FOTOVOLTAİK ARIZA TESPİTİ İÇİN DOĞRUSAL VE DOĞRUSAL OLMAYAN PCA, LDA VE ICA YÖNTEMLERİNİN KARŞILAŞTIRMALI DEĞERLENDİRMESİ

Elektrik-Elektronik Müh., Gökhan KARACA[1]/ Doç. Dr., Ü. Çiğdem TURHAL[2]/ Doç. Dr., Yasemin ÖNAL[3]

[1]Eti Makine Sanayi ve Ticaret A.Ş. (BSEU Lisansüstü Eğitim Enstitüsü), Mühendis, 1966682@ogrenci.bilecik.edu.tr, https://orcid.org/0009-0004-4072-2558

[2]Bilecik Şeyh Edebali Üniversitesi (BSEU) Müh. Fak. Elektrik Elektronik Mühendisliği Bölümü, Öğretim Üyesi, ucigdem.turhal@bilecik.edu.tr , https://orcid.org/0000-0003-2387-1637

[3]Bilecik Şeyh Edebali Üniversitesi (BSEU) Müh. Fak. Elektrik Elektronik Mühendisliği Bölümü, Öğretim Üyesi, yasemin.onal@bilecik.edu.tr, https://orcid.org/0000-0003-0173-0948

Özet

Fotovoltaik (FV) sistemlerdeki arızaların doğru tespiti, operasyonel güvenilirliği artırmak, enerji verimini en üst düzeye çıkarmak ve bakım maliyetlerini en aza indirmek için esastır. Bu çalışma, akıllı FV arıza sınıflandırması için ön işleme yöntemleri olarak hem doğrusal hem de doğrusal olmayan boyut azaltma tekniklerini (Temel Bileşen Analizi (PCA), Doğrusal Ayırımcı Analiz (LDA), Bağımsız Bileşen Analizi (ICA), Çekirdek PCA (KPCA), Çekirdek LDA (KLDA) ve Çekirdek ICA) değerlendiren birleşik bir karşılaştırmalı çerçeve sunmaktadır. Bu teknikler, FV elektrik sinyallerinin tamamlayıcı istatistiksel özelliklerini yakalar: doğrusal PCA küresel varyansı modeller, LDA denetimli sınıf ayrılabilirliğini optimize eder ve ICA gizli arıza örüntülerine karşılık gelebilecek istatistiksel olarak bağımsız modları çıkarır. Doğrusal olmayan çekirdek karşılıkları, bu yetenekleri gerçek FV çalışma koşullarında karşılaşılan karmaşık, doğrusal olmayan şekilde ayrılabilir özellik yapılarına genişletir. Ön işlemeden sonra, her yöntemin aynı deneysel ayarlar altında ayırt edici etkinliğini değerlendirmek üzere bir Destek Vektör Makinesi (SVM) sınıflandırıcısı eğitilir. Normal davranış, kısmi gölgelenme, bozulma ve elektriksel arızaları içeren gerçek operasyonel veri kümeleri, tüm yaklaşımların kıyaslanması için kullanılmıştır. Performans, doğruluk, F1 puanı, karışıklık matrisleri ve hesaplama karmaşıklığı kullanılarak karşılaştırılmıştır. Ön sonuçlar, PCA'nın sürekli olarak en yüksek genel doğruluk ve kararlılığı sağladığını, KLDA'nın ise kıvrımlar arasında düşük varyansla güçlü bir denetimli sınıf ayrımı sunduğunu göstermektedir. KPCA, verilerdeki doğrusal olmayan yapıları yakalayarak rekabetçi bir performans sergilemektedir, ancak biraz daha fazla değişkenliğe sahiptir. Buna karşılık, ICA ve Kernel-ICA sınırlı bir ayrım yeteneği ve gürültüye karşı yüksek hassasiyet göstermektedir, bu da onları PV arıza sınıflandırması için daha az uygun hale getirmektedir. Genel olarak, bulgular her bir ön işleme yönteminin göreceli güçlü yönlerini açıkça ortaya koymakta ve yeni nesil PV izleme ve arıza teşhis sistemlerinde etkili boyut azaltma yöntemlerinin seçilmesi için pratik rehberlik sağlamaktadır.

Anahtar kelimeler: Doğrusal olmayan özellik çıkarımı, Sinyal işleme, Akıllı izleme sistemleri

COMPARATIVE EVALUATION OF LINEAR AND NONLINEAR PCA, LDA, AND ICA METHODS FOR PHOTOVOLTAIC FAULT DETECTION

Abstract

Accurate detection of faults in photovoltaic (PV) systems is essential for enhancing operational reliability, maximizing energy yield, and minimizing maintenance costs. This study provides a unified comparative framework that evaluates both linear and nonlinear dimensionality reduction techniques—Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA), Kernel PCA (KPCA), Kernel LDA (KLDA), and Kernel ICA—as preprocessing methods for intelligent PV fault classification. These techniques capture complementary statistical properties of PV electrical signals: linear PCA models global variance, LDA optimizes supervised class separability, and ICA extracts statistically independent modes that may correspond to hidden fault patterns. Their nonlinear kernel counterparts extend these capabilities to complex, nonlinearly separable feature structures encountered in real PV operating conditions. After preprocessing, a Support Vector Machine (SVM) classifier is trained to assess the discriminative effectiveness of each method under identical experimental settings. Real operational datasets containing normal behavior, partial shading, degradation, and electrical faults are used to benchmark all approaches. Performance is compared using accuracy, F1-score, confusion matrices, and computational complexity. Preliminary results show that PCA consistently yields the highest overall accuracy and stability, while KLDA offers strong supervised class separation with low variance across folds. KPCA demonstrates competitive performance by capturing nonlinear structures in the data, though with slightly more variability. In contrast, ICA and Kernel-ICA show limited discriminative ability and high sensitivity to noise, making them less suitable for PV fault classification. Overall, the findings clearly outline the relative strengths of each preprocessing method and provide practical guidance for choosing effective dimensionality reduction methods in next-generation PV monitoring and fault-diagnosis systems.

Keywords: Nonlinear feature extraction, Signal processing, Intelligent monitoring systems

1. INTRODUCTION

Photovoltaic (PV) energy has rapidly become one of the fastest-growing renewable technologies, with global installations surpassing 1.2 terawatts (TW) as of 2023. This accelerated growth is driven by decreasing module costs, improvements in conversion efficiency, and supportive policy frameworks (IEA, 2023). As PV plants expand in capacity and structural complexity, the demand for accurate and reliable fault detection systems has become increasingly critical. Operational anomalies—such as partial shading, dirt accumulation (soiling), module degradation, arc faults, inverter failures, and electrical mismatches—can significantly reduce energy yield and threaten both system safety and grid stability (Prasanna et al., 2022; Gawre, 2022). Consequently, intelligent, data-driven approaches for automated PV fault diagnosis have emerged as a pivotal research area.

Modern PV systems generate vast amounts of high-dimensional, nonlinear, and often noisy time-series data. Traditional linear feature extraction methods frequently struggle to capture the nonlinear dynamics inherent in PV operation, especially under fault conditions (Yao et al., 2021). This limitation has motivated the extensive adoption of kernel-based nonlinear dimensionality reduction techniques, which use kernel functions to map input data into high-dimensional feature spaces where nonlinear relationships become linearly separable.

Among these approaches, Kernel PCA (KPCA) has demonstrated strong capabilities in extracting nonlinear variance components and improving diagnostic robustness under challenging conditions such as partial shading and soiling (Zhang et al., 2021). Kernel LDA (KLDA), as a supervised variant, enhances class separability by maximizing between-class variance, making it particularly suitable for labeled PV fault datasets (Zhang et al., 2021). Kernel ICA, by contrast, isolates statistically independent components, a property advantageous for identifying independent electrical fault signatures that may not align with traditional variance- or discriminant-based projections (Hyvärinen & Oja, 2000).

Despite the growing interest in these kernel techniques, comparative studies evaluating KPCA, KLDA, and Kernel ICA within a unified PV-fault classification framework remain limited. Recent reviews highlight that kernel-based preprocessing can substantially improve the performance of downstream classifiers such as SVMs, yet the magnitude of improvement is strongly dependent on the characteristics of the PV dataset and the type of fault under analysis (Et-Taleby et al., 2023; Dhibi et al., 2020). SVMs are widely regarded as one of the most reliable classifiers in PV diagnostics due to their strong generalization capability, robustness to limited training data, and ability to model complex nonlinear boundaries (Lu et al., 20204).

Motivated by this gap, the present study conducts a systematic comparative evaluation of KPCA, KLDA, and Kernel ICA as preprocessing steps for PV fault classification. After transforming features into kernel space, an SVM classifier is trained to differentiate between normal operation and multiple fault categories. Real-world PV monitoring datasets are employed, and the performance of each kernel method is assessed using accuracy, precision, recall, F1-score, confusion matrices, and computational efficiency. The findings offer practical insights into selecting the most effective kernel-based preprocessing strategy for next-generation intelligent PV diagnostic systems.

1.1. Related Work

Fault diagnosis in photovoltaic (PV) systems has been an active research area for more than a decade, driven by the need to ensure reliable operation, maximize energy yield, and mitigate safety risks. Early diagnostic strategies relied on current–voltage (I–V) curve analysis, signal-based indicators, and threshold-based detection schemes (Hajji et al., 2023). Although these approaches provided initial insights into system health, their performance deteriorated under nonlinear operating conditions and varying environmental factors. As PV datasets have become increasingly high-dimensional, noisy, and nonlinear, machine learning (ML) and kernel-based feature extraction techniques have gained prominence for capturing complex fault signatures.

Linear dimensionality reduction methods such as PCA, LDA, and ICA were among the earliest tools adopted in PV diagnostics. PCA has been used to detect deviations in voltage and current characteristics but is limited by its assumption of linear variance structure. LDA improves supervised class separability but struggles when class distributions overlap nonlinearly. ICA identifies statistically independent latent components and has shown potential in separating mixed electrical behaviors; however, its performance declines when independence assumptions are violated or noise levels are high.

To overcome these limitations, kernel-based nonlinear dimensionality reduction techniques have become increasingly popular. KPCA has demonstrated strong capabilities in modeling nonlinear variance in PV systems. Zhang et al. (2021) reported that KPCA outperforms linear PCA in detecting subtle anomaly patterns such as partial shading and early-stage degradation. Dhibi et al. (2021) further applied KPCA for inverter fault identification, showing improved sensitivity to weakly separable patterns. KLDA extends traditional LDA into a high-dimensional feature space, enhancing supervised class discrimination. Zhang et al. (2021) demonstrated its effectiveness in wind–PV hybrid energy systems, while Shi (2016) reported improved classification performance under noisy PV monitoring data and overlapping fault categories. Kernel ICA, although less frequently used in PV applications, remains a powerful tool when faults arise from mixtures of independent electrical sources. Hyvärinen and Oja (2000) established the theoretical foundation for ICA in nonlinear feature learning, and later studies applied Kernel ICA for anomaly detection in energy systems (Zhang et al., 2008). More recently, Qureshi et al. (2020) showed that Kernel ICA can isolate fault-specific independent components in solar irradiance and power quality signals, offering complementary diagnostic information to variance-based techniques.

In classification-based PV diagnostic systems, SVMs remain one of the most robust and widely adopted algorithms due to their strong generalization ability, capacity to model nonlinear decision boundaries, and effectiveness with limited training data (Lu et al., 2024). Numerous studies have paired SVMs with various feature extraction methods, including wavelet-transform features (Cai and Wai,., 2022), PCA, and deep-learning-based representations (Mansouri et al., 2021). However, despite the widespread use of kernel-based dimensionality reduction techniques, few studies have systematically compared KPCA, KLDA, and Kernel ICA under identical experimental conditions, particularly when integrated as preprocessing steps for SVM-based PV fault classification.

Recent reviews underscore the need for rigorous, unified evaluations of nonlinear preprocessing methods to understand their relative strengths across different PV operating conditions and fault scenarios (Et-Taleby et al., 20223; Harrou et al., 2021). This gap directly motivates the present study, which provides a comprehensive comparative analysis of KPCA, KLDA, and Kernel ICA for extracting discriminative features prior to SVM classification. By benchmarking these techniques on real-world PV datasets and assessing performance across multiple diagnostic metrics, this work offers valuable guidance for selecting effective kernel-based preprocessing strategies in next-generation intelligent PV monitoring systems.

2. METHODOLOGY

2.1. Dataset Description

This study generated the dataset through simulations conducted in MATLAB/Simulink, a well-established platform for modeling photovoltaic (PV) systems (Radhakrishnan et al., 2021; Ghoneim et al., 2021; Amiri et al., 2024). The core of the simulated system consisted of 7 series-connected and 88 parallel-connected PV panels, forming a 250 kW grid-connected PV plant (Ghoneim et al., 2021). To facilitate grid integration, a three-level inverter was used, employing a control strategy that implemented Maximum Power Point Tracking (MPPT) using the Perturb and Observe (P&O) method (Ghoneim et al., 2021). While other MPPT strategies, such as incremental conductance, are often mentioned in the literature, P&O was chosen for its simplicity and effectiveness (Radhakrishnan et al., 2021). Additionally, the system included a DC-DC boost converter to stabilize the voltage output (Radhakrishnan et al., 2021; Aziz et al., 2020).

The simulations encompassed four distinct scenarios to emulate realistic operating conditions: one healthy (fault-free) state and three prevalent fault types—string fault, grounding fault, and string-to-string fault (Ghoneim et al., 2021). The dataset is detailed in Table 1.

Table 1. Distribution of Dataset by Fault Class

Description	Record Percentage	Number of Samples	Class Label
No Fault	16.67%	100	0
String Fault	25.50%	153	1
Grounding Fault	24.83%	149	2
String-to-String Fault	33.00%	198	3

This study primarily concentrated on the three most prominent fault types; however, the literature also highlights other PV faults, including open circuits, short circuits, high-resistance connections, and partial shading, all of which can undermine system performance (Lazzaretti, et al.,2020).

The dataset comprises a total of 30 features that encompass both physical and electrical measurements essential for precise fault classification (Table 2). Among these features, four have been recognized as primary indicators of system status. These variables—current (I), voltage (V), power (P), irradiance (G), and temperature (T)—are frequently cited in the literature as reliable inputs for fault detection and classification models in photovoltaic (PV) systems (Ghoneim et al., 2021).

Table 2. Description of Physical and Electrical Measurements Used in the Dataset

Category	Feature	Description
Physical Measurements	Temperature (T)	Ranges from 10°C to 35°C; significantly affects PV panel efficiency (Lazzaretti, et al.,2020) .
	Irradiance (IR)	Varies from 100 W/m² to 1000 W/m²; directly impacts PV energy output (Lazzaretti, et al.,2020) .
	Fault Resistance	Simulated using random values between 1 Ω and 2000 Ω to represent different fault severities (Ghoneim et al., 2021).
Electrical Measurements	String Currents	Max/min average currents for three strings (I1A–I3A and I1B–I3B); includes inter-string current ranges (Range1–Range4) (Ghoneim et al., 2021).
	Total Current (ITOTAL)	Total average current of the PV array.
	Voltage (VDC)	Average DC voltage of the system.
	Power (PDC)	Average DC power; key metric for system efficiency and fault detection.

2.2. Linear Data Preprocessing Techniques

To enhance model generalization and computational efficiency by transforming the original 30-dimensional feature space into a more compact and informative form, this study applied and compared the three most common dimensionality reduction techniques. The methods examined include PCA, LDA, and ICA, compared in Table 3, all intended to improve classification performance and reduce computational load.

PCA is a linear transformation technique that is unsupervised and focuses on projecting data into a new space of uncorrelated components, known as principal components. This method retains the most significant variance in the dataset while eliminating redundancy and noise. The number of principal components is determined by aiming for a cumulative explained variance threshold of 95%. PCA effectively eliminates redundancy and noise, allowing classifiers to perform more efficiently.

Contrary to PCA, LDA focuses on maximizing the separability of the data among given classes of the data rather than treating the data as one class. It focuses on finding linear combinations of features that best separate the classes. Fewer classes that have Gaussian-distributed sample sets. There are two ways LDA was used in the stated study: LDA was used as a classifier on its own, and LDA was also used as a preliminary step to project the data into a lower-dimensional space to aid in the classification. As for ICA, it is an unsupervised approach that separates the given multivariate datasets into components that are independently statistically and not Gaussian and which differ from PCA, which is focused on maximizing the variance. In other words, ICA is trying to make the signals as independent as possible. Nonlinear mixing of signals is problematic. In the study, ICA was used to extract certain features and the generated components were used to implement some of the classical ML classifiers.

The following sections present a comparative analysis of PCA, LDA, and ICA, combined with various classifiers. The application of these dimensionality reduction methods showed improvements in computational efficiency and classification accuracy, varying by classifier and fault scenario.

Table 3. Comparison of Dimensionality Reduction Techniques

Aspect	PCA	LDA	ICA
Type	Unsupervised	Supervised	Unsupervised
Main Goal	Maximize total variance in the data	Maximize class separability	Maximize statistical independence among components
Feature Output	Principal components (orthogonal directions of maximum variance)	Linear combinations that best separate classes	Statistically independent source signals
Data Assumptions	Linear relationships, Gaussian-like distribution	Normally distributed classes with equal covariance	Non-Gaussian, statistically independent sources
Interpretability	Moderate	High (components aligned with class separation)	Low (components are often abstract and hard to interpret)
Noise Sensitivity	Relatively low	Moderate	High
Computational Efficiency	High	High (with fewer classes)	Moderate to Low (depends on implementation and convergence)
Limitation	May discard class-discriminative features	Limited to number of classes − 1 dimensions	Sensitive to scaling and initial conditions

2.3. Nonlinear Data Preprocessing Techniques

Consider the PV dataset:

$$X = [x_1, x_2, \dots, x_N]^{\top} \in \mathbb{R}^{Nxd} \tag{1}$$

where each sample x_i corresponds to PV operational measurements (voltage, current, irradiance, temperature, power, etc.). A nonlinear mapping function

$$\phi: \mathbb{R}^d \to \mathcal{F} \tag{2}$$

projects the original data into a high-dimensional feature space $\mathcal{F}$. The inner product in F is computed by a kernel:

$$\mathrm{K}(\mathrm{x_i}, \mathrm{x_j}) = \langle \phi(\mathrm{x_i}), \phi(\mathrm{x_j}) \rangle \tag{3}$$

Common kernels include the RBF kernel:

$$K(\mathrm{x_i}, \mathrm{x_j}) = \exp\left(-\frac{\|\mathrm{x_i} - \mathrm{x_j}\|^2}{2\sigma^2}\right) \tag{4}$$

Kernel Principal Component Analysis (KPCA)

KPCA finds eigenvectors of the covariance matrix in $\mathcal{F}$ without explicitly computing $\phi(x)$. The centered kernel matrix is

$$\widetilde{K} = K - 1K - K1 + 1K1 \tag{5}$$

The dual eigenvalue problem is

$$\widetilde{K}\alpha = \lambda\alpha \tag{6}$$

Projected KPCA features are computed as

$$\mathrm{y_i} = \sum_{\mathrm{j}=1}^{\mathrm{N}} \alpha_{\mathrm{j}} \mathrm{K}(\mathrm{x_i}, \mathrm{x_j}) \tag{7}$$

KPCA captures nonlinear variance structures useful for detecting shading, soiling, and degradation patterns.

Kernel Linear Discriminant Analysis (KLDA)

KLDA maximizes class separability in kernel space. Let the between-class and within-class scatter in $\mathcal{F}$ be:

$$S_b^{\emptyset} = \sum_{c=1}^{C} n_c \left(\mu_c^{\emptyset} - \mu^{\emptyset}\right)\left(\mu_c^{\emptyset} - \mu^{\emptyset}\right)^T \tag{8}$$

$$S_w^{\emptyset} = \sum_{c=1}^{C} \sum_{x_i \in c} \left(\emptyset(x_i) - \mu_c^{\emptyset}\right)\left(\emptyset(x_i) - \mu_c^{\emptyset}\right)^T \tag{9}$$

KLDA solves the generalized eigenvalue problem:

$$S_b^{\emptyset}\alpha = \lambda S_w^{\emptyset}\alpha \tag{10}$$

In kernel form, this becomes:

$$(KMK + \epsilon I)\alpha = \lambda(KNK)\alpha \tag{11}$$

where M and N encode class statistics.

KLDA is especially effective for supervised PV fault datasets with clear class boundaries.

Kernel Independent Component Analysis (Kernel ICA)

Kernel ICA seeks statistically independent components in kernel space. Let:

$$Z = \Phi(X) = [\phi(x_1), \dots, \phi(x_N)] \tag{12}$$

KICA finds a weight vector www maximizing non-Gaussianity (e.g., negentropy):

$$J(w^\top Z) = H(v) - H(w^\top Z) \tag{13}$$

where H(·)

The kernelized version uses Gram matrices:

$$w = \sum_{i=1}^{N} \beta_i \, \phi(x_i) \tag{14}$$

leading to a fixed-point iteration:

$$\beta \leftarrow \mathbb{E}[Zg(w^\top Z)] - \mathbb{E}[g'(w^\top Z)]w \tag{15}$$

with whitening using kernel PCA.

KICA is powerful for isolating independent PV fault signatures, such as inverter oscillations, partial shading modes, or sensor-specific anomalies.

2.4. SVM-Based PV Fault Classification

After kernel preprocessing, features f_i are fed into an SVM classifier. Given labeled samples (f_i, y_i), the SVM optimization problem is:

$$min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{N} \text{ξi} \tag{16}$$

subject to

$$y_i(w^T \emptyset(f_i) + b) \geq 1 - \text{ξi}, \;\; \text{ξi} \geq 0 \tag{17}$$

The dual form uses kernel functions:

$$max \left(\sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j K\left(f_i, f_j\right) \right) \tag{18}$$

$$s.t. 0 \leq \alpha_i \leq C, \; \sum_i \alpha_i y_i = 0 \tag{19}$$

The decision function is:

$$\widehat{y}(x) = sign(\sum_{i=1}^{N} \alpha_i y_i \, K(f_i, x) + b) \tag{20}$$

SVM is chosen due to its robustness, strong generalization capability, and suitability for nonlinear PV fault patterns.

2.5. Performance Evaluation

Each preprocessing method (KPCA, KLDA, KICA) is evaluated using the same SVM classifier and same dataset. Performance metrics include:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{21}$$

$$F1 - score = \frac{2.Precision.Recall}{Precision+Recall} \tag{22}$$

and computational cost (training time, feature extraction time). Comparisons reveal which kernel method offers the most discriminative features for PV fault classification.

3. EXPERIMENTAL STUDIES, RESULTS AND DISCUSSION

This study examines fault classification in photovoltaic (PV) systems by systematically evaluating how various preprocessing and dimensionality reduction techniques—both linear and nonlinear—affect the performance of SVMs. By utilizing a combined dataset of training and test measurements, six preprocessing strategies (PCA, LDA, ICA, Kernel-ICA, KPCA, and KLDA) were assessed through a 10-fold cross-validation approach, with all implementations performed in Python. The experimental pipeline was developed using Scikit-learn, allowing for standardized scaling, feature transformation, and SVM-based classification.

This study focuses on isolating the impact of preprocessing by keeping the classifier constant. This design facilitates a clear evaluation of how each transformation—variance-based (PCA), supervised (LDA, KLDA), independence-based (ICA), and kernel-based nonlinear mappings (KPCA, Kernel-ICA)—influences decision boundary formation in SVM. The results reveal significant performance variation among the preprocessors, highlighting PCA and KLDA as the most effective feature extraction methods for PV fault diagnosis, while ICA and Kernel-ICA show limited discriminative capability.

The comparative evaluation of linear and nonlinear preprocessing techniques offers valuable insights into the behavior of fault signatures in PV systems. As shown in Fig.1, PCA achieved the highest mean accuracy of 0.71 among all preprocessing methods. It was closely followed by KLDA at 0.66 and LDA at 0.67. These results suggest that, despite the nonlinear characteristics typically associated with PV faults, linear transformations can effectively create well-separated feature subspaces—especially when the dataset exhibits strong global variance patterns, as captured by PCA. In contrast, both ICA and Kernel-ICA demonstrated significantly lower accuracies, approximately 0.41.

This trend is further illustrated in the performance ranking shown in Fig.2. The reduced accuracy may result from the sensitivity of ICA-based methods to noise and the necessity for statistical independence, a condition that is often not met in real-world PV electrical data where fault effects frequently overlap. While Kernel-ICA has the theoretical capacity to model nonlinear dependencies, it did not provide a significant performance improvement over ICA, indicating that the selected kernel or component count may not effectively capture the underlying nonlinear structure of the dataset.

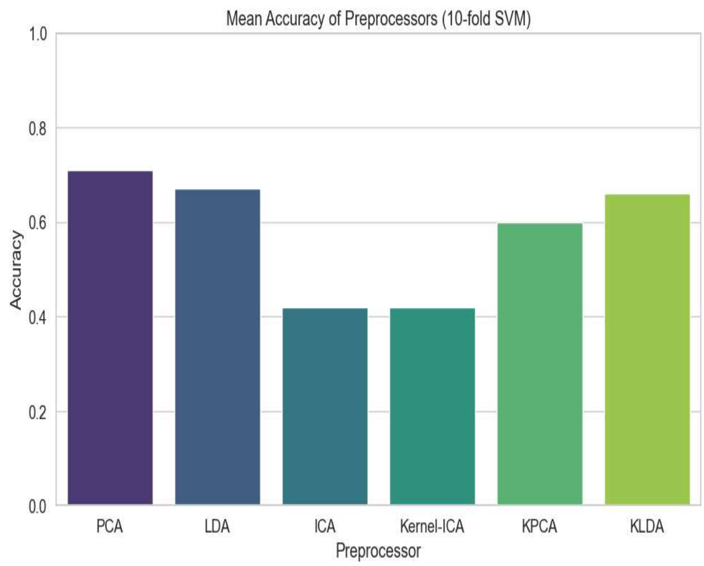

Fig.1 Mean accuracy values of all preprocessing methods evaluated using 10-fold SVM classification.

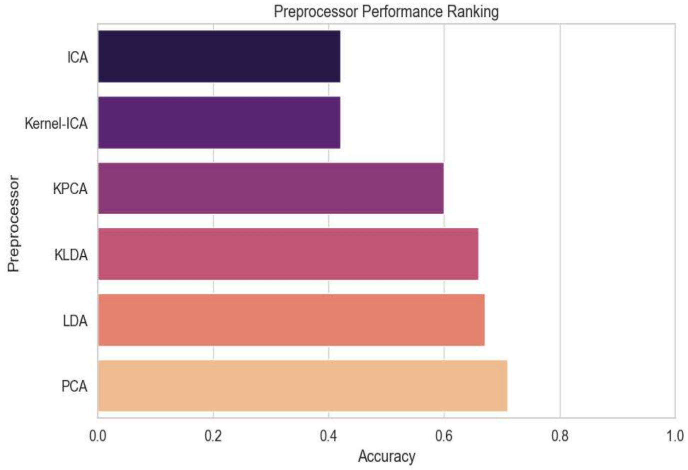

Fig. 2 Rank ordering of preprocessors based on average classification accuracy.

A broader analysis comparing linear and nonlinear families, presented in Fig.3, reveals that nonlinear preprocessing (mean accuracy approximately 0.63) shows a noticeable improvement over linear methods (mean accuracy approximately 0.56). This difference supports the hypothesis that nonlinear variations—such as fluctuations in irradiance, transitions due to partial shading, and distortions caused by inverters—play a crucial role in shaping fault-related patterns within PV systems. KPCA, highlighted in Figs. 1 and 2, outperformed many linear methods by effectively capturing these nonlinear relationships, achieving an average accuracy of 0.60. This demonstrates that expanding the feature space through the Radial Basis Function (RBF) kernel enhances SVM performance by helping to create more discriminative hyperplanes.

The performance heatmap in Fig.4 further emphasizes the multidimensional behavior of the preprocessing methods. PCA consistently produces high precision, recall, and F1 scores, confirming its robustness across various evaluation metrics. KLDA also shows balanced performance, benefiting from supervised dimensionality reduction that aligns feature projections with class boundaries. On the other hand, ICA-based methods exhibit low recall and F1 scores, reinforcing the conclusion that independence-based decomposition is not well-suited for PV fault structures.

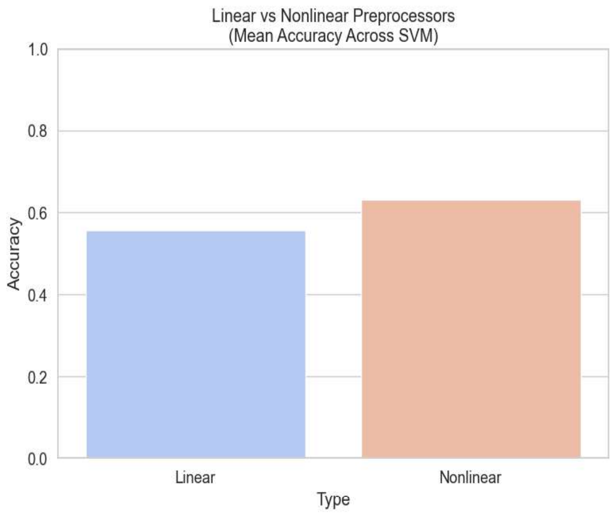

Fig.3 Comparison of mean accuracy between linear and nonlinear preprocessing families.

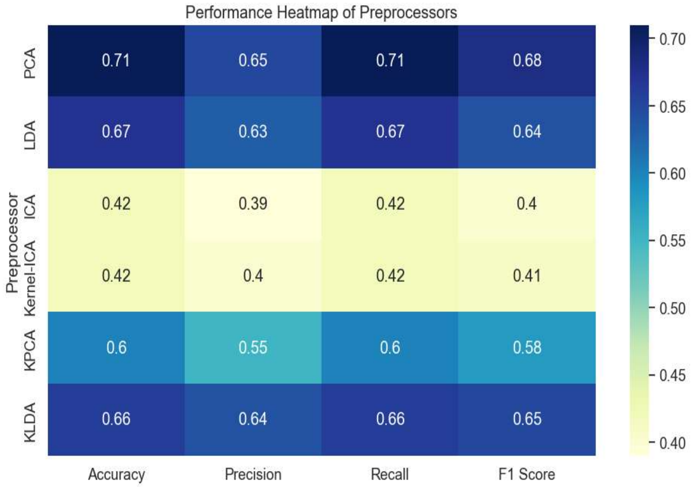

Fig.4 Heatmap illustrating accuracy, precision, recall, and F1-score across all preprocessing methods.

Finally, the fold-wise accuracy distributions in Fig.5 provide insight into the stability of each preprocessing method across 10-fold cross-validation. PCA and KLDA show minimal variance across folds, reflecting their reliability and generalization capacity. Although KPCA achieved moderate accuracy, it demonstrates slightly higher dispersion, indicating sensitivity to the composition of the folds. Meanwhile, both ICA and Kernel-ICA display the largest variability, suggesting limited robustness when faced with changes in data partitioning. In conclusion, the results indicate that while nonlinear preprocessing generally enhances performance, PCA remains the most effective method for this dataset. This is likely due to its ability to capture dominant global variance patterns that closely correspond with fault behavior. KLDA also emerges as a competitive alternative by incorporating class information during dimensionality reduction. Kernel-based approaches, particularly KPCA, show promise but require further optimization, while ICA-based approaches seem ill-suited for PV fault classification due to their instability and limited discriminative ability.

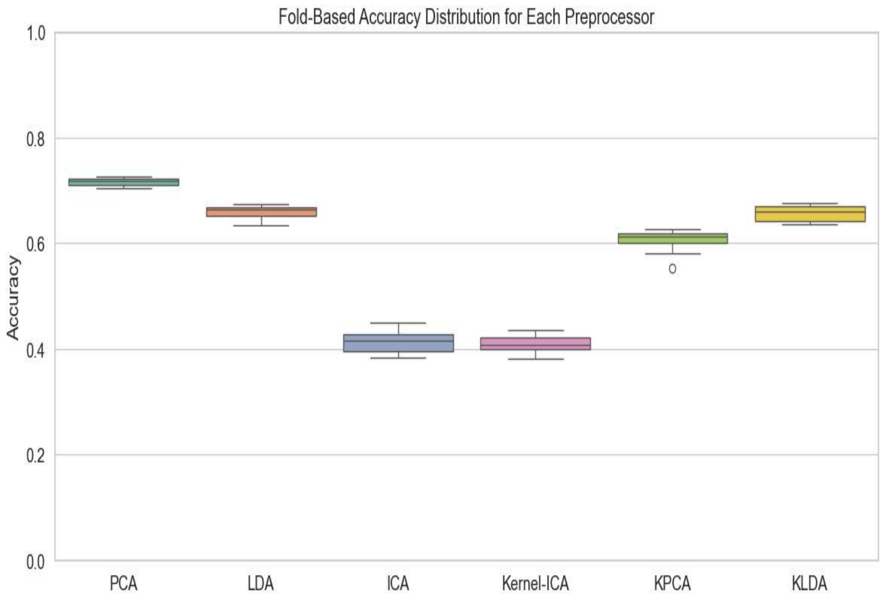

Fig.5 Fold-wise accuracy distribution (boxplots) showing the stability of each preprocessing method across 10-fold validation.

4. CONCLUSION

This study conducted a systematic comparison of linear and nonlinear preprocessing techniques for classifying faults in PV systems using a unified evaluation framework based on SVM. The results clearly show that the choice of preprocessing method significantly influences classification performance. Among the various methods tested, PCA achieved the highest overall accuracy and stability. This indicates that variance-maximizing projections effectively capture the key fault-related patterns in PV electrical signals. KLDA also performed well, benefiting from supervised class-aligned projections that improve the separability between different operational states. The nonlinear approaches yielded mixed results. KPCA provided competitive accuracy by revealing the nonlinear structures present in the data. In contrast, Kernel ICA and classical ICA exhibited low and unstable performance, suggesting that independence-based decompositions are not well suited for the correlated measurements typically found in PV systems. Further analysis of fold-wise variability confirmed that PCA and KLDA were the most robust methods across different cross-validation partitions. Overall, the findings underscore the critical role of dimensionality reduction in diagnosing faults in PV systems and indicate that linear transformations—especially PCA—remain highly effective, even though many PV faults display nonlinear characteristics. This study offers practical guidance for selecting preprocessing strategies for real-time monitoring applications and lays the groundwork for future research involving hybrid deep learning pipelines, adaptive kernel optimization, and physics-informed feature extraction for next-generation PV diagnostic systems.

REFERENCES

Amiri, A. F., Kichou, S., Oudira, H., Chouder, A., & Silvestre, S. (2024). Fault detection and diagnosis of a photovoltaic system based on deep learning using the combination of a convolutional neural network (cnn) and bidirectional gated recurrent unit (Bi-GRU). *Sustainability*, *16*(3), 1012.

Aziz, F., Haq, A. U., Ahmad, S., Mahmoud, Y., Jalal, M., & Ali, U. (2020). A novel convolutional neural network-based approach for fault classification in photovoltaic arrays. *IEEE Access*, *8*, 41889-41904.

Cai, X., & Wai, R. J. (2022). Intelligent DC arc-fault detection of solar PV power generation system via optimized VMD-based signal processing and PSO–SVM classifier. *IEEE Journal of Photovoltaics*, *12*(4), 1058-1077.

Dhibi, K., Fezai, R., Mansouri, M., Trabelsi, M., Kouadri, A., Bouzara, K., ... & Nounou, M. (2020). Reduced kernel random forest technique for fault detection and classification in grid-tied PV systems. *IEEE Journal of Photovoltaics*, *10*(6), 1864-1871.

Dhibi, K., Mansouri, M., Bouzrara, K., Nounou, H., & Nounou, M. (2021). An enhanced ensemble learning-based fault detection and diagnosis for grid-connected PV systems. *IEEE Access*, *9*, 155622-155633.

Et-Taleby, A., Chaibi, Y., Benslimane, M., & Boussetta, M. (2023). Applications of machine learning algorithms for photovoltaic fault detection: a review. *Statistics, Optimization & Information Computing*, *11*(1), 168-177.

Gawre, S. K. (2022). Advanced fault diagnosis and condition monitoring schemes for solar PV systems. In *Planning of Hybrid Renewable Energy Systems, Electric Vehicles and Microgrid: Modeling, Control and Optimization* (pp. 27-59). Singapore: Springer Nature Singapore.

Ghoneim, S. S., Rashed, A. E., & Elkalashy, N. I. (2021). Fault detection algorithms for achieving service continuity in photovoltaic farms. *Intelligent Automation & Soft Computing*, *29*(3), 467-479.

Hajji, M., Yahyaoui, Z., Mansouri, M., Nounou, H., & Nounou, M. (2023). Fault detection and diagnosis in grid-connected PV systems under irradiance variations. *Energy Reports*, *9*, 4005-4017.

Harrou, F., Saidi, A., Sun, Y., & Khadraoui, S. (2021). Monitoring of photovoltaic systems using improved kernel-based learning schemes. *IEEE Journal of Photovoltaics*, *11*(3), 806-818.

Hyvärinen, A., & Oja, E. (2000). Independent Component Analysis: Algorithms and Applications. *Neural Networks*, 13(4–5), 411–430.

IEA. (2023). *World Energy Outlook 2023*. International Energy Agency. https://iea.blob.core.windows.net/assets/86ede39e-4436-42d7-ba2a-edf61467e070/WorldEnergyOutlook2023.pdf Last Access: 06.12.2025.

Lazzaretti, A. E., Costa, C. H. D., Rodrigues, M. P., Yamada, G. D., Lexinoski, G., Moritz, G. L., ... & Santos, R. B. D. (2020). A monitoring system for online fault detection and classification in photovoltaic plants. *Sensors*, *20*(17), 4688.

Lu, S. D., Liu, H. D., Wang, M. H., & Wu, C. C. (2024). A novel strategy for multitype fault diagnosis in photovoltaic systems using multiple regression analysis and support vector machines. *Energy Reports*, *12*, 2824-2844.

Mansouri, M., Trabelsi, M., Nounou, H., & Nounou, M. (2021). Deep learning-based fault diagnosis of photovoltaic systems: A comprehensive review and enhancement prospects. *IEEE Access*, *9*, 126286-126306.

Prasanna, R., Karthik, C., Chowdhury, S., & Khan, B. (2022). Comprehensive review on modelling, estimation, and types of faults in solar photovoltaic system. *International Journal of Photoenergy*, *2022*(1), 3053317.

Qureshi, F. A., Uddin, Z., Satti, M. B., & Ali, M. (2020). ICA-based solar photovoltaic fault diagnosis. *International Transactions on Electrical Energy Systems*, *30*(8), e12456.

Radhakrishnan, P., Ramaiyan, K., Vinayagam, A., & Veerasamy, V. (2021). A stacking ensemble classification model for detection and classification of power quality disturbances in PV integrated power network. *Measurement*, *175*, 109025.

Shi, H., Liu, J., Wu, Y., Zhang, K., Zhang, L., & Xue, P. (2016). Fault diagnosis of nonlinear and large-scale processes using novel modified kernel Fisher discriminant analysis approach. *International Journal of Systems Science*, *47*(5), 1095-1109.

Yao, S., Kang, Q., Zhou, M., Abusorrah, A., & Al-Turki, Y. (2021). Intelligent and data-driven fault detection of photovoltaic plants. *Processes*, *9*(10), 1711.

Zhang, Y. (2008). Fault detection and diagnosis of nonlinear processes using improved kernel independent component analysis (KICA) and support vector machine (SVM). *Industrial & Engineering Chemistry Research*, *47*(18), 6961-6971.

Zhang, K., Tang, B., Deng, L., & Yu, X. (2021). Fault detection of wind turbines by subspace reconstruction-based robust kernel principal component analysis. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1-11.