

Secured protection of transmission line by distance relay using data mining approach

M. Kiruthika, Bindu. S

Department of Electrical Engineering, Agnel Charities' Fr. Conceicao Rodrigues Institute of Technology, Vashi, Navi Mumbai, India

Article Info

Article history:

Received Oct 14, 2020

Revised May 19, 2021

Accepted Jun 18, 2021

Keywords:

Distance relay

Fault identification

Phasor measurement unit

Protection

ABSTRACT

Distance relay is one of the most important protection element of a transmission line used in protection schemes. Relay may malfunction if it is not able to distinguish faults from system stressed conditions. This work mainly focusses on enhancing the performance of the distance relay in a secured manner based on data mining approach which uses two phases of classification. Level 1 classifier identifies the system conditions like normal, fault, and power swing and level 2 classifier gets initiated when there is a power swing and distinguishes between the persistence of power swing condition and a three-phase fault. In both the phases, the protection scheme in the respective zone where the fault occurred gets activated. The proposed methodology is tested for an IEEE 9-bus system wherein the data is collected from phasor measurement units placed in optimal locations. Optimal PMU placement is economical since it overcomes issues like cost, communication infrastructure issues, maintenance and complexity. The results proved that the proposed method is effective with good efficiency and higher accuracy with less number of PMUs.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

M. Kiruthika

Department of Electrical Engineering

Agnel Charities' Fr. Conceicao Rodrigues Institute of Technology

Vashi, Navi Mumbai, India

Email: m.kiruthika@fcrit.ac.in

1. INTRODUCTION

In today's world scenario, usage of power keeps rising day by day. High power usage demands for a proper infrastructure for transmitting power from the generation centers to the regions in remote locations through long distance lines of transmission. The main objective of all power systems is to maintain continuous power supply and minimize power transmission losses. However, natural events like lightning, weather, and equipment failure may lead to faults. Under the worst scenario, if the fault persists, it may lead to long term power loss, blackouts, and also cause permanent damage to some equipment. Thus, to prevent such undesirable situations the temporary isolation of the live system has to be done as early as possible. To achieve this transmission line is divided into zones and protection is arranged in each zone. These zones are generally overlapping each other so that none of the parts of the power system is left unprotected and also to provide back-up protection to remote sections.

Distance relays are commonly employed for primary as well as back-up security for transmission lines. It makes a decision based on the local current and voltage measures. During stressed conditions, the distance relay struggles to differentiate between a stress condition and a fault, and as a result there exists a possibility that the relay may mal-operate.

During stressed conditions, the distance relay can misinterpret it as a fault and can issue a trip decision [1]. Hence, the research community is focusing to develop approaches that can improve the performance of the distance relay. Lin *et al.* developed a mho type distance relay model to operate correctly under various types of fault [2]. Prashant Gawande *et al.* proposed an approach based on the prediction of impedance, for detecting the 3 phase faults under system stresses like load encroachment and power swing [3]. This method works on the basis of local measurements of voltage and current.

Li Y *et al.* presented an approach to identify the faulty line with only sparse synchronized measurements [4]. Samantaray *et al.* presented a scheme, to obtain a full observable wide area backup scheme of protection by employing cubature kalman filtering based DSE with the phasor measurement unit [5]. Sharafi A. *et al.* and Jose T. *et al.* proposed approaches to improve the relay's performance under the system stressed circumstances [6], [7]. There are various backup protection schemes based on synchrophasor measurements and adaptive in nature that are proposed in the literature [8]-[13]. Various techniques to avoid zone 3 maloperation under various stressed conditions are presented in the literature [14]-[19].

Rao *et al.* proposed a power swing blocking algorithm to prevent unintended tripping of the transmission lines because of the malfunctioning of the distance relays. This algorithm uses synchrophasor measurements for distinguishing power swing from fault. In this research work, an impedance trajectory prediction model was developed for supervising distance relay under power swing with the help of synchrophasor measurements [20]. Lavand, S. A *et al.* proposed a synchrophasor-based scheme of supervision for improving the security of protection of zone 1 at the time of power swing by employing a prediction technique. Current differential protection is employed in this work to confirm the presence of fault by placing synchrophasors at the two ends of the line [21]. This approach needs measuring devices at all the buses and so this approach is not viable practically.

Also, issues like increase in data size, uncertainty associated with data, complex and nonlinear behavior of power systems are need to be currently addressed. These issues can be handled by computationally intelligent methods like data mining techniques. Data mining approaches addressing discrimination of fault and system stress conditions and enhancing the distance relay's performance for protection are available in the literature but are very limited. An adaptive technique of protection to enhance the performance of relay at the time of power swing for both uncompensated and compensated lines based on data mining is proposed in [22]. The scheme uses random forests (RF) and decision tree (DT) to provide supervisory control, distinguishing power swing, faults during power swings, and also identifying fault zone for series compensated lines.

Dubey, R. *et al.* presented an online adaptive technique of protection using data mining algorithms such as, DT and RF to supervise and enhance the output of distance relay [23]. The above approach used Phasor measurement units (PMUs) in all the buses for measurements. A random forest algorithm is more complex, require more computational resources and is time-consuming. Support vector machine (SVM) is efficient but not suitable for large datasets and also it may underperform if the target classes overlap. The decision tree gives higher accuracy, however, the value achieved is less than 100% for the specified application. Placing PMUs in all the buses not only increases the installation cost, but also maintenance, communication infrastructure related issues and computational issues. Even though data analytic approaches can handle complex data, a cost effective solution is required to be addressed. This shows scope of testing the suitability of other data mining algorithms for discriminating faults and stressed conditions and protection by placing PMU's in optimal locations to achieve higher efficiency and accuracy.

In this research work, secure operation of distance relay is proposed for the IEEE 9-bus system. The PMUs are placed in optimal positions/buses from where the whole system can be observed achieving full observability. Using the data collected from these PMUs, two levels of classifiers are used. Level 1 Classifier is used to classify system conditions like normal, fault and stress condition (power swing). If a power swing condition is identified, then it goes to the next level classifier to differentiate conditions like power swing and symmetrical fault during power swing. Also, further, it identifies the zone where the fault occurred and the distance relay then sends a trip signal to the circuit breaker to isolate the faulty section from the healthy section.

After testing the suitability of five classifiers like convolutional neural network (CNN), decision tree, naïve bayes, SVM and K nearest neighbors (KNN) for the above mentioned research work it is found that decision tree and CNN were more suitable. Then, the system is further tested for large datasets with different splitting percentage of training and testing to validate the results. The rest of the research paper is formulated in the following manner: Section 2 includes the description of the methodology proposed; Section 3 includes the results and discussion; Section 4 presents the conclusion of the research work.

2. RESEARCH METHOD

2.1. Modelling of the IEEE-9 bus system

The system taken for the research work is IEEE-9 bus system. The IEEE-9 bus system is composed of 3 generators, 3 loads, 3 transformers, and 9 buses. The 3 generators are connected to the buses 1, 2, and 3, whereas, the 3 loads are linked to the buses 5, 6, and 8. The single line representation of the IEEE-9 bus system is displayed in Figure 1 and bus data is given in Table 1.

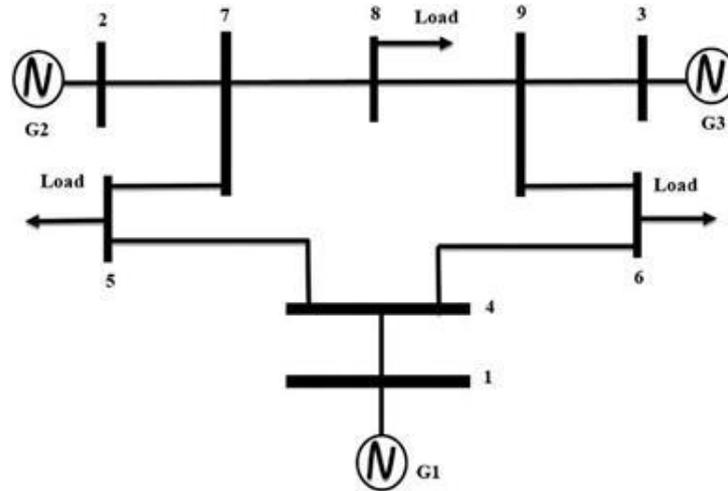


Figure 1. Single line representation of IEEE-9 bus system

Table 1. IEEE-9 bus system bus data

Sr. No.	Gen (MW)	Gen (Mvar)	Load (MW)	Load (Mvar)	Nom (kV)	Volt (kV)	P.U (volt)	Angle (Deg)
1	247.5	27.91			16.5	17.16	1.04	0
2	192	4.9			18	18.45	1.03	9.173
3	128	11.45			13.8	14.15	1.03	5.14
4					230	235.8	1.03	-2.22
5			125	50	230	229.9	1	-3.68
6			90	30	230	232.8	1.01	-3.57
7					230	236.2	1.03	3.8
8			100	35	230	234	1.02	1.34
9					230	237.5	1.03	2.44

2.2. Optimal placement of phasor measurement units (PMU)

A phasor measurement unit (PMU) is a device employed for estimating the phase angle and magnitude of an electric phasor quantity (like current or voltage) in the utility grid with a help of common source of time for synchronization. In this work, DFT (discrete fourier transform) based PMU is modelled along with P-class PMU algorithm that is designed using adaptive filters. The design reflects the Frequency tracking IEEE C37.118.1a compliant P-class PMU algorithm and this P-class PMU algorithm is intended for applications such as protection that requires fast response [24].

The technique to determine the optimal placement of PMUs is inspired from balas additive algorithm (BAA) [25]. The main aim of the optimal PMU placement (OPP) is to reduce the total installation cost of PMUs without compromising the information collected which is really essential for strategic decisions. The solution obtained for placing PMUs in optimal locations based on the above algorithm is that the full observability of the system is maintained.

Two cases have been identified for the bus system under study (IEEE-9 bus system) given in Table 2 i.e. (a) Case 1-without considering zero injection bus (ZIB) measurements and (b) Case 2-considering zero injection bus measurements. Zero injection buses are the ones which are not connected to either load or generator and they are buses 4, 7, 9 in the test system. With PMUs installed only at these buses, the system still ensures full observability. The data obtained from these PMUs are useful for analysis without any loss of information and can be processed with ease.

Table 2. Optimal placement of PMUs

Cases	Optimal no of PMUs	Optimal location of PMUs
Case 1: Without Considering Zero Injection Bus (ZIB) Measurements (Base case)	3	4, 7, 9
Case 2: Considering Zero Injection Bus measurements	2	6, 8

2.3. Proposed framework

The flow diagram of the proposed work is provided in Figure 2. The IEEE 9 bus system under study has been modelled using MATLAB/simulink. Data required for analysis is obtained using PMUs installed in optimal locations as per the cases discussed in Table 2. The data from the PMU are received at 20 kHz of sampling frequency. A total of 13 parameters are extracted from each PMU, they are: positive sequence voltage phase angle, positive sequence voltage magnitude, positive sequence current phase angle, positive sequence current magnitude, positive sequence impedance magnitude, positive sequence impedance phase angle, real power, reactive power, time stamp, positive sequence magnitude of voltage, frequency, phase reference, and ROCOF (rate of change of frequency). The above 13 parameters are obtained by integrating values from DFT based PMU (first eight) and P class PMU (next five).

The parameters are collected for various fault conditions for a time duration of 2s. These features are preprocessed and given as input to the classifiers of two levels. Level 1 classifier identifies whether the condition is normal, fault, or power swing. Furthermore, if the condition obtained is fault, then it identifies which zone is being affected (Zone 1 or 2 or 3) and protects the system by activating the relay. If this classifier identifies power swing condition, then the next level classifier is instigated which identifies whether the zone is affected only by power swing or power swing with symmetrical fault. Once the fault or fault under power swing is identified, protection is carried out.

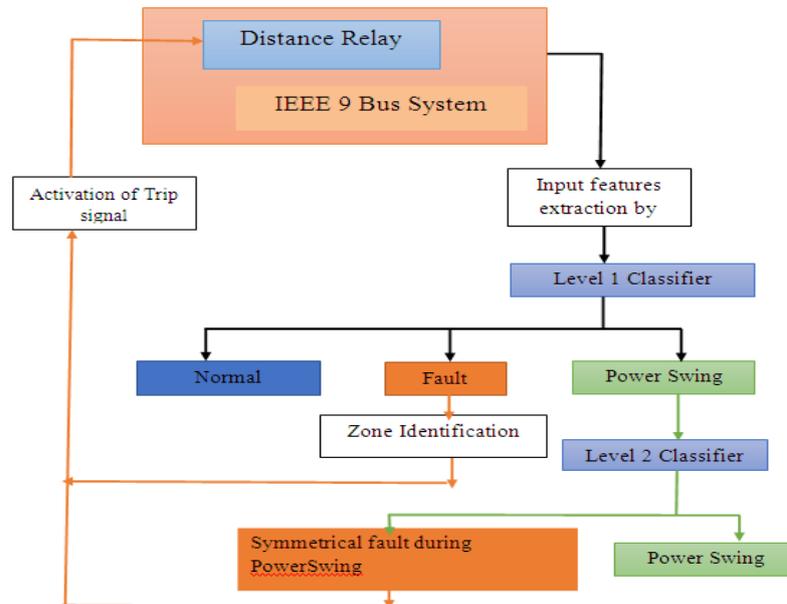


Figure 2. Framework of the proposed methodology

2.3.1. Datasets

The datasets used for the proposed data analytics approach contains vast amount of data which are being captured from the PMUs placed in the IEEE 9 bus system. By considering ZIB bus, PMUs placed at bus locations 4, 7, 9 produce data of size 40001×13 each (as there are 13 parameters mentioned earlier). So for the whole system we get data of dimension 40001×39 (approximately 40000 rows because of simulation time 2s and 39 columns because 13 parameters per PMU- 13×3) with full observability. By ignoring ZIB, the data size of 40001×13 each is obtained from the PMUs placed in optimal bus locations 6 and 8. So for the whole system, data obtained is of dimension 40001×26 (approximately 40000 rows because of simulation time 2s and 26 columns because 13

parameters per PMU-13x2). For both the cases, total count of the dataset set/sample collected is 213. This large data size proves that data analytics approach can help analyzing the system conditions in an efficient manner. Due to large size of the dimension of the data collected there is a need for preprocessing of the above data.

2.3.2. Data preprocessing

The main objective behind preprocessing of the dataset is to make the raw dataset more convenient and efficient for analysis without losing any information so that data quality is maintained. In this work, reconstructed independent component analysis (RICA) is used for preprocessing. This technique is used to reduce the size of the dataset i.e. to do dimensionality reduction. The classifiers will be using this dataset for efficient classification.

2.3.3. Classifiers

The dataset collected after preprocessing is fed to Level 1 classifier which detects the system conditions like normal, fault and power swing conditions. Also, this classifier identifies the respective zone where fault occurs activates the respective relay to protect the system. In case if the system condition is power swing then the next level classifier is instigated. The second level classifier then discriminates the system condition as power swing (if the power swing itself persists) or symmetrical fault in the existence of power swing. The above dataset is used to model different classifiers like CNN, decision tree, KNN and naïve bayes. The choice of the classifier for each level is done based on the evaluation metrics such as accuracy, sensitivity, error rate and specificity of the above classifier models.

2.3.4. Distance relay

Distance relay functions depending on the impedance ($Z=V/I$) between the point of the fault and the point where the relay is located. For the protection of the system during faults, the trip signal for the distance relay is derived based on the information from the classifiers.

2.3.5. Zone protection

Transmission line is segregated into several zones of protection normally zone 1, zone 2, and zone 3. Zone 1 is being secured by the distance relay kept in zone 1 while it acts as backup protection for zone 3 and zone 2. If a fault occurred in zone 1, the relay will instantaneously send trip signal to open the circuit breaker of zone 1. For faults in zone 2 and zone 3 tripping signal will be sent at a delayed time and the tripping time for zone 3 reach is more than that of zone 2.

3. RESULTS AND DISCUSSION

The proposed framework is modelled and simulated in MATLAB/Simulink platform. This segment displays the results of the simulation obtained for the proposed system.

3.1. Data collection

Initially, the system is simulated under normal condition and the data is collected for both the cases as mentioned in Table 2. The dataset is also collected for the following conditions namely, fault (1 phase to ground fault, 2 phase to ground fault, and 3 phase to ground fault), power swing and 3 phase to ground fault under power swing at various time durations of 0-0.5s, 0.5-1s, 1-1.5 s and 1.5-2s. Faults have been applied between 7th and 8th bus. Zones are segregated with reference to the relay placed near bus 4 for study purpose. A total of 213 datasets are collected for the following conditions:

- a) Normal Condition (1)
- b) Bus 4-5: Fault Conditions at Zone 1 (2-29)
- c) Bus 5-7 and Bus 6-9: Fault Conditions at Zone 2 (30-85)
- d) Bus 7-8, 8-9: Fault Conditions at Zone 3 (86-141)
- e) Power Swing Condition (142-154)
- f) Symmetrical Fault during Power swing (155-213)

Input thus generated has a matrix dimension of 40001 x 39 for one condition stated above for case 1 and for case 2 the matrix dimension would be 40001 x 26. The rows of the dataset depend on the simulation time and columns depend on number of parameters and PMUs. Since, the dimension is too large the above dataset has to be preprocessed. A total of 213 datasets for case 1 and case 2 are collected and preprocessed.

The waveforms of few system conditions taken into consideration for analysis are shown from Figures 3-8. All the waveforms presented are the recordings of PMU placed at bus 4. The Figures 3 and 4 show the voltage and current, respectively, when there is a presence of 3 phase to ground fault is created between 7th and 8th bus and the duration ranges from 1-1.5 s out of simulation time 2s. Figures 5 and 6 show

the voltage and current of a power swing condition on the line 5-7 simulated by tripping the line 7-8 at 0.7s. The Figures 7 and 8 illustrate the voltage and current observed during a power swing with symmetrical fault condition between 5th and 7th bus.

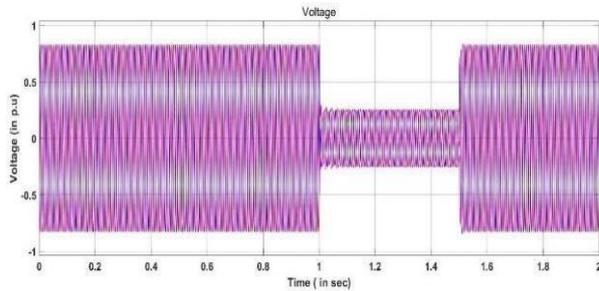


Figure 3. Voltage - 3 phase to ground fault

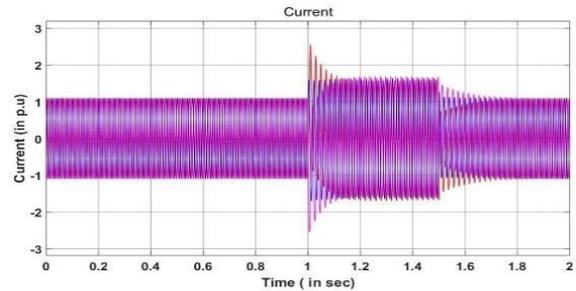


Figure 4. Current - 3 phase to ground fault

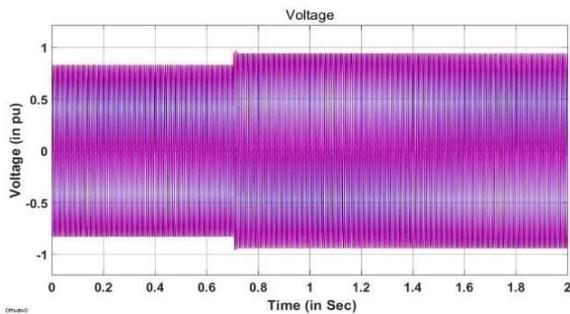


Figure 5. Voltage during power swing

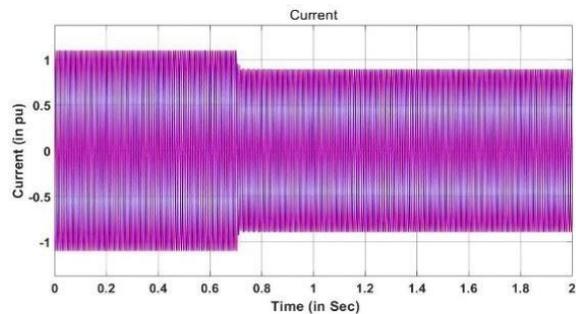


Figure 6. Current during power swing

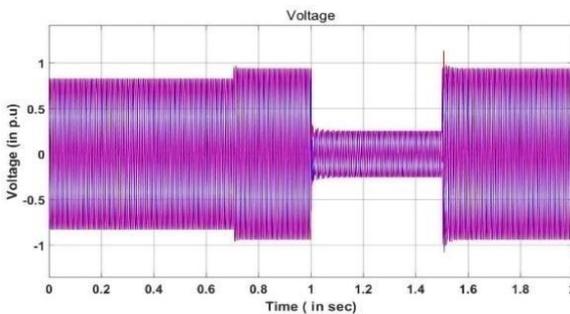


Figure 7. Voltage during power swing with symmetrical fault

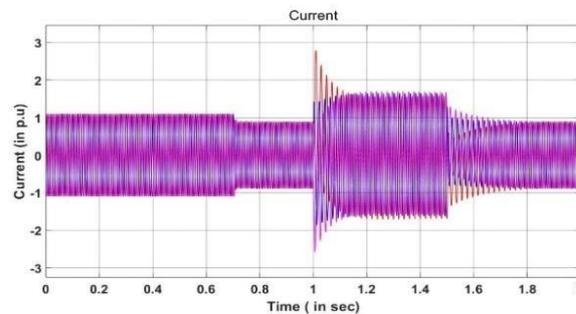


Figure 8. Current during power swing with symmetrical fault

3.2. Performance analysis

Performance analysis of the system has been performed for two cases as mentioned in Table 2.

3.2.1. Case 1: Optimal placement of PMUs in Bus locations 4, 7 and 9.

1) Preprocessing

As discussed, since the data size is large preprocessing becomes essential. This step helps in dimensionality reduction by keeping the quality of the data same. RICA-Algorithm (reconstruction independent component analysis) has been used for this purpose. For PMU locations at buses 4,7, 9 the input data size for one dataset is 40001 X 39 which interprets that there are 40001 rows and 39 columns/features (13 parameters x 3 PMUs). This now gets reduced to 13 X 39 after preprocessing. RICA algorithm reduces 40001 rows to 13 rows by maintaining the data efficiency. The data i.e. the feature matrix is sent as input to the classifier for level 1 classification of system conditions. There are 5 class labels allotted for classification

because the conditions to be identified are normal, fault conditions along with their zones (zone 1, zone 2 and zone 3) and power swing.

2) Level 1 classifier

All the required data collected from the system under study is now ready for analysis after preprocessing. Training and testing data comprises of all the functioning situations for various time frames. This data is fed to different classifier models for classification and to discriminate system conditions. Input gets classified into one of the 5 labels as mentioned above by this classifier. The best datamining model is selected for the classifier based on the evaluation metrics. Performance of different classifiers based on accuracy, sensitivity, specificity, and error rate are tabulated in Table 3. In general, accuracy is given by the ratio of correct classifications to the total count of instances evaluated and error rate is the ratio of incorrect classifications to the total count of instances evaluated. The capability of the classifier to recognize negative outcomes is specificity and to identify positive results is sensitivity. From Table 3, it is clear that CNN and Decision tree give higher accuracy compared to others. Hence CNN or Decision tree can be chosen as level 1 classifier and decision tree performance is presented in terms of confusion matrix in Figure 9(a) which is again used for evaluating the classifier performance. It can be seen from the matrix that all the samples have been correctly classified. Similarly, the confusion matrix of classifier CNN is presented in Figure 9(b). Diagonal cells point out the number of circumstances that have been classified properly by the classifier. Off diagonal elements specify the number of circumstances that have been mistakenly classified by the classifier. The last column in the matrix specifies the precision (or positive predictive value) and false discovery rate, respectively and last row specifies the recall (or true positive rate) and false negative rate, respectively.



Figure 9(a). Confusion matrix obtained for decision tree



Figure 9(b). Confusion matrix obtained for CNN

Table 3. Comparison for level 1 classifier

Sr. no	Classifiers	Accuracy (%)	Sensitivity	Error rate (%)	Specificity
1	Decision Tree	100	1	0	1
2	CNN	95.31	1	4.69	1
3	KNN	59.15	0	40.85	0.995
4	NAIVE BAYES	26.29	0	73.71	1
5	SVM	66.2	0	33.8	1

3) Level 2 classifier

Level 2 classifier is trained with data with the following labels: power swing and symmetrical fault during power swing. Similar to level 1 after feeding the data to different models, it is observed that CNN and decision tree give superior performance as shown in Table 4. After testing the trained decision tree, a confusion matrix is obtained to assess the performance of the classifier and is displayed in Figure 10. It can be seen that all the samples have been correctly classified.

Table 4. Comparison for level 2 classifier

Sr. no	Classifiers	Accuracy (%)	Sensitivity	Error rate	Specificity
1	Decision Tree	100	1	0	1
2	CNN	91.67	1	8.33	0.9091
3	KNN	81.94	0.0833	18.06	0.9667
4	NAIVE BAYES	83.33	0	16.67	1
5	SVM	83.33	0	16.67	1



Figure 10. Confusion matrix obtained for decision tree

4) Zone protection

For evaluating the performance of the proposed framework, a fault is initiated in a particular zone for a particular time duration and tested. Now this situation is sensed by the level 1 classifier as a fault condition and it also identifies its respective zone. As a result, the distance relay is stimulated to send a trip signal. In case of power swing, classifier 2 is initiated and when a three phase fault occurs during power swing this level 2 classifier initiates the required protection scheme and avoids mal-operation of the relay. The above methodology is explained with help of a test scenario by considering one of the relay near bus 4. The segregation of zones for the analysis of the IEEE system with reference to this relay is listed below.

- Zone 1: Bus No.4-5
- Zone 2: Bus No.5-7
- Zone 3: Bus No.7-8

Now, the results recorded are presented for the above scenario. For zone 1 detection and protection, 3 phase fault is given at 1s between bus 4 and 5 (zone 1). The relaying information is collected by PMU at bus 4. Depending on the impedance ($Z=V/I$) between the point of the fault and the point where the relay is located (in this case near bus 4), classifier identifies the zone as zone 1. As a result, distance relay gets activated and the trip signal is sent instantaneously to isolate the line i.e. at 0.0006 sec after the fault. Figure 11 shows the dialog box depicting the classifier output, Figure 12 the trip signal generation, and Figure 13 the trip signal for zone 1. The Figure 14 shows zone 1 current waveform which clearly indicates the faulty line is isolated and protected by this algorithm. To understand zone 2 and zone 3 protection it is assumed that the relay near Bus 7 is blocked such that relay near bus 4 is deliberately allowed to exceed the threshold time and it acts as a backup protection. Now, to simulate zone 2 protection 3 phase fault is given at 1s between bus 5 and 7. Since the fault occurs in zone 2, the relay near bus 4 gets activated by the classifier and the trip signal is sent to the breakers after a time delay of 0.0025s to isolate the faulty line 5-7. The following figures shows the protection of zone 2. Figure 15 shows the dialog box depicting the classifier output, Figure 16 shows the trip signal generation, Figure 17 the trip signal for zone 2, and Figure 18 zone 2 current waveform depicting the line is tripped.

To illustrate zone 3 protection, a three phase fault is applied at 1s between bus 7 and 8. The relay near bus 4 is deliberately allowed to exceed the threshold time and it acts as a backup protection same as in zone 2. Now, the classifier identifies as zone 3 with reference to the relay near bus 4 and the trip signal is sent to the breakers after a time delay of 0.03s to isolate the faulty line 7-8. The following figures shows the

protection of zone 3. Figure 19 shows the dialog box depicting the classifier output, Figure 20 the trip signal generation, Figure 21 the trip signal for zone 3 and Figure 22 the current flow showing tripping. The performance of the classifier in differentiating the persistence of power swing and a fault occurrence in the presence of power swing is shown in Figures 23-25. Tripping of line 7-8 at 0.7s creates power swing which is sensed in the line 5-7 as shown in Figures 5 and 6. This is identified as a power swing condition by classifier 1 as shown in Figure 23. As a result, classifier 2 gets instigated and continues to show the same condition and the relay near bus 4 is blocked till 1s. At 1s, classifier 2 senses the occurrence of the fault as shown in Figure 24 and sends a signal to activate the blocked relay which in turn sends a trip signal to the breaker of the faulty zone. This scenario can be understood from Figure 8 showing the symmetrical fault condition during power swing and from Figure 25 showing the tripping of the faulty line after the fault occurred at the time of power swing.



Figure 11. Fault condition on zone 1



Figure 12. Trip signal generation-zone 1

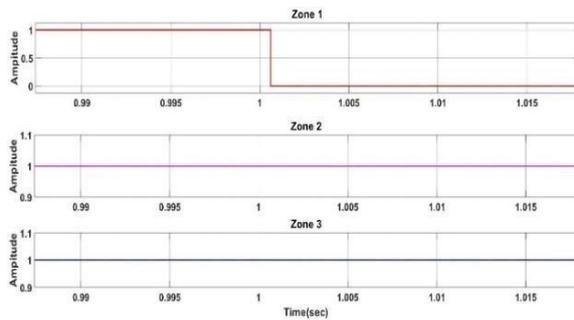


Figure 13. Trip signal for zone 1

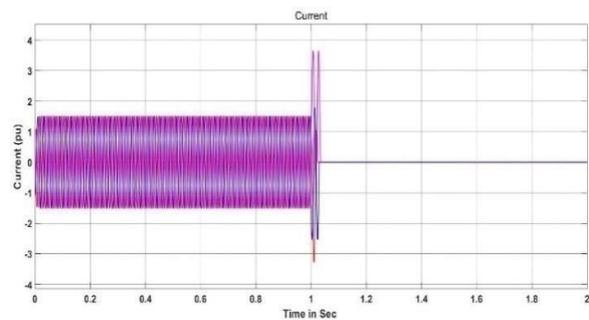


Figure 14. Zone 1 current waveform



Figure 15. Fault condition on zone 2



Figure 16. Trip signal generation-zone 2

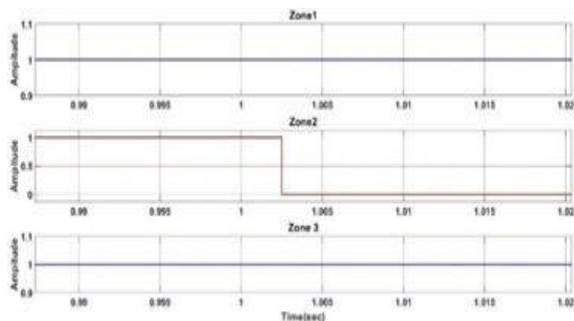


Figure 17. Trip signal for Zone 2

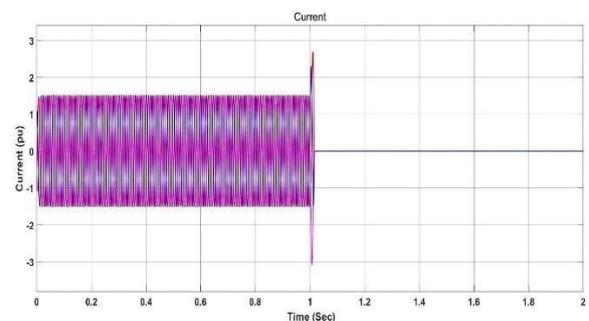


Figure 18. Zone 2 current waveform



Figure 19. Fault condition on zone 3



Figure 20. Trip signal generation-zone 3

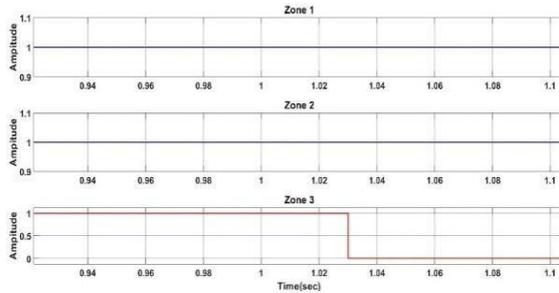


Figure 21. Trip signal for zone 3

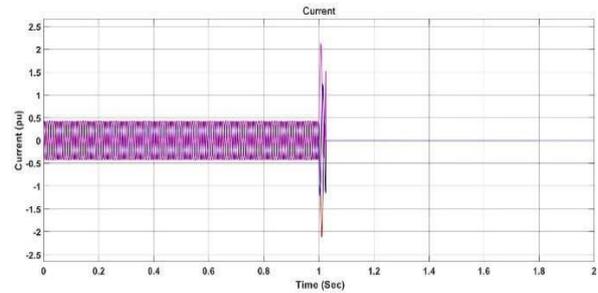


Figure 22. Current waveform of zone 3



Figure 23. Power swing



Figure 24. Power swing with fault condition

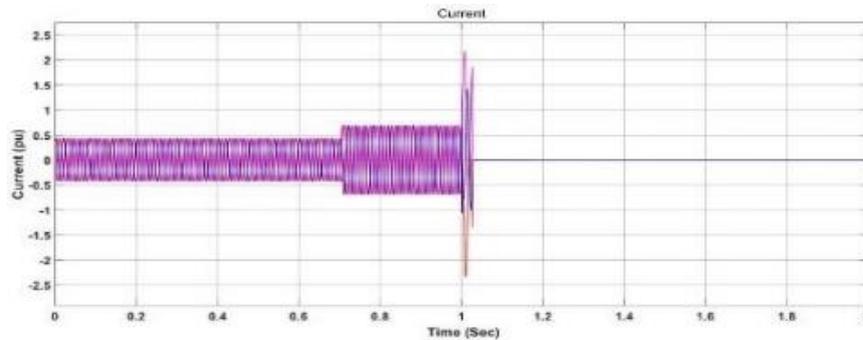


Figure 25. Current waveform for fault during power swing

3.2.2. Case 2: Optimal placement of PMUs in Bus locations 6 and 8.

The performance evaluation of the system has been conducted effectively in the similar manner as case 1 for case 2. After preprocessing, the data has been fed to the various classifier models and evaluated based on the metrics. In this case also, CNN and decision tree performed superior than other classifiers for both the levels. CNN and DT performed consistently for both the cases and KNN performed better compared to case1 for level1. The performance of the protection technique has been tested and found to be coherent to the proposed framework in identifying the system conditions, respective zones and generating the trip signal for protection. Table 5 and Table 6 shows the performance of different classifiers for level 1 and level 2 respectively. Even though it is known that by considering zero injection bus measurements the number of PMUs reduce, this analysis has shown that the parameters accuracy, sensitivity, error rate and specificity are not compromised with the advantage of less computation, complexity, cost and communication infrastructure related issues.

Table 5. Comparison for level 1(case2)

Sr. No	Classifiers	Accuracy (%)	Sensitivity	Error rate (%)	Specificity
1	Decision Tree	100	1	0	1
2	CNN	94.84	1	5.16	1
3	KNN	62.91	0	37.09	1
4	NAIVE BAYES	26.29	0	73.71	1
5	SVM	33.8	0	66.2	1

Table 6. Comparison for level 2(case2)

Sr. No	Classifiers	Accuracy (%)	Sensitivity	Error Rate (%)	Specificity
1	Decision Tree	100	1	0	1
2	CNN	94.44	1	5.56	0.9375
3	KNN	84.72	0.083	15.28	1
4	NAIVE BAYES	83.33	0	16.67	1
5	SVM	83.33	0	16.67	1

3.2.3. Performance analysis with different levels of testing

From the above analysis, wherein testing was done on full training data, it was found that decision tree and CNN were performing better among the five classifiers for both the cases. These models have been further tested by increasing the datasets as well as with different splitting percentage of training and testing (ie. 80%-20%, 70%-30%). Dataset are collected by placing PMUs at locations 4, 7, 9 (case 1) and 6, 8 (case 2). Also, dataset is collected by placing PMUs in all the 5 buses. This is done to do comparative analysis in detail.

1) Data collection:

- a) A total of 446 samples of dataset are collected for the following conditions:
- b) Normal Condition (22 samples)
- c) Bus 4-5: Fault Conditions at Zone 1 (56 samples)
- d) Bus 5-7 and Bus 6-9: Fault Conditions at Zone 2 (112 samples)
- e) Bus 7-8, 8-9: Fault Conditions at Zone 3 (112 samples)
- f) Power Swing Condition (24 samples)
- g) Symmetrical Fault during Power swing (120 samples)

2) Results of different percentage split:

The performance of each of the classifier (CNN and decision tree) for all cases are presented below in the Tables 7-10. The tables represent the training and testing accuracy of different levels of split percentage and for different PMU locations. The results of level 1 classifiers are presented in Table 7 & 8 for both 80% - 20% and 70% - 30% split. Similarly, the results of level 2 classifiers are also presented in Table 9 & 10. From the results, it is clearly understood that for all the cases with different PMU locations, for level1 classifier and level 2 classifier, the training accuracy and testing accuracy marginally improved with the data split of 80% – 20% compared to 70% – 30%. For level 1 and level 2 CNN showed better performance compared to decision tree. Also the performance of CNN is better for large datasets. In the existing system, the efficiency of decision tree is 71% by placing PMU’s on all buses wherein in the proposed research work, the efficiency of Decision tree and CNN is above 80% for 70-30 percentage split and 85% and 90% above for Decision tree and CNN respectively for 80-20 percentage split with PMU’s placed in optimal locations.

Table 7. Result of 80% - 20% (Level-1)

Training Accuracy	Classifiers Models	PMU Locations 4, 7, 9		PMU Locations 6, 8		PMU Locations 4, 7, 9, 6, 8	
		Accuracy(%)	Error rate(%)	Accuracy(%)	Error rate(%)	Accuracy(%)	Error rate(%)
Testing Accuracy	CNN	100	0	99.7	0.3	99.4	0.6
	Decision Tree	96.9	3.1	94.4	5.6	95.8	4.2
Testing Accuracy	CNN	93.3	6.7	94.4	5.6	91	9
	Decision Tree	85.4	14.6	84.3	15.7	87.6	12.4

Table 8. Result of 70% - 30% (Level-1)

Training Accuracy	Classifiers Models	PMU Locations 4, 7, 9		PMU Locations 6, 8		PMU Locations 4, 7, 9, 6, 8	
		Accuracy(%)	Error rate(%)	Accuracy(%)	Error rate(%)	Accuracy(%)	Error rate(%)
Testing Accuracy	CNN	98.7	1.3	98.7	1.3	99	1
	Decision Tree	94.2	5.8	92.9	7.1	93.58	6.4
Testing Accuracy	CNN	82.1	17.9	85.1	14.9	82.8	17.2
	Decision Tree	80.6	19.4	79.1	20.9	79.1	20.9

Table 9. Result of 80% - 20% (Level-2)

Training Accuracy	Classifiers Models	PMU Locations 4, 7, 9		PMU Locations 6, 8		PMU Locations 4, 7, 9, 6, 8	
		Accuracy(%)	Error Rate(%)	Accuracy(%)	Error Rate(%)	Accuracy(%)	Error Rate(%)
Testing Accuracy	CNN	100	0	98.3	1.7	99.1	0.9
	Decision Tree	99.14	0.86	99.1	0.9	98.2	1.8
Testing Accuracy	CNN	96.6	3.4	93.1	6.9	96.6	3.4
	Decision Tree	89.7	10.3	96.6	3.4	93.1	6.9

Table 10. Result of 70% - 30% (Level-2)

Training Accuracy	Classifiers Models	PMU Locations 4, 7, 9		PMU Locations 6, 8		PMU Locations 4, 7, 9, 6, 8	
		Accuracy(%)	Error Rate(%)	Accuracy(%)	Error Rate(%)	Accuracy(%)	Error Rate(%)
Testing Accuracy	CNN	100	0	100	0	100	0
	Decision Tree	99	1	98	2	99	1
Testing Accuracy	CNN	95.3	4.7	90.7	9.3	95.3	4.7
	Decision Tree	88.4	11.6	90.7	9.3	93.1	6.9

4. CONCLUSION

Discriminating faults from stressed conditions (power swing) and distinguishing symmetrical fault during power swing conditions is addressed in this paper. This paper also proposes a data mining technique to operate the distance relay for providing efficient protection to the power system. The IEEE 9-bus Network is taken for study. Two classification levels are employed for discriminating the system conditions and three phase fault in the presence of a power swing. Furthermore, the classifier identifies the zone being affected and activates the distance relay to send trip signal to make the corresponding circuit breaker open. The outcomes from the simulation study define the adaptability of the proposed methodology. The above study has proved that good efficiency can be achieved with less number of PMUs by placing them in optimal locations thus reducing the cost, communication infrastructure related issues and complexity of the system, hence making it economical. The results are further validated by testing for different splitting percentages and it is found that accuracy is not compromised. CNN Classifier is found to be more suitable for the above research problem because of the higher accuracy it has shown for all cases. Thus, the objective of accomplishing good accuracy by reducing the number of PMU's in comparison with the ones addressed in the literature is achieved.

REFERENCES

- [1] Horowitz, S. H., and Phadke, A. G., "Third zone revisited. IEEE Transactions on power delivery," *IEEE Transactions on power delivery*, vol. 21, no. 1, pp. 23-29, 2005.
- [2] Lin, O. Z., and Yin, H. S. "Modelling and Simulation of Mho Type Distance Relay for High Voltage Transmission Line Protection Using MATLAB Software," *JAREE (Journal on Advanced Research in Electrical Engineering)*, vol. 3, no. 2, 2019, doi: 10.12962/j25796216.v3.i2.91.
- [3] Gawande, P., and Dambhare, S., "Secure third zone operation of distance relay using impedance prediction approach," *In 2016 National Power Systems Conference (NPSC) IEEE*, 2016, pp. 1-6, doi: 10.1109/NPSC.2016.7858933.
- [4] Li Y, Zhang C, Wang X, Liu Y, and Shi D., "A Fast Faulted Line Identification Method for Wide-Area Backup Protection with PMU Optimal Placement Strategy," *IET Generation, Transmission & Distribution*, 2018, doi: 10.1049/iet-gtd.2018.5865.
- [5] Samantaray SR and Sharma A., "Enhancing Performance of Wide-Area Back-Up Protection Scheme using PMU assisted Dynamic State Estimator," *IEEE Transactions on Smart Grid*, October 2018, doi: 10.1109/TSG.2018.2874946.
- [6] Sharafi A, Sanaye-Pasand M, and Jafarian P. "Improvement of distance relay zone-3 security using fault and breaker opening generated traveling waves," *International Transactions on Electrical Energy Systems*, vol. 27, no. 10, p. e2414, 2017, doi: 10.1002/etep.2414.
- [7] Jose T, Biswal M, Venkatanagaraju K., Malik OP., "Integrated approach based third zone protection during stressed system conditions," *Electric Power Systems Research*, vol. 161, pp. 199-211, 2018, doi: 10.1016/j.epsr.2018.04.011.
- [8] Neyestanaki, M. K., and Ranjbar, A. M., "An adaptive PMU-based wide area backup protection scheme for power transmission lines," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1550-1559, 2015, doi: 10.1109/TSG.2014.2387392.
- [9] Chen, M., Wang, H., Shen, S., and He, B., "Research on a distance relay-based wide-area backup protection algorithm for transmission lines," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 97-105, 2016, doi: 10.1109/TPWRD.2016.2599198.
- [10] Thakre MP and Kale VS. "An adaptive approach for three zone operation of digital distance relay with Static VarCompensator using PMU," *International Journal of Electrical Power & Energy Systems*, vol. 77, pp. 327-36, 2016, doi: 10.1016/j.ijepes.2015.11.049.

- [11] Phadke AG, Peter WA, Lei DI, and Terzija V., "Improving the performance of power system protection using wide area monitoring systems," *Journal of Modern Power Systems and Clean Energy*, vol. 4, no. 3, pp. 319-31, 2016, doi: 10.1007/s40565-016-0211-x.
- [12] Elghazaly H, Emam A, and Saber A., "A backup wide-area protection technique for power transmission network," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 12, no. 5, pp. 702-9, 2017, doi: 10.1002/tee.22456.
- [13] Saber A, Emam A, and Elghazaly H., "Wide-Area Backup Protection Scheme for Transmission Lines Considering Cross-Country and Evolving Faults," *IEEE Systems Journal*, vol. 9, no. 99, pp. 1-0, May 2018, doi: 10.1109/JSYST.2018.2827938.
- [14] Kang D, and Gokaraju R., "A new method for blocking third-zone distance relays during stable power swings," *IEEE Transactions on Power Delivery*, pp. 1836-43, 2016, doi: 10.1109/PESGM.2017.8273837.
- [15] Boussadia F, and Belkhiat S., "A new algorithm to prevent maloperation of distance protection zone 3 during wide-area disturbances," *International Transactions on Electrical Energy Systems*, vol. 29, no. 1, p. e2670, 2019, doi: 10.1002/etep.2670.
- [16] Das S, and Panigrahi BK., "Real-Time Secured Third Zone Relay Operation Under Dynamic Stressed Conditions," *IEEE Systems Journal*, November 2018, doi: 10.1109/JSYST.2018.2879874.
- [17] Moravej Z, and Bagheri S., "Assessment of the maximum loadability point of a power system after third zone of distance relay corrective actions," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 24, no. 5, pp. 4174-92, 2016, doi: 10.3906/elk-1404-511.
- [18] Nikolaidis VC., "Emergency zone 3 modification as a local response-driven protection measure against system collapse," *IEEE Transactions on Power Delivery*, vol. 31, no. 5, pp. 2114-22, 2016, doi: 10.1109/TPWRD.2016.2552644.
- [19] Nikolaidis VC, Chliara AM., "Emergency Zone 3 modification to prevent cascaded outages: Local vs. centralized approach," 2016, doi: 10.1049/cp.2016.0100.
- [20] Rao, J. G., and Pradhan, A. K., "Supervising distance relay during power swing using synchro-phasor measurements," *IET Generation, Transmission & Distribution*, vol. 11, no. 17, pp. 4136-4145, 2017, doi: 10.1049/gtd.2016.1110.
- [21] Lavand, S. A., and Soman, S. A., "Predictive analytic to supervise zone 1 of distance relay using synchro-phasors," *IEEE Transactions on Power Delivery*, vol. 31, no. 4, pp. 1844-1854, 2016, doi: 10.1049/iet-gtd.2016.1110.
- [22] Dubey, R., Samantaray, S. R., Panigrahi, B. K., and Venkoparao, V. G., "Data-mining model based adaptive protection scheme to enhance distance relay performance during power swing," *International Journal of Electrical Power & Energy Systems*, vol. 81, pp. 361-370, 2016, doi: 10.1016/j.ijepes.2016.02.014.
- [23] Rahul Dubey, Bijay Ketan Panigrahi, Subhransu Ranjan Samantaray, and Sayari Das, "Secured zone-3 protection during power swing and voltage instability: an online approach," *IET Generation, Transmission & Distribution*, vol. 11, no. 2, pp. 437-446, 2017, doi: 10.1049/iet-gtd.2016.0974.
- [24] Roscoe, A., "Frequency tracking, C37. 118.1 a compliant, P and M class PMU (Phasor Measurement Unit) algorithm examples for MATLAB Simulink," 2014.
- [25] Raju, V. V. R., and Kumar, S. J., "An optimal PMU placement method for power system observability," In 2016 *IEEE Power and Energy Conference at Illinois (PECI)*, IEEE, pp. 1-5, 2016, doi: 10.1109/PECI.2016.7459248.

BIOGRAPHIES OF AUTHORS



Ms. M. Kiruthika is a Research Scholar at Fr. C. Rodrigues Institute of Technology, Vashi, Navi Mumbai, Maharashtra, India. She has received her B.E. in ECE from Bharathidasan University, Tamilnadu and M.E. degree in Computer Science Engineering from NIT, Tiruchirapalli, Tamilnadu. She is currently working as an Associate professor at Fr. C. Rodrigues Institute of Technology, Navi Mumbai, Maharashtra, India. Her research interests include Power systems, Data Mining, Distributed Computing and Networking.



Dr. Bindu. S has received her Ph. D. degree from VeermataJijabai Technological Institute (VJTI), Mumbai Maharashtra in 2014. She received her B. Tech. from M G Univ. Kerala and M.E in Power systems from VJTI. She is currently working as Professor and Head, Department of Electrical Engineering at Fr. C. Rodrigues Institute of Technology, Vashi, Navi Mumbai. Her research interests include Power systems, High Voltage Engineering and HVDC. She has good research publications in International Journals, conferences and had taken up various research projects funded by BRNS and University of Mumbai.