



Detection of Faults in Electrical Power Grids Using an Enhanced Anomaly-Based Method

Wisam Elmasry¹ · Mohammed Wadi¹

Received: 23 October 2021 / Accepted: 29 March 2022 / Published online: 19 July 2022
© King Fahd University of Petroleum & Minerals 2022

Abstract

The increasing demand on electrical power consumption all over the world makes the need of stable and reliable electrical power grids is indispensable. However, one of hostile obstacles which delays reaching out to that desired goal is occurrence of faults. Despite to fact that dozens of studies have been put forward to detect electrical faults, these studies still suffer from several downsides such as validation and automation. In this paper, an electrical fault detection system based on the concept of anomaly detection is presented. The main salient advantages of the proposed system are overcoming the limitations of existed counterpart systems and its compatibility with real-world power grids. To enhance the performance of the proposed system, two vital stages are involved in its design prior to training, namely, data preprocessing and pre-training. Whereas the former is to prepare raw signals to be modeled, the latter is dedicated for model's hyperparameter selection using the particle swarm optimization metaheuristic. Moreover, two well-known anomaly detection models, namely, One-Class Support Vector Machines and principal component analysis are utilized to validate the proposed system as well as real-time data (VSB dataset) are used to train and test models. Finally, the experimental results and discussion emphasize that there is a performance improvement in detecting of electrical faults when using the proposed system.

Keywords Anomaly detection · Electrical faults · Signal filtering and decomposition · Extraction of features · Selection of hyperparameters · Metaheuristic optimization methods

1 Introduction

For the time being, we live in the era of widespread electrical power grids, whereby the demand for electrical energy increases all over the world. This increment of demand does not only make a rapid growth for the size of electrical power grid, but it also makes the electrical power grid itself more complicated [1]. In such a case and for avoiding economic losses and reducing maintenance cost, keeping the electrical power grid in a stable state for the longest possible period is the most desired aspiration of a power grid's administrators [2]. One of well-known conditions, that lead to instability in the electrical power grids, is the unexpected occurrence of faults in transmission or distribution power lines [3].

The electrical faults are sudden changes in both current and voltage values of the electrical signal to abnormal levels that deviate extremely from what normally is expected. Regardless to the cause of electrical faults which can be generated by human, environment or equipment, they have very detrimental effects on electrical network and devices [4]. In other words, once the electrical fault is occurred, a high current propagates through the electrical network causing a damage to equipment, cutting-off service, or even losing lives [5]. Thereby, there is a dire need for developing an effective and early detection systems to cope with such faults [6].

Due to the importance of electrical fault detection in power grids, dozens of studies have been published and many methods have been proposed as well [7]. Despite that, it was reported that the existed electrical fault detection systems are subjected to two critical flaws in real-world power grids regarding their automation and validation [8]. Automation refers to the methodology which is used to detect faulty signals. The vast majority of the previous studies adopted the binary classification methods and neglected the signal preprocessing [6,7]. Validation points to the data which is

✉ Wisam Elmasry
wisam.elmasry@izu.edu.tr

Mohammed Wadi
mohammed.wadi@izu.edu.tr

¹ Electrical and Electronics Engineering Department, Istanbul Sabahattin Zaim University, 34303 Istanbul, Turkey



utilized to validate the proposed methods. The overwhelming amount of previous studies used simulated data for the electrical signals by suggesting some scenarios [3,8]. Sections 1.1 and 1.2 discuss these limitations and put forward alternative solutions.

1.1 Anomaly-Based Detection

Indeed, faulty signals in real-world power grids are not common comparing to normal ones. Based on several theoretical studies, only 5% of electrical signals are faults, i.e., faults in power transmission lines are considerably rare [9]. Thus, adopting binary classification methods which suppose that the two classes have reasonable amount of samples to detect faults is inaccurate choice [3,6–8]. Instead, someone might utilize another approach to detect faulty signals by concentrating on the overwhelming normal samples and their characteristics. This approach is known as anomaly-based detection and it endeavors to learn from data of a class that has the majority number of samples and after that it identifies anomalous points (outliers) which stay away from the majority samples [10,11]. This scenario is typical to electrical fault detection problem, that is, anomaly detection is the fittest approach to be used when dealing with electrical fault detection task [12].

To achieve the concept of anomaly detection, some anomaly fault detection approaches are proposed [13]. The working idea behind these models that they are trained only on the majority samples (normal) to construct a profile of these samples which contains their properties. Thereafter, the trained anomaly-based models can detect anomalous samples (faulty) if they are located away from the properties in the built profile [14]. In this paper, two anomaly detection models are exploited based on two approaches the OC-SVM and PCA. The OC-SVM approach is a variant of the conventional SVM model in such a way that the OC-SVM model is compatible with the concept of anomaly-based detection [15]. Specifically, the OC-SVM is only trained on the unfaulty samples to draw the boundaries of the smallest hypersphere that groups all these samples inside it or on its boundaries. Hence, any unseen data that deviate away from this hypersphere will be considered as faulty samples. Furthermore, the ν hyperparameter of OC-SVM model adjusts the distance between the boundaries of the hypersphere and the outliers [16], and the Epsilon (ϵ) hyperparameter controls the stopping criteria of the OC-SVM model [15]. The same working concept of anomaly detection is also applied when using the PCA model. Practically, the PCA model compacts the space of the normal samples into smallest feature subset that can represent the normal samples accurately. Then, the PCA uses this feature subset to decide whether unseen data are outliers or not. Likewise, the PCA approach mainly

depends on three hyperparameters to tune its performance such as oversampling, rank, and center [17].

1.2 VSB Dataset

Validation of proposed systems is a cornerstone in machine learning applications to prove the efficiency of these systems. Basically, the validation process can be handled by using some public benchmark datasets. In electrical fault detection problem, most of previous studies exploited simulated data for validation process. The simulated data in the previous works contain the values of current and voltage of a simulated power grid. They obtained these values by using some emulators under specified conditions and within a fixed period of time. However, many theoretical studies in electrical fault detection area stated that the simulated data are insufficient and do not reflect the reality of electrical faults [3,6–8].

To overcome this drawback, the ENET center of research at the University of Technical Ostrava created a fresh, and real dataset for fault detection problems in power systems [18]. It is named as VSB dataset and it is available online at Kaggle website [19]. They utilized a new equipment for recording a voltage signal every 20 milliseconds that crossed transmission lines of a real 3-phase electrical power grid (50 Hz) in Ostrava city. Thereafter, they sampled each voltage signal into 800K voltage measurements and stored them as 1-byte integer values. The total number of samples in the VSB dataset is 8712 with each sample has 800K voltage measurements and the class label to indicate whether the signal is normal (0) or faulty (1). Since the VSB dataset is actual data, it clearly demonstrates faults in real-world power systems as it contains 8187 normal signals and 525 faulty signals. This realistic structure of the VSB dataset confirms the need of using anomaly detection approach instead of using conventional binary classification methods.

1.3 Contributions

Based on aforementioned introduction, this paper includes three main contributions to be in line with aroused issues.

- To get along with the concept of fault detection in real-world power system, a novel anomaly fault detection model is proposed.
- To boost the performance of the proposed system, data preprocessing and pre-training stages are involved as preliminary procedures of electrical signals.
- To ensure the reliability of the proposed system, two well-known anomaly fault detection approaches are validated on the VSB dataset.

The rest of the paper is structured as follows. Section 2 reviews a wide range of recent works in the field of electri-

cal fault detection. The full details of the proposed system and its methodology are introduced in Sect. 3. In Sect. 4, the experimental results are presented along with a description of the used evaluation measures. Then, Sect. 5 discusses the obtained results from different aspects. The conclusions, limitations, and suggested future work are drawn in Sect. 6.

2 Literature Review

Due to the significance of detecting faults in power networks, researchers have focused on developing fault detection systems during the last decade. However, most of their works employed one or more binary classification classifiers on simulated data. For instance, the artificial neural network (ANN) [20–24], Concurrent Fuzzy-Logic (CFL) [25], and support vector machine (SVM) [26,27] were widely-used in the literature. The methodology of their works was to simulate an electrical power grid using simulation platforms such as MATLAB Simulink and then to emulate faults in different phases. Afterwards, they recorded the voltage and current values and used Wavelet transform to sample these values to be stored in a dataset. Finally, they trained and tested the used binary classifiers on the simulated dataset. As mentioned in Sect. 1, the simulated dataset was the weakness point of their works [3,6].

To combat this limitation, some recent studies in the literature used a modern and real dataset for fault detection such as the VSB dataset [19]. Dong et al. utilized a new method known as the Seasonal and Trend decomposition using Loess (STL) to preprocess the signals in the VSB dataset [28]. Then, they detected the activities of Partial Discharge (PD) by using a Radial Basis Function (RBF)-based SVM classifier. In a similar article regarding PDs, a Discrete Wavelet Transform (DWT)-based Long Short-Term Memory network (LSTM) was proposed to process signals and detect Insulated Overhead Conductor (IOC) [29]. Wadi et al. exploited two anomaly detection-based models, namely, the OC-SVM and PCA to detect faults in a reduced version of the VSB dataset [30]. Two machine learning-based techniques were used along with four binary classification-based classifiers to detect faults in a balanced version of the VSB dataset [31].

Moreover, there are another papers that explored fault detection problem in various electrical power generation and transmission sites [32]. Biswas et al. investigated the impact of transmission lines on distance relay performance by proposing a new fault detection scheme and fault classification method as well [33]. They validated their fault detection and classification methods on two real-time wind farm systems as well as on simulated data. In a study [34], a hybrid intrusion detection system, which based on anomaly detection, was introduced to detect possible attacks. They utilized K-means and SVM methods in the design of the

introduced system. Likewise, an anomaly-based intrusion detection system was suggested and enhanced by using the Grey Wolf Optimization (GWO) to select the optimal features [35]. However, they only used a SVM classifier to measure the accuracy of their system.

Himeur et al. suggested two different schemes for detecting abnormality of smart power consumption [36]. The first scheme based on an unsupervised OC-SVM model, whereas the second scheme relied on a supervised K-Nearest Neighbours (KNN) classifier. The DWT and Naive Bayes (NB) classifier were combined together to classify the type of electrical faults in transmission lines using simulated dataset [37]. Chen et al. proposed an improved detection and diagnosis approach for incipient electrical faults using a convolutional neural network (CNN) model based on data-driven [38]. Finally, an anomalous energy consumption detection system, based on micro-moments feature extractor, was presented and validated using a Deep Neural Network (DNN) classifier [39].

3 Proposed Methodology

The main task of the suggested anomaly-based detection system in this study is to identify the abnormal voltage signals associated with the energy transmission lines. It is difficult for the anomaly-based detection system to deal with the collected raw data voltage signals. Therefore, these data need to be handled via many phases, such as preprocessing and splitting into chunks. Data preprocessing refers to filtering the collected voltage measurements from noise, whereas data splitting into chunks means decomposing data into segments. The next stage is the feature extraction by which the voltage measurements patterns are characterized; afterward, these extracted features are combined into a data record. Afterward, the data records are normalized. The PSO algorithm and the input signals' data records are employed to identify the best hyperparameter values of the utilized anomaly fault detection model within the pre-training phase. Subsequently, the performance of the best anomaly-based detection model is improved to construct an accurate profile for the normal signals by training it over the data records of these signals. Eventually, the qualified anomaly-based detection model can accurately differentiate between the normal and faulty signals. Figure 1 illustrates the steps of working-flow of the suggested anomaly-based detection system.

The proposed methodology in this article is designed to be characterized by clarity and simplicity. It contains four consecutive stages: data preprocessing, pre-training, training, and testing. Figure 2 displays the working mechanism of the proposed methodology.

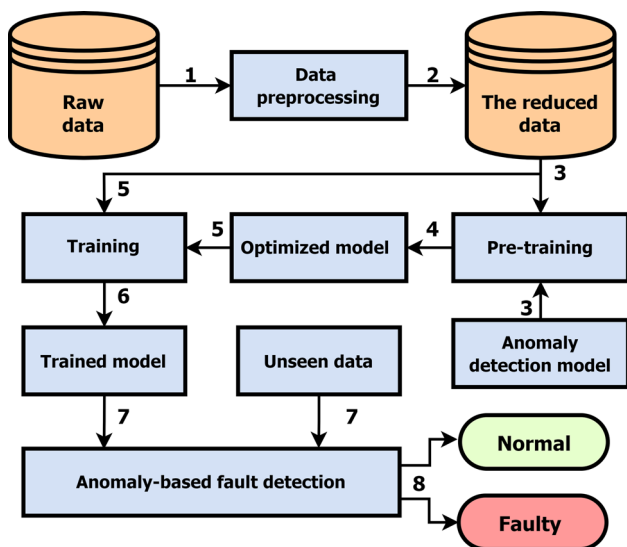


Fig. 1 Working-flow steps of the suggested anomaly-based detection system

3.1 Data Preprocessing Stage

In this stage, all the fundamental processes on the samples of the VSB dataset are executed. This stage is crucial since it organizes the data for examining and modeling by the utilized fault detection model.

3.1.1 Signal Denoising

Machine learning experts are generally concerned with using machine learning techniques on the dataset rather than gathering samples. Hence, they assume that the gathered data

are ready for analysis and that no additional tasks is needed. However, this is not the valid case because the gathered data, in general, are messy and include noise. There are many reasons for the presence of noise in the gathered data, such as measurement devices failure, environmental conditions, or other unpredictable events. Undoubtedly, there is no way to avoid the noise by collecting the data, and noise is unavoidable. On the one hand, data processing depending on noisy data without eliminating this noise can lead to poor data quality and analysis performance. This poorness can be illustrated by noticing that all the statistical descriptors, such as the average standard deviation, and variance values, are significantly sensitive to noise, leading to test failure or distorting the actual outcomes. Accordingly, the most practical key for noise is data filtering [40].

Noise is another face of anomaly irregularity in data as referred to in Sect. 1. The noise in data can be represented by the samples whose characteristics are significantly different from the majority of data samples. Unfortunately, generally, there is no specific way to specify noise. While with the support of the concept of noise described overhead, some statistical techniques can be employed to discover noise samples. Interquartile Range (IQR) is one of the popular noise filtering techniques. The real benefits of utilizing the IQR technique are not only because it does not rely on the data distribution, but also it is comparatively robust to the existence of noise compared to the other quantitative techniques [41].

The utilized VSB dataset is filtered from noise by scanning signal by signal. Voltage measurements copies of a specific signal are initially kept and categorized in ascendance way. Then, the first quartile $Q1$ (25%) of the signal measurements

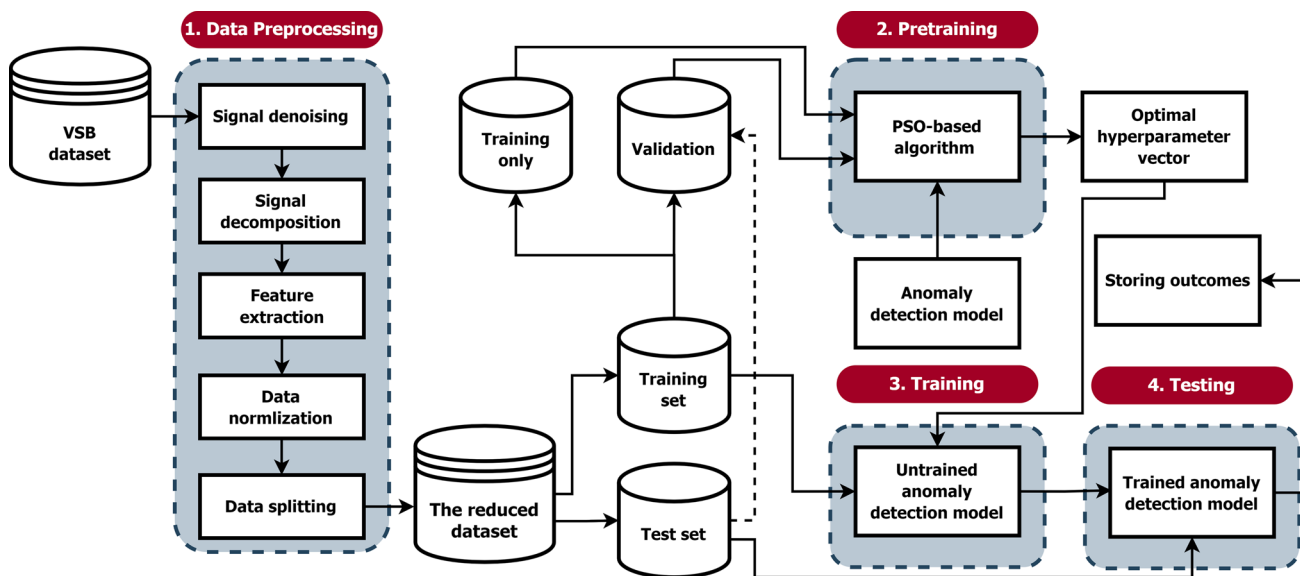


Fig. 2 Working mechanism of the proposed methodology

and the third quartile $Q3$ (75%) are computed. Then, the IQR value, the average of the signal (50%), is calculated.

$$IQR = Q3 - Q1 \tag{1}$$

The computed IQR is then multiplied by the adjustment factor k . The purpose of utilizing such a factor is to specify the strength of outliers. Statistically two commonly values of k , 1.5 and 3 are used [41]. The 1.5 value refers to weak outliers, whereas the 3 value refers to is the strong outliers in data. There are two classes of outliers for the electrical data: invalid measurements that negligibly vary from the normal values and noise measurements that excessively deviate from the normal values. Accordingly, the fault measurements must be maintained to conduct fault detection, while the noise measurements must be eliminated. Thus, the used k value in this study is selected to be 3.

$$Threshold = 3 * IQR \tag{2}$$

Subsequently, the *Threshold* value is manipulated to decide the boundaries of the noise measurements. The lower and upper noise boundaries can be calculated as follows:

$$Lower_Boundary = Q1 - Threshold \tag{3}$$

$$Upper_Boundary = Q3 + Threshold \tag{4}$$

Eventually, if the voltage measurements of a certain signal are less than the lower boundary or greater than the upper boundary, this signal will be eliminated from the dataset. The filtering process moves to the following signal and repeats the exact process until all dataset signals are successfully purified.

3.1.2 Signal Decomposition

Signal denoising refers to the process of removing the noise samples from the measurement signals. In this study, the used dataset after removing all noise samples and the i th signal remaining voltage measurements equals $(800,000 - l)$ are used in the following stages. l indicates the eliminated number of voltage noise measurements in the i th signal. Nevertheless, catching the faults signals after signal denoising is still challenging since their numbers are smaller than the unfaulty signals. Consequently, the signal decomposition process is crucial to get rid of this problem, [42].

The signal decomposition refers to dividing the surviving voltage measurements into smaller segments named chunks at which faults detection is less complex. Moreover, dividing the dataset into chunks shrink the range of voltage measurements which significantly improves the ability to detect faults [43]. Subsequently, the more chunks, the better performance is [43]. In this article, each signal data is split into four chunks,

such as 1, 2, 4, and 8, in order to examine the performance differences at each chunk. If M is the number of chunks, then the VSB dataset can be divided into M chunks, as follows.

$$chunk_size = ROUND((remaining_measurements_of_signal_i)/M) \tag{5}$$

where $chunk_size$ is the chunk size and $ROUND()$ is a function that changes a floating number into an integer.

$$Chunk_j^i = X_{[(j-1)*chunk_size]+1}^i, \dots, X_{[(j-1)*chunk_size]+chunk_size}^i$$

$$j = 1, 2, \dots, M \tag{6}$$

where X_d^i is the d th voltage measurement of the i th signal and $Chunk_j^i$ is the j th chunk of the i th signal.

3.1.3 Feature Extraction

Due to the huge size of the original VSB dataset, which is about 800k measurements in each signal. Even after performing the signal denoising and decomposition reduction processes, the data size is still massive, which is called the remaining data. These remaining data will be provided to the input of the machine learning model. Since these data are still very huge, this reduces the performance of most machine learning models [12].

The feature extraction technique is one of the most used techniques to overcome the dilemma of huge data. The main aim of the feature extraction is to decrease the dimensions of the feature space by selecting those that have a dominant effect on the fault detection model. In this article, the most effective 19-feature set from the remaining voltage samples are obtained individually per chunk. Then, these extracted features are merged with the "Class" tag to construct the data record per signal. This continues until all the voltage measurements are handled. The resultant dataset after feature extraction is called the reduced dataset. The features number is directly proportional to the chunk number of M and can be determined by the $(19 * M) + 1$ formula.

The extracted features from chunks per signal are, in general, statistical descriptors that can display a comprehensive picture of the data distribution. Besides, these features are numeric values as given below:

- Mean is the average of a collection of numbers and can be calculated by Equation (7).

$$Mean_j = \frac{\sum_{i=1}^{chunk_size} X_i^j}{chunk_size} \tag{7}$$

where $Mean_j$ refers to the average of the j th chunk and X_i^j is the i th voltage measurement of the j th chunk.

- Standard deviation refers to the degree of data scattering of a discrete set of numbers with reference to the mean value. A high standard deviation value indicates a high deviation of mong dataset numbers and vice versa.

$$Standard_deviation_j = \sqrt{\frac{\sum_{i=1}^{chunk_size} (X_i^j - \bar{X}_j)^2}{chunk_size}} \quad (8)$$

where $Standard_deviation_j$ and \bar{X}_j are the standard deviation and mean of the j th chunk of the signal, respectively.

- Maximum value refers to the highest value of the voltage measurements within a specific chunk.
- Minimum value refers to the smallest voltage measurements within a specific chunk.
- Percentile is a statistical criterion that refers to a specific percentage of scores that fall below that value [44]. For example, the p th percentile is a value where $p\%$ of the voltage measurements fall into. Five percentiles such as 1%, 25%, 50%, 75%, and 99% per chunk are employed and assessed in this study.

$$z = \lceil \frac{P}{100} \times chunk_size \rceil \quad (9)$$

where P is the value of percentage.

- Relative percentile is the degree of divergence of certain data from the average value. Seven relative percentiles like 0%, 1%, 25%, 50%, 75%, 99%, and 100% per chunk are estimated as follows.

$$\begin{aligned} P\%_Relative_Percentile_j \\ = P\%_Percentile_j - Mean_j \end{aligned} \quad (10)$$

where $P\%_Relative_Percentile_j$ and $P\%_Percentile_j$ are the $P\%$ relative percentile of the j th chunk per signal.

- Lower and Upper boundaries are the least and greatest bounds of the voltage measurements of a specific chunk and can be computed as follows.

$$Lower_Boundary_j = Mean_j - Standard_deviation_j \quad (11)$$

$$Upper_Boundary_j = Mean_j + Standard_deviation_j \quad (12)$$

where $Lower_Boundary_j$ and $Upper_Boundary_j$ are the lower and upper values of the j th chunk, respectively.

- Height is the different distance between the minimum and the maximum voltage measurement values of a specific chunk of the signal.

$$Height_j = Maximum_j - Minimum_j \quad (13)$$

where $Height_j$, $Maximum_j$, and $Minimum_j$ indicate the height, maximum, and minimum values of the j th chunk of the signal, respectively.

3.1.4 Data Normalization

Data normalization plays a crucial role in improving the homogeneity of data. In this study, for all data records within the compressed dataset, the features excluding the "Class" are converted into the range [0, 1] based on the normalization min-max method.

$$x_i = \frac{x_i - Min}{Max - Min} \quad (14)$$

where x_i stands for the value of feature of the i th data record within the compressed dataset, Min and Max stand for the minimum and maximum values per feature, respectively.

3.1.5 Dataset Splitting

In this stage, the compressed dataset received from the prior stage is divided based on the anomaly detection approach, by which the training set includes only unfaulty samples, while the testing set includes faulty and unfaulty samples. Accordingly, a training set of 6140 randomly-selected unfaulty samples (70.48%) is created using the holdout without replacement selection method. The remaining unfaulty (2047) and all faulty (525) samples are inserted in the testing set that reach a total of 2572 samples (29.52%). Table 1 explains the major features of the compressed dataset after the preprocessing data stage. Figure 3 depicts the proposed data preprocessing stage flowchart.

Table 1 Major features of the reduced dataset

Item	Value(s)
Released	2018
Total number of samples	8712
Classes number	2
Type of data	Numerical
# features/chunk	For 1-chunk 20 For 2-chunk 39 For 4-chunk 77 For 8-chunk 153
Training set structure	Normal=6140 Faulty=0 Total=6140
Test set structure	Normal=2047 Faulty=525 Total=2572

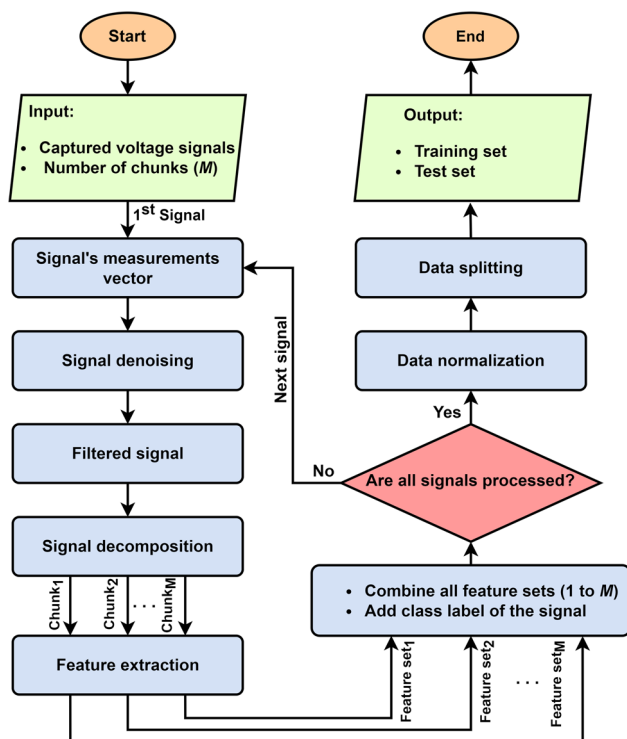


Fig. 3 Flowchart of the proposed data preprocessing stage and its sub-stages

3.2 Pre-training Phase

The primary duty of the pre-training stage is to improve the performance of the proposed detection model by specifying its optimal hyperparameter values that match the fault detection problem. Many metaheuristics optimization methods can be employed for hyperparameter selection, while the PSO algorithm presented by Elmasry et al. [13] has drawn high attraction due to its robustness, stability, and simplicity [14]. Figure 4 shows the operational diagram of the PSO method.

Fig. 4 Operational block of the PSO method for hyperparameters optimization

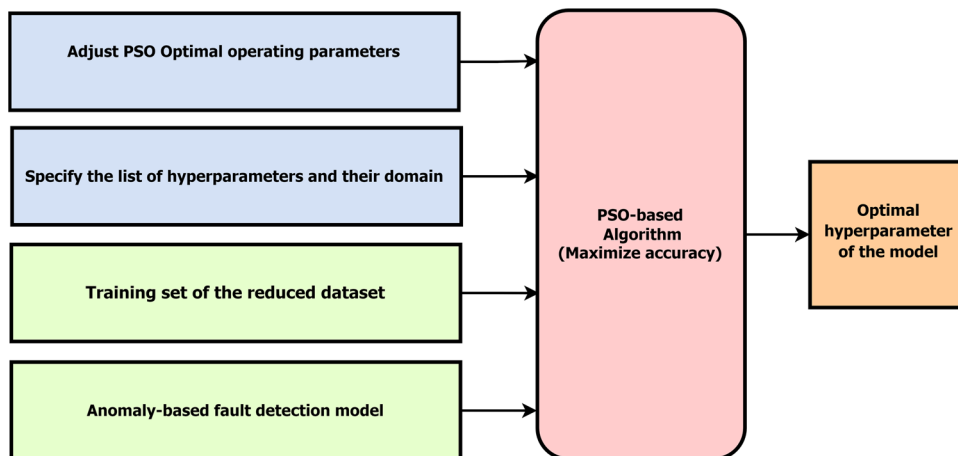


Table 2 A list of the PSO parameters, ranges, and used values [12,13]

Parameter	Range	Used value
Swarm size	[10, 200]	100
Minimum velocity	[0, 1]	0
Maximum velocity	[0, 1]	1
Acceleration factors	[0.7, 9.95]	2.68
Constant of inertia weight	[0.1, 1.5]	0.84
Maximum iteration number	[10, 120]	60
Termination criterion	[0.01, 0.1]	0.01

PSO algorithm plays a crucial role by optimizing the hyperparameter vector of the fault detection model that maximizes the model’s performance and accuracy. Hence, the initial step of the PSO algorithm is to adapt its parameters to the best operating ones. Table 2 depicts the central values of the PSO algorithm operating values. Table 2 illustrates the optimized values per PSO parameter in its advised range after performing a grid search. Besides, the parameter ranges in Table 2 are advised by many research works in the literature [12,13]. Then, the user can specify the number of hyperparameters of a model and the recommended ranges for each hyperparameter.

The following step begins with dividing the training set into two portions: the training-only and validation portions. 6850 normal samples are selected randomly for the training-only portion based on the hold-out without replacement selection method. 375 (250 normal+125 faulty) samples are also randomly selected for the validation portion. Afterward, the PSO algorithm attempts all possible combinations to specify the best model hyperparameters within their defined domains. As the hyperparameters vector adjusts the model, the training-only and validation portions will be provided to the training and testing, respectively. Consequently, the model accuracy value will be calculated and held.

Table 3 The best hyperparameters per fault detection model

The utilized model	Parameter	Domain	Best value
OC-SVM	ν	[0.001, 0.1] Step=0.01	0.1
	ϵ	[0.001, 0.1] Step=0.01	0.001
PCA	Rank	[2, 10] Step=2	2
	Oversampling	[2,10] Step=2	4
	Center	{True,False}	False

The third and final stage of the PSO method begins when the termination criterion is fulfilled. Then, the PSO method generates the best hyperparameter values, which improves the precision of the utilized fault detection model. Table 3 shows the hyperparameter ranges and the best values of the employed models after the pre-training stage is completed.

3.3 Training and Testing Phases

The training stage is initiated after the optimized fault detection model is created based on the best hyperparameter vector and trained over the whole training set. Then, the optimized and trained fault detection model is tested using the testing set. Ultimately, the obtained results are kept for the following processing stages.

4 Experimental Results

The designing and execution of our experiments are handled using the Azure Machine Learning (AML) studio [45]. AML is a distinctive cloud-based tool that can supply users freely with various modules for modeling, training, testing, and analysis machine learning methods such as the OC-SVM and PCA models [46]. Moreover, the AML dedicates enormous storage and computation resources for the users to carry out their experiments [47]. On the other hand, both data pre-processing and pre-training stages are programmed using the Python version 3.10.2¹ augmented by the NumPy package.² The following subsections describe the evaluation measures and elaborate the experimental results.

4.1 Evaluation Measures

Although anomaly-based detection is used in the proposed approach, but the result of the testing phase is in the form of a binary classification. This can be explained by the fact that we basically have two classes in the electrical fault detection, i.e., "normal" and "faulty". After the testing phase is finished, the outcome consists of four values, namely, True Negative (TN), False Negative (FN), True Positive (TP), and False Positive

(FP). The terms "Negative" and "Positive" here means "normal" and "faulty" classes, respectively. On the other hand, the terms "False" and "True" corresponds to misclassification and true classification of samples in the testing set, respectively. For instance, TP is merely the number of "faulty" samples which is classified correctly. Furthermore, the values of TP, TN, FP, and FN are the basis of computing many evaluation measures that can be exploited to evaluate the performance of the used models. In this paper, eight well-known evaluation measures are utilized to assess the performance of the OC-SVM and PCA models, namely, accuracy, precision, recall, F1-score, False Alarm Rate (FAR), Specificity, False Negative Rate (FNR), Matthews Correlation Coefficient (MCC) [48,49]. The list of equations, of how these evaluation measures are calculated, is given below.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (18)$$

$$FAR = \frac{FP}{FP + TN} \quad (19)$$

$$Specificity = \frac{TN}{TN + FP} \quad (20)$$

$$FNR = \frac{FN}{FN + TP} \quad (21)$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FN) \times (TP + FP) \times (TN + FP) \times (TN + FN)}} \quad (22)$$

4.2 Performance Analysis

The performance of the proposed anomaly-based detection model is evaluated utilizing the evaluation measures explained in Sect. 4.1. The accuracy of any model increases by higher and lower of the true and false classifications, respectively. Table 4 presents the computed evaluation measures in percentage per fault detection model. The bold and underlined values to simplify the readability of Table 4 are used to describe the best results of the employed models at the same experiment and among all experiments, respectively. Moreover, the 'Original' and 'Enhanced' columns in Table 4 illustrate the outcomes of the utilized models without

¹ <https://www.python.org>

² <https://www.numpy.org>

Table 4 The obtained results of the empirical experiments

Model	Evaluation measures	Original	Enhanced			
			Chunks Number			
			based-1	based-2	based-4	based-8
OC-SVM	Accuracy	76.51	84.18	88.27	91.15	95.24
	Precision	65.37	79.54	83.39	86.41	95.03
	Recall	56.42	70.55	79.76	85.59	89.91
	F1-Score	60.76	74.16	81.04	86.63	92.41
	FAR	13.94	8.07	6.19	5.18	2.31
	Specificity	86.65	91.08	93.90	94.56	98.24
	FNR	43.67	29.67	20.11	14.50	10.18
PCA	MCC	43.17	63.42	72.27	79.43	88.81
	Accuracy	74.93	83.31	86.03	93.26	96.87
	Precision	61.12	77.72	80.77	91.82	97.37
	Recall	54.09	66.25	74.66	87.76	91.30
	F1-Score	57.99	71.94	77.64	89.22	94.92
	FAR	16.02	8.48	7.29	3.15	1.74
	Specificity	83.60	91.61	92.80	96.94	98.35
FNR	45.01	33.84	27.76	12.44	9.06	
MCC	38.48	59.31	66.49	84.50	91.43	

The bold and underline values show the best results of the used models in the same experiment and among all experiments, respectively

and with using the proposed anomaly-based fault detection approach.

It can be observed from the obtained results in Table 4 that the suggested fault detection model significantly improved the performance of all models and made them qualified compared to the same models without using the proposed model. For example, the recall and FAR metric values based on the two approaches OC-SVM and PCA for 1-chunk are (70.55%,8.07%) and (66.25%,8.48%), respectively, which means an improvement by 14% and 8%, respectively regarding the original case. These satisfactory outcomes can be justified by the influence of data preprocessing and pre-training stages that enable anomaly fault detection models to identify faults accurately. In addition, by increasing the number of chunks, the accuracy measures for all models are greatly enhanced. For instance, based on approaches OC-SVM and PCA for 8-chunks, the recall and FAR metrics are significantly improved (89.91%,2.31%) and (91.30%,1.74%).

Unfortunately, the more the number of chunks, the more complexity and computation are. This makes the detection model structure more complicated. The model complexity increases with the increase of the chunks numbers originating from the fact that expanding the number of chunks leads to many extracted features per chunk, which can be defined as a linear relationship between the complexity and the chunk numbers. In the case of high-dimensional features, machine learning becomes unable to treat like these data, and deep learning is indispensable [9,13].

Comparing the performance of the two approaches, OC-SVM and PCA, it can be noticed that the OC-SVM approach is outstanding to the PCA approach at chunk numbers less than four, while the PCA model exceeded the OC-SVM approach at the chunk numbers equal to or bigger than four. These results can be attributed to the tendency of the PCA approach to operate efficiently in the high-dimensional spaces to connect the smallest group of features to identify the features of the most occurrence samples. On the one hand, this characteristic of the PCA approach is not available in the SVM model. Despite that, all the utilized models offer satisfactory electrical fault detection results.

To put it together, the suggested fault detection model adequately enhanced the detection problem; however, it faces a primary challenge when the feature space is increased [12]. Due to space restrictions, some of the evaluation measures mentioned above are shown in Fig. 5. It depicts a graphical comparison between the employed approaches at all utilized chunk numbers. There is also another way to compare the performance of the utilized models called the Critical Difference Diagram (CDD) [50]. Figure 6 presents the CDD of the utilized models at all used numbers of chunks (M). Besides, the Critical Difference (CD) is drawn at the top of Fig. 6 with a value equals 6.8675.

Another challenge that may encounter the fault detection systems is the ability of these systems to identify the faults regardless of their types. Accordingly, the performance of detection systems can be measured by the capability of these systems to operate with all types of faults accurately. The

Fig. 5 Visualization of four evaluation measures based on number of chunks for the used models: (a) OC-SVM, and (b) PCA

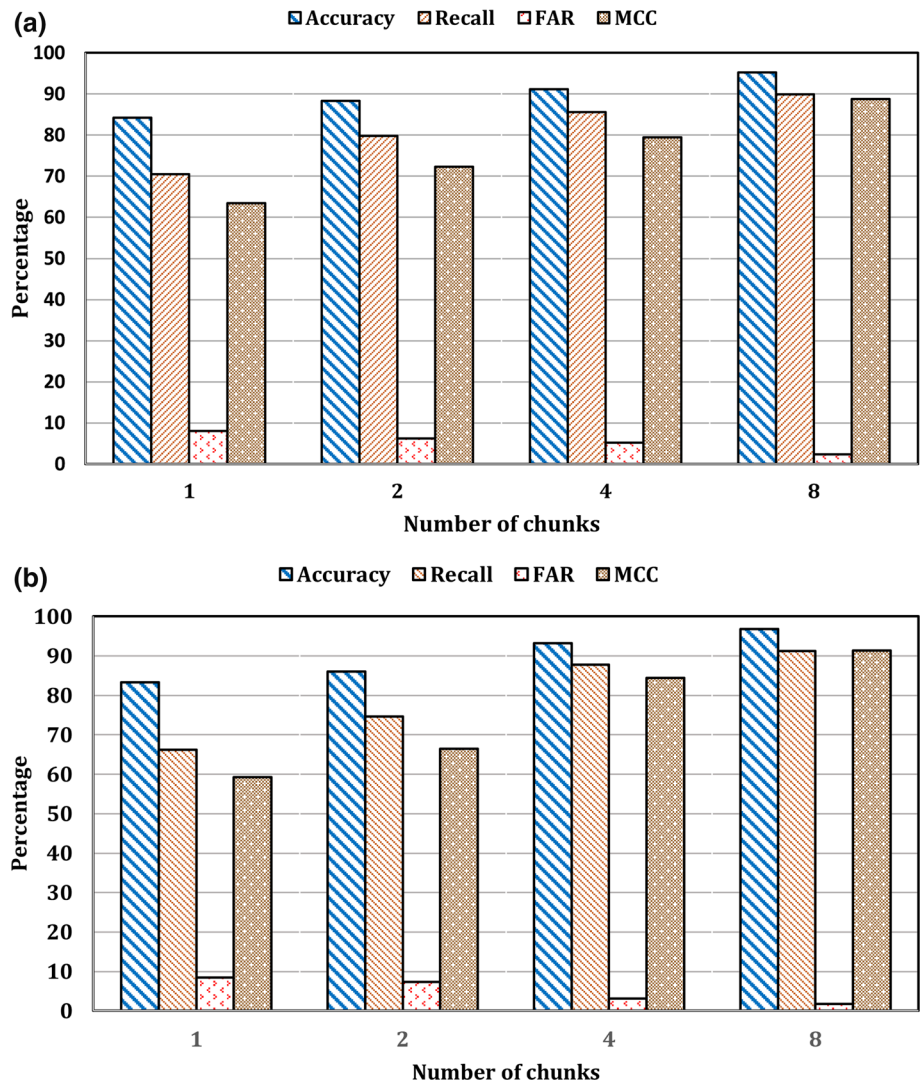
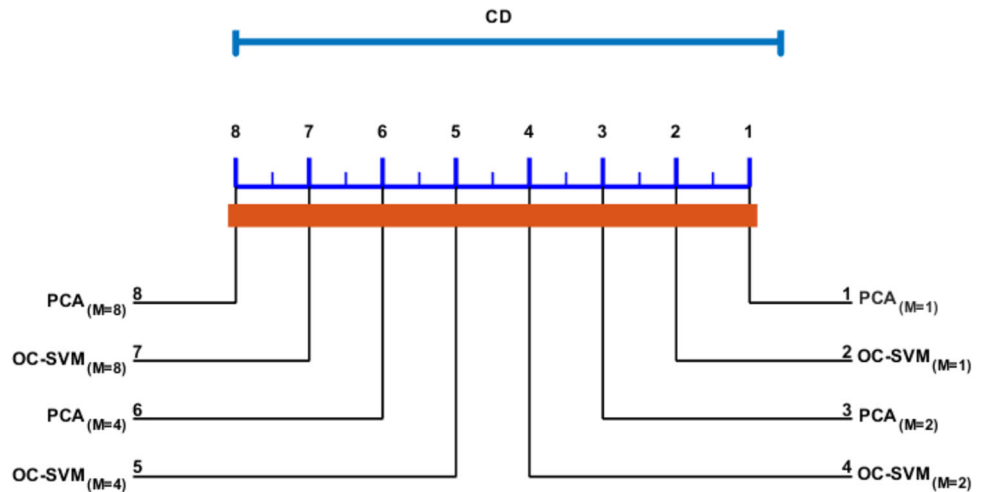


Fig. 6 Critical difference diagram for all models based on M



faults in power systems can be categorized into two main types: symmetrical (balanced) and unsymmetrical (unbalanced) faults [3]. Symmetrical faults are also divided into two types: 3-phase (LLL) fault and 3-phase to the ground (LLLG) fault [8]. These faults are the most severe fault that causes a tremendous current within the power grid, though it infrequently occurs [6] with occurrence ranges from 2% to 5% [7].

The unsymmetrical faults, on the other hand, are more dominant and less hazardous than the symmetrical faults [8]. Unsymmetrical faults consist of three types: line-to-ground (LG), line-to-line (LL), and double line-to-ground (LLG) faults [7]. As reported in many studies, the probability of occurrence is (65%-70%), (5%-10%), and (15%-20%) for the LG, LL, and LLG faults, respectively [3]. Even though they are less dangerous than symmetrical faults, unbalanced currents cause increased temperature, which reduces the power system equipment life leading to early failure [6].

The used VSB dataset in this article clearly illustrates the ratio of symmetrical faults to the unsymmetrical faults. The dataset contains 525 faulty samples are divided into 4.5% and 95.5% symmetrical and unsymmetrical faults, respectively. The percentages of the unsymmetrical faults are also 69%, 19%, and 7.5% for LG, LLG, and LL, respectively. The percentage matching of faults in the VSB dataset with the published studies confirms the correctness and reliability of this dataset. Figure 7(a) displays the detection rate in percentage of symmetrical and unsymmetrical faults with and without utilizing the proposed detection model. It can be observed from the obtained results that the symmetrical faults detection based on the proposed model is improved by 60.32%, whereas the unsymmetrical faults detection are also improved by 52.15% compared to the case without using the proposed model. Besides, Fig. 7(b) shows the effect of chunks number in identifying all types of faults. As the number of the chunk is increased, the OC-SVM and PCA approaches provide better accuracy and higher performance in fault detection. Subsequently, it can be said that the proposed fault detection model is effective with all fault types in power systems.

4.3 ROC Analysis

It is essential to differentiate between different machine learning models in most cases. This process is called ranking and trade-off. For a fair comparison and to select the best (optimal) classifier, the models must be applied to the same dataset. One commonly used method is the ranking method using the Receiver Operating Characteristic (ROC) curves. The ROC curves are plots of the recall measure as a function of the FAR measure of a classifier [51]. The line from (0, 0) to (1, 1) divides the graph into two areas, upper and lower.

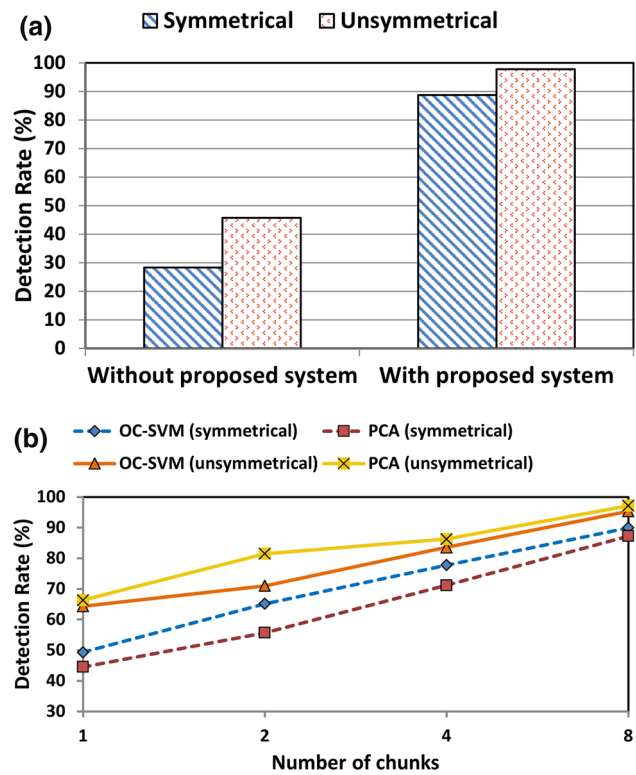


Fig. 7 Impact of using the suggested system in identifying symmetrical and unsymmetrical faults: (a) Detection rate of models with and without the suggested system, and (b) Detection rate of models based on chunks number

As the curve of the classifier approaches the upper left border of the plot, the performance is better. Figure 8 presents the ROC curves of the utilized models for different chunk numbers.

The Area Under the ROC Curve (AUC) is a quantitative metric that characterizes the performance of the utilized classifier [52]. It takes values within the period [0, 1]. As the AUC values approach the one, the more satisfactory the classifier is. Besides, the AUC can be calculated utilizing the formula (23) [53]. Table 5 shows the obtained AUC metric values.

$$AUC = \frac{1}{2} \times (Recall + Specificity) \tag{23}$$

5 Discussions

This Section discusses the obtained results in Sect. 4 utilizing different comparative analyses and many statistical measures, as follows:

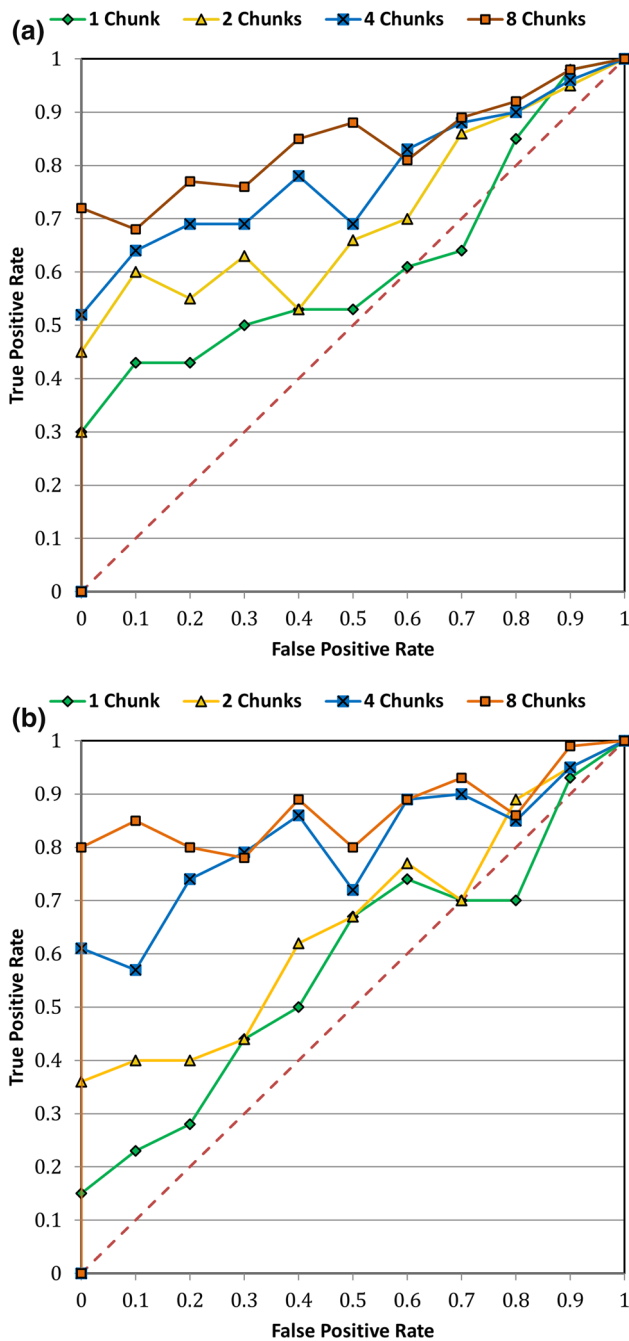


Fig. 8 ROC curves based on chunks number for: (a) OC-SVM, and (b) PCA

5.1 Stability and Sensitivity Analyses

In this subsection, many statistical tests are applied to verify the stability of the proposed fault detection model. Table 4 illustrates the Friedman test, which applied to the results to show their consistency. This test identifies the divergences between different repeated treatments [54]. The Friedman test characterizes by many advantages, such as generality and simplicity. Moreover, one of its outstanding characteristics is

Table 5 Values of AUC performance measure for ROC curves

Chunks number	Model	AUC
1	OC-SVM	0.8170
	PCA	0.7848
2	OC-SVM	0.8683
	PCA	0.8302
4	OC-SVM	0.9078
	PCA	0.9280
8	OC-SVM	0.9318
	PCA	0.9513

The bold and underline values show the best results of the used models in the same experiment and among all experiments, respectively

Table 6 Result of stability test (Friedman test)

Outcome	<i>FC</i>	<i>FS</i>	α	<i>P_value</i>
TN	7	9	0.05	0.00165
FP	7	9	0.05	0.00165
TP	7	9	0.05	0.00165
FN	7	9	0.05	0.00165

the ability to deal with any data set regardless of its distribution. Four experiments based on four different chunks are performed as given in Table 4. The Friedman test is mainly based on the null hypothesis assuming that the performed experiments have identical effects. There are two constraints to reject the null hypothesis: firstly, the critical value (*FC*) must be less than the computed statistical value (*FS*). The second constraint is that the significance level value (α) must be bigger than the computed probability value (*P_value*). In this article, the standard value of α is decided to be 0.05. Table 6 displays the obtained results of the Friedman test when it is applied to TP, FP, TN, and FN. As shown in Table 6, the null hypothesis of the Friedman test is rejected because the two constraints are met in all cases ($FC < FS$ and $\alpha > P_value$). Accordingly, the obtained empirical results of the utilized models are meaningful and divergent.

Sensitivity analysis refers to how the target (dependent) variables are affected by the variations in the input (independent) variables within a certain number of constraints [55]. In this study, the number of chunks as input variables and the recall values as the target variables are utilized to perform the sensitivity analysis. The sensitivity analysis can be performed based on different techniques, while the One-At-a-Time (OAT) technique is most familiar. Initially, in the OAT technique, the model’s reference case has been defined by the recall values of the employed models based on one chunk. In the next step, the recall values are computed for 2-chunk, 4-chunk, and 8-chunk combinations without changing any other constraints. Ultimately, the statistical sensitivity values are estimated by the following formula [56]:

Table 7 The OAT sensitivity analysis results

Chunks number	Sensitivity Statistic (%)	
	OC-SVM	PCA
1	–	–
2	14.74	12.29
4	23.35	33.82
8	28.40	38.65

$$Sensitivity\ statistic = \frac{\Delta\ output\ variable}{\Delta\ input\ variable} \quad (24)$$

As the sensitivity statistic increases, the recall is more sensitive to variations of the chunk number. Table 7 gives the results of the OAT sensitivity analysis at four different chunks number. Examining Table 7, it can be observed that the recall values of the utilized anomaly-based fault detection models significantly increase as the number of chunks increases.

5.2 Feature Selection Methods

This subsection discusses the effect of applying the approach of feature selection in the data preprocessing stage. Many optimization methods are proposed to modify the process of feature selection such as Biogeography-Based Optimizer (BBO) [57], Firefly Algorithm (FA) [58], Fitness Proportionate Selection Binary Particle Swarm Optimization and Entropy (FPSBPSO-E) [12,59,60], hybrid of Grey Wolf Optimization and Genetic Algorithm (GWO-GA) [61], hybrid of Grey Wolf Optimization and Particle Swarm Optimization (GWO-PSO) [61], Satin Bowerbird Optimizer (SBO) [62], and Stochastic Fractal Search-based Guided Whale Optimization Algorithm (SFS-Guided WOA) [63]. These appeared methods in the literature are compared based on the AUC and Feature Reduction Rate (FRR) metrics [12]. The FRR metric is defined as the complement of the ratio of selected features number to all features numbers, and it can be computed by the following formula [12].

$$FRR = 1 - \frac{Selected\ features\ number}{All\ features\ number} \quad (25)$$

The first experiment is performed on the data at 1-chunk, at which the total size of the feature set is 20. Besides, the anomaly fault detection models are taught and examined by the feature subsets created by the feature selection techniques. Table 8 shows the obtained results of the feature selection based on the above-mentioned different selection methods. The results clearly show that the FPSBPSO-E method outperformed all the other methods regarding performance and the size of the feature subset. It provided the smallest feature subset.

Table 8 Feature selection results based on different methods

Method	Selected features number	FRR (%)	AUC (%)
BBO	15	25	84.56
FA	17	15	82.79
FPSBPSO-E	12	40	87.78
GWO-GA	16	20	83.28
GWO-PSO	14	30	85.37
SBO	18	10	81.33
SFS-Guided WOA	13	35	86.65

The bold values show the best results of the used models in the same experiment and among all experiments, respectively

Table 9 Hyperparameter optimization results based on different methods

Method	AUC (%)
BA	76.45
GA	79.93
GOA	77.24
GWO	80.50
MVO	78.07
PSO	81.67
WOA	79.51

The bold value shows the best results of the used models in the same experiment and among all experiments, respectively

5.3 Hyperparameter Optimization Methods

The AUC-based performance of the robust PSO algorithm to select the optimal hyperparameter values is compared with many methods appeared in the literature such as Bat Algorithm (BA) [64], Genetic Algorithm (GA) [65], Grasshopper Optimization Algorithm (GOA) [66], GWO [67], Multiverse Optimization (MVO) [68], and Whale Optimization Algorithm (WOA) [69]. The obtained results are displayed in Table 9. It can be noticed that the utilized PSO algorithm for hyperparameter selection surpassed all the other optimization methods.

5.4 Anomaly-based Vs. Binary Models

This subsection illustrates and compares the ability and the performance of anomaly-based detection and two-class (binary) classification models in detecting the faults within power systems. Most of the utilized binary classification methods are ANN [47], Boosted Decision Tree (BDT) [70], Decision Forest (DF) [71], Decision Jungle (DJ) [72], NB [73], Quantum Support Vector Machine (QSVM) [74], and SVM [47] are presented in this study. The performance of these binary classification methods is compared to anomaly-

Table 10 AUC percentages of anomaly detection and binary classification models for 1-chunk

Detection method	Utilized model	AUC (%)
Binary classification	ANN	53.61
	BDT	62.32
	DF	66.66
	DJ	69.70
	NB	58.88
	QSVM	71.13
	SVM	61.75
Anomaly-based	OC-SVM	81.70
	PCA	79.48

The bold value shows the best results of the used models in the same experiment and among all experiments, respectively

based detection utilizing the two the OC-SVM and PCA approaches, and their results are depicted in Table 10. The proposed anomaly fault detection models outperformed the binary classification models, and this is attributed to the outstanding performance of the combination of PSO method and the OC-SVM and PCA models.

5.5 Comparison With Related Works

Generally, most fault detection studies which were appeared in the literature, as presented in Sect. 2, mainly based on simulated data that cannot identify the real pattern of real power systems. Accordingly, the suggested anomaly fault detection system in this article is compared with related studies that utilized the same VSB fault detection dataset [28–31]. The main aim of this comparison is to highlight the significance of the suggested system for the readerships. The compared results are given in Table 11. The suggested anomaly fault detection system significantly improved the recall values of

the used models when compared to the corresponding values in related works. These outcomes indicate the capability of the suggested system to detect the faults in the power systems effectively.

6 Conclusion

No one can deny the significance of effective electrical fault detection either on financial cost or quality of service. Thereby, there is a research focus on detecting faults using various perspectives. However, the lack of available real-time data and a superficial understanding of the problem of detecting electrical faults prevented significant progress in this field. Accordingly, this study sheds light on the concept of fault detection in real-world power systems and proposes a new detection system based on anomaly detection. The methodology of the proposed system consists of four cascaded stages, they are data preprocessing stage, pre-training stage, training stage, and testing stage. The data preprocessing stage has five essential sub-stages in order to filter noise, decompose signal's measurement, extract useful features, normalize data, and create training and testing sets. On the other hand, the pre-training stage is responsible to select the optimal hyperparameters of the detection classifier. Furthermore, two anomaly detection classifiers along with the VSB dataset are exploited in the training and testing phases to ensure the effectiveness of the proposed system. The gained results indicate that, when using the proposed system, all positive evaluation measures are increased by 14% and the negative ones are decreased by 8%. Regarding the performance of the used models, the OC-SVM classifier shows its superiority in a low number of chunks and vice versa for the PCA. However, the proposed system still has limitations such as its deficiency in a high-dimensional input space and automatic selection of number of chunks. To the sake of sur-

Table 11 Recall percentages of the used models and related works

Study	Method	Classifier	Recall (%)
[28]	STL	RBF-based SVM	73
[29]	–	LSTM	75
		LSTM	81
[30]	Anomaly-based detection	OC-SVM	67.05
		PCA	72.57
[31]	Binary classification	NB	80.08
		ANN	76.73
		BDT	82.18
		DF	81.98
		OC-SVM	89.91
This study	PSO+Anomaly-based detection	OC-SVM	89.91
		PCA	91.30

The bold value shows the best results of the used models in the same experiment and among all experiments, respectively

pass these drawbacks, employing deep learning methods and automating the selection of number of chunks and the best classifier could be considered as promising future works.

Funding The authors did not receive support from any organization for the submitted work.

Declarations

Conflicts of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

- Wadi, M.; Elmasry, W.: Statistical analysis of wind energy potential using different estimation methods for Weibull parameters: a case study. *Electr. Eng.* **103**, 2573–2594 (2021). <https://doi.org/10.1007/s00202-021-01254-0>
- Wadi, M.; Elmasry, W.: Modeling of wind energy potential in marmara region using different statistical distributions and genetic algorithms. In: 2021 International Conference on Electric Power Engineering–Palestine (ICEPE-P) IEEE. pp. 1–7 (2021)
- Raza, A.; Benrabah, A.; Alquthami, T.; Akmal, M.: A Review of Fault Diagnosing Methods in Power Transmission Systems. *Appl. Sci.* **10**(4), 1312 (2020)
- Elmasry, W.; Wadi, M.: Enhanced anomaly-based fault detection system in electrical power grids. *Int. Trans. Electr. Energy Syst.* (2022). <https://doi.org/10.1155/2022/1870136>
- Elmasry, W.; Wadi, M.: EDLA-EFDS: a novel ensemble deep learning approach for electrical fault detection systems. *Electr. Power Sys. Res.* **207**, 107834 (2022). <https://doi.org/10.1016/j.epr.2022.107834>
- Prasad, A.; Edward, J.B.; Ravi, K.: A review on fault classification methodologies in power transmission systems: Part-I. *J. Electrical Syst. Inf. Technol.* **5**(1), 48–60 (2018)
- Prasad, A.; Edward, J.B.; Ravi, K.: A review on fault classification methodologies in power transmission systems: Part-II. *J. Electr. Syst. Inf. Technol.* **5**(1), 61–67 (2018)
- Chen, K.; Huang, C.; He, J.: Fault detection, classification and location for transmission lines and distribution systems: a review on the methods. *High Volt.* **1**(1), 25–33 (2016)
- Elmasry, W.; Akbulut, A.; Zaim, A.H.: Empirical study on multiclass classification-based network intrusion detection. *Comput. Intell.* **35**(4), 919–954 (2019). <https://doi.org/10.1111/coin.12220>
- Chandola, V.; Banerjee, A.; Kumar, V.: Outlier detection: a survey. *ACM Comput. Surv.* **14**, 15 (2007)
- Hodge, V.; Austin, J.: A survey of outlier detection methodologies. *Artif. Intell. Rev.* **22**(2), 85–126 (2004)
- Elmasry, W.; Akbulut, A.; Zaim, A.H.: Evolving deep learning architectures for network intrusion detection using a double PSO metaheuristic. *Comp. Netw.* **168**, 107042 (2020)
- Elmasry, W.; Akbulut, A.; Zaim, A.H.: Deep learning approaches for predictive masquerade detection. *Security Commun. Netw.* (2018). <https://doi.org/10.1155/2018/9327215>
- Elmasry, W.; Akbulut, A.; Zaim, A.H.: A design of an integrated cloud-based intrusion detection system with third party cloud service. *Open Comput. Sci.* **11**(1), 365–379 (2021). <https://doi.org/10.1515/comp-2020-0214>
- One-Class Support Vector Machine. Microsoft Docs. <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/one-class-support-vector-machine> (2019)
- Schölkopf, B.; Platt, J.C.; Shawe-Taylor, J.; Smola, A.J.; Williamson, R.C.: Estimating the support of a high-dimensional distribution. *Neural Comput.* **13**(7), 1443–1471 (2001)
- PCA-Based Anomaly Detection. Microsoft Docs. <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/pca-based-anomaly-detection> (2019)
- ENET Centre. VSB. <https://cenet.vsb.cz/en/> (2021)
- VSB Power Line Fault Detection. Kaggle. <https://www.kaggle.com/c/vsb-power-line-fault-detection/data> (2018)
- Singh, R.: Fault detection of electric power transmission line by using neural network. *Int. J. Emerg. Technol. Adv. Eng.* **2**(12), 530–538 (2012)
- Tayeb, E.B.M.: Faults detection in power systems using artificial neural network. *Am. J. Eng. Res.* **2**(6), 69–75 (2013)
- Jamil, M.; Sharma, S.K.; Singh, R.: Fault detection and classification in electrical power transmission system using artificial neural network. *SpringerPlus* **4**(1), 1–13 (2015)
- Koley, E.; Verma, K.; Ghosh, S.: An improved fault detection classification and location scheme based on wavelet transform and artificial neural network for six phase transmission line using single end data only. *Springerplus* **4**(1), 551 (2015)
- Uzubi, U.; Ekwue, A.; Ejiogu, E.: Artificial neural network technique for transmission line protection on Nigerian power system. In: IEEE PES PowerAfrica. IEEE. pp. 52–58 (2017)
- Eboule, P.S.P.; Pretorius, J.H.C.; Mbuli, N.; Leke, C.: Fault detection and Location in power transmission line using concurrent neuro fuzzy technique. In: IEEE Electrical Power and Energy Conference (EPEC). IEEE, pp. 1–6 (2018)
- Singh, M.; Panigrahi, B.; Maheshwari, R.: Transmission line fault detection and classification. In: 2011 International Conference on Emerging Trends in Electrical and Computer Technology. IEEE (2011) pp. 15–22
- Gururajapathy, S.S.; Mokhlis, H.; Illias, H.A.B.: Classification and regression analysis using support vector machine for classifying and locating faults in a distribution system. *Turkish J. Electr. Eng. Comput. Sci.* **26**(6), 3044–3056 (2018)
- Dong, M.; Sun, Z.; Wang, C.A.: pattern recognition method for partial discharge detection on insulated overhead conductors. In: IEEE Canadian Conference of Electrical and Computer Engineering (CCECE). IEEE. pp. 1–4 (2019)
- Qu, N.; Li, Z.; Zuo, J.; Chen, J.: Fault detection on insulated overhead conductors based on DWT-LSTM and partial discharge. *IEEE Access* **8**, 87060–87070 (2020)
- Wadi, M.; Elmasry, W.: An anomaly-based technique for fault detection in power system networks. In: 2021 International Conference on Electric Power Engineering - Palestine (ICEPE- P). IEEE. pp. 1–6 (2021)
- Wadi, M.: Fault detection in power grids based on improved supervised machine learning binary classification. *J. Electr. Eng.* **72**(5), 315–322 (2021)
- Himeur, Y.; Ghanem, K.; Alsalemi, A.; Bensaali, F.; Amira, A.: Artificial intelligence based anomaly detection of energy consumption in buildings: a review, current trends and new perspectives. *Appl. Energy* **287**, 116601 (2021)
- Biswas, S.; Nayak, P.K.: A fault detection and classification scheme for unified power flow controller compensated transmission lines connecting wind farms. *IEEE Syst. J.* **15**(1), 297–306 (2020)
- Rose, T.; Kifayat, K.; Abbas, S.; Asim, M.: A hybrid anomaly-based intrusion detection system to improve time complexity in the Internet of Energy environment. *J. Parallel Distrib. Comput.* **145**, 124–139 (2020)
- Alamiyedy, T.A.; Anbar, M.; Alqattan, Z.N.; Alzubi, Q.M.: Anomaly-based intrusion detection system using multi-objective



- grey wolf optimisation algorithm. *J. Ambient Intell. Humaniz. Comput.* **11**(9), 3735–3756 (2020)
36. Himeur, Y.; Alsalemi, A.; Bensaali, F.; Amira, A.: Smart power consumption abnormality detection in buildings using micromoments and improved K-nearest neighbors. *Int. J. Intell. Syst.* **36**(6), 2865–2894 (2021)
 37. Aker, E.; Othman, M.L.; Veerasamy, V.; Aris, I.B.; Wahab, N.I.A.; Hizam, H.: Fault detection and classification of shunt compensated transmission line using discrete wavelet transform and naive bayes classifier. *Energies* **13**(1), 243 (2020)
 38. Chen, H.; Jiang, B.; Zhang, T.; Lu, N.: Data-driven and deep learning-based detection and diagnosis of incipient faults with application to electrical traction systems. *Neurocomputing* **396**, 429–437 (2020)
 39. Himeur, Y.; Alsalemi, A.; Bensaali, F.; Amira, A.: A novel approach for detecting anomalous energy consumption based on micro-moments and deep neural networks. *Cognit. Comput.* **12**(6), 1381–1401 (2020)
 40. Naveed, K.; Akhtar, M.T.; Siddiqui, M.F.; Ur Rehman, N.: A statistical approach to signal denoising based on data-driven multiscale representation. *Digital Signal Process.* **108**, 102896 (2021)
 41. García-Gil, D.; Luengo, J.; García, S.; Herrera, F.: Enabling smart data: noise filtering in big data classification. *Inf. Sci.* **479**, 135–152 (2019)
 42. Yue, X.: Data decomposition for analytics of engineering systems: Literature review, methodology formulation, and future trends. In: 58745 of International Manufacturing Science and Engineering Conference. American Society of Mechanical Engineers (2019) V001T02A011
 43. Yasir, M.N.; Koh, B.H.: Data decomposition techniques with multiscale permutation entropy calculations for bearing fault diagnosis. *Sensors* **18**(4), 1278 (2018)
 44. Hyndman, R.J.; Fan, Y.: Sample quantiles in statistical packages. *Am. Stat.* **50**(4), 361–365 (1996)
 45. Azure machine learning studio. Microsoft. <https://studio.azureml.net/> (2022)
 46. Barga, R.; Fontama, V.; Tok, W.H.: Introducing microsoft azure machine learning. In: Predictive Analytics with Microsoft Azure Machine Learning (pp. 21–43) Springer (2015)
 47. Elmasry, W.; Akbulut, A.; Zaim, A.H.: Comparative evaluation of different classification techniques for masquerade attack detection. *Int. J. Inf. Comput. Security* **13**(2), 187–209 (2020)
 48. Boughorbel, S.; Jarray, F.; El-Anbari, M.: Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PloS One* **12**(6), e0177678 (2017)
 49. Matthews, B.W.: Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica Biophysica Acta BBA-Protein Struct.* **405**(2), 442–451 (1975)
 50. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **7**, 1–30 (2006)
 51. Fawcett, T.: An introduction to ROC analysis. *Pattern Recognit. Lett.* **27**(8), 861–874 (2006)
 52. Bradley, A.P.: The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.* **30**(7), 1145–1159 (1997)
 53. Sokolova, M.; Lapalme, G.: A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **45**(4), 427–437 (2009)
 54. Friedman, M.: The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J. Am. Stat. Assoc.* **32**(200), 675–701 (1937)
 55. Sahmoud, S.; Elmasry, W.; Abudalfa, S.: Enhancement the Security of AES Against Modern Attacks by Using Variable Key Block Cipher. *Int. Arab. J. Technol.* **3**(1), 17–26 (2013)
 56. Cem kasapbaşı, M.; Elmasry, W.: New LSB-based colour image steganography method to enhance the efficiency in payload capacity, security and integrity check. *Sādhanā* (2018) 43(5): 1–14. <https://doi.org/10.1007/s12046-018-0848-4>
 57. Simon, D.: Biogeography-based optimization. *IEEE Trans. Evol. Comput.* **12**(6), 702–713 (2008)
 58. Fister Jr, I.; Yang, X.S.; Fister, I.; Brest, J.: Memetic firefly algorithm for combinatorial optimization. *arXiv preprint arXiv:1204.5165* (2012)
 59. Cervante, L.; Xue, B.; Zhang, M.; Shang, L.; Binary particle swarm optimisation for feature selection: A filter based approach. In: IEEE Congress on Evolutionary Computation. IEEE, 1–8 (2012)
 60. Zhou, Z.; Liu, X.; Li, P.; Shang, L.: Feature selection method with proportionate fitness based binary particle swarm optimization. In: Asia-Pacific Conference on Simulated Evolution and Learning. Springer (2014) 582–592
 61. Şenel, F.A.; Gökçe, F.; Yüksel, A.S.; Yiğit, T.: A novel hybrid PSO-GWO algorithm for optimization problems. *Eng. Comput.* **35**(4), 1359–1373 (2019)
 62. Moosavi, S.H.S.; Bardsiri, V.K.: Satin bowerbird optimizer: a new optimization algorithm to optimize ANFIS for software development effort estimation. *Eng. Appl. Artif. Intell.* **60**, 1–15 (2017)
 63. El-Kenawy, E.S.M.; Ibrahim, A.; Mirjalili, S.; Eid, M.M.; Hussein, S.E.: Novel feature selection and voting classifier algorithms for COVID-19 classification in CT images. *IEEE Access* **8**, 179317–179335 (2020)
 64. Karakonstantis, I.; Vlachos, A.: Bat algorithm applied to continuous constrained optimization problems. *J. Inf. Optim. Sci.* **42**(1), 57–75 (2021)
 65. Holland, J.H.; et al.: Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. MIT Press, Cambridge (1992)
 66. Saremi, S.; Mirjalili, S.; Lewis, A.: Grasshopper optimisation algorithm: theory and application. *Adv. Eng. Softw.* **105**, 30–47 (2017)
 67. Mirjalili, S.; Mirjalili, S.M.; Lewis, A.: Grey wolf optimizer. *Adv. Eng. Softw.* **69**, 46–61 (2014)
 68. Mirjalili, S.; Mirjalili, S.M.; Hatamlou, A.: Multi-verse optimizer: a nature-inspired algorithm for global optimization. *Neural Comput. Appl.* **27**(2), 495–513 (2016)
 69. Mirjalili, S.; Lewis, A.: The whale optimization algorithm. *Adv. Eng. Softw.* **95**, 51–67 (2016)
 70. Two-Class Boosted Decision Tree. Microsoft Docs. <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-boosted-decision-tree> (2019)
 71. Two-Class Decision Forest. Microsoft Docs. <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-decision-forest> (2019)
 72. Two-Class Decision Jungle. Microsoft Docs. <https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/two-class-decision-jungle> (2019)
 73. Zeineddine, H.; Braendle, U.; Farah, A.: Enhancing prediction of student success: automated machine learning approach. *Comput. Electr. Eng.* **89**, 106903 (2021)
 74. Havlíček, V.; Córcoles, A.D.; Temme, K.; et al.: Supervised learning with quantum-enhanced feature spaces. *Nature* **567**(7747), 209–212 (2019)