International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Three-Tier-Deep-Learning Model for Fault Classification in Power System using PMU Data: A Hybrid Meta-Heuristic Optimization

Mahesh Yenagimath[1]*, Dr. Shekhappa Ankaliki[2], Dr. Girish V.[3]

Submitted: 18/01/2024 **Revised**: 27/02/2024 **Accepted**: 04/03/2024

Abstract: The process of categorising and recognising various sorts of defects that occur inside a power system is referred as fault classification. It involves classifying and identifying the faults that occur in the system like Short Circuit errors, Frequency Disturbance events, High Impedance faults and Phase-balance errors. To overcome these faults, we proposed a three-tier deep learning model using PMU Data. The proposed a methodology consist of Pre-Processing stage followed by Feature extraction and Feature Selection. In Pre-Processing phase, the data are collected from the PMU Dataset and then pre-processed via Missing Data Imputation, Outlier Detection and Handling and Data normalization via Z-score normalization. The pre-processed PMU data are used to extract the features using Higher-Order Statistical Moments and Multivariate Statistical Measures. The extracted optimal features are used to select the relevant features using the Hybrid meta-heuristic optimization model which is combination of Darts Game Optimization (DGO) and Monarch Butterfly Optimization (MBO). In this paper work a Three-Tier-Deep Learning-Based model is developed for fault diagnosis which combines Artificial Neural Network (ANN), Optimized Convolutional Neural Network (CNN) and Bi-LSTM. And finally, the model will be fine-tuned by adjusting hyper parameters, architecture, or the optimization process to enhance fault diagnosis accuracy. The proposed model is executed using PYTHON Platform.

Keywords: *PMU Dataset, fault detection in power distribution system, ANN, Bi-LSTM and CNN.*

1. Introduction

The demand for electricity is increasing significantly as the world's population grows; as a result, the power systems (PS) must be capable of satisfy the demands of delivering harmless and ensure power for the expanding amount of users [1]. A significant portion of a nation's economy goes towards the production, transmission, and distribution of electricity. As a result, it's critical that all of the equipment in the power system run at maximum efficiency and be protected from accidents [2]. Power grids must include a protection system that can identify and isolate faults. A protective system should determine the necessary levels of sensitivity and selectivity. The capacity to identify and isolate defects quickly enough to prevent their propagation is referred to as sensitivity. By being selective, you can reduce the number of consumers who lose power throughout the isolation process [3]. In order to reduce the risk of power outages during severe weather and for aesthetic reasons, power distribution cables are often installed underground rather than overhead [4]. The frequency disturbance events (FDEs), which have an impact on the reliability of power systems, include generator trips (GT), line outages (LO), and load disconnections (LD) [5]. Transmission lines in the electric power system network account for 80% of all power system defects because of exposure to various environments and contact with humans or animals [6].For several reasons, underground (UG) cables are superior to overhead transmission networks and are a necessity in smart cities. One of the primary causes of power outages is reportedly problems in underground distribution networks [7].The upkeep and growth of modern activities have become more and more dependent on the steady flow of electricity because so many processes depend heavily on it. Power distribution system dependability indices are a major concern in this situation. High-impedance faults (HIFs) are one of the many fault types that frequently occur in distribution networks and therefore require extra care [8].Power system lines that are in contact with a high impedance surface while they are energised develop high impedance faults (HIFs). Arcing and flashover could happen at fault spots because of the significant potential difference created between the two contact sites [9]. Transformers are regarded as the essential piece of machinery for power transmission and distribution since they significantly affect how safely and reliably power systems run [10]. In electricity generating systems, transformers play a crucial function [11]. Bushfire eruptions have been significantly influenced by powerline failures, which have an impact on both the economy and people's way of life [12].

The prime contribution of this research work is arranged as follows,

- To propose Optimized-CNN to learn and extract spatial features from the power system data.

[1]*Department of Electrical and Electronics Engineering, Hirasugar Institute of Technology, Nidasoshi, Karnataka, India.*
[2]*Department of Electrical and Electronics Engineering, S.D.M College of Engineering and Technology, Dharwad, Karnataka, India.*
[3]*Executive Engineer (EI), ALDC, HESCOM Corporate office, HESCOM, Navanagar, Hubli, Karnataka, India*

- To propose the hybrid meta-heuristic optimization model (DGO+MBO) to choose the features from the excavated features and to optimize the weights of CNN.

2. Literature Review

In 2021, Gilanifaret al. [13] proposed a novel to improve knowledge and address concerns with the lack of labelled data for fault classification, a unique latent structure learning method was used within a multi-task learning framework. The recommended method avoids the restraints of dormant framework learning by stopsorters from being overfit to unlabeled data in addition to exploiting the dormantframeworkin underutilised unlabeled data.

In 2021, Rai*et al.* [14] presented a CNN, a deep learning technique, was tailored for fault sorting in dispersed networks linked with DGs. For the first time, CNN was utilized to identify faults in three-stage voltage and current signals of various defect classes and defect-free classes utilising raw and sampled data.

In 2013, Zhang*et al.* [15] proposed a defect classification method for neutral non-effectively earthed allocation systems on the basis of ANFIS. After faults occur, the temporary currents are retrieved using a wavelet transform. The fault identifiers are described in accordance with the statistical feature of transient currents in various defect kinds. The fault detectors can describe the characteristics of a fault sort and demonstrate various despot in various fault kinds.

In 2021, Belagoune*et al.* [16] introduced a unique DL classification and regression models for FRI, FTC, and FLP. These special methods examine the whole transient data from pre- and post-fault cycles while power and voltage signals are acquired by PMUs at various depots and used as input traits to the DRNN models.

In 2023, Goni *et al.* [17] introduced anML-based system for spontaneous FD and FC has been suggested. To imitate two distinct TLs and produce ordinary and defect data of 10 different sorts, MATLABSimulink was used. In contrast to TL-1, which had one generator and one load, TL-2 had two generators and three loads.

In 2021, Rafique*et al.* [18] suggested a fresh machine learning method for identifying and categorising faults in electrical power transmission networks. By using LSTM units that operate right on the functional data rather than features, this approach takes advantage of the secular sequence of the functional data for the PS to construct an "end to end" model.

In 2020, Moloi*et al.* [19] suggested a power distribution system fault protection diagnostic strategy. The strategy uses a WPD for signal handling, assessment and SMV for defect categorization and placement. Following the use of the WPD to extract important fault signatures, the SVM is utilised to classify faults and locate various fault states.

In 2022, Mo*et al.* [20] proposed a technique for defect categorization and localization in the power system network based on SR and GNN. Only by measuring a few important buses in the distribution network of power system can it reliably determine the fault kind and location accurately, certainly which lowers the distribution system's building costs inturn lowers the overall cost.

Table 1 shows that various authors reviews about research gaps.

Research Gap

Table 1: Reviews by various authors of research gaps

Author	Aim	Research Gap
Gilanifar*et al.*2021	To enhance knowledge and address the difficulties associated with the lack of sufficient labelled data for fault classification.	The link between labelled and unlabeled data was assumed incorrectly.
Rai*et al*. 2021	To classify all fault types in a distribution network with DG integration, including no-fault.	Determining faults in the actual distribution grid system Computation cost
Zhang*et al.*2013	To distinguish between the transient signals in the desired frequency range.	Should not be improved through the use of field samples with heavy loads. Failed to use the history fault data from the power system to validate the proposed approach.
Belagoune*et al.*2021	To verify the FRI and FTC model in the Large-Scale Multi-Machines Power Systems	Did not tested on larger power systems

Goni *et al.*2023	Using the ELM to make sure the electricity system is dependable and safe.	Did not analyse within a real-world scenario.
Rafique*et al.*2021	Eliminating the requirement for a difficult feature descent process and working rightly with functional data	The data acquired from the PS, both in actual-time and as noted data, was not tested by the proposed model.
Moloi*et al.* 2020	Utilising a half-cycle from the past-defect signal to categorise and find the defect	The proposed model did not take into account defect detection as renewable energy technologies were more widely adopted.
Mo*et al.*2022	To merely measure a few important buses to establish the nature and position of the defect.	It lacked distributed power access, strong fault resistance change, anti-noise capability, and the capacity to reconfigure distribution networks.

3.Proposed Methodology

3.1 Overview

The primary target of this research paper is fault classification in power distribution using a three-tier deep learning model. Which is anamalgamation ofANN, optimized CNN, and Bi-LSTM networks. the proposed model developed by following four major phases:'(1) Pre-processing, (2) Feature Extraction, (3) Feature Selection, (4)Three-Tier-Deep Learning-Based Fault Diagnosis and (5) Model Evaluation and Fine-tuning'. Fig. 1 illustrate the general architecture of the suggested model.

Step 1- Pre-processing:

Data collection: Gather PMU data from the power distribution system.

Missing Data Imputation: Apply techniques to estimate missing values in the PMU data.

Outlier Detection and Handling: Use statistical methods like Median Absolute Deviation (MAD) to identify and handle outliers in the PMU data.

Data normalization via Z-score normalization: Scale the PMU data using Z-score normalization to ensure consistent input for the model.

Step 2- Feature Extraction:

Higher-Order Statistical Moments: Calculate features such as skewness and kurtosis to capture non-Gaussian behaviour and asymmetry in the pre-processed PMU data.

Multivariate Statistical Measures: Explore cross-correlation matrices to capture dependencies and relationships between multiple PMUs or variables in the power distribution system.

Step 3- Feature Selection:

Hybrid meta-heuristic optimization model: The Darts Game Optimization (DGO) and Monarch Butterfly Optimization (MBO) techniques will be combined to select the most informative features for fault diagnosis.

Step 4- Three-Tier-Deep Learning-Based Fault Diagnosis:

ANN: An ANN will be utilized as the initial layer of the hybrid model to capture complex relationships between features and faults.

Optimized CNN: A CNN with weights optimized using the hybrid meta-heuristic optimization model (DGO+MBO) will effectively learn and extract spatial features from the PV array system data.

Bi-LSTM: A Bi-LSTM layer will be employed to catchsecularreliance and sequential patterns in the data for fault detection.

Ensemble and decision-making: The outputs of the ANN, optimized CNN, and Bi-LSTM will be combined to form an ensemble model. Appropriate decision-making techniques such as voting or averaging will be applied to make the final fault diagnosis based on the ensemble predictions.

Step 5- Model Evaluation and Fine-tuning:

Performance metrics: The execution of the hybrid deep learning model will be estimated using performance metrics

Hyperparameter tuning: The model will be fine-tuned by adjusting hyperparameters, architecture, or the optimization process to enhance fault diagnosis accuracy.

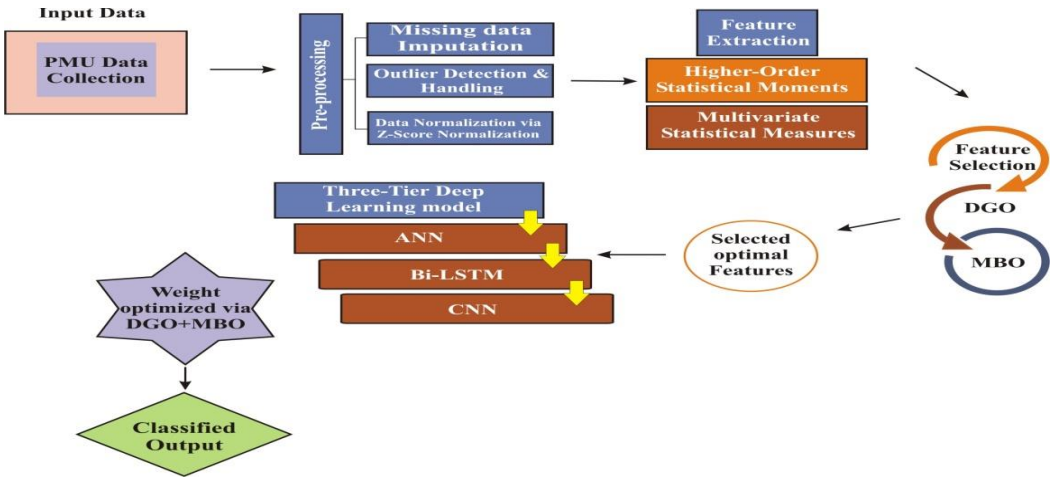

Fig. 1 General architecture of the proposed model

3.1.1 Pre-processing

The collected PMU data are pre-processed via Missing Data Imputation, Outlier Detection and Handling and Data normalization via Z-score normalization. Data preparation is the process of transforming raw data into an effective, understandable format. Real-world or fresh data is generallydeficient, prostrate to manualfaults, and has roving formatting. Data pre-processing corrects these issues while also enhancing the efficacy and completeness of datasets used for data analysis. Fig 2 shows the techniques of pre-processing.

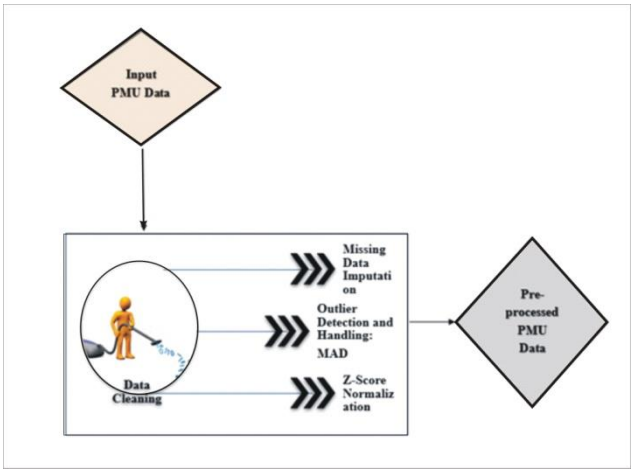

Fig. 2 Pre-Processing

3.1.1.1 Data collection

In this research paper the data are gathered by PMU dataset. Three separate datasets are taken into consideration to evaluate the noise information of PMUs used in systems with 'high, medium, and low voltages'. The first dataset contains information from a utility in the United States. It consists of 'voltage, current, and frequency information' from four PMUs that are situated on 345 KV buses. The data was collected at a declaration rate of 30 frames per second (fps). The monitoring system for the smart-grid guideplan on the Ecole Polytechnique Federale de Lausanne (EPFL) campus provided the second dataset. It is made up of data on 'voltage, current, and frequency' that was gathered from five PMUs positioned in a 20 kV active distribution network at a sample rate of 50 fps. The third set of datainclude of the voltage and frequency data sets (without current measurements) from the NCSU Synchro Phasor Lab. The data was recorded at 60 frames per second from PMUs installed on a laboratory's 120 v distribution system.

3.1.1.2 Missing Data Imputation

Imputation is a method for holding the bulk of the data and information in the dataset by stuffing in any intervals with a rate. Because it would be impractical to eliminate data from a dataset repeatedly, different types of methods are used. Additionally, adopting it will significantly reduce the dataset's size, increasingproblems about bias and impeding research.

3.1.1.3 Outlier Detection and Handling

Outliers are measuring points that stand out from the rest of the record. This unusual inspection, which generally come from imprecise observations or poor data entry, oftenwrap the data allocation. Detecting outliers Using

MAD (Median absolute deviation):It substitutes the median for the mean and the standard deviation (SD) for the variance.The mathematical representation of the MAD as shown in Eq (1),

$$MAD = B\, m_i\left(\left|x_i - m_j(x_j)\right|\right) \quad (1)$$

Where x_j is the $n$realremarks and m_i is the median of the series. $B = 1.4826$is a constant that refers to the assumption of the data's normality, which ignores the abnormality brought on by outliers.

3.1.1.4 Data normalization via Z-score normalization

Data normalisation is a method used in data mining to convert a dataset's values onto a standard scale. This is significant because many ML algorithms are sensitive to the magnitude of the input features and benefit from the normalisation of the data by producing more accurate results.

Z-score normalization: When a dataset is normalised using z-scores, each value is changed to have a mean of 0 and SD of 1. The formula of Z-score normalization is shown in the Eq (2)

$$Z = \frac{x-\mu}{\sigma} \quad (2)$$

Where x is the real value, σ is the standard deviation and μ mean value of the data.

3.1.2 Feature Extraction

The pre-processed PMU data are move on to the feature extraction section to extract the features. The feature extraction techniques include Higher-Order Statistical Moments (HOS) and Multivariate Statistical Measures. Feature extraction used to make the process more accurate and it increases the prediction power of the algorithm to select the most relevant features. Figure 3 shows the feature extraction techniques.

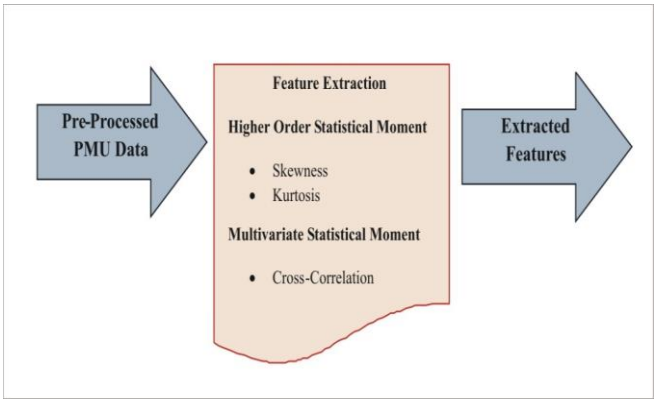

Fig. 3 Feature Extraction

3.1.2.1 Higher-Order Statistical Moments (HOS)

HOS are particularly helpful for evaluating skewness and kurtosis, shape parameters that are used to gauge how far a distribution deviates from the normal distribution. In this work the pre-processed data used to extract the HOS features like skewness and kurtosis. Table 2 represents the explanation and the formula of skewness, kurtosis.

Table 2: Mathematical Expression and Description of HOS

Metrics	Description	Mathematical Expression
skewness	Measures the symmetry, or the lack of symmetry	$skew = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(x_i - \underline{x})}{\sigma}\right]^3$
kurtosis	Measures, whether the data are heavy-tailed or light-tailed	$kurt = \frac{1}{N}\sum_{i=1}^{N}\left[\frac{(x_i - \underline{x})}{\sigma}\right]^4$

3.1.2.2 Multivariate Statistical Measures

To clarify and explain relationships between many factors connected to this data, multivariate analysis techniques are utilised in the evaluation and collecting of statistical data. In this work the cross-correlation matrices to capture dependencies and relationships between multiple PMUs or variables in the power distribution system.

3.1.3 Feature Selection

The procedure of choosing a division of pertinent features to be used in a model's building is known as feature selection. It increases the prediction power by selecting the relevant features. The extracted features are move on to feature selection process to select the most informative features for fault diagnosis using the hybrid meta-heuristic optimization model which is a combination of Darts Game

Optimization (DGO) and Monarch Butterfly Optimization (MBO). Fig 4 shows the feature selection of the proposed methodology. Fig.4 illustrate the feature selection process.

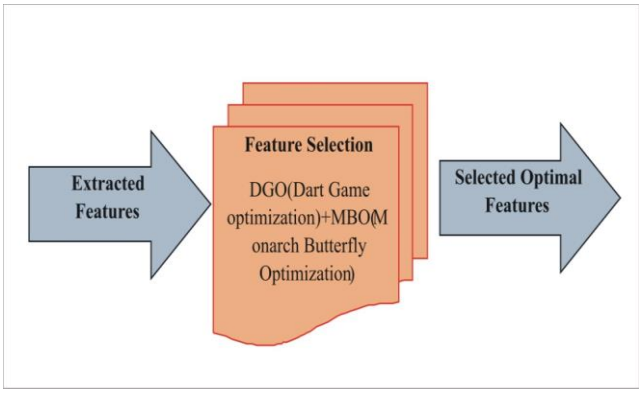

Fig 4 Feature Selection

3.1.3.1 DGO(dart Game Optimizer)

The dartboard and dart are the only two pieces of equipment needed to play this indoor game. Each of the six rings that make up the dashboard's 20 sectors holds a unique scoring point. The players score points by launching darts onto the board. The bull's-eye in the dartboard's centre is regarded to have the highest point total. The inside bull ring holds 50 points, the outer bull ring holds 25, the internalcontrast ring triples the points, and the external contrast ring doubles the points. The player population is shown as a matrix, where the row matrix shows players and the column matrix shows the individual player characteristics. To build and introduce a new optimizer, the power of the darts game is utilised.Players in this game are searcher agents from DGO, and their objective is to receive the best score possible.

Step 1:Generating the initial player population.

A matrix is used to represent the population of players, with each row denoting a single individual and each column denoting that player's many traits. The number of columns in this matrix actually reflects both the number of problem variables and their suggested values. Eq (1) specifies the matrix of players.

$$x = \left[x_1^1\ x_1^d\ x_1^m\ x_i^1\ x_i^d\ x_i^m\ x_N^1\ x_N^d\ x_N^m\ \right] \quad (3)$$

Here, m is the amount of variables, N is the count of players, x is the players' matrix, and x_i^d is the ith player's d-th dimension.

Step 2: Calculating the fitness function

By placing x_i in the fitness function, beneficialdata is acquired,

Step 3: Updating $f_{BEST}, x_{BEST}, F_{WORST}, and\ x_{WORST}$

Where f_{BEST} represents the best fitness value,

$$f_{BEST} = min\ (fit)_{N\times 1} \quad (4)$$

Here x_{BEST} is the best variables,

$$x_{BEST} = x(location\ of(fit)\ , 1{:}\,m) \quad (5)$$

F_{WORST} represents the worst fitness function value,

$$F_{WORST} = max\ (fit)_{N\times 1} \quad (6)$$

Where x_{WORST} is the worst variable values,

$$x_{WORST} = x(location\ of(fit)\ , 1{:}\,m) \quad (7)$$

Step 4: Updating f^n and p_i

f^n is the standardised value of fitness function, here, p_i is the probability function of i th player,

$$f^n = \frac{fit - F_{WORST}}{\sum_{j=1}^{N}\ (fit_j - F_{WORST})} \quad (8)$$

$$p_i = \frac{f_i^n}{max\ (f^n)} \quad (9)$$

Step 5: Calculating s_i^n

sc_iindicate the score candidates for ith player,s is the score matrix, this is arranged in order of high scores to low ones, s_i is the score for each throw of ith player, s_i^n is the standardised score of i th player,

$$c_i = round(82 \times (1 - p_i)) \quad (10)$$

$$sc_i = \{s(1{:}\,c), Rand < p_i\ s(c + 1.82), else \quad (11)$$

$$s_i = sc_i(k) \& 1 \le k \le 82 \quad (12)$$

$$s_i^n = \frac{\sum_{throws=1}^{3}\ s_i^{throws}}{180} \quad (13)$$

Step 6: Updating x_i

Upgrade the x_i using the following Eq(14)

$$x_i = x_i + Rand(1, m) \times (x_{BEST} - 3s_i^n x_i) \quad (14)$$

Step 7: Checking the stop condition

The stop condition establishes when the algorithm should stop looking for solutions and return the best one so far found.

Step 8: Printing solution

The best solution discovered during the optimisation process is often displayed by printing the solution.

3.1.3.2 MBO (Monarch Butterfly Optimization)

The migration pattern of the monarch butterfly served as inspiration for the algorithm. The MBO was initially introduced in a study where the creatorsmatched it to five other metaheuristic methodsand assessed it against 38 benchmark functions. In every instance, the MBO outperformed the other five metaheuristics.Monarch butterflies migrate to Mexico from their two homes in N-USA and S-Canada. They can move via either the immigration operator or the butterfly adjusting operator, which are two different behaviours. In other words, these two operators specify the instruction of search for each member of the community.

The algorithm's fundamental guidelines, which approximate the real system, are as follows:

There are people in both Lands 1 and 2, which are separate locales.By using the migration operator, the offspring are produced in both locations. If the new person is more physically fit than the parent monarch butterfly, the previous solution will be replaced. The best-fitting solutions don't change for the next iteration.

Migration operator

Monarch butterfly individuals (solutions), referred to collectively as the sn (solution number), are split into two subpopulations: Subpopulation 1 ($sp1$) and Subpopulation ($sp2$).The two lands where they are situated, Lands 1 and 2, respectively, are designated as $sp1$ and $sp2$.

$$sp1 = ceil(p.sn) \quad (15)$$

$$sp2 = sp - sp1 \quad (16)$$

where p is the Sub-population 1 individual's migration ratio.The following method is used to carry out the migration process:

$$x_{i,j}^{t+1} = \{x_{r1,j}^{t} \ if \ r \le p, x_{r2,j}^{t} \ otherwise \quad (17)$$

where $x_{i,j}^{t+1}$stands for the jth component of the ith individual at iteration $t+1$, and $x_{r1,j}^{t}$ and $x_{r2,j}^{t}$ stand for the locations of the $r1$ and $r2$singles, which are aimlessly chosen from $sp1$ and sp2, respectively, at iteration t.

Depending on the value of the parameter r, either Sub-population 1 or Sub-population 2 will be used to select the jth element of the new solution. A random number between 0 and 1 ($rand$) and the migration time ($peri$) are multiplied to find the value of r.

$$r = rand.peri \quad (18)$$

where the suggested $peri$ value is 1.2.

Butterfly Adjusting Operator

The butterfly adjusting operator is the second mechanism that directs people towards the best possible position within the search space. $If \ rand \le p$ during this procedure, the position will be adjusted using the next formula:

$$x_{i,j}^{t+1} = x_{best,j}^{t} \quad (19)$$

where $x_{best,j}^{t}$ stands for the jth parameter of the suitableresolution at iteration t, and $x_{i,j}^{t+1}$ represents the monarch butterflies' updated position.On the other hand, if the random number is greater than the migration ratio ($rand > p$), the position update follows the formula as follows:

$$x_{i,j}^{t+1} = x_{r3,j}^{t} \quad (20)$$

where $x_{r3,j}^{t}$ denotes the jth component of the current iteration's randomly chosen resolution from Sub-population 2.The person is updated according to the following equation if the equallyshared random number is higher than the adjusting rate (also known as the butterfly adjusting rate, BAR):

$$x_{i,j}^{t+1} = x_{i,j}^{t+1} + \propto \times (dx_j - 0.5) \quad (21)$$

where dx_j denotes the $i-$th individual walk step and dx is determined by the Lévy flight:

$$dx = Levy(x_i^t) \quad (22)$$

Following are the calculations for the scaling factor,

$$\propto = s_{max}/t^2 \quad (23)$$

where s_{max} is the maximum number of steps a person can take while walking, and t denotes the current iteration.The ideal balance between intensification and diversity is determined by the parameter$\propto$.When the value of $\propto$ is less, the search is conducted in a way that favours intensification (exploitation), whereas when the value of $\propto$ is bigger, exploration (diversification) predominates.

3.1.3.3 Hybrid

3.1.4 Three-Tier-Deep Learning-Based Fault Diagnosis

The selected features are move on to the Fault Diagnosis phase. A three-tire-deep-learning is used for fault diagnosis which is a combination ofANN, Bi-LSTM and O-CNN. Fig. 5 illustrate the Three-Tier-Deep Learning model. Figure 5 shows the Three-Tier-Deep Learning model-based fault diagnosis.

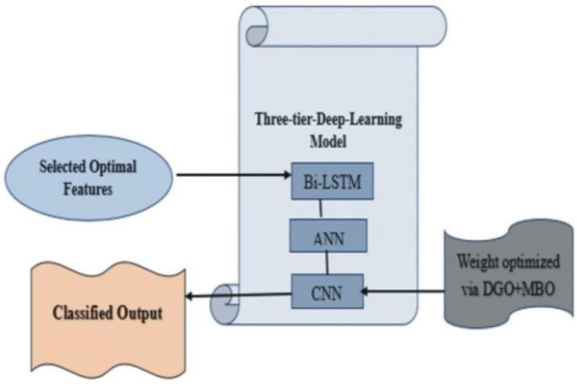

Fig. 5 Three-Tier-Deep Learning model

3.1.4.1 ANN

Neural networks are algorithms whose basic operation is modelled after how the brain operates. The input layer, hidden layer, and output layer make up the structure's three basic layers. There are a certain number of nodes (or neurons) that connect each layer to the next. Weights that are determined repeatedly during a training stage define the link between the neurons. Learning is obtained based on weight modification until a predetermined criterion is reached, i.e., a criterion to map the non-linearity between the inputs and the output. The networks are first initialised randomly during the training phase.Layers are fully interconnected in an MLP network, which is controlled using supervised learning. Equation (23) models a network with K inputs, a single output, Y and H hidden neurons.

$$Y = Y(X; w) = \sum_{j=0}^{H} \left[w_j F\left(\sum_{i=0}^{K} w_{ij}.X_i \right) \right] \quad (24)$$

The network parameters are represented by the vector w, while the weights and biases are represented by the values of i and j, which connect the layers.We can get various results from each mapping configuration since the neural network typically initialises its weights at random and has a number of parameters that can be set in a variety of ways.

3.1.4.2 Bi-LSTM

Bi-directional LSTM anticipate or labels the series of each component using a limitedseries, the setting of both the recent past and present, and other factors. Two parallel LSTMs, one moving from left to right and the other from right to left, resulted in this. Composite output refers to the prediction of a certain target signal. Using Equations (24) and (25), the furtheroperation of a bi-LSTM with inputs of L units and H as the number of hidden units is determined. In its hidden layer, the bidirectional LSTM network holds two values. a is used in the furthercomputation, and a is used in the reverse calculation. The values of a and a transposed determine the value of the output, y.

$$a_h^t = \sum_{L=1}^{l} x_L^t w_{lH} + \sum_{H',t>0}^{h} b_{H'}^{t-1} w_{H'H} \quad (25)$$

$$a_h^t = \theta_H(a_h^t) \quad (26)$$

3.1.4.3 CNN

Using grid-like matrices, convolutional neural networks (CNN), which have evolved from ANN, are mostly used to extract features from datasets. Convolutional, pooling, and fully linked layers are among the layers that make up a CNN in most cases. Convolutional layers perform the work of extracting features from the input data, and the task of down sampling the feature maps by pooling layers reduces their spatial dimensions. The fully linked layers at the network's end carry out classification or regression tasks based on the learned properties. Eq (26) shows the mathematical representation of CNN,

$$x_{i,j}^l = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} w_{ab} y_{(i+a)(j+b)}^{l-1} \quad (27)$$

3.1.5 Model Evaluation and Fine-tuning

Finally, the hybrid model is assessed using performance metrics such as “accuracy, precision, F-Measure, Sensitivity, Specificity, NPV, MCC, FPR, FNR”. Also, the hybrid model is fine-tuned by adjusting the hyperparameters, architecture, or the optimization process to enhance fault diagnosis accuracy.

3.1.5.1 Performance metrics

Several metrics are used to measure the performance including “Accuracy, Precision, F-Measure, Sensitivity, Specificity, NPV (Negative prediction value), MCC (Matthew’s correlation coefficient), FPR(False positive ratio), FNR (False Negative ratio)”.

Accuracy:

The accuracy is the ratio of properlycategorized data to all of the data in the log. The precision is described as,

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (28)$$

Precision:

By employing the entire count of examples used in the classification process, precision is the representation of the total number of genuine samples that are appropriately taken into consideration during the classification process.

$$Precision = \frac{TP}{TP+FP} \quad (29)$$

F-Measure:

The definition of the F-measure is the consonant mean of recollect rate and accuracy.

$$F_{Measure} = \frac{2\,Precision \times Recall}{Precision+Recall} \quad (30)$$

Sensitivity

The sensitivity value is obtained by just dividing the total positives by the proportion of true positive predictions.

$$Sensitivity = \frac{TP}{TP+FN} \quad (31)$$

Specificity

Specificity is premeditated by dividing the count of accurately anticipated negative outcomes by the total number of negatives.

$$Specificit\ y\ = \frac{TN}{TN+FP} \quad (32)$$

NPV

NPV determines the effectiveness of ananalytical test or another quantifiable metric.

$$NPV = \frac{TN}{TN+FN} \quad (33)$$

MCC

Below is a representation of the two-by-two binary variable association measure known as MCC,

$$MCC = \frac{(TP\times TN-FP\times FN)}{\sqrt{(TP+FN)(TN+FP)(TN+FN)(TP+FP)}} \quad (34)$$

FPR

The count of negative occurrences divided by the count of negative events that were incorrectly ranked as positive yields the false positive rate (false positives).

$$FPR = \frac{FP}{FP+TN} \quad (35)$$

FNR

The likelihood that an actual positive may be overlooked by the test is known as the false-negative rate, sometimes referred to as the "miss rate".

$$FNR = \frac{FN}{FN+TP} \quad (36)$$

3.1.5.2 Hyperparameter tuning

In this process the proposed model is fine-tuned by adjusting the hyperparameter, architecture or the optimization process, it can enhance the fault diagnosis accuracy.

4. Result and Discussion

This research presents a Three-Tier-Deep Learning-Based model for defect diagnosis that incorporates Bi-LSTM, Optimized Convolutional Neural Network (CNN), and Artificial Neural Network (ANN). A Hybrid meta-heuristic optimization model which is combination of Darts Game Optimization (DGO) and Monarch ButterflyOptimization (MBO) used to extract relevant features. In this section, the suggested approach is compared with different existing technique to know the performance efficiency of recommended method.

Comparative Analysis and Discussion

A various matrices like Accuracy, Sensitivity, Precision, specificity, MCC, F-measure, FPR, NPV, and FNR were taken for computation. To prove the performance improvement of newly developed model, it is compared with some existing methods such as ANN, Bi-LSTM, and ANN for each metrics. The evaluation results are displayed in Table 4.1. The performance of suggested technique us find out from the values given in table below:

Table 4.1. Evaluation of metrices for existing and recommended technique

	ANN	CNN	Bi-LSTM	Proposed
Accuracy	0.905210	0.878527	0.863822	0.941472
Precision	0.897769	0.921931	0.827439	0.933722
Sensitivity	0.912832	0.832212	0.907962	0.949389
Specificity	0.897688	0.924229	0.820266	0.933660
F-Measure	0.877015	0.849065	0.837852	0.912137
MCC	0.945138	0.870040	0.869607	0.989882
NPV	0.912801	0.843254	0.907400	0.949378
FPR	0.027645	0.021202	0.025918	0.020641
FNR	0.005920	0.004390	0.006943	0.003618

The data given in Table 4.1 stated that, a suggested technique reached best value in all evaluated metrics. Therefore, the suggested model has great performance than all other compared techniques. Furthermore, each metrics evaluation is displayed in the form of graph from Figure 4.1 to 4.9.

The most important metrics in computation is accuracy. It has been compared for proposed and various existing algorithm. The result of accuracy is displayed in Fig. 4.1

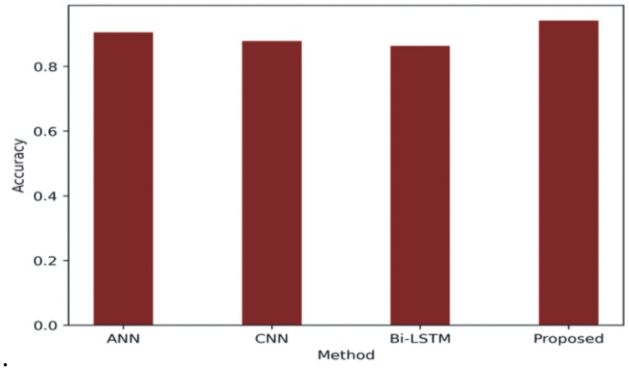

Fig. 4.1 Accuracy analysis of various algorithm

The above graph shown in Fig. 4.1 was made from values given in Table 4.1. The accuracy values are indicated as 0.905210, 0.878527, 0.863822, and 0.941472 for the methods ANN, CNN, Bi-LSTM, and developed model in this study respectively. The suggested approach has a highest accuracy value than all other values. Thus, the newly designed model has great accuracy when compared to other existing model which are in practice currently

The precision of recommended method is compared with various methods and the obtained values are graphically represented in below Fig. 4.2.

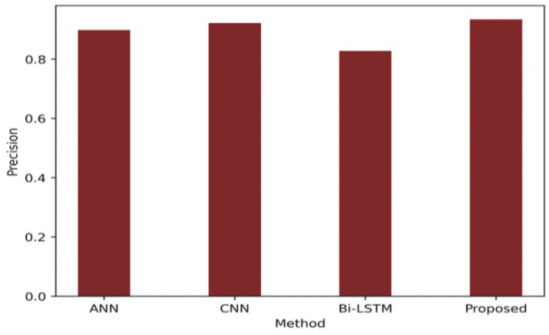

Fig. 4.2 Analysis of precision for several existing and proposed method

The values from Table 4.1 were used to create above graph in Fig. 4.2. The values of precision for ANN, Bi-LSTM, CNN, and created model in this investigation are given as 0.897769, 0.827439, 0.921931, and 0.933722correspondingly. The proposed method is the most precise of all the options. In comparison to previous models that are currently in use, the newly created model has a high level of precision.

The sensitivity of a suggested approach has been compared with other existing methods, and outcomes are visually displayed in Fig. 4.3 below.

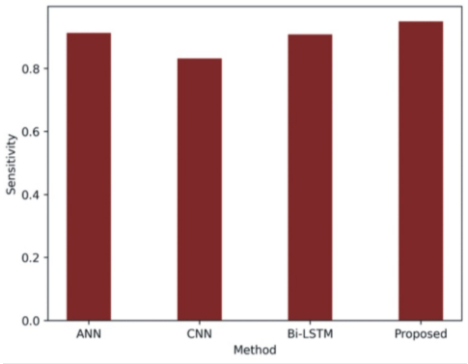

Fig 4.3 Comparison of sensitivity for suggested and existing techniques

The graph in Fig 4.3 above was made using the numbers from Table 4.1. In this study, the sensitivity values for CNN, Bi-LSTM, ANN and the generated model are presented as 0.907962, 0.832212, 0.912832, and 0.949389 respectively. The suggested approach is the most sensitivity available. The newly developed model offers a

high degree of sensitivity in contrast to earlier models that are currently in use.

The results of comparing a proposed approach's specificity to those of other techniques are shown graphically in Fig. 4.4 below.

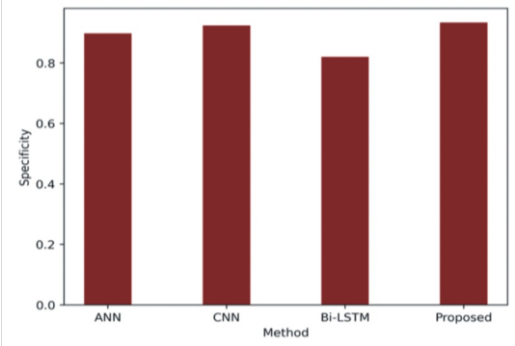

Fig 4.4.Analysis of specificity for a number of current and suggested methods

The data from Table 4.1 were used to create the graph in Figure 4.4 above. The specificity values in this study are 0.897688 for ANN, 0.924229 for CNN, 0.820266 for Bi-LSTM, and 0.933660 for the created model. The proposed strategy offers the highest level of specificity. Compared to older versions still in use, the recently designed model offers a high level of specificity.

In Fig. 4.5 below, the findings of comparing the F-measure of a suggested strategy to those of alternative approaches are represented visually.

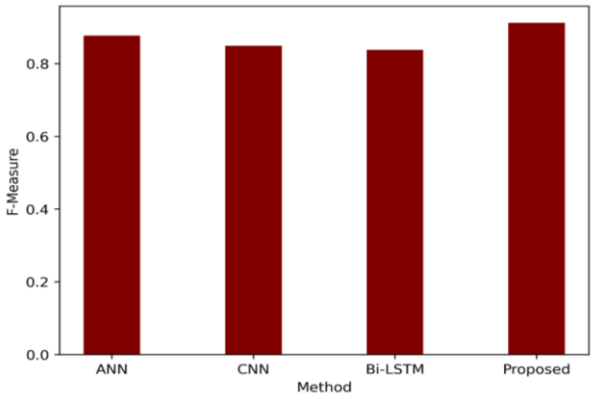

Fig 4.5 F-measure evaluation of proposed and current methods

The graph in Fig. 4.5 above was made using the information from Table 4.1. The F-measure values for ANN is 0.877015, Bi-LSTM is 0.837852,CNN is 0.849065, and the constructed model in this study is 0.912137. The highest F-measure value achieved by approach is the one that is suggested. The freshly developed model delivers a high level of F-measure as compared to previous versions that are still in use.

The results of contrasting an MCC of a recommended strategy to those of alternative techniques are illustrated in Fig. 4.6 below.

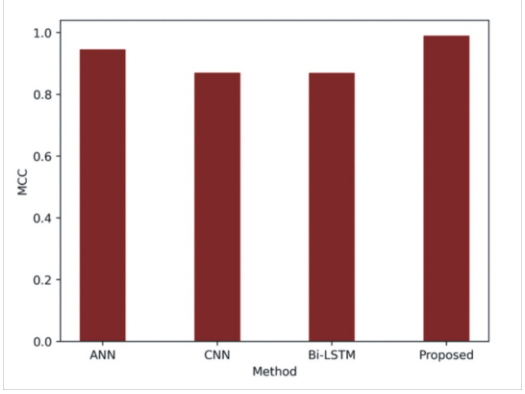

Fig 4.6 Analysis of MCC for different technique

The data from Table 4.1 were utilizedfordraw the graph shown in Fig.4.6 above. The MCC values for recommended model, CNN, Bi-LSTM, and ANN are 0.989882,0.870040, 0.869607, and 0.945138 respectively. The recommended strategy results in the greatest MCC score. Comparing the recently produced model to the older versions that are still in use, the newly developed model provides a highest MCC rate.

The Fig. 4.7 below shows an outcome of comparing an NPV of a suggested strategy to those of alternative methods.

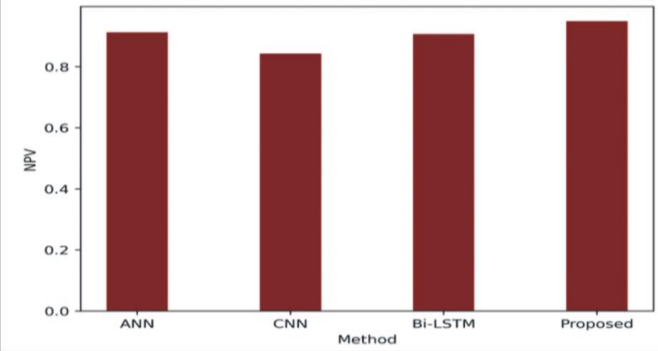

Fig. 4.7 Analysis of NPV for several existing and proposed method

The graph in Fig. 4.7 above was created using the data from Table 4.1. CNN, Bi-LSTM, ANN, and the suggested model all have NPV values of 0.843254, 0.907400, 0.912801, and 0.949378respectively. The best NPV score is obtained using a suggested approach. The designed model used in this study offers the greatest NPV rate as compared to the previous ones that are still in use.

Fig. 4.8 below illustrates the results of comparing the FPR of a proposed plan to those of competing strategies.

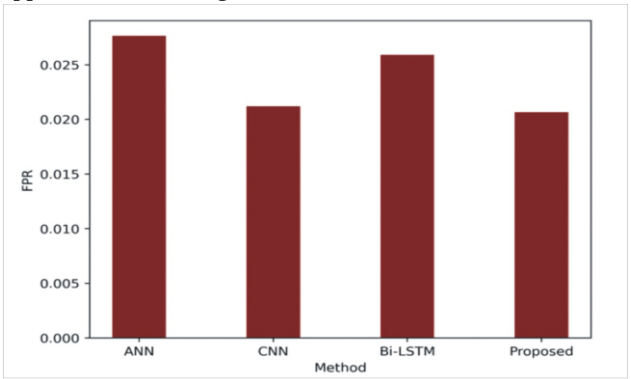

Fig. 4.8 FPR evaluation of current and proposed methods

The information from Table 4.1 was used to construct the graph in Figure 4.8 above. The FPR values for Bi-LSTM, CNN, ANN, and the proposed model are 0.025918,0.021202, 0.027645, and 0.020641respectively. The proposed method yields a lowest FPR value. In comparison to the earlier models that are still in use, the created model employed in this study has the lowest FPR rate. The model with least FPR value will perform more efficiently. Since the suggested technique has least value, it is a most efficient method than other models.

The outcomes of contrasting the FPR of a proposed plan to those of competing strategies are shown in Figure 4.9 below.

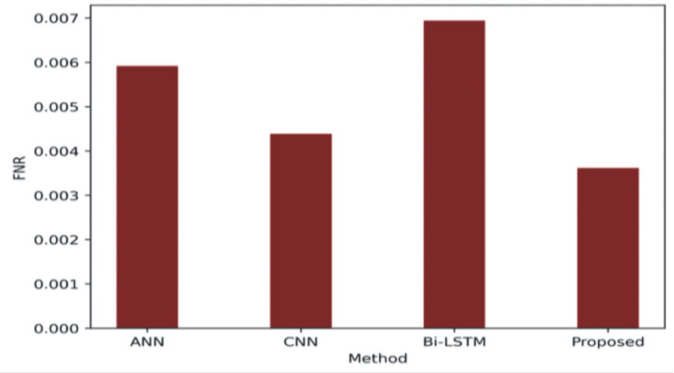

Fig. 4.9 Analysis of FNR for different technique

The graph in Fig. 4.9 above was drawn using the data from Table 4.1. The result of FNR values is given as 0.003618 for proposed method, 0.004390 for CNN, 0.005920 for ANN, and 0.006943 for Bi-LSTM. The developed model used in this study had the lowest FNR rate when compared to the prior models that are still in use. The model that performs more effectively has a lower FNR value. The recommended strategy is more effective than other models since it has the least value.

The proposed method of this research achieved best value in all metrices discussed above. From those analysis, the newly developed method has best performance than all other compared existing technique.

Evaluation For Class Accuracy of Recommended Method

Class accuracy, sometimes referred to as class-wise accuracy or per-class accuracy, is a statistic used to assess how well a classifier performs for each specific class. It calculates the percentage of cases that are successfully categorized into each class, giving information about how well the model is doing for each particular category. The class accuracy of suggested model is analyzed and the values are depicted in Table 4.2.

Table 4.2. Result for Class accuracy of suggested technique

No Fault Accuracy	0.5710
LG Fault Accuracy	0.7821
LL Fault Accuracy	0.7546
LLG Fault Accuracy	0.7575
LLL Fault Accuracy	0.7584
LLLG Fault Accuracy	0.7444

A metric called “No fault accuracy” or "overall accuracy" counts how many examples in the dataset were properly categorized as a percentage of all instances. It offers a broad overview of the classifier's overall performance across all classes. 636 labels are falls in no fault among total 715 labels. The No fault accuracy of suggested model is classified as 0.5710. In 296 predicted label, 255 labels are predicted in LG fault. This LG fault accuracy is listed as 0.7821 in proposed method. 285 labels are correctly predicted in LL fault out of 316 total predicted labels. The LL fault class accuracy of newly designed method is indicated as 0.7546. 356 predicted labels are considered for LLG fault and 304 labels are correctly predicted in this. The class accuracy of this fault was achieved 0.7575. 278 of the 328 predicted labels that are taken into account for the LLL fault are those that were predicted properly. This error's classification accuracy was 0.7584.289 of the 334 anticipated labels that are taken into consideration for the LLLGfault were correctly predicted. The classification accuracy for this fault was 0.7444.

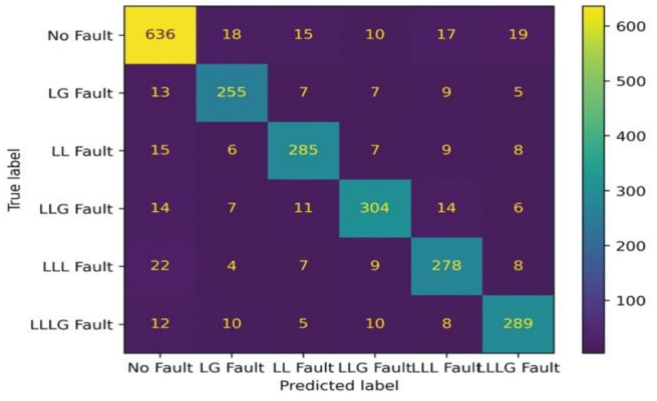

Fig.4.10 Analysis of True label Vs Predicted label

5.Conclusion

The paper has developed a three-tier-deep-learning model for fault classification in power system. The data were collected by PMU Dataset and then pre-processed via some techniques (Missing Data Imputation, Outlier Detection and Handling and Data normalization via Z-score normalization). From the pre-processed PMU data, the features were extracted using (HOS and Multivariate Statistical Measures). Using the hybrid model (DGO+MBO), the features were selected from the extracted features. A three-tier-deep-learning based model was used for fault diagnosis which is combination of ANN, Bi-LSTM and O-CNN. Also, the CNN’s weights were optimized using the hybrid model.(ResultsAccuracy and precision). The proposed model was executed using PYTHON Platform.

References

[1] Ghaemi, A., Safari, A., Afsharirad, H. and Shayeghi, H., 2022. Accuracy enhances of fault classification

and location in a smart distribution network based on stacked ensemble learning. Electric Power Systems Research, 205, p.107766.

[2] Thomas, J.B. and Shihabudheen, K.V., 2023. Neural architecture search algorithm to optimize deep transformer model for fault detection in electrical power distribution systems. Engineering Applications of Artificial Intelligence, 120, p.105890.

[3] Kurup, A.R., Summers, A., Bidram, A., Reno, M.J. and Martínez-Ramón, M., 2023. Ensemble models for circuit topology estimation, fault detection and classification in distribution systems. Sustainable Energy, Grids and Networks, 34, p.101017.

[4] Zhu, K. and Pong, P.W., 2019. Fault classification of power distribution cables by detecting decaying DC components with magnetic sensing. IEEE Transactions on Instrumentation and Measurement, 69(5), pp.2016-2027.

[5] Shadi, M.R., Ameli, M.T. and Azad, S., 2022. A real-time hierarchical framework for fault detection, classification, and location in power systems using PMUs data and deep learning. International Journal of Electrical Power & Energy Systems, 134, p.107399.

[6] Xi, Y., Zhang, W., Zhou, F., Tang, X., Li, Z., Zeng, X. and Zhang, P., 2023. Transmission line fault detection and classification based on SA-MobileNetV3. Energy Reports, 9, pp.955-968.

[7] Mishra, S., Swain, S.C., Biswal, T. and Routray, A., 2023. Tree Based Fault Classification in Underground Cable. Procedia Computer Science, 218, pp.524-531.

[8] Carvalho, J.G.S., Almeida, A.R., Ferreira, D.D., dos Santos Jr, B.F., Vasconcelos, L.H.P. and de Oliveira Sobreira, D., 2022. High-impedance fault modeling and classification in power distribution networks. Electric Power Systems Research, 204, p.107676.

[9] Shihabudheen, K.V. and Gupta, S., 2023. Detection of High Impedance Faults in Power Lines using Empirical Mode Decomposition with Intelligent Classification Techniques. Computers and Electrical Engineering, 109, p.108770.

[10] Hong, L., Chen, Z., Wang, Y., Shahidehpour, M. and Wu, M., 2022. A novel SVM-based decision framework considering feature distribution for Power Transformer Fault Diagnosis. Energy Reports, 8, pp.9392-9401.

[11] Sudha, B., Praveen, L.S. and Vadde, A., 2022. Classification of faults in distribution transformer using machine learning. Materials Today: Proceedings, 58, pp.616-622.

[12] Pirmani, S.K. and Mahmud, M.A., 2023. Advances on fault detection techniques for resonant grounded power distribution networks in bushfire prone areas: Identification of faulty feeders, faulty phases, faulty sections, and fault locations. Electric Power Systems Research, 220, p.109265.

[13] Gilanifar, M., Wang, H., Cordova, J., Ozguven, E.E., Strasser, T.I. and Arghandeh, R., 2021. Fault classification in power distribution systems based on limited labeled data using multi-task latent structure learning. Sustainable Cities and Society, 73, p.103094.

[14] Rai, P., Londhe, N.D. and Raj, R., 2021. Fault classification in power system distribution network integrated with distributed generators using CNN. Electric Power Systems Research, 192, p.106914.

[15] Zhang, J., He, Z.Y., Lin, S., Zhang, Y.B. and Qian, Q.Q., 2013. An ANFIS-based fault classification approach in power distribution system. International Journal of Electrical Power & Energy Systems, 49, pp.243-252.

[16] Belagoune, S., Bali, N., Bakdi, A., Baadji, B. and Atif, K., 2021. Deep learning through LSTM classification and regression for transmission line fault detection, diagnosis and location in large-scale multi-machine power systems. Measurement, 177, p.109330.

[17] Goni, M.O.F., Nahiduzzaman, M., Anower, M.S., Rahman, M.M., Islam, M.R., Ahsan, M., Haider, J. and Shahjalal, M., 2023. Fast and accurate fault detection and classification in transmission lines using extreme learning machine. e-Prime-Advances in Electrical Engineering, Electronics and Energy, 3, p.100107.

[18] Rafique, F., Fu, L. and Mai, R., 2021. End to end machine learning for fault detection and classification in power transmission lines. Electric Power Systems Research, 199, p.107430.

[19] Moloi, K., Ndlela, N.W. and Davidson, I.E., 2022. Fault Classification and Localization Scheme for Power Distribution Network. Applied Sciences, 12(23), p.11903.

[20] Mo, H., Peng, Y., Wei, W., Xi, W. and Cai, T., 2022. SR-GNN Based Fault Classification and Location in Power Distribution Network. Energies, 16(1), p.433.