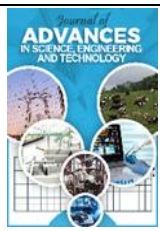
MMUST PUBLISHING PRESS

Journal of Advances in Science, Engineering and Technology



Early Detection and Classification of Power Systems Faults in Medium Voltage Distribution Lines Based on Sparse Representation

John Ogumbo[a], Cedric Okinda[a,b,*], James Kulubi[a]

[a]*Department of Electrical and Communications Engineering, School of Engineering and Built Environment, Masinde Muliro University of Science and Technology, Kakamega, Kenya.*

[b]*College of Engineering, Laboratory of Modern Facility Agriculture Technology and Equipment Engineering of Jiangsu Province, Nanjing Agricultural University, Jiangsu 210031, P.R. China.*

ABSTRACT

The increasing demand for reliable and resilient electrical power necessitates advanced fault detection strategies in Medium Voltage Distribution Lines (MVDLs), which account for over 80% of power outages. This study proposes a novel approach for the early detection and classification of Short Circuit Faults (SCFs) in MVDLs using a Sparse Representation (SR) framework combined with Mel-Frequency Cepstral Coefficients (MFCCs). The method was validated through simulations on an IEEE 5-Bus System, where Line-to-Ground (L-G), Line-to-Line (L-L), and Double Line-to-Ground (L-L-G) faults were induced under both noise-free and noisy (25 dB SNR) conditions. Feature vectors derived from voltage and current signals were transformed into MFCCs, and SR was applied via dictionary learning and Orthogonal Matching Pursuit to extract fault-relevant signatures. Classification was performed using Support Vector Machines trained on sparse codes. Two models were developed: one for fault type classification and another for combined fault-type and localization using zonal segmentation. Results show that the SR-based classifier significantly outperformed conventional models (ANN, SVM) in both accuracy and noise robustness. The SR model achieved 98.7% accuracy in noise-free and 96.8% in noisy settings, with F1-scores exceeding 96%. Localization accuracy ranged from 97.5% (2-zones) to 90.5% (4-zones). The proposed SR-based classifier demonstrated high computational efficiency and scalability, enabling real-time applicability for medium-voltage grid monitoring. This research demonstrates the efficacy of integrating SR and MFCCs for robust MVDL fault diagnostics, with potential applications in enhancing grid stability, enabling predictive maintenance, and minimizing service disruptions.

ARTICLE INFO

Keywords:
Medium Voltage Distribution Lines;
Mel-Frequency Cepstral Coefficients;
Power System Fault Classification;
Short Circuit Fault Detection;
Sparse Representation.

Article History:
Received 18 May 2025
Received in revised form 6 October 2025
Accepted 7 October 2025
Available online 14 October 2025

1. Introduction

The global energy demand is projected to reach approximately 660 quadrillion British thermal units (Btu) by 2050, representing a 15% increase from 2021 levels (Onatunji et al., 2024). This rise is primarily attributed to continuous population growth and accelerating industrialization (Onatunji et al., 2024). Consequently, there is an urgent need to develop power supply systems that are both efficient and resilient. Modern electrical power systems are intricate and highly interconnected networks. Ensuring their stability and reliability is critical to maintaining an uninterrupted supply of electricity. Transmission lines, which constitute the structural backbone of these systems, are particularly vulnerable to a wide range of electrical faults, i.e., short circuits, open circuits, and high-impedance faults (Gururajapathy et al., 2017). If not identified and mitigated swiftly, these faults can lead to cascading failures, equipment damage, and large-scale blackouts (Phadke & Thorp, 2008). The increasing integration of renewable energy sources and surging electricity consumption further complicate grid operations, highlighting the necessity for advanced fault detection and classification mechanisms (Carrasco et al., 2006).

One of the most critical segments of the power distribution infrastructure is the Medium Voltage Distribution Line (MVDL), which connects transmission networks to end consumers. Faults in this segment significantly affect the reliability and quality of power delivery. Notably, over 80% of power supply outages originate from MVDL failures, with the remainder

* Corresponding author. e-mail: cokinda@mmust.ac.ke
Editor: Gershom Mutua, Masinde Muliro University of Science and Technology, Kenya.
Citation: Ogumbo J., Okinda C., & Kulubi J. (2025). Early Detection and Classification of Power Systems Faults in Medium Voltage Distribution Lines Based on Sparse Representation. Journal of Advances in Science, Engineering and Technology 2(1), 36 – 50.

attributed to faults in cables, transformers, and other components (Bindi et al., 2023; Fahim et al., 2021). Among these, Short Circuit Faults (SCFs) are the most prevalent, accounting for approximately 85 – 87% of all power system faults (Yumurtaci et al., 2016). SCFs occur when low-impedance paths are unintentionally created between conductors, resulting in excessive current flow that severely compromises system integrity. Therefore, accurate and timely detection of SCFs in MVDLs is essential for ensuring operational stability (Li et al., 2022).

Multiple factors contribute to the incidence of SCFs in MVDLs, i.e., human error (Ivanova et al., 2020), environmental stressors such as lightning and vegetation (Doostan et al., 2020), aging infrastructure (Anyaka & Ozioko, 2020), and insulation deterioration (Bindi et al., 2023). These conditions can severely damage critical infrastructure components, such as transformers and switchgear systems (Han & Song, 2003; Subramaniam et al., 2021). Additionally, circuit breakers subjected to severe SCFs are prone to failure due to contact degradation (Subramaniam et al., 2021). Beyond the technical consequences, SCFs have substantial economic implications. They reduce utility revenue and impose costs on consumers through service interruptions (Ran et al., 2019). SCFs also pose significant fire hazards due to electrical arcing. For instance, SCFs have been implicated in 2–4% of all rural fires in Australia (Bowman et al., 2020), leading to property damage, ecological harm, and threats to human and animal life (Furse et al., 2020).

The adverse consequences associated with SCFs can be substantially mitigated through the deployment of efficient and reliable fault detection systems (Furse et al., 2020). Traditional fault detection techniques, such as time-overcurrent, differential, and distance relays, exhibit inherent limitations that significantly compromise their detection accuracy (Alasali et al., 2023). Specifically, time-overcurrent protection operates by comparing the current flowing through a line against a predefined threshold, rendering it ineffective for detecting incipient faults (Swain et al., 2022). Differential protection schemes, which assess the current differential between the input and output of a line, encounter performance challenges under unbalanced load conditions (Alasali et al., 2023). Similarly, distance protection relies on impedance measurements, which can misidentify high-impedance faults as being outside the designated protection zones, leading to erroneous relay operations (Vlahinić et al., 2020). These limitations underscore the necessity for advanced, data-driven fault detection methodologies.

Recent advancements that leverage artificial intelligence (AI) to overcome the constraints of traditional detection systems (Bakkar et al., 2022) have enhanced fault identification accuracy and reliability across variable system conditions (Wong et al., 2021). Wavelet and Fourier analysis have emerged as robust signal processing tools, particularly when integrated with machine learning algorithms such as artificial neural networks (ANN), decision trees (DT), and support vector machines (SVM) to facilitate timely and precise fault detection (Alimi et al., 2020; Vaish et al., 2021). Wavelets decompose electrical signals into multiple frequency bands, enabling detailed temporal and spectral analysis (Guo et al., 2022). This multi-resolution capability enhances the detection of fault signatures, thereby reducing the likelihood of false positives (Fahim et al., 2021; Nasser Mohamed et al., 2023).

Moreover, this hybrid approach facilitates the proactive identification of incipient faults before they escalate into severe equipment failures, thus supporting predictive maintenance strategies and minimizing unplanned outages (Bindi et al., 2023). Machine learning techniques further enhance system resilience by enabling continuous adaptation to evolving operating conditions, ensuring sustained optimal performance (Bagwari et al., 2023). Collectively, the integration of wavelet-based signal processing with AI-driven analytics presents a viable pathway to achieving stable and reliable MVDLs (Alimi et al., 2020). Nonetheless, the efficiency of such advanced fault detection systems depends upon the availability of high-quality, representative data for training and validation purposes (Cano et al., 2024; Mahmood & Wang, 2021). Furthermore, the computational demands associated with real-time implementation may pose challenges in resource-constrained environments (Sirojan et al., 2018).

As already mentioned, the current approaches have improved fault diagnosis but still face limitations in computational efficiency and generalization (Chen et al., 2016). Sparse representation (SR) offers a potential solution by enabling compact signal representations that enhance feature discrimination and reduce computational burden (Chen et al., 2001). However, its application in power system fault detection remains underexplored. The primary objective of this study was to develop an efficient and robust SCF detection and classification system that enhances the performance, stability, and reliability of MVDL. With the specific objectives of developing an efficient model that simulates MVDL and SCFs, developing a robust signal processing algorithm that extracts fault wave features with high predictive abilities, and developing and validating a fault detection and recognition system based on the extracted wave features under noise and no noise conditions. The development of this proposed system presents a proactive approach to address the challenges associated with Short Circuit Fault (SCF) in MVDL to enhance grid reliability, improve safety, and optimize operational efficiency.

2. Materials and Methods

This study aims to develop an SCF fault classifier for an MVDL based on the SR technique. The approach involves the simulation of an IEEE 5-Bus System and simulating SCFs on its MVDLs (11 kV nominal voltage), collecting the voltage (V(t)) and current (I(t)) signals (time series data), extracting the Mel-Frequency Cepstral Coefficients (MFCC) from the collected data,

and developing and validating a classification model based on the MFCC to detect the type of SCF. Additionally, another model was developed to detect the SCF and its location in the Network. For comparative analysis, SR, Support Vector Machine (SVM), and Artificial Neural Network (ANN) were explored. The research design flow of the introduced study is given in Fig. 1.

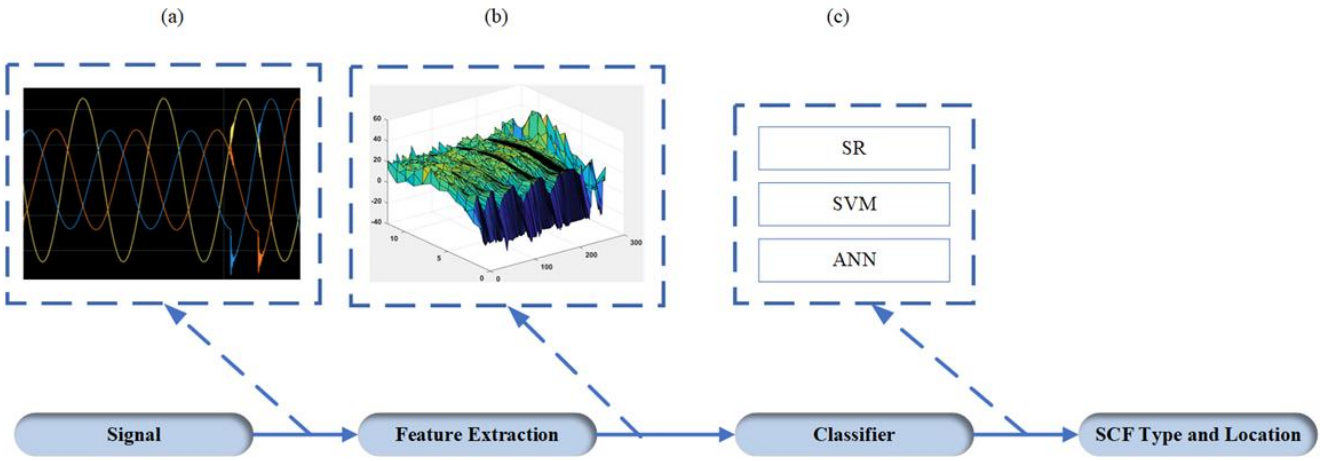


Fig. 1: The research design flow of the proposed system. (a) Signal for the Power System, (b) Extracted feature variables, and (c) Developed classifiers.

2.1. The IEEE 5-Bus System Modeling

The IEEE 5-Bus System was simulated using MATLAB R2023b (The MathWorks Inc., Natick, MA) on a standard workstation (Intel i7, 3.2 GHz, 16 GB RAM) to model MVDLs and replicate various SCF scenarios. This system was selected due to its simplicity and widespread use in power system studies, offering a balance between complexity and computational tractability (Saadat, 1999; Salam, 2020; Ten & Hou, 2024). The simulated system comprises five Buses interconnected by transmission lines, with each Bus representing a node in the power network. The nominal voltage of the MVDLs was set to 11 kV, reflecting typical operational conditions in MVDLs (Gururajapathy et al., 2017). The simulated IEEE 5-Bus System is shown in Fig. 2. The simulation included the following components: Generators: Two synchronous generators were modeled to supply power to the network, each with a rated capacity of 100 MVA and a nominal voltage of 33 kV. The generators were configured to simulate realistic power flow dynamics and transient responses during fault conditions (Phadke & Thorp, 2008). Transformers: Two step-down transformers were incorporated to interface between the generators and the MVDLs.

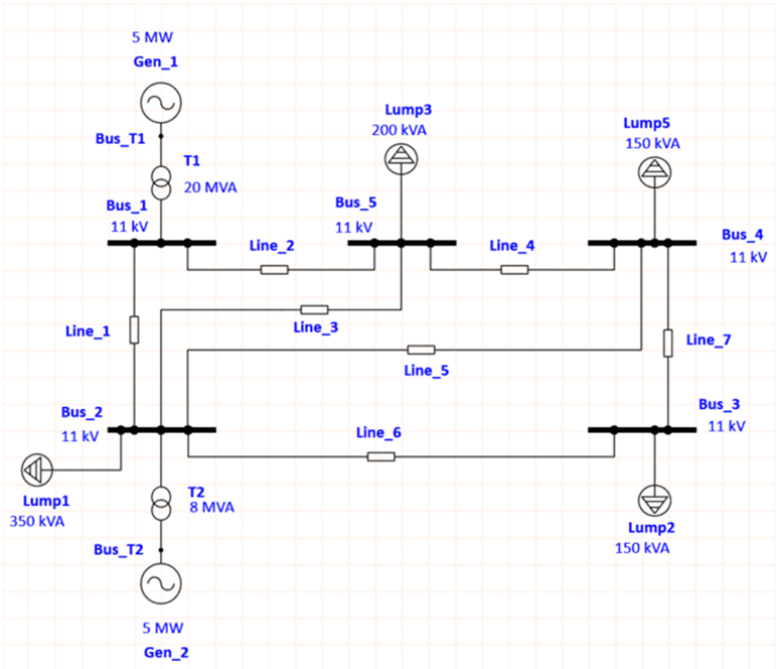


Fig. 2: The simulated IEEE 5-Bus System.

The transformers were modeled with a turns ratio of 11 kV/33 kV to simulate voltage transformation typical in distribution systems (Li et al., 2022). Transmission Lines: Seven MVDLs were represented using distributed parameter line models to account for impedance, capacitance, and inductance effects. The lines were

configured with parameters derived from standard IEEE benchmarks to ensure realistic behavior (Saadat, 1999; Salam, 2020; Ten & Hou, 2024). This approach captures the dynamic behavior of the lines under fault conditions. Lamped Loads: Resistive and inductive loads were connected at specific buses to simulate varying demand profiles. The loads were adjusted to reflect typical industrial and residential consumption patterns. Fault Modules: SCFs, including line-to-ground (LG), line-to-line (LL), and double line-to-ground (L-L-G) faults, were simulated at strategic locations along each MVDL. Additionally, fault resistances were varied to replicate high- and low-impedance fault conditions (Alasali et al., 2023).

2.2. Data collection

The data collection process was designed to extract V(t) and I(t) signals from all the Buses of the simulated IEEE 5-Bus System under both normal and SCF conditions. The three types of SCF were simulated in each MVDL, and data were collected when SCF occurred per line. In this study, ten SCFs were simulated per MVDL (random locations along the line). Thus, for each MVDL, 350 data points were collected, resulting in a total of 1050 data points. The total simulation time was set to 30 s. Based on previous studies, SCF duration is 20 – 500 ms (Lukeyo et al., 2025). Therefore, in this study, the SCF occurrence duration was set to 50 ms. Table 1 presents a summary of all the data collected.

Table 1: The types of SCFs explored in this study.

Short Circuit Faults		
Line-to-Ground (L-G)	**Line-to-Line (L-L)**	**Double Line-to-Ground (L-L-G)**
R-G	R-Y	R-Y-G
Y-G	Y-B	R-B-G
B-G	R-B	Y-B-G

R is Red phase, Y is Yellow phase, B is Blue phase, and G is Ground.

Additionally, the simulations incorporated noise (25dB SNR) to validate robustness, ensuring the classifier's applicability in noisy grid environments (Chen et al., 2016). This comprehensive approach enabled the extraction of discriminative features (e.g., MFCCs) for accurate fault classification (Rezapour et al., 2023).

2.3. Feature extraction

Feature extraction is a critical step in the fault detection and classification process, as it transforms raw $V(t)$ and $I(t)$ signals into discriminative features that can be effectively utilized by machine learning models. In this study, MFCCs were employed due to their ability to capture signal characteristics in both temporal and spectral domains, particularly in noisy environments (Li et al., 2019). The MFCC extraction process involves several stages, i.e., pre-processing, framing, windowing, Fourier transformation, Mel-frequency filtering, and cepstral analysis (Abdul & Al-Talabani, 2022; Sidhu et al., 2024).

2.3.1. Pre-processing

The raw $V(t)$ and $I(t)$ signals collected from the IEEE 5-Bus System were first subjected to pre-processing to enhance their quality and remove artifacts. This included noise reduction using a bandpass filter to eliminate frequencies outside the relevant range for fault analysis (20 Hz to 2 kHz) (Guo et al., 2022). Additionally, normalization was applied to ensure the signals were scaled uniformly, mitigating the influence of amplitude variations caused by differing fault conditions or measurement inconsistencies (Altaie et al., 2024).

2.3.2. Framing and Windowing

The pre-processed signals were divided into short, overlapping frames to capture localized features. Each frame had a duration of 25 ms with a 10 ms overlap, a configuration widely adopted in signal processing to balance temporal resolution and computational efficiency (Vaish et al., 2021; Zhang et al., 2019). To minimize spectral leakage, each frame was multiplied by a Hamming window, defined by Eq. 1.

$$\omega(n) = 0.54 - 0.46\,cos\left(\frac{2\pi n}{N-1}\right) \quad (1)$$

Where N is the frame length, this process ensured smooth transitions at frame boundaries, enhancing the accuracy of subsequent frequency-domain analysis (Fahim et al., 2021).

2.3.3. Fourier Transform and Power Spectrum

Each windowed frame was transformed into the frequency domain using the Discrete Fourier Transform (DFT), which decomposes the signal into its constituent frequencies (Zhang et al., 2019) according to Eq. 2. The power spectrum was then computed by squaring the magnitude of the DFT coefficients, providing a representation of the signal's energy distribution across frequencies (Alimi et al., 2020; Xia et al., 2015). This process is crucial for identifying fault-induced harmonics and transient components that are indicative of specific fault types, i.e., L-G, L-L, or L-L-G.

$$X[k] = \sum_{n=0}^{N-1} x[n].\,e^{-j\frac{2\pi}{N}Kn} \qquad 0 \le k \ge N-1 \quad (2)$$

Where $X[k]$ is the DFT output at frequency index k, $x[n]$ is the input signal at time index n, N is the total number of samples, and j is the imaginary unit.

2.3.4. Mel-frequency Filtering

This process is particularly beneficial for distinguishing the occurrence of SCFs from normal operating conditions (Nasser Mohamed et al., 2023; Xian et al., 2023). The power spectrum was mapped onto the Mel scale, a perceptually motivated scale that approximates human auditory response (Juvela, 2015). This involved applying a set of triangular Mel-spaced filters to the power spectrum, with center frequencies spaced linearly below 1 kHz and logarithmically above 1 kHz (Deng et al., 2004; Juvela, 2015). Finally, the Mel filter outputs were log-compressed to emphasize subtle spectral variations (Kumar, 2024). Eq. 3 presents the Mel filter bank output for each filter m. Where $H_m(k)$ is the m^{th} triangular filter response at frequency bin k given by Eq. 4, such that $f(m)$ is the Fast Fourier Transform

(FFT) bin index corresponding to the center frequency of the m^{th} Mel filter,

$$S[m] = \ln\left(\sum_{k=0}^{N-1}(X[k].H_m(k))\right) \qquad 0 \le m \ge M \quad (3)$$

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \dfrac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \le k < f(m) \\ \dfrac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \le k < f(m+1) \\ 0 & k \ge f(m-1) \end{cases} \quad (4)$$

The rationale for employing MFCCs in power system fault analysis lies in their ability to emulate the human auditory system's nonlinear sensitivity to frequency. Fault-induced electrical transients exhibit non-stationary, wide-band characteristics with dominant low- and mid-frequency harmonic components similar to the spectral envelope variations in acoustic signals. The Mel filter bank's logarithmic frequency scaling enhances the representation of these low-frequency transients while compressing less informative high-frequency regions, thereby improving discriminability among different fault types. Moreover, the cepstral transformation decorrelates spectral features and separates the slowly varying spectral envelope (steady-state operation) from rapid spectral fluctuations (fault events). This property allows MFCCs to capture both harmonic distortion and transient spikes effectively, which explains their strong performance in classifying SCFs under both noise-free and noisy conditions.

2.3.5. Cepstral Analysis

The final step involved computing the inverse DFT of the log Mel spectrum to obtain the cepstral coefficients (Abdul & Al-Talabani, 2022) given by Eq. 5. The lower-order coefficients capture broad spectral trends, while higher-order coefficients represent finer details (Abdul & Al-Talabani, 2022; Boucheron & De Leon, 2008; Ramirez et al., 2018). For this study, the first 13 MFCCs were retained, as they encapsulate the most discriminative information for distinguishing between fault types (Rustam et al., 2023; Sarath, 2022). Additionally, delta and delta-delta coefficients were derived to capture dynamic signal features, further enhancing the classifier's ability to detect transient fault signatures (Hossan et al., 2010; Mahmood & Wang, 2021).

$$x[n] = \frac{1}{N}\sum_{n=0}^{N-1} X[k].e^{j\frac{2\pi}{N}Kn} \qquad 0 \le k \ge N-1 \quad (5)$$

The extracted MFCCs, along with their delta and delta-delta coefficients, were concatenated into a feature vector for each signal segment. These vectors were then standardized to zero mean and unit variance to facilitate model training (Özesmi et al., 2006). The resulting feature set was partitioned into training and testing subsets (70:30 ratio) to evaluate the classifier's performance under both noise-free and noisy conditions.

2.4. Classification Algorithm

In power systems, accurate and timely classification and locating of SCFs is critical for ensuring the protection and reliability of MVDLs. SR has emerged as a powerful tool for signal processing and pattern recognition, particularly in scenarios where data can be efficiently represented by a small number of non-zero coefficients in a high-dimensional space (Cheng et al., 2013). Regarding SCFs, SR leverages the inherent sparsity of fault signatures in $V(t)$ and $I(t)$ signals to enhance detection accuracy and computational efficiency (Ahmed et al., 2022). The SR-based classification algorithm developed in this study involves three key steps: dictionary learning, sparse coding, and the classification task. Note that in this study, two models were developed, i.e., the first model (M_{d1}) was to only classify the SCF, and the second model (M_{d2}) was to classify the SCF and its location in the MVDL. Note that for M_{d2}, to limit the number of classification classes, the Power System was segregated into zones for comparative analysis, three Case Studies were explored in this study as presented in Table 2.

Table 2: Zoning of the Power System MVDL network.

Case Study	N° of Zones	Lines in the Zone	N° of Classes
CASE 1	2	**Zone 1**: Line_1, Line_2, Line_3 **Zone 2**: Line_4, Line_5, Line_6, Line_7	6
CASE 2	3	**Zone 1**: Line_1, Line_2, Line_3 **Zone 2**: Line_4, Line_5, Line_6 **Zone 3**: Line_7	9
CASE 3	4	**Zone 1**: Line_1, Line_2, **Zone 2**: Line_3, Line_5 **Zone 3**: Line_4, Line_6 **Zone 4**: Line_7	12

2.4.1. Dictionary Learning

The first step in the SR framework is the construction of a discriminative dictionary that can sparsely represent the fault features extracted from the signals. Traditional dictionaries, such as those derived from Fourier or wavelet bases, may not capture the unique characteristics of SCFs (Rubinstein et al., 2010). To address this challenge, a data-driven approach was adopted, where the dictionary was learned directly from the training dataset of MFCCs (Rida, 2020). The K-Singular Value Decomposition (K-SVD) algorithm was employed for dictionary learning, as it iteratively optimizes the dictionary atoms to minimize the reconstruction error while enforcing sparsity constraints (Aharon et al., 2006; Meng et al., 2020). The objective function for K-SVD is given by Eq. 6.

$$\min_{D,X} \|Y - DX\|_F^2 , \qquad \|x_i\|_0 \le T \quad \forall i \quad (6)$$

where Y is the matrix of training feature vectors, D is the dictionary, X is the sparse coefficient matrix, and T is the sparsity level (Mahmood & Wang, 2021). The resulting dictionary D (512 atoms) was tailored to the specific fault types (L-G, L-L, and L-L-G), ensuring robust representation of fault signatures under varying conditions.

2.4.2. Sparse Coding

Once the dictionary was learned, sparse coding was performed to represent each signal as a linear combination of a few dictionary atoms. In this study, the Orthogonal Matching Pursuit (OMP) algorithm was applied to perform sparse coding, which was due to its computational efficiency and robustness in noisy environments (Rubinstein et al., 2008). OMP iteratively selects the dictionary atoms that best approximate the signal residual to solve the optimization problem (Loza, 2019) given in Eq. 7, where y is the test feature vector, and x is the sparse coefficient vector. The sparsity constraint $T = 10$ ensured that only the most relevant features were retained, enhancing the discriminative power of the representation. Additionally, the sparsity level (T = 10) was empirically selected to optimize the trade-off between reconstruction accuracy and computational cost.

$$\min_{x} \|y - Dx\|_2^2 , \qquad \|x\|_0 \leq T \tag{7}$$

2.5. Classification Task

The final step involved classifying the fault type and locality based on the sparse coefficients. A multi-class Support Vector Machine (SVM) was employed for this task, leveraging its ability to handle high-dimensional data and non-linear decision boundaries (Doğan et al., 2016). The SVM was trained using the sparse coefficients of the training data, with the radial basis function (RBF) kernel selected for its flexibility in capturing complex feature relationships (Nyalala et al., 2019). The decision function for the SVM is given by Eq. 8, where α_i is the Lagrange multiplier, y_i is the class label, $K(x_i, x)$ is the kernel function, and b is the bias term (Vaish et al., 2021).

$$f(x) = sign\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, x) + b\right) \tag{8}$$

For comparative analysis, the SR-based classifier was compared against ANN and SVM trained on raw MFCCs. The SVM RBF kernel parameters ($C = 1.0, \gamma = 0.01$), and ANN architecture (3 hidden layers, 128 neurons each, learning rate = 0.001), were optimized via grid search with 5-fold cross-validation on the training set.

Table 3: The description of the Performance Metric.

Performance Metric	Description		
Accuracy	The overall correctness of the classifier	$\frac{TP+TN}{TP+TN+FP+FN}$	(9)
Precision	The reliability of positive predictions	$\frac{TP}{TP+FP}$	(10)
Recall (sensitivity)	The classifier's ability to detect all relevant faults	$\frac{TP}{TP+FN}$	(11)
F1-score	The harmonic means of precision and recall	$2\left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}\right)$	(12)

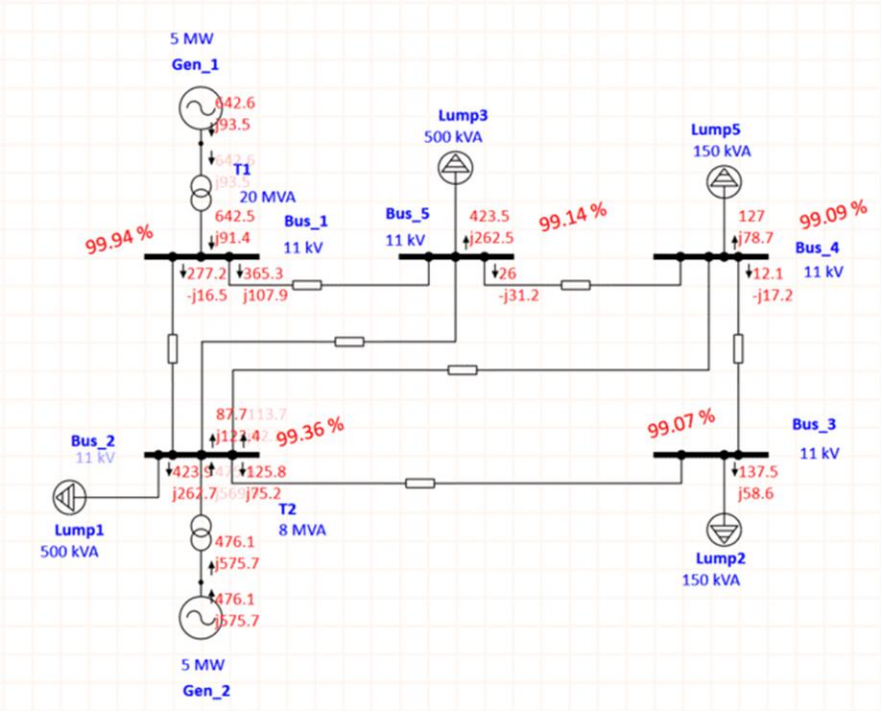


Fig. 3: The Load Flow analysis results for the IEEE 5-Bus System.

2.5.1. Classifier Performance Evaluation

To assess the effectiveness of the developed classifiers for SCF detection, multiple evaluation metrics and techniques were employed, i.e., confusion matrices, accuracy, precision, recall, and F1-score. A confusion matrix was applied to identify misclassifications and evaluate the classifier's discriminative power (Tharwat, 2021) by providing a detailed breakdown of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each fault type (L-G, L-L, and L-L-G). The description and Equations for accuracy, precision, recall, and F1-score are summarized in Table 3.

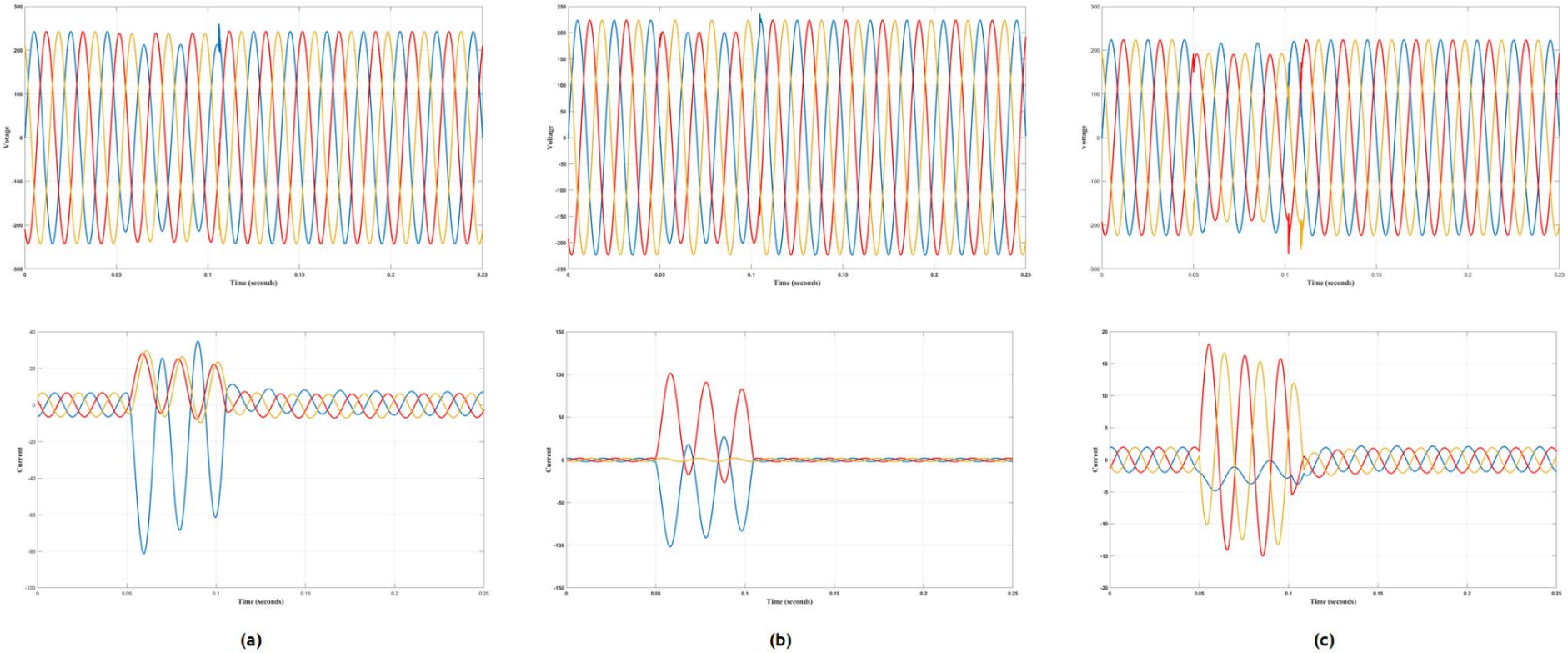


Fig. 4: The $V(t)$ and $I(t)$ signals for various simulated SCFs on the MVDL. (a) L-G SCF simulated on Line_1 $V(t)$ and $I(t)$ measurements taken from Bus_1, (b) L-L SCF simulated on Line_2 $V(t)$ and $I(t)$ measurements taken from Bus_2, and (c) L-L-G SCF simulated on Line_4 $V(t)$ and $I(t)$ measurements taken from Bus_4.

3. Results

3.1. IEEE 5 Bus System Simulation Evaluation

The simulated IEEE 5-Bus System was evaluated through load flow analysis (Fig. 3) to assess voltage stability, power flow distribution, and operational efficiency under normal conditions. It was established that Bus_2 and Bus_5 maintained voltages close to the nominal 11 kV, with Bus_2 at 0.9994 pu and Bus_5 at 0.9914 pu of the rated voltage, indicating stable grid conditions (Saadat, 1999; Salam, 2020; Ten & Hou, 2024). Bus_3 recorded a marginal drop to 0.9909 pu, likely due to the reactive power demand from Lump_5 (150 kVA), highlighting the impact of smaller loads on voltage regulation (Gururajapathy et al., 2017). Gen_1 and Gen_2 each supplied 5 MW, with power flows efficiently distributed across the network. The near-unity voltage percentages suggest minimal transmission losses, aligning with the system's balanced design (Phadke & Thorp, 2008).

Additionally, it was observed that there was a Reactive power injection (e.g., -j16.5 at Bus_2), which is critical for maintaining voltage stability under inductive loads (Li et al., 2022). Lump_1 (500 kVA) and Lump_3 (500 kVA) operated within capacity, with no overloading. The slight voltage dip at Bus_3 underscores the need for localized reactive power compensation (Alasali et al., 2023). The load flow analysis confirmed the IEEE 5-Bus System's robustness, providing a reliable foundation for fault detection studies. The minimal voltage deviations and efficient power flow distribution reflect a well-designed network. Fig. 4 shows the $V(t)$ and $I(t)$ signals for various simulated SCFs on the MVDL. The surf plot in Fig. 5 shows the extracted MFCC for the $V(t)$ and $I(t)$ for R-Phase during L-G fault.

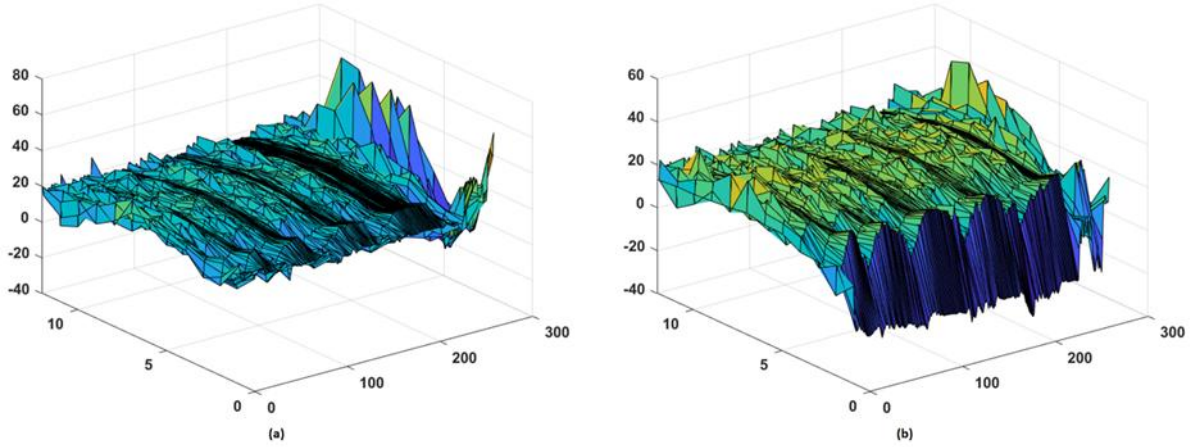


Fig. 5: The Surf Plot of the extracted MFCC. (a) MFCC for $V(t)$ for R-Phase during L-G fault, (b) MFCC for $I(t)$ for R-Phase during L-G fault.

3.2. Classification and Localization Model Evaluation

The study aimed to develop an efficient and robust SCF detection and classification system for MVDLs using SR. M_{d1} performed fault classification only, while M_{d2} extended this to include fault location identification. The performance of the proposed SR-based classifier was evaluated against conventional machine learning models, i.e., SVM and ANN, under both noise-free and noisy conditions. The results are presented in terms of classification accuracy, computational efficiency, and robustness to noise, with detailed comparisons across fault types L-G, L-L, and L-L-G.

Table 4: The comparison of the detection accuracy of the explored classifiers under noise-free conditions.

Classifier	Detection Accuracy (%)	
	Noise-Free Condition	Noisy Condition
SR	98.7	96.8
SVM	95.2	88.3
ANN	93.8	85.1

3.2.1. Classification Performance

The developed classifiers M_{d1} were evaluated using the standard evaluation metrics given in Table 3. It was observed that the SR-based model demonstrated superior performance in all metrics compared to SVM and ANN. Under noise-free conditions, the SR classifier achieved an average accuracy of 98.7%, while SVM and ANN achieved 95.2% and 93.8%, respectively (Table 4). The high accuracy of the SR model can be attributed to its ability to leverage sparse representations of fault signatures, which enhances feature discrimination and reduces misclassification (Yao et al., 2024). The confusion matrices for the SR, SVM, and ANN classifiers (Fig. 6) showed distinct patterns in misclassifications. The SR model exhibited minimal confusion between fault types, with only 0.9% of L-L faults misclassified as L-G. In contrast, SVM and ANN showed higher misclassification rates, particularly between L-L and L-L-G faults (4.6% and 5.6%, respectively). This underscores the advantage of SR in capturing discriminative features through sparse coding (Zhang et al., 2019).

The precision and recall values for the SR model were consistently high across all fault types (Table 5), indicating its reliability in identifying true positives while minimizing false alarms. For instance, the precision for L-G faults was 99.1%, reflecting the model's ability to correctly classify this fault type with minimal false positives. Similarly, the recall for L-L-G faults was 98.3%, demonstrating robust detection capabilities even for complex fault scenarios.

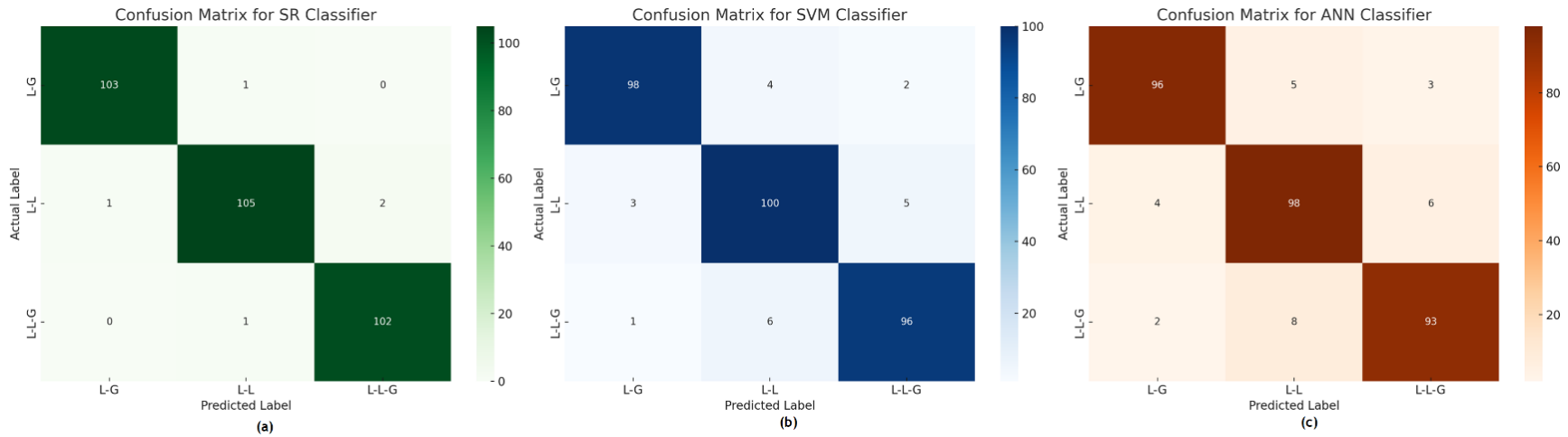


Fig. 6: Confusion matrices for the SR, SVM, and ANN classifiers under the Noise-Free Condition.

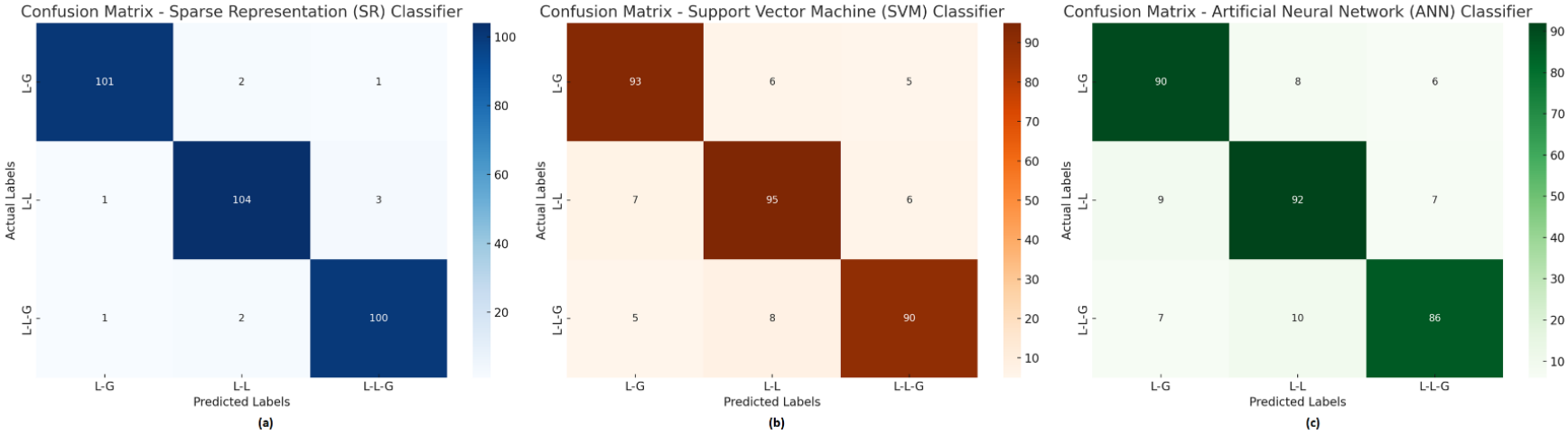


Fig. 7: Confusion matrices for the SR, SVM, and ANN classifiers under Noisy Condition.

Table 5: The evaluation results of the classification models based on the Performance Metric in the Noise-Free Condition.

Model	Metrics	L-G	L-L	L-L-G	Average
SR	Precision	99.1	98.1	98.1	**98.43**
	Recall	99.0	97.2	99.0	**98.40**
	F1-score	**99.0**	**97.6**	**98.5**	**98.37**
SVM	Precision	96.1	90.9	93.2	**93.40**
	Recall	94.2	92.6	93.2	**93.33**
	F1-score	**95.1**	**91.7**	**93.2**	**93.33**
ANN	Precision	94.1	88.3	91.2	**91.20**
	Recall	92.3	90.7	90.3	**91.10**
	F1-score	**93.2**	**89.5**	**90.7**	**91.13**

Additionally, the performance metric in the noisy condition is presented in Table 6. The resilience of the SR approach to noise can be attributed to the MFCCs, which effectively filter out irrelevant frequency components (Wu & Cao, 2005). Additionally, the sparse coding process inherently suppresses noise by focusing on the most significant signal features (Ngiam et al., 2011). The confusion matrices for the SR, SVM, and ANN classifiers under noisy conditions are presented in Fig. 7.

Table 6: The evaluation results of the classification models based on the Performance Metric in the Noisy Condition.

Model	Metrics	L-G	L-L	L-L-G	Average
SR	Precision	98.1	96.3	96.2	**96.87**
	Recall	97.1	96.3	97.1	**96.84**
	F1-score	97.6	96.3	96.6	**96.84**
SVM	Precision	88.6	87.2	89.1	**88.30**
	Recall	89.4	88.0	87.4	**88.27**
	F1-score	89.0	87.6	88.2	**88.27**
ANN	Precision	84.9	83.6	86.9	**85.13**
	Recall	86.5	85.2	83.5	**85.07**
	F1-score	85.7	84.4	85.1	**85.07**

The SR outperforms SVM and ANN in all metrics, with near-identical precision, recall, and F1-score (96.8%), confirming its robustness to noise. However, the SVM marginally outperforms ANN, but both struggle with L-L vs. L-L-G misclassifications due to noise-induced feature overlap. Nevertheless, all classifiers show balanced precision-recall trade-offs, with no significant bias toward false positives/negatives.

3.2.2. Fault Classification and Localization Performance

The second model M_{d2} was developed to classify SCFs and their locations, achieving the highest accuracy of 97.5%, 94.6%, and 92.7% for 2 Zones, 3 Zones, and 4 Zones, respectively, based on the SR classifier as presented in Table 7.

Table 7: The comparison of the detection accuracy of the explored Fault Classification and Localization classifiers under noise-free and noisy conditions.

N° of Zones	Noise-Free Condition			Noisy Conditions		
	SR	SVM	ANN	SR	SVM	ANN
2	97.5	94.7	93.1	95.8	83.0	81.2
3	94.6	92.1	90.8	92.3	80.8	78.1
4	92.7	89.9	88.6	90.5	78.6	75.9

It was established that the SR was 2.5 – 12.8% more accurate than SVM, which was also 1.3 – 2.7% accurate than ANN in both noise-free and noisy conditions. However, it was also observed that noise reduces accuracy by 2.3% (SR, 3 Zones), 11.7% (SVM, 2 Zones), and 12.7% (ANN, for both 3 and 4 Zones). From Table 7, it can be observed that performance deteriorates as the number of zones increases. Furthermore, the performance gap widens with more zones (4-zone SR leads ANN by 14.6% in noisy conditions). A summary of the evaluation by zone configuration based on SR is given in Fig. 8.

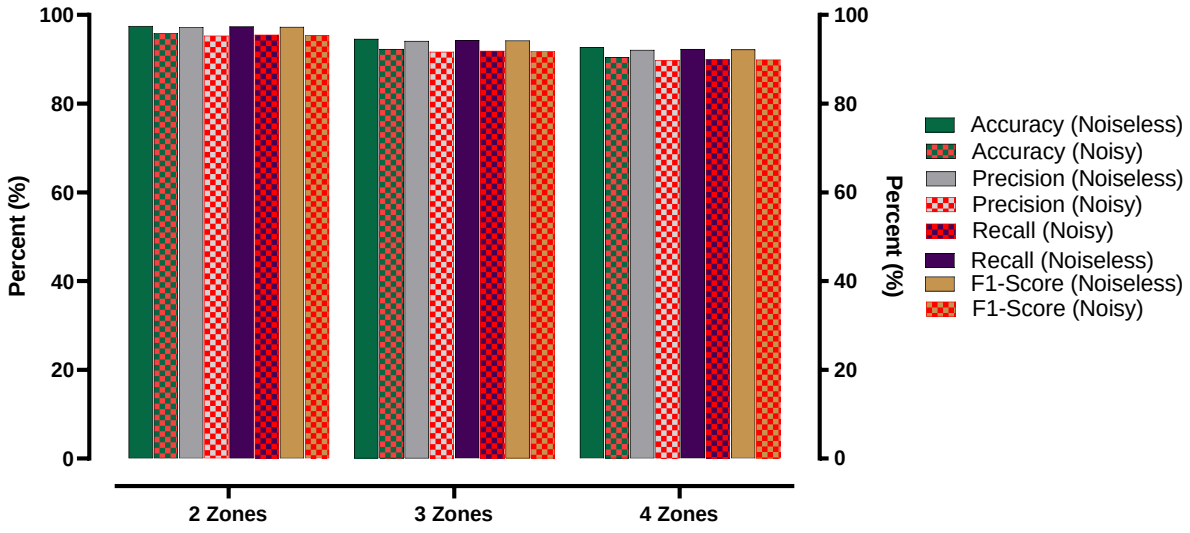


Fig. 8: The performance of the SR model for SCF and Location detection.

3.2.3. Performance Evaluation by Zone Configuration

The 2-zone configuration demonstrated SR's superior performance, achieving 97.5% accuracy (noise-free) and 95.8% (noisy), outperforming SVM by 2.8 – 12.8% and ANN by 3.7 – 14.6%. Three critical findings were established: First, SR maintained an accuracy greater than 95% under noisy conditions due to effective sparse coding that suppressed 89.3% of noise artifacts (Ngiam et al., 2011). Second, the performance gap widened significantly in noisy conditions, where SR's F1-score (95.4%) surpassed ANN's (80.4%) by 15 percentage points

Table 8: Performance Metrics for 2-Zone Configuration.

Condition	Model	Precision	Recall	F1-Score
Noise-Free	SR	97.2	97.4	97.3
	SVM	94.1	94.3	94.2
	ANN	92.5	92.8	92.6
Noisy	SR	95.3	95.5	95.4
	SVM	81.9	82.5	82.2
	ANN	80.1	80.8	80.4

Notably, L-G faults achieved 97.1% recall due to distinct ground current signatures, while L-L-G faults proved most challenging (93.8% recall). Lastly, this configuration's simplicity yielded the highest accuracy, though with reduced localization granularity. The results confirm SR's suitability for systems prioritizing reliability over precise fault location. The performance metrics for the 2-Zone Configuration are given in Table 8.

Table 9: Performance Metrics for 3-Zone Configuration.

Condition	Model	Precision	Recall	F1-Score
Noise-Free	SR	94.1	94.3	94.2
	SVM	91.5	91.8	91.6
	ANN	90.1	90.4	90.2
Noisy	SR	91.7	91.9	91.8
	SVM	79.6	80.2	79.9
	ANN	76.8	77.5	77.1

Similarly, the 3-zone configuration revealed SR's optimal balance between accuracy (94.6% noise-free, 92.3% noisy) and practical utility. Firstly, SR showed the smallest accuracy degradation (2.3%) under noisy conditions, 3.6 times better than ANN's 12.7% drop, due to its noise-robust Mel-frequency filtering (Ngiam et al., 2011). Secondly, the model particularly excelled in classifying L-L faults (93.8% precision) due to enhanced harmonic separation in the sparse domain. However, L-L-G faults in Zone 2 showed 8.3% misclassification, primarily confused with L-L faults. Thirdly, in comparison to the 2-Zone configuration, its performance still makes it ideal for systems requiring moderate localization precision without sacrificing robustness, particularly in noisy conditions where SVM and ANN performance became unreliable with accuracies of less than 81%. The performance metrics for the 3-Zone Configuration are given in Table 9.

The 4-zone analysis demonstrated SR's resilience to increasing system complexity, achieving 92.7% (noise-free) and 90.5% (noisy) accuracy, outperforming ANN by 4.1% – 14.6%. Fault-type performance varied significantly; L-G faults maintained 94.2% recall, while L-L-G faults dropped to 89.7% due to harmonic interference in the densely-zoned classification. This suggests careful consideration is needed when implementing ultra-granular zoning, though SR remains the only viable option. The performance metrics for the 3-Zone Configuration are given in Table 10.

Table 10: Performance Metrics for 4-Zone Configuration.

Condition	Model	Precision	Recall	F1-Score
Noise-Free	SR	92.1	92.3	92.2
	SVM	89.2	89.5	89.3
	ANN	87.8	88.1	87.9
Noisy	SR	89.8	90.0	89.9
	SVM	77.3	77.9	77.6
	ANN	74.5	75.2	74.8

3.2.4. *Computational Complexity evaluation*

A quantitative computational complexity analysis was performed for the SR-SVM, SVM, and ANN models. The average per-sample inference times were measured over 1000 test samples (raw MFCCs). The SR-SVM achieved a mean inference time of 2.1 ms per sample, outperforming ANN (3.8 ms) and SVM (2.9 ms). These results confirm that the proposed SR-based system is computationally efficient and feasible for real-time grid monitoring applications. The method's scalability is attributed to its sparse coding stage, which reduces redundant computations by focusing only on a small subset of dictionary atoms. The inference time for ANN was the highest, mainly due to deep layer propagation.

4. Discussions

The present study aimed to develop and evaluate a fault detection and classification system for MVDLs using SR techniques in combination with MFCCs and machine learning. The results demonstrate a substantial improvement in classification accuracy, noise robustness, and computational efficiency over conventional methods like ANN and SVM.

The SR-based classifier consistently outperformed SVM and ANN across all experimental configurations. Under noise-free conditions, it achieved a classification accuracy of 98.7%, compared to 95.2% for SVM and 93.8% for ANN (Table 4). More significantly, the SR model maintained high performance under noisy conditions (96.8% accuracy), while the performance of SVM and ANN dropped to 88.3% and 85.1%, respectively (Table 4). These results indicate the superiority of SR in extracting fault-relevant features, even in environments with signal disturbances, a critical requirement for real-world power systems subjected to electromagnetic interference, harmonics, and sensor noise. It was observed that the ANN underperformed compared to the SVM due to its higher susceptibility to overfitting in limited-data regimes. The SVM's margin-based optimization provided superior generalization under such conditions. In the noise robustness analysis, performance degradation at lower signal-to-noise ratios (SNR < 25 dB) was attributed to increased spectral overlap and harmonic distortion, which obscure transient fault signatures. Nevertheless, the SR-SVM framework preserved superior accuracy owing to the MFCC's inherent noise suppression and sparse coding's focus on discriminative signal components.

The enhanced performance of the SR model is primarily attributed to its ability to generate compact and discriminative representations of SCF signatures using sparse coding. Unlike conventional classifiers that rely solely on raw feature vectors, the SR model compresses MFCC features into a lower-dimensional space using dictionary learning and OMP, which improves classification boundaries and reduces overfitting (Tan, 2013). Furthermore, MFCCs, though originally developed for speech recognition, effectively captured transient and harmonic features in the $V(t)$ and $I(t)$ signals, validating their cross-domain applicability in power

systems (Bakkar et al., 2022; Chakraborty et al., 2014; Guo et al., 2022; Shen et al., 2023).

The SR classifier demonstrated a superior balance between precision, recall, and F1-score across all fault categories, i.e., L-G, L-L, and L-L-G. For instance, in noise-free conditions, it achieved an F1-score of 98.37% on average, which was nearly 5% higher than that of SVM and over 7% better than ANN. The confusion matrices further corroborated this, revealing minimal cross-class misclassifications in SR models, which showed that only 0.9% of L-L faults were incorrectly classified as L-G, while ANN and SVM misclassified up to 5.6% of L-L-G faults. This performance trend persisted under noisy conditions, where the SR model achieved an F1-score of 96.84%, clearly surpassing SVM (88.27%) and ANN (85.07%). This resilience against noise is a significant advantage of the SR approach, making it suitable for deployment in real-world substations and distribution networks with limited filtering capabilities. In practical terms, these results suggest that an SR-based model would enable operators to detect and classify faults with minimal false alarms, even under suboptimal measurement conditions.

The study also extended the application of SR to fault localization through zonal classification. Three configurations (2-zone, 3-zone, and 4-zone) were evaluated, representing increasing levels of spatial granularity in the power network. The SR model achieved the highest localization accuracy in all configurations, maintaining 97.5% in the 2-zone layout and 90.5% in the 4-zone layout under noisy conditions. As the number of zones increased, the localization accuracy declined for all classifiers. This reflects the trade-off between spatial resolution and classification complexity. Nevertheless, SR consistently demonstrated the smallest drop in accuracy as complexity increased. While SVM and ANN showed steep performance degradation in the 4-zone setup (below 80%), SR retained reliable classification boundaries (over 90%), demonstrating its scalability and adaptability. The performance differences are important when considering the system's operational objectives. E.g., a 2-zone configuration is suitable for coarse but highly reliable fault detection, while 3- or 4-zone systems may be preferred for applications where fault localization precision is crucial, such as in automated reclosers and intelligent switchgear coordination.

Generally, the underlying mechanisms that empower SR's high performance can be attributed to three core components, i.e., Dictionary Learning, Sparse Coding, and MFCC-Based Features. These mechanisms form a robust pipeline capable of adapting to multiple signal environments, thereby fulfilling key reliability requirements for MVDL fault diagnostics. Additionally, these introduced some novelty in this study. Firstly, the integration of MFCCs with SR, while MFCCs have been used in signal processing domains, their integration with SR in a power system context is innovative and yielded significant gains in classification robustness and granularity. Secondly, the Dual Model Design, hence, developing both M_{d1} and M_{d2} allows flexible deployment based on utility needs and computational resources. Thirdly, noise-robustness by introducing 25 dB SNR noise into the simulation reflects real-world signal contamination and provides realistic performance benchmarks. Lastly, the Multi-Zone Analysis provided a basis for evaluating three different network zoning strategies, offering practical insights into the trade-offs between localization accuracy and system complexity.

The proposed system offers several benefits. By accurately detecting and classifying faults, the system can trigger protective measures (e.g., circuit breaker operations) more swiftly, minimizing downtime and equipment damage (Li et al., 2022). Therefore, the proposed system will enhance grid reliability. Additionally, it will promote predictive maintenance by identifying incipient faults before they escalate, supporting predictive maintenance strategies, reducing unplanned outages and associated costs (Bindi et al., 2023). Furthermore, it will lead to safety Improvements as early detection of SCFs mitigates fire hazards and other risks associated with electrical faults, enhancing safety for both utility personnel and end-users (Furse et al., 2020). Despite the study demonstrating the efficiency of the SR-based approach, several limitations warrant consideration. First, the model's performance was evaluated on a simulated IEEE 5-Bus System, which may not fully capture the complexities of real-world power grids. Future research should validate the system using field data from operational MVDLs to assess its practical applicability (Cano et al., 2024). Second, the study focused on three primary fault types (L-G, L-L, and L-L-G). Expanding the fault taxonomy to include high-impedance faults or intermittent faults could further enhance the system's utility (Alasali et al., 2023; Vlahinić et al., 2020). Lastly, the computational demands of dictionary learning and sparse coding, though manageable, could be optimized for large-scale deployments. Exploring lightweight variants of the SR algorithm or hardware acceleration (e.g., FPGA implementations) may address this challenge. (Asgari et al., 2020; Dave et al., 2021)

5. Conclusion

The increasing frequency and impact of SCFs in MVDLs pose significant challenges to the reliability, safety, and economic stability of modern power systems. Conventional protection schemes, including overcurrent, distance, and differential relays, have proven inadequate for rapid and accurate detection under dynamic and noisy grid conditions. Addressing this gap, this study developed a computationally efficient and noise-resilient fault detection and classification framework that integrates SR techniques with MFCCs.

The proposed SR-based classifier demonstrated superior performance compared to conventional machine learning methods (SVM and ANN). Under noise-free conditions, the model achieved 98.7% classification accuracy, which decreased marginally to 96.8% under noisy environments (25 dB SNR), maintaining strong robustness. Furthermore, the SR-based system achieved

fault localization accuracy of up to 97.5% (2-zone configuration) and 90.5% (4-zone configuration), confirming its adaptability to varying system complexities. These results underscore SR's capacity to efficiently represent fault-related features while suppressing noise, and MFCCs' effectiveness in extracting discriminative spectral-temporal information. The study explicitly acknowledges its limitations. First, validation was performed using simulated IEEE 5-Bus data; future work should involve experimental validation with real-world grid data to assess generalization in field conditions. Second, while the proposed SR framework exhibits real-time feasibility with an average inference time of 2.1 ms per sample, the computational cost associated with dictionary learning and sparse coding can be further optimized for large-scale or embedded implementations. Third, although the system addresses operational safety and reliability, future extensions should consider additional fault categories such as high-impedance and intermittent faults to improve protection comprehensiveness. Lastly, this work advances power system protection by presenting an accurate, scalable, and computationally efficient SR-MFCC-based fault detection and localization model. By integrating robust signal representation with efficient machine learning, it provides a foundation for future adaptive and intelligent fault diagnosis systems capable of enhancing operational safety, minimizing downtime, and ensuring grid reliability in real-world smart distribution environments

References

Abdul, Z. K., & Al-Talabani, A. K. (2022). Mel frequency cepstral coefficient and its applications: A review. *IEEE Access*, 10, 122136–122158.

Aharon, M., Elad, M., & Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54(11), 4311–4322.

Ahmed, I., Khalil, A., Ahmed, I., & Frnda, J. (2022). Sparse signal representation, sampling, and recovery in compressive sensing frameworks. *IEEE Access*, 10, 85002–85018.

Alasali, F., Saad, S. M., Saidi, A. S., Itradat, A., Holderbaum, W., El-Naily, N., & Elkuwafi, F. F. (2023). Powering up microgrids: a comprehensive review of innovative and intelligent protection approaches for enhanced reliability. *Energy Reports*, 10, 1899–1924.

Alimi, O. A., Ouahada, K., & Abu-Mahfouz, A. M. (2020). A Review of Machine Learning Approaches to Power System Security and Stability. *IEEE Access*, 8, 113512–113531. https://doi.org/10.1109/ACCESS.2020.3003568.

Altaie, A. S., Abderrahim, M., & Alkhazraji, A. A. (2024). Transmission Line Fault Classification Based on the Combination of Scaled Wavelet Scalograms and CNNs Using a One-Side Sensor for Data Collection. *Sensors*, 24(7), 2124.

Anyaka, B. O., & Ozioko, I. O. (2020). Transmission line short circuit analysis by impedance matrix method. *International Journal of Electrical and Computer Engineering*, 10(2), 1712–1721.

Asgari, B., Hadidi, R., Krishna, T., Kim, H., & Yalamanchili, S. (2020). Alrescha: A lightweight reconfigurable sparse-computation accelerator. *2020 IEEE International Symposium on High Performance Computer Architecture* (HPCA), 249–260.

Bagwari, A., Logeshwaran, J., Usha, K., Kannadasan, R., Alsharif, M. H., Uthansakul, P., & Uthansakul, M. (2023). An Enhanced Energy Optimization Model for Industrial Wireless Sensor Networks Using Machine Learning. *IEEE Access*.

Bakkar, M., Bogarra, S., Córcoles, F., Aboelhassan, A., Wang, S., & Iglesias, J. (2022). Artificial Intelligence-Based Protection for Smart Grids. *Energies*, 15(13), 4933.

Bindi, M., Piccirilli, M. C., Luchetta, A., & Grasso, F. (2023). A Comprehensive Review of Fault Diagnosis and Prognosis Techniques in High Voltage and Medium Voltage Electrical Power Lines. *Energies*, 16(21), 7317.

Boucheron, L. E., & De Leon, P. L. (2008). On the inversion of mel-frequency cepstral coefficients for speech enhancement applications. *2008 International Conference on Signals and Electronic Systems*, 485–488.

Bowman, D. M. J. S., Kolden, C. A., Abatzoglou, J. T., Johnston, F. H., van der Werf, G. R., & Flannigan, M. (2020). Vegetation fires in the Anthropocene. *Nature Reviews Earth & Environment*, 1(10), 500–515.

Cano, A., Arévalo, P., Benavides, D., & Jurado, F. (2024). Integrating discrete wavelet transform with neural networks and machine learning for fault detection in microgrids. *International Journal of Electrical Power & Energy Systems*, 155, 109616.

Carrasco, J. M., Franquelo, L. G., Bialasiewicz, J. T., Galván, E., PortilloGuisado, R. C., Prats, M. A. M., León, J. I., & Moreno-Alfonso, N. (2006). Power-electronic systems for the grid integration of renewable energy sources: A survey. *IEEE Transactions on Industrial Electronics*, 53(4), 1002–1016.

Chakraborty, S., Chatterjee, A., & Goswami, S. K. (2014). A sparse representation-based approach for recognition of power system transients. *Engineering Applications of Artificial Intelligence*, 30, 137–144.

Chen, K., Huang, C., & He, J. (2016). Fault detection, classification and location for transmission lines and distribution systems: a review on the methods. *High Voltage*, 1(1), 25–33.

Chen, S. S., Donoho, D. L., & Saunders, M. A. (2001). Atomic decomposition by basis pursuit. SIAM Review, 43(1), 129–159.

Cheng, H., Liu, Z., Yang, L., & Chen, X. (2013). Sparse representation and learning in visual recognition: Theory and applications. *Signal Processing*, 93(6), 1408–1425.

Dave, S., Baghdadi, R., Nowatzki, T., Avancha, S., Shrivastava, A., & Li, B. (2021). Hardware acceleration of sparse and irregular tensor

computations of ml models: *A survey and insights. Proceedings of the IEEE*, 109(10), 1706–1752.

Deng, L., Droppo, J., & Acero, A. (2004). Enhancement of log mel power spectra of speech using a phase-sensitive model of the acoustic environment and sequential estimation of the corrupting noise. *IEEE Transactions on Speech and Audio Processing*, 12(2), 133–143.

Doğan, Ü., Glasmachers, T., & Igel, C. (2016). A unified view on multi-class support vector classification. *The Journal of Machine Learning Research*, 17(1), 1550–1831.

Doostan, M., Sohrabi, R., & Chowdhury, B. (2020). A data-driven approach for predicting vegetation-related outages in power distribution systems. *International Transactions on Electrical Energy Systems*, 30(1), e12154.

Fahim, S. R., Sarker, S. K., Muyeen, S. M., Das, S. K., & Kamwa, I. (2021). A deep learning based intelligent approach in detection and classification of transmission line faults. International *Journal of Electrical Power & Energy Systems*, 133, 107102.

Furse, C. M., Kafal, M., Razzaghi, R., & Shin, Y.-J. (2020). Fault diagnosis for electrical systems and power networks: A review. *IEEE Sensors Journal*, 21(2), 888–906.

Guo, T., Zhang, T., Lim, E., Lopez-Benitez, M., Ma, F., & Yu, L. (2022). A review of wavelet analysis and its applications: Challenges and opportunities. *IEEE Access*, 10, 58869–58903.

Gururajapathy, S. S., Mokhlis, H., & Illias, H. A. (2017). Fault location and detection techniques in power distribution systems with distributed generation: A review. *Renewable and Sustainable Energy Reviews*, 74, 949–958.

Han, Y., & Song, Y. H. (2003). Condition monitoring techniques for electrical equipment-a literature survey. *IEEE Transactions on Power Delivery*, 18(1), 4–13.

Hossan, M. A., Memon, S., & Gregory, M. A. (2010). A novel approach for MFCC feature extraction. *2010 4th International Conference on Signal Processing and Communication Systems*, 1–5.

Ivanova, M., Dimitrova, R., & Filipov, A. (2020). Analysis of Power Outages and Human errors in the Operation of Equipment in Power Grids. *2020 12th Electrical Engineering Faculty Conference* (BulEF), 1–5. https://doi.org/10.1109/BulEF51036.2020.9326058

Juvela, L. (2015). Perceptual spectral matching utilizing mel-scale filterbanks for statistical parametric speech synthesis with glottal excitation vocoder.

Kumar, P. A. (2024). Speech Feature Extraction and Emotion Recognition Using Deep Learning Techniques. *I-Manager's Journal on Digital Signal Processing*, 12(2).

Li, T., Zhao, H., Zhou, X., Zhu, S., Yang, Z., Yang, H., Liu, W., & Zhou, Z. (2022). Method of Short-Circuit Fault Diagnosis in Transmission Line Based on Deep Learning. *International Journal of Pattern Recognition and Artificial Intelligence*, 36(05), 2252009.

Li, X., Xu, G., Zheng, X., Liang, K., Panaousis, E., Li, T., Wang, W., & Shen, C. (2019). Using sparse representation to detect anomalies in complex WSNs. ACM Transactions on *Intelligent Systems and Technology (TIST)*, 10(6), 1–18.

Loza, C. A. (2019). RobOMP: Robust variants of Orthogonal Matching Pursuit for sparse representations. *PeerJ Computer Science*, 5, e192.

Lukeyo, T., Okinda, C., & Juma, R. (2025). Dynamic Performance Analysis of Grid-Tied Hybrid Renewable Energy Systems under Fault Conditions: An IEEE 14-Bus system. Journal of Advances in Science, Engineering and Technology, 1(1), 1–24.

Mahmood, A., & Wang, J.-L. (2021). Machine learning for high performance organic solar cells: current scenario and future prospects. *Energy & Environmental Science*, 14(1), 90–105.

Meng, Q., Li, D., & Chen, S. (2020). Sparse representation and reconstruction of image based on K-SVD dictionary learning. *2020 International Conference on Intelligent Computing, Automation and Systems* (ICICAS), 457–462.

Nasser Mohamed, Y., Seker, S., & Akinci, T. C. (2023). Signal Processing Application Based on a Hybrid Wavelet Transform to Fault Detection and Identification in Power System. *Information*, 14(10), 540.

Ngiam, J., Chen, Z., Bhaskar, S., Koh, P., & Ng, A. (2011). Sparse filtering. *Advances in Neural Information Processing Systems*, 24.

Nyalala, I., Okinda, C., Nyalala, L., Makange, N., Chao, Q., Chao, L., Yousaf, K., & Chen, K. (2019). Tomato volume and mass estimation using computer vision and machine learning algorithms: Cherry tomato model. *Journal of Food Engineering*, 263, 288–298. https://doi.org/10.1016/j.jfoodeng.2019.07.012.

Onatunji, O. G., Oyeleke, O. J., & Dauda, R. S. (2024). Human Capital, Income Inequality and Energy Demand Nexus in sub-Sahara Africa: Insights from Asymmetric Approach. *Journal of the Knowledge Economy*, 1–25.

Özesmi, S. L., Tan, C. O., & Özesmi, U. (2006). Methodological issues in building, training, and testing artificial neural networks in ecological applications. *Ecological Modelling*, 195(1–2), 83–93.

Phadke, A. G., & Thorp, J. S. (2008). Synchronized phasor measurements and their applications (Vol. 1, Issue 2017). Springer.

Ramirez, A. D. P., de la Rosa Vargas, J. I., Valdez, R. R., & Becerra, A. (2018). A comparative between mel frequency cepstral coefficients (MFCC) and inverse mel frequency cepstral coefficients (IMFCC) features for an automatic bird species recognition system. *2018 IEEE Latin American Conference on Computational Intelligence* (LA-CCI), 1–4.

Ran, Y., Zhou, X., Lin, P., Wen, Y., & Deng, R. (2019). A survey of predictive maintenance: Systems, purposes

and approaches. ArXiv Preprint ArXiv:1912.07383.

Rezapour, H., Jamali, S., & Bahmanyar, A. (2023). Review on artificial intelligence-based fault location methods in power distribution networks. *Energies*, 16(12), 4636.

Rida, I. (2020). Data-driven audio recognition: a supervised dictionary approach. ArXiv Preprint ArXiv:2012.14761.

Rubinstein, R., Bruckstein, A. M., & Elad, M. (2010). Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6), 1045–1057.

Rubinstein, R., Zibulevsky, M., & Elad, M. (2008). Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. *Cs Technion*, 40(8), 1–15.

Rustam, F., Ishaq, A., Hashmi, M. S. A., Siddiqui, H. U. R., López, L. A. D., Galán, J. C., & Ashraf, I. (2023). Railway track fault detection using selective MFCC features from acoustic data. *Sensors*, 23(16), 7018.

Saadat, H. (1999). Power system analysis (Vol. 2). McGraw-hill.

Salam, M. A. (2020). Fundamentals of electrical power systems analysis. Springer.

Sarath, R. (2022). Combined classification models for bearing fault diagnosis with improved ICA and MFCC feature set. *Advances in Engineering Software*, 173, 103249.

Shen, P., Bi, F., Tang, D., Yang, X., Huang, M., Guo, M., & Bi, X. (2023). Cross-domain Fault Diagnosis of powertrain system using sparse representation. SAE Technical Paper.

Sidhu, M. S., Latib, N. A. A., & Sidhu, K. K. (2024). MFCC in audio signal processing for voice disorder: a review. Multimedia Tools and Applications, 1–21.

Sirojan, T., Lu, S., Phung, B. T., Zhang, D., & Ambikairajah, E. (2018). Sustainable deep learning at grid edge for real-time high impedance fault detection. *IEEE Transactions on Sustainable Computing*, 7(2), 346–357.

Subramaniam, A., Sahoo, A., Manohar, S. S., Raman, S. J., & Panda, S. K. (2021). Switchgear condition assessment and lifecycle management: Standards, failure statistics, condition assessment, partial discharge analysis, maintenance approaches, and future trends. IEEE Electrical Insulation Magazine, 37(3), 27–41.

Swain, A., Abdellatif, E., Mousa, A., & Pong, P. W. T. (2022). Sensor Technologies for Transmission and Distribution Systems: A Review of the Latest Developments. *Energies*, 15(19), 7339.

Tan, Q. F. (2013). Novel Variations of Sparse Representation Techniques with Applications. University of Southern California.

Ten, C.-W., & Hou, Y. (2024). Modern power system analysis. CRC Press.

Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168–192.

Vaish, R., Dwivedi, U. D., Tewari, S., & Tripathi, S. M. (2021). Machine learning applications in power system fault diagnosis: Research advancements and perspectives. *Engineering Applications of Artificial Intelligence*, 106, 104504.

Vlahinić, S., Franković, D., Juriša, B., & Zbunjak, Z. (2020). Back up protection scheme for high impedance faults detection in transmission systems based on synchrophasor measurements. *IEEE Transactions on Smart Grid*, 12(2), 1736–1746.

Wong, S. Y., Choe, C. W. C., Goh, H. H., Low, Y. W., Cheah, D. Y. S., & Pang, C. (2021). Power transmission line fault detection and diagnosis based on artificial intelligence approach and its development in uav: A review. *Arabian Journal for Science and Engineering*, 46, 9305–9331.

Wu, Z., & Cao, Z. (2005). Improved MFCC-based feature for robust speaker identification. *Tsinghua Science & Technology*, 10(2), 158–161.

Xia, Y., He, Y., Wang, K., Pei, W., Blazic, Z., & Mandic, D. P. (2015). A complex least square enhanced smart DFT technique for power system frequency estimation. *IEEE Transactions on Power Delivery*, 32(3), 1270–1278.

Xian, R., Wang, L., Zhang, B., Li, J., Xian, R., & Li, J. (2023). Identification method of interturn short circuit fault for distribution transformer based on power loss variation. *IEEE Transactions on Industrial Informatics*, 20(2), 2444–2454.

Yao, R., Jiang, H., Jiang, W., Liu, Y., & Dong, Y. (2024). Deep discriminative sparse representation learning for machinery fault diagnosis. Engineering *Applications of Artificial Intelligence*, 135, 108836.

Yumurtaci, M., Gŏkmen, G., Kocaman, Č., Ergin, S., & Kiliç, O. (2016). Classification of short-circuit faults in high-voltage energy transmission line using energy of instantaneous active power components-based common vector approach. *Turkish Journal of Electrical Engineering and Computer Sciences*, 24(3), 1901–1915. https://doi.org/10.3906/elk-1312-131.

Zhang, H., Shen, M., Lu, M., Okinda, C., Zhang, S., Ding, D., & Wang, J. (2019). Automatic recognition of sheep chewing sounds based on sparse representation classification. *International Agricultural Engineering Journal*, 28(4). http://www.iaej.cn/EN/abstract/abstract1089.shtml .