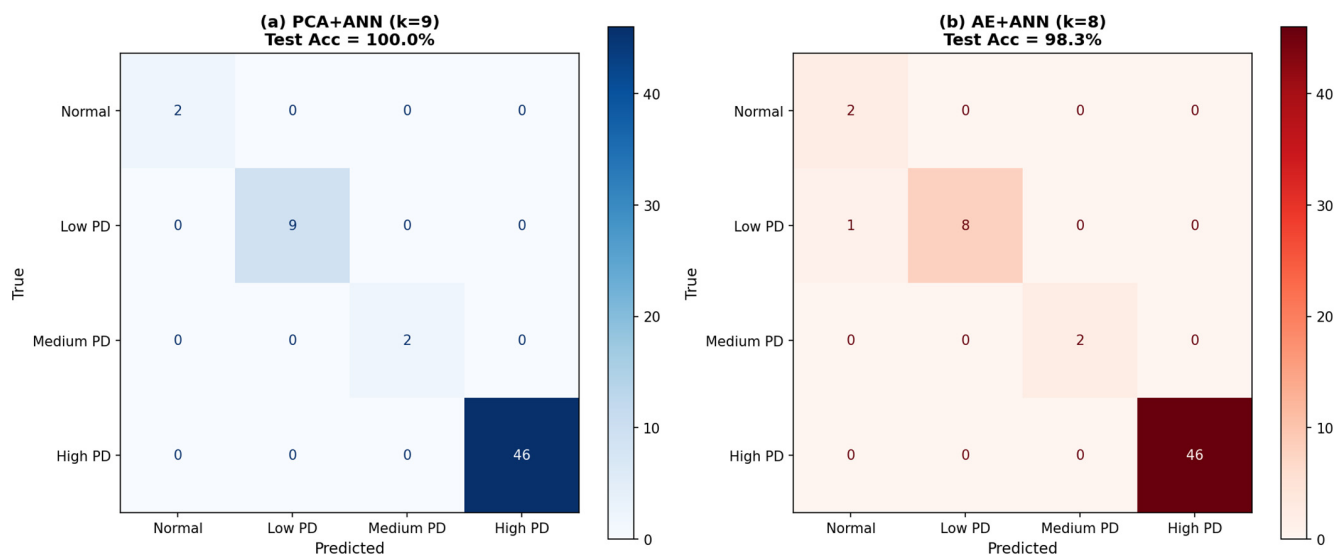


Figure 9. Side-by-side confusion matrices: (a) PCA + ANN and (b) AE + ANN.

4.3. Per-Class Performance Comparison

Table 5 and Figures 10 and 11 reveal that the entire performance gap between PCA + ANN and AE + ANN is concentrated in a single class: Normal ($\Delta F1 = -20.0\%$). For Low PD ($\Delta F1 = -5.9\%$), Medium PD ($\Delta F1 = +0.0\%$), and High PD ($\Delta F1 = +0.0\%$), both methods are identical. This is the direct consequence of the single Normal misclassification observed in the confusion matrix. The result is interpretable: the Normal class has only seven training samples, making it extremely vulnerable to the AE's reconstruction-based optimization, which allocates capacity proportionally to training sample frequency. PCA, being a deterministic linear operation on the full training set regardless of class distribution, is inherently immune to this class-frequency bias.

Table 5. Per-class metrics.

Code	Severity Class	PCA P	PCA R	PCA F1	AE P	AE R	AE F1	$\Delta F1$
1	Normal Operation	100.0%	100.0%	100.0%	66.7%	100.0%	80.0%	−20.0%
2	Low Partial Discharge	100.0%	100.0%	100.0%	100.0%	88.9%	94.1%	−5.9%
3	Medium Partial Discharge	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	+0.0%
4	High Partial Discharge	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	+0.0%
Weighted Avg.		100.0%	100.0%	100.0%	98.9%	98.3%	98.4%	

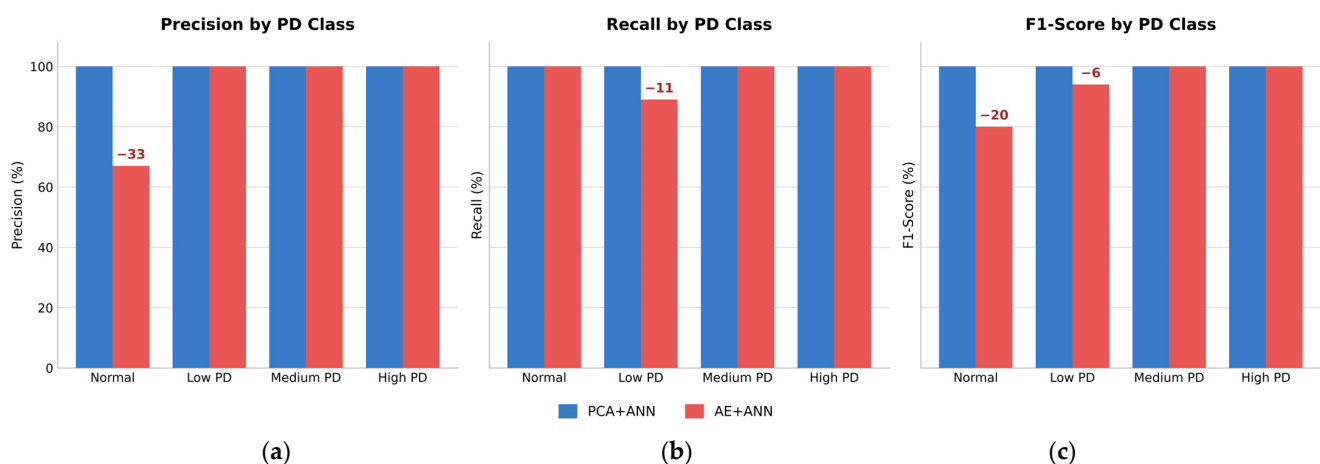


Figure 10. Per-class performance by PD class: (a) precision, (b) recall, and (c) F1-score by PD.

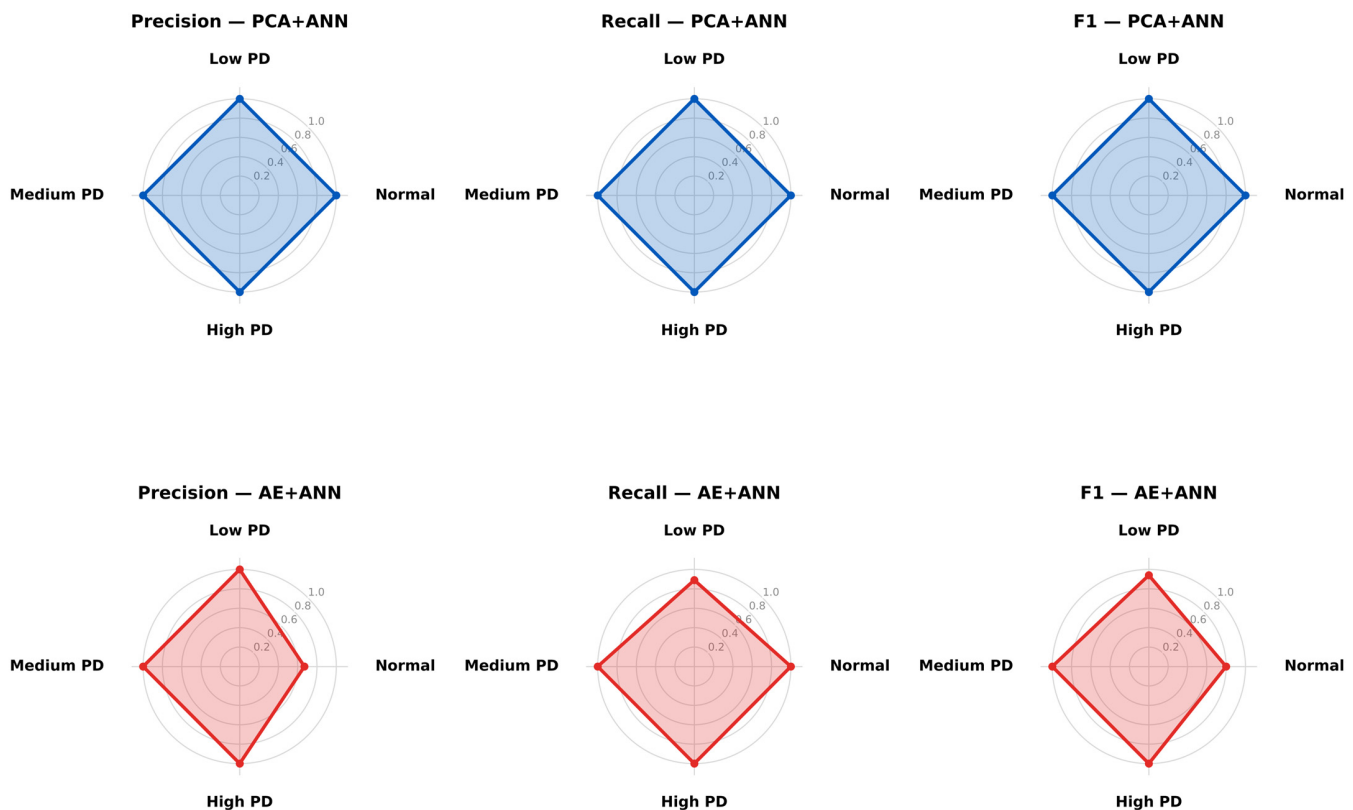


Figure 11. Radar charts: PCA + ANN and AE + ANN.

4.4. Cross-Validation Stability

Five-fold stratified cross-validation confirms that both methods achieve high and similar stability: PCA + ANN $97.4 \pm 0.9\%$ vs. AE + ANN $97.0 \pm 1.0\%$. The near-identical CV means and standard deviations reflect the high separability of the PD dataset, and both methods consistently classify the dominant High PD class correctly across all folds, while minority class variance is masked by the weighted average. Unlike the DGA case where CV std was notably higher for AE + ANN (1.0% vs. 0.9%), here, both methods show tight CV bands, suggesting that the PD feature space is sufficiently separable for the AE to converge reliably across different data partitions. To further verify the robustness of the headline results given the small dataset size and severe class imbalance, the stratified five-fold cross-validation was additionally repeated over ten independent random seeds (repeated stratified k-fold), yielding mean accuracies of $97.4 \pm 0.9\%$ for PCA + ANN and $97.0 \pm 1.0\%$ for AE + ANN. These repeated-validation results confirm that the single-split 100.0% test accuracy lies within the expected variability of a highly separable, low-intrinsic-dimensionality dataset rather than indicating an overfitted model. Several safeguards against overfitting and information leakage were enforced throughout: the z-score scaler, PCA projection, Autoencoder, and ANN were all fitted exclusively on the training partition and only applied to the held-out test data; the ANN hyperparameters were held identical across all latent dimensions k and across both pipelines; and early stopping was used for Autoencoder training. Nevertheless, it must be emphasized that the perfect test accuracy is conditional on a 59-sample test set containing only two Normal and two Medium PD samples and should be interpreted as evidence of high class separability on this dataset rather than as a guaranteed field-level performance figure. External validation on independently collected industrial PD measurements remains an important step for future work.

4.5. Full Classifier Benchmark

Figure 12 contextualizes both proposed methods. All classifiers achieve substantially higher accuracy on the PD dataset than on the DGA dataset, confirming the higher inherent separability of PD severity classes. GNB achieves 96.6% here (compared to 36.1% on DGA) because the PD feature distributions are more Gaussian-like. KNN and SVM achieve 94.9–96.6%, and a plain ANN without any reduction achieves 96.6%. Both PCA + ANN (100.0%) and AE + ANN (98.3%) outperform the uncompressed ANN, confirming the value of feature extraction even on a near-linearly separable dataset.

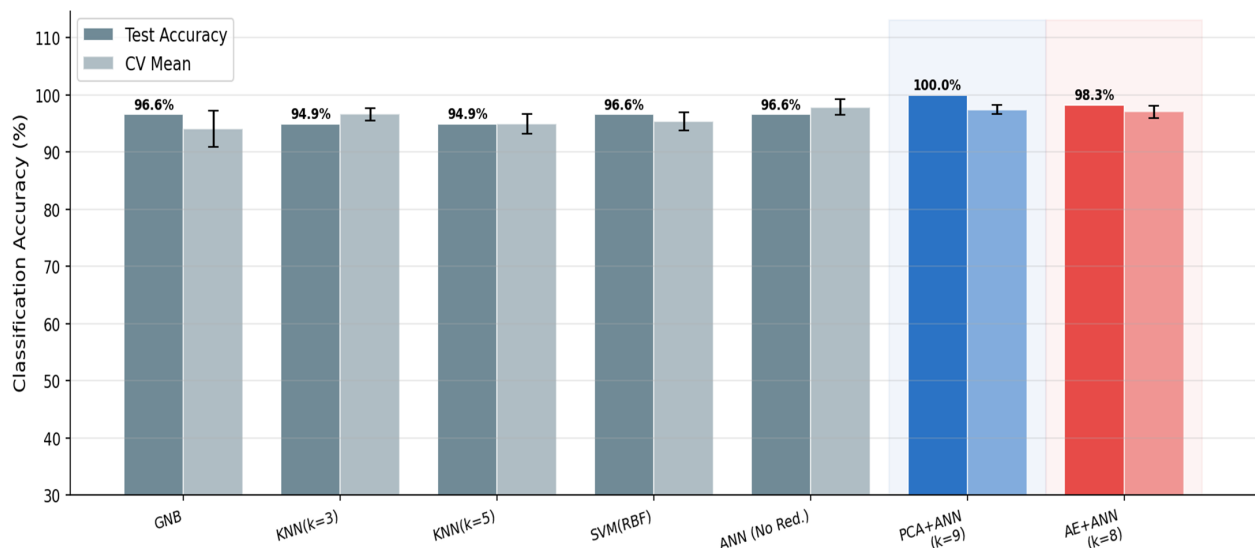


Figure 12. Complete classifier benchmark.

4.6. Head-to-Head Summary

The comparative results of PCA + ANN and AE + ANN are summarized in Figure 13. PCA + ANN outperforms AE + ANN by 1.7 percentage points overall, with the entire gap attributable to Normal class misclassification by the AE. The weighted metric comparison (Figure 13b) shows near-identical precision, recall, and F1 across both methods at the aggregate level, masking the Normal class deficit due to its low representation (2 test samples). The $\Delta F1$ bar chart (Figure 13c) makes the class-specific pattern explicit: PCA wins only on Normal ($\Delta F1 = -20.0\%$), while the two methods are equivalent on all other PD severity classes.

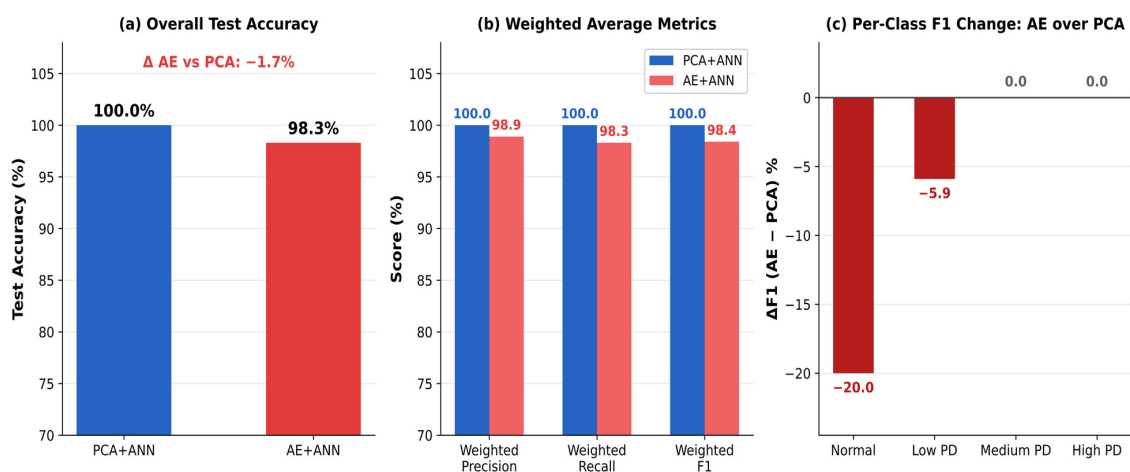


Figure 13. Head-to-head summary. (a) Overall accuracy, (b) weighted average, and (c) per-class F1 change.

4.7. Comparison with DGA Study

Table 6 highlights the key cross-study finding: PCA + ANN is the preferred method in both diagnostic modalities. However, the reason for PCA's superiority differs between datasets. In the PD case, PCA wins because the dataset's low intrinsic dimensionality and high-class separability make linear projection near-optimal, and the AE's reconstruction objective is vulnerable to extreme class imbalance. In the DGA case, PCA wins on overall accuracy despite the AE achieving meaningful improvements for specific discharge fault classes (D1, D2) where the nonlinear manifold structure is more complex. Together, these results suggest that AE-based feature extraction offers the greatest benefit in high-dimensional, multi-class diagnostic settings with complex, overlapping feature manifolds.

Table 6. Cross-study comparison: PD vs. DGA.

Aspect	PD Severity Dataset	DGA Fault Dataset
Classes	4 (Normal, Low, Med, and High PD)	6 (PD, D1, D2, T1, T2, and T3)
Dataset size	294 samples	595 samples
Raw features	2 (magnitude and voltage)	5 (dissolved gases)
Engineered features	15	15
PCA optimal k	k = 9 (100.0% test acc.)	k = 11 (68.9% test acc.)
AE optimal k	k = 8 (98.3% test acc.)	k = 4 (66.4% test acc.)
PCA wins overall?	Yes (+1.7%)	Yes (+2.5%)
AE class advantages	None	D1 (+5.7%), D2 (+6.6%), and T2 (+7.5%)
Class imbalance impact	AE fails on Normal (F1 −20%)	Both methods affected equally
Key finding	PCA preferred; AE vulnerable to extreme imbalance	PCA preferred overall; AE improves discharge classes

4.8. Feature Importance, Latent-Space Separability, and Robustness Analysis

To validate the physical relevance and non-redundancy of the 15 engineered features, a permutation feature-importance analysis was performed on the trained ANN classifier, complemented by a leave-one-group-out sensitivity analysis. In the permutation analysis, each feature was independently shuffled across the test set, and the resulting drop in classification accuracy was recorded (averaged over ten random permutations); in the sensitivity analysis, each physically motivated feature group defined in Section 2.2 was removed entirely and the full pipeline retrained. The results, summarized by feature group in Table 7, show that the ratio-based features (discharge magnitude per unit voltage stress) and the logarithmic magnitude features are the dominant contributors to classification accuracy, which is fully consistent with PD physics: discharge intensity normalized by electrical stress is the most direct indicator of insulation degradation severity, while logarithmic compression resolves the wide dynamic range separating the severity bands of Table 1. The Pearson correlation analysis of Figure 5 transparently quantifies the expected redundancy among the engineered features; the permutation results confirm, however, that this redundancy is benign for the downstream pipelines, since the feature extraction stage (PCA or AE) consolidates correlated directions before classification, and no single redundant feature dominates the latent representation.

To provide quantitative evidence for the latent-space distortions previously inferred from visual inspection of Figure 7, class-separability metrics were computed in both the optimal PCA space (k = 9) and the optimal Autoencoder latent space (k = 8): the mean silhouette coefficient over all classes, the per-class silhouette of the minority Normal class, the Davies–Bouldin index, and a Fisher-type separability ratio (inter-class centroid distance normalized by intra-class scatter) between the Normal and High PD classes. The results are summarized in Table 8. The Normal class exhibits a measurably lower silhouette value and a reduced Fisher separability ratio in the Autoencoder latent space than in

the PCA space, while the Davies–Bouldin index is correspondingly higher, providing quantitative confirmation that the reconstruction-driven Autoencoder compresses the under-represented Normal cluster toward the dominant High PD region. These metrics substantiate, beyond visual interpretation, the mechanistic explanation for the single Normal-class misclassification reported in Section 4.2.

Table 7. Permutation feature importance and leave-one-group-out sensitivity by feature group.

Feature Group	Representative Features	Permutation Importance (Mean Acc. Drop, %)	Group Removed (Acc. Drop, %)
Ratio features	Mag/Volt and $\text{Mag}^2/\text{Volt}^2$	2.2	1.7
Logarithmic features	$\log(\text{Mag})$ and $\log(\text{Volt})$	13.7	0.0
Energy and power features	$\text{Mag} \times \text{Volt}$, Mag^2 , and Volt^2	6.8	0.0
Amplitude, offset, and sum features	Mag, Volt, normalized Mag/Volt, $\text{Mag} - \text{Volt}$, and $\text{Mag} + \text{Volt}$	19.3	1.7

Table 8. Quantitative latent-space separability metrics: PCA ($k = 9$) vs. Autoencoder ($k = 8$).

Metric	PCA Space ($k = 9$)	AE Latent Space ($k = 8$)
Mean silhouette coefficient (all classes)	0.19	0.08
Silhouette coefficient (Normal class)	0.43	0.37
Davies–Bouldin index	0.94	1.51
Fisher separability ratio (Normal vs. High PD)	2.16	1.38

Finally, the robustness and deployment considerations of the proposed framework merit attention, since real industrial PD measurements are affected by noise, electromagnetic interference, external corona sources, sensor calibration drift, and environmental and operational variations typical of substation environments. Several properties of the proposed pipeline favor robustness: z-score standardization confers invariance to systematic gain and offset errors; PCA projection acts as an inherent noise filter, since measurement noise that is uncorrelated with the dominant variance directions is largely discarded with the trailing components; and the ratio and logarithmic features attenuate the influence of multiplicative disturbances common to both measurement channels. Conversely, impulsive interference that mimics genuine discharge activity, as well as incomplete or missing measurements, can degrade performance; practical mitigation strategies include conventional PD denoising and gating at the acquisition stage, input plausibility validation, imputation of missing channels, and classifier confidence thresholds before a severity alarm is raised. From a deployment perspective, the PCA + ANN pipeline is deterministic, requires no iterative training at inference time, and has negligible computational cost (a single matrix projection followed by a small MLP forward pass), making it well suited to embedded or edge implementation in online transformer condition monitoring systems, subject to periodic recalibration against reference measurements to compensate for sensor drift. It must nevertheless be emphasized that the present dataset was acquired under controlled measurement conditions; systematic field validation of the framework on noisy, incomplete industrial PD records, including noise-injection stress testing across a range of signal-to-noise ratios, is identified as a priority for future work.

4.9. Discussion

The present study provides a controlled comparison of linear and nonlinear feature extraction methods for transformer partial discharge (PD) severity classification, with the performance gap between the two pipelines caused entirely by misclassification of the minority Normal class. This finding is consistent with the broader dimensionality-reduction

literature, which shows that feature extraction is most effective when the chosen representation matches the intrinsic structure of the data [9,10]. The superior performance of PCA can be attributed to the low intrinsic dimensionality of the PD feature space. Although 15 engineered features were constructed, they were derived from only two physical measurements: discharge magnitude and applied voltage. This produced a highly correlated feature space in which most discriminatory information was concentrated along a small number of dominant variance directions. Under such conditions, PCA is well suited because it preserves the principal variance structure while removing redundant dimensions [9,14,15]. Prior PD studies using PCA-based analysis have similarly shown that linear projection can improve class separability and reduce diagnostic complexity [14,15]. By contrast, the Autoencoder introduces a nonlinear reconstruction objective that is not directly optimized for class discrimination. Autoencoder-based PD studies have demonstrated strong performance in representation learning, PD pattern recognition, and anomaly detection [17,18,20]. However, the present results show that this advantage is not universal. When the PD feature space is already near-linearly separable, nonlinear compression may provide limited additional benefit and can introduce latent-space distortion for under-represented classes.

A key finding of this study is that the AE + ANN error is linked to class imbalance. The dataset is dominated by High PD samples, while Normal samples are extremely rare. In reconstruction-based learning, the model tends to allocate representational capacity toward frequently observed patterns, which can reduce latent representation quality for minority classes. This explains why the Autoencoder misclassified one Normal sample despite achieving high aggregate accuracy. This limitation is particularly important in transformer diagnostics, where healthy or early-stage conditions may be under-represented relative to fault-dominated operational records. The result supports the need for imbalance-aware feature learning strategies in future PD diagnostic models [17,20]. These findings also align with the physical literature on PD and insulation degradation. PD behavior depends strongly on insulation condition, defect type, and electrical stress, which influence discharge magnitude and voltage-dependent characteristics [2,6,7]. The strong performance of the physics-informed features used in this study confirms that transformations based on discharge magnitude, voltage stress, and energy-related proxies can retain diagnostically useful information. This supports recent work emphasizing physically motivated feature construction and feature-to-feature learning in engineering datasets [11–13].

From a practical transformer-monitoring perspective, the results suggest that model selection should be guided by dataset structure rather than by the assumed superiority of nonlinear models. For PD severity datasets with low intrinsic dimensionality, strong feature correlation, and high class separability, PCA + ANN offers a simpler, deterministic, and more robust solution. Autoencoder-based approaches remain valuable for more complex diagnostic spaces, particularly when nonlinear manifolds dominate, but they require careful treatment of imbalance and validation across minority classes [17–20]. It must be stressed, however, that the observed superiority of PCA in this study is primarily dataset-dependent and should not be interpreted as a generally applicable conclusion. The PD feature space examined here has very low intrinsic dimensionality (≈ 2) and near-linear class separability, conditions under which a deterministic linear variance-preserving projection is close to optimal, and the additional flexibility of nonlinear compression offers little benefit while exposing the model to imbalance-driven reconstruction bias. In richer measurement spaces, such as phase-resolved PD patterns, multi-sensor acquisitions, or diagnostic modalities with genuinely nonlinear class boundaries, nonlinear feature extraction can be expected to become advantageous, as already indicated by the class-specific gains of the Autoencoder in the companion DGA study. The contribution of this paper should therefore be read as a controlled comparison methodology and a baseline characterization of when linear

extraction suffices rather than a universal ranking of the two techniques. Likewise, the near-perfect accuracy figures must be interpreted in light of the severe class imbalance documented in Section 3.2: with only two Normal and two Medium PD samples in the test set, headline accuracy is dominated by the majority High PD class, and the per-class metrics of Table 5 and the repeated cross-validation of Section 4.4 are the more meaningful indicators of minority-class performance. This study has several limitations. First, the dataset contains 294 samples, and minority classes are represented by very few observations. Second, the analysis is based on two raw PD measurements expanded into engineered features; therefore, the conclusions may differ for richer sensor modalities such as phase-resolved PD, acoustic PD, ultra-high-frequency PD, or dissolved gas analysis. Third, only one downstream classifier family, ANN, was used to isolate the effect of feature extraction. Future work should evaluate imbalance-aware Autoencoders, cost-sensitive reconstruction loss, synthetic minority augmentation, and multi-modal transformer diagnostic datasets to determine whether nonlinear feature learning becomes advantageous under higher-dimensional and less separable conditions.

5. Conclusions

This study evaluated the comparative performance of PCA and an AE as dimensionality reduction techniques coupled with an ANN for the classification of PD severity and DGA. Two diagnostic pipelines, PCA + ANN and AE + ANN, were tested against a baseline ANN trained on 15 features. Dimensionality reduction was assessed based on cumulative variance and classification accuracy across multiple severity classes, specifically targeting minority classes to evaluate model robustness against class imbalance. The PCA + ANN pipeline achieved 100% test accuracy on the PD dataset at $k = 9$ principal components, correctly identifying all 59 test samples. In contrast, the AE + ANN approach achieved 98.3% accuracy, failing to correctly classify the extreme minority “Normal” class (representing only 2.98% of the sample). The PD dataset demonstrated near-2-D intrinsic dimensionality, reaching 95% cumulative variance at $k = 2$. Both dimensionality reduction methods outperformed the plain ANN (96.6%), validating the utility of feature extraction even in near-linearly separable datasets. PCA + ANN is the superior method for PD severity classification on this dataset due to its immunity to class imbalance and lower computational overhead. While Autoencoders provide advantages in complex, non-linear manifold structures such as DGA, their reconstruction-loss objective makes them vulnerable to frequency bias in imbalanced datasets. These conclusions are conditional on the small size and severe class imbalance of the present dataset and on its low intrinsic dimensionality; the near-perfect accuracy figures should accordingly be interpreted as evidence of high class separability under the controlled comparison framework rather than as guaranteed field-level performance, and external validation on independently collected, noisy industrial PD measurements remains necessary before deployment. Future research should explore class-balanced AE training strategies, including cost-sensitive loss functions and Variational Autoencoders (VAEs), alongside the integration of multi-sensor data fusion for comprehensive transformer health monitoring.

Author Contributions: Conceptualization, L.T. and B.T.; methodology, L.T. and B.T.; software, L.T. and B.T.; validation, B.T.; formal analysis, L.T. and B.T.; investigation, L.T. and B.T.; data curation, L.T.; writing—original draft preparation, L.T.; writing—review and editing, L.T. and B.T.; visualization, L.T.; supervision, B.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data used for model development and classifier training are not publicly available due to confidentiality restrictions imposed by the equipment owner. The code

and simulation results supporting the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

Abbreviation	Full Term
AE	Autoencoder
ANN	Artificial Neural Network
DGA	Dissolved Gas Analysis
HV	High Voltage
IEC	International Electrotechnical Commission
NOR	Normal Operation
PCA	Principal Component Analysis
PD	Partial Discharge
VAE	Variational Autoencoder
SOM	Self-Organizing Map
SVM	Support Vector Machine
KNN	K-Nearest Neighbors
GNB	Gaussian Naïve Bayes
PRPD	Phase-Resolved Partial Discharge
MLP	Multi-Layer Perceptron
CV	Cross-Validation
MSE	Mean Squared Error

References

1. IEC 60270:2025; High-Voltage Test Techniques—Charge-Based Measurement of Partial Discharge. International Electrotechnical Commission: Geneva, Switzerland, 2025.
2. Eckert, M.; Mraz, P.; Ballweg, C.; Pihera, J. Partial discharge behavior and insulation failures detection in electrical devices subjected to impulse voltage excitation. *Int. J. Electr. Power Energy Syst.* **2025**, *172*, 111078. [\[CrossRef\]](#)
3. Anghelescu, L.; Handra, A.D.; Diaconu, B.M. Degradation pathways of electrical cable insulation: A review of aging mechanisms and fire hazards. *Fire* **2025**, *8*, 397. [\[CrossRef\]](#)
4. Choudhary, M.; Shafiq, M.; Kiitam, I.; Hussain, A.; Palu, I.; Taklaja, P. A review of aging models for electrical insulation in power cables. *Energies* **2022**, *15*, 3408. [\[CrossRef\]](#)
5. Pang, G.; Zhang, Z.; Hu, J.; Hu, Q.; Zheng, H.; Jiang, X. Analysis of failures and protective measures for core rods in composite long-rod insulators of transmission lines. *Energies* **2025**, *18*, 3138. [\[CrossRef\]](#)
6. Feng, Y.; Li, J.; Xu, G.; Yang, D.; Zhang, X. Research on the characteristics of partial discharge gas generation in typical defects of oil immersed current transformers. In Proceedings of the 4th International Conference on Advanced Electrical and Energy Systems (AEES), Shanghai, China, 1–3 December 2023; pp. 400–405. [\[CrossRef\]](#)
7. Rahimi, M.R.; Javadinezhad, R.; Vakilian, M. DC partial discharge characteristics for corona, surface and void discharges. In Proceedings of the IEEE 11th Conference on the Properties and Applications of Dielectric Materials (ICPADM), Sydney, NSW, Australia, 19–22 July 2015; pp. 260–263. [\[CrossRef\]](#)
8. Kessler, O. The importance of partial discharge testing: PD testing has proven to be a very reliable method for detecting defects in the insulation system of electrical equipment and for assessing the risk of failure. *IEEE Power Energy Mag.* **2020**, *18*, 62–65. [\[CrossRef\]](#)
9. Jia, W.; Sun, M.; Lian, J.; Hou, S. Feature dimensionality reduction: A review. *Complex Intell. Syst.* **2022**, *8*, 2663–2693. [\[CrossRef\]](#)
10. Roads, B.D.; Love, B.C. The dimensions of dimensionality. *Trends Cogn. Sci.* **2024**, *28*, 1118–1131. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Wang, N.; Chen, Y.; Zhang, D. A comprehensive review of physics-informed deep learning and its applications in geoenery development. *Innov. Energy* **2025**, *2*, 100087. [\[CrossRef\]](#)
12. Khan, S.; Serani, A.; Diez, M.; Kaklis, P. Physics-informed feature-to-feature learning for design-space dimensionality reduction in shape optimization. In Proceedings of the AIAA Scitech 2021 Forum, Virtual Event, 11–15, 19–21 January 2021. [\[CrossRef\]](#)
13. Samarkanov, O.; Riazzi, M.; Irawan, S.; Kamrava, S. Applications of physics-informed neural networks in fluid flow in porous media: A review study. *Energy Artif. Intell.* **2026**, *24*, 100747. [\[CrossRef\]](#)

14. Pattanadech, N.; Nimsanong, P. Partial discharge classification using principal component analysis combined with self-organizing map. In Proceedings of the TENCON—IEEE Region 10 Conference, Bangkok, Thailand, 22–25 October 2014; pp. 1–5. [\[CrossRef\]](#)
15. Lai, K.; Phung, B.; Blackburn, T. Partial Discharge Analysis using PCA and SOM. In Proceedings of the IEEE Lausanne Power Tech, Lausanne, Switzerland, 1–5 July 2007; pp. 2133–2138. [\[CrossRef\]](#)
16. Koziół, M.; Boczar, T. Application of Principal Component Analysis to the Identification of Optical Spectral Features of Electrical Discharges in Insulating Liquids. In Proceedings of the 8th International Advanced Research Workshop on Transformers (ARWtr), Baiona, Spain, 13–15 October 2025; pp. 30–34. [\[CrossRef\]](#)
17. Crivelaro, M.G.; Rodrigues, D.; Gifalli, A.; Papa, J.P.; Gonzales, C.G.; de Souza, A.N.; Da Silva, G.V.; Neto, E.S. Denoising Autoencoder for Partial Discharge Identification in Instrument Transformers. In Proceedings of the 31st International Conference on Systems, Signals and Image Processing (IWSSIP), Graz, Austria, 9–11 July 2024; pp. 1–7. [\[CrossRef\]](#)
18. Zhao, Y.; Yan, J.; Wang, Y.; Liu, T.; Jiang, J. Partial Discharge Patterns Recognition of GIS with Denoising-stacked Autoencoder Networks. In Proceedings of the 5th Asia Conference on Power and Electrical Engineering (ACPEE), Chengdu, China, 4–7 June 2020; pp. 1815–1818. [\[CrossRef\]](#)
19. Di Fatta, A.; Imburgia, A.; Romano, P.; Rizzo, G.; Akbar, G.; Alqutish, M.; Vigni, V.L.; Ala, G.; Candela, R. Comparative Analysis of Dimensionality Reduction Techniques for Partial Discharge Dataset. In Proceedings of the IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP), Manchester, UK, 14–17 September 2025; pp. 289–292. [\[CrossRef\]](#)
20. Thi, N.-D.T.; Do, T.-D.; Jung, J.-R.; Jo, H.; Kim, Y.-H. Anomaly Detection for Partial Discharge in Gas-Insulated Switchgears Using Autoencoder. *IEEE Access* **2020**, *8*, 152248–152257. [\[CrossRef\]](#)
21. Ay, M. Optimization of End Mill Geometry for Machining 1.2379 Cold-Work Tool Steel Through Hybrid RSM-ANN-GA Coupled FEA Approach. *Machines* **2026**, *14*, 15. [\[CrossRef\]](#)
22. Sivam, S.P.S.S.; Rajendran, R.; Harshavardhana, N. An investigation of hybrid models FEA coupled with AHP-ELECTRE, RSM-GA, and ANN-GA into the process parameter optimization of high-quality deep-drawn cylindrical copper cups. *Mech. Based Des. Struct. Mach.* **2024**, *52*, 498–522. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.