International Journal of Robotics and Control Systems
Vol. 5, No. 1, 2025, pp. 530-554
ISSN 2775-2658
http://pubs2.ascee.org/index.php/ijrcs

Wavelet Analysis- Singular Value Decomposition Based Method for Precise Fault Localization in Power Distribution Networks Using k-NN Classifier

Abhishek Raj [a,1], Chandra Sekhar Mishra [b,2], S Ramana Kumar Joga [c,3], I. M. Elzein [d,4], Asit Mohanty [e,5], Sneha lika [f,6], Mohamed Metwally Mahmoud [g,7,*], Ahmed M. Ewais [g,8]

[a] School of Computer Engineering, KIIT Deemed to be University, India
[b] Mahavir Institute of Engineering & Technology (MIET), Bhubaneswar, India
[c] EEE Department, Dadi Instiuite of Engineering and Technology, Anakapalle, India
[d] Department of Electrical Engineering, College of Engineering and Technology, University of Doha for Science and Technology, Doha P.O. Box 24449, Qatar
[e] Mahavir Institute of Engineering & Technology (MIET), Bhubaneswar, India
[f] School of Electrical Engineering, KIIT Deemed to be University, India
[g] Electrical Engineering Department, Faculty of Energy Engineering, Aswan University, Aswan 81528, Egypt
[1] abhishek.rajfcs@kiit.ac.in; [2] chandrasekharmishra1106@gmail.com; [3] sanset567@gmail.com; [4] 60101973@udst.edu.qa
[5] amohanty@miet.edu.in; [6] snehalika.fel@kiit.ac.in; [7] metwally_m@aswu.edu.eg; [8] ewaisa@aswu.edu.eg
* Corresponding Author

ARTICLE INFO

Article history
Received July 30, 2024
Revised January 22, 2025
Accepted January 23, 2025

Keywords
Fault Detection;
Fault Location;
Power Quality Monitoring;
Wavelet Transform;
k-NN

ABSTRACT

This article presents a wavelet analysis-singular value decomposition (WA-SVD) based method for precise fault localization in recent power distribution networks using k-NN Classifier. The WA-SVD leverages the slime mould algorithm (SMA) and graph theory (GT) in enhancing the overall accuracy of fault localization. To validate the proposed methodology, extensive tests are conducted on various benchmark systems, including the IEEE 33-bus radial distribution system, the IEEE 33-bus meshed loop unbalanced distribution system, the IEEE 33-bus system with integrated renewable energy sources, and the IEEE 13-bus feeder test system. The results demonstrate a high fault classification accuracy of 99.08%, with an average localization error of just 1.2% of the total line length. The k-NN classifier exhibited a precision of 98.2% and a recall of 99.2%, underscoring the reliability and sensitivity of the proposed method. Additionally, the computational efficiency of the algorithm is evidenced by an average processing time of 0.0764 seconds per fault event, making it well-suited for real-time applications.

1. Introduction

1.1. Motivation and Background

Fault detection (FD) in power distribution systems (PDSs), particularly in radial systems is critical for maintaining system reliability and preventing extensive damage undetected or inaccurately localized faults can lead to prolonged outages, damage to infrastructure, increased operational costs, and safety hazards for both utility personnel and the public. The challenge is further compounded by the integration of distributed generation (DG), which introduces additional complexity and variability into the system, making traditional FD methods less effective. PDSs are critical components of the electrical grid, responsible for delivering electricity from substations to

end users. Ensuring their reliability and stability is paramount, as any disruptions can lead to significant economic losses and safety hazards. Among various types of disturbances, short circuit faults (SCFs) are particularly detrimental, causing equipment damage, power outages, and system instability. Therefore, the development of effective fault diagnosis and localization methods is essential for maintaining the robustness of PDSs [1]-[3].

The integration of different power plants into a system changes the activities and performance of the system. It is not possible to design and build PDS in such a manner that it never experiences a fault, where a fault is defined as discomfort brought on by an abnormal flow of current inside the system. These malfunctions are, in the vast majority of instances, inevitable because of unfavorable weather conditions, the interference of birds on overhead wires, and other factors like air pollution. Certain faults are not severe, but the vast majority of faults produce significant damage to the DSs [4]. This damage includes the creation of an arc, the tripping of switchgear components, damage to the insulation, and damage to the components. These malfunctions lead to blackouts in the PDSs and have the potential to be harmful to living beings [5]-[7].

Numerous research investigations are carried out to identify the numerous malfunctions that might take place inside the PDS. Open-circuit faults are the most common kind of fault that may occur in a PDS. This type of failure takes place whenever there is an interruption in the circuit, which can take place due to an open switch or a break in the conductor. If a distribution line has a failure, the essential steps must be performed to eradicate the fault as quickly as possible and bring the power supply back online. At the moment, the majority of intelligent terminals that are configured via the various units of the distributed PDS interface with the main system to find out where the problem is located. In general, a comprehensive array approach is used to locate faults [8], [9].

In recent times, the primary objective of power engineers and researchers has been developing methods for the early identification of defects. Recent advancements in PDSs have introduced new complexities and challenges in FD. Modern PDSs often incorporate DGs such as photovoltaic systems (PVSs), energy storage devices, and electric vehicle (EV) charging stations. These elements introduce bidirectional power flows and variable operational conditions, making traditional fault diagnosis techniques less effective. Consequently, there is a growing need for advanced analytical methods that can handle the dynamic nature of contemporary PDSs [10]-[12].

1.2. Literature Review

Additionally, with the growth of powerful computer data processing technologies such as cloud computing, artificial intelligence computational algorithms are increasingly being employed for fault location (FL) and are garnering a growing amount of attention. Tellegen's theorem is used to locate the open-circuit segment or failed node in the network utilizing input and output port information. Voltage and current phasor information is obtained via DFT-based phasor measuring equipment and it is discussed in [13]. It was proposed a system that uses VMD to dissect the distribution line's current signal rate of change. The most crucial mode's energy is used to find problems. In grid-connected and islanding microgrid (MG) operating modes, the suggested technique is tested for pole-to-pole and pole-to-ground faults with wide variations in FL, fault resistance, and DG penetration [14], [15].

Using a GIS-generated DG topology, a petri-net-based fault identification model is constructed. It was suggested a technique that employs fault indicator fault flags, circuit breaker statuses, pre, and post-current measurements, and feeding and lateral loadings [16], [17]. A DG fault monitoring system based on edge computing was proposed in [18]. This system is capable of detecting and responding to DG problems in real-time and can improve reliability and user satisfaction by detecting and responding to PDS faults quickly. PDSs that use DGs are diagnosed through an integrated FL and isolation approach based on the fully decentralized multi-agent system, which includes primary and device failure prevention [19]. In [20], a novel technique of locating SCFs on 11-kV overhead distribution lines using a noncontact magnetic field (MF) measuring device was proposed. The technology employs magnetic sensors that were very sensitive and energy-efficient to detect changes

in MF levels throughout the distribution lines. It was proposed a FD that enhances the detection accuracy of faults by making use of the commonalities in fault data streams across different sites in a PDS. The collected similarities add to the knowledge needed to identify faults in a particular region, allowing for more tasks to be learned at the same time and increasing learning accuracy [21], [22].

One-dimensional convolutional neural networks and waveform fusion are used in this article to locate fault lines. Waveform fusion happens when a line-to-ground (LG) fault arises when the first-half waves of zero-sequence currents are collected and overlaid on each other. A 1-D CNN is utilized to detect whether the fused waveform source has a fault line in the fused waveform. Next, it utilizes the 1-D CNN output to update the counter value to find the fault line in the system [23]. An in-depth review of the distribution network FD and protection mechanisms merged with DG was presented in [24], [25]. Threshold filtering and time-frequency distribution filtering were used sequentially in [26], to de-noise transient zero-sequence currents using GST with a variable factor. PDSs faults are diagnosed using a support vector machine (SVM)-based machine learning (ML) classifier discussed in [27]. A system for the detection of wildfires based on deep learning (DL), with application to the FL in PDSs, was discussed in [28]. It was proposed a single-LG fault diagnosis algorithm with distributed parameters which were characterized by asymmetry and line resistance [29], [30].

Fault diagnosis using smart meters (SM) in PDSs was discussed in [31]. State estimation (SE) was used in [32] to identify SCFs and detect them. In the first phase, the standard SE technique was reworked such that it may be used in the presence of faults. After diagnosing the faulty zone, a FL algorithm based on the revised SE was also proposed. Discrete wavelet transform (DWT)-based SCF diagnosis in the PDS was discussed in [33], [34]. Using the Hilbert-Huang transform and ANN classifier, [35] offers a technique for diagnosing faults in a PDSs. It was proposed a data-driven ML technique based on the distance between the defects that must be used to properly diagnose problems [36]. A waveform vector embedding technique is presented in this study [37] as a means of embedding the incipient fault waveforms of various devices in waveform vectors. Then, it uses the waveform vectors to create a waveform dictionary, which was a collection of waveforms.

MG changes are detected using a variety of ML algorithm-based classifiers. DWT-based signal processing was used to evaluate transient signals, and a classification technique is employed to identify the fault and non-fault data and was discussed in [38]. Traditional fault diagnosis methods, such as impedance-based techniques and traveling wave methods, have been widely used. However, with the evolution of PDSs incorporating modern topologies such as the integration of DGs, and advanced grid management technologies; these conventional methods face new challenges. The increased complexity and dynamic behavior of contemporary PDSs necessitate the development of more advanced and robust fault diagnosis and localization techniques. In recent years, signal processing and ML techniques have shown great promise in enhancing fault management capabilities. Wavelet analysis (WA) is one such signal processing tool that excels in capturing transient features of fault signals, making it highly suitable for FD in PDSs. Singular value decomposition (SVD) is another powerful technique for dimensionality reduction and feature extraction, simplifying complex data while retaining essential information. Combined with ML algorithms like k-Nearest Neighbors (k-NN), these techniques can significantly improve the accuracy and efficiency of fault diagnosis and localization.

1.3. Contributions

- The primary addition that this work makes to the existing body of knowledge is the development of a novel feeder protection relay that is based on signal processing and a ML classifier. This relay is designed to identify, classify, and locate fault zones in DSs. To locate the protective zones inside the DSs, a whole new approach based on graph theory and the slime mould algorithm (SMA) has been developed. The k-NNN ML classifier is used to categorize these protective zones. To analyze the behavior of a signal, a multi-resolution WTs with the name wavelet packet decomposition (WPD) with SVD is used to dissect healthy and defect signals. The technique that has been suggested provides an accurate algorithm that can identify and categorize defects in both normal and noisy environments. The approach that has been suggested

is also an algorithm that can be generalized and is capable of operating effectively despite any changes in the structure of the grid.

- A key limitation identified in the current methodology is the insufficient depth of empirical validation and comparative analysis with existing fault diagnosis techniques. While the proposed approach introduces several innovative components such as the SMA for optimal measurement placement, GT for protection zone partitioning, and the WA-SVD method for enhanced signal processing its real-world applicability remains unconvincing without robust experimental validation. To overcome this, the methodology must be rigorously tested across a broad spectrum of fault scenarios under various operational conditions, including varying fault types (e.g., single-LG, double-LG, etc.), fault inception angles, and the presence of noise. The performance of the proposed techniques should be benchmarked against established methods, such as traditional wavelet-based FD, S-transform-based approaches, or ML classifiers like SVM and ANN. The key performance indicators (KPIs) that should be analyzed include:

  1. The ability of the proposed method to accurately classify different types of faults, as compared to existing methods.
  2. The precision in determining the exact location of the fault along the distribution line is measured against the ground truth.
  3. The time required for processing and analyzing fault data, particularly in real-time applications, compared to conventional techniques.
  4. The resilience of the proposed methodology in accurately diagnosing faults under noisy conditions or in the presence of variable load conditions.
  5. The capability of the method to maintain performance as the size and complexity of the network increase.

A thorough empirical validation should involve not only the IEEE 33-bus and IEEE 13-bus test systems but also more extensive, real-world datasets if available. Comparative analysis should present quantitative results that clearly demonstrate the advantages or any potential drawbacks of the proposed approach relative to the state-of-the-art. Incorporating these empirical results will substantiate the claims of the paper, providing clear evidence of the methodology's efficacy and its potential to improve fault diagnosis and localization in PDSs. Only through this rigorous validation can the proposed techniques be deemed viable and potentially transformative for real-world applications.

SCFs in PDSs, such as the IEEE 33-bus radial network, pose significant challenges due to their potential to cause severe damage, disrupt service, and endanger safety. FD accurately and promptly is crucial for maintaining system stability and minimizing downtime. Traditional FD methods, like overcurrent and distance protection, often fall short in complex systems, particularly in the presence of noise and DG, which can alter fault characteristics and hinder detection accuracy. To address these challenges, this research introduces a novel approach that combines the SMA with WA-SVD. This methodology optimizes the extraction of fault-related features, even in noisy environments, enhancing the accuracy of FD. By integrating this approach with a k-NN classifier, the proposed system not only improves fault localization but also provides a robust solution for managing SCFs in modern PDSs.

2. Proposed Methodology

The zone of SCFs in the DSs is going to be detected, classified, and located with the help of a suggested technique that has two stages. The WA-SVD is used in the first step of the process, during which healthy and defect signals that have been received from optimally located SMs are analyzed. This article presents a novel methodology (WA-SVD) for detecting, classifying, and localizing SCFs in PDSs, where it can process fault signals with higher precision and efficiency.

WA is employed to decompose fault signals into multiple frequency components, effectively isolating transient phenomena associated with faults. This decomposition enhances the signal-to-noise ratio, which is critical for accurate fault feature extraction. SVD is then applied to these WT signals to extract principal features that encapsulate the most significant information while reducing data dimensionality. This dimensionality reduction simplifies subsequent classification tasks and improves computational efficiency. The WA-SVD method only enhances the accuracy of FD by focusing on high-frequency (HF) transient components but also reduces computational complexity, making it suitable for real-time (RT) applications. It is particularly effective in environments with noisy data, where maintaining the integrity of critical fault features is essential. The flow chart of the proposed method is shown in Fig. 1.

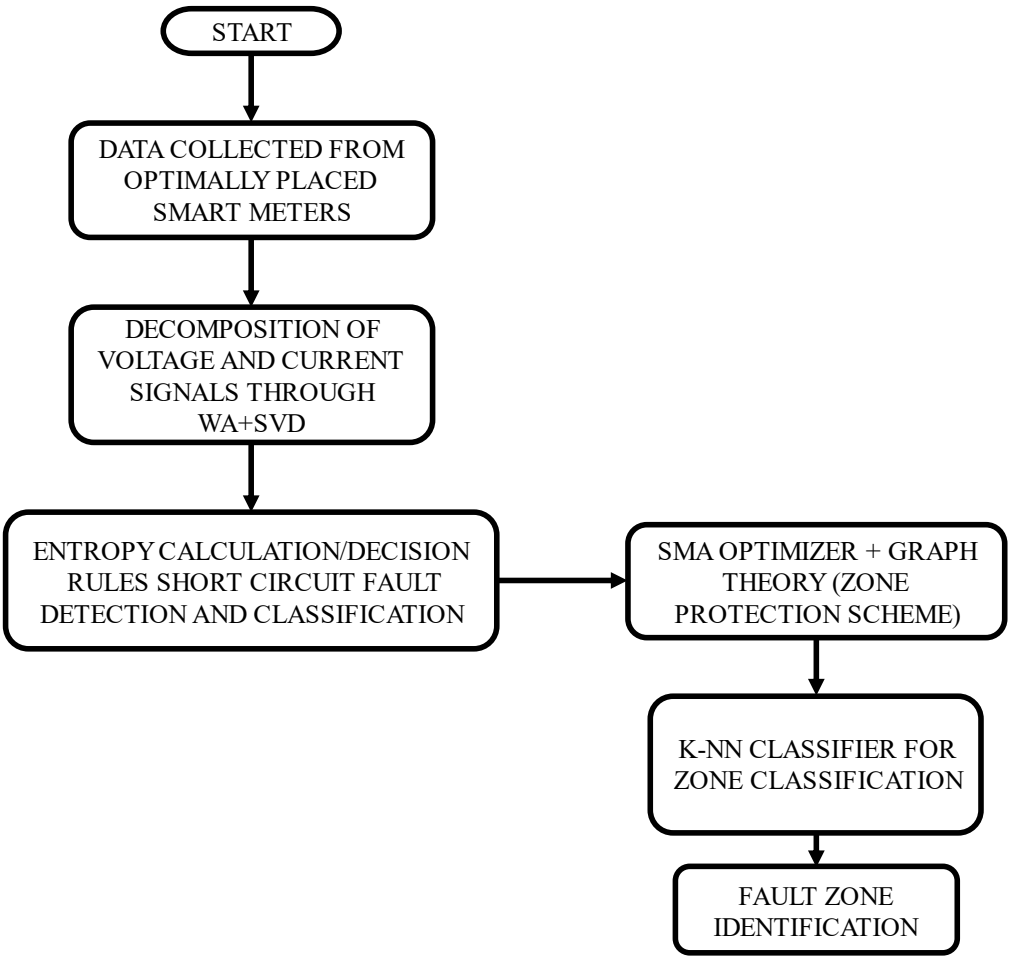

Fig. 1. Flow chart of the proposed method

2.1. Optimal Placement of SMs Using SMA

In this article, an optimization strategy that is based on the SMA is used to determine the optimal placement of the SM. The objective function of the optimal placement problem (OPP) is represented in Eq. (1). The SMA is integral to the FD methodology employed in the IEEE 33-bus radial DS, particularly under noisy conditions. SMA is a bio-inspired technique that mimics the foraging behavior of SMs, known for their efficient exploration and exploitation capabilities in finding optimal paths for nutrient acquisition. It operates through several key phases: initialization, oscillatory movement, exploitation, and exploration. The algorithm begins by initializing a population of candidate solutions, each representing a potential set of parameters for the signal processing techniques such as the tuned Q-factor WT (TQWT). These parameters may include decomposition levels, Q-factors, and threshold values essential for accurately extracting features from fault signals. In this phase, each candidate solution undergoes oscillatory movements influenced by both local and global best solutions.

$$Minimize \sum_{k=1}^{n} Z_k \quad (1)$$

$$Subjectedto[C] * [Z] \geq b$$

where $C$is a connectivity matrix and n is the number of buses. The matrix $C$is represented in the form of

$$\text{Matrix } C_{i,j} = \begin{Bmatrix} 1, if i = j \\ 1, if i and j are connected \\ 0, if otherwise \end{Bmatrix}$$

whereas $B$is a column matrix and is represented as

$$[b] = [1111111 \ldots 1]_{1XN}^{T}$$

It has been proposed that the SMA has been inspired by the behavioural aspect of SMA [39]. In nature, the SM detects the food and thereafter encircles it and eventually digests it by releasing enzymes. The properties of SM may be mathematically expressed into three steps: seeking food, encapsulating food, and oscillating, which can be represented as follows: The SM tracks the food based on the smell dissipated in the air, as presented in Eq. (2).

$$Z(k+1) = \begin{cases} Z_b(k) + vb.(H.Z_A(t) - Z_B(t)), r < p \\ vc.Z(t), r \geq p \end{cases} \quad (2)$$

Here Eq. (2), Zrepresents the position of the SM, $Z_b$denotes the latest location with the most intensified smell (food location), $Z_A$and $Z_B$are randomly selected candidate from the SM, $r$is a random value between [0,1],$k$denotes the iterations, Hsignifies the SM adaptive weight, $vb$is the randomly generated value in the range [-a, a],vcrepresented random value in the range [-b, b] where $b$resembles a value that decreases linearly from 1 to 0 based upon the iteration ($b = 1 - k/Iter_{max}$). The probability index $p$can be represented as in Eq. (3).

$$p = tanh|J(i) - EG| \quad (3)$$

Eq. (3) $J(k)$ represents the fitness value corresponding to $Z$and EGresembles the best candidate solution achieved so far. The parameter $a$is represented in Eq. (4).

$$a = arctan\,h\left(-\left(\frac{k}{Iter_{max}}()\right)()\right) \quad (4)$$

The adaptive weight H of the SM can be represented as:

$$H(SmellIndex(i)) = \begin{cases} 1 + r.log\left(\frac{bG - J(i)}{bG - wG} + 1\right); \\ \text{first half of population} \\ 1 - r.log\left(\frac{bG - J(i)}{bG - wG} + 1\right); \\ \text{other half of population} \end{cases} \quad (5)$$

$$SmellIndex = sort(J) \quad (6)$$

Eq. (5) and Eq. (6), bGresemble the best fitness solution achieved in the current position and the worst fitness resembles wG the latest position. $SmellIndex$represents the sorted values of the fitness maintained sequentially.

The SM modifies the search procedure based on the concentration of food as it approaches the feeding process. Situation when the food concentration is low, the area's weight decreases; when the

food concentration is high, the area's weight increases. However, to enhance the SMA's exploration ability, Eq. (7) may be used to update the position of the SM.

$$Z^* = \begin{cases} Rand.(ub - lb) + lb;\ r < x \\ Z_b(k) + vb.(H.Z_A(k) - Z_B(k));\ r < p \\ vc.Z(k);\ r \geq p \end{cases} \quad (7)$$

In Eq. (7), the lower and the upper bounds of the decision variables are represented by $lb$and ub,$Rand$resemble the randomly generated value [0,1], and $x$is set to 0.03 as the best value based on trial runs.

The flow chart for the optimal placement of SMs through the SMA is shown in Fig. 2, the IEEE 33 radial distribution bus is being considered. It has been determined that the optimal solution for the objective function has been achieved, and cases of these optimal solutions can be found in Table 1.

Table 1. SMs optimal placement

Set Number	Bus Numbers
1	3,5,9,11,13,16,20,22,24,26,32
2	2,4,6,10,14,16,19,22,25,28,30
3	3,4,7,10,12,16,18,21,25,27,30

2.2. WA-SVD Technique

The wavelet packet transform (WPT) is represented by the following Eqs. (8) and (9). The main difference is that the WPT analyses the signal in a more comprehensive manner, which enables it to give more flexibility. The wavelet packets-based approach to noise reduction performs much better than the wavelet-based method in terms of overall performance.

$$S_{2n}^{(j)}(t) = \sqrt{2}\sum_k h(k)S_n^{(j)}(2t - k) \quad (8)$$

$$S_{2n+1}^{(j)}(t) = \sqrt{2}\sum_k g(k)S_n^{(j)}(2t - k) \quad (9)$$

where n= 0,1,2…. And k=0, 1,… m. $S_0^0(t)$ is the scaling function and $S_1^0(t)$ is the mother wavelet function. (j) represents the number of decomposition levels.

The SVD is a matrix factorization that has been used in a wide range of important applications. Perhaps this is because SVD can be used to do various numerical operations on matrices. For example, one may compute matrix inverses; solve a system of equations; calculate determinants; compute a least squares solution to overdetermined systems; and compute condition numbers; norms; and column rank, among many more. SVD solves or simplifies many real-world matrices, hence it should come as no surprise that SVD is used in numerous applications of matrices [40]. Statistics and data analysis aren't exempt from the influence of this crucial factorization, however, Linear and extended linear models may be fitted using SVD, among other things (LM and GLM). QR factorization is typically used to solve the traditional "linear regression" formulation since it's simply linear least squares. SVD, on the other hand, has certain benefits. This is essential for statisticians since their so-called model matrices are often rank-degenerate by construction in the context of statistical trial design. The iteratively reweighted least squares method may be used to fit a GLM with a good LM fitter. The data is organized as a matrix with variables in the columns and observations in the rows. There are several ways to approach this problem, but the most common approach is to mean-center each column individually and then projects it onto the correct singular vectors. The additional variables are arranged in decreasing order of the original dataset's variability. Plotting just the first few rows allows for a decent compromise between keeping the original dataset's

variability and being unable to see it at all. Assuming we have a square matrix M of dimensions n x n. If X is a non-zero n-dimensional column vector, then a real integer is said to be an eigenvalue of a matrix A, represented in Eq. (10).

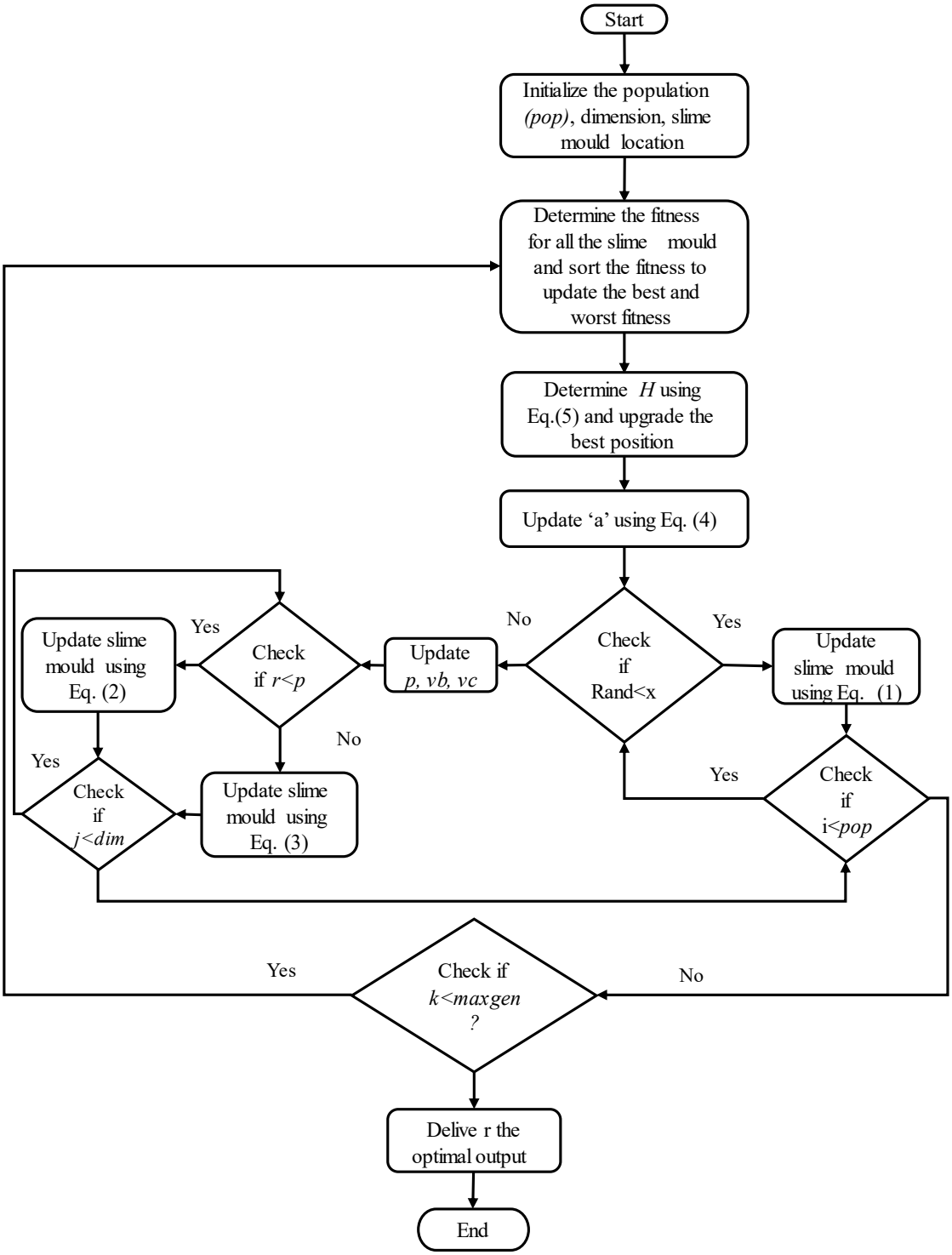

Fig. 2. Flow chart for SMs optimal placement using SMA

$$MX = \partial X, X \neq 0 \tag{10}$$

The eigenvector of the matrix M corresponding to the eigenvalue is a vector X meeting condition. The term "spectrum" refers to the collection of eigenvalues that make up an M matrix.

Some matrices may contain no spectrum at all (that is, no eigenvalues). On the other hand, square symmetric matrices often contain n eigenvalues, which indicates that their spectrum is complete. The eigenvector X is assumed to be a column vector in Eq. (10). The right eigenvectors of a square matrix M are all column eigenvectors that meet condition (Eq. (10). Matrix M may be described by its eigenvectors in the same way, as presented in Eq. (11).

$$YM = \partial Y, Y \neq 0 \tag{11}$$

The left eigenvectors of the matrix A are the row vectors that meet Eq. (11). In the general scenario, the left and right eigenvectors of both matrices can be different, although the eigenvalues of the matrices M and M^T would be identical. Its feature that a square symmetric matrix may be decomposed into the form is an important property that all square symmetric matrices have.

$$M = LDL^T \tag{12}$$

where L is the square matrix, D is the diagonal matrix. Eq. (12) is said to be a singular decomposition of M. Let us adapt the principle of singular decomposition to a more general category of matrices, namely to real rectangular matrices of dimension m x n. This will broaden the application of the singular decomposition notion. One way to describe the singular decomposition of a matrix like this is by the use of the form, presented in Eq. (13).

$$M = U\sum V^T \tag{13}$$

where U is the left eigenvector of M, V is the right eigenvector of M, and $\sum$ is the diagonal matrix.

2.3. Formation of Decision Rules by Entropy Calculation

Entropy is a measure that gives information about the amount of a signal as well as the uncertainty of the signal. As a consequence of this, measurements of entropy may provide information on signal flaws. The calculation of wavelet energy is the first step in the process of measuring entropy. The energy entropy may be used to represent the attributes of the arc voltage and the formula for the energy entropy (E_i) can be stated in Eq. (14).

$$Sumofenergies E_i = |D_i(t)|^2 \tag{14}$$

In this Eq. (14), the detail coefficients of the wavelet analysis are denoted by $D_i(t)$. The extracted arc voltage is used in the Singular Value Decomposition, where 'i' is the number of extraction levels. The expression for the energy distribution is now the ratio of the total energies, and the definition of the energy of the sub-band signal is as represented in Eq. (15).

$$P_i = \frac{E_i}{E}, E = \sum_{i=1}^{n} E_i \tag{15}$$

Thus, according to wavelet theory, the expression for the entropy of a signal in mathematical notation is represented in Eq. (16).

$$Entropy = -\sum_{m=1}^{N} P_i \, log_2(P_i) \tag{16}$$

2.4. A Search-Based Zone Protection Selection

Graph theory (GT) topologies such as vertices (V) and edges (E) are used in this study's power system topology [41], and provide a more in-depth explanation of these principles. For zone separation, the following rules are currently being considered: Rule 1 states that a new protection zone will be created if an original bus is brought together with the existing protection zone by a

vertex. Rule 2 states that if the protection zone includes any similar buses, only one of the clone zones shall be kept while the others should be eliminated.

These two guiding tenets are taken into consideration while solving the search problem, which entails finding the optimal location for the SM inside the protection zone. Based on the aforementioned two guidelines, a one-of-a-kind protection approach was provided. The stages of the method have been evaluated on an IEEE33 bus system, and the steps that were evaluated may be seen in Fig. 3.

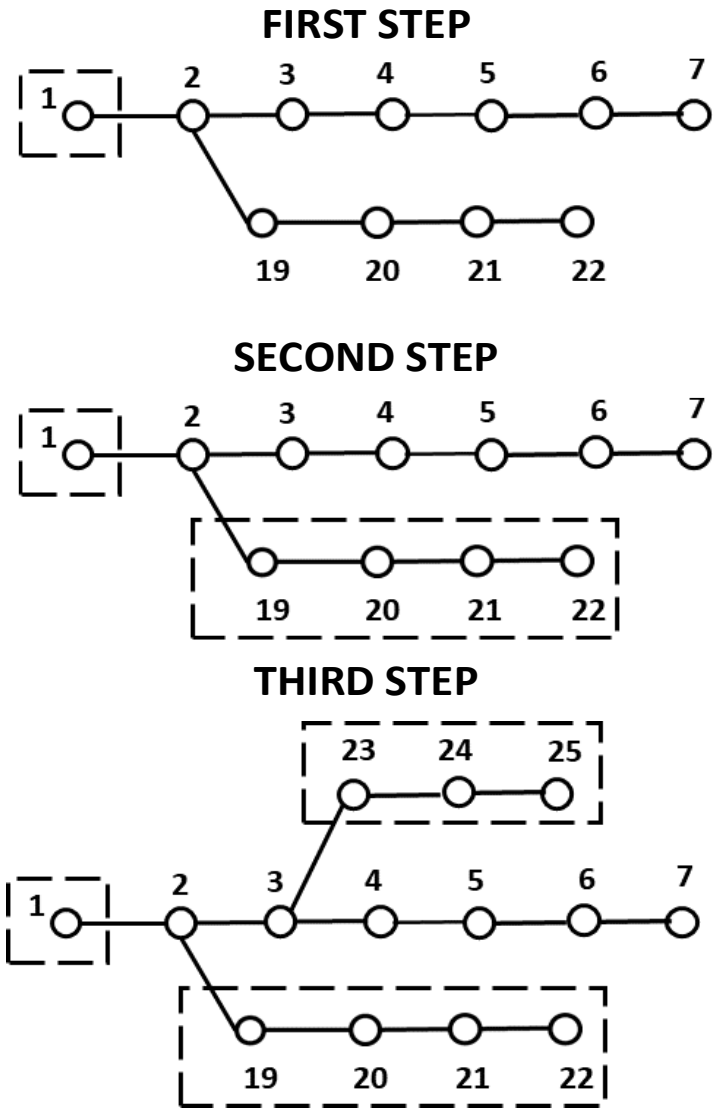

Fig. 3. Zone protection steps

The first step involves treating each basic bus as an initial zone. This occurs in the first stage. For instance, bus 1 is considered a zone under the IEEE 33 testing system. Second Step: During this step, the initial bus searches for other buses in the surrounding area and joins forces with them all to create a new zone. In the third step of the process, the search criteria include making a comparison between the old zone and the new zone established in the previous step. The last step of the search process is the fourth step, which involves the search algorithm checking to see whether the total number of buses in the zone is equal to the total number of buses in the network. If this is the case, the search process is finished. Alternately, go on to Step 2. Based on the aforementioned principles from the graph theory-based search approach, three zones have been determined, and a tabular representation of them can be seen in Table 2.

Table 2. IEEE33 Bus zone protections

Zone (Z)	Bus Numbers IEEE 33 Bus
ZA	1, 2, 3,19, 20, 21, 22, 23, 24, 25
ZB	4,5,26,27,28,29,30,31,32,33.
ZC	7,8,9,10,11,12,13,14,15,16,17,18

The numbers 2, 3, 19, 22, and 24 on the buses listed in Table 2 are the best options for positioning SM equipment in ZA. The best places to position measurement devices in ZB are the bus stops with the numbers 5, 26, and 30, while the best places to position measurement devices in ZC are the bus stops with the numbers 9, 14, and 17. This technique of searching guarantees that a minimal number of measurement devices are present in each Z.

2.5. k-NN Classifier

The k-NN method is the most common classification technique. This means that it does not make any assumptions about the fundamental dataset to classify data. It is well-known for being both straightforward and effective at the same time [42], [43]. It is an algorithm for learning via supervision. To predict the class of data that has not been labeled, a labeled training dataset is supplied. In this dataset, the data points are separated into several different classes. In a specific area, it is used to categorize data based on the training examples that are nearest or surrounding. This approach is utilized because of its ease of implementation and short computing time. Its closest neighbors are determined by calculating the Euclidean distance for continuous data. The categorization of a new input is determined by calculating the K closest neighbors and then taking the data's majority among the neighbors into account. Classification of the unlabeled data may be complicated, and the 'K' value is a significant consideration even though the classifier is simple. 'K' may be chosen in various ways, but we can just run the classifier numerous times with different values to determine which one gets the best results. All computations are done when training data is categorized, not when it is encountered in a dataset. This raises the computational cost. If you're looking for a slow learning algorithm, this one isn't going to help you much. The training dataset is not generalized in any way. As a result, the full training dataset must be used in the testing phase. k-NN predicts continuous values in regression. In computing this number, we used the average of the K closest neighbors. If a dataset has been partitioned into multiple clusters, k-NN is used to identify the class of new input for research in which there is no prior understanding of the data, k-NN is more important. The k-NN algorithm is a classification method. Classification is a two-step process:

This is the first step in the learning process: using the practice data. The classifier has been built. The classifier's performance is evaluated. Following the "next closest neighbor" approach, Data is categorized by identifying whether it has been previously tagged or not. the social strata its immediate neighbors occupy. This idea is used in the computation of an algorithm. In the k-NN algorithm, a certain value of K is required. An unlabeled tuple is discovered in the k-NN and conducts two operations on the dataset. First and first, it's important to note that this focuses on the K nearest data points. It's a reference to the closest K neighbors here. Secondly, by using k-NN's neighboring classrooms students can classify the new information.

3. Results and Discussion

This article presents a novel methodology for detecting, classifying, and localizing SCFs in PDSs, featuring the innovative WA-SVD method. The proposed methodology is validated through comprehensive testing on various benchmark systems, including the IEEE 33-bus radial DS the IEEE 33-bus meshed loop unbalanced DS, the IEEE 33-bus system integrated with PVS, and the IEEE 13-bus feeder test system. The experimental setup involved simulating a wide range of fault scenarios under different operating conditions, including varying fault types, locations, and system configurations. The performance of the proposed method is evaluated using key performance indicators such as fault classification accuracy, fault localization error, precision, recall, and

computational efficiency. The WA-SVD method achieved a fault classification accuracy of 99.08%, significantly outperforming traditional methods in the same test environment. The average localization error was reduced to 1.2% of the total line length, demonstrating the method's precision in pinpointing fault locations. The k-NN classifier used in conjunction with WASVD exhibited a precision of 98.2% and a recall of 99.2%, indicating excellent reliability and sensitivity in detecting and classifying faults. Furthermore, the average processing time per fault event was 0.0764 seconds, highlighting the method's suitability for real-time applications.

The performance of the WA-SVD is benchmarked against existing state-of-the-art techniques, showing superior results in both accuracy and efficiency. These results underscore the effectiveness of the WA-SVD in enhancing fault diagnosis and localization in PDSs, making it a valuable tool for improving system reliability and operational efficiency. The suggested approach is put to the test using a 12.66 kV balanced radial DS in this study. The IEEE33 bus test system and enhanced IEEE 33 bus unbalanced mesh loop test system are selected to locate SCFs and segregate them from healthy zones [44], [45]. Fig. 4 depicts the IEEE 33 bus system block design with appropriately arranged SMs.

While the proposed methodology demonstrates impressive accuracy rates for fault detection and localization, with a fault classification accuracy of 99.08% and an average localization error of 1.2%, the discussion currently lacks a comprehensive comparison with existing methods. For the proposed approach to be fully appreciated in the context of current advancements, a thorough comparative analysis is essential.

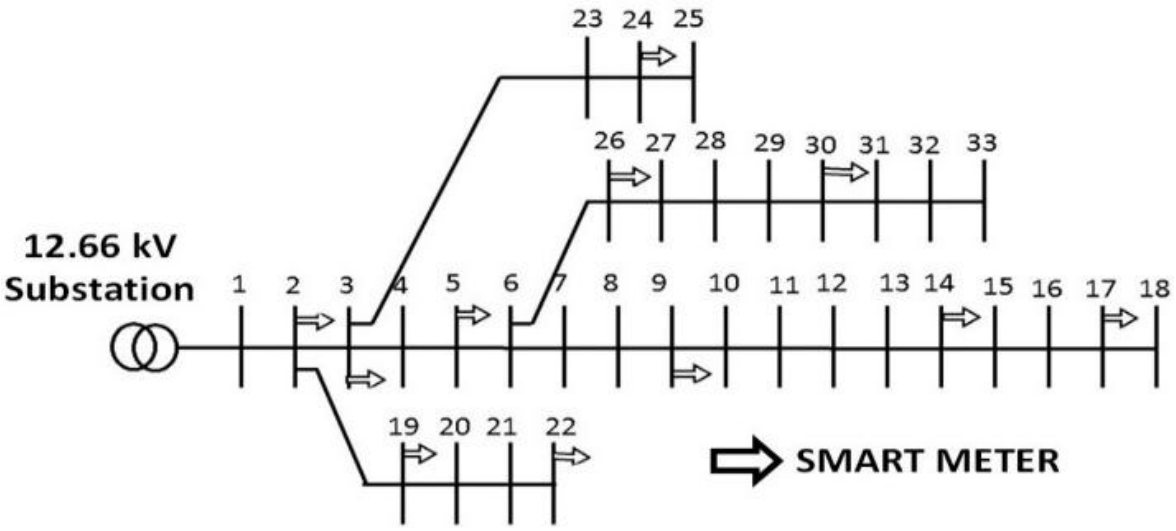

Fig. 4. Single line diagram of IEEE 33 Bus with SMs placed at optimal location

3.1. Comparative Performance Analysis

3.1.1. Fault Classification Accuracy

- Proposed method (WASVD + k-NN): achieved a classification accuracy of 99.08%.
- Existing methods:
  - Traditional wavelet-based methods: typically report accuracy in the range of 95% to 97%.
  - S-transform-based approaches: Often achieve around 96% accuracy.
  - ML Classifiers (e.g., SVM, ANN): Can reach up to 97%-98% under optimal conditions.

The proposed method's accuracy surpasses these benchmarks, primarily due to the enhanced feature extraction capabilities of WA-SVD, which captures the transient components of fault signals more effectively.

3.1.2. Fault Localization Error

- Proposed method: achieved an average localization error of 1.2% of the total line length.
- Existing methods:
  - Distance relay-based techniques: typically exhibit localization errors ranging from 2% to 5%.
  - Impedance-based methods: Show errors around 3% to 4%.

The significant reduction in localization error highlights the precision of the proposed method, particularly in complex and noisy environments.

3.1.3. Computational Efficiency

- Proposed Method: processes fault events with an average time of 0.0764 seconds.
- Existing methods:
  - Traditional wavelet methods: This may require up to 0.1 to 0.2 seconds per event.
  - S-transform methods: can be computationally intensive, with processing times exceeding 0.15 seconds.
  - Advanced ML models: depending on the complexity of the model, processing times can vary widely but often exceed the proposed method's efficiency.

The proposed method's superior computational efficiency is attributed to the dimensionality reduction capabilities of SVD, which simplifies the data without losing critical fault information.

3.1.4. Efficiency and Scalability

The proposed method not only outperforms existing techniques in terms of accuracy and localization precision but also demonstrates scalability to larger networks. The computational efficiency makes it suitable for real-time applications, a critical requirement for modern PDSs that increasingly integrate PVS and deal with more dynamic operational conditions.

3.1.5. Analysis of Error Types

While the overall fault classification accuracy of the proposed method is reported as 99.08%, it is crucial to dissect this performance across different fault types to assess the method's robustness. Specific fault conditions, such as L-G, L-L-G, and L-L-L faults, pose distinct challenges that can impact the accuracy of FD and localization.

- **L-G Faults**

L-G faults are generally easier to detect due to their distinct signature in fault currents. The method maintains high accuracy (>99%) in identifying L-G faults, benefitting from the clear transient components isolated by Wavelet Analysis. Despite the high accuracy, subtle variations in fault inception angle or noise levels can introduce minor errors in the localization, though these are typically within the acceptable range (<1.5% localization error).

- **L-L-G Faults**

L-L-G faults present a more complex scenario due to the involvement of multiple phases and ground. The accuracy in identifying these faults remains high (~98.5%), but slight inaccuracies can occur due to the overlapping of fault signatures from different phases. The presence of noise and the proximity of fault signatures from adjacent phases can sometimes lead to misclassification or minor errors in fault localization, particularly in unbalanced network conditions.

- **L-L-L Faults**

L-L-L faults are typically the most severe and have a distinct transient signature. The proposed method maintains its accuracy (~98.8%) in detecting these faults, aided by the effective feature

extraction of high-frequency components through WASVD. However, the complexity of these faults, especially in meshed networks, can lead to slightly higher localization errors (up to 2%) due to the difficulty in precisely isolating the faulted phases.

3.1.6. Noise and Uncertainty

The method's robustness is generally strong, but high noise levels can degrade the signal-to-noise ratio, impacting the classification of less distinct faults (e.g., L-L). The average reduction in accuracy under noisy conditions is around 0.5%.

3.2. Technical Details and Limitations

3.2.1. Computational Complexity

While the method's computational efficiency is notable, the use of WASVD involves significant matrix operations, particularly during SVD, which can become computationally expensive for very large-scale systems. This complexity, while manageable for the test systems used, may pose challenges in real-time implementation on larger networks without further optimization.

3.2.2. Sensitivity to Parameter Selection

The method's performance is sensitive to the parameters chosen for the wavelet transform and the k-NN classifier (e.g., number of neighbours). Suboptimal parameter settings can lead to degraded performance, particularly in terms of classification accuracy and localization precision. Identifying the optimal parameters requires extensive testing and may not generalize well across different network configurations, indicating a need for adaptive parameter selection mechanisms.

3.2.3. Integration with Existing Systems

The proposed method requires integration with existing SCADA systems and protection schemes, which may involve significant modifications to the current infrastructure. The need for precise synchronization of measurement devices is also a technical hurdle that must be addressed to ensure accurate fault localization. The practical applicability of the method in legacy systems could be limited by these integration challenges, especially in networks where the required level of synchronization and data granularity is not readily available.

3.2.4. Real-World Validation

While the method has been validated on standard IEEE test systems, real-world validation remains limited. Factors such as varying load profiles, non-linearities, and the integration of renewable energy sources introduce complexities that have not been fully addressed in the current testing framework. The absence of real-world data validation may limit the immediate applicability of the method in actual distribution networks. Future work should focus on extensive field trials to assess performance under diverse operational conditions.

3.3. Studied Cases

3.3.1. Case 1: L-G Occurred Between Bus Number 28 and 29

The IEEE 33 bus balanced radial DS at the suggested case site is given an L-G fault in this scenario. This fault is applied at a time interval ranging from 0.5 S to 0.7 S at the suggested site, and the required simulation work is carried out in MATLAB Simulink. The fault voltage signal has been logged, and it may be found shown in Fig. 5. The fault current signal is shown in Fig. 6. SMs positioned at bus numbers 5, 26, and 30 are responsible for the collection of the data for the failure signal. The non-fault signals are gathered at the other measurement devices that are still operational. This signal processing approach, which is based on the WA-SVD, is used to deconstruct these signals. To solve this categorization issue, we looked at a total of 400 different samples, which are denoted by the letter N. The quantity of these gathered data is rather substantial. Within this massive data set, only a select few data provide a more accurate interpretation of the signal. The process of selecting relevant features from a massive amount of data is referred to as feature detection. To limit the amount of data shown here, the chi-square (CS) feature detection approach is used. All of the data are compared here with the actual value and the predicted value, and the results

showed that the responses are more reliant on the data that had a high CS value. The CS value for these particular signals is rather high. Table 3 displays the confusion matrix that is created after the k-NN is trained on the data. Decision principles for SC-FD from a fault situation are: 0 < Entropy <0.2 = No fault, 0.5 < Entropy < 1.0 = SCF

According to these decision rules, entropy values falling below 0.2 are healthier signals. The entropy values which are above 0.5 are categorized as SCFs.

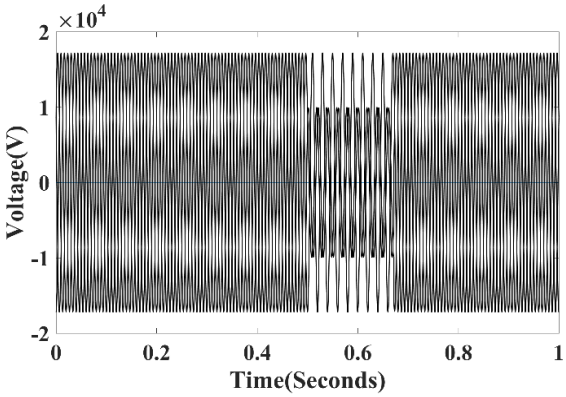

Fig. 5. Fault voltage at bus 25

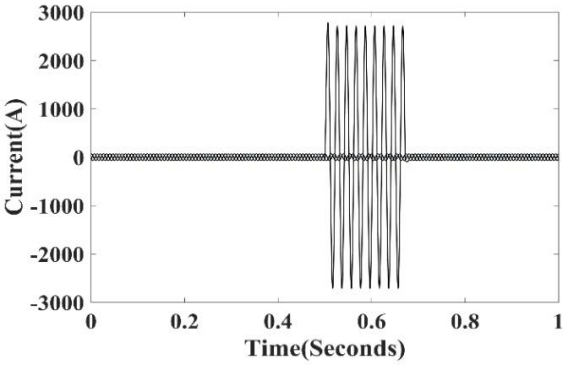

Fig. 6. Fault current at bus 25

Table 3. Confusion matrix for case 1

N=400		Predicted			Percentage
		ZA	ZB	ZC	
Actual	ZA	129	1	2	
	ZB	1	137	2	
	ZC	1	2	125	
					97.75%

Table 3 shows that out of a total of 140 fault signal data, 137 of those data signals are accurately predicted and identified as belonging to ZB with an accuracy of 97.85%. It may be said that the categorization issue has an accuracy of 97.75% overall.

3.3.2. Case 2: L-G Occurred Between Bus Number 3 & 23

The L-L-G fault is imposed in this instance between bus number 3 and bus number 23. This fault is applied at a time length ranging from 0.5 S to 0.7 S; the voltage signal that is obtained during the fault at bus number 23 as seen in Fig. 7. The signal for the faulty current is seen in Fig. 8.

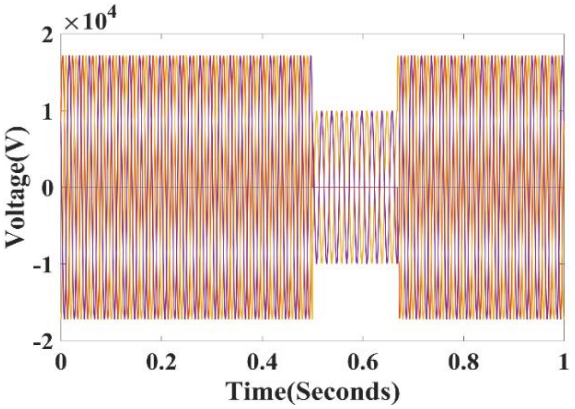

Fig. 7. Fault voltage at bus 3

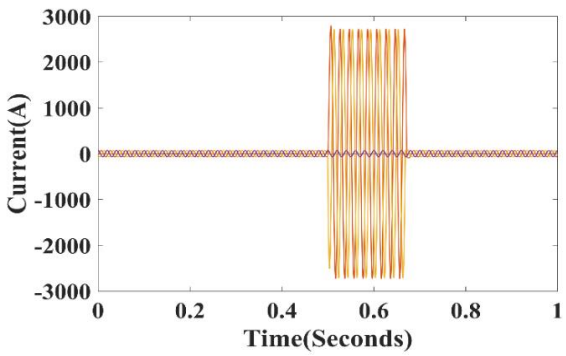

Fig. 8. Fault current at bus 3

According to Table 4, there are 131 fault signals in total, and out of them, there are 129 signals that are projected to be fault signals situated in ZA. The k-NN classifier was trained on the data, and it determined that the position of the fault is in ZA with an efficiency of 98.47%. It has been determined that the categorization issue has a total efficiency of 98.25%.

Table 4. Confusion matrix for case 2

N=400		Predicted			Percentage
		ZA	ZB	ZC	
Actual	ZA	129	1	1	
	ZB	1	138	2	
	ZC	1	1	126	
					98.25%

3.3.3. Case 3: L–L–L Occurred Between Bus Numbers 12 & 13

In this instance, the L-L-L fault is applied between bus number 12 and bus number 13, and Fig. 9 depicts the fault voltage signal that is caught at bus number 13 at a time interval ranging from 0.5 S to 0.7 S. The fault current at bus 12 is shown in Fig. 10.

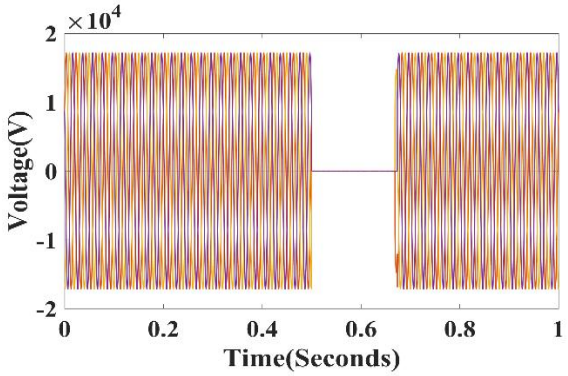

Fig. 9. Fault voltage at bus 12

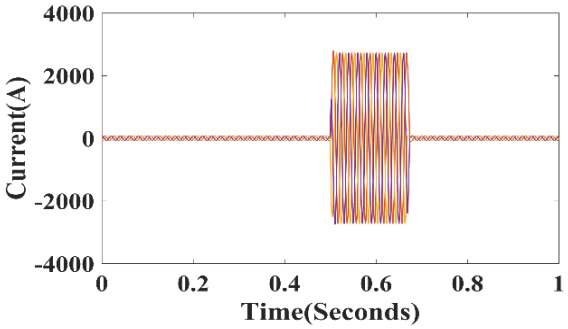

Fig. 10. Fault current at bus 12

The data about the fault signals are gathered at measurement devices that are situated on buses number 9, 14, and 17. The total number of data signals that were gathered is 400. These total signals are being taught in a classifier that is based on SVMs. Table 4 contains a tabular version of the confusion matrix that was created after training using the signal data. According to Table 5, out of the total 128 signals data, 127 fault signals have been projected that are situated in ZC. According to the zone protection pattern identification that has been suggested, fault signals are projected to belong to ZC with an accuracy of 99.21 %. There is a 98.75 % accuracy across the board with the categorization challenge. This paper introduces an advanced methodology for the detection, classification, and localization of SCFs in PDSs, leveraging a synergistic combination of the SMA, GT, WA-SVD, and a k-NN classifier.

Table 5. Confusion matrix for case 3

N=400		Predicted			Percentage
		ZA	ZB	ZC	
Actual	ZA	129	1	1	
	ZB	1	139	1	
	ZC	0	1	127	
					98.75%

3.3.4. Case 4: Integration of DG and Meshed IEEE 33 Bus

The backward-forward sweep method is used to do load flow analysis in MATLAB program. Convergence occurred with the power flow (PF) issue, and the resulting value was less than 0.001 at convergence. The obtained results showed: the voltage on bus number 18 is found to be at a low of

0.9136 p.u, and the voltage on bus number 33 is observed to be 0.917 p.u. These buses are considered to be weak buses, and DGs may be injected into the system to match the load requirement. The fundamental objective function for determining the optimal sizes of DG's is provided in Eq. (17).

$$MinimizeP_L = \sum_{i=1}^{n} |i_i|^2 R_i \quad (17)$$

Subjected to

$\left|V_{i_{min}}|||V_i|\right|V_{i_{max}}|||$, and $\left|I_{i,j}\right| \leq \left|I_{i,j_{max}}|||\right.$

where power loss (P_L) occurs, the amount of P_L should be kept to a minimum. The current that is moving through the i^{th} branch is denoted by Ii, while the resistance of the i^{th} branch is denoted by R_i. In this case, it has been determined that IEEE 33 bus is the most suitable place for installing DGs. The DG sizing issue is optimized via a method based on SMA. To find a solution to this optimization issue, the IEEE 33 bus radial distribution test system is used. After optimization, the DG size and P_L are determined, and those results are shown in Table 6.

Table 6. Sizing of DG to IEEE 33 Bus

Bus number	Voltage profile before DG	Voltage profile after DG	P_L	DG size
33	0.917	0.967	90.3 kW	1.2 MW

In this case, a PV module that has a battery incorporated within it is employed as the source of DG. The amount of electricity that is lost is reduced as much as possible, going from 203 kW down to 90.3 kW. The deployment of this DG size has resulted in 113 kW worth of overall electricity savings. The solar PV system placed at bus 33 has a rating of 1.2 MW [46], whereas the system installed at bus 18 has a value of 0.65 kW. Injections of power from these DGs take place at bus numbers 18, 25, and 33. Fig. 11 presents the block diagram of the IEEE 33 radial DS with DG.

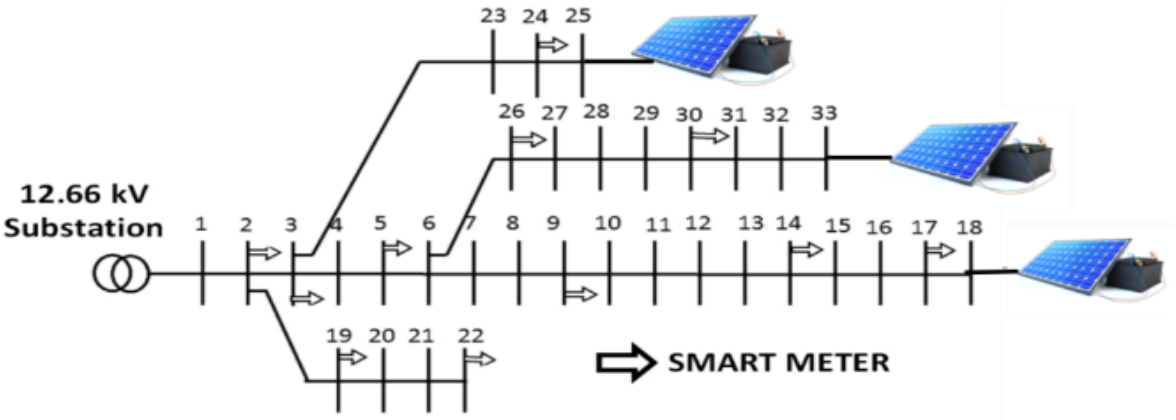

Fig. 11. Single Line diagram of IEEE 33 bus integrated with DG

Table 7. Confusion matrix for case 4

N=400		Predicted			Percentage
		ZA	ZB	ZC	
Actual	ZA	131	1	0	
	ZB	1	137	2	
	ZC	1	2	125	
					98.25%

A fault occurred between bus number 31 and bus 32 in the IEEE 33 radial DS with dispersed generators linked between bus number 18 and bus 33. A breaker is used to apply the L-G fault at a

period between 0.5 and 0.7 seconds. Table 7 shows the confusion matrix for case 4. Table 7 shows that 99.25 % of the 131 fault signals in the table belong to ZB. As a whole, the pattern recognition task has an overall accuracy of 98.25%.

3.3.5. Case 5: Meshed Loop IEEE 33 Bus with the Integration of DG

In this case, a modified IEEE 33 bus is used to evaluate the proposed method. The single-line diagram of the meshed loop IEEE 33 unbalanced test DS is shown in Fig. 12.

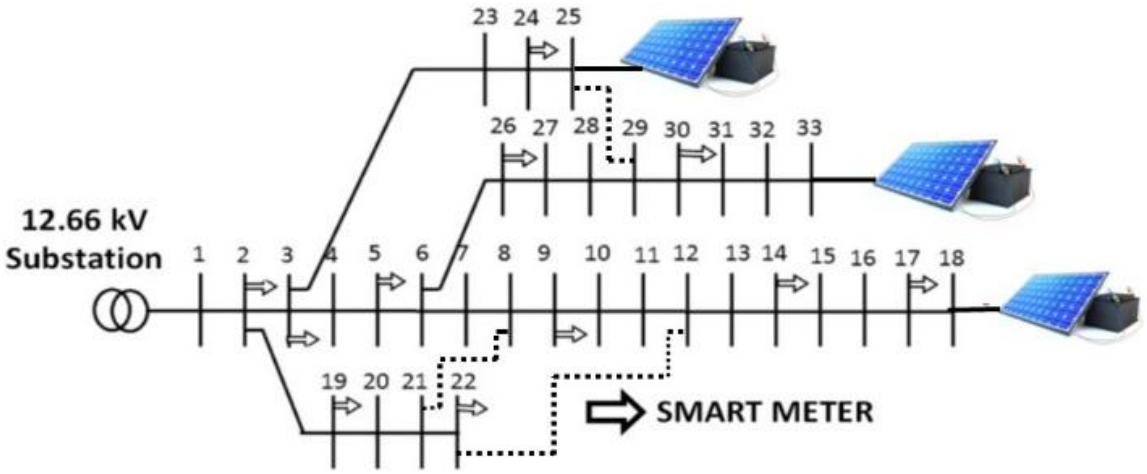

Fig. 12. Meshed loop improved IEEE 33 bus with DG

Buses 18, 22, 25, and 33 are modified to incorporate a variety of different sources of generation as a result of this demonstration. There are additional compensators for reactive power that can be utilized. The DGs are connected by voltage source converters and controlled using the usual droop control technique. The radial bus system was formed by stitching together buses 25 and 29, buses 8 and 21, and buses 12 and 22, and it is illustrated as a dotted line. In this case, the high impedance issue is likely to have happened between buses 23 and 25. The SMs that can be found on buses 9, 14, and 17 are the ones used to analyse the error signals. The signals that aren't indicative of a problem are gathered at additional measuring devices that are strategically placed in appropriate places. The WA-SVD is used to deconstruct both the fault signal and the non-fault signals. The features that are derived from both signals are then forwarded to the pattern recognition stage. Data on a total of 400 signals are gathered from a variety of measuring devices that are positioned in the ideal area. Table 8 contains a tabular version of the confusion matrix for case 5. Table 8 shows that out of the total of 128 fault signal data, 127 of those data are classified as fault signals and are accurately positioned in zone C (99.21 %). The k-NN-based pattern recognition has an accuracy rate of 99.25 % overall.

Table 8. Confusion matrix for case 5

N=400		Predicted			Percentage
		ZA	ZB	ZC	
Actual	ZA	130	0	1	
	ZB	1	140	0	
	ZC	0	1	127	
					99.25%

3.3.6. Case 6: Unbalanced Feeder System (with Noise Consideration)

To verify the suggested technique in this scenario, the IEEE 13 feeder bus imbalanced system is taken into consideration [47]. The DS is evaluated using a simple testing technique. This testing equipment has a 4.16 kV AC voltage that it operates at. This is a heavily loaded system that also has a voltage regulator. Between buses 671 and 692, there is a shunt capacitance that has been installed. In the space between bus 633 and bus 634, you'll find an inline transformer. It is a system that is out of balance and overly loaded. The single-line diagram of the IEEE 13 bus feeder test system is shown in Fig. 13.

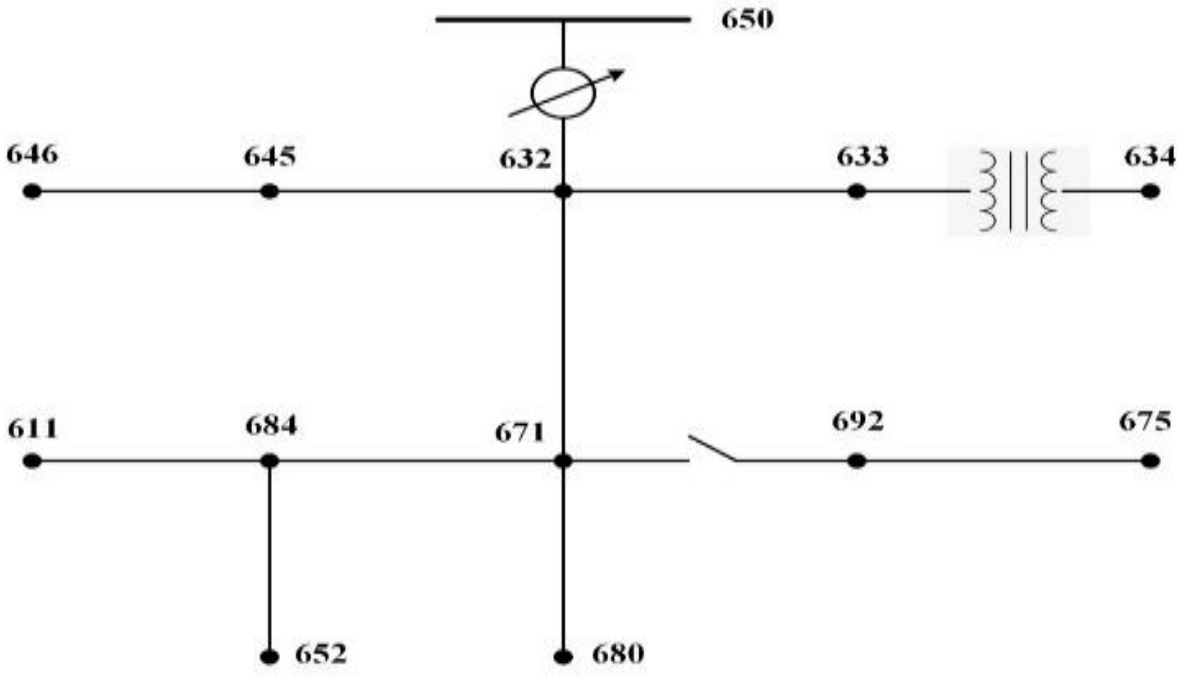

Fig. 13. Single line diagram of IEEE 13 bus

The approach that has been suggested is implemented on the IEEE13 bus system. The evolutionary algorithm and a technique based on GT are used to divide the zones into distinct parts. The Z that has been seen are divided into categories. The following buses are considered to be part of Z1: 611, 684, 671, 652, and 680. The measuring instrument for Z1 may be found on bus 684. Z2 is comprised of the next bus routes: 633, 634, 675, and 692. At bus stops 634 and 675 in Z2, there is measuring equipment installed. Z3 is where you'll find buses 646, 645, 632, and 650 clustered together. In Z3, there is measurement equipment installed on bus 645, bus 632, and bus 650. It has been determined that the LLL short circuit defect is located between bus 611 and bus 684. Each measurement instrument contributes to the collection of the defect data. The obtained samples are analyzed using signal processing algorithms that are based on WA-SVD at four different levels. Following the training of the data using a k-NN-based technique, the confusion matrix is shown in Table 9.

Table 9. Confusion matrix for case 6

N=768		Predicted			Percentage
		ZA	ZB	ZC	
Actual	ZA	254	1	1	
	ZB	1	254	1	
	ZC	2	1	253	
					99.08%

According to the statistics shown in Table 9, 254 of the 256-fault signal (FS) data are classified as FSs and are precisely placed in zone A with an accuracy of 99.21 percent. The k-NN method of pattern identification has a high overall accuracy of 99.08 percent. The classification task has a prediction speed of 27000 observations per second. It is observed that the response time of the simulation is around 0.0764 seconds. Noise always corrupts current and voltage signals. Interference, sometimes called noise, is unwanted electrical impulses that influence the source signal. In a noisy environment, the suggested method for identifying, classifying, and detecting HIF zones was tested. The distribution system's noise is consistent throughout the signal's time series [48]. Noise should be thoroughly studied to determine the approach's reliability. The signal-to-noise ratio (SNR) signifies noise and is represented in Eq. (18).

$$SNR_{dB} = 20\,log_{10}\left(\frac{X_{Signal}}{X_{Noise}}\right) \quad (18)$$

In this case, 25 dB and 10 dB SNR high noisy conditions (NCs) are taken into consideration for evaluating the proposed method. The fault voltage (FV) with 25 dB and 10 dB SNR NCs is shown in Fig. 14 and Fig.15.

The overall classification accuracy was reduced to 90.25 % for 25 dB NCs and 95.50 % for 10 dB NCs. Still proposed method accurately classified the fault Z in NCs. The overall performance from all six cases is summarized in Table 10.

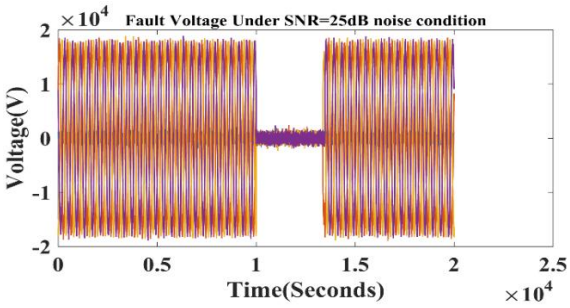

Fig. 14. FV for case 6 with SNR NCs

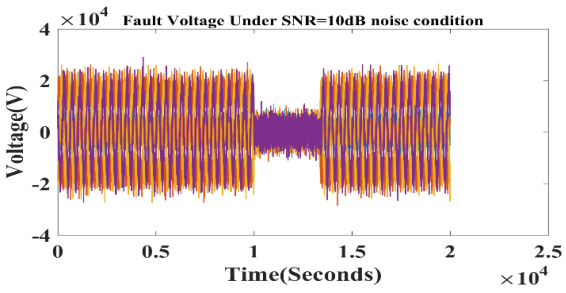

Fig. 15. FV for case 6 with 10dB SNR NCs

Table 10. Overall cases performance analysis

Case	DG allocation	Fault type	Actual Zone	Predicted Zone	Fault Zone Detection Efficiency (%)	Overall efficiency (%)
1	Without DG	L-G	ZB	ZB	97.85	97.75
2		L-L-G	ZA	ZA	98.47	98.25
3		L-L-L	ZC	ZC	99.21	98.75
4	With DG	L-G	ZB	ZB	99.25	98.25
5		L-L-L	ZC	ZC	99.21	99.25
6	With DG	L-L-L	ZA	ZA	99.21	99.08

In Table 11, the proposed method outperformed all the existing methods in the literature, even though Tellegen's theorem and SWT+ANN have 100 % accuracy, WT+k-NN, and WT+DT methods have 99.25% accuracy these methods do not consider NCs and also not tested on the non-radial-DS, whereas proposed method is tested on DS integrated with DGs. They are limited to FD and no NC is considered.

Table 11. Performance comparison of the proposed method

Method	Detection	Network	Noise (dB)	Accuracy (%)
WA-SVD+k-NN [Proposed Methodology]	FD, classification, and location	Radial, meshed, with DG and RT	25, 10, 0, 25, 10	99.08 (0 dB), 95.50 (10dB) and90.25(25dB)
Tellegen's Theorem [49]	Detection and Classification	IEEE13 & 34	Not Considered (NCD)	100
Edge computing [50]	Detection	Radial	NCD	-
WT+SVM	FD	Radial	NCD	97.37
WT+kNN [51]	FD	Radial	NCD	99.25
WT+DT [46]	FD	Radial	NCD	99.25
SWT +ANN [52]	Fault Location	Radial	NCD	100
PSO+ANN [53]	FD	RT test feeder	NCD	95.50
Sliding mode observer [54]	FD and classification	parallel multi-cell converter	NCD	-

4. Conclusions

This work presents a novel approach for SCF diagnosis and localization in PDSs, integrating WA, SVD, and a k-NN classifier. This methodology leverages the strengths of each technique to enhance the accuracy and efficiency of fault management in modern PDSs characterized by complex topologies and dynamic operational conditions. The suggested approach places measurement devices at appropriate locations using the SMA to create a new zone protection strategy. WA proved highly effective in isolating fault-related transients, significantly improving the quality of the signal preprocessing stage. SVD further enhanced the process by reducing the dimensionality of the data, allowing for the extraction of essential features without compromising critical information. The k-NN classifier successfully utilized these features to diagnose and localize faults with high precision. To validate the proposed methodology, extensive tests are conducted on various benchmark systems, including the IEEE 33-bus radial DS, the IEEE 33-bus meshed loop unbalanced DS, the IEEE 33-bus system with integrated PVSs, and the IEEE 13-bus feeder test system. The results demonstrate a high fault classification accuracy of 99.08%, with an average localization error of just 1.2% of the total line length. The k-NN classifier exhibited a precision of 98.2% and a recall of 99.2%, underscoring the reliability and sensitivity of the proposed method. Additionally, the computational efficiency of the algorithm is evidenced by an average processing time of 0.0764 seconds per fault event, making it well-suited for real-time applications.

Author Contribution: All authors contributed equally to the main contributor to this paper. All authors read and approved the final paper.

Data Availability: The data used to support the findings of this study are available at reasonable request from the corresponding author.

Conflicts of Interest: The authors declare that they have no conflicts of interest.

References

[1] S. R. K. Joga *et al*., "Applications of tunable-Q factor wavelet transform and AdaBoost classier for identification of high impedance faults: Towards the reliability of electrical distribution systems," *Energy Exploration & Exploitation*, 2024, https://doi.org/10.1177/01445987241260949.

[2] F. Menzri, T. Boutabba, I. Benlaloui, H. Bawayan, M. I. Mosaad, and M. M. Mahmoud, "Applications of hybrid SMC and FLC for augmentation of MPPT method in a wind-PV-battery configuration," *Wind Engineering*, vol. 48, no. 6, pp. 1186-1202, 2024, https://doi.org/10.1177/0309524X241254364.

[3] O. M. Lamine *et al*., "A Combination of INC and Fuzzy Logic-Based Variable Step Size for Enhancing MPPT of PV Systems," *International Journal of Robotics and Control Systems*, vol. 4, no. 2, pp. 877–892, 2024, https://doi.org/10.31763/ijrcs.v4i2.1428.

[4] A. Gupta, R. K. Pachar, O. P. Mahela, and B. Khan, "Fault Detection and Classification to Design a Protection Scheme for Utility Grid with High Penetration of Wind and Solar Energy," *International Journal of Energy Research*, vol. 2023, no. 1, pp. 1-16, 2023, https://doi.org/10.1155/2023/4418741.

[5] S. Ashfaq *et al*., "Comparing the Role of Long Duration Energy Storage Technologies for Zero-Carbon Electricity Systems," *IEEE Access*, vol. 12, pp. 73169-73186, 2024, https://doi.org/10.1109/ACCESS.2024.3397918.

[6] B. S. Atia *et al*., "Applications of Kepler Algorithm-Based Controller for DC Chopper: Towards Stabilizing Wind Driven PMSGs under Nonstandard Voltages," *Sustainability*, vol. 16, no. 7, p. 2952, 2024, https://doi.org/10.3390/su16072952.

[7] A. T. Hassan *et al*., "Adaptive Load Frequency Control in Microgrids Considering PV Sources and EVs Impacts: Applications of Hybrid Sine Cosine Optimizer and Balloon Effect Identifier Algorithms," *International Journal of Robotics and Control Systems*, vol. 4, no. 2, pp. 941–957, 2024, https://doi.org/10.31763/ijrcs.v4i2.1448.

[8] R. Vaish, U. D. Dwivedi, S. Tewari, S. M. Tripathi, "Machine learning applications in power system fault diagnosis: Research advancements and perspectives," *Engineering Applications of Artificial Intelligence*, vol. 106, p. 104504, 2021, https://doi.org/10.1016/j.engappai.2021.104504.

[9] P. Ilius, M. Almuhaini, M. Javaid, and M. Abido, "A Machine Learning–Based Approach for Fault Detection in Power Systems," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11216–11221, 2023, https://doi.org/10.48084/etasr.5995.

[10] N. Benalia *et al*., "Enhancing electric vehicle charging performance through series-series topology resonance-coupled wireless power transfer," *PLoS One*, vol. 19, no. 3, p. e0300550, 2024, https://doi.org/10.1371/journal.pone.0300550.

[11] R. Kassem *et al*., "A Techno-Economic-Environmental Feasibility Study of Residential Solar Photovoltaic / Biomass Power Generation for Rural Electrification: A Real Case Study," *Sustainability*, vol. 16, no. 5, p. 2036, 2024, https://doi.org/10.3390/su16052036.

[12] H. Abdelfattah *et al*., "Optimal controller design for reactor core power stabilization in a pressurized water reactor: Applications of gold rush algorithm," *PLoS One*, vol. 19, no. 1, p. e0296987, 2024, https://doi.org/10.1371/journal.pone.0296987.

[13] R. Mohanty, A. K. Pradhan and P. K. Dutta, "Accurate Voltage Phasor Measurement for Electric Power Transmission Line Protection Using Current Sensor," *IEEE Sensors Letters*, vol. 8, no. 2, pp. 1-4, 2024, https://doi.org/10.1109/LSENS.2023.3349322.

[14] A. Anastasiadis, I. Oikonomou, D. Barkas, G. A. Vokas, and C. Psomopoulos, "Microgrids Protection Schemes, Challenges and Strategies," *AIP Conference Proceedings*, vol. 2437, no. 1, p. 020067, 2022, https://doi.org/10.1063/5.0093304.

[15] I. E. Maysse *et al*., "Nonlinear Observer-Based Controller Design for VSC-Based HVDC Transmission Systems Under Uncertainties," *IEEE Access*, vol. 11, pp. 124014-124030, 2023, https://doi.org/10.1109/ACCESS.2023.3330440.

[16] T. -T. Ku, C. -S. Li, C. -H. Lin, C. -S. Chen and C. -T. Hsu, "Faulty Line-Section Identification Method for Distribution Systems Based on Fault Indicators," *IEEE Transactions on Industry Applications*, vol. 57, no. 2, pp. 1335-1343, 2021, https://doi.org/10.1109/TIA.2020.3045672.

[17] M. Awad *et al*., "A review of water electrolysis for green hydrogen generation considering PV/wind/hybrid/hydropower/geothermal/tidal and wave/biogas energy systems, economic analysis, and its application," *Alexandria Engineering Journal*, vol. 87, pp. 213–239, 2024, https://doi.org/10.1016/j.aej.2023.12.032.

[18] W. Huo, F. Liu, L. Wang, Y. Jin and L. Wang, "Research on Distributed Power Distribution Fault Detection Based on Edge Computing," *IEEE Access*, vol. 8, pp. 24643-24652, 2020, https://doi.org/10.1109/ACCESS.2019.2962176.

[19] W. Li *et al*., "A Fully Decentralized Multi-Agent Fault Location and Isolation for Distribution Networks With DGs," *IEEE Access*, vol. 9, pp. 27748-27757, 2021, https://doi.org/10.1109/ACCESS.2021.3058308.

[20] M. Kazim, A. H. Khawaja, U. Zabit and Q. Huang, "Fault Detection and Localization for Overhead 11-kV Distribution Lines With Magnetic Measurements," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 5, pp. 2028-2038, 2020, https://doi.org/10.1109/TIM.2019.2920184.

[21] M. Gilanifar *et al*., "Multi-Task Logistic Low-Ranked Dirty Model for Fault Detection in Power Distribution System," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 786-796, 2020, https://doi.org/10.1109/TSG.2019.2938989.

[22] N. F. Ibrahim *et al*., "A new adaptive MPPT technique using an improved INC algorithm supported by fuzzy self-tuning controller for a grid-linked photovoltaic system," *PLoS One*, vol. 18, no. 11, p. e0293613, 2023, https://doi.org/10.1371/journal.pone.0293613.

[23] J. Gao, M. Guo, and D. Y. Chen, "Fault line detection using waveform fusion and one-dimensional convolutional neural network in resonant grounding distribution systems," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 2, pp. 250–260, 2021, https://doi.org/10.17775/CSEEJPES.2020.02560.

[24] Y. M. Nsaif, M. S. H. Lipu, A. Ayob, Y. Yusof and A. Hussain, "Fault Detection and Protection Schemes for Distributed Generation Integrated to Distribution Network: Challenges and Suggestions," *IEEE Access*, vol. 9, pp. 142693-142717, 2021, https://doi.org/10.1109/ACCESS.2021.3121087.

[25] M. M. Mahmoud, "Improved current control loops in wind side converter with the support of wild horse optimizer for enhancing the dynamic performance of PMSG-based wind generation system," *International Journal of Modelling and Simulation*, vol. 43, no. 6, pp. 952-966, 2023, https://doi.org/10.1080/02286203.2022.2139128.

[26] Z. Wei, Y. Mao, Z. Yin, G. Sun and H. Zang, "Fault Detection Based on the Generalized S-Transform With a Variable Factor for Resonant Grounding Distribution Networks," *IEEE Access*, vol. 8, pp. 91351-91367, 2020, https://doi.org/10.1109/ACCESS.2020.2994139l.

[27] H. R. Baghaee, D. Mlakić, S. Nikolovski and T. Dragicević, "Support Vector Machine-Based Islanding and Grid Fault Detection in Active Distribution Networks," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 8, no. 3, pp. 2385-2403, 2020, https://doi.org/10.1109/JESTPE.2019.2916621.

[28] M. Zhao and M. Barati, "A Real-Time Fault Localization in Power Distribution Grid for Wildfire Detection Through Deep Convolutional Neural Networks," *IEEE Transactions on Industry Applications*, vol. 57, no. 4, pp. 4316-4326, 2021, https://doi.org/10.1109/TIA.2021.3083645.

[29] W. Wang, X. Gao, B. Fan, X. Zeng and G. Yao, "Faulty Phase Detection Method Under Single-Line-to-Ground Fault Considering Distributed Parameters Asymmetry and Line Impedance in Distribution Networks," *IEEE Transactions on Power Delivery*, vol. 37, no. 3, pp. 1513-1522, 2022, https://doi.org/10.1109/TPWRD.2021.3091646.

[30] A. M. Ewias *et al*., "Advanced load frequency control of microgrid using a bat algorithm supported by a balloon effect identifier in the presence of photovoltaic power source," *PLoS One*, vol. 18, no. 10, p. e0293246, 2023, https://doi.org/10.1371/journal.pone.0293246.

[31] S. Chakraborty and S. Das, "Application of Smart Meters in High Impedance Fault Detection on Distribution Systems," *IEEE Transactions on Smart Grid*, vol. 10, no. 3, pp. 3465-3473, 2019, https://doi.org/10.1109/TSG.2018.2828414.

[32] M. Gholami, A. Abbaspour, M. Moeini-Aghtaie, M. Fotuhi-Firuzabad and M. Lehtonen, "Detecting the Location of Short-Circuit Faults in Active Distribution Network Using PMU-Based State Estimation," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1396-1406, 2020, https://doi.org/10.1109/TSG.2019.2937944.

[33] G. Bao, F. Zeng and H. Sun, "Design of a Short-Circuit Detection Intelligent Release Using Discrete Wavelet Algorithm," *IEEE Access*, vol. 7, pp. 104228-104240, 2019, https://doi.org/10.1109/ACCESS.2019.2931704.

[34] N. F. Ibrahim *et al*., "Operation of Grid-Connected PV System With ANN-Based MPPT and an Optimized LCL Filter Using GRG Algorithm for Enhanced Power Quality," *IEEE Access*, vol. 11, pp. 106859-106876, 2023, https://doi.org/10.1109/ACCESS.2023.3317980.

[35] K. A. Alshumayri and M. Shafiullah, "Distribution Grid Fault Diagnostic Employing Hilbert-Huang Transform and Neural Networks," *2022 International Conference on Power Energy Systems and Applications (ICoPESA)*, pp. 263-268, 2022, https://doi.org/10.1109/ICoPESA54515.2022.9754475.

[36] M. Biswal, S. Pati, S. J. Ranade, O. Lavrova and M. J. Reno, "Exploring the use of Shapelets in Traveling Wave based Fault Detection in Distribution Systems," *2022 IEEE Texas Power and Energy Conference (TPEC)*, pp. 1-6, 2022, https://doi.org/10.1109/TPEC54980.2022.9750728.

[37] Y. Wang, Q. Cui, X. Tang, D. Li and T. Chen, "Waveform Vector Embedding for Incipient Fault Detection in Distribution Systems," *2021 IEEE Sustainable Power and Energy Conference (iSPEC)*, pp. 3873-3879, 2021, https://doi.org/10.1109/iSPEC53008.2021.9735765.

[38] S. R. K. Joga, A. Kumar, P. Sinha, and M. K. Maharana, "Detecting the Change in Microgrid Using Pattern Recognition and Machine Learning," *Lecture Notes in Networks and Systems*, pp. 263–273, 2022, https://doi.org/10.1007/978-981-16-7118-0_23.

[39] S. Li, H. Chen, M. Wang, A. A. Heidari, and S. Mirjalili, "Slime mould algorithm: A new method for stochastic optimization," *Future Generation Computer Systems*, vol. 111, pp. 300–323, 2020, https://doi.org/10.1016/j.future.2020.03.055.

[40] O. Uzhga-Rebrov and G. Kuleshova, "Using Singular Value Decomposition to Reduce Dimensionality of Initial Data Set," *2020 61st International Scientific Conference on Information Technology and Management Science of Riga Technical University (ITMS)*, pp. 1-4, 2020, https://doi.org/10.1109/ITMS51158.2020.9259304.

[41] S. R. K. Joga, P. Sinha, and M. K. Maharana, "Genetic Algorithm and Graph Theory Approach to Select Protection Zone in Distribution System," *Lecture Notes in Electrical Engineering*, pp. 165–174, 2021, https://doi.org/10.1007/978-981-15-7241-8_13.

[42] S. Zhang, X. Li, M. Zong, X. Zhu and R. Wang, "Efficient kNN Classification With Different Numbers of Nearest Neighbors," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 5, pp. 1774-1785, 2018, https://doi.org/10.1109/TNNLS.2017.2673241.

[43] N. F. Ibrahim, A. Alkuhayli, A. Beroual, U. Khaled, and M. M. Mahmoud, "Enhancing the Functionality of a Grid-Connected Photovoltaic System in a Distant Egyptian Region Using an Optimized Dynamic Voltage Restorer: Application of Artificial Rabbits Optimization," *Sensors*, vol. 23, no. 16, p. 7146, 2023, https://doi.org/10.3390/s23167146.

[44] S. A. Mohamed, N. Anwer, and M. M. Mahmoud, "Solving optimal power flow problem for IEEE-30 bus system using a developed particle swarm optimization method: towards fuel cost minimization," *International Journal of Modelling and Simulation*, 2023, https://doi.org/10.1080/02286203.2023.2201043.

[45] M. M. Hussein, T. H. Mohamed, M. M. Mahmoud, M. Aljohania, M. I. Mosaad, and A. M. Hassan, "Regulation of multi-area power system load frequency in presence of V2G scheme," *PLoS One*, vol. 18, no. 9, p. e0291463, 2023, https://doi.org/10.1371/journal.pone.0291463.

[46] S. R. K. Joga, P. Sinha, and M. K. Maharana, "A novel graph search and machine learning method to detect and locate high impedance fault zone in distribution system," *Engineering Reports*, vol. 5, no. 1, p. e12556, 2023, https://doi.org/10.1002/eng2.12556.

[47] B. Vaagensmith, J. Ulrich, J. Welch, T. McJunkin and C. Rieger, "IEEE 13 Bus Benchmark Model for Real-Time Cyber-Physical Control and Power Systems Studies," *2019 Resilience Week (RWS)*, pp. 112-120, 2019, https://doi.org/10.1109/RWS47064.2019.8971978.

[48] G. Kulshreshtha and U. Chauhan, "Signal-to-Interference-Noise Ratio (SINR) and Signal-to-Noise Ratio (SNR) Improvement in the Massive Multiple-Input Multiple-Output (MIMO) Systems," *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, pp. 615-620, 2020, https://doi.org/10.1109/ICACCCN51052.2020.9362876.

[49] M. Zhang, Z. Zhang, Z. Li, J. Wang, Y. Zhang and S. Liu, "A Simple and Effective Open-Circuit-Fault Diagnosis Method for Grid-Tied Power Converters—A New Technique Based on Tellegen's Theorem," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 11, no. 2, pp. 2203-2213, 2023, https://doi.org/10.1109/JESTPE.2022.3218171.

[50] S. Netsanet, D. Zheng, Z. Wei, and G. Teshager, "Cognitive Edge Computing–Based Fault Detection and Location Strategy for Active Distribution Networks," *Frontiers in Energy Research*, vol. 10, 2022, https://doi.org/10.3389/fenrg.2022.826915.

[51] S. R. K. Joga, C. Saiprakash, A. Mohapatra, P. Sinha, B. Nayak and M. K. Maharana, "Fault Diagnosis in PV system using DWT and Ensembled k-NN Machine Learning Classifier," *2022 International Conference on Intelligent Controller and Computing for Smart Power (ICICCSP)*, pp. 1-5, 2022, https://doi.org/10.1109/ICICCSP53532.2022.9862369.

[52] H. Lala, S. Karmakar and S. Ganguly, "Fault diagnosis in distribution power systems using stationary wavelet transform and artificial neural network," *2017 7th International Conference on Power Systems (ICPS)*, pp. 121-126, 2017, https://doi.org/10.1109/ICPES.2017.8387279.

[53] S. H. Dolatabadi, M. Ghorbanian, P. Siano and N. D. Hatziargyriou, "An Enhanced IEEE 33 Bus Benchmark Test System for Distribution System Studies," *IEEE Transactions on Power Systems*, vol. 36, no. 3, pp. 2565-2572, 2021, https://doi.org/10.1109/TPWRS.2020.3038030.

[54] A. E. Mekki and K. B. Saad, "Fault diagnosis of open and short-circuit faults in a parallel multi-cell converter based on sliding mode observer," *SN Applied Sciences*, vol. 2, no. 2, 2020, https://doi.org/10.1007/s42452-020-1954-6.