IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 3 July 2024, accepted 13 August 2024, date of publication 16 August 2024, date of current version 6 September 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3445287

RESEARCH ARTICLE

Machine Learning-Based Multiclass Anomaly Detection and Classification in Hybrid Active Distribution Networks

SADULLAH CHANDIO[1], JAVED AHMED LAGHARI[1], (Senior Member, IEEE), MUHAMMAD AKRAM BHAYO[1], MOHSIN ALI KOONDHAR[1], YUN-SU KIM[2], (Senior Member, IEEE), BESMA BECHIR GRABA[3], AND EZZEDDINE TOUTI[4]

[1]Department of Electrical Engineering, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah, Sindh 67450, Pakistan
[2]Graduate School of Energy Convergence, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, South Korea
[3]Department of Physics, College of Science, Northern Border University, Arar 91431, Saudi Arabia
[4]Department of Electrical Engineering, College of Engineering, Northern Border University, Arar 91431, Saudi Arabia

Corresponding author: Yun-Su Kim (yunsukim@gist.ac.kr)

This work was supported by Gwangju Institute of Science and Technology (GIST) Research Project grant funded by the GIST in 2024. Also the authors extend their appreciation to the Deanship of Scientific Research at Northern Border University, Arar, KSA for funding this research work through the project number "NBU-FFR-2024-2226-07".

ABSTRACT Anomaly detection in power systems is crucial for operational reliability and safety, often addressed through binary classification in existing research. However, a research gap exists in multiclass classification for enhanced reliability. To bridge this gap, this study employs four machine learning (ML) classifiers: Random Forest (RF), Decision Tree, Naive Bayes (NB), and Support Vector Machine (SVM) using comprehensive testing on a dataset comprising sixteen indices and their pair combinations (totaling 136 pairs). These classifiers, trained on a dataset derived from simulating a test system with hybrid DGs, exhibit superior anomaly detection, especially with the $\frac{dv}{dq}\&\frac{dv}{dp}$ pair. Among them, RF and DT classifier achieves precision, recall, and F score of unity and outperforming NB and SVM. The performance of the proposed RF and DT classifiers with $\frac{dv}{dq}\&\frac{dv}{dp}$ pair is compared with existing research papers in terms of accuracy and data division. The comparison shows that the proposed RF and DT classifiers with $\frac{dv}{dq}\&\frac{dv}{dp}$ pair achieve 100% accuracy even with 50% data division, whereas other techniques fail to achieve it even at 20% for testing and 80% for training. The study underscores the critical role of pair selection and classifier combinations in effective anomaly detection, facilitating the implementation of robust mitigating strategies for power system stability.

INDEX TERMS Anomaly detection, islanding detection, machine learning classifiers, hybrid active distribution network, PV.

I. INTRODUCTION

The amalgamation of Distributed Energy Resources (DER) into distribution networks has marked a significant shift towards sustainable power generation, diminishing reliance on non-renewable sources [1], [2]. Among these DERs, inverter-based DGs have gained prominence for their environmental benefits. However, alongside their advantages, these systems pose challenges in terms of control, privacy, stability, and the detection of faults and anomalies [3], [4]. The reliable operation of inverter-based DGs is crucial, given their integration into the grid's closest points to the load. The detection and classification of anomalies such as faults and islanding events in power infrastructure have garnered the attention of researchers aiming for reliable and safe operations [5]. Detecting these faults and islanding events promptly can prevent substantial financial losses, maintain uninterrupted power supply, and contribute to the economic stability

The associate editor coordinating the review of this manuscript and approving it for publication was Ahmed F. Zobaa.

TABLE 1. Comparative assessment of anomaly detection and classification in HADN.

Ref	Multiclass Classification	Faults	DG Power Variations	Load Variations	Induction Motor Starting Events	Capacitor Switching	Islanding Events	Testing Data ≥ 50 %
[3, 7-13]	×	✓	×	×	×	×	×	×
[5, 14-17]	✓	✓	×	×	×	×	×	×
[18-21]	×	✓	×	×	×	×	✓	×
[22-27]	×	✓	×	✓	×	✓	✓	×
[28, 29] [30]	×	✓	×	✓	✓	✓	✓	×
[31]	×	✓	✓	✓	×	×	✓	×
[32]	✓	✓	×	×	×	✓	✓	×
[33]	✓	✓	✓	✓	×	✓	✓	×
Proposed Work	✓	✓	✓	✓	✓	✓	✓	✓

✓: Considered, ×: Not considered

of a country. In power systems, events are categorized as normal or abnormal, with normal occurrences including routine actions like switching capacitor banks, DG power variations, and sudden static and dynamic load changes. Abnormal events encompass faults at various locations, islanding scenarios with active/reactive power mismatches, and other deviations that can lead to voltage, frequency, and stability issues [6].

Detecting and categorizing these events play a critical role in preemptively addressing potential disruptions.

Various techniques are employed for anomaly detection, ranging from statistical methods like principal component analysis to signal processing techniques like the Fourier and wavelet transform [16]. Signal processing techniques are particularly useful in analyzing power system data to identify abnormal patterns. For instance, islanding events, which can cause voltage fluctuations and power quality issues, can be detected by examining voltage, current, or frequency data using Fourier transform, wavelet transform, or ML algorithms [34]. Early detection enables timely response actions to maintain grid stability and reliability. Additionally, ML algorithms complement these methods by modeling normal system behavior and flagging deviations. These approaches have demonstrated efficacy in detecting anomalies like line outages, transformer failures, islanding events and generator malfunctions [12], [14].

The existing literature reveals that most research on anomaly detection in power systems focuses on binary classification, distinguishing between the existence and non-existence of faults or islanding events. However, this approach overlooks several critical factors, such as the types of anomaly events. Many studies have concentrated solely on fault classification, merely identifying the presence or absence of a fault, as referenced in [3], [4], [7], [8], [9], [10], [11], [12], and [13] research studies.

These research papers are limited to binary classification and do not address other significant events essential to modern power systems, such as distributed generation (DG) power variations, islanding events, capacitor switching, load variations, and induction motor starting events. Consequently, the findings of these papers are not entirely applicable to contemporary power systems.

Additionally, the literature indicates that only a few studies have explored multiclass classification of anomaly events in power systems [5], [14], [15], [16], [17]. Even in these cases, the focus has predominantly been on fault types and islanding events, without a comprehensive approach to classifying a wider range of anomaly events. The detailed literature reviews for detecting anomalies in power systems are elucidated in Table 1.

From Table 1, it is evident that various techniques have been successfully implemented for detecting anomalies in power systems. However, there is a notable gap in current research, particularly in the area of multiclass classification of anomalies. This classification is crucial for ensuring the power system reliability in the presence of DGs, contributing significantly to resource efficiency and overall system reliability. Additionally, it has been observed that researchers often use training and testing datasets that are insufficient in size, falling below the 50% threshold. This situation raises concerns regarding potential biases in performance evaluations. To tackle this challenge, a practical approach is proposed by splitting the datasets evenly at 50%. This division aims to reduce the risk of overfitting, a common issue when dealing with limited dataset sizes, thus enhancing the robustness of performance evaluations.

A. OBJECTIVES OF THE RESEARCH

To address the existing gap in the literature, the primary aim of this research is to apply a ML approach to multiclass classification of anomalies in Hybrid Active Distribution Networks (HADN). The specific objectives of proposed approach are presented as below:

1. To design and implement an innovative multiclass detection and classification system for anomaly events using ML techniques.
2. To perform a comprehensive analysis of 16 fundamental indices and their combinations to identify the most

TABLE 2. Types of indices employed for anomaly detection.

S. No.	Indices	Symbol	S. No.	Indices	Symbol
1	Active Power Variation w.r.t time	dP/dt	9	Active Power Variation w.r.t Reactive Power	dP/dQ
2	Reactive Power Variation w.r.t time	dQ/dt	10	Voltage Variation w.r.t Reactive Power	dV/dQ
3	Frequency Variation w.r.t time	df/dt	11	Reactive Power Variation w.r.t Frequency	dQ/df
4	Voltage Variation w.r.t time	dV/dt	12	Active Power Variation w.r.t Voltage	dP/dV
5	Reactive Power Variation w.r.t Active Power	dQ/dP	13	Active Power Variation w.r.t Frequency	dP/df
6	Reactive Power Variation w.r.t Voltage	dQ/dV	14	Voltage Variation w.r.t Frequency	dV/df
7	Frequency Variation w.r.t Reactive Power	df/dQ	15	Frequency Variation w.r.t Active Power	df/dP
8	Voltage Variation w.r.t Active Power	dV/dP	16	Frequency Variation w.r.t Voltage	df/dV

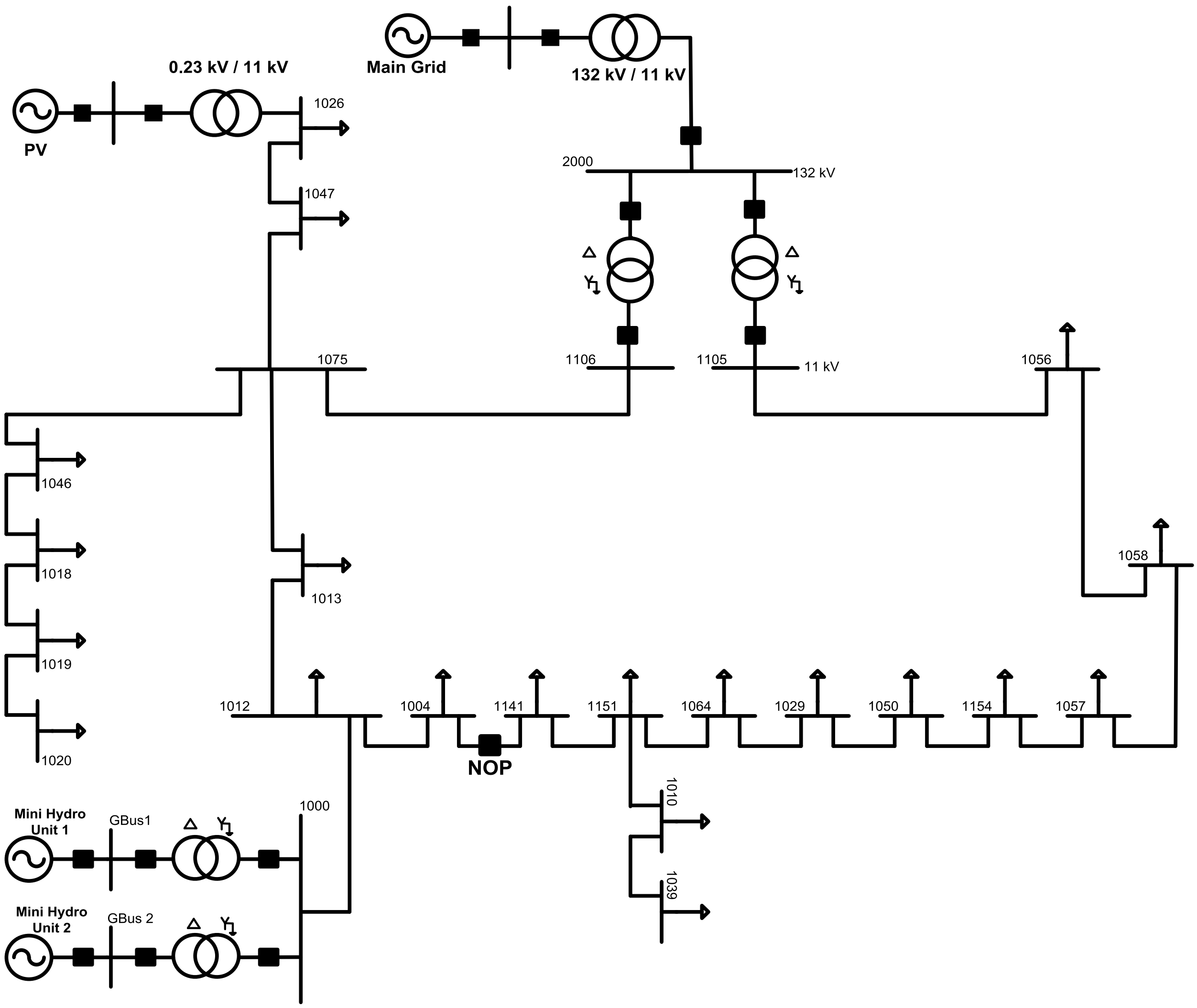

FIGURE 1. Hybrid active distribution network.

suitable ones that yield the highest precision, recall, F-score, and accuracy.

3. To conduct a thorough comparative analysis of various ML algorithms to determine the most effective classifier for distinguishing between normal and anomaly events in HADN, with a focus on achieving optimal precision, recall, F-score, and overall accuracy.
4. To improve the reliability of the HADN system by accurately detecting and classifying anomalies in the system.

By achieving these goals, this study intends to make a substantial contribution to the progression of anomaly detection techniques in HADN, ultimately enhancing their overall reliability.

B. RESEARCH CONTRIBUTIONS

This research paper makes several noteworthy contributions:

1. By individually examining 16 fundamental indices and their pair wise combinations, totaling 136 combinations, this study aimed to identify the most effective indices for achieving high accuracy in anomaly detection. Notably, among these combinations, the pairing of $\frac{dv}{dq}$ & $\frac{dv}{dp}$ emerged as the most suitable.
2. The research introduces a robust detection and classification framework for anomalies in HADN, leveraging four prominent ML classifiers (RF, DT, SVM, and NB). Among these classifiers, RF and DT achieved impressive accuracy rates of 100% for the $\frac{dv}{dq}$ & $\frac{dv}{dp}$, pair respectively.
3. A significant milestone achieved in this study is the successful implementation of multiclass classification for anomaly events, marking a substantial advancement in ensuring the reliable operation of HADN systems.

Collectively, these contributions significantly enhance the understanding and capability of detecting and classifying anomalies in HADN systems. Subsequent sections of the paper will delve into a detailed explanation of the methodology utilized in this research. The rest of the paper is structured as follows: Section II presents the entire research methodology of the proposed technique, including the modeling of the test system in PSCAD software, the selection of indices and their pair combinations, and the methodology for the proposed machine learning (ML) classifiers. Section III details the data generation methodology for classifier training and evaluation. This section covers the generation of various normal and anomaly events at different locations and types within the test system, concluding with the data division for training and testing. Section IV discusses the different evaluation metrics used to assess the performance of the ML classifiers. Section V presents the simulation results of the ML classifiers for various training and testing scenarios involving different indices and their pairs, identifying the most suitable indices and ML classifier. Section VI compares the proposed technique with existing classifiers in terms of data division and accuracy, demonstrating its effectiveness. Section VII provides the conclusion, summarizing the key findings and suggesting directions for future research.

II. RESEARCH METHODOLOGY

A. MODELING HYBRID ACTIVE DISTRIBUTION NETWORK (HADN) FOR ANALYSIS

To perform a comparative analysis of different indices and validate the proposed method, an 11 kV HADN located in Malaysia is chosen for study. This HADN includes a combination of hybrid DG units, specifically three DGs (2 Mini hydro DGs and a PV-based DG), in addition to an electric grid. The system comprises 28 buses and 20 lumped loads, as depicted in Figure 1. The connection between the transmission grid and the HADN is facilitated by a transformer rated at 132 kV/11 kV with a capacity of 30 MVA. Both Mini hydro DGs having 2 MVA capacity, operates at 3.3 kV and employs 2 MVA transformers to raise its voltage to 11 kV. Similarly, the PV generation unit, with a capacity of 1 MW, is linked to a 1 MVA transformer to increase the voltage from 0.23 kV to 11 kV. To design the models for the Mini hydro DGs and PV generation units, standard models available in the PSCAD/EMTDC library for exciter, governor, hydraulic turbine, and PV modules are utilized. The entire line is structured following a nominal π form for accurate representation. The specific parameters of the HADN DGs and load data are taken from [35].

B. SELECTION OF INDICES

Extensive research has been dedicated to exploring a range of indices for feature extraction within the anomaly detection domain. Notably, in [35] conducted a focused investigation in this field and identified 16 indices that were primarily utilized for detecting islanding events. However, this current study diverges by employing the same set of 16 indices for anomaly detection in power systems, as outlined in Table 2.

The principal objective of this study is to ascertain the minimum requisite number of indices for effective anomaly detection within power systems. Initially, the study endeavors to identify individual indices capable of autonomously detecting all anomaly types. If a single parameter proves insufficient, the study will then explore combinations of two indices from the pool of 16, with the ultimate goal of achieving comprehensive anomaly detection coverage (100%). In pursuit of this goal, the study will systematically assess all conceivable combinations of the 16 indices utilizing the combination formula as given below:

$$C = \frac{n!}{k! \times (n-k)!} \tag{1}$$

This results in a total of 120 combinations when selecting 2 indices at a time. These combinations will undergo comprehensive testing using various classifiers to determine the optimal combination of indices that maximizes anomaly detection capability in power systems. Table 3 provides a comprehensive listing of all 120 combinations, offering a basis for in-depth analysis and assessment.

C. ML CLASSIFIERS

Within this study, four widely recognized ML classifiers are explored. The assessment of their effectiveness will rely on various metrics including precision, recall, accuracy, F-measure, and confusion matrix analysis. Through a comparative analysis, the performance differences among these classifiers will be highlighted. Subsequent sections provide concise descriptions of each classifier.

1) RANDOM FOREST (RF) CLASSIFIER

The RF classifier is a widely used technique in supervised ML, suitable for classification challenges. Its approach is based on ensemble learning, where multiple decision trees are employed to different subsets of the data, enhancing predictive accuracy. The performance and problem-solving

TABLE 3. Index combinations for anomaly detection.

S. No.	Pair	S. No.	Pair	S. No.	Pair	S. No.	Pair
1	$dP/dt, dQ/dt$	31	$df/dt, dP/df$	61	$dP/df, dP/dV$	91	$df/dV, dV/dP$
2	$dP/dt, df/dt$	32	$df/dt, dQ/df$	62	$dP/df, dV/dQ$	92	$df/dV, dQ/dP$
3	$dP/dt, dV/dt$	33	$df/dt, dP/dQ$	63	$dP/df, dQ/dV$	93	$df/dQ, df/dP$
4	$dP/dt, dP/df$	34	$df/dt, df/dV$	64	$dP/df, dV/dP$	94	$df/dQ, dV/df$
5	$dP/dt, dQ/df$	35	$df/dt, df/dQ$	65	$dP/df, dQ/dP$	95	$df/dQ, dP/dV$
6	$dP/dt, dP/dQ$	36	$df/dt, df/dP$	66	$dP/df, dP/dQ$	96	$df/dQ, dV/dQ$
7	$dP/dt, df/dV$	37	$df/dt, dV/df$	67	$dQ/df, df/dV$	97	$df/dQ, dQ/dV$
8	$dP/dt, df/dQ$	38	$df/dt, dP/dV$	68	$dQ/df, df/dQ$	98	$df/dQ, dV/dP$
9	$dP/dt, df/dP$	39	$df/dt, dV/dQ$	69	$dQ/df, df/dP$	99	$df/dQ, dQ/dP$
10	$dP/dt, dV/df$	40	$df/dt, dQ/dV$	70	$dQ/df, dV/df$	100	$df/dP, dV/df$
11	$dP/dt, dP/dV$	41	$df/dt, dV/dP$	71	$dQ/df, dP/dV$	101	$df/dP, dP/dV$
12	$dP/dt, dV/dQ$	42	$df/dt, dQ/dP$	72	$dQ/df, dV/dQ$	102	$df/dP, dV/dQ$
13	$dP/dt, dQ/dV$	43	$dV/dt, dP/df$	73	$dQ/df, dQ/dV$	103	$df/dP, dQ/dV$
14	$dP/dt, dV/dP$	44	$dV/dt, dQ/df$	74	$dQ/df, dV/dP$	104	$df/dP, dV/dP$
15	$dP/dt, dQ/dP$	45	$dV/dt, dP/dQ$	75	$dQ/df, dQ/dP$	105	$df/dP, dQ/dP$
16	$dQ/dt, df/dt$	46	$dV/dt, df/dV$	76	$dP/dQ, df/dV$	106	$dV/df, dP/dV$
17	$dQ/dt, dV/dt$	47	$dV/dt, df/dQ$	77	$dP/dQ, df/dQ$	107	$dV/df, dV/dQ$
18	$dQ/dt, dP/df$	48	$dV/dt, df/dP$	78	$dP/dQ, df/dP$	108	$dV/df, dQ/dV$
19	$dQ/dt, dQ/df$	49	$dV/dt, dV/df$	79	$dP/dQ, dV/df$	109	$dV/df, dV/dP$
20	$dQ/dt, dP/dQ$	50	$dV/dt, dP/dV$	80	$dP/dQ, dP/dV$	110	$dV/df, dQ/dP$
21	$dQ/dt, dP/dV$	51	$dV/dt, dV/dQ$	81	$dP/dQ, dV/dQ$	111	$dP/dV, dV/dQ$
22	$dQ/dt, df/dQ$	52	$dV/dt, dQ/dV$	82	$dP/dQ, dQ/dV$	112	$dP/dV, dQ/dV$
23	$dQ/dt, df/dP$	53	$dV/dt, dV/dP$	83	$dP/dQ, dV/dP$	113	$dP/dV, dV/dP$
24	$dQ/dt, dV/df$	54	$dV/dt, dQ/dP$	84	$dP/dQ, dQ/dP$	114	$dP/dV, dQ/dP$
25	$dQ/dt, dP/dV$	55	$dP/df, dQ/df$	85	$df/dV, df/dQ$	115	$dV/dQ, dQ/dV$
26	$dQ/dt, dV/dQ$	56	$dP/df, dP/dQ$	86	$df/dV, df/dP$	116	$dV/dQ, dV/dP$
27	$dQ/dt, dQ/dV$	57	$dP/df, df/dV$	87	$df/dV, dV/df$	117	$dV/dP, dQ/dP$
28	$dQ/dt, dV/dP$	58	$dP/df, df/dQ$	88	$df/dV, dP/dV$	118	$dQ/dV, dV/dP$
29	$dQ/dt, dQ/dP$	59	$dP/df, df/dP$	89	$df/dV, dV/dQ$	119	$dQ/dV, dQ/dP$
30	$df/dt, dV/dt$	60	$dP/df, dV/df$	90	$df/dV, dQ/dV$	120	$dV/dP, dQ/dP$

capabilities of Random Forest improve with the addition of more trees, establishing it as a foundational principle in ML. Mathematically, RF classifier is represented as below [4]:

$$\hat{Y} = RF\left(X\right) \tag{2}$$

where $\widehat{Y}$ represents the predicted output or class label, X denotes the input features or attributes used for prediction and $RF(X)$ signifies the Random Forest model applied to the input features X to predict the output $\widehat{Y}$.

2) DECISION TREE (DT) CLASSIFIER

DT classifier is a powerful and extensively employed ML model that is popular for their interpretability and ease of understanding. They are useful when dealing with complex datasets. The mathematical equation for entropy in Decision Trees is given by [10].

$$E\left(s\right) = \sum_{i=1}^{c} -p_i \times \log_2\left(p_i\right) \tag{3}$$

where, $E\left(s\right)$ represents the entropy at a specific node in the decision tree, measuring the uncertainty or disorder in the dataset, c denotes the number of classes or possible outcomes in the dataset, p_i indicates the probability of the dataset belonging to class i, $\log_2(p_i)$ refers to the logarithm of p_i to the base 2. The entropy calculation quantifies the level of disorder or randomness in the dataset at a given node. Lower entropy values designate more purity in the dataset, while higher entropy values suggest more diversity or mixed classes.

Decision Trees use entropy as a key criterion to make splitting decisions, aiming to reduce entropy and improve

TABLE 4. Implementation of fault scenarios.

No.	Types of fault	No.	Types of fault	No.	Types of fault
1	$A-G$ Fault	5	$BC-G$ Fault	9	$C-A$ Fault
2	$B-G$ Fault	6	$CA-G$ Fault	10	$ABC-G$ Fault
3	$C-G$ Fault	7	$A-B$ Fault	11	$A-B-C$ Fault
4	$AB-G$ Fault	8	$B-C$ Fault	-	-

TABLE 5. Scenarios of solar power variability.

S. No.	Solar Irradiance (W/m^2)	Temperature	Step Change	Cases
1	1000	0°C - 55°C	1°C	56
2	1000 W/m^2-10 W/m^2	25°C	10 W/m^2	100
3	1000-100	1°C - 50°C	1°C & 10 W/m^2	100

information gain at each node, leading to a more accurate and reliable model.

3) NAÏVE BAYES (NB) CLASSIFIER

The NB classifier is a fundamental algorithm in ML, dating back to the 1950s. It has its roots in the principles of Bayes Theorem and is known for its simplicity and effectiveness in various fields, including statistics and computer science. The NB algorithm is based on a probabilistic approach, specifically Bayesian probability, which calculates the probability of an event based on prior knowledge and conditional probabilities. The mathematical equation for the Naive Bayes Classifier is derived from Bayesian probability and is expressed as [10]:

$$P\left(C_k \left| x\right.\right) = \frac{p\, C_k \times p\left(x \left| C_k\right.\right)}{p\left(x\right)} \tag{4}$$

In this context, $P\left(C_k \left| x\right.\right)$ represents the class C_k posterior probability which is what the classifier predicts, $p\left(C_k\right)$ symbolizes the prior probability, which is the probability of class C_k occurring before observing the input $x.p\left(C_k \left| x\right.\right)$ represents the likelihood, which is the probability of observing input x for true class C_k $p(x)$ denotes the evidence, which is the probability of observing input x across all classes.

4) SUPPORT VECTOR MACHINE (SVM) CLASSIFIER

(SVM) is an extensively employed technique in the field of ML, particularly in supervised learning tasks. It is used for classification and regression tasks on datasets that can be either discrete or continuous. SVM focuses on maximizing the margin between classes, aiming to achieve better classification accuracy for new data points. The primary optimization goal in SVM is to diminish the following cost function [36], [37]:

$$\left[\frac{1}{n}\sum_{i=1}^{n}\max\left(0, 1-y_i\left(w^T x_i - b\right)\right)\right] + \lambda \left\|w\right\|^2 \tag{5}$$

In this context, n represents the data points quantity, x_i and y_i denotes the i^{th} data point features and labels, w depicts the weight vector, b symbolizes the bias term, and λ is the regularization parameter, $\max(0, 1 - y_i(w^T x_i - b))$ is the hinge loss function, penalizing misclassifications, $\|w\|^2$ is the regularization term that helps in controlling over fitting. SVM can handle various problem types by using different types of kernels such as linear, polynomial, Gaussian, and radial basis function (RBF). These algorithms empower SVM to grasp intricate data relationships, making it a versatile and powerful algorithm for classification and regression tasks. This research employs polynomial kernel for better results.

III. DATA GENERATION METHODOLOGY FOR CLASSIFIER TRAINING AND EVALUATION

To generate the dataset for training and testing purposes, simulations of the HADN system were conducted using PSCAD/EMTDC software. These simulations aimed to replicate a wide range of anomaly events as well as normal operational events. For each case, all 16 parameters were meticulously measured, covering various scenarios such as symmetrical and unsymmetrical faults, load variations, capacitor switching events, effects of induction motor starting, fluctuations in solar power output, and occurrences of islanding. This thorough methodology guarantees that the ML classifier is fully capable of effectively detecting a wide variety of anomalies. The detailed methodology for data collection and categorization for each event type is elaborated upon in subsequent sections.

A. FAULT SCENARIOS

Within this classification, there are 248 distinct fault scenarios, encompassing both symmetrical and unsymmetrical faults. These scenarios were generated by implementing various fault types across different buses, as outlined in Table 4.

B. SOLAR POWER VARIATION SCENARIOS

For fluctuations in solar power, 256 scenarios incorporating alterations in irradiance and temperature are examined, as detailed in Table 5. It's important to emphasize that solar energy is variable and influenced by atmospheric factors, posing challenges for the ML classifier in accurately identifying anomalies within the power system.

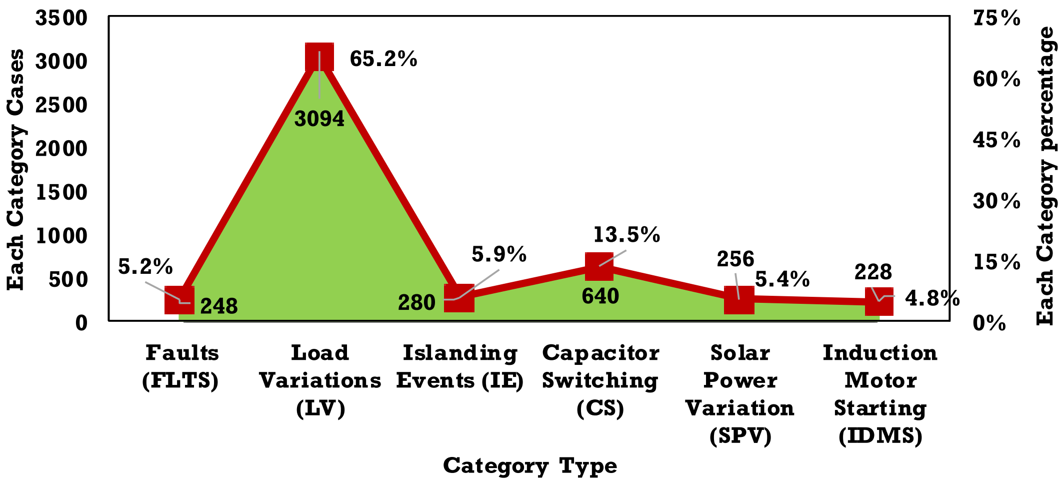

FIGURE 2. Scenarios distribution for each category.

C. LOAD VARIATION SCENARIOS

Within this classification, a total of 3094 load variation scenarios encompassing load increases and decreases are derived by connecting and disconnecting loads spanning from 0.01 MW to 0.7 MW at different buses.

D. CAPACITOR SWITCHING SCENARIOS

Within this classification, 640 instances of capacitor switching scenarios are derived by implementing the capacitor switching effect, spanning from 0.01 Mvar to 0.3 Mvar at different buses.

E. INDUCTION MOTOR STARTING EFFECT SCENARIOS

In this categorization, data is synthesized to evaluate the classifier's effectiveness in detecting anomalies amidst the initial transients of an induction motor. The induction motor exhibits a notably diminished power factor during its initial startup phase, posing a potential challenge to the anomaly detection technique. To address this, the startup impact of various induction motors with capacities ranging from 50 kW to 500 kW, incremented in steps of 50 kW, is simulated across different buses.

Consequently, 228 scenarios are generated to simulate the startup effects of induction motors on diverse buses, aiming to comprehensively evaluate their impact on the anomaly detection method.

F. ISLANDING EVENTS SCENARIOS

Within this classification, a collection of 280 islanding cases is generated through simulating grid disconnections under various power mismatches, spanning from 0.01 to 1 MW and 0.01 to 1 Mvar. This wide range of scenarios enhances the classifier's effectiveness in distinguishing anomalies within power system detection. A total of 4746 instances are analyzed for the ML classifier. The breakdown of different scenarios is illustrated in Figure 2 accordingly.

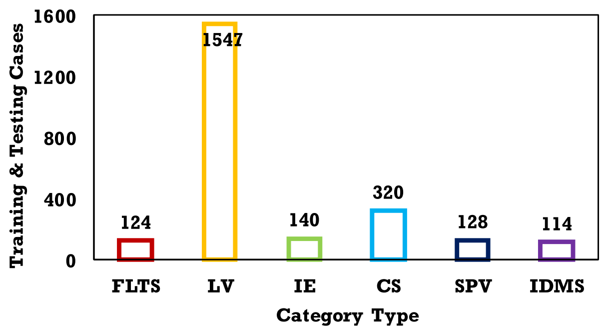

FIGURE 3. 50% division for training and testing cases of each category.

The subsequent stage entails partitioning the data into training and testing sets. This study employs a 50% split, with half of the cases randomly allocated for training and the other half for testing. Specifically, 2373 cases are randomly chosen for training. The allocation of cases from each category for both training and testing is depicted in Figure 3. The testing dataset is withheld from the classifier to ensure impartial evaluation.

IV. COMPARARTIVE PERFORMANCE ASSESSMENT OF CLASSIFIERS

This research paper will assess the performance of the machine learning classifiers using diverse metrics, including the confusion matrix, precision, recall, F score, and accuracy, to gauge their efficacy in anomaly detection. The respective mathematical equations of the various indices are given below [38].

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \times 100 \tag{6}$$

$$Precision = \frac{TP}{TP + FP} \tag{7}$$

$$Recall = \frac{TP}{TP + FP} \tag{8}$$

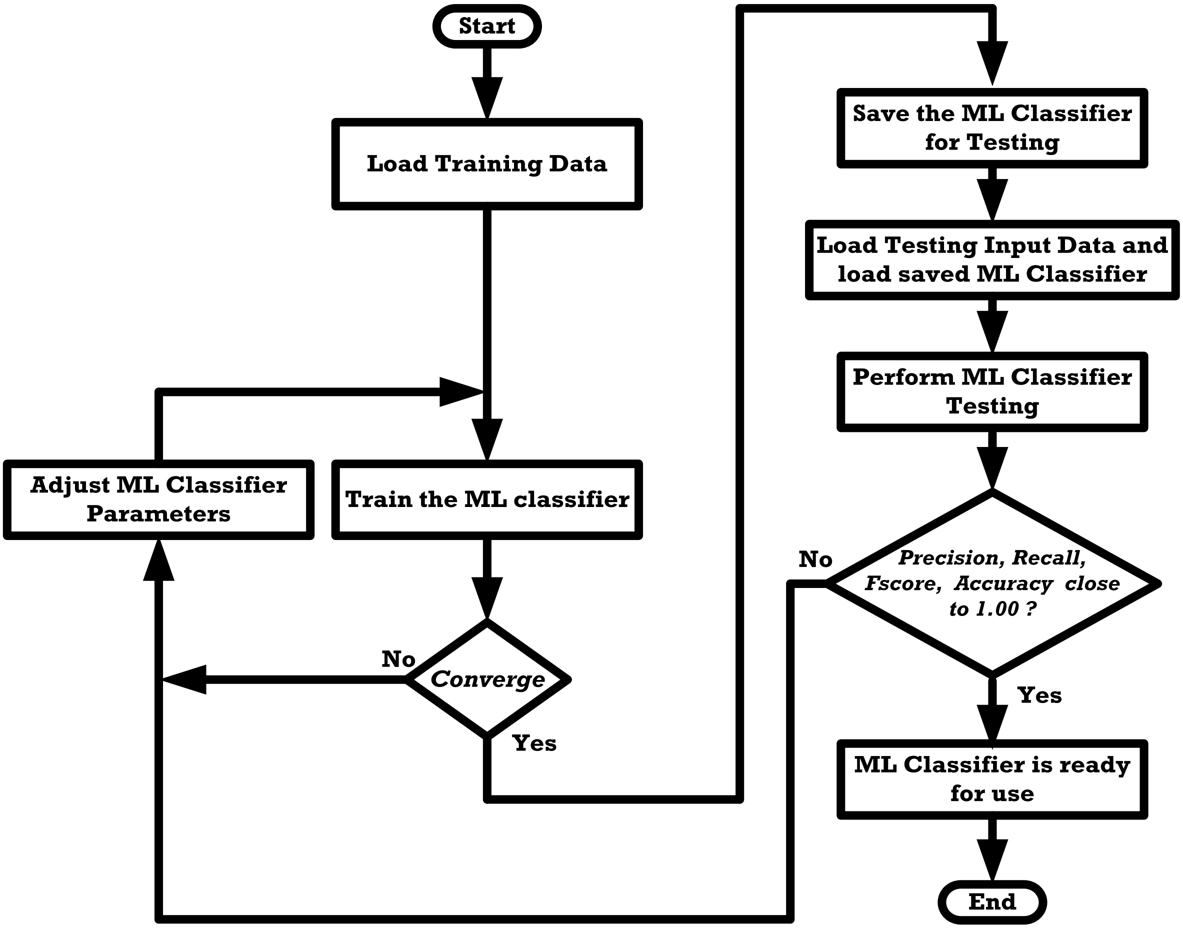

FIGURE 4. Flowchart illustrating ML approach to anomaly detection.

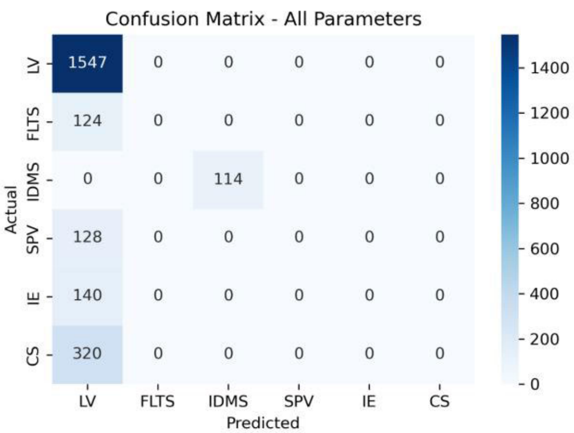

FIGURE 5. SVM classifier confusion matrix for $\frac{dv}{dq}$ & $\frac{dv}{dp}$ indices pair.

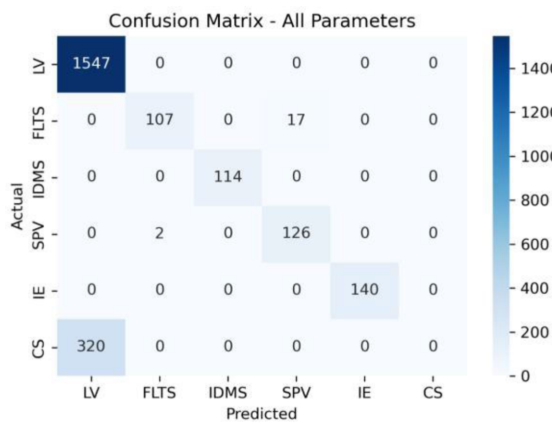

FIGURE 6. NB classifier confusion matrix for $\frac{dv}{dq}$ & $\frac{dv}{dp}$ indices pair.

$$Fscore = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{9}$$

In this context, True Positive (TP) refers to the count of accurately predicted positive instances, True Negative (TN) refers to the count of accurately predicted negative instances, False Positive (FP) refers to the count of erroneously predicted positive instances, and False Negative (FN) refers to the count of erroneously predicted negative instances. The flow chart illustrating the ML-based classifier designed for anomaly detection is depicted in Figure 4.

V. SIMULATION RESULTS

This research employs four prominent ML classifiers: RF, DT, NB, and SVM. Following successful training, these classifiers undergo rigorous evaluation with previously unseen testing data. The study focuses on 16 indices and their pair combinations, resulting in a total of 136 combinations to find the most suitable pair for achieving the highest accuracy. Among these combinations, it has been found that the $\frac{dv}{dq}$ & $\frac{dv}{dp}$ pair provides the best results among all other pairs. The confusion matrix for SVM and NB classifiers is presented in Figure 5-6, respectively, while the confusion matrix for the RF and DT classifiers, which exhibit similar responses, is shown in Figure 7.

Figures 5-7 demonstrate that RF and DT classifiers successfully and accurately detect all anomalies and normal events, achieving an accuracy of 100%. They exhibit the ability to distinguish between anomalies effectively.

In contrast, NB shows comparatively better performance than the SVM classifier across all evaluation metrics. Figure 8 presents the precision, recall, F score, and accuracy of all four classifiers for the $\frac{dv}{dq}$ & $\frac{dv}{dp}$ pair.

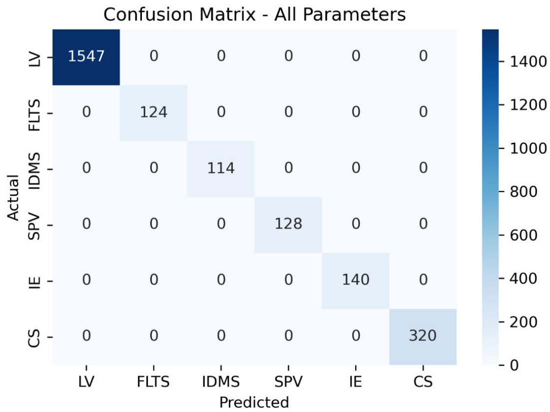

FIGURE 7. DT and RF classifier confusion matrix for $\frac{dv}{dq}$ & $\frac{dv}{dp}$ indices pair.

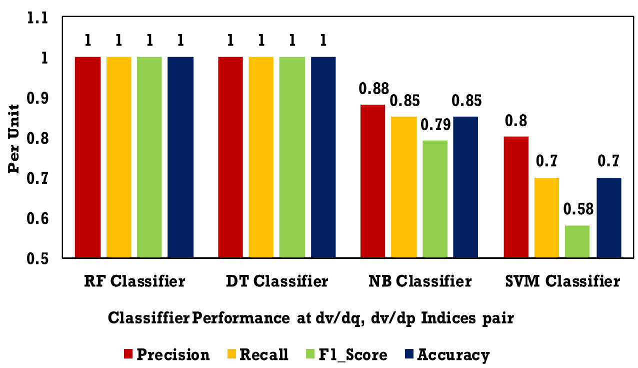

FIGURE 8. Comparison of classifier performance for $\frac{dv}{dq}$ & $\frac{dv}{dp}$ indices pair.

It is observed that both RF and DT classifiers demonstrate precision, recall, F score, and accuracy unity, showcasing superior performance compared to NB. SVM exhibits the poorest performance among all four classifiers. In conclusion, the proposed RF and DT classifiers with the $\frac{dv}{dq}$ & $\frac{dv}{dp}$ pair demonstrate the ability to detect and classify all anomalies in the power system, leading to enhanced reliability of power system operations by proposing suitable mitigating strategies for various anomalies.

VI. COMPARISON

Table 6 provides a comparative assessment of accuracy between the proposed RF and DT classifier and existing research methods. The results highlight the significance of the proposed ML algorithms (RF and DT), which have achieved 100% accuracy for all types of anomaly events. It is noteworthy that none of the existing techniques discussed in previous research papers have achieved 100% accuracy.

TABLE 6. Comparative assessment of accuracy among different methods.

Ref	Accuracy	Data Division
[3]	96.66%	30/70
[5]	99.45%	12/88
[14]	99.92%	10/90
[7]	99.74%	30/70
[8]	99.85%	30/70
[4]	99.4%	25/75
[16]	99.9%	20/80
[17]	98%	34/66
[32]	95.23%	30/70
Proposed RF and DT Classifier	100%	50/50

For instance, the highest accuracy obtained by [16] is 99.9%. However, it is important to mention that [16] employed a data division of 20% for testing and 80% for training, whereas the proposed research utilized a 50% data division for both training and testing. This approach provides more robust and challenging scenarios for evaluating classifier performance, especially for complex and real power system anomalies. The comparison underscores the superiority of the proposed research in accurately detecting and classifying all types of anomaly events with 100% accuracy, showcasing an advantage over existing techniques.

VII. CONCLUSION

This study introduces a multiclass detection and classification approach for identifying anomaly events in power systems with hybrid DGs. The research systematically investigates 16 indices and their pair combinations, totaling 136 combinations, to determine the most suitable pair for precise anomaly detection and classification. The study employs four well-known classifiers—RF, DT, NB, and SVM. Training and testing data are collected from simulating a test system with 4746 cases of various normal and anomaly events, with 50% allocated for training and an equivalent portion for evaluating classifier performance. Notably, the results demonstrate that the $\frac{dv}{dq}$ & $\frac{dv}{dp}$ pair exhibits superior efficiency in detecting and classifying anomalies across all classifiers, with RF and DT achieving perfect precision, recall, and F-measure scores of unity for this pair. Furthermore, a comparative analysis with existing techniques validates the efficacy of the proposed RF and DT classifier, achieving 100% accuracy even with 50% data division. This positions them as robust solutions for multiclass anomaly detection and classification in power systems featuring hybrid DGs. The study emphasizes the crucial role of pair selection and classifier combinations in enhancing anomaly detection effectiveness and implementing mitigating strategies for power system stability.

REFERENCES

[1] S. Sarwar, H. Mokhlis, M. Othman, M. A. Muhammad, J. A. Laghari, N. N. Mansor, H. Mohamad, and A. Pourdaryaei, "A mixed integer linear programming based load shedding technique for improving the sustainability of islanded distribution systems," *Sustainability*, vol. 12, no. 15, p. 6234, Aug. 2020.

[2] J. A. Laghari, H. Mokhlis, M. Karimi, A. H. A. Bakar, and A. Shahriari, "Artificial neural network based islanding detection technique for mini hydro type distributed generation," in *Proc. 3rd IET Int. Conf. Clean Energy Technol. (CEAT)*, Nov. 2014, pp. 1–6.

[3] M. Baker, A. Y. Fard, H. Althuwaini, and M. B. Shadmand, "Real-time AI-based anomaly detection and classification in power electronics dominated grids," *IEEE J. Emerg. Sel. Topics Ind. Electron.*, vol. 4, no. 2, pp. 549–559, Apr. 2023.

[4] A. F. Amiri, H. Oudira, A. Chouder, and S. Kichou, "Faults detection and diagnosis of PV systems based on machine learning approach using random forest classifier," *Energy Convers. Manage.*, vol. 301, Feb. 2024, Art. no. 118076.

[5] F. Rafique, L. Fu, and R. Mai, "End to end machine learning for fault detection and classification in power transmission lines," *Electric Power Syst. Res.*, vol. 199, Oct. 2021, Art. no. 107430.

[6] S. Banik, S. K. Saha, T. Banik, and S. M. M. Hossain, "Anomaly detection techniques in smart grid systems: A review," in *Proc. IEEE World AI IoT Congr.*, Jun. 2023, pp. 0331–0337.

[7] K. Chen, J. Hu, and J. He, "Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, Jul. 2017, pp. 1–10.

[8] A. Harish, A. Prince, and M. V. Jayan, "Fault detection and classification for wide area backup protection of power transmission lines using weighted extreme learning machine," *IEEE Access*, vol. 10, pp. 82407–82417, 2022.

[9] G.-Y. Li and G. W. Chang, "A multiple anomaly detection scheme and random forest algorithm for identifying the incipient faults in transmission network," in *Proc. IEEE Power Energy Soc. Gen. Meeting (PESGM)*, Jul. 2022, pp. 01–05.

[10] L. Yavuz, A. Soran, A. Önen, X. Li, and S. M. Muyeen, "Adaptive fault detection scheme using an optimized self-healing ensemble machine learning algorithm," *CSEE J. Power Energy Syst.*, vol. 8, no. 4, pp. 1145–1156, Jul. 2022.

[11] B. P. Tangirala, A. Yadav, and B. K. Chaitanya, "Bus-bar fault detection and classification using fast S-transform and artificial neural networks," in *Proc. 1st Int. Conf. Power Electron. Energy (ICPEE)*, Jan. 2021, pp. 1–6.

[12] J. U. Hassan and I. Fareed Nizami, "Machine learning algorithm analysis for detecting and classification faults in power transmission system," in *Proc. 2nd Int. Conf. Digit. Futures Transformative Technol.*, May 2022, pp. 1–5.

[13] W. Elmasry and M. Wadi, "Detection of faults in electrical power grids using an enhanced anomaly-based method," *Arabian J. for Sci. Eng.*, vol. 47, no. 11, pp. 14899–14914, Nov. 2022.

[14] M. O. F. Goni, M. Nahiduzzaman, M. S. Anower, M. M. Rahman, M. R. Islam, and M. Ahsan, "Fast and accurate fault detection and classification in transmission lines using extreme learning machine," *e-Prime Adv. Electr. Eng., Electron. Energy*, vol. 3, May 2023, Art. no. 100107.

[15] C. F. Mbey, V. J. Foba Kakeu, A. T. Boum, and F. G. Y. Souhe, "Fault detection and classification using deep learning method and neuro-fuzzy algorithm in a smart distribution grid," *J. Eng.*, vol. 2023, Jun. 2023, Art. no. e12324.

[16] A. S. Alhanaf, H. H. Balik, and M. Farsadi, "Intelligent fault detection and classification schemes for smart grids based on deep neural networks," *Energies*, vol. 16, no. 22, p. 7680, Nov. 2023.

[17] M. Ganjkhani, A. Gholami, J. Giraldo, A. K. Srivastava, and M. Parvania, "Multi-source data aggregation and real-time anomaly classification and localization in power distribution systems," *IEEE Trans. Smart Grid*, vol. 15, no. 2, pp. 2191–2202, Mar. 2024.

[18] A. Jaiswal, S. Chandra, A. Priyadarshi, S. Sachan, and S. Deb, "Reinforcement learning based islanding detection technique in distributed generation," *Energy Rep.*, vol. 9, pp. 6006–6019, Dec. 2023.

[19] S. S. Mohapatra, M. K. Maharana, A. Pradhan, P. K. Panigrahi, and R. C. Prusty, "Detection and diagnosis of islanding using artificial intelligence in distributed generation systems," *Sustain. Energy, Grids Netw.*, vol. 29, Mar. 2022, Art. no. 100576.

[20] A. Hussain, C. Kim, and S. Admasie, "An intelligent islanding detection of distribution networks with synchronous machine DG using ensemble learning and canonical methods," *IET Gener., Transmiss. Distrib.*, vol. 15, no. 23, pp. 3242–3255, Dec. 2021.

[21] J. R. Reddy, A. Pandian, and C. R. Reddy, "An efficient learning based RFMFA technique for islanding detection scheme in distributed generation systems," *Appl. Soft Comput.*, vol. 96, Nov. 2020, Art. no. 106638.

[22] F. Mumtaz, K. Imran, H. Rehman, and S. B. A. Bukhari, "1D recursive median filter based passive islanding detection strategy for grid-connected distributed generations network," *IET Renew. Power Gener.*, vol. 17, no. 7, pp. 1731–1746, May 2023.

[23] S. Sarangi, B. K. Sahu, and P. K. Rout, "Detection and classification of islanding by using variational mode decomposition and adaptive multi-kernel based extreme learning machine technique," *Sustain. Energy, Grids Netw.*, vol. 30, Jun. 2022, Art. no. 100668.

[24] A. Arif, K. Imran, Q. Cui, and Y. Weng, "Islanding detection for inverter-based distributed generation using unsupervised anomaly detection," *IEEE Access*, vol. 9, pp. 90947–90963, 2021.

[25] H. Mohamad, A. N. Ab Salam, N. Md Razali, N. A. Salim, and Z. Mat Yasin, "A new islanding detection technique based on passive parameter using a combination of artificial neural network and evolutionary programming algorithm," *J. Electr. Electron. Syst. Res.*, vol. 18, no. APR2021, pp. 1–8, Apr. 2021.

[26] A. Hussain, A. Mehdi, and C.-H. Kim, "A communication-less islanding detection scheme for hybrid distributed generation systems using recurrent neural network," *Int. J. Electr. Power Energy Syst.*, vol. 155, Jan. 2024, Art. no. 109659.

[27] R. Nale, M. Biswal, N. Kishor, J. Merino, A. Perez-Basante, and J. E. Rodriguez-Seco, "Real-time analysis of islanding detection scheme developed for AC microgrid system," *Electric Power Syst. Res.*, vol. 226, Jan. 2024, Art. no. 109926.

[28] S. Mishra, R. K. Mallick, D. A. Gadanayak, and P. Nayak, "A novel hybrid downsampling and optimized random forest approach for islanding detection and non-islanding power quality events classification in distributed generation integrated system," *IET Renew. Power Gener.*, vol. 15, no. 8, pp. 1662–1677, Jun. 2021.

[29] K. Chatterjee, A. De, and P. K. Saha, "Application of computer intelligence method to detect islanding phenomena in microgrid network," in *Proc. IEEE Appl. Signal Process. Conf. (ASPCON)*, Oct. 2020, pp. 277–282.

[30] T. S. Menezes, D. V. Coury, and R. A. S. Fernandes, "Islanding detection based on artificial neural network and S-transform for distributed generators," in *Proc. IEEE Milan PowerTech*, Jun. 2019, pp. 1–6.

[31] V. L. Merlin, R. C. Santos, A. P. Grilo, J. C. M. Vieira, D. V. Coury, and M. Oleskovicz, "A new artificial neural network based method for islanding detection of distributed generators," *Int. J. Electr. Power Energy Syst.*, vol. 75, pp. 139–151, Feb. 2016.

[32] A. Hussain, S. Mirza, and C.-H. Kim, "Islanding detection and classification of non-islanding disturbance in multi-distributed generation power system using deep neural networks," *Electric Power Syst. Res.*, vol. 224, Nov. 2023, Art. no. 109807.

[33] A. Gholami, A. Tiwari, C. Qin, S. Pannala, A. K. Srivastava, R. Sharma, S. Pandey, and F. Rahmatian, "Detection and classification of anomalies in power distribution system using outlier filtered weighted least square," *IEEE Trans. Ind. Informat.*, vol. 20, no. 5, pp. 7513–7523, May 2024.

[34] K. Chen, J. Hu, and J. He, "Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 1748–1758, May 2018.

[35] S. Raza, H. Mokhlis, H. Arof, J. A. Laghari, and H. Mohamad, "A sensitivity analysis of different power system parameters on islanding detection," *IEEE Trans. Sustain. Energy*, vol. 7, no. 2, pp. 461–470, Apr. 2016.

[36] O. T. Ibitoye, M. O. Onibonoje, and J. O. Dada, "Machine learning based techniques for fault detection in power distribution grid: A review," in *Proc. 3rd Int. Conf. Electr. Eng. Informat. (ICon EEI)*, Oct. 2022, pp. 104–107.

[37] F. M. Shakiba, S. M. Azizi, M. Zhou, and A. Abusorrah, "Application of machine learning methods in fault detection and classification of power transmission lines: A survey," *Artif. Intell. Rev.*, vol. 56, no. 7, pp. 5799–5836, Jul. 2023.
[38] B. K. Chaitanya, A. Yadav, and M. Pazoki, "Reliable islanding detection scheme for distributed generation based on pattern-recognition," *IEEE Trans. Ind. Informat.*, vol. 17, no. 8, pp. 5230–5238, Aug. 2021.

SADULLAH CHANDIO was born in Larkana, Sindh, Pakistan, in 1985. He received the B.E. and M.E. degrees from the Electrical Engineering Department, Quaid-e-Awam University of Engineering, Science and Technology, in 2007 and 2018, respectively. He is currently an Assistant Professor with the Quaid-e-Awam University of Engineering, Science and Technology. His research interests include power system protection, islanding detection, and power quality.

JAVED AHMED LAGHARI (Senior Member, IEEE) received the bachelor's degree in electrical engineering from BUET Khuzdar, Pakistan, in 2007, and the master's and Ph.D. degrees in electrical engineering from the University of Malaya, Malaysia, in 2012 and 2015, respectively. He joined as a Lecturer with the Quaid-e-Awam University of Engineering, Science and Technology (QUEST), Nawabshah, Sindh, Pakistan, in 2008. In 2015, he joined back his university as an Assistant Professor. He is currently an Associate Professor with the Department of Electrical Engineering, QUEST. He has more than 30 different publications in various ISI indexed based journals and conferences. His main research interests include intelligent power system control, power system optimization, islanding operation in distributed generation and load shedding techniques, and smart grid.

MUHAMMAD AKRAM BHAYO was born in Khairpur, Sindh, Pakistan, in 1980. He received the B.E. degree in electrical from QUEST, Nawabhshah, Pakistan, in 2004, and the M.Sc. degree in electrical and electronic engineering from Universität Duisburg-Essen, Germany, in 2013, and the Ph.D. degree in electrical engineering from Universiti Teknologi Malaysia (UTM), Malaysia, in 2021. Since 2006, he has been with the Department of Electrical Engineering, QUEST, where he is currently an Assistant Professor. His current research interests include modeling and simulation of wind energy conversion systems, with focus on implementation of adaptive neuro-fuzzy inference system (ANFIS) based controllers in wind turbine emulator.

MOHSIN ALI KOONDHAR was born in Nawabshah, Sindh, Pakistan, in 1985. He received the B.E. and M.E. degrees from the Electrical Engineering Department, Quaid-e-Awam University of Engineering, Science and Technology, in 2007 and 2016, respectively. He is currently a Lecturer with the Quaid-e-Awam University of Engineering, Science and Technology. His research interests include the control of dc and ac machines, renewable energy, and programmable logic controllers. He has served as a Reviewer for *IET Generation, Transmission and Distribution* journal, and *IIUM Engineering Journal*, Malaysia.

YUN-SU KIM (Senior Member, IEEE) received the B.S. and Ph.D. degrees in electrical engineering from Seoul National University, Seoul, South Korea, in 2010 and 2016, respectively. He was with Korea Electrotechnology Research Institute (KERI) as a Senior Researcher, from 2015 to 2017. He joined the Faculty of Gwangju Institute of Science and Technology (GIST), in 2018, where he is currently an Associate Professor with the Graduate School of Energy Convergence. His research interests include distribution networks, distributed energy resources, microgrid, artificial intelligence, and wireless power transfer. He was the Director of Korean Society for New and Renewable Energy and Korean Institute of Electrical Engineers. He has been an Associate Editor of IEEE TRANSACTIONS ON SUSTAINABLE ENERGY, since 2023.

BESMA BECHIR GRABA received the master's degree in energy engineering from the National Engineering College of Monastir, Tunisia, in 1999, and the Ph.D. degree specializing in the study of heat and mass transfer in forced turbulent convection within a horizontal channel from the National Engineering College of Monastir, in 2006. Currently, she holds the position of an Assistant Professor with the College.

EZZEDDINE TOUTI received the B.S. degree in electrical engineering from the Higher National College of Engineers of Tunis, Tunisia, in 1997, the master's degree in electrical engineering from the National College of Engineers of Tunis, Tunisia, in 2005, and the dual Ph.D. degree in induction generators wind turbine, power quality and electric drives from the National College of Engineers of Monastir, Tunisia, and Artois University, France, in 2013. In 2005, he joined the Laboratory of Industrial Systems Engineering and Renewable Energies (LISIER), University of Tunis, Tunisia, as a Researcher Faculty Member. Currently, he is an Associate Professor with the College of Engineering, Northern Border University. His research interests include renewable energy, control of electrical systems smart grid, electric vehicle, and power electronic.