Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Machine Learning -Based Data-Driven Fault Classification on Electric Power Transmission System

Faiz Ahmad[1], Dharmendra Kumar Dheer[1] and Ajit Kumar[1]
[1] *Electrical Engineering Department, NIT, Patna, India*
E-mail: faiza.ph21.ee@nitp.ac.in

ARTICLE INFO

Received: 15 Oct 2024

Revised: 28 Nov 2024

Accepted: 18 Dec 2024

ABSTRACT

This paper proposes a novel scheme to improve the performance metrics of machine learning (ML) models for the classification of the short circuit fault (SCF) on electric power transmission line (TL) by enhancing the quantity and quality of dataset. The dataset consists of 24 features and 55289 observations, where each observation represents the maximum and minimum values of the signals during the fault and post fault conditions. A comparative analysis is conducted against several ML to showcase the efficacy of proposed dataset. The ability to classify the faults using proposed dataset is also compared with Phasor Measurement Unit (PMU) based dataset on Kundur two-area four-Machine Power System. The various simulations, dataset collection and ML algorithm have been performed in MATLAB environment. The performance metrics such as accuracy, precision, recall and F1-score and training time of various ML algorithm trained by proposed dataset is much superior than PMU based dataset for the classification of short circuit fault on Transmission Line.

Keywords: ML Algorithm, Alpha-dataset, Beta-dataset, Short Circuit Fault, Fault Classification, Phasor Measurement Unit, Performance Metrics.

1. INTRODUCTION

The electrical fault causes disturbance and destabilization, resulting in very serious problems for the proper and real functioning of the power system. The transmission system interconnects all major generating stations and main load centres in the system, which forms the backbone of the integrated power system and operate at higher voltage levels [1]. In transmission lines, occurrence of power outages is significantly attributed to unpredictable and irregular faults [2], [3], [4]. The electrical faults on transmission line should be detected and classified accurately as fast as possible so that the stability of the system must be maintained [5]. Rapid fault classification can also accelerate the restoration process, reduces the duration of outages, and help the working personnel to repair faulty transmission line as quickly as possible.

In recent years ML techniques have shown significant promise in fault detection and classification in transmission line [6]. As a branch of artificial intelligence, ML enables computers to learn rules and patterns automatically from massive data by training and optimizing algorithms [7]. The major bottleneck in ML is data collection, and it has been an active research topic in multiple communities. The recent research in collection and handling of large data comes not only from ML, natural language, and computer vision, but also from the data management community [8]. As big data continues to grow in scale and influence, the quality of datasets has become a key factor in determining the effectiveness of machine learning models. Reliable and well-structured data is vital for extracting meaningful insights and ensuring the accuracy of model prediction [9]. The attention of researchers and practitioners has gradually shifted from advancing model design to enhancing the quality and quantity of data [10].

Jamil et. al. applied feedforward neural network along with back propagation algorithm on 300 km TL using 8712 data samples by measuring three phase voltage and current for fault classification [5]. Tong et. al. applied Graph Convolutional Neural Network (GCNN) on IEEE-39 bus system using 11900 data samples by measuring graphical based fault signals for detection and classification of faults [11].Nikhil et. al. employed KNN, SVM using 12000 kaggle dataset to classify the types of shunt faults [12]. Harish et. al. employed Weighted extreme ML algorithm on

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

IEEE-39 bus system using 24654 data samples by measuring PMU based voltages to classify the faults [13]. Elmasry et. al. applied anomaly-based fault detection on Kaggle dataset using 8712 data samples by measuring three phase voltage and current for fault identification and classification [14]. Fornas et. al. applied Time Domain Reflectometry (TDR) methodology on complex distribution network modelled in PSCAD using 200 simulated signals collected by impedance measurement to identify the faults [15]. Roy et. al. applied Long Short-Term Memory (LSTM), combination of LSTM network and Feed-Forward NN(FFNN) with back-propagation on Microgrid model by measuring three phase voltage and current to classify the faults [16]. Abed et. al. employed KNN using 12000 kaggle dataset to classify the faults [17]. Joythula et. al. applied Linear, Logistic, Polynomial, Naïve Bayes, DT using 12000 Kaggle dataset to detect and classify the faults [18]. Bouaziz et. al. applied RF, KNN, SVM on grid connected Doubly Fed Induction Generator based Wind Turbine using 48001 data samples by measuring three phase voltage and current for fault classification [19]. Najafzadeh et. al., applied Fuzzy Logic, Adaptive Fuzzy Neural Network (AFNN), Random Forest (RF), and Decision Tree (DT) on 600 Km TL using 1100 data samples by measuring three phase voltage and current for fault classification [20]. Alhanaf et. al., employed CNN, LSTM, Hybrid CNN-LSTM on IEEE-6 and IEEE-9 bus system using 62501 data samples by measuring three phase and current [6]. Kanwal et. al. applied discrete wavelet transform (DWT), Google Net, and probabilistic Neural Network (PNN) on Vietnam 220-500 kV, 8302 Km line using colour image with size 224×224×3 by measuring three phase voltage and current waveform for fault classification [21].

In this paper, unlike earlier works, instead of focusing on the algorithms, the quantity and quality of dataset are enhanced to improve the performance metrics of ML models to detect and classify SCF on TL. To achieve this aim, the following objectives are accomplished:

1) Simulation of various types of SCF on TL in a Two-area Four-Generator System for different values of fault resistance and different locations using MATLAB Simulation and m-file code.

2) Collection of 55,289 samples and 8 features using PMU based data named as Beta-dataset and 55,289 samples and 24 features by adding three phase voltage, three phase current, zero sequence voltage, zero sequence current to Beta-dataset named as Alpha-dataset.

3) Training and testing of various available ML model using Beta-dataset and Alpha-dataset.

4) Finding of Confusion Matrix (CM) of the trained and tested ML model.

5) Comparison of the ability of classifying the faults of various ML model when trained and tested by both the dataset.

6) Comparison of the accuracy of the proposed dataset with the existing works.

This study comprises five sections. Section 2 describes the modelling of power system. Section 3 presents the Machine Learning Approach. The process of fault classification is presented in section 4. Section 5 concludes this study.

2. MODELING OF POWER SYSTEM

The sub-transient model of machines with four rotor coils in each machine, known as Model 2.2 [23] has been used to study power system dynamics. Model of static voltage regulator and power system stabilizer are also included with the model of each machine. Also, all the loads in the system are assumed to be modelled as constant impedance loads. The dynamic equations are given as follows [24] (where i refers to the system's ith machine, $1 \leq i \leq m$).

$$\Delta\dot{\delta}^i = \left(\omega^i - \omega_0\right) = \Delta\omega^i \ (where\ \Delta\delta^i = \delta^i - \delta_0^i) \quad (1)$$

$$\Delta\dot{\omega}^i = \frac{\omega_0}{2H^i}\left(T_m^i - T_e^i\right) - \frac{D_i}{2H^i}\Delta\omega^i \quad (2)$$

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$\dot{E}_d^{i\,\prime} = \frac{1}{T_{q0}^{i\,\prime}}\left[-E_d^{i\,\prime} - \left(X_q^i - X_q^{i\,\prime}\right)\left(K_{q1}^i I_q^i + K_{q2}^i \frac{\Psi_{2q}^i + E_d^{i\,\prime}}{X_q^{i\,\prime} - X_l^i}\right)\right] \tag{3}$$

$$\dot{E}_q^{i\,\prime} = \frac{1}{T_{q0}^{i\,\prime}}\left[E_{fd}^i - E_q^{i\,\prime} + \left(X_d^i - X_d^{i\,\prime}\right)\left(K_{d1}^i I_d^i + K_{d2}^i \frac{\Psi_{1d}^i + E_q^{i\,\prime}}{X_d^{i\,\prime} - X_l^i}\right)\right] \tag{4}$$

$$\dot{\Psi}_{1d}^i = \frac{1}{T_{d0}^{i\,\prime\prime}}\left[E_q^{i\,\prime} + \left(X_d^{i\,\prime} - X_l^i\right) I_d^i - \Psi_{1d}^i\right] \tag{5}$$

$$\dot{\Psi}_{2q}^i = \frac{1}{T_{q0}^{i\,\prime\prime}}\left[-E_d^{i\,\prime} + \left(X_q^{i\,\prime} - X_l^i\right) I_q^i - \Psi_{2q}^i\right] \tag{6}$$

$$\dot{E}_{dc}^{i\,\prime} = \frac{1}{T_c^i}\left[\left(X_d^{i\,\prime\prime} - X_q^{i\,\prime\prime}\right) I_q^i - E_{dc}^{i\,\prime}\right] \tag{7}$$

$$\dot{V}_r^i = \frac{1}{T_r^i}\left[V^i - V_r^i\right] \tag{8}$$

$$E_{fd}^i = K_a^i\left[V_{ss}^i + V_{ref}^i - V_r^i\right] \tag{9}$$

$$E_{fdmin}^i \le E_{fd}^i \le E_{fdmax}^i \tag{10}$$

$$\begin{bmatrix} I_d^i \\ I_q^i \end{bmatrix} = \begin{bmatrix} R_s^i & X_q^{i\,\prime\prime} \\ -X_d^{i\,\prime\prime} & R_s^i \end{bmatrix}^{-1} \begin{bmatrix} E_d^{i\,\prime} K_{q1}^i - \Psi_{2q}^i K_{q2}^i - V_d^i \\ E_q^{i\,\prime} K_{d1}^i + \Psi_{1d}^i K_{d2}^i - V_q^i \end{bmatrix} \tag{11}$$

$$T_e^i = \frac{\omega_0}{\omega_i} P_G^i \tag{12}$$

$$P_G^i = V_d^i I_d^i + V_q^i I_q^i \tag{13}$$

$$Q_G^i = V_d^i I_q^i - V_q^i I_d^i \tag{14}$$

$$V_d^i = -V^i Sin(\delta^i - \theta^i) \tag{15}$$

$$V_q^i = V^i Cos(\delta^i - \theta^i) \tag{16}$$

$$I_g^i = \frac{\left[E_q^{i\,\prime} K_{d1}^i + \Psi_{1d}^i K_{d2}^i + j\left\{E_d^{i\,\prime} K_{q1}^i - \Psi_{2q}^i K_{q2}^i - E_{dc}^{i\,\prime}\right\}\right] e^{j\delta^i}}{R_s^i + jX_d^{i\,\prime\prime}} \tag{17}$$

The bus voltages $(V^i, \theta^i), 1 \le i \le M$ (total number of buses are M), can be obtained by the power flow equations (18) and (19). Equation (20) gives the network equations.

$$P^i = \sum_{j=1}^{M} V^i V^j \left[G^{ij} \cos\left(\theta^i - \theta^j\right) + B^{ij} \sin\left(\theta^i - \theta^j\right)\right] \tag{18}$$

$$Q^i = \sum_{j=1}^{M} V^i V^j \left[G^{ij} \sin\left(\theta^i - \theta^j\right) - B^{ij} \cos\left(\theta^i - \theta^j\right)\right] \tag{19}$$

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

$$Y_A = Y_N + Y_G + Y_L;\ Z_A = [Y_A]^{-1}; V = Z_A I_g \quad (20)$$

Here, V is column vector of bus voltages, I_g is column vector of current injections with ith element equal to I_g^i if i is a machine bus, else it is equal to zero; Y_N is network admittance matrix; Y_G is diagonal matrix of machine admittances with ith diagonal element is equal to $\frac{1}{(R_s^i + jX_d^{i\,\prime\prime})}$ if i is a machine bus, else it is equal to zero; similarly Y_L is a diagonal matrix of load admittances. Y_A and Z_A are augmented matrices of admittance and impedance respectively.

3. MACHINE LEARNING APPROACH:

In this section, the steps for model design and performance have been carried out. The problem statements encountered during the model construction are as follows:

3.1 Data Set Creation

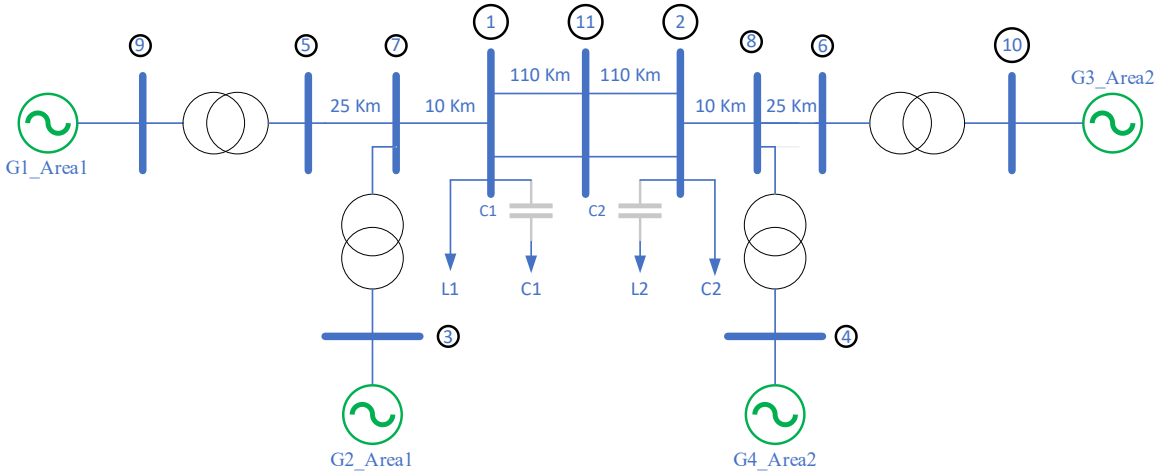

Figure 1: Kundur 2-Area Power System

The 2-area, 4-machines, 11-bus system is used as a case study in this paper. This case study system is publicly available in MATLAB and it is a well-known benchmark for conducting power system studies which is detailed in [1]. The system as shown in Figure 1 consists of two fully symmetrical areas connected by double circuit 220 Km transmission line. The synchronous generators are represented by their detailed models. For the analysis of classification of shunt fault, the faults are applied at different locations at a step of 2 Km between Bus 1 and Bus 2. The fault resistance in ohm is varied as [0.001 0.005 0.01 0.05 0.1:0.2:20] to collect the large data. The different types of SCF and its representation in m-file is shown in Table 1.

Table 1: Type of SCF and its representation in *m-file*

Fault	Fault type	Details	Representation in *m-file*			
			A	B	C	G
1	Single Line-to-Ground	Phase A to Ground	1	0	0	1
2		Phase B to Ground	0	1	0	0
3		Phase C to Ground	0	0	1	0
4	Double Line-Ground	Phase A, B to Ground	1	1	0	1
5		Phase B, C to Ground	0	1	1	1
6		Phase C, A to ground	1	0	0	1
7	Line-Line Fault	Phase A, B	1	1	0	0

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

8		Phase B, C	0	1	1	0
9		Phase C, A	1	0	1	0
10	Three phase Fault	Phase A, B, C	1	1	1	0
11	Three phase to Ground Fault	Phase A, B, C-Ground	1	1	1	1
12	No Fault	No phase involved	0	0	0	0

For the implementation of classification algorithms, two datasets $\alpha_{dataset}$ *and* $\beta_{dataset}$) with 24 parameters $[V_{m,n}^{abc}, I_{m,n}^{abc}, V_{m,n}^{0}, I_{m,n}^{0}, V_{m,n}^{PMU}, \delta V_{m,n}^{PMU}, I_{m,n}^{PMU}, \delta I_{m,n}^{PMU}]$ and 8 parameters $[V_{m,n}^{PMU}, \delta V_{m,n}^{PMU}, I_{m,n}^{PMU}, \delta I_{m,n}^{PMU}]$ respectively were created for all SCF (11 types) and No-Fault conditions.

$V_{m,n}^{abc} = [V_m^a \; V_n^a \; V_m^b \; V_n^b \; V_m^c \; V_n^c]$

$I_{m,n}^{abc} = [I_m^a \; I_n^a \; I_m^b \; I_n^b \; I_m^c \; I_n^c]$

$V_{m,n}^{0} = [V_m^0 \; V_n^0]$

$I_{m,n}^{0} = [I_m^0 \; I_n^0]$

$V_{m,n}^{PMU} = [V_m^{PMU} \; V_n^{PMU}]$

$\delta V_{m,n}^{PMU} = [\delta V_m^{PMU} \; \delta V_n^{PMU}]$

$I_{m,n}^{PMU} = [I_m^{PMU} \; I_n^{PMU}]$

$\delta I_{m,n}^{PMU} = [\delta I_m^{PMU} \; \delta I_n^{PMU}]$

$V_m^a, V_m^b, V_m^c =$ The peak positive voltage values observed for phases A, B, and C respectively during the fault condition are recorded and analyzed.

V_n^a, V_n^b, V_n^c = Peak negative values of voltage of phases A, B, and C respectively during the fault

I_m^a, I_m^b, I_m^c = Maximum positive values of current in phases A, B, and C respectively during the fault

I_n^a, I_n^b, I_n^c = Peak negative values of current in phases A, B, and C respectively during the fault

$V_m^0, \; V_n^0$ = Peak and minimum values of zero sequence voltage respectively during the fault

$I_m^0, \; I_n^0$ = Maximum and minimum values of zero sequence voltage respectively during the fault

V_m^{PMU}, V_n^{PMU} = Peak and minimum value of voltage of PMU during the fault

$\delta V_m^{PMU} \; \delta V_n^{PMU}$ = Maximum and minimum value of voltage angle of PMU during the fault

I_m^{PMU}, I_n^{PMU} = Maximum and minimum value of current of PMU from during the fault

$\delta I_m^{PMU} \; \delta I_n^{PMU}$ = Maximum and minimum value of current angle of PMU during the fault

Algorithm: Pseudocode for dataset creation

1. *Simulate 2-area, 4-Generator, and 11 Bus system as shown in Figure 1 in MATLAB Simulink environment*
2. $L = length\ of\ TL$
3. $r_f = fault\ resistance$
4. $x - \beta_{dataset} = [V_{m,n}^{PMU} \; \delta V_{m,n}^{PMU} \; I_{m,n}^{PMU} \; \delta I_{m,n}^{PMU}]$
5. $x - \alpha_{dataset} = [V_{m,n}^{abc} \; I_{m,n}^{abc} \; V_{m,n}^{0} \; I_{m,n}^{0} \; V_{m,n}^{PMU} \; \delta V_{m,n}^{PMU} \; I_{m,n}^{PMU} \; \delta I_{m,n}^{PMU}]$
6. $initiate\ x - \beta_{dataset}\ and\ x - \alpha_{dataset}\ as\ zero\ vector$
7. $counter = 0$
8. *simulate the fault*
9. $for1\ L = [1{:}2{:}110\ Km]$
10. $for2\ r_f = [0.001\ 0.005\ 0.01\ 0.05\ 0.1{:}0.2{:}20]$
11. $target = [A\ B\ C\ G]$*; replace the value of ABCG by 0 or 1 depending on type of fault as shown in Table 1*
12. $find\ the\ time\ duaration\ of\ fault$

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

13. *measure* $V_{m,n}^{abc}$ $I_{m,n}^{abc}$ $V_{m,n}^{0}$ $I_{m,n}^{0}$ $V_{m,n}^{PMU}$ $\delta V_{m,n}^{PMU}$ $I_{m,n}^{PMU}$ $\delta I_{m,n}^{PMU}$
14. *counter=counter+1*
15. *add measured values to above zero vector*
16. $end - for2$
17. $end - for1$
18. *Run above code for every fault*
19. *Tabulate all dataset to train and test the MLA*

3.2 Multi-class classification ML model

Fault classification of SCF on TL using $\alpha_{dataset}$ *and* $\beta_{dataset}$ and its comparison is the main objective of this work. To do so the 25 ML models have been trained and tested using training and testing datasets in the proportion of 80:20. Figure 2 describes the process flow of training and testing of MLA.

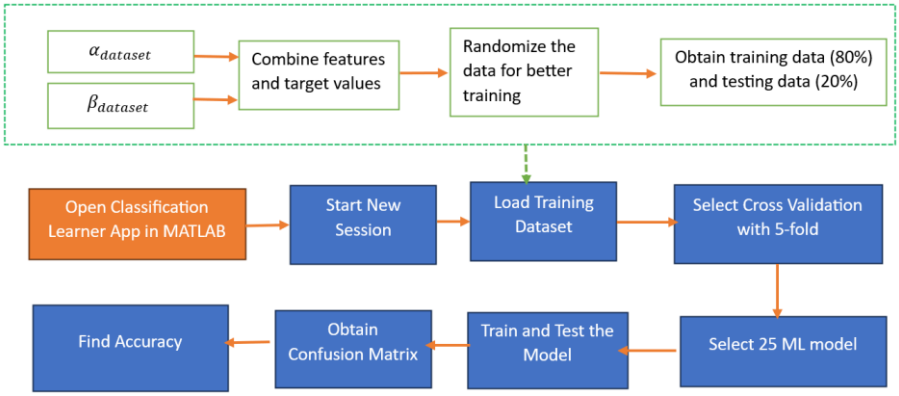

Figure 2: Process flow diagram to train and test the ML Model

4. RESULTS AND DISCUSSION

This section presents the results, organized into three distinct categories. The first category involves the simulation of various fault types using MATLAB. The second focuses on the classification of Single Circuit Faults (SCF) through the application of machine learning algorithms (MLA). The third category provides a comparative analysis of different models based on their performance when trained and tested using the proposed dataset.

4.1 Simulation of Various Types of Faults in MATLAB

To begin the fault analysis, a load flow study is conducted using the Newton-Raphson method. This approach is used to calculate key system parameters such as bus voltages, phase angles, generated power, and load demand under normal operating conditions—prior to the occurrence of any faults. The computed values are summarized in Table 2.

Faults may occur at any point of power system components such as transmission lines, transformers, generators, HVDC Convertors etc. The faults have been applied at different location with different combinations of fault resistance. The faults have been applied at 0.2 second and removed at 0.3 second. The three phase measurement block is used to measure the three-phase voltage & current, sequence analyser block is used to measure zero sequence voltage and current, and PMU (PLL based, positive sequence) block is used to measure voltage magnitude and its angle, current magnitude, and its angle. To workspace "structure with time" block is used to store data in the workspace. The three outputs of the PMU block return the magnitude, the phase (in degrees relative to the PLL phase), and the frequency. For the data collection purpose only the magnitude and phase outputs are used. The simulation results for different types of faults (11 SCF and 1 No Fault) are shown in Figure 3-14 for the setting of fault resistance as 0.001 ohm and length of transmission line as 20 km. Moreover, it can be observed from Figure 3-14 that the magnitude of fault current is maximum for three phase fault and minimum for single phase to ground fault. Such simulation results were obtained 4992 times for every SCF and 377 times for No Fault as shown in

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Table 3. Maximum and minimum values of voltage, current and angle are measured during the fault which is in between 0.2 to 0.3 second. It can be observed from the Table 3 that 55289 times simulation were run and 442312 data samples are collected for $\beta_{dataset}$ and 1326936 data samples are collected for $\beta_{dataset}$.

Table 2: Power flow analysis for Kundur 2-area system, NR Method (Converged in 4 iterations)

Bus No.	Voltage mag. (p.u.)	Voltage angle (deg)	Generation (MW)	Generation (MVAr)	Demand (MW)	Demand (MVAr)
1.	0.991	15.03	-	-	950.3372	-282.053
2.	1.003	-11.59	-	-	1776.536	-439.355
3. (Swing Bus)	1.000	0.00	703.2315	108.8019	2.800011	-8.20001
4. (PV Bus)	1.000	-26.71	700	72.69255	2.800011	-8.20001
5.	0.992	33.45	-	-	0	0
6.	0.994	7.02	-	-	0	0
7.	0.987	23.21	-	-	0	0
8.	0.993	-3.43	-	-	0	0
9. (PV Bus)	1.000	-16.08	700	82.00615	2.800011	-8.20001
10. (PV Bus)	1.000	10.17	719	72.05932	2.800011	-8.20001
11.	0.987	12.60	-	-	0.056534	0
Total			2822.232	335.5599	2738.129	-754.208

Table 3: Number of simulations data

Fault Type				Simulations run	No. of observations		Total No of data samples	
A	B	C	G		$\beta_{dataset}$ (8 features)	$\alpha_{dataset}$ (24 features)	$\beta_{dataset}$	$\alpha_{dataset}$
0	0	0	0	377	377	377	3016	9048
1	0	0	1	4992	4992	4992	39936	119808
0	1	0	1	4992	4992	4992	39936	119808
0	0	1	1	4992	4992	4992	39936	119808
1	1	0	1	4992	4992	4992	39936	119808
0	1	1	1	4992	4992	4992	39936	119808
1	0	1	1	4992	4992	4992	39936	119808
1	1	0	0	4992	4992	4992	39936	119808
0	1	1	0	4992	4992	4992	39936	119808
1	0	1	0	4992	4992	4992	39936	119808
1	1	1	0	4992	4992	4992	39936	119808
1	1	1	1	4992	4992	4992	39936	119808
Total				**55289**	**55289**	**55289**	**442312**	**1326936**

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

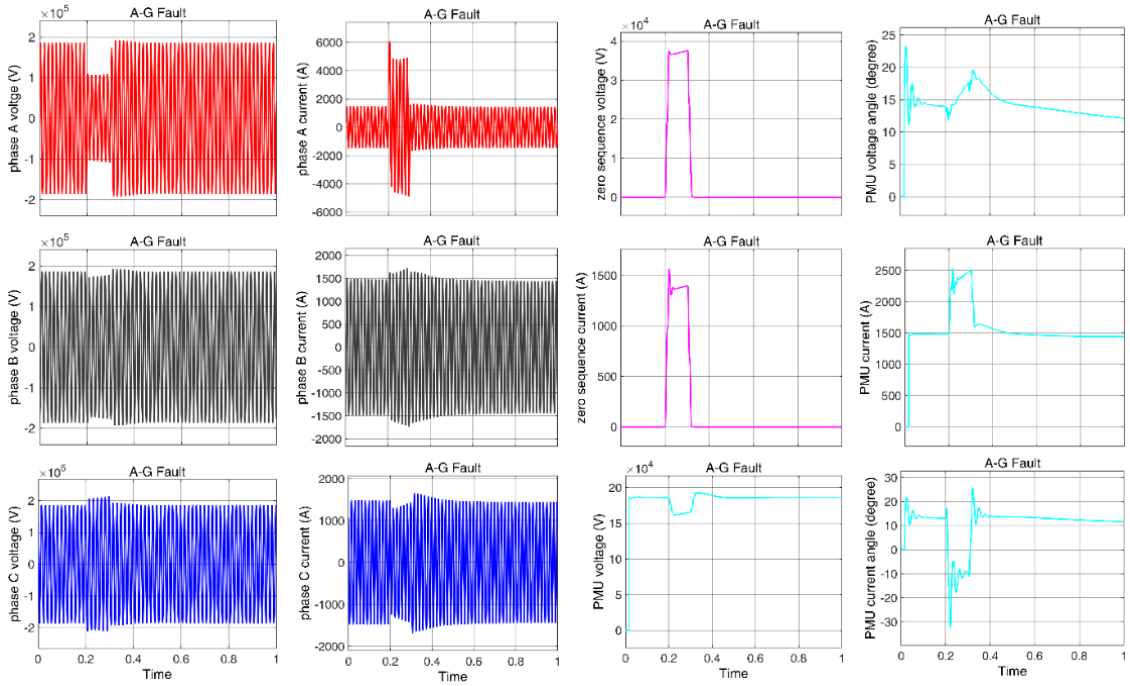

Figure 3: Phase A to Ground Fault

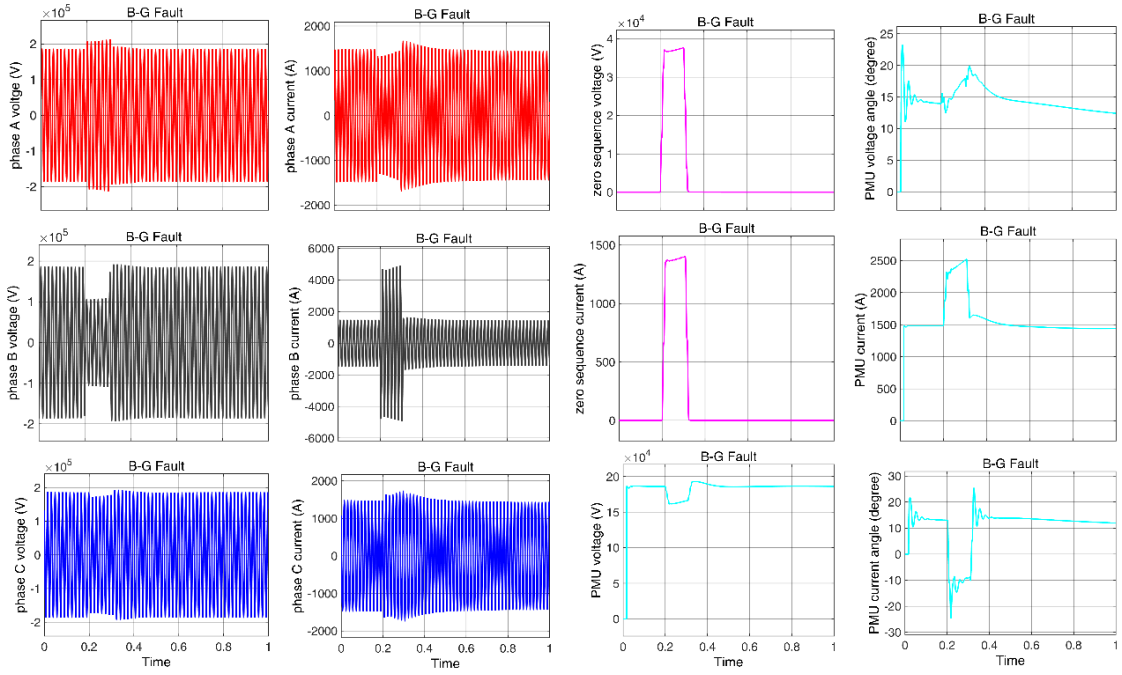

Figure 4: Phase B to Ground Fault

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

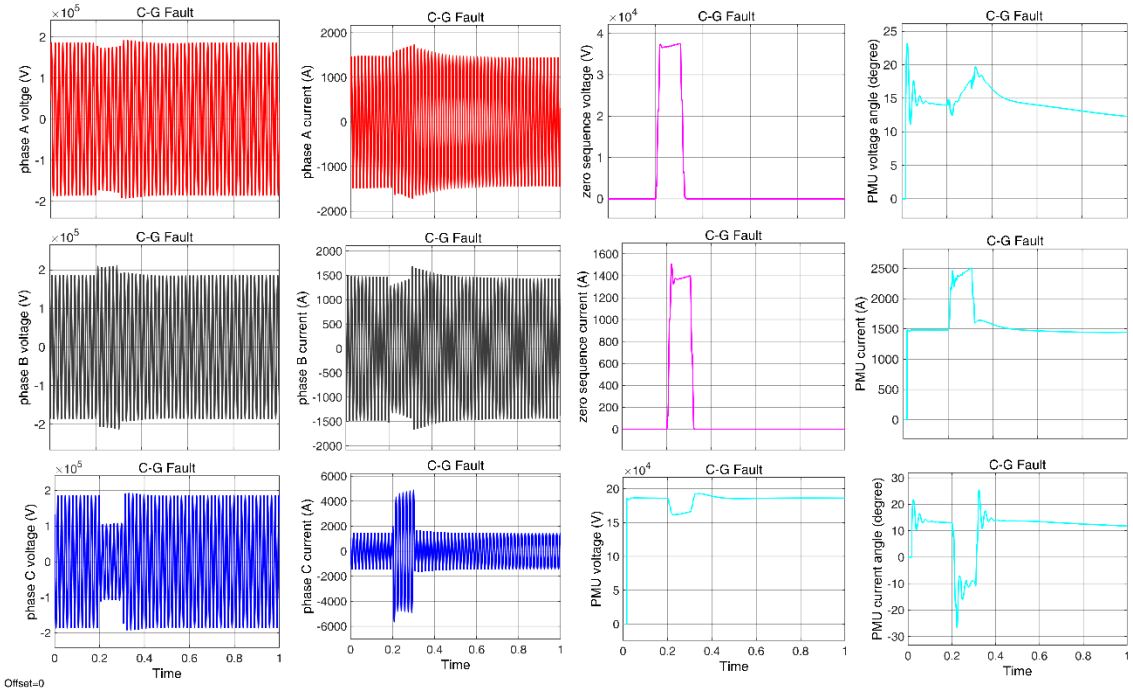

Figure 5: Phase C to Ground Fault

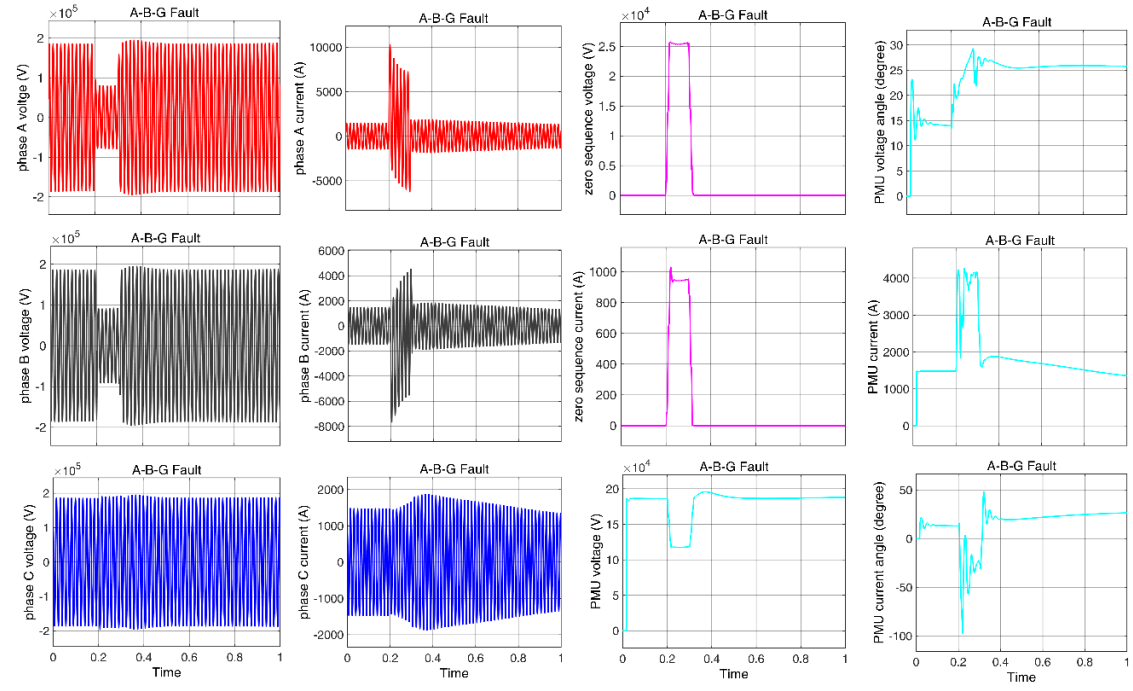

Figure 6: Phase A-phase B to Ground Fault

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

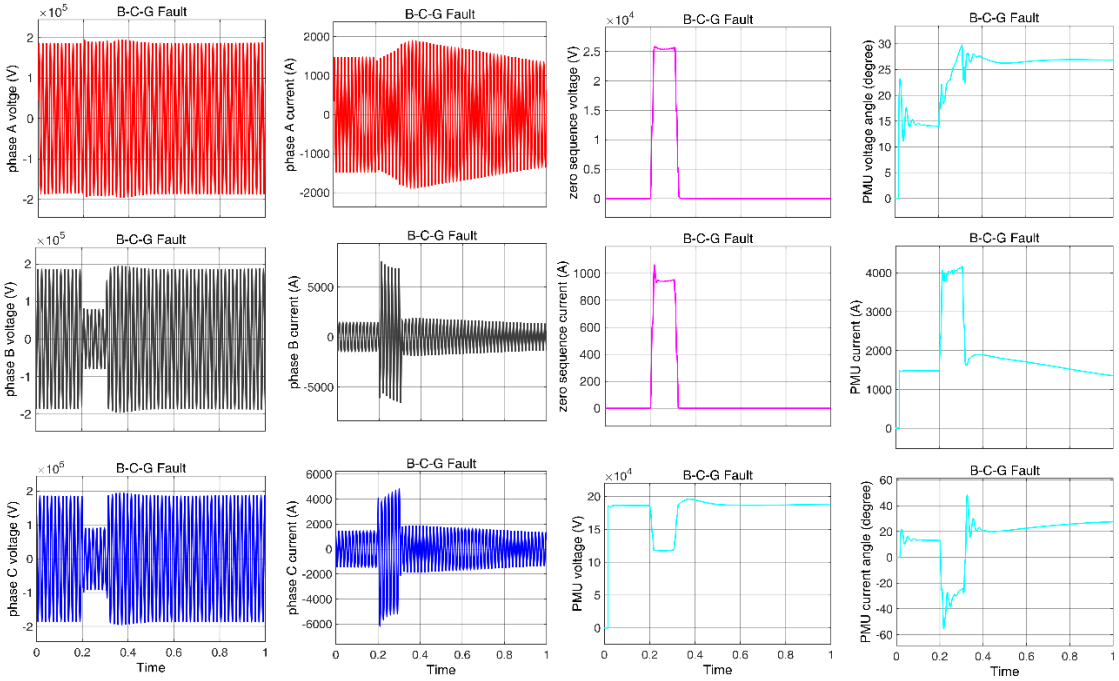

Figure 7: Phase B-phase C to Ground Fault

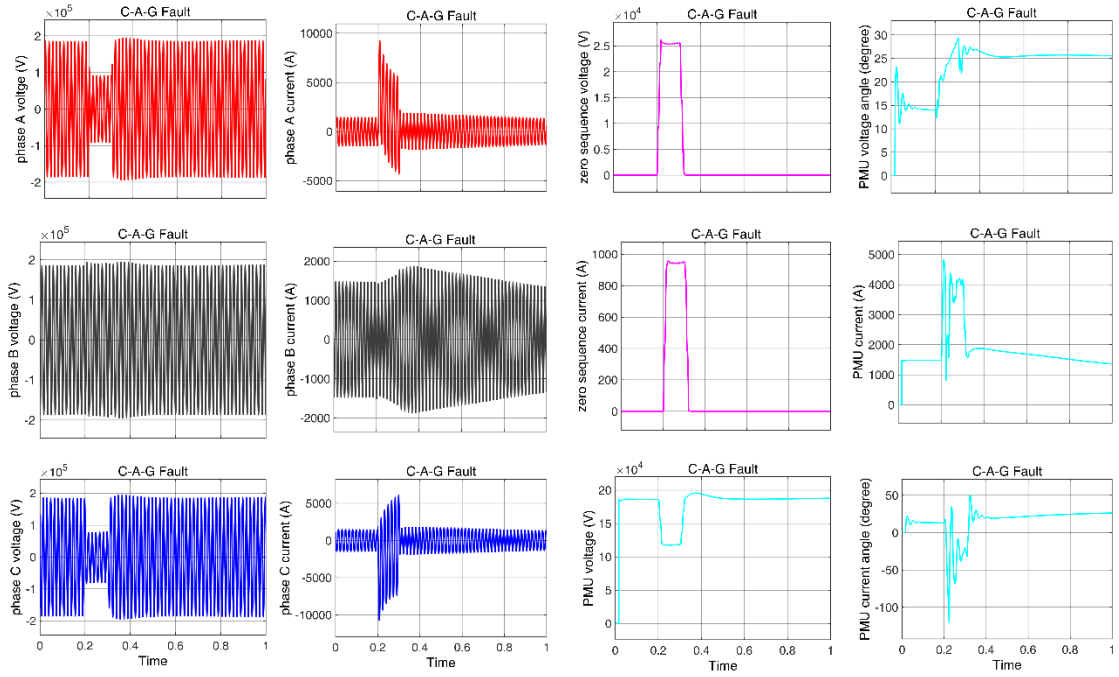

Figure 8: Phase C-phase A to Ground Fault

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

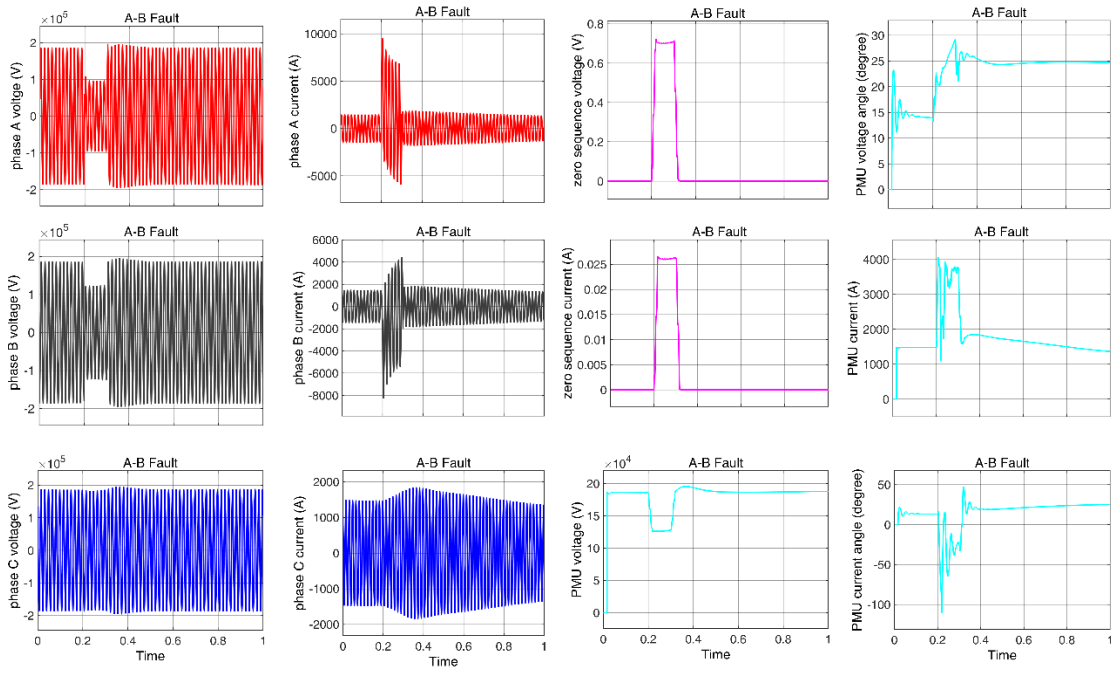

Figure 9: Phase A-phase B Fault

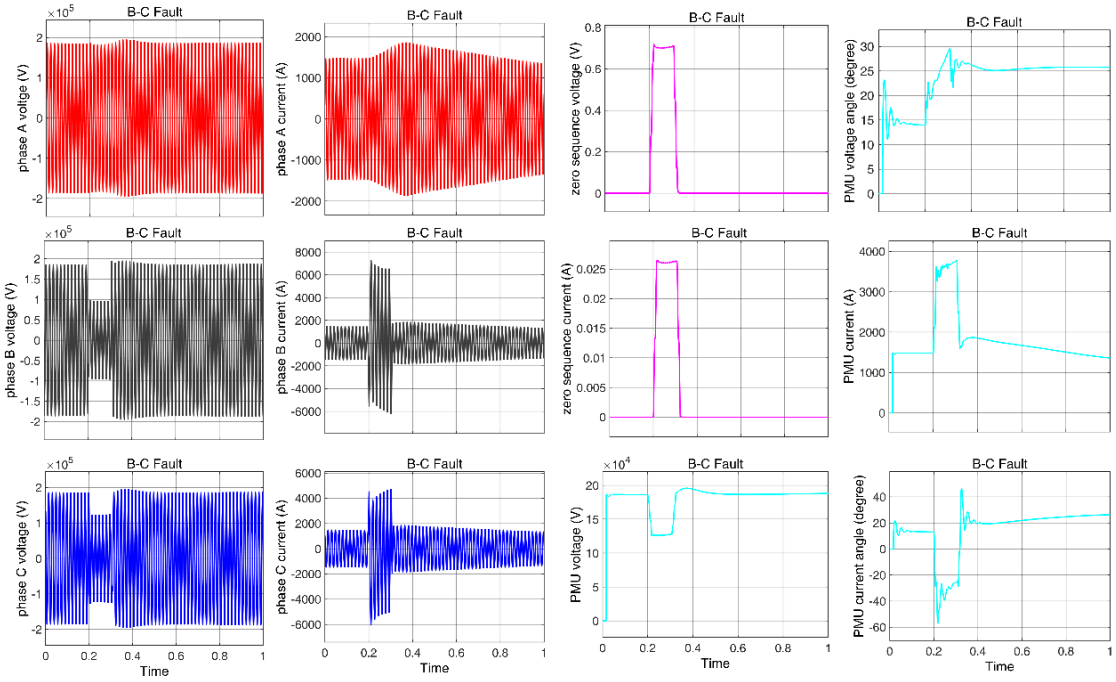

Figure 10: Phase B-phase C Fault

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

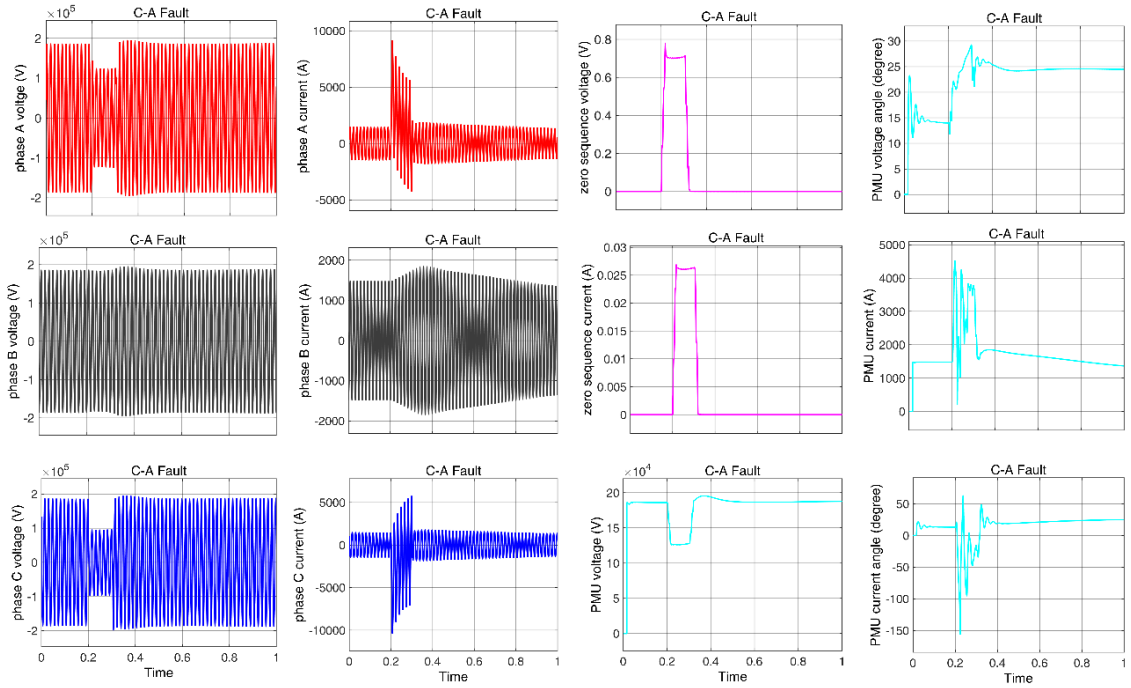

Figure 11: Phase C-phase A Fault

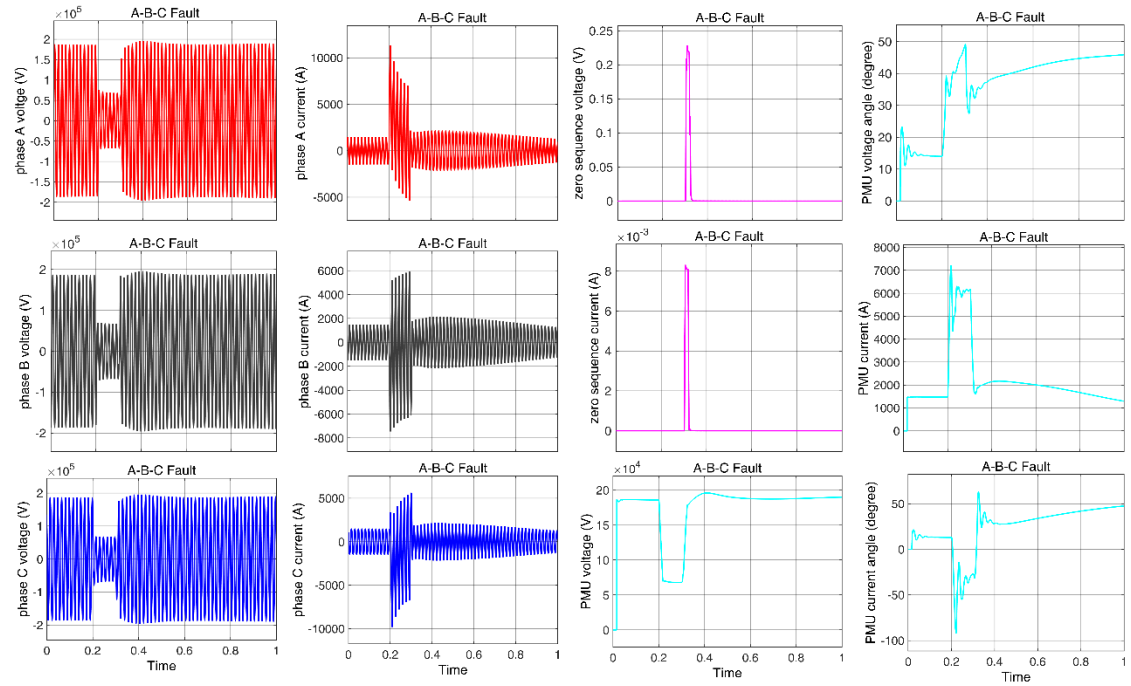

Figure 12: Phase A-phase B-phase C Fault

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

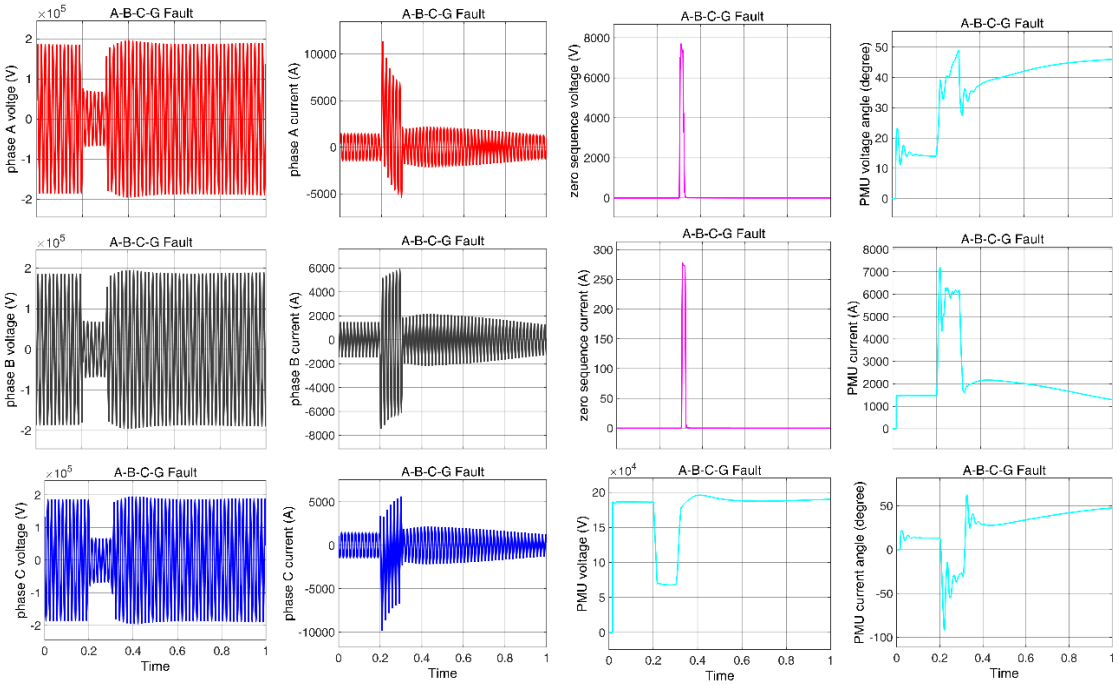

Figure 13: Phase A-phase B-phase C to Ground fault

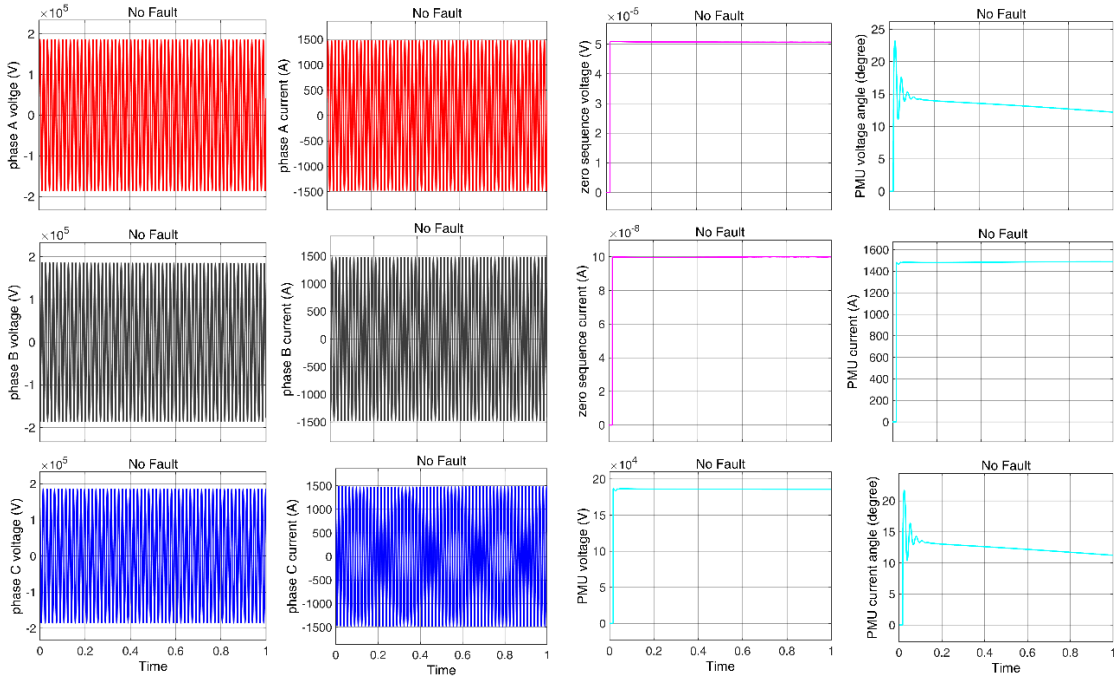

Figure 14: No Fault

4.1 Performance Indicators

During the analysis phase, several evaluation metrics are employed to measure the performance of the model trained on the proposed dataset. The confusion matrix, presented in Table 4, serves as a key tool for this assessment. In the matrix, True Positives (TP) refer to correctly identified positive instances, while True Negatives

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

(TN) are correctly identified negative instances. False Positives (FP) are negative instances incorrectly classified as positive, and False Negatives (FN) are positive instances incorrectly labeled as negative. To further evaluate the model, metrics such as Accuracy, Precision, Recall, and F1-Score are utilized, providing a comprehensive understanding of its effectiveness during training.

Table 4: Confusion matrix

Confusion Matrix		Predicted Values	
		Positive	Negative
Actual values	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$	(21)
$Precision = \frac{TP}{TP + FP}$	(22)
$Recall = \frac{TP}{TP + FN}$	(23)
$F1 - Score = \frac{2TP}{2TP + FP + FN}$	(24)

4.2 Test Results

In this paper, the effectiveness of fault classification was verified using the dataset as listed in Table 3 for training and testing the various MLA. The confusion matrix of both dataset for Fine Tree MLA is shown in Figure 15 and 16. The confusion matrix of other ML model is also found and the performance metrics are also calculated. Figure 15 and 16 represent the confusion matrix of Fine Tree when tested and trained by $\beta_{dataset}$ and $\alpha_{dataset}$ for multiclass fault classification. The values of TP, FP, FN, TN are calculated from confusion matrix and listed in Table 5. When the model is trained through $\beta_{dataset}$, it is observed that the values of FP and FN for NF, AG, BG, CA, and CG faults are zero respectively which indicate that these types of faults are accurately predicted. The other remaining type of faults are not predicted accurately as the values of FP and FN are not zero. For example, if the case of AB fault classification is considered then it is observed that out of 3986 faults only 3832 faults classified correctly as AB fault, 14 faults misclassified as ABG fault and 140 faults misclassified as CAG fault. The value of FP is 124 which indicates that these were not the AB fault, but mistakenly predicted as AB fault. The value of FN is154 which indicates that these were AB fault but mistakenly predicted as other types of faults. The value of TN is 40121 which indicates that these were not the AB fault and correctly classified as not the AB fault. When the same model is trained through the proposed dataset then it is observed that the values of FP and FN for all type of faults are zero, which indicate that all type of faults are accurately classified and none of the faults are misclassified.

Figure 17 represents the accuracy of every type of faults when the fine tree model is trained and tested through both the dataset. 100% accuracy is achieved only for NF, A-G, B-G, C-A and C-G fault and for other type of faults is less than 100%. When the same model is trained through proposed dataset the accuracy of every fault classification is 100%. Figure 18, Figure 19, and Figure 20 represent the Precision, Reall, and F1-Score comparison of every types of faults for Fine Tree Model only. Similar procedures adopted for every SCF and 25 ML algorithm to find the accuracy comparison as shown in Figure 21. It can be observed that 21 ML algorithm gives 100% accuracy for $\alpha_{dataset}$.

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

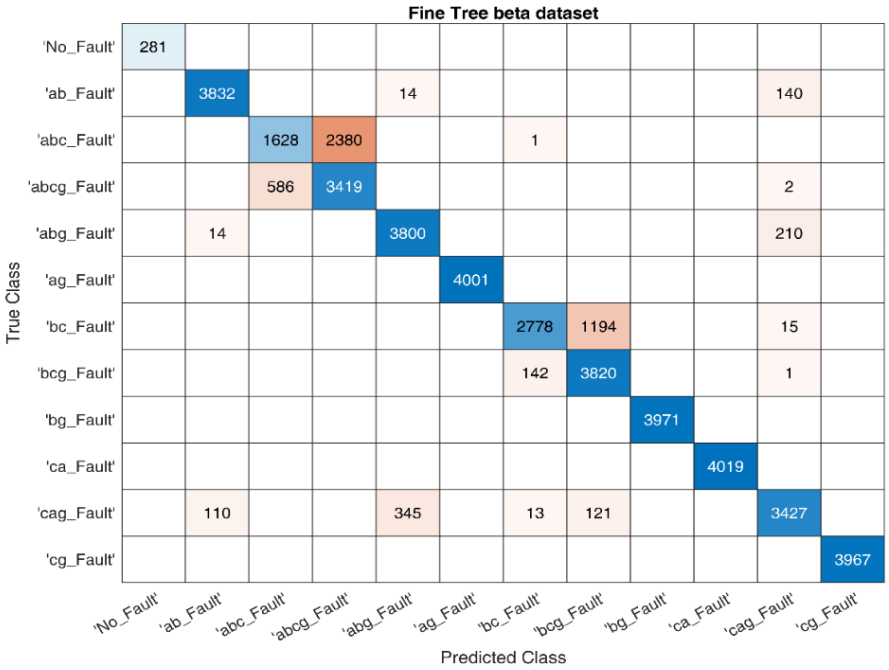

Figure 15: Confusion Matrix for validation of Fine Tree using $\beta_{dataset}$

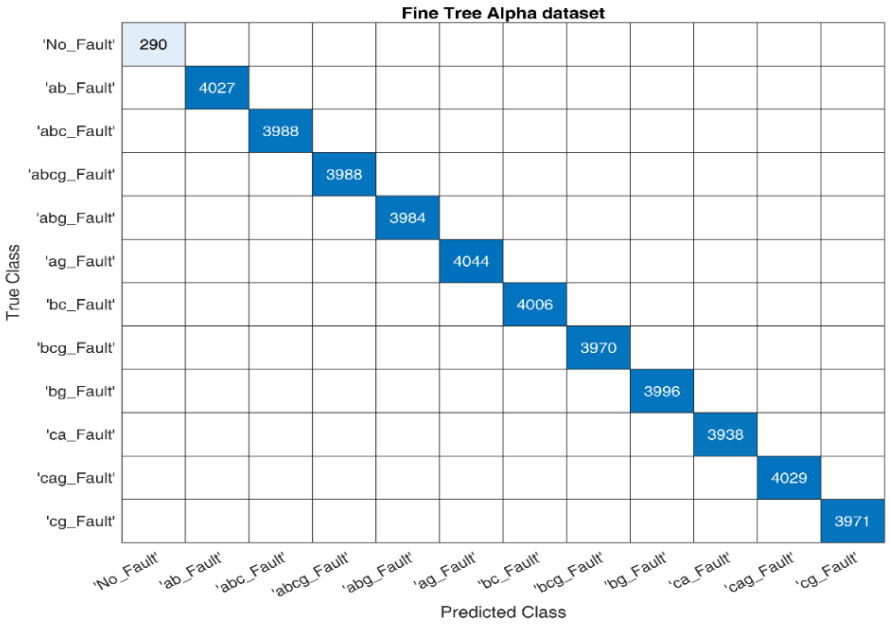

Figure 16: Confusion Matrix for validation of Fine Tree using $\alpha_{dataset}$

Table 5: TP, FP, FN, TN of every fault for Fine Tree Model

Fault type	**Performance indicators**							
	Beta-dataset				**Proposed Alpha-dataset**			
	TP	**FP**	**FN**	**TN**	**TP**	**FN**	**FP**	**TN**
NF	281	0	0	43950	290	0	0	43941
AB	3832	124	154	40121	4027	0	0	40204

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

ABC	1628	586	2381	39636	3988	0	0	40243
ABCG	3419	2380	588	37844	3988	0	0	40243
ABG	3800	359	224	39848	3984	0	0	40247
AG	4001	0	0	40230	4044	0	0	40187
BC	2778	156	1209	40088	4006	0	0	40225
BCG	3820	1315	143	38953	3970	0	0	40261
BG	3971	0	0	40260	3996	0	0	40235
CA	4019	0	0	40212	3938	0	0	40293
CAG	3427	368	589	39847	4029	0	0	40202
CG	3967	0	0	40264	3971	0	0	40260

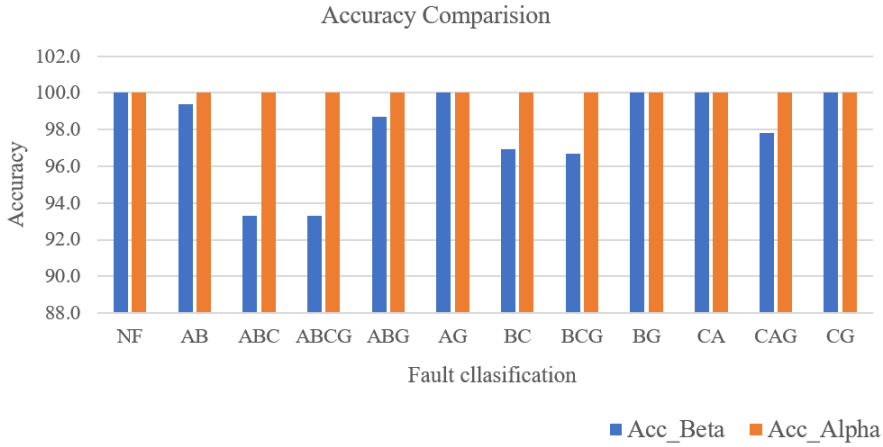

Figure 17: Accuracy comparison of every fault

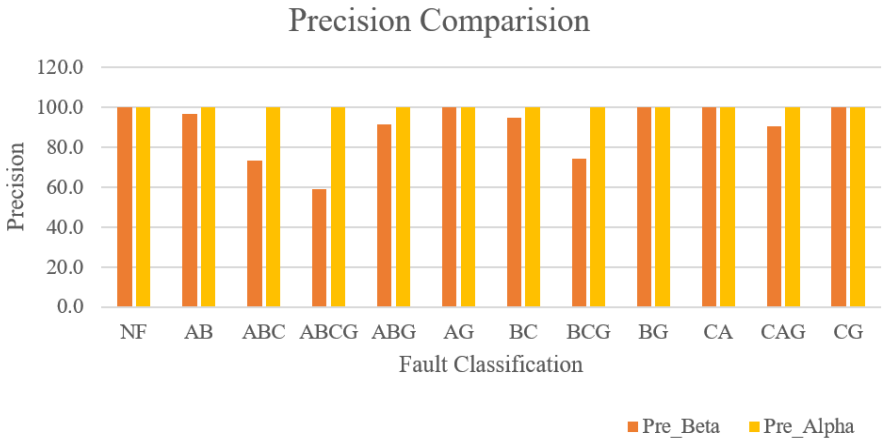

Figure 18: Precision comparison of every fault

Journal of Information Systems Engineering and Management
2024, 9(4s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

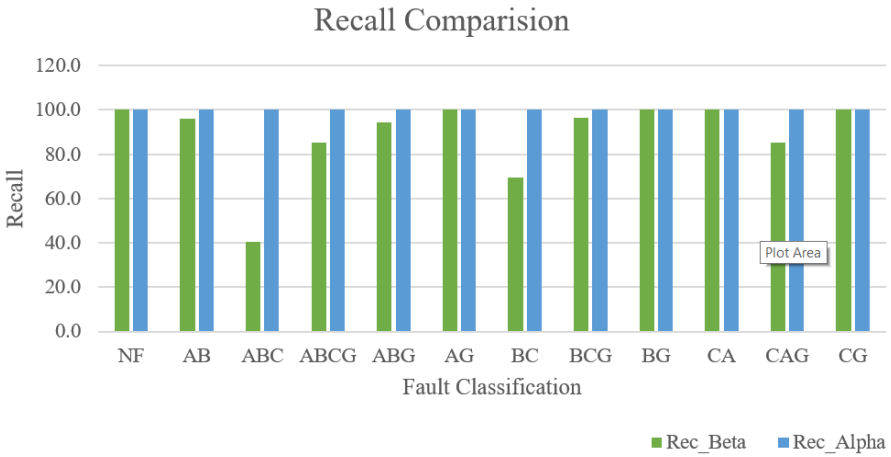

Figure 19: Recall comparison of every fault

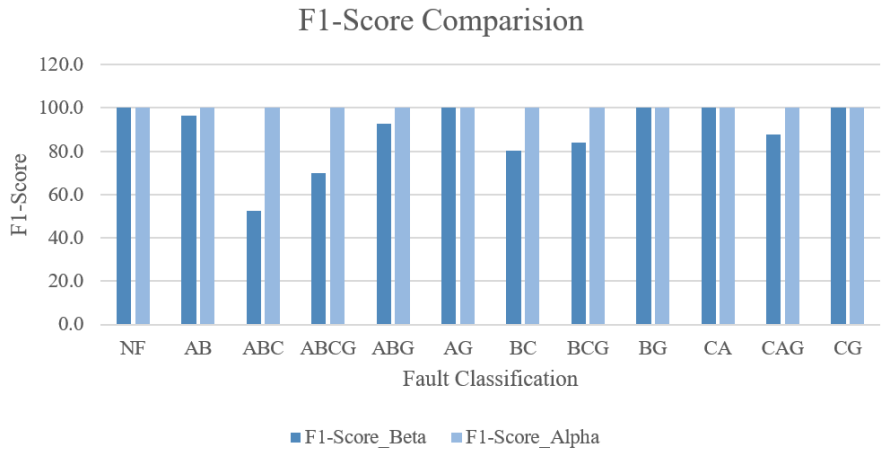

Figure 20: F1-Score comparison of every fault

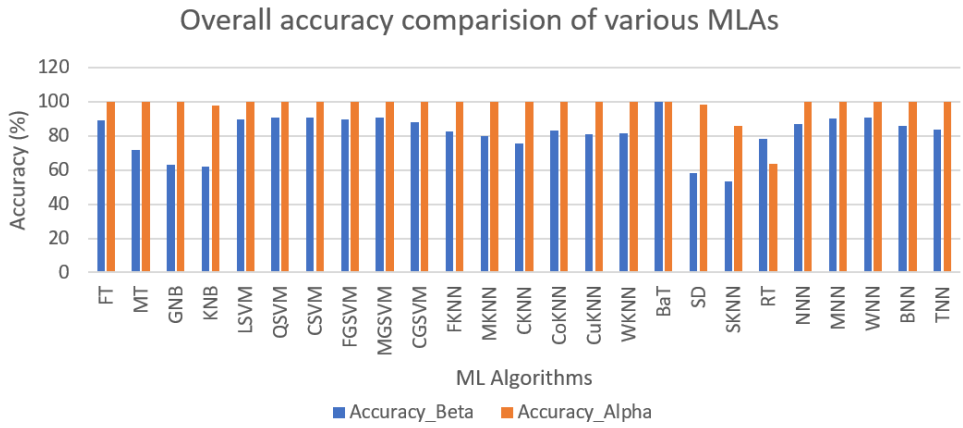

Figure 21: Accuracy Comparison of various ML Algorithm

Table 6 represents the accuracy comparison of SCF on TL between the previous work and the present work. The 21 ML model gives 100% accuracy due to good quality and large dataset.

Table 6: Accuracy Comparison with Previous Work

Ref.	Year	Methodology	No. of Data Samples	Accuracy (in %)
[12]	2021	Long Short-Term Memory	4380	96.77
[11]	2021	Graph Convolutional Neural Network	11900	98.28

Research Article

[14]	2022	Anamoly based One-Class SVM	8712	94.42
		Anamoly based One-Class SVM PCA	8712	95.78
[13]	2022	Weighted Extreme Learning Machine	22451	97
[17]	2023	KNN	12000	93.2
[16]	2023	CNN	20000	99.73
[21]	2024	Hybrid CNN-LSTM	62501	99.86
This work	-	FT, MT, GNB, KNB, LSVM, QSVM, CSVM, FGSVM, MGSVM, CGSVM, FKNN, MKNN, CKNN, CoKNN, CuKNN, WKNN, BaT, SD, SKNN, NNN, MNN, WNN, BNN, TNN	1326936 (Alpha data set),	100 % from 21 ML Models

5. CONCLUSION

This paper proposed a novel scheme to improve the accuracy for SCF classification using Classification Learner APP in MATLAB by enhancing the quality and quantity of the dataset. The various SCF characteristics were observed on Kundur Two-Area System at different locations for various combination of fault resistance. 1326936 data samples were collected by measuring maximum and minimum values of three phase voltage, current, zero sequence voltage and current, PMU based voltage, current and its angle during the fault and no-fault conditions. The accuracy of 21 ML models is found to increase to 100% when 25 available ML models are trained and tested using such huge data samples.

REFERENCES

[1] P. Kundur, *Power System Stability and Control*. McGraw-Hill Pvt. Limited, 1994.
[2] M. Wadi and W. Elmasry, "Modelling of wind energy potential in Marmara 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 Generation, Transmission & Distribution*, vol. 13, no. 5, pp. 724–732, 2019.
[3] W. Wang, H. Yin, C. Chen, A. Till, W. Yao, X. Deng, and Y. Liu, "Frequency disturbance event detection based on synchro phasors and deep learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3593–3605, 2020
[4] Y. Xi, W. Zhang, F. Zhou, X. Tang, Z. Li, X. Zeng, and P. Zhang, "Transmission line fault detection and classification based on sa-mobilenetv3," *Energy Reports*, vol. 9, pp. 955–968, 2023.
[5] M. Jamil, S. K. Sharma, and R. Singh, "Fault detection and classification in electrical power transmission system using artificial neural network," *Springer Plus*, vol. 4, pp. 1–13, 2015.
[6] A. S. Alhanaf, M. Farhadi, and H. H. Balik, "Fault detection and classification in ring power system with dg penetration using hybrid cnn-lstm," *IEEE Access*, 2024.
[7] Z. Wang, "The application and optimization of machine learning in big data analysis," *Computer Life*, vol. 12, no. 1, pp. 8–11, 2024.
[8] Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: a big data-ai integration perspective," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, no. 4, pp. 1328–1347, 2019.
[9] Y. Gong, G. Liu, Y. Xue, R. Li, and L. Meng, "A survey on dataset quality in machine learning," *Information and Software Technology*, vol. 162, p. 107268, 2023.
[10] D. Zha, Z. P. Bhat, K.-H. Lai, F. Yang, Z. Jiang, S. Zhong, and X. Hu, "Data-centric artificial intelligence: A survey," *arXiv preprint arXiv:2303.10158*, 2023.
[11] H. Tong, R. C. Qiu, D. Zhang, H. Yang, Q. Ding, and X. Shi, "Detection and classification of transmission line transient faults based on graph convolutional neural network," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 3, pp. 456–471, 2021.
[12] N. Nikhil, A. Kumar, P. Amrit, *et al.*, "Fault detection and classification in power system using machine learning," in *2021 2nd International Conference on Smart Electronics and Communication (ICOSEC)*, pp. 1801–1806, IEEE, 2021.
[13] A. Harish, A. Prince, and M. Jayan, "Fault detection and classification for wide area backup protection of power transmission lines using weighted extreme learning machine," *IEEE Access*, vol. 10, pp. 82407– 82417, 2022.

Research Article

[14] W. Elmasry and M. Wadi, "Enhanced anomaly-based fault detection system in electrical power grids," *International Transactions on Electrical Energy Systems*, vol. 2022, no. 1, p. 1870136, 2022.
[15] J. G. Fornas, E. H. Jaraba, H. Bludszuweit, D. C. Garcia, and A. L. Estopinan, "Modelling and simulation of time domain reflectometry signals on a real network for use in fault classification and location," *IEEE Access*, vol. 11, pp. 23596–23619, 2023.
[16] B. Roy, S. Adhikari, S. Datta, K. J. Devi, A. D. Devi, F. Alsaif, S. Alsulamy, and T. S. Ustun, "Deep learning-based relay for online fault detection, classification, and fault location in a grid-connected microgrid," *IEEE Access*, vol. 11, pp. 62674–62696,2023.
[17] N. K. Abed, F. T. Abed, H. F. Al-Yasriy, and H. T. S. AL-Rikabi, "Detection of power transmission lines faults based on voltages and currents values using k-nearest neighbors," *International Journal of Power Electronics and Drive Systems (IJPEDS)*, vol. 14, no. 2, pp. 1033– 1043, 2023.
[18] V. R. Jyothula, Y. Purohit, and M. Paraye, "Fault detection in power transmission line using machine learning techniques," in *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1–5, IEEE, 2023.
[19] F. Bouaziz, A. Masmoudi, A. Abdelkafi, and L. Krichen, "Applying machine learning algorithms for fault detection and classification in transmission lines," in *2023 IEEE 11th International Conference on Systems and Control (ICSC)*, pp. 207–212, IEEE, 2023.
[20] M. Najafzadeh, J. Pouladi, A. Daghigh, J. Beiza, and T. Abedinzade, "Fault detection, classification and localization along the power grid line using optimized machine learning algorithms," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, p. 49, 2024.
[21] S. Kanwal and S. Jiriwibhakorn, "Advanced fault detection, classification, and localization in transmission lines: A comparative study of anfis, neural networks, and hybrid methods," *IEEE Access*, 2024.
[22] A. S. Meliopoulos, G. J. Cokkinides, P. Myrda, Y. Liu, R. Fan, L. Sun, R. Huang, and Z. Tan, "Dynamic state estimation-based protection: Status and promise," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 320–330, 2016.
[23] "IEEE Guide for Synchronous Generator Modelling Practices and Applications in Power System Stability Analyses," in *IEEE Std 1110-2002 (Revision of IEEE Std 1110-1991)*, pp.1-80, 12 Nov. 2003, doi: 10.1109/IEEESTD.2003.94408.
[24] A. K. Singh and B. C. Pal, "Decentralized Nonlinear Control for Power Systems Using Normal Forms and Detailed Models," in *IEEE Transactions on Power Systems*, vol. 33, no. 2, pp. 1160-1172, March 2018, doi: 10.1109/TPWRS.2017.2724022.