Journal of Operation and Automation in Power Engineering

vol. 12, no. 3, Aug. 2024, Pages: 206-214

http://joape.uma.ac.ir

Data Mining and SVM Based Fault Diagnostic Analysis in Modern Power System Using Time and Frequency Series Parameters Calculated From Full-Cycle Moving Window

*P. Venkata**, *V. Pandya, A.V. Sant*

Electrical Engineering Department, Pandit Deendayal Energy University, Gandhinagar, Gujarat, India.

Abstract— *This paper proposes a complete diagnostic analysis of faults in a typical modern power system's transmission line using the support vector machine (SVM) with time-series parameters and frequency series parameters as features. The training and testing data of the proposed method are collected by simulating all types of faults with all possible variations on a transmission line (TL) in the IEEE-9 bus system using the PSCAD/EMTDC software. While simulating one type of fault, fault resistances and fault inception angles are also varied to account for the various behaviours of the fault. The three-phase instantaneous currents and voltages on both sides of TL are recorded at 32 samples per cycle. A thirty-two sample moving window is used to compute time-series and frequency-series parameters applied as features to the SVM. Ten-fold cross-validation is used to evaluate the performance of the proposed algorithm with evaluation metrics such as accuracy, precision, recall and F1 score. Features generation, training and testing of the proposed method, and performance comparison are done using PYTHON software. The proposed method has achieved an average accuracy of 99.996%, even in the most contaminated environment of 30 dB noise. Compared with the performance of the other popular machine learning algorithms, the proposed method has achieved more accuracy. The performance of the proposed method is also tested with different noise levels, which account for the measurement errors of 30 dB, 35 dB and 40 dB.*

Keywords— Data Mining, Fault classification, FFT, Machine Learning, SVM, Transmission line.

1. INTRODUCTION

Transmission lines (TL), an integral part of the power system, improve the power reliability by interconnecting different parts of the electrical grid. This interconnection helps export power to the generation deficient regions from the excess power region. In addition to the power system interconnections, TL permits cost-effective electrical power dispatch from generating stations –located on cheaply available land on the periphery of a city– to load canters [1, 2]. These TLs cover different geographic terrains such as forests, rivers, deserts etc. Unfortunately, because TLs are exposed to the atmosphere, they are more susceptible to various faults [3, 4]. Transmission line faults and subsequent transmission line shutdowns impact the reliability and stability of power systems [5].

A transmission line fault is a problem caused by an unintentional path of power flow, which results in an anomalous flow of current and, if not addressed, burns the line [13]. The fire created by line burning could spread to other elements of the power system, posing a considerable risk of death. Faults cause dangerous and insecure power system operations and physical damage to line infrastructure [14]. Based on how conductors have reacted to external disturbances like breaking a line or a tree falling on the line, faults in the system can be categorized as open conductor faults, series faults, or shunt faults [15]. Proper protection of the transmission lines ensures the future expansion of the power system [16].

Received: 16 May 2022
Revised: 16 Nov. 2022
Accepted: 05 Dec. 2022
* Corresponding author:
E-mail: pavan.venkata@sot.pdpu.ac.in (P. Venkata)
DOI: 10.22098/JOAPE.2023.10819.1789
Research Paper

1.1. Literature Review

Fault classification is necessary to estimate the amount of repair work at the fault site [17]. Fault classification results are used as inputs to the fault location identification algorithms and as decision points for the circuit breaker's single pole or double pole operation to maintain the power system's reliability, security, and stability [18]. Because of the huge importance of fault classification in transmission lines, much research is being carried out. The methods used for the transmission line classification can be broadly classified into two categories. Namely,

1) Analytical Methods or conventional methods
2) Intelligent Methods

The analytical methods can be further subdivided into three main categories 1. Impedance-based methods 2. Travelling wave-based Methods 3. Modal transformation methods

In the modal transformation methods [19–21], three-phase quantities of voltages and currents in the stationary reference frame represented by a, b and c axes are transformed into another stationary reference frame α,β, and 0 axes with the help of Modal transformation such as Clarke transformation (CT). Based on this transformation, fault types were characterized by describing the relationships between phase quantities and modal component

The entire classification is completed in this analytical method in a single step. In analytical methods, with a solid mathematical background, the computation time and complexity of the solution depend on the system's size. However, the second classification method, i. e. intelligent, mainly uses two fault classification stages. The first stage is feature generation from the front-end data. The second stage is applying the features to different classification algorithms. Different signal processing techniques are used in the first stage, such as Discrete Fourier Transform (DFT), S-Transform (ST), Wavelet Transform (WT), Hilbert Huang Transform (HHT), Principal Component Analysis (PCA) and Empirical Mode Decomposition (EMD). The second stage of fault classification may consist of artificial intelligent methods such as Artificial Neural Network (ANN), Fuzzy Inference Systems (FIS), Support Vector Machine (SVM), Decision Tree (DT), k-nearest neighbors (KNN), Random Forest (RF).

Strong pattern recognition capabilities make Artificial Neural Network (ANN) a good transmission line fault classification method. Researchers and scientists have used different types of ANN methods for classification. Radial basis function neural network is used in [22], [23], multi neural network is used in [24]. Authors in [25] used Chebyshev neural networks for the protection of ISPST [25], and authors in [26] used data mining and SVM-based technique to protect the microgrid [26]. Similarly, authors in [27] used data mining techniques to improve vehicle charging on imbalance index in unbalanced distribution networks [27]. The drawback of the ANN is that it takes a long time for training, and also overfitting the trained data will give wrong results to data out of the trained data [22].

With the help of only current samples, fuzzy logic-based fault classification is done in [28]. Data mining-based fuzzy logic is used in [29] to identify the cause of fault in unbalanced distributed lines. Some authors have used the Adaptive Network-Based Fuzzy Inference System [30] for fault classification. Authors in [31] have used the ANFIS algorithm to classify the faults in the transmission line. However, the fault classification metrics are not properly mentioned. Authors in [32] used a fuzzy neuro approach [33] to classify the faults with the help of symmetrical components along with the line currents. Moreover, several studies have ignored the effect of noise in training or test data.

Authors in [34] used a cross-correlation aided fuzzy-based scheme for fault classification. This method has a Low burden for calculating features, is less immune to noise, and only utilizes single-end parameters. However, fewer fault and non-fault scenarios are used in the method [34]. Authors in [35] used ratio-based and probabilistic neural networks with the help of Principal component analysis for the transmission line fault classification. This method requires the data of only one end of the line, but the transmission line with single-end feeding is considered in this work [35]. Chen et al. used novel integrated feature extraction for the transmission line fault classification. This method is used for both fault location and classification [36]. Authors in [37] used an unsupervised feature learning and convolutional spares auto encoder-based fault classification of the transmission line. The effect of sampling frequency is considered in this work [37].

The authors in [38-40], [17], and [41] used the Support vector machines algorithm [42] for classifying the faults in a transmission line or distribution line. With the help of SVM, wavelet-based fault classification is performed in [17]. The average classification accuracy is around 96%, and no proper procedure is available to select the mother wavelet. A radial basis neural network is used along with SVM in [39], where the average classification is around 98%. However, the data set used for testing the algorithm is very less. The method used in [43] is very complex, and the classification accuracy is around 95%. The accuracy of the classification in [40] is around 95%.

1.2. Significant Contribution and paper organization

As the faults happen on different phases of the transmission line, the corresponding phase currents and phase voltages get affected severely. Different faults will result in different current and voltage value variations. Fault resistance, fault inception angle, and fault location will contribute to this variation.

- All samples in one full cycle, i.e. thirty-two samples, are gathered from the instant of the fault, the three-phase currents and three-phase voltages on both sides of the transmission line to calculate time series parameters and frequency series parameters up to the seventh harmonic using FFT.
- The time-series and frequency-series parameters are applied as features to the SVM algorithm.
- To the original data, the noise of various signal-to-noise ratio (SNR) levels such as 30dB, 35dB, and 40dB is introduced. Next, the algorithm's robustness is tested by calculating the time and frequency series parameters.
- The performance of the proposed method is compared with that of the other popular Machine Learning algorithms such as Decision Tree, Random Forest, K-Nearest neighbours, Adaptive Boosting Classifier and Gaussian Naive Bayes.

The rest of this paper is organized as follows. Section 2 explains the basics and formulation of the Support Vector Machine Algorithm. Section 3 explains how the application of the SVM can solve the present problem. Section 4 describes the simulation of the IEEE9 bus system in PSCAD / EMTDC and the data generation process. Section 5 talks about validating the generated data. Section 6 provides the results and discussions. Finally, Section 7 talks about the conclusions and recommendations.

2. Mathematical Description of SVM

SVM is a classification algorithm that provides a clear and possibly large boundary between the two classes [42]. SVM can be applied to distinguish between two classes and among multiple classes. SVM is an algorithm which bifurcates two classes. However, the same theory can be applied to multiple classes by converting the multiple classes into two classes - one class is all the instances that belong to one of the classes in the given multiple classes. The other class is all the instances that do not belong to the above-said class. This process can be repeated x-1 times, where x is the number of unique classes in a given data [42].

Let there be two classes, namely C_1 and C_2, in the complete training data of a classification problem. Let "n" be the number of features of a class and "m" be the number of instances of those two classes in the given problem. Out of these m instances, let us assume that for k instances, the data belong to C_1, and for the remaining (m-k) instances, the data belong to C_2. Here,

$$0 \le m \le \infty \ and\ m\ \in\ \mathcal{Z}\ (\ A\ set\ of\ Integers) \tag{1}$$

$$0 \le n \le \infty \ and\ n\ \in\ \mathcal{Z}\ (\ A\ set\ of\ Integers) \tag{2}$$

$$0 \le k \le n \ and\ k \in\ \mathcal{Z}\ (\ A\ set\ of\ Integers) \tag{3}$$

The total number of occurrences and features can be represented in a matrix of the type $[X]_{mXn}$, where m represents the number of rows (representing instances), and n represents the number of columns (representing features). The following equation can be used to solve the given categorization problem.

$$g(x) = w^t X + b,\ where\ w = weights\ and\ b = bias \tag{4}$$

$$g(X_1) = w^t X_1 + b > 0,\ when\ X_1\ is\ on\ RHS\ of\ the\ line \tag{5}$$

$$g(X_1) = w^t X_1 + b < 0,\ when\ X_1\ is\ on\ LHS\ of\ the\ line \tag{6}$$

The orientation and bias of the line are dynamically changed to accommodate all the instances of class C_1 such that

$$g(X_i) = wX_i + b > 0\ then\ X_i \in C_1 \tag{7}$$

$$g(X_i) = wX_i + b < 0\ then\ X_i \in C_2 \tag{8}$$

For any instance, if $X_i \in C_1$, then $wX_i + b > 0$, and if $X_i \in C_2$, then, $wX_i + b < 0$, and from the given labels, we can encode as

$$y_i = \pm 1\ such\ that\ y_i = +1\ if\ X_i \in C_1\ and\ \ y_i = -1\ if\ X_i \in C_2 \tag{9}$$

Combining the above three equations, we can get another equation as follows

$$y_i(wX_i + b) > 0 \tag{10}$$

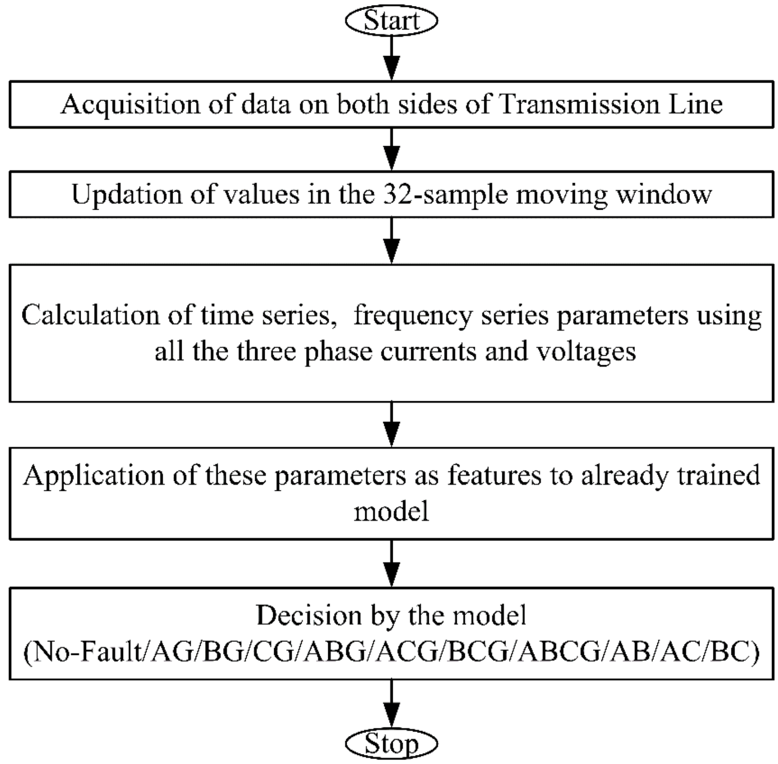

Fig. 1. the proposed moving window of one cycle (thirty-two samples) for calculating the features applied to the already trained model for fault classification

If p is any vector from the testing data set, and if $(wp + b) > 0$ then p belongs to C_1; otherwise, p belongs to C_2.

Let x be the support vector in the given training data set. In order to get good classification results irrespective of the pollution in the testing data, we need to maximize the margin between the support vectors of the two classes.

The distance from the support vector to the boundary is defined as

$$\frac{w.x + b}{|w|} \geq \gamma \tag{11}$$

$$w.x + b \geq \gamma|w| \tag{12}$$

After proper scaling, the above equation can be rewritten as

$$w.x + b \geq 1 \ if \ x \in C_1 \tag{13}$$

$$w.x + b \leq -1 \ if \ x \in C_2 \tag{14}$$

In order to maximize the margin, γ, |w| should be the minimum value, and b should be the maximum value

To get the minimum value of w, let us define a function

$$\varphi(w) = w^t.w \cong \frac{1}{2}w.w \tag{15}$$

Now, let us minimize the function $\varphi(w)$ such that $y_i(w.x_i + b) = 1$

We all know that a Lagrangian multiplier can be used to convert a constrained optimization problem to an unconstrained optimization problem.

$$L(w,b) = \frac{1}{2}(w.w) - \sum_{i=1}^{n} \alpha_i [y_i(w.x_i + b) - 1] \tag{16}$$

where, $\alpha_i \geq 0$ is the Lagrangian multiplier.

$$L(w,b) = \frac{1}{2}(w.w) - \sum_{i=1}^{n} \alpha_i y_i w x_i - \sum_{i=1}^{n} \alpha_i y_i b + \sum_{i=1}^{n} \alpha_i \tag{17}$$

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{n} \alpha_i.y_i = 0 \tag{18}$$

Where n= No. of features. Similarly,

$$\frac{\partial L}{\partial w} = w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0 \tag{19}$$

Substituting (12) and (13) in (11), and after further simplification, we get

$$L = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_j y_i y_j (x_i.x_j) \tag{20}$$

Once the value of L is known, the values of w and b can be computed from (16). This enables the identification of the support vectors in the training data with the maximum possible distance between the support vectors.

3. Methodology Used for Fault Diagnostic Analysis in the Present Work

There are 11 classes in the problem statement: no-fault, AG, BG, CG, AB, AC, BC, ABG, ACG, BCG, and ABCG. The purpose of this paper's work is to classify the classes as precisely as possible. The thirty-two sample moving window is refreshed at a sampling frequency of 1920 Hz, as shown in Fig. 1. The moving window stores the three-phase instantaneous line currents and three-phase instantaneous line to ground voltages on both sides of the TL. The moving window is updated with the latest thirty-two samples after the fault instant in the current study, as illustrated in Fig. 1. Time series and frequency series parameters are generated from the values recorded in the moving window. These parameters are used as features in the SVM algorithm for training and testing. Minimum, maximum, average, root mean square, peak to peak values of all three-phase voltages and currents on both sides of the transmission line are among the time-series properties.

The DC value, magnitudes, and phase angles of the fundamental signal or first harmonic up to the seventh harmonic are all

Table 1. Variation in important fault parameters for simulation of the fault and non-fault scenarios

Parameter	Range of Values	Step size	Count
Fault Type	no-fault, AG, BG, CG, AB, AC, BC, ABG, ACG, BCG, and ABCG	NA	11
Fault Resistance	0 Ω to 60 Ω	5 Ω	13
Fault Location	0 km to 100 km	10 km	11
Fault Inception Angle	0° deg to 360° deg	22.5° deg	16
Total Fault scenarios studied in this work (10x13x11x16)			22,880
Total Non-Fault scenarios studied in this work (1x13x11x16)			2,288

Table 2. Formulae of evaluation metrics

Metric	Accuracy	Precision	F1-score	Recall
Order of Preference	1	2	3	4
Formula	$\frac{(TP+TN)}{(T+N)}$	$\frac{TP}{(TP+FP)}$	$\frac{(2*Precision*Recall)/}{(Precision\ +\ Recall)}$	$\frac{TP}{(TP+FN)}$

Table 3. Quantitative values of evaluation metrics for the full cycle moving window data sets without any noise

Fold	Accuracy	F1 score	Precision	Recall
1	1	1	1	1
2	1	1	1	1
3	1	1	1	1
4	1	1	1	1
5	1	1	1	1
6	1	1	1	1
7	1	1	1	1
8	1	1	1	1
9	1	1	1	1
10	1	1	1	1
Avg	1	1	1	1

Table 4. Ten-fold average quantitative values of evaluation metrics for data with different noise levels

Noise Level	Accuracy	F1 score	Precision	Recall
30 dB	0.99996	0.99996	0.99996	0.99996
35 dB	1	1	1	1
40 dB	1	1	1	1

included in the frequency series. These frequency series features are estimated for all three-phase voltages and currents on both sides of the transmission line.

4. Data Generation for Training and Testing of the SVM Model

The transmission line connecting buses 7 and 8 is considered to simulate a variety of fault and non-fault scenarios. To simulate faults at various locations along the transmission line, including bus seven and bus eight, the line comprises ten equal portions. All possible faults, i. e. No-fault, A phase to Ground fault (AG), B phase to Ground (BG), C phase to Ground (CG), A phase to B phase to Ground (ABG), A phase to C phase to Ground (ACG), B phase to C phase to Ground (BCG), A phase to B phase to C phase to Ground (ABCG), A phase to B phase (AB), A phase to C phase (AC) and B phase to C phase (BC) are simulated at each location of the transmission line. Each type of fault at each location is simulated to accommodate all possible variations of fault parameters like fault resistance and fault inception angle. Fault resistance varies from 0 Ω to 60 Ω in steps of 5 Ω, and fault inception angle varies from 0 deg to 360 deg in 22.5 deg. The parameters, varied to simulate all fault and non-fault scenarios, are summarized in Table 1.

5. Data Validation Before Employing the Data in Training and Testing

In any machine learning or artificial intelligence method, data is of utmost importance as the output of the model is as good as the data used for the training and testing. This section will describe the visual inspection of the generated data before training and testing. All other parameters, such as fault location and inception angle, should remain constant when the fault resistance values grow from 0 to 60 in step 5. The variations in the fault voltages and fault currents measured on TL near bus 7 for fault resistances of 0 Ω and 60 Ω, respectively, are shown in Fig. 2. Fig. 2(a) and Fig. 2(b) show the variations in fault voltages and currents for the LG fault. Fig. 2(a) and Fig. 2(b) show that the fault current values will decrease and fault voltage values will increase, respectively, for an increase in fault resistance value. Figures 2(c) and 2(d) show the variations in fault voltages and fault currents for LLG fault. Figures 2(c) and 2(d) show that the fault voltages and currents behave the same way as the LG fault. Figures 2(e) and 2(f) show the fault voltages and fault current variations for the LLLG fault. Figures 2(e) and 2(f) reiterate that the fault voltage magnitude increases with the increase of fault resistance and vice versa. Figures 2(e) and 2(f) also reiterated that the fault current magnitudes decrease with the increase of fault resistance value and vice versa. Similar behaviour will be observed in the fault currents and voltages measured at bus 8 for different types of faults.

In a fault scenario, as the fault inception angles are increased, keeping the fault resistance and fault location constant- depending on the load and other conditions in the grid, the values of the fault currents and fault voltages also change as shown in Fig. 3. Fig. 3 shows the variation in the fault voltages and fault currents measured at bus seven due to the change of fault inception angle for LG, LLG and LLG faults. Similar behaviour will be observed in the fault currents and voltages measured at bus 8 for different types of faults.

6. Results and Discussions

In this work, a total of 11 classes are present: no-fault, AG, BG, CG, AB, AC, BC, ABG, ACG, BCG, and ABCG. Suppose the classification algorithm predicts a data point that belongs to AG as AG. In that case, this result can be called true positive (TP). Suppose the classification algorithm predicts a data point that belongs to AG, as any other class except AG. In that case, this result can be called true negative (FN). Suppose the classification algorithm predicts a data point that belongs to any other class except AG, as AG. In that case, this result can be called a false positive (FP). Suppose the classification algorithm predicts a data point that belongs to any other class except AG as any other class. In that case, this result can be called true negative (TN).

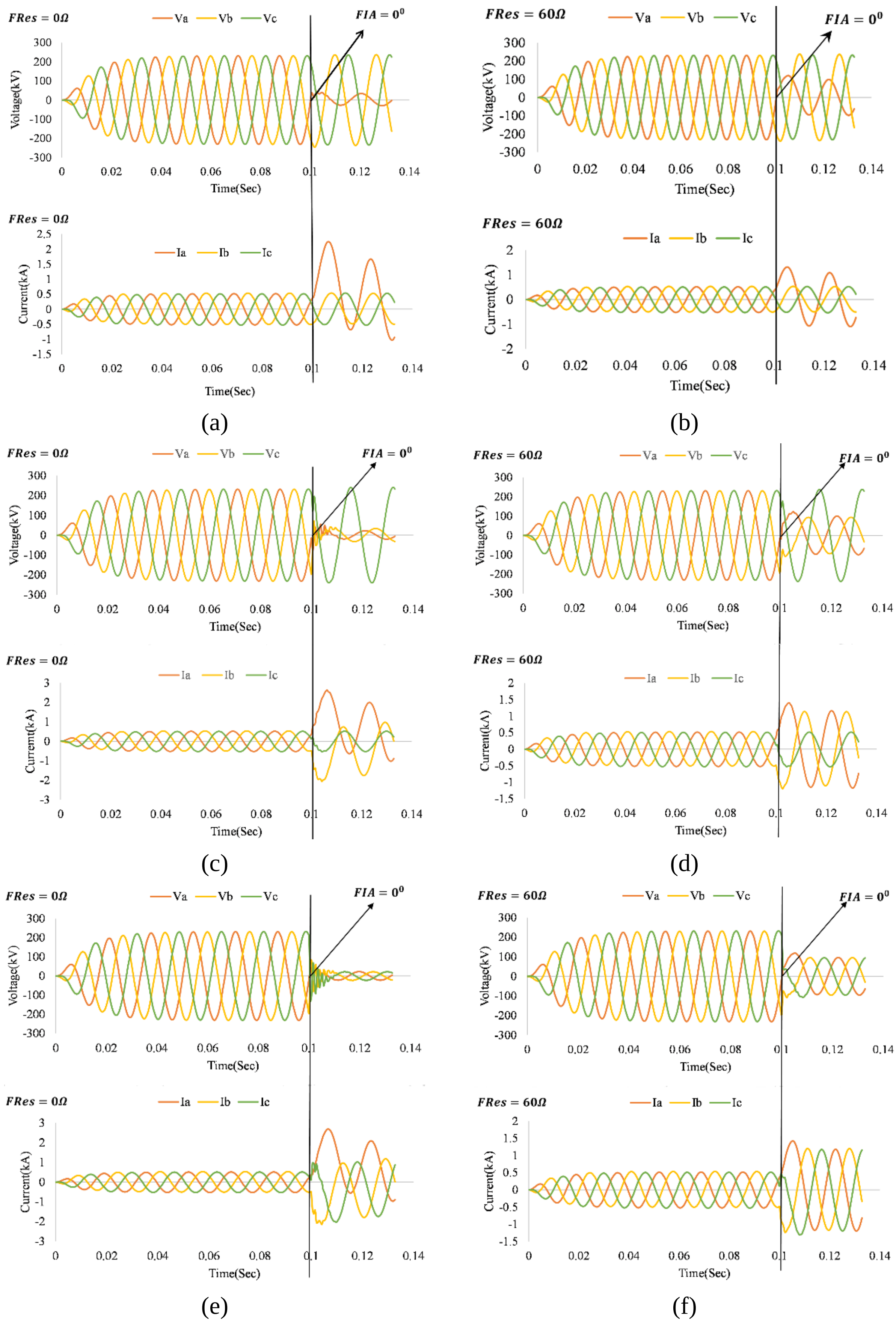

Fig. 2. Variations in voltage and current signals measured on TL near bus seven due to change in fault resistance values for LG, LLG, and LLLG faults

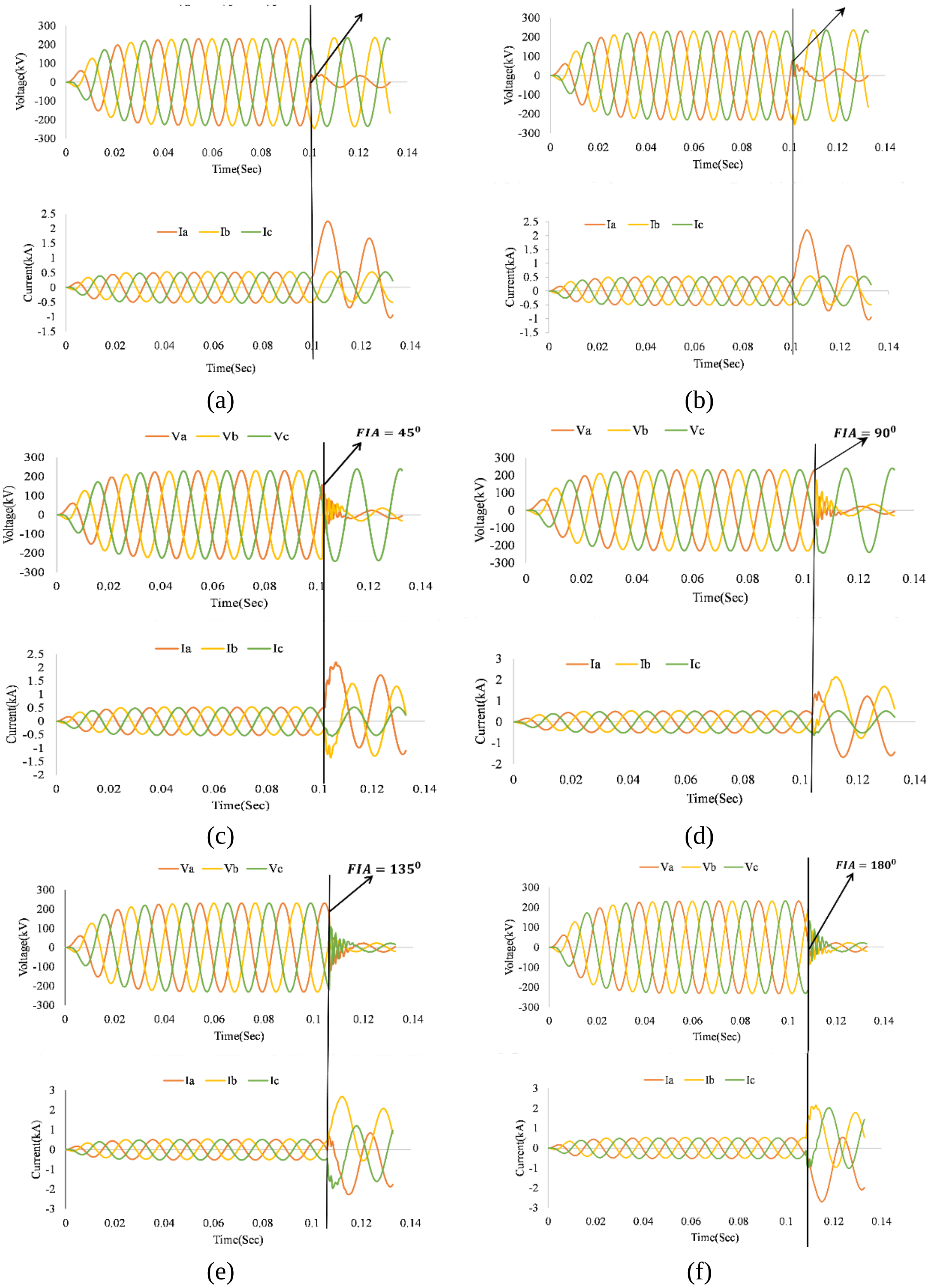

Fig. 3. Variations in voltage and current signals measured on TL near bus seven due to change in fault inception angle values for LG, LLG, and LLLG faults

Table 5. Performance comparison of the proposed method with other ML algorithms for (a) data set without any noise, (b) data set with 30dB SNR noise level, (c) data set with 35dB SNR noise level and (d) data set with 40dB SNR noise level

Noise	Algorithm	Accuracy	F1 Score	Precision	Recall
No noise	ABC	0.782	0.70888	0.67253	0.78163
No noise	DT	0.99984	0.99984	0.99984	0.99984
No noise	GNB	1	1	1	1
No noise	KNN	0.99217	0.99217	0.99221	0.99217
No noise	RF	0.99869	0.99869	0.99871	0.99869
No noise	SVM	1	1	1	1
30 dB	ABC	0.72703	0.6361	0.59067	0.72703
30 dB	DT	0.98057	0.98057	0.98065	0.98057
30 dB	GNB	0.93818	0.93766	0.94511	0.93818
30 dB	KNN	0.99074	0.99074	0.9908	0.99074
30 dB	RF	0.74837	0.68218	0.84737	0.74837
30 dB	SVM	0.99996	0.99996	0.99996	0.99996
35 dB	ABC	0.72703	0.6361	0.59067	0.72703
35 dB	DT	0.98411	0.98411	0.98417	0.98411
35 dB	GNB	0.94271	0.94242	0.94677	0.94271
35 dB	KNN	0.99146	0.99146	0.99149	0.99146
35 dB	RF	0.74289	0.67201	0.80756	0.74289
35 dB	SVM	1	1	1	1
40 dB	ABC	0.72703	0.6361	0.59067	0.72703
40 dB	DT	0.98633	0.98633	0.98643	0.98633
40 dB	GNB	0.95311	0.95304	0.95446	0.95311
40 dB	KNN	0.99166	0.99166	0.9917	0.99166
40 dB	RF	0.74348	0.67359	0.84067	0.74348
40 dB	SVM	1	1	1	1

Table 6. Performance comparison of the proposed method with other existing methods in the literature

Parameter	Proposed Method	[34]	[35]	[36]	[37]
Number of the fault and non-fault scenarios	25168	12474	1485	59825	24948
Accuracy	100%	99.34	100	98.9	99.29
Noise Levels Considered	30 dB, 35 dB and 40 dB	20 dB, 30 dB and 40 dB	Not considered	Not considered	Not considered

The performance of the proposed model is evaluated with the help of popular classification metrics such as accuracy, precision, F1-score and Recall. The formulae for the metrics are provided in Table 2. Classification accuracy is the ratio of total correct predictions to the total number of predictions.

This metric is the most popular and frequently used metric to quantify the classification accuracy of any classification algorithm. Accuracy talks about the classification algorithm's ability to make correct predictions. Precision is the ratio of true positive predictions to the sum of true positive predictions and false-positive predictions. Precision quantifies the ability of the classification algorithm in terms of correct positive predictions. Recall or sensitivity is the ratio of true positive predictions to the sum of true positive predictions and false-negative predictions. Recall quantifies the ability of the classification algorithm in terms of exact prediction. F1-score is the harmonic mean of precision and recall. F1-score is considered a very important metric for quantifying the ability of the classification algorithm in some situations where the trade-off between precision and recall cannot be decided. Here in this work, the order of preference for the evaluation metrics is Accuracy, Recall, Precision and F1-score. The evaluation is carried out using a ten-fold cross-validation procedure. The range of values for all the evaluation metrics is 0 to 1.0, and 0 indicates the worst performance. In contrast, 1 indicates the best possible performance.

6.1. Performance of the proposed method with actual data

The result of the proposed method for full-cycle data is provided in Table 3. As shown in Table 3, all the metrics, i. e., accuracy, precision, recall and F1-score, are equal to 1, indicating that all the classes are accurately classified.

6.2. Performance of the proposed method with polluted data

To test the robustness of the proposed model, the model is tested with data sets which consist of polluted data of different levels. The pollution in any given data is measured by a signal-to-noise ratio (SNR) generally measured in dB. The lower the value of the SNR in dB, the higher the pollution in the given data. Three data sets are created for full-cycle data with different noise levels such as 30dB, 35dB and 40dB. The average values (for the ten folds) of the classification evaluation metrics full-cycle moving window with the noise level of 30dB, the most polluted data set, are provided in Table 4. The accuracy, F1-score, precision and recall values are the same for the full cycle data and equal to 0.999960. For 35 dB and 40 dB, all the evaluation metrics have a value of one. The proposed method can classify the faults with 100 per cent accuracy for the data sets with 35 dB and 40 dB noise.

Table 4 Ten-fold average quantitative values of evaluation metrics for data with different noise levels.

6.3. Comparison of the performance of the proposed method with that of the other Machine Learning Algorithms

The performance of the proposed method is compared with that of the other popular Machine Learning classification algorithms such as the Decision Tree algorithm (DT), Random Forest algorithm (RF), k-Nearest Neighbours algorithm (KNN), Adaptive Boosting Classifier algorithm (ABC) and Gaussian Nave Bayes classification algorithm (GNB) to validate the performance of the proposed method. The average values of the classification evaluation metrics for full-cycle moving window data without any noise are presented in Table 5. The net values of the classification evaluation metrics for full-cycle moving window data with 30dB, 35 dB and 40 dB noise are also provided in Table 5. From the table, it can be seen that the proposed algorithm outperformed the remaining ML algorithms in each case. GNB performed well in the case of data with no noise, but its performance, compared to the performance of SVM, is rather poor. The proposed SVM achieved 100 % accuracy in every case except with high polluted noise of 30 dB noise case.

6.4. Comparison of the performance of the proposed method with that of the other existing methods in the literature

Table 6 shows the proposed method's performance with that of the other existing literature. The proposed method uses 25168 fault and non-fault scenarios for the training and testing of the proposed method. Noise analysis for noise with 30 dB, 35 dB and 40 dB is also performed. The proposed method gives a 100 % accurate classification of faults and non-faults. Authors in [34] used 12474 scenarios and also performed noise analysis with 20 dB, 30dB and 40 dB. They have achieved an accuracy of 99.34%. Noise analysis is not performed in [35–37]. Authors in [36] used 59825 scenarios and achieved 98.9% of accurate results. Performance comparison of the proposed method with other existing methods in the literature

7. CONCLUSIONS

This work proposes a new method for fault classification of transmission line faults using SVM with the help of time-series and frequency series parameters. The performance of the proposed method is checked with data derived from all possible situations. This data includes the variations of fault resistance, fault locations and inception angles. The algorithms performance is also tested by adding noise of different SNR levels.

The proposed model can classify the transmission line faults with 100% accuracy for all the full cycle moving window data without any noise. The proposed model can achieve a classification accuracy of approximately 100% even for 35 dB and 40 dB noise data. The proposed model achieves a classification accuracy of 99.996%, even with the most polluted data set. The proposed model can classify the transmission line faults even under the wide variations of fault resistance, fault inception angle and location. The proposed model performs well even under noise in the data sets. The performance of the proposed method is also tested by comparing it with the performance of other popular classification methods. The classification accuracy of the proposed method is far better than that achieved with ABC, DT, GNB, kNN, and RF algorithms.

REFERENCES

[1] J. Allen Wood, F. Bruce Wollenberg, and B. Gerald Sheblé, “Power generation, operation, and control,” *John Wiley & Sons*, 2013.

[2] K. Atul Raturi, “Renewables 2019 global status report,” *Tech. Rep, REN21 Secretariat*, 2019.

[3] B. Chen, “Fault statistics and analysis of 220-kV and above transmission lines in a southern coastal provincial power grid of China,” *IEEE Open Access J. Power Energy*, Vol. 7, pp. 122–129, 2020.

[4] R.M. Arias Velásquez, “Performance improvement in long overhead lines associated to single-phase faults due to atmospheric discharges,” *Eng. Fail. Anal.*, Vol. 105, pp. 347–372, 2019.

[5] H. Haes Alhelou, M E. Hamedani-Golshan, T. Cuthbert Njenda, and P. Siano, “A survey on power system blackout and cascading events: Research motivations and challenges,” *Energies*, Vol. 12, No. 4, pp. 682, 2019.

[6] US DOE, “Enabling modernization of the electric power system,” *Quadrennial techno. review*, Vol. 22, 2015.

[7] J M. Maza-Ortega, E. Acha, S. García, and A. Gómez-Expósito, “Overview of power electronics technology and applications in power generation transmission and distribution,” *J. Mod. Power Syst. Clean Energy*, Vol. 5, No. 4, pp. 499–514, 2017.

[8] M. I. Henderson, D. Novosel, and M L. Crow, “Electric power grid modernization trends, challenges, and opportunities,” *IEEE Advancing Techno. Humanity*, 2017.

[9] National Academies of Sciences Engineering Medicine et al., The power of change: Innovation for development and deployment of increasingly clean electric power technologies, *National Academies Press*, 2016.

[10] E. Naderi, M. Pourakbari-Kasmaei, and M. Lehtonen, “Transmission expansion planning integrated with wind farms: A review, comparative study, and a novel profound search approach,” International *Int. J. Electr. Power Energy Syst.* , Vol. 115, pp. 105460, 2020.

[11] P L. Joskow, “Transmission capacity expansion is needed to decarbonize the electricity sector efficiently,” *Joule*, Vol. 4, No. 1, pp. 1–3, 2020.

[12] N M. Zainuddin, MS. Abd Rahman, MZA. Ab Kadir, NH. Nik Ali, A. Zaipatimah, Miszaina Osman, M. Mansor, A. Mohd Ariffin, M. Syahmi Abd Rahman, SFM Nor, et al, “Review of thermal stress and condition monitoring technologies for overhead transmission lines: Issues and challenges,” *IEEE Access*, Vol. 8, pp. 120053–120081, 2020.

[13] B. Chatterjee and Sudipta Debnath, “A new protection scheme for transmission lines utilizing positive sequence fault components,” *Electr. Power Syst. Res.*, Vol. 190, pp. 106847, 2021.

[14] YI. Xiao-Ran, ZHOU. En-Zhe, YE. Li, DU. Shuang-Yu, and YU. Zhan-Qing, “Monitoring of transmission line wildfires using satellite remote sensing,”*3rd Annual Int. Conf. Electron. Electri. Eng. Inf. Sci. (EEEIS 2017), Atlantis Press*, 2017.

[15] M .Anthony Sleva, Protective relay principles, *CRC Press*, 2018.

[16] R. K.Avvari, and V. DM. Kumar ‘A novel hybrid multi-objective evolutionary algorithm for optimal power flow in wind, PV, and PEV systems,” *J. Oper. Autom. Power Eng.*, 2022.

[17] B. Bhalja and RP. Maheshwari, “Waveletbased fault classification scheme for a transmission line using a support vector machine,” *Electric Power Compon. Sys.*, Vol. 36, No. 10, pp. 1017–1030, 2008.

[18] E. Godoy, A. Celaya, H.J. Altuve, N. Fischer, and A. Guzmán, “Tutorial on single-pole tripping and reclosing," *Western Protective Relay Conf.*, pp. 1–21, 2012.

[19] J.A. Jiang, J.Z. Yang, Y-H. Lin, C-W. Liu, and J-C. Ma, “An adaptive pmu based fault detection/location technique for transmission lines.i.theory and algorithms,” *IEEE Trans. Power Delivery*, Vol. 15, No. 2, pp. 486–493, 2000.

[20] J-A. Jiang, C-S. Chen, and C-W. Liu, “A new protection scheme for fault detection, direction discrimination, classification, and location in transmission lines,” *IEEE Trans. Power Delivery*, Vol. 18, No. 1, pp. 34–42, 2003.

[21] A. Asuhaimi Mohd Zin, M.Saini, M. Wazir Mustafa, A. Rizal Sultan, and Rahimuddin, “New algorithm for detection and fault classification on parallel transmission line using dwt and bpnn based on clarke's transformation,” *Neurocomputing*, Vol. 168, pp. 983–993, 2015.

[22] PK. Dash and SR. Samantaray, “An accurate fault classification algorithm using a minimal radial basis function neural network,” *Int. J. Eng. Intelligent Sys. Electri. Eng. Commun.*, 2004.

[23] RN Mahanty and PB Dutta Gupta, “Application of rbf neural network to fault classification and location in transmission lines”, *IEE Proc. Gener. Transm. Distrib.*, Vol. 151, No. 2, pp. 201–212, 2004.

[24] T. Dalstein and B. Kulicke, “Neural network approach to fault classification for high speed protective relaying,” *IEEE Trans. Power Delivery*, Vol. 10, No. 2, pp. 1002–1011, 1995.

[25] P.Venkata, V. Pandya, and A.V. Sant. “Data mining model

based differential microgrid fault classification using SVM considering voltage and current distortions," *J. Oper. Autom. Power Eng.*, 2022.

[26] M. A. Baherifard, R. Kazemzadeh, A.S. Yazdankhah, and M. Marzband, "Improving the effect of electric vehicle charging on imbalance index in the unbalanced distribution network using demand response considering data mining techniques," *J. Oper. Autom. Power Eng.*, Vol. 11, No. 3, pp. 182–192, 2023.

[27] Bhasker, S. K., et al. "Differential protection of ISPST using Chebyshev neural network," *J. Oper. Autom. Power Eng.*, Vol. 11, No. 2, pp. 123-129, 2023.

[28] RN. Mahanty and PB. Dutta Gupta, "A fuzzy logic-based fault classification approach using current samples only," *Electr. Power Syst. Res.*, Vol. 77, No. 5–6, pp. 501–507, 2007.

[29] Le. Xu, M-Y. Chow, and L.S. Taylor, "Power distribution fault cause identification with imbalanced data using the data miningbased fuzzy classification e-algorithm," *IEEE Trans. Power Syst.*, Vol. 22, No. 1, pp. 164–171, 2007.

[30] J-SR. Jang, "Anfis: adaptive-network-based fuzzy inference system," *IEEE Trans. Syst. Man Cybern.*, Vol. 23, No. 3, pp. 665–685, 1993.

[31] TSKMM. Hassan, "Adaptive neuro fuzzy inference system (anfis) for fault classification in the transmission lines," *Online J. Electron. Electr. Eng. (OJEEE)*, Vol. 2, pp. 2551–2555, 2010.

[32] H. Wang and WWL. Keerthipala, "Fuzzyneuro approach to fault classification for transmission line protection," *IEEE Trans. Power Delivery*, Vol. 13, No. 4, pp. 1093–1104, 1998.

[33] J-SR. J.and C-T. Sun, "Neuro-fuzzy modeling and control," *Proc. the IEEE*, Vol. 83, No. 3, pp. 378–406, 1995.

[34] Ch. Biswapriya, and S. Debnath. "Cross correlation aided fuzzy based relaying scheme for fault classification in transmission lines," *Eng. Sci. Technol. Int. J.*, Vol. 23.3, pp. 534-543, 2020.

[35] M. Alok, Palash Kumar Kundu, and A. Das. "Application of principal component analysis for fault classification in transmission line with ratio-based method and probabilistic neural network: a comparative analysis," *J. Inst. Eng. India Ser. B*, Vol. 101.4, pp. 321-333, 2020.

[36] Ch.Y. Qi, O. Fink, and G. Sansavini. "Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction," *IEEE Trans. Ind. Electron.*, Vol. 65.1, pp. 561-569, 2017.

[37] Chen, Kunjin, Jun Hu, and Jinliang He. "Detection and classification of transmission line faults based on unsupervised feature learning and convolutional sparse autoencoder," *IEEE Trans. Smart Grid*, Vol. 9, No. 3, pp. 1748–1758, 2016.

[38] A. Jamehbozorg and S. M. Shahrtash, "A decision-tree-based method for fault classification in single-circuit transmission lines," *IEEE Trans. Power Delivery*, Vol. 25, No. 4, pp. 2190–2196, 2010.

[39] SR Samantaray, PK Dash, and G Panda, "Distance relaying for transmission line using support vector machine and radial basis function neural network," International *Int. J. Electr. Power Energy Syst.*, Vol. 29, No. 7, pp. 551–556, 2007.

[40] U. B. Parikh, B. Das, and R. Maheshwari, "Fault classification technique for series compensated transmission line using support vector machine," International *Int. J. Electr. Power Energy Syst.*, Vol. 32, No. 6, pp. 629–636, 2010.

[41] M. Manohar and E. Koley, "Svm based protection scheme for microgrid,"*Int. Conf. Intell. Comput. Instrum. Control Technol. (ICICICT)*, *IEEE*, pp. 429–432, 2017.

[42] D. Boswell, "Introduction to support vector machines," *Dep. Computer Sci. Eng. Univ. California San Diego*, 2002.

[43] H. Livani and C.Y. Evrenosoğlu, "A fault classification and localization method for three-terminal circuits using machine learning," *IEEE Trans. Power Delivery*, Vol. 28, No. 4, pp. 2282–2290, 2013.