Fault Identification with High SPV Penetrated Distribution Grid Using Modified Ceemdan Method

Ritu Singh[1]*, Soorya Prakash Shukla[2]

[1]Research Scholar, Department of Electrical Engineering, BIT Durg, CG, India
[2]Professor, Department of Electrical Engineering, BIT Durg, CG, India
*Corresponding author

Received: 18.04.2024 Revised : 13.05.2024 Accepted: 28.05.2024

ABSTRACT
Microgrids are not a prospective future solution but an existing paradigm, ensuring the energy sufficiency through decentralized and resilient power generating systems. Considering the techno-economical benefits, SPV is widely used in the present day as decentralised power sources. Along with the energy security, SPV contributes the system complexity in the active distribution system, which adds further complexities in the protection systems to identify the faults. To address this complex issue, this paper proposes a new technique for fault detection in a solar-integrated distribution grid system that combines Modified Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (MCEEMDAN) for feature extraction and sophisticated machine learning classifiers for classification. The Modified MCEEMDAN approach successfully decomposes fault signals into intrinsic mode functions (IMFs), allowing for the extraction of entropy-based features that represent the nonlinear properties of these signals. Several entropy measures, including Approximate Entropy and Sample Entropy, were investigated to improve fault classification performance. The findings show that the XGBoost classifier beat other models, obtaining an accuracy of 97.5% using approximation entropy features, demonstrating its ability to detect flaws. This study emphasizes the importance of feature selection in optimizing computing efficiency and model performance. The results provide important insights into the integration of signal processing methods and machine learning for improved problem identification in electrical systems, opening the way for future advances in the reliability and efficiency of power distribution networks. The entire analysis is done on a IEEE-33 bus distribution system modelled in Matlab Simulink.

Keywords: Deep learning, Fault identification, Modified Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (MCEEMDAN), SPV, Active Distribution Network.

1. INTRODUCTION

The distribution network structure is becoming increasingly complex, and it is inevitably impacted by a variety of defects in real-world operating conditions [1]. The most critical mission in distribution network troubleshooting technology is the identification and detection of faults whenever a failure occurs. Fault classification algorithms have garnered considerable attention in recent years. Nevertheless, these classification algorithms are primarily employed on transmission lines. The malfunction classification methods for the transmission system cannot be directly applied to the distribution system due to the distinct operation schemes of the electric systems. The neutral grounding mode, fault resistance, and other factors can all affect the characteristics of the electric signal. These factors contribute to the complexity of defect identification in the distribution network. The distribution network's fault identification is of immense importance in terms of the reliability of the power supply, the safety and stable operation of the system, and the time required to recover from a failure.

Power outages in transmission lines are mostly caused by unanticipated and irregular faults[2]. Power system problems are inevitable and must not be disregarded. Fault detection and classification are crucial for ensuring the stability of both traditional and intelligent power grids[3]. Transmission line faults and equipment breakdowns may result in substantial interruptions to the power system, resulting in power outages and equipment damage. Hence, it is crucial to ensure precise and prompt identification and categorization of faults in order to maintain the stability and safety of the smart grid[4]. A smart grid is a sophisticated and ever-changing system that necessitates ongoing surveillance and upkeep to guarantee dependability and effectiveness. Fault detection and categorization are essential activities in the operation and administration of smart grids. According to [5], most of the problems in the transmission part of the power system occur in transmission lines. Short-circuit faults are common and considered the

most severe form, presenting significant dangers to transmission lines[6]. These dangers include reducing the operational lifetime of components, increasing power losses, causing cable heat, and damaging insulators.
Over the last two decades, there has been a swift advancement in numerous domains related to identifying, classifying, and detecting power system malfunctions. Growing numbers of researchers are now able to conduct studies with a high breadth and depth because of advancements in signal processing techniques, artificial intelligence and machine learning, global positioning system (GPS), and communications, which have allowed the boundaries of conventional fault protection techniques to be stretched. Precisely identifying and categorising transmission line defects can lower the cost of replacing power lines and improve the likelihood of power grid safety. Customers experience power outages due to transmission line failures [7].

1.1 Overview of Fault Identification and classification

Transmission line faults can result in a variety of disruptions, including overheating, mechanical stress, and unbalanced power flow. Furthermore, reliable fault detection and classification (FDC) is crucial to maintaining grid system stability. While the process of recovering from failure phases is contingent upon human intervention and the detection and classification method employed to pinpoint the specific type of failure and its location within the network. This is important because a quick and accurate FDC guarantees prompt repair, increases the likelihood of separating problematic phases from the transmission system, and improves the transient stability and power quality of the interconnected power network. Many algorithms are used to classify faults in transmission lines. Two popular approaches are artificial intelligence (AI) and machine learning (ML), which are chosen for their ability to learn quickly, produce accurate results, and identify patterns in input training data [8]. Historical fault classification methods are classified as well-known and contemporary methods. The popular methods include fuzzy logic-based approaches for fault classification, commonly referred to as hybrid methods, and Wavelet Transform (WT) based analysis combined with Artificial Neural Networks.
In the last several decades, solar energy has become a well-liked solution for energy shortages and a competitive substitute for fossil fuels. This ecologically benign and renewable energy source offers a limitless and sustainable supply of power [9]. Consequently, solar energy has the capacity to supply all of the world's energy needs. It is mainly dependent on the weather, though, and any changes in that regard could have a big effect on its output power[10].
Artificial neural networks (ANNs) have gained attention for their ability to learn complex functions through nonlinear transformations. Deep learning methods have made them effective in tasks like voice and picture recognition, and fault detection. Convolutional neural networks (CNNs) are a supervised learning approach that can be tailored to address complex issues in exploratory geophysics due to their high architectural flexibility. CNNs are particularly useful for identifying seismic faces of interest, which are unique edges in seismic data. CNNs have shown remarkable effectiveness in detecting faults, but most studies have used hierarchical or shallow neural networks. Further research is needed to fully explore the potential of deep neural networks for defect diagnostics.
The feature extraction of deep learning is characterized by a remarkable performance in the areas of image classification, speech recognition, and machine translation, which is achieved through the use of multiple levels of abstraction[11]. Deep Neural Networks (DNNs) can automatically extract spatial and temporal features from input data without any traditional signal processing stages by utilizing special layers and supervised training of multiple samples. This allows them to complete tasks such as classification and regression. There is no doubt that the efficacy of features extracted through DNNs is significantly superior to that achieved through artificial feature engineering, as evidenced by the performance of deep learning in speech recognition and image classification.

1.2 Relevance of Deep Learning in Power Systems

Faults in power transmission lines may arise due to a variety of factors, including short circuits, tree or animal contact, lightning strikes, earthquakes, conductor clashing, and equipment corrosion. Some are under human influence, while others are naturally occurring. When protective relays identify a defect, they must clear it promptly [12]. While defects may develop for a variety of causes, locating and analysing them remains a key problem. Reducing post-fault analysis time allows for speedier system maintenance and restoration, perhaps leading to lower failure costs. To improve power system dependability, it's crucial to quickly and efficiently classify problem types and locations. advocated fault-location observability and a novel approach for transmission networks using synchronised phasor measuring units (PMUs). developed deterministic and stochastic methods for locating faults in power systems using a low number of PMUs [13].

Fault type and location categorization involves three steps: (1) importing transient fault data, (2) pre-processing using suitable methods, and (3) analysing the data. To convert three-phase voltage and current fault signals, pre-data processing methods such as Transform, wavelet transform, and Fast Fourier Transform (FFT) may be utilised. Fault data may be analysed using several methods, including machine learning and waveform-based correlation coefficients. Several academics have published articles on using machine learning and deep neural networks to detect and locate power system issues.Fault techniques are classified into model-based, knowledge-based, and data-driven approaches, which rely on numerical data analysis and interpretation rather than personal observation or experience [14].

Data-driven strategies base ideas and solutions on verifiable facts, rather than assumptions or personal experience. Various machine learning algorithms, including decision trees, support vector machines, and k-nearest neighbours (k-NN), have been suggested for fault classification[15]. The study found that processing high-dimensional data requires computational complexity and reduction strategies for reconstruction. However, reducing data dimensions might lead to information loss and undermine the accuracy of outcomes. In Ref [16], the authors suggest a way to extract PV cell attributes from thermal images and compare the results using the SVM algorithm. In [17] the author presents a strategy for distinguishing PQ disturbances from pure sinusoidal signals using time-domain descriptor fusion (FTDD). The recommended technique is evaluated using multiclass SVM and Naive Bayes (NB) classifiers. [18] developed the Modified Multi-Class Support Vector Machines (MMC-SVM) approach to categorise open-circuit faults in power distribution networks. Simulation findings indicate the usefulness and resilience of the proposed machine learning model [19].

In [20], a classification technique is suggested that utilizes convolutional neural networks (CNNs) with varying sample frequencies. The use of wavelet transform for extracting fault harmonics in the input of CNNs has been observed. However, the accuracy of the classification judgements and the results are affected by data generalization difficulties, as mentioned in[21]. Deep neural networks use convolutional neural networks (CNNs) as a powerful technique for image categorization. CNNs are also employed as fundamental components of ResNet and VGG16. Convolutional Neural Networks (CNNs) have the capability to categories extensive picture collections collected from ImageNet. Various convolutional layers, pooling layers, and fully connected layers are used to extract the fundamental characteristics of the data from the pictures and categories them via supervised learning.

1.3 Background

1.3.1 MCEEMDAN

Complete ensemble empirical mode decomposition with adaptive noise(CEEMDAN) is noise assisted EMD technique. It decomposes Non stationary signals (Time Series Data) into Intrinsic mode functions(IMF).CEEMDAN is an EMD algorithm that provides an exact reconstruction of the original signal and a better spectral separation of the IMFs. While the EMD and EEMD are unable to accurately identify the various oscillatory modes that are present in the signal. Consequently, the issue of mode blending arises. As a result, an IMF inaccurately displays various physical processes that are depicted in the mode. The CEEMDAN algorithm has several disadvantages, including the presence of residual noise in the decomposed modes and the decomposition of the provided signal into spurious modes [23]. The modified CEEMDAN (MCEEMDAN) scheme has been devised to address this issue.The MCEEMDAN accurately finds oscillatory modes in non-linear and non-stationary signals. The CEEMDAN and MCEEMDAN algorithms are identical, save for noise reduction capabilities. The MCEEMDAN approach effectively addresses spurious modes and residual noise issues. This approach extracts many intrinsic mode functions (IMFs) and selects the most suited one for further analysis to determine fault index.

1.3.1.a. Partial Mean of Multi-Scale Permutation Entropy

Bandt and Pompe introduced the concept of permutation entropy (PE) as a means to quantify the time series complexity. Probabilistic analysis (PE) may evaluate signals by comparing neighboring data, without taking into account the particular value of the signal [24][25]. The method has the benefits of straightforward procedures, consistent resilience, and considerable practicality, enabling the measurement of nonlinear signals. One may succinctly outline the procedural stages of PE as follows:

For a time -series $\{x(i), i = 1, 2, \ldots, N\}$ of length N, the phase space reconstruction can be described as

$$X(1) = \{x(1), x(1+\tau), \ldots., x(1+(m-1)\tau)\} \quad (1)$$

$$X(i) = \{x(i), x(i+\tau), \ldots., x(i+(m-1)\tau)\} \quad (2)$$

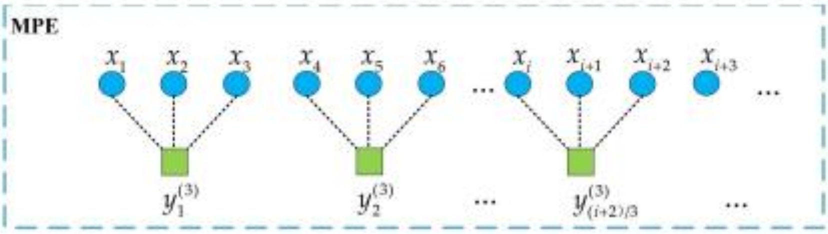

Figure 1. Process of coarse-grained time series with s = 3.

$X(N-(m-1)\tau) = \{x(N-(m-1)\tau),\ x(N-(m-2)\tau),....,x(N)\}$ (3)

where m and τ represent the embedding dimension and time delay, respectively. In the paper, the m is set to 5, and the τ is set to 1.

The m-dimension space vector in reconstructed X(i) can be rearranged in ascending order as

$x(i+(j_1-1)\tau) \le x(i+(j_2-1)\tau) \le ... \le x(i+(j_m-1)\tau)$ (4)

Where $j_1, j_2....j_m$ denote the index value in the sequence. If there existed $x(i+(j_{i1}-1)\tau) = c(i+(j_{i2}-1)\tau),$ the permutation could be arranged according to j value size determination. The sequence can be ordered as $x(i+(j_{i1}-1)\tau) \le x(i+(j_{i2}-1)\tau)\ j_{i1} < j_{i2}.x(i)$ when $j_{i1} < j_{i2}$. This can let vector X(i) get a symbol sequence S(g) as follows:

$S(g) = [j_1, j_2...., j_m]$ (4.1)

Where $g = 1\square\ k, k \le m!. S(g)\ m!$) is one of the m! symbol sequences (m different symbols have m! kinds of arranging form).

According to the Shannon entropy, the PE for the time series can be defined as

$H_p(m) = -\sum_{g=1}^{k} P_g \ 1n\ P_g$ (5)

Where $P_g(g = 1, 2, ..., k)$ can denote the probability of each symbol sequence presence. As $0 \le H_p \le 1n(m!),$ the PE can be normalised as

$H_p = \frac{H_p(m)}{1n(m!)}$ (6)

Where $0 \le H_p \le 1,$ The power spectral density (PE) may characterize the nuanced fluctuations in the signal. The use of multi-scale permutation entropy (MPE) may enhance the precision of signal analysis. It represents an enhancement compared to the PE. In the computation of Mean Percentage Error (MPE), the coarse-graining of time series is the most critical step.

$y^s(j) = \frac{1}{s}\sum_{i=(j-1)s+1}^{js} x(i), 1 \le j \le \frac{N}{S}$ (7)

where s is the scale factor, and ys(j) is the sequence with a coarse grain size. Figure 1 outlines the procedure for computing the coarse-grained time series when s is equal to 3. The Mean Precondition Error (MPE) may be defined by computing the Precondition Error (PE) of each coarse-grained sequence

$MPE(x, s, m, \tau) = PE(y^{(s)}, m, \tau)$ (8)

The partial mean of multi-scale permutation entropy can be calculated to enhance the MPE's integrated analysis capabilities. The mean value and variation tendency of a sequence can be described using the partial mean. Under various scales, the PE of a time series can be transformed into a sequence M_{PE} as

$M_{PE} = [M_{PE}(1), M_{PE}(2),...., M_{PE}(s)]$ (9)

The skewness S_{ke} can be expressed as

$S_{ke} = \frac{3(M_{PEm} - M_{PEn})}{M_{PEt}}$ (10)

Where, M_{PEm} is the mean value, M_{PEn} is the median level, and M_{PEt} is the standard deviation. The partial mean of multi-scale permutation entropy P_{MMPE} can be defined as

$$P_{MMPE} = (1+S_{ke} / 3\backslash)M_{PEm} \quad (11)$$

The original sequence is more regular as the value of P_{MMPE} decreases. The complexity degree of the time series can be determined by modifying the P_{MMPE}. Several experiments have demonstrated that the P_{MMPE} can effectively differentiate between normal and aberrant signals when it is set to 0.8~0.9. This paper establishes a threshold value of 0.85.

1.3.2 IEEE -33 Bus system

The IEEE 33 Bus system is a standardised test case used in power engineering research to assess the performance and dependability of distribution networks. It is made up of 33 buses (nodes) and 32 radial lines (branches) that represent a typical distribution network. Baran and Wu suggested a 33-bus distribution system in 1989. It is often used to evaluate many electrical engineering issues, including load flow analysis, fault analysis, and network optimisation. The system is distinguished by its single feeder and radial architecture, which means that all lines originate from a single substation and stretch outward, like a tree structure. The IEEE 33-Bus radial distribution system is used to test and evaluate different kinds of DG units. This system is made up of 33 buses and 32 lines, with a voltage of 12.66kV, a load capacity of 3.715MW, and 2.3MVar[26]. The distributed generating unit utilised represents 30% of the total load. The DG unit voltage is 12.66kV, and the system's lower and higher voltages are set at 0. 95p.u and 1.05p.u. This will allow us to see how the different DG units affect the electricity system's load ability margin. To provide a clear assessment of the various DG units' effects on the distribution system, the research will be conducted with a set optimum location and DG penetration level. The DG unit location was decided using an optimisation approach with a set penetration level (30% of the total load). This arrangement serves as a critical benchmark for academics and engineers developing and testing new algorithms, methods, and technologies to improve the efficiency, stability, and resilience of electrical distribution networks.

2. LITERATURE REVIEW

In [27]the authors discussed the development and deployment of Machine Learning (ML)-based algorithms for fault classification and detection in electrical distribution systems. The methodology uses higher computational accuracy than traditional algorithms. The parameters for fault detection include fundamental frequency, fault voltage, and current components. The Wavelet Decomposition technique is used to break down transient signals during faults. The performance of the algorithms is investigated using an IEEE 33 bus system and faults generated in Matlab/Simulink. K Nearest Neighbour (KNN), Decision Tree (DT), and Support Vector Machine (SVM) methodologies are used. The results show exceptional accuracy.

Study in [28] proposed a fault diagnostic model for distribution systems based on deep graph learning, considering the physical structure of the power network as a significant constraint during training. This model enhanced information perception and resistance to abnormal data input and unknown application conditions. A spatiotemporal convolutional block enhances waveform feature extraction, making it more effective in dealing with fault waveform changes and spatial effects. A multi-task learning framework is constructed for fault location and fault type analysis, improving performance and generalization ability. The model's effectiveness is verified using IEEE 33-bus and IEEE 37-bus test systems, and its anti-interference and generalization performance is evaluated under different fault conditions, topological changes, and interference factors. Experimental results showed that the proposed model outperforms other state-of-the-art methods.

The authors in [29]applied deep learning, specifically the Faster R-CNN model, to improve the efficiency and reliability of drone imagery-based inspections for component identification and defect detection of transmission lines. The research found that deep learning is highly effective in identifying defects in high-voltage transmission line components, with a recognition speed of 0.17 seconds per image and a recognition rate of 96.8% for pressure-equalizing rings.

The paper presented a data-driven fault location identification and types classification application using continuous wavelet transformation and convolutional neural networks optimized through Bayesian optimization in [30]. The application can identify short-circuit faults and classify them into eleven types. Its intrinsic models understand spatial characteristics and convert them into frequency domain temporal measurements, increasing network visibility in real-time. Simulations using synthetic data replicated fault

occurrence scenarios under noise conditions and load variability, achieving an accuracy of 91.4% for fault detection, 93.77% for correct branch identification, and 94.93% for fault type classification.
The study in [31] utilized transmission line voltage and current data to identify and classify power transmission issues, employing Artificial Neural Network (ANN) and Convolutional Neural Network (CNN) models to evaluate their accuracy. This study uses Kaggle voltage and current data to locate power transmission issues. Data is analysed and fed into two Deep Learning (DL) networks: an ANN and a CNN. These models identify defects using input data. The CNN exceeds the ANN in accuracy. This shows that CNNs handle voltage and current data better for power transmission line problem detection and classification.
The article examined an adaptive scheme that accurately detects volatile changes in micro grids under normal and fault conditions in [33]. The scheme uses a two-level SVM classifier model to identify mode of operation, variations in renewable energy generation, random addition of EV charging load, and fault occurrence. The model uses RMS values of three-phase voltage and current measurements as data inputs. The efficacy of the scheme is tested on an IEEE 9-bus micro grid test bed using simulation experiments and real-time experiments on hardware-in-loop simulators, OPAL-RT, and Raspberry Pi microcontrollers.

2.1 Deep Learning Approaches in Fault Identification

[34]focuses on to detect and classify defects in electrical distribution networks using deep learning techniques. It considers fault voltage, fundamental frequency, and current components for fault identification and categorization. An IEEE 33 bus system is used to model distribution networks, and fault conditions are created in simulation. Discrete Wavelet Transform (DWT) and Deep Learning (DL) approaches are used to detect and classify faults in distribution systems. The proposed Deep Neural Network (DNN) model has high accuracy in recognizing and classifying faults. Simulations are conducted using MATLAB software.
A reinforcement learning algorithm has been developed to address the challenge of managing voltage profile in distribution systems due to complex network configurations[35]. The algorithm involves placing distributed generation units at different locations in IEEE 33 bus radial distribution networks, resulting in a 69% reduction in active power losses. The method uses a convolutional neural network, a long short-term memory network, and attention mechanisms to detect and classify faults. The algorithm's performance is compared with other deep learning techniques and the time taken for fault detection is determined.
The study in [36]aims to enhance fault detection, diagnosis, identification, and location in large-scale multi-machine power systems by introducing novel Deep Learning models for fault region identification, fault type classification, and fault location prediction. Three new Deep Recurrent Neural Networks (DRNN) models with Long Short-Term Memory (LSTM) are used to analyse transient data from pre- and post-fault cycles. The proposed algorithms show superior performance in fault detection, classification, and location prediction, achieving high accuracy and robustness compared to existing techniques.
The electrical power system is complex and susceptible to faults, especially in the secondary distribution network. Various methods have been explored for fault detection and classification, including mathematical approaches, expert systems, and artificial neural networks integrated with SCADA and PMU systems. However, there is limited research on the application of deep learning approaches in fault detection and classification. This study compared several deep learning approaches, revealing that Recurrent Neural Network (RNN) is the most efficient in detecting and classifying faults in the electrical secondary distribution network, with accuracy increasing with complexity as depict in [37].

2.2 Modified CEEMDAN and its Applications in Fault Detection

The data mining method for fault diagnosis in distribution networks is limited due to the unbalanced fault sample problem. A fault identification method for distribution networks is proposed by combining modified complete ensemble empirical mode decomposition with adaptive noise (MCEEMDAN) and conditional generative adversarial network (CGAN)[38]. The MCEEMDAN decomposes the electric signal into intrinsic mode functions, transforming the raw time-domain signal into a two-dimensional gray-level image. The fault gray image is labelled and put into CGAN to generate new samples for data augmentation. The proposed method effectively learns distribution characteristics and improves fault recognition accuracy. It has good stability, fast convergence speed, and high precision, making it suitable for fault identification in distribution networks.
A method for single-phase grounding fault line selection in small current grounding systems is proposed using modified CEEMDAN and convolutional neural network. The algorithm uses random forest and multiscale permutation entropy to modify the Complete Ensemble Empirical Mode Decomposition Adaptive Noise algorithm (CEEMDAN)[39]. The zero-sequence current of each line is decomposed into

intrinsic mode functions using the modified CEEMDAN (MCEEMDAN) algorithm. The colour images are fused into the GoogLeNet network, and fault line selection is realized as a probability output using the Softmax function. The method has strong feature extraction ability, high recognition accuracy, and good anti-noise and robustness.

A novel fault diagnosis method for rolling bearings is proposed, based on wavelet thresholding denoising, complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) energy entropy, and particle swarm optimization least-squares support vector machine (PSO-LSSVM). This method reduces noise-induced interference in vibration signals, obtains multiple groups of intrinsic mode functions (IMFs), and selects feature vectors by combining correlation coefficient and variance contribution rate. The energy entropy of the selected IMF component is then used in the PSO-LSSVM classifier for fault diagnosis and classification. The method achieves a 100% identification rate for various fault states of rolling bearings as shown in [40].

The paper proposed a fault diagnosis approach for gas pressure regulators, which is crucial for optimizing safety and reliability in natural gas pipeline networks[41]. The approach combines complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) and fuzzy c-means (FCM) clustering to classify three typical faults of gas regulators. The CEEMDAN approach decomposes intrinsic mode functions (IMFs), establishes feature vectors using the Hilbert marginal spectrum (HMS) of IMFs, and uses cluster centers and feature clustering algorithm to distinguish fault types. The experimental results showed high performance, with membership degrees optimized to be within 0.9 to 1.

2.3 Research Gap

Despite significantadvances in fault identification and classification by using different deep learning models in electrical distribution systems essential insufficient studies still exist. While some studied have incorporated convolutional neural networks(CNN), Graph learning, Recurrent Neural Network(RNN) to enhance fault detection, classification, limited research has focused on integrating modified CEEMDAN with deep learning for fault identification. Moreover, the challenge of handling noise, variability in fault conditions and complex signals in micro grids with distributed solar energy has not thoroughly addressed.Although some studieshave done in CEEMDAN and its modifications for fault identification they have not thoroughly examined and they do not fully utilize the potential deep learning models such as CNN and LSTM inconjunction with signal decomposition with CEEMDAN to improve accuracy and robustness in complex micro grid environments. Additionally, there is a need for further research to explore the Modified CEEMDAN with advanced deep learning frameworks to enhance fault identification and classification in solar integrated micro grids.

3. METHODOLOGY

The main contribution of this study is the proposal of entropy-based features from the time-frequency (T-F) decomposed fault signals, which aid in accurately quantifying and capturing the faults' properties. These features are lead to precise recognition performance of a deep learning model. In this paper, we looked at the T-F signal analysis tool Modified MCEEMDAN, which helps us pinpoint frequency rhythms important to the defects by breaking down the fault signals into mode functions. The Modified MCEEMDAN, an enhanced form of EMD, effectively addresses mode mixing difficulties and noisy IMFs. The popular deep learning models are used to categorize the entropy characteristics and choose a model that greatly aids in properly finding flaws.

A. Selection of relevant mode function for fault prediction

The T-F analysis of fault signals using Modified MCEEMDAN methodologies has proved its fault prediction capability, as evidenced by its mode functions, i.e., IMFs, in the preceding background section we have briefy discussed about the Modified MCEEMDAN method which is for feature extraction or the fault and no fault classes, as shown in Fig 2. It is worth observing that the faults of interest exhibit various oscillatory characteristics in terms of their amplitudes and frequencies, as shown in Fig. 11(a-e). Thus, the mode functions carry the discriminating characteristics of each fault. But MCEEMDAN was used to decompose all of the mode functions are irrelevant to their fault signals. As a result, choosing relevant mode functions while eliminating unnecessary ones is critical for reducing the suggested system's computing cost. We used a periodogram-based technique to determine the number of mode functions that represent the bulk of the frequency rhythms associated with fault signals

B. Entropy based feature extraction using mode function

In the previous section, we demonstrated the suitability of Modified MCEEMDAN decomposition for fault analysis of the signals. We have employed entropy features to further localize the discriminating characteristics of the T-F decomposed fault signals (mode functions), which assist us in capturing the non-linear characteristics of the mode functions. Entropy is a widely used non-linear time series metric that is utilized in numerous applications to analyse fault signals. This research examined the entropy computations of the T-F decomposed time-series fault signals using Modified MCEEMDAN. The results of this investigation may assist in resolving the mode-mixing and noise issues associated with conventional EMD-IMFs. Over the mode functions, various forms of entropies have been examined, including approximate, sample, and permutation entropy. The equations that are detailed below are used to calculate the entropy metrics for each mode function.

- **Approximate Entropy (ApEn)**

ApEn is a regularity statistic that quantifies the irregularity of time series variations for mode functions (IMFs).
Mathematically, the following equation is used to compute it:

$$ApEn(m,r,N) = 1n\left[\frac{X_m(r)}{X_{m+1}(r)}\right] \quad (12)$$

X_m and $X_{m+1}$are, respectively, the subsequence means of lengths m and m+1.

- **Sample Entropy (SaEn)**

SaEn measures the time series complexity of the IMFs. If template vectors of lengths m and m+1, respectively, are represented by the counters Xm and Xm+1, which are calculated as follows.

$$SaEn(m,r,N) = -\ln\left[\frac{X^{m+1}(r)}{X^{m}(r)}\right] \quad (13)$$

We choose m=2 and r=0.2 to compute the ApEn and SaEn of IMFs, as per previous work. The suggested strategy improved performance with these settings.

C. Analysis of the discriminating capabilities of the entropy features

Before moving on to the fault classification task, it is critical to demonstrate the discriminating skills (ability to detect distinct faults) of the various suggested entropy characteristics (ApEn and SaEn) in the Modified MCEEMDAN domain. This section compares the ability of various entropies to pinpoint fault-relevant features in other fault signal patterns. We generated the F-score to assess the discriminating capability of the entropy characteristics of various signal patterns. The F-score for a channel profile may be defined using its entropy feature vector.

$$F_w \frac{\sum_{c=1}^{4}((\bar{X}_w^{(c)} - \bar{X}_w)^2)}{\sum_{c=1}^{4}\left(\frac{1}{n_c - 1}\sum_{k=1}^{n_c}(\bar{X}_{k,w}^{(c)} - \bar{X}_w)^2\right)} \quad (14)$$

Where $\bar{X}_w^{(c)}$ and $\bar{X}_w$ are the fault category's and other fault categories' average attributes. $\bar{X}_{k,w}^{(c)}$ represent the characteristics of the kth fault signal sample of the c_{th} category, and nc indicates the total number of samples in the c_{th} category. The (F-score) is calculated for entropy characteristics to examine if the predicted values of a quantitative variable change across feature types. The characteristics with the highest F-score show strong discrimination ability and may be used to categorize problems. Additionally, we used the post-hoc analysis of the ANOVA test to carry out the statistical analysis, as shown in Fig. 8.

D. Classification using Machine learning model

The next step after constructing the feature matrix is to classify the faults. We investigated well-known deep learning techniques for identifying entropy features, including probabilistic, neural network, distance, and ensemble models. The findings demonstrate that ensemble-based classifiers perform well, with XGBoost beating all of the investigated models as reported in the results section.

E. Tools & Techniques

In this study TensorFlow/Keras used to train the MCEEMDAN model. Scikit-learn is utilised for data preparation, which includes cleaning, normalization, and noise reduction, as well as producing evaluation metrics like accuracy, precision, recall, and F1-score. Pandas and NumPy are crucial for data processing, huge dataset management, and integration with other tools like as TensorFlow and Scikit-learn. Visualization tools like as Matplotlib and Seaborn help in model performance analysis by generating confusion matrices and ROC curves.

To integrate Solar into an IEEE-33 bus system for load flow analysis, utilize MATLAB/SIMULINK software. The IEEE 33-bus distribution system's bus voltage and branch loss are analysed using the MATLAB programming environment. According to one source, this program used to simulate and evaluate the performance of a PV-integrated IEEE-33 bus test system. This complete technique guarantees reliable model assessment and fault identification in the solar-integrated IEEE 33-bus system.

F. Library facilities

- Utilise TensorFlow/Keras to construct and train the model.
- Scikit-learn is used for data pre-processing and evaluating metrics.
- Pandas and NumPy are utilised for the purpose of manipulating and analysing data.
- Used MATLAB2023a version to build the model.

4. RESULTS

4.1 Confusion Matrices

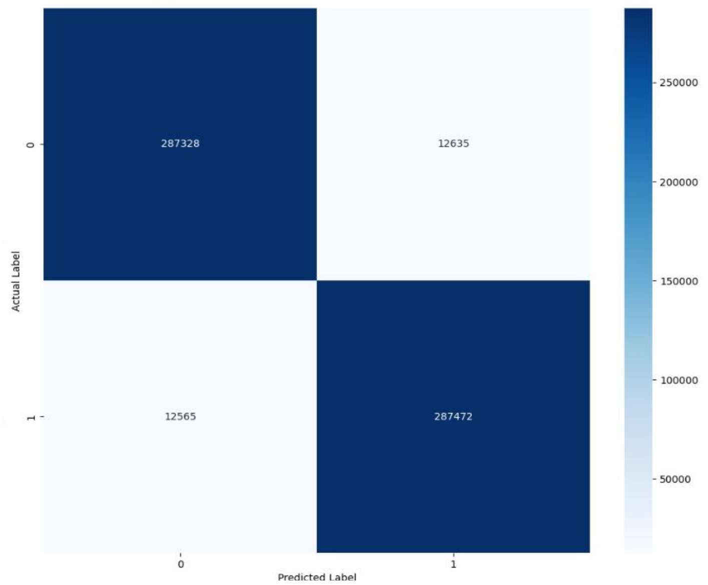

Figure 2. Confusion metrics of XgBoost

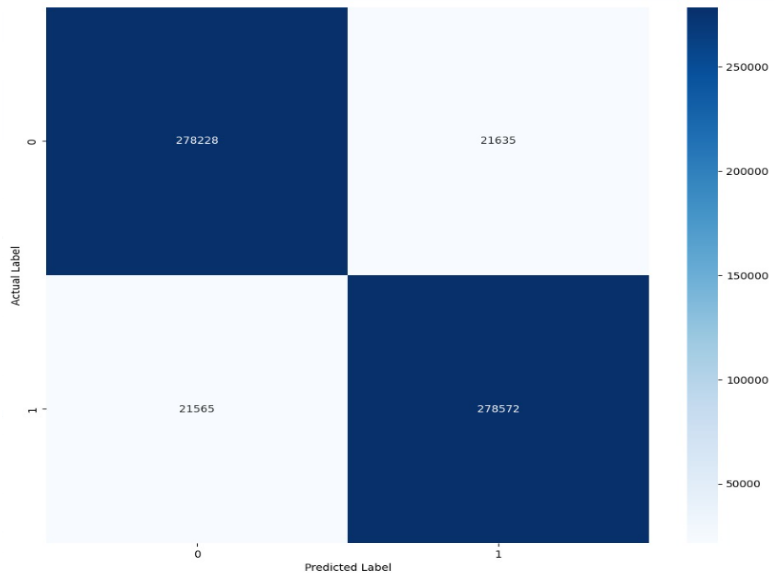

Figure 3. Confusion metrics of ANN

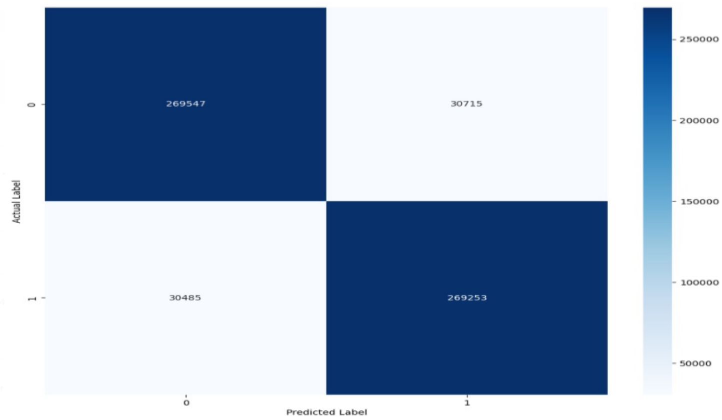

Figure 4. Confusion metrics of LS-SVM

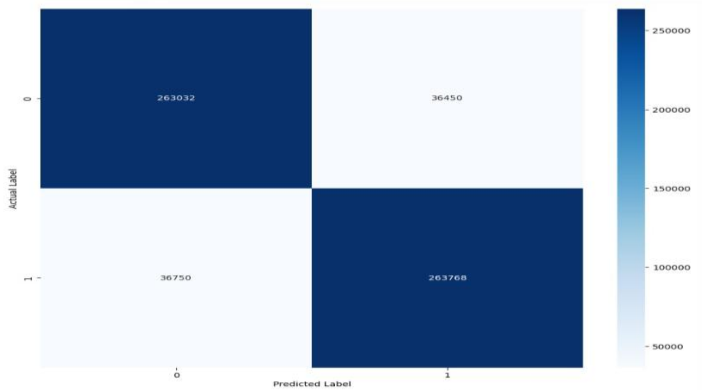

Figure 5. Confusion metrics of CNN

The confusion matrices for the XGBoost,LS-SVM, and CNN, ANN models demonstrate their performance in fault detection for the solar-integrated IEEE 33-bus system. The XGBoost model outperforms the other models in terms of true positive and false negative rates, showing that it is more accurate at identifying flaws. The XGBoost matrix performs moderately, with somewhat greater false negatives, indicating worse accuracy and recall compared to its modified equivalent. The LS-SVM and CNN, ANN models perform comparably poorly, as shown by their confusion matrices with larger false positives and false negatives, indicating that these models are less successful at detecting flaws in non-stationary and non-linear data. Overall, the Modified XGBoost surpasses the other models by providing higher fault detection accuracy and resilience in this setting.

4.2 ROC CURVES

The Receiver Operating Characteristic (ROC) curve is a graphical diagram that shows the diagnostic capabilities of a binary classifier when its discrimination threshold is changed. The ROCcurve is created by calculating the true positive rate (TPR) and false positive rate (FPR) for each conceivable threshold (in practice, at predetermined intervals) and then plotting TPR vs FPR.

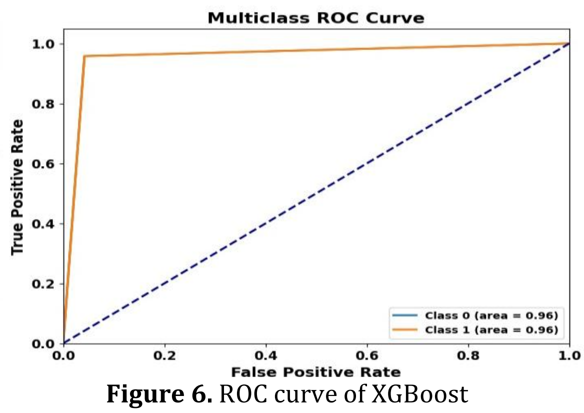

Figure 6. ROC curve of XGBoost

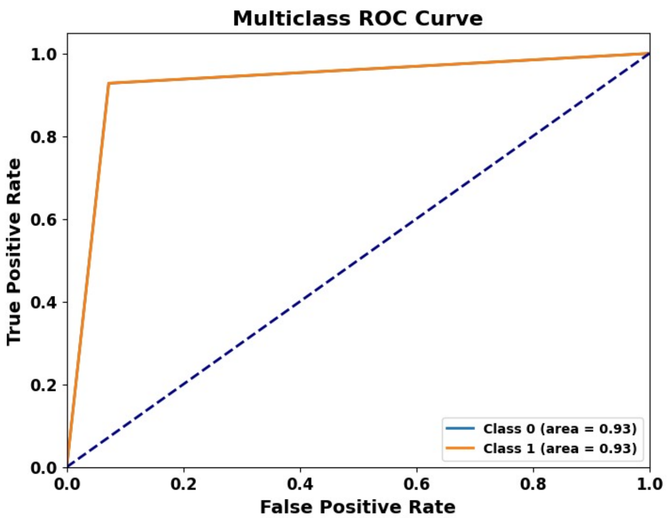

Figure 7. ROC curve of ANN

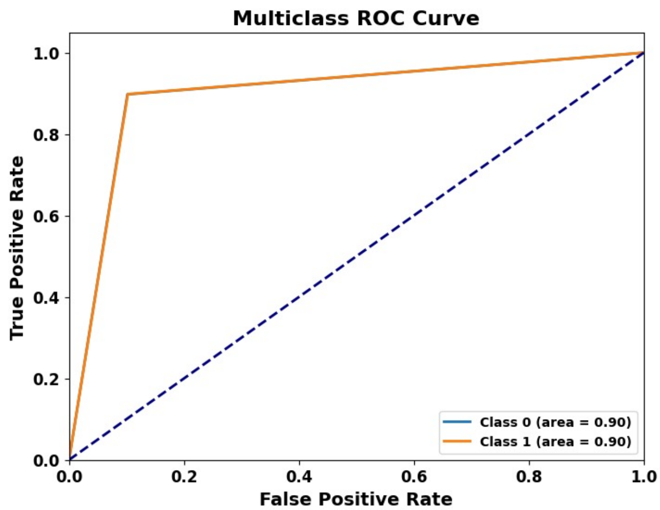

Figure 8. ROC curve of LS-SVM

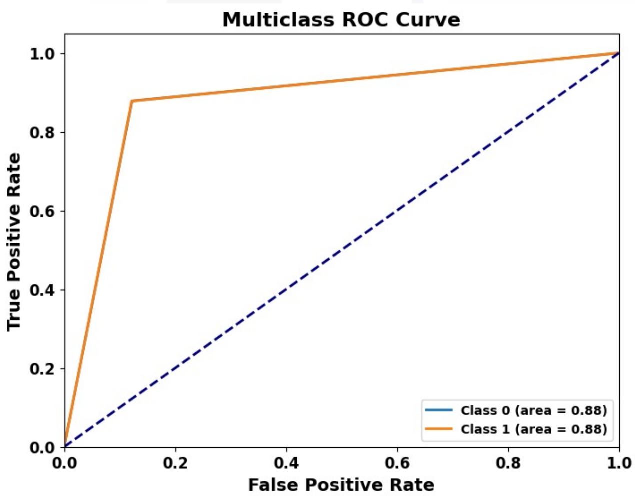

Figure 9. ROC curve of CNN

The ROC curve study shows that the XGBoost model surpasses the other classifiers in defect identification, with an AUC (Area Under the Curve) of 0.96, indicating superior discriminative capacity. ANN follows with an AUC of 0.93, demonstrating strong but slightly lesser performance than XGBoost. LS-SVM gets an AUC of 0.90, indicating high classification capabilities but not as resilient as earlier models.

Finally, CNN has the lowest AUC of 0.88, showing a lesser ability to discern between fault and no-fault scenarios. Overall, XGBoost is the best classifier, continuously surpassing ANN, LS-SVM, and CNN in terms of accuracy and reliability in defect identification.

4.3 Evaluation Metrics

Evaluation metrics are critical for determining the success of machine learning models. They give quantitative measurements for model selection and Hyperparameter adjustment. Different tasks need different metrics, and knowing which one to employ is critical to properly interpreting model findings.

Table 1. Evaluation Metrics of different Models

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
XGBoost	0.95	0.93	0.96	0.94
LS-SVM	0.89	0.87	0.9	0.88
CNN	0.87	0.84	0.86	0.85

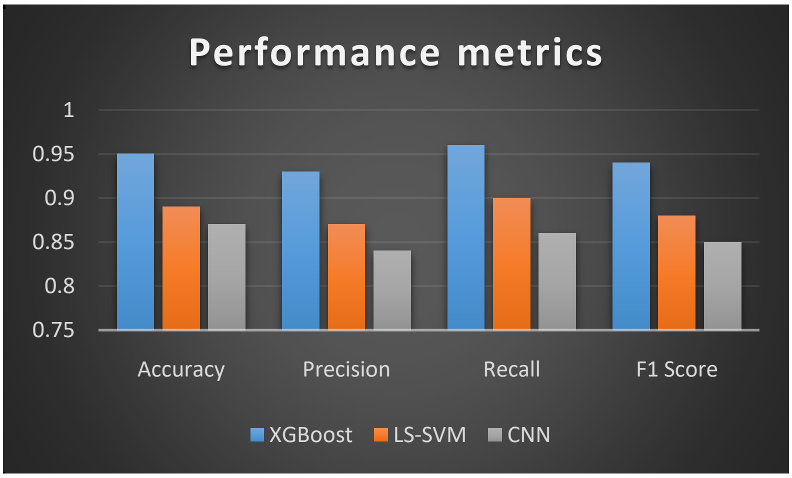

Figure 10. Performance metrics of various models

This table presents the performance metrics for the different models tested for fault classification. XGBoost has an excellent 95% accuracy, as well as good precision (0.93), recall (0.96), and an F1 score of 0.94, showing that it is adept at precisely finding flaws while reducing mistakes. In comparison, the LS-SVM model performs somewhat worse, with an accuracy of 89% and metrics for precision (0.87), recall (0.90), and F1 score (0.88). The CNN model has the lowest performance of the three, with an accuracy of 87% and precision, recall, and F1 scores of 0.84, 0.86, and 0.85, respectively. These findings show XGBoost as the preferred model owing to its improved ability to balance accuracy and precision in defect identification.

Table 2. Entropy Table for various faulty conditions

faults	**Approximate Entropy (ApEn) volatge**			**Sample Entropy (SaEn) voltages**			**Approximate Entropy (ApEn) currents**			**Sample Entropy (SaEn) currents**		
	A-phase	B-phase	c-phase	A-phase	B-phase	c-phase	A-phase	B-phase	c-phase	A-phase	B-phase	c-phase
AG	0.0147	0.019	0.0147	0.014695	0.01895	0.01475	0.0066	0.0162	0.0124	0.00662	0.01622	0.01243
AB	0.0121	0.0163	0.0131	0.012149	0.01626	0.0131	0.0058	0.0058	0.0123	0.00582	0.00577	0.01229
ABC	0.012	0.013	0.0127	0.011996	0.01305	0.01268	0.0066	0.0058	0.0057	0.00664	0.00576	0.00565
ABCG	0.012	0.0127	0.0135	0.012039	0.01274	0.01347	0.0065	0.0057	0.0057	0.00646	0.00574	0.00567
no	0.016	0.019	0.01	0.016	0.018	0.012	0.013	0.016	0.01	0.012	0.016	0.012

fault			3	039	95	96		1	26	98	1	59

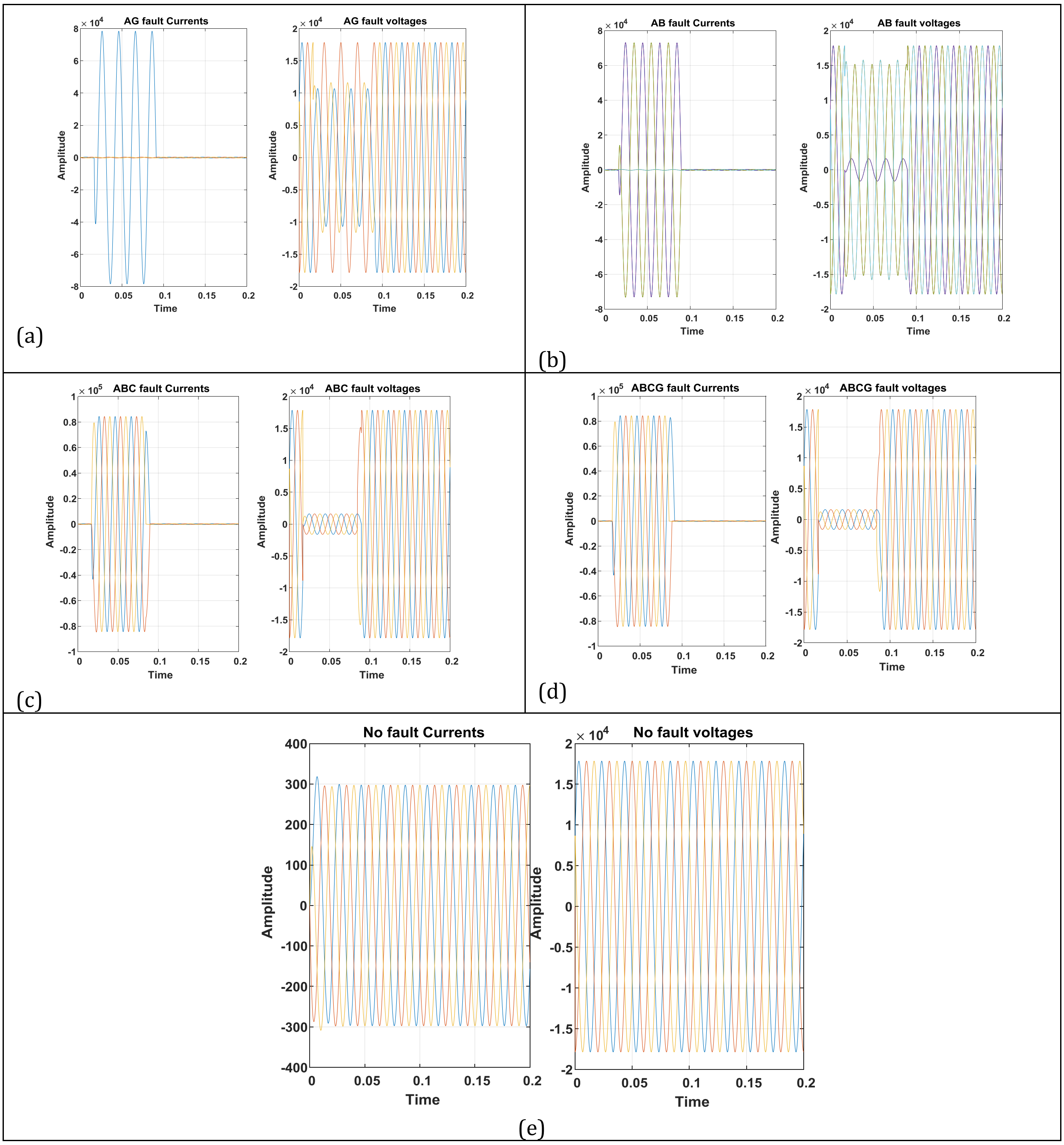

Figure 11 (a-e). Various IMFs obtained after CEEMDAN for five fault categories of a single participant show variations in the amplitudes.

The above table and graphs presents the Approximate Entropy (ApEn) and Sample Entropy (SaEn) values for voltage and current across various fault conditions, including AG (single-phase), AB (phase-to-phase), ABC (three-phase), ABCG (three-phase-to-ground), and a 'no fault' scenario. For voltages, ApEn values range from 0.012 to 0.0147, indicating lower complexity under fault conditions, while SaEn shows slightly higher variability, particularly in the A-phase. Current measurements reflect even lower entropy values, suggesting less complexity in fault-induced current signals. The 'no fault' condition displays higher entropy values in both voltage and current, signifying a more complex signal structure during normal operation. These results demonstrate that entropy measures, particularly when derived using the Modified MCEEMDAN method, effectively capture and distinguish between fault and no-fault conditions, providing valuable insights for fault detection and system diagnostics.

Table 3. Reported accuracy (%) of the MCEEMDAN approach with All features with various classifiers.

Classifiers	ApEn	SaEn
Naive Bayes	64.5	67.3
ANN	69.6	66.5
KNN	70.3	72.9
ELM	73.89	74.2
LS-SVM	78.8	81.1
DT	81.3	84.4
RF	83.2	84.2
Adaboost	85.6	86.5
XGBoost	97.5	95.61

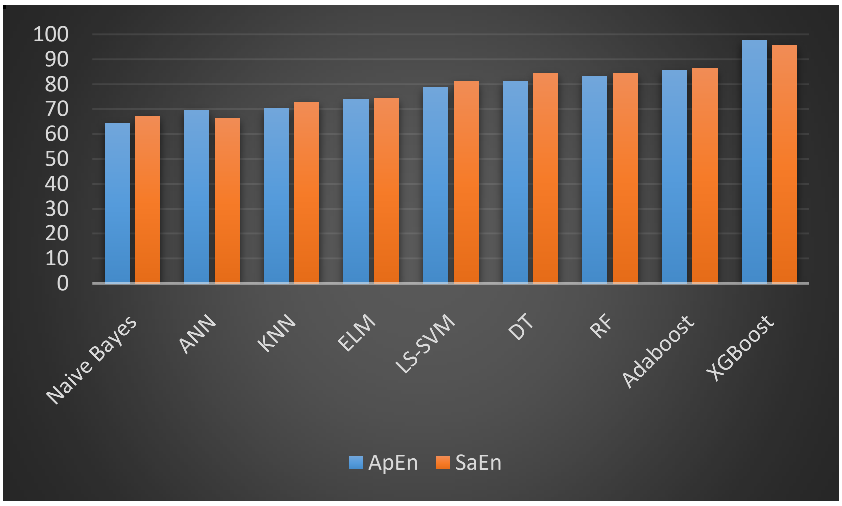

Figure 12. Reported accuracy (%) of the MCEEMDAN method using All-Channel features and several classifiers

This table compares the performance of several classifiers based on two entropy features: approximate entropy (ApEn) and sample entropy (SaEn). Each classifier's accuracy is expressed as a percentage. Notably, XGBoost performs very well, obtaining the greatest accuracy of 97.5% with ApEn and 95.61% with SaEn, suggesting a significant capacity to identify faults efficiently. Other classifiers, such as Adaboost and Random Forest (RF), exhibit competitive results; nonetheless, XGBoost surpasses all models investigated in this research.

4.4 Discussion

The findings of this investigation demonstrate the XGBoost classifier's strong performance, especially when combined with features acquired using the Modified CEEMDAN approach. By successfully decomposing fault signals into intrinsic mode functions, Modified CEEMDAN improved the extraction of entropy-based characteristics that are critical for discriminating between various fault states. The high accuracy of 97.5% attained by XGBoost, particularly with approximation entropy features, demonstrates the model's ability to use these properly chosen features to improve fault classification.

The comparison research found that, although other classifiers such as LS-SVM and ANN performed well, they fell behind XGBoost, which consistently shown superior accuracy and recall. This emphasizes the significance of integrating sophisticated feature extraction techniques with strong classification algorithms in improving fault detection accuracy in complicated systems like as the solar-integrated IEEE 33-bus system. The research underlines the need of further exploring feature extraction approaches and machine learning models to improve problem detection and ensure dependable performance under changing operating settings.

5. CONCLUSION

Finally, this work proved the efficacy of integrating the Modified CEEMDAN feature extraction approach with sophisticated machine learning classifiers for defect detection in a solar-integrated IEEE 33-bus system. The Modified CEEMDAN technique established a strong foundation for decomposing fault signals into intrinsic mode functions, allowing for the extraction of entropy-based characteristics that accurately define fault signals' nonlinear and dynamic properties.

The results showed that the XGBoost classifier beat other models, with an outstanding 97.5% accuracy using approximation entropy features. This performance demonstrates the classifier's capacity to harness the discriminative potential of the retrieved features, resulting in enhanced fault classification skills. The comparison research revealed that, although other classifiers, such as LS-SVM and ANN, performed satisfactorily, they lacked the accuracy and dependability of XGBoost.

Furthermore, the research emphasized the relevance of feature selection in reducing computing costs while improving model performance. The use of entropy measures, such as Approximate Entropy and Sample Entropy, proven to be helpful in capturing the intricacies of fault signals, allowing for more accurate detection and categorization.

Overall, this study provides important insights into the integration of modern signal processing methods with machine learning approaches for defect detection in electrical systems. It emphasizes the possibility for future breakthroughs in this subject by investigating more feature extraction techniques and classification algorithms. Future research might look at the applicability of this framework to various kinds of electrical systems and failure scenarios, thereby improving the reliability and efficiency of power distribution networks.

REFERENCES

[1] L. Niu, G. Wu, and Z. Xu, "Single-phase fault line selection in distribution network based on signal injection method," Ieee Access, vol. 9, pp. 21567–21578, 2021.

[2] H. Liang, Y. Liu, G. Sheng, and X. Jiang, "Fault-cause identification method based on adaptive deep belief network and time–frequency characteristics of travelling wave," IET Gener. Transm. Distrib., vol. 13, no. 5, pp. 724–732, 2019.

[3] W. Wang et al., "Frequency disturbance event detection based on synchrophasors and deep learning," IEEE Trans. Smart Grid, vol. 11, no. 4, pp. 3593–3605, 2020.

[4] L. Xiang, P. Wang, X. Yang, A. Hu, and H. Su, "Fault detection of wind turbine based on SCADA data analysis using CNN and LSTM with attention mechanism," Measurement, vol. 175, p. 109094, 2021.

[5] Q.-L. Zhang and Y.-B. Yang, "Sa-net: Shuffle attention for deep convolutional neural networks," in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2021, pp. 2235–2239.

[6] Y. Xi et al., "Transmission line fault detection and classification based on SA-MobileNetV3," Energy Reports, vol. 9, pp. 955–968, 2023.

[7] F. Rafique, L. Fu, and R. Mai, "End to end machine learning for fault detection and classification in power transmission lines," Electr. Power Syst. Res., vol. 199, p. 107430, 2021.

[8] A. Raza, A. Benrabah, T. Alquthami, and M. Akmal, "A review of fault diagnosing methods in power transmission systems," Appl. Sci., vol. 10, no. 4, p. 1312, 2020.

[9] R. Asghar, F. R. Fulginei, M. Quercio, and A. Mahrouch, "Artificial Neural Networks for Photovoltaic Power Forecasting: A Review of Five Promising Models," IEEE Access, 2024.

[10] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization," Renew. Sustain. Energy Rev., vol. 124, p. 109792, 2020.

[11] S. Wang and H. Chen, "A novel deep learning method for the classification of power quality disturbances using deep convolutional neural network," Appl. Energy, vol. 235, no. September 2018, pp. 1126–1140, 2019, doi: 10.1016/j.apenergy.2018.09.160.

[12] L. Zheng, K. Jia, T. Bi, Z. Yang, and Y. Fang, "A novel structural similarity based pilot protection for renewable power transmission line," IEEE Trans. Power Deliv., vol. 35, no. 6, pp. 2672–2681, 2020.

[13] N. P. Theodorakatos, "Fault location observability using phasor measurement units in a power network through deterministic and stochastic algorithms," Electr. Power Components Syst., vol. 47, no. 3, pp. 212–229, 2019.

[14] R.-A. Tîrnovan and M. Cristea, "Advanced techniques for fault detection and classification in electrical power transmission systems: An overview," in 2019 8th International Conference on Modern Power Systems (MPS), IEEE, 2019, pp. 1–10.

[15] V. Le, X. Yao, C. Miller, and B.-H. Tsao, "Series DC arc fault detection based on ensemble machine learning," IEEE Trans. Power Electron., vol. 35, no. 8, pp. 7826–7839, 2020.

[16] K. Natarajan, P. K. Bala, and V. Sampath, "Fault detection of solar PV system using SVM and

thermal image processing," Int. J. Renew. Energy Res, vol. 10, no. 2, pp. 967–977, 2020.
[17] O. J. Singh et al., "Robust detection of real-time power quality disturbances under noisy condition using FTDD features," Autom. časopis za Autom. Mjer. Elektron. računarstvo i Komun., vol. 60, no. 1, pp. 11–18, 2019.
[18] F. Mohammadi, G.-A. Nazri, and M. Saif, "A fast fault detection and identification approach in power distribution systems," in 2019 International Conference on Power Generation Systems and Renewable Energy Technologies (PGSRET), IEEE, 2019, pp. 1–4.
[19] F. Mohammadi, C. Zheng, and R. Su, "Fault diagnosis in smart grid based on data-driven computational methods," in Proceedings of the 5th international conference on applied research in electrical, mechanical, and mechatronics engineering, Tehran, Iran, 2019.
[20] H. A. Shiddieqy, F. I. Hariadi, and T. Adiono, "Effect of sampling variation in accuracy for fault transmission line classification application based on convolutional neural network," in 2018 International symposium on electronics and smart devices (ISESD), IEEE, 2018, pp. 1–3.
[21] S. Chan, I. Oktavianti, V. Puspita, and P. Nopphawan, "Convolutional adversarial neural network (CANN) for fault diagnosis within a power system: addressing the challenge of event correlation for diagnosis by power disturbance monitoring equipment in a smart grid," in 2019 International Conference on Information and Communications Technology (ICOIACT), IEEE, 2019, pp. 596–601.
[22] J. Hong, Y. H. Kim, H. Nhung-Nguyen, J. Kwon, and H. Lee, "Deep-Learning Based Fault Events Analysis in Power Systems," Energies, vol. 15, no. 15, pp. 1–16, 2022, doi: 10.3390/en15155539.
[23] M. Biswal, "Modified Complete Ensemble Empirical Mode Decomposition based HIF detection approach for microgrid system," Int. J. Electr. Power Energy Syst. 141 108254, 2022.
[24] W. Xue, X. Dai, J. Zhu, Y. Luo, and Y. Yang, "A noise suppression method of ground penetrating radar based on EEMD and permutation entropy," IEEE Geosci. Remote Sens. Lett., vol. 16, no. 10, pp. 1625–1629, 2019.
[25] Z. Guo, M. Liu, Y. Wang, and H. Qin, "A new fault diagnosis classifier for rolling bearing united multi-scale permutation entropy optimize VMD and cuckoo search SVM," IEEE Access, vol. 8, pp. 153610–153629, 2020.
[26] K. Ras Guerriche and T. Bouktir, "Maximum loading point in distribution system with renewable resources penetration," Proc. 2014 Int. Renew. Sustain. Energy Conf. IRSEC 2014, pp. 481–486, Mar. 2015, doi: 10.1109/IRSEC.2014.7059767.
[27] B. K. Ponukumati, P. Sinha, M. K. Maharana, A. V. Pavan Kumar, and A. Karthik, "An Intelligent Fault Detection and Classification Scheme for Distribution Lines Using Machine Learning," Eng. Technol. Appl. Sci. Res., vol. 12, no. 4, pp. 8972–8977, 2022, doi: 10.48084/etasr.5107.
[28] J. Hu et al., "Fault Location and Classification for Distribution Systems Based on Deep Graph Learning Methods," J. Mod. Power Syst. Clean Energy, vol. 11, no. 1, pp. 35–51, 2023, doi: 10.35833/MPCE.2022.000204.
[29] X. Zheng, R. Jia, Aisikaer, L. Gong, G. Zhang, and J. Dang, "Component identification and defect detection in transmission lines based on deep learning," J. Intell. Fuzzy Syst., vol. 40, no. 2, pp. 3147–3158, 2021, doi: 10.3233/JIFS-189353.
[30] V. Rizeakos, A. Bachoumis, N. Andriopoulos, M. Birbas, and A. Birbas, "Deep learning-based application for fault location identification and type classification in active distribution grids," Appl. Energy, vol. 338, p. 120932, 2023, doi: https://doi.org/10.1016/j.apenergy.2023.120932.
[31] T. S. Kumar et al., "Deep Learning based Fault Detection in Power Transmission Lines," in 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), IEEE, 2022, pp. 861–867.
[32] S. Hou and W. Guo, "Fault identification method for distribution network based on parameter optimized variational mode decomposition and convolutional neural network," IET Gener. Transm. Distrib., vol. 16, no. 4, pp. 737–749, 2022, doi: 10.1049/gtd2.12324.
[33] A. R., D. S. Nair, R. T., and V. V., "A novel SVM based adaptive scheme for accurate fault identification in microgrid," Electr. Power Syst. Res., vol. 221, p. 109439, 2023, doi: https://doi.org/10.1016/j.epsr.2023.109439.
[34] P. Balamurali Krishna, A. V Pavan Kumar, and A. Karthik, "An Investigation of Fault Detection in Electrical Distribution Systems Using Deep Neural Networks BT - Soft Computing Applications in Modern Power and Energy Systems," O. H. Gupta, N. P. Padhy, and S. Kamalasadan, Eds., Singapore: Springer Nature Singapore, 2024, pp. 317–333.
[35] M. Bhatnagar, A. Yadav, and A. Swetapadma, "Integrating Distributed Generation and Advanced Deep Learning for Efficient Distribution System Management and Fault Detection," Arab. J. Sci. Eng., vol. 49, no. 5, pp. 7095–7111, 2024, doi: 10.1007/s13369-023-08663-2.
[36] E. Varney, "Institute of electrical and electronic engineers, inc," Handb. Transnatl. Econ. Gov.

Regimes, pp. 561–571, 2019, doi: 10.1163/ej.9789004163300.i-1081.483.
[37] D. Mnyanghwalo, H. Kundaeli, E. Kalinga, and N. Hamisi, “Deep learning approaches for fault detection and classifications in the electrical secondary distribution network: Methods comparison and recurrent neural network accuracy comparison,” Cogent Eng., vol. 7, p. 1857500, Jan. 2020, doi: 10.1080/23311916.2020.1857500.
[38] S. Z. Hou, W. Guo, Z. Q. Wang, and Y. T. Liu, “Deep-Learning-Based Fault Type Identification Using Modified CEEMDAN and Image Augmentation in Distribution Power Grid,” IEEE Sens. J., vol. 22, no. 2, pp. 1583–1596, 2022, doi: 10.1109/JSEN.2021.3133352.
[39] X.-R. Cheng, B.-J. Cui, and S.-Z. Hou, “Fault line selection of distribution network based on modified CEEMDAN and GoogLeNet neural network,” IEEE Sens. J., vol. 22, no. 13, pp. 13346–13364, 2022.
[40] W. Chen, J. Li, Q. Wang, and K. Han, “Fault feature extraction and diagnosis of rolling bearings based on wavelet thresholding denoising with CEEMDAN energy entropy and PSO-LSSVM,” Measurement, vol. 172, p. 108901, 2021, doi: https://doi.org/10.1016/j.measurement.2020.108901.
[41] S. Tian, X. Bian, Z. Tang, K. Yang, and L. Li, “Fault Diagnosis of Gas Pressure Regulators Based on CEEMDAN and Feature Clustering,” IEEE Access, vol. 7, pp. 132492–132502, 2019, doi: 10.1109/ACCESS.2019.2941497.