ORIGINAL CONTRIBUTION

A Differential Signal-Based Fault Classification Scheme Using PCA for Long Transmission Lines

Alok Mukherjee[1] · Palash Kumar Kundu[2] · Arabinda Das[2]

Abstract Transmission line fault classification is one of the most studied themes of power system analysis and research. This is of utmost importance to isolate the faulted phase for avoiding undue drainage of bulk power during fault. This paper presents a simple and effective method for classification of power system faults in a transmission line using a multivariate statistical method like Principal Component Analysis (PCA). Half-cycle post-fault sending end fault transient current signals are used as the working data for the work, which are normalized, scaled, filtered and finally differentiated. Differentiation of the fault currents is a key method used here in order to highlight particularly the post-fault transient oscillations compared to the original fault transients. These increased oscillations of the modified signals are analyzed using PCA, which extracts fault features in terms of Principal Component scores, which, in turn, are remodeled to develop Principal Component Indices (PCIs). These PCI values as obtained from the analysis of differentiated signals are found higher compared to un-modulated signals in most of the cases; thus yielding more prominent features to distinguish different fault classes, as well as fault and no-fault conditions. The algorithm is tested by varying fault locations at an interval of 10 km. Besides, the proposed scheme is made more practical by incorporating power system noise, as well as varying fault inception angle at intervals of 45°, finally to yield an overall classifier accuracy of 99.41%. This, in turn, validates the robustness of the proposed model.

Introduction

Long transmission lines are one of the most widely expanded engineering systems. These lines spread over miles across different terrains and often come across different natural unpleasant phenomena like strong wind, storm, rain, snow and others. Living creatures like birds, animals, etc., often cause short circuit among the lines or between lines and ground, causing temporary or permanent types of faults. More often, these faults cause major severity to the system. Hence, isolation of the faulted phase is very important to prevent damage to the system, the associated equipments, as well as the working personnel. Besides, huge power outage during fault is required to be restricted with utmost importance. So, immediate discontinuation of power flow is absolutely essential to prevent damage to the system as well as prevent large of power loss. Hence, classification of fault is very essential in order to isolate the faulted line by opening the respective circuit breaker of the affected phases.

Researchers have investigated numerous schemes for the analysis of power system faults: their identification,

✉ Arabinda Das
adas_ee_ju@yahoo.com

Alok Mukherjee
alokmukherjee.ju@gmail.com

Palash Kumar Kundu
palashm.kushi@gmail.com

[1] Government College of Engineering and Ceramic Technology, Kolkata 700010, India

[2] Department of Electrical Engineering, Jadavpur University, Kolkata 700032, India

classification and localization [1]. This paper presents a simple PCA-based approach for classification of power line faults in a long overhead transmission system. PCA is a multivariate statistical analysis which is mostly used to determine the principal directions of variation in a multivariate system. Besides, PCA is well capable of reducing the dimension of a large and spread data set [2]. Hence, PCA is widely used in different engineering fields including power system analysis. PCA is often used as a standalone method of analysis tool or in combination with several other methodologies as a hybrid system [3–12]. Electrical power system is one such multivariate system which consists of several electrical parameters like voltage, current, frequency, etc. These parameters are severely affected during faults. Besides, the three individual phases are affected in different levels for each fault prototype. Hence, PCA becomes a major handful tool, especially for pattern recognition of the fault classes: often alone [3–5] or in combination with other methodologies as hybrid scheme [6–12]. Fault signals are analyzed in this work using PCA-based algorithm to extract key fault features in terms of three-phase PCI. Half-cycle post-fault three-phase current signals of the sending end are used as the test data for the proposed scheme. It is observed that voltage or current waveforms carry very high-frequency transient oscillations immediately followed by fault. These high frequencies are one of the most vital features of faults used for fault diagnosis. In order to enhance these features further, the scaled and filtered signals are differentiated. It is observed that this method of differentiation enhances the oscillations to a higher level compared to the original signal, which contain increased fault features in terms of more pronounces high frequency fault oscillations. Hence, in this work, the method of differentiation is adopted. These highlighted fault features are analyzed using PCA to obtain three-phase PCI for each fault class independently. These PCI on analysis revealed three distinct classes of values, based on which two thresholds are developed, which in turn are modeled in the form of direct classifier rule base. The unknown fault PCI is compared to this rule base table to predict fault class. The major advantage of using PCA in fault researches is its inherent property of identification of the major direction among diversely directed variables [2]. PCA is used effectively to bring out the fault features from the three-phase current signals as Principal Component Indices (PCIs). Since PCA extracts only the key directions of variation, it has the inherent property of reducing the less significant effect of power line noise, which is another advantage of using PCA in this research. Hence, the extent of disturbance caused in each phase during each class of fault is mapped in the PCI values, which are processed to obtain the fault classes.

The proposed work is simple in design incorporating multivariate statistical analysis like PCA. The simulated fault signals are contaminated using externally impregnating noise in order to develop robustness of the model. Besides, the proposed model is investigated by varying the fault locations and fault inception angle (FIA) for introducing practicality of the classifier model as well as develop higher level of robustness to the algorithm. The method of differentiation adopted in this scheme is able to highlight the fault transients more prominently compared to the original signal, which in turn produces more distinct PCI features and enables distinct differentiation among the fault classes. These PCI values are basically indicate the extent of disturbance in each line. Inclusion of PCA as the single method of feature extraction reduces computational burden of the algorithm, even compared to more popular supervised schemes like neural network or mathematical analysis-based schemes like wavelet or fuzzy inference schemes, which mostly suffer from heavier computation. Besides, PCA largely disregards the effect of noise as it concentrates only the most prominent fault signal variations. The threshold-based analysis is also easily implementable, simple, yet very much effective. Altogether, the proposed PCI threshold-based work shows a high accuracy of 99.41% toward classification faults using fault signals simulated with diverse combination of practically varying combinations of fault location, FIA and power line noise. This primarily highlights the practical suitability of the proposed work. Many researchers have used PCA models for transmission line fault analysis, although the proposed method of signal differentiation has rarely been used. It is observed that this differentiated signal contains pronounced fault transient oscillations which are more distinctly identified as well as classified using PCA, which is one of the key features of this work. Besides, non-susceptibility toward practical fault parameter variations and the presence of simulated power line noise further justifies the superiority and robustness of the proposed work.

Survey of Existing Literatures

Researchers have worked on several methodologies for determining accurate fault classes since long [1]. Artificial intelligence-based methods like neural network (ANN), signal analysis-based mathematical method like wavelet transform (WT) and computational method like fuzzy inference system have been most widely practiced methods in this field of research traditionally. ANN and several of its variants have been extensively used in fault diagnosis [13], but being a supervised method, ANN requires healthy level of training using a large and well distributed data set for its successful modeling, which is also subject to considerable training time. Wavelet transform (WT) and

wavelet entropy-based methods have also been applied in numerous researches as a key feature extraction tool [14, 15]. WT is also used with ANN for developing hybrid fault diagnosis to give high level of accuracy, further incorporating wavelet entropy and wavelet packet transform for researches regarding fault analysis very frequently [16–19]. But wavelet analysis suffers from its disadvantage that it becomes progressively intensive in computation, especially for higher level of decomposition. Fuzzy inference system (FIS) has also been highly effective in this regard [20, 21]. Different combinations of these three fundamental methodologies like ANN, WT and FIS have often been combined in several researches for developing efficient hybrid algorithms for complete or partial fault analysis [22–24]. Adaptive neuro-fuzzy inference system or ANFIS methods have come up with accurate algorithm development [24]. Traveling wave-based analysis is one of the few methods which use wave propagation time to detect faults, especially aiding in fault localizer design. These methods of analysis have been acquired effectively in many researches, especially for distance estimation of faults apart from classification [11, 12, 25, 26], but the detection time of fault is mostly dependent on the propagation time and hence, is a variable quantity. Support vector machine (SVM) has been another major approach in fault analysis for its ability of zone segmentation, which influenced many researchers to design SVM-based transmission line fault analysis schemes: either as a major standalone feature extractor [27–29] or as a hybrid model incorporating other analysis methods like discrete orthogonal S-transform or DOST [30], wavelet analysis [31] and supervised analysis like neural network [32] to develop excellent fault features. PCA-based schemes are also popular among fault analysis and research. PCA is often combined with several other methodologies like SVM [10], WT-aided traveling wave [11, 12] or a combination of WT, ANN, SVM and others [6, 7] to develop fault analysis algorithms. Probabilistic neural network (PNN), being a major variant of neural network, is very effective for pattern recognition problems [13, 33]. Power system faults, being one such kind, have huge applications of these methodologies. PCA is often combined with PNN for developing effective fault diagnosis schemes [8, 9]. Hence, the overall power system fault analysis is researched over using several schemes.

Apart from these major conventional schemes, several other methods have come up in latest researches. Phasor measurement unit (PMU) has been one of the latest additions to the power system research which includes computation of magnitude and phase angle of electrical parameter using common time source for synchronization [34, 35], although the cost of implementation becomes higher due to the incorporation of additional hardware

support, apart from the computational heaviness. Ensemble Kalman filter-based analysis has been applied in similar research [36]. Sequence analysis of voltage or current parameters has been traditionally applied in research, although these have been investigated with latest computational analysis with positive and negative sequence circuit-based analysis [37, 38]. ANN methods have been investigated extensively in recent few years to develop machine learning and deep learning algorithm, which have come up with major challenges to existing schemes. Extreme learning machine (ELM)-based neural networks have been investigated as a part of such research [39]. Other artificial intelligence-based methods like genetic algorithm too have been applied as hybrid combination with neural networks to develop GA-GNN [40]. Decision tree, a powerful statistical-mathematical method, has also found application in few studies [41]. Frequency domain analysis including Fourier transform-based power spectrums is very useful in fault analysis. The high-frequency fault transients bear significant fault information which is investigated in several researches [42]. Recent developments in fault analysis include mathematical morphology-based feature extraction filtering [43]. These altogether gives an overview of the trend of contemporary fault analysis researches.

Methods and Analysis

System Design and Data Collection

A 132 kV, 50 Hz, 150 km long transmission line is stimulated in EMTP/ATP software, similar to the one developed in [9]. The simulated model is shown in Fig. 1. Different classes of faults have been simulated at different intermediate points. The sending end phase current signals are collected as the working data for the proposed work. The data so collected are normalized followed by scaling with respect to no-fault signal. These signals are filtered
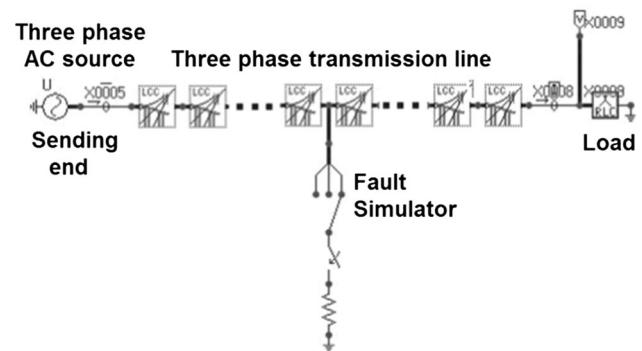

Fig. 1 Simulated transmission line model in EMPT/ATP

further to reduce the effects of power line noise to some extent. The sampling frequency is kept at 100 kHz, i.e., each cycle of the nominal frequency produces 2000 samples.

The faults are segmented into four major classes, each of which are divided further into different classes to constitute the ten different fault prototypes as follows: single line to ground faults or SLG (contains AG, BG and CG faults), double line faults or DL (contains AB, BC and CA faults), double line to ground faults or DLG (contains ABG, BCG and CAG), and LLL, i.e., ABC faults. The proposed work uses only a half-cycle post-fault signal for necessary analysis, which means the length of the vector of each phase signal becomes 1000 samples. Thus, the input vector of each set of fault signal becomes:

$$[Xa_i] = [Ia_i Ib_i Ic_i]_{1000 \times 3};$$

where Ia, Ib and Ic are the sending end three-phase fault signals and i indicates the fault class; hence, $i = 1, 2, 3, …, 11$, i.e., ten fault classes and one healthy condition.

Development of Idea of Signal Modulation

The proposed work uses the three-phase normalized, scaled and filtered current signals, denoted by $f(nT)$, for development of the classifier model. Instead of using the signals directly, these are differentiated to develop the modified working signal $d(nT)$. The idea behind such analysis lies in the fact that the edges of the high-frequency post-fault transient oscillations of the current signals are found to enhance further on differentiation. These enhanced oscillations create a high difference between the fault and no-fault conditions, as well as the individual fault classes. These are highlighted difference between the faulted and non-faulted line signals which are analyzed through the PCA algorithm. It is known that PCA reduces the dimensionality of a signal by finding out the most important directions of variation in the descending order of importance in terms of the principal components (PC). In this work, we have considered only the two primary PCs for the analysis, from which the PCIs are developed, which are primarily a measure of the extent of disturbance of the line currents from their respective healthy conditions. Hence, the enhanced oscillations of the differentiated signals $d(nT)$ signals are also observed accordingly from the PCI values. These PCI values are primarily the vector the measures of the disturbances of each of the signals from the healthy condition. It is further required to mention here that in this work the large data size of 1000 samples per signal corresponding to half-cycle post-fault signal reduces to a simple scalar PCI value only from the analysis of the two most major PCs. Hence, PCA in our work effectively reduces the dimension of the large current data set

associated with fault, thus reducing the memory requirement as well.

It is very important to smooth out the noise contamination to the highest possibility, without losing vital information regarding the fault oscillations. Differentiation enhances the edges of oscillation in a signal. Noise has an inherent nature of prompt and complete random change in magnitude. Hence, these noise signals, on differentiation, are also prone to cause abrupt changes in signal, which may even hide the highlighted post-fault transient oscillations. Thus, selection of the cut-off frequency of filter is highly important for this work. Spectrum analysis shows that noise frequencies are very high, even compared to most of the high-frequency post-fault transients. Hence, the cut-off frequency of the low-pass filter is selected such a way that it only allows even the highest frequency of the fault transients observed for the 150 km long line and discards most of the noise contamination.

Analysis of Modulated Fault Signals

The differential line A currents $d(nT)$ are shown in Fig. 2 for different categories of faults, as an example case. It is evident that the differentiation of signals enhances the fault transient edges. It is further observed that in case of AB, AG and ABG faults, line A is the directly affected; hence, the transients are most prominent and of high magnitude. Again, the $d(nT)$ for line A for BC fault and the healthy condition are much similar and hardly possess any disturbances other than noise. Finally, in case of BG fault, the directly un-faulted line A produces higher disturbance compared to the same for BC fault, but less than AG fault, immediately after the fault.

It is understood that the directly affected lines, since are disturbed the most, produce maximum value of the PCI. The un-faulted line in case of double line faults behaves much like the healthy phases. The un-grounded lines in case of ground faults are disturbed to higher extent than the un-faulted lines of DL faults. These phenomena are clearly observed from the sub-figures of Fig. 2. These un-faulted lines of ground faults like SLG or DLG are indirectly connected to ground by the grounded neutral and the directly grounded lines. This causes the zero sequence currents to flow through these un-grounded lines too, causing higher disturbances in these lines compared to the third line of DL fault, where ground involvement is nearly absent. This causes almost no such disturbances in the un-faulted lines. This distinction is also observed from the BC (DL fault) and BG (SLG fault) of Fig. 2, where phase A behaves as the un-faulted phase in both the cases. These inferences are interpreted mathematically in terms of the PCI values, based on which some fault classifier rules are set, followed by development of fault signature table. The
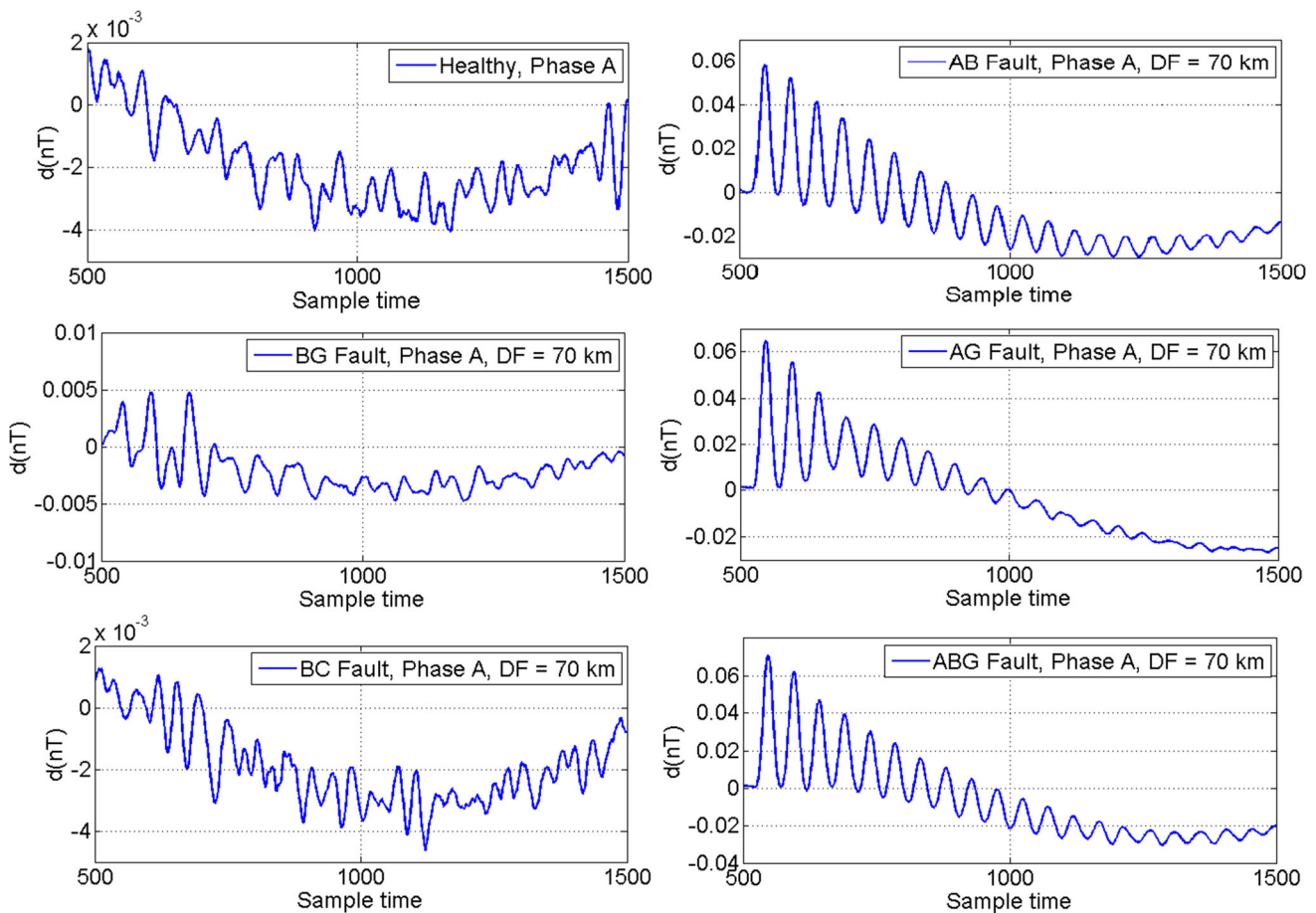
Fig. 2 Immediate post-fault differential signal $d(nT)$ of line currents for phase A under different fault conditions, for fault at 70 km from the sending end

unknown fault is also analyzed similarly which produces three-phase test PCI values. These three-phase test PCI values are compared directly with the fault signatures to predict fault classes.

Numerical Analysis of PCI

The ten different fault classes for faults carried out at three intermediate locations along the length of line and the three-phase PCI values are obtained which are described in Table 1. The above discussion, in connection with the observations from Fig. 2, is well interpreted in terms of the PCI values from this table. It is found from the combined analysis of these two that that the PCI values are basically measure of the extent of disturbance caused during fault in each of the line currents. It is mentioned already that an attempt has been made to make the proposed method more robust by incorporating power system noise as well as by varying the fault geometric location and fault inception angle. Hence, for the sake of initial analysis of the PCI values, the FIA is kept fixed at 90° and fault location is

only varied in three intermediate locations: 30 km, 70 km and 110 km. Level of noise is also kept at 30 dB, which is quite often a near practical situation. The PCI values are obtained as described in Table 1 using these fault parameters.

It is duly observed that the directly affected lines in case of all fault classes produce a very high PCI and the un-faulted lines of DL faults produce the least PCI, which is near to no-fault value. The un-faulted lines of ground faults like DLG or SLG produce PCI values higher than the un-faulted lines of DL faults, although this difference is narrow. Hence, the DL and DLG faults are required to be separated using some suitable method. Faults have been carried out for three times for each fault class and at each of the fourteen intermediate locations. These repetitions are done in order to reduce the effect of randomness due to noise. Hence, a total of (14 locations × 3 repetitions) i.e., 42 unknown conditions of each fault class, are tested with this proposed algorithm. It is further observed that the PCI of the directly affected phases is almost monotonic in nature with variation in fault location, following a certain pattern. Hence, the two terminal locations of the line, i.e.,

Table 1 Three-phase PCI of $d(nT)$ for ten fault prototypes for faults at locations 30 km, 70 km and 110 km, fault resistance: 1Ω, fault inception angle (FIA): 90°

Fault location	Fault location 30 km			Fault location 70 km			Fault location 110 km		
Fault class	PCI_A	PCI_B	PCI_C	PCI_A	PCI_B	PCI_C	PCI_A	PCI_B	PCI_C
AG	1281.4	18.619	18.475	593.06	23.77	26.311	429.37	24.482	28.654
BG	7.4539	672.35	7.3784	9.6177	392.55	12.142	11.858	288.01	17.341
CG	11.013	11.117	795.85	16.161	16.047	405.74	15.06	15.636	271.78
AB	929.56	930.81	1.4718	612.57	616.56	2.0751	518.14	525.59	0.7527
BC	1.1659	680.27	680.39	1.5036	392.01	393.26	0.4426	277.7	282.49
CA	1055.6	0.6206	1057.3	610.29	1.8019	615.07	463.9	1.5968	483.19
ABG	1088.3	870.78	9.9127	682.64	585.1	13.44	560.45	507.54	10.469
BCG	9.7413	688.68	883.01	18.365	439.66	456.14	15.05	337.91	315.85
CAG	1195.5	6.5369	980.32	688.93	8.9181	582.32	528.86	8.846	461.36
ABC	1245.8	827.53	1009.9	773.15	553.12	558.07	644.38	463.54	397.72

10 km and 140 km, are appended with the three above-mentioned locations of 30 km, 70 km and 110 km in order to enhance the accuracy of the proposed algorithm. The highest and the lowest PCIs for each of the major classes of faults obtained at these five locations, for two successive repetitions at each location, are described in Table 2. The values which are of least importance, i.e., which are not required for the development of the scheme, are described as NR in the table.

The FIA is altered next for testing the above limiting values of each of the fault classes. It is rather found that these limiting values are altered to some extent with variation in FIA. It is needful to mention here that since all the quarters of a complete sinusoidal waveform are symmetric, FIA has been varied in the range from 0° to 90° in three discrete steps. The PCI obtained for SLG-AG fault for FIA 0°, 45° and 90° at these three locations and with the same 30 dB noise level is shown in Table 3 as follows. These values are observed carefully to note that PCI values vary almost in proportion to the values we had obtained for 90° only; although the PCI values vary with variation in FIA values. PCI values seem to reduce gradually as the FIA

closes towards 0°, primarily for the directly affected lines which is line A in this case. PCI of the indirectly affected lines, i.e., line B and C, does not show any such prominent trend. Hence, considering all the ten fault prototypes, the limiting values of PCI for different major classes of faults, as described in Table 2, are further modified in Table 4. The PCI values obtained from analyzing the terminal points of the line, i.e., 10 km and 140 km, are also considered while constructing Table 4, as done during construction of Table 2. These limiting values corresponding to Table 4 are considered as the final limiting values for the faulted and un-faulted line.

Analysis of PCI Leading to the Development of Two Threshold Values

The four major observations of PCI values in Table 4 are marked in bold. These reveal the following facts:

i. The lowest PCI of the direct faulted line is obtained for SLG fault among all classes and it measures **180.451**.
ii. The highest PCI for the un-faulted line among ground faults (SLG and DLG) is **84.574** obtained for SLG fault.
iii. The least PCI for the un-faulted line of ground faults (SLG and DLG together) is **6.017** obtained for SLG fault.
iv. The highest PCI obtained for the un-faulted lines of DL faults is **3.463**.

These PCIs corresponding to the four major fault categories are deterministic for the development of the proposed classifier. A factor of safety of 25% is further applied over these values to incorporate more reliability and stability of the classifier scheme. Hence, applying the

Table 2 The limiting values of PCI for different major classes of faults

Fault class	Faulted line		Un-faulted line	
	Highest PCI	Lowest PCI	Highest PCI	Lowest PCI
SLG	NR	**216.875**	**31.574**	6.652
DL	NR	235.183	**3.251**	NR
DLG	NR	253.372	22.485	**6.273**
LLL	NR	325.641	–	–

Table 3 Three-phase PCI of $d(nT)$ for SLG-AG fault for faults at locations 30 km, 70 km and 110 km, fault resistance: 1Ω, fault inception angle (FIA): 0°, 45°, 90°

Fault location	Fault Inception Angle (FIA) (Degree)								
	0 PCI-A	45	90	0 PCI-B	45	90	0 PCI-C	45	90
30	1024.1221	1229.9626	1278.514	6.875	28.1256	20.485	7.486	20.145	28.475
70	478.154	560.145	588.457	8.458	19.458	27.184	10.458	32.745	20.164
110	324.821	371.245	431.247	12.485	25.145	32.147	6.984	19.458	35.475

Table 4 The modified final limiting values of PCI for different major classes of faults

Fault class	Faulted line		Un-faulted line	
	Highest PCI	Lowest PCI	Highest PCI	Lowest PCI
SLG	NR	**180.451**	**84.574**	**6.017**
DL	NR	192.782	**3.463**	NR
DLG	NR	201.645	76.354	6.132
LLL	NR	249.418	–	–

Table 5 Rule base for fault classification: fault classifier table

PCI Fault line status	PCI ≤ ΔL No-fault	ΔL < PCI < ΔH Un-faulted line of ground fault	PCI ≥ ΔH Direct fault	Predicted Fault Class
Number of PCI values in this range	3	0	0	No-fault
	0	0	3	LLL
	1	0	2	DL
	0	1	2	DLG
	0	2	1	SLG

mentioned 25% tolerance limit, the modified boundary limits are reformed as described next:

Case 1: If a line has PCI higher than (*180.451– 25% of 180.451*), i.e., **135.338**, it is assumed to be directly affected.

Case 2: If a line has PCI lesser than (*3.463 + 25% of 3.463*), i.e., **4.329**, it is assumed to be indirectly affected in case of DL fault, or under no-fault.

Case 3: If a line has PCI in between **4.329** and **135.338**, it may be treated as un-grounded line of ground fault.

In order to further verify these two levels, the obtained limiting values of the directly un-grounded line of ground faults are tested. These limiting values are found from Table 4 as **84.574** and **6.017**, including both SLG and DLG faults. Thus, applying the tolerance limits of 25% again, the limiting PCIs for these two limits become:

Fig. 3 Selection of lower and higher thresholds (with 25% tolerance) for the proposed classifier
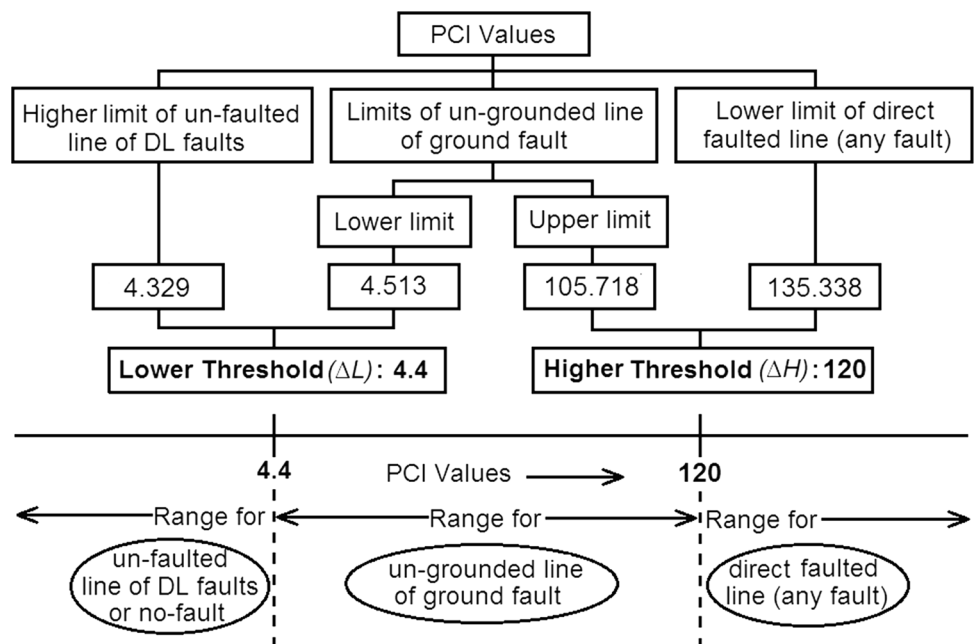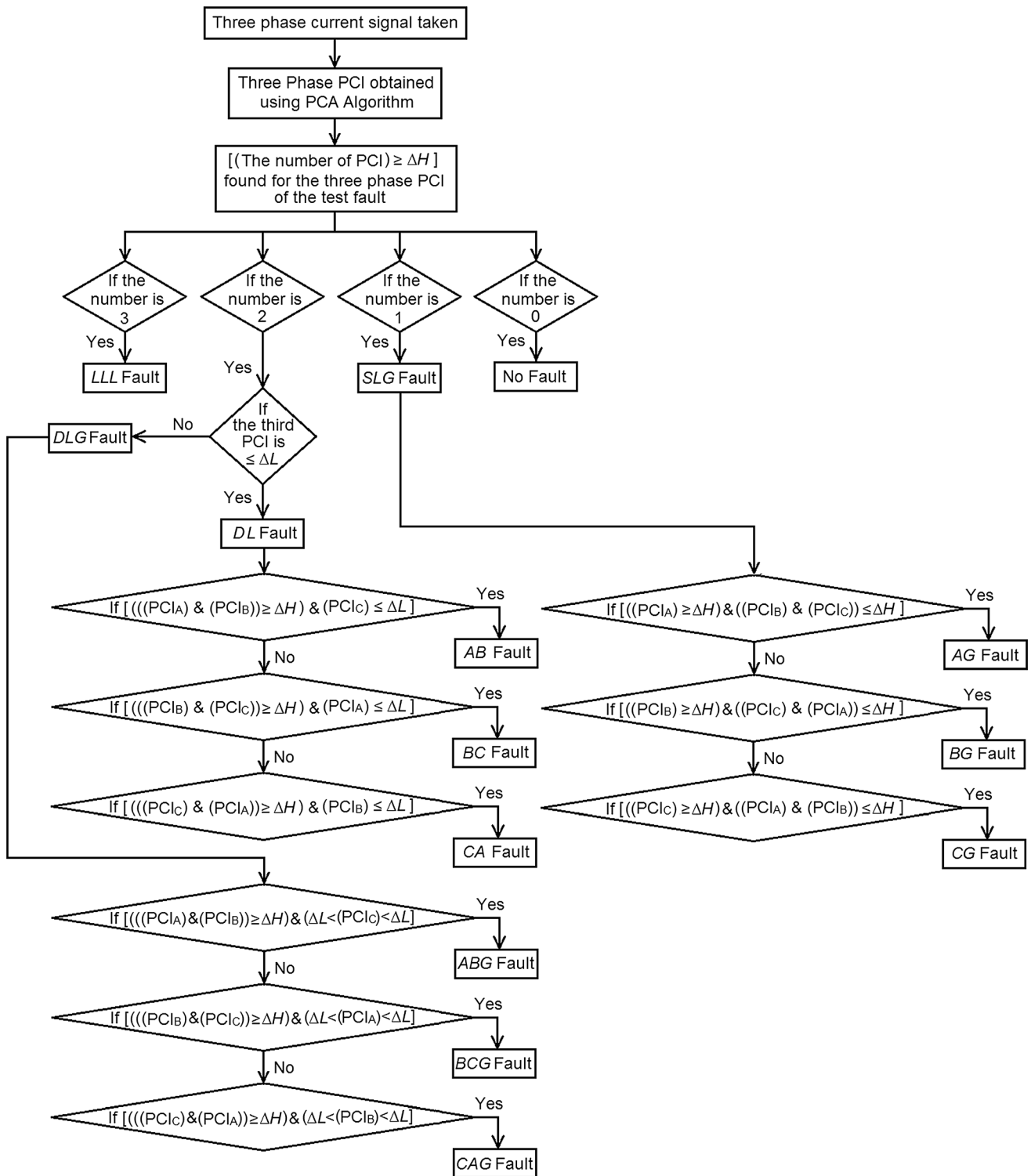
Fig. 4 Flowchart of the proposed classifier algorithm

Lowest value with tolerance: $(6.017 - 25\%\ of\ 6.017)$ i.e., **4.513**
Highest value with tolerance: $(84.574 + 25\%\ of\ 84.574)$ i.e., **105.718**

Selection of threshold limits

It is observed that the higher limit of the un-faulted lines of the ground faults obtained above with tolerance, i.e.,

Table 6 Overall classification accuracy of the proposed classifier

Fault class	SLG	DL	DLG	LLL	Overall Accuracy
No. of observations	252	252	252	84	840
Correct prediction	252	249	250	84	835
Wrong prediction	0	3	2	0	5
Classifier accuracy	100	98.81	99.21	100	99.41

Table 7 Detailed analysis of classifier results for each class of fault

True class	Predicted class										
	No fault	AG	BG	CG	AB	BC	CA	ABG	BCG	CAG	ABC
No fault	20										
AG		84									
BG			84								
CG				84							
AB					83			1			
BC						82			2		
CA							84				
ABG					1			83			
BCG									84		
CAG							1			83	
ABC											84

105.718, is much less than the limit obtained in case 1 as *135.338*. Again, the lower limit of un-faulted lines of the ground faults is found as *4.513*, which is higher than that obtained in case 2, i.e., *4.329*. Hence, both these limits lie well within the in-between limits of **fault** and **no-fault** levels, i.e., *4.329* and *135.338*. Hence, two thresholds: *lower threshold* (ΔL) and the *higher threshold* (ΔH), are selected based on these ranges. It could be inferred from Fig. 3 that the *higher threshold* or ΔH must lie to segregate PCI values of two fault levels: directly faulted line having highest disturbance level and un-grounded line of ground faults, which has lesser level of disturbance. This ΔH is instrumental to identify the direct faulted line. Hence, any line having PCI higher than ΔH is indicated to be directly faulted. Similarly, the *lower threshold* (ΔL) is used to distinguish between the un-grounded line of ground fault and un-faulted line of DL faults (or no-fault). Hence, any line having PCI less than ΔL is indicated as belonging to either un-faulted line of DL faults or no-fault class. Hence, it could be summarized that these two thresholds ΔH and ΔL signify, respectively, the predicted minimum level PCI of the directly faulted line and the predicted maximum level of the un-faulted line of DL faults or no-fault PCI. The preceding discussion regarding selection of these two thresholds is illustrated diagrammatically in Fig. 3.

Fault Classifier Algorithm

The two thresholds are selected from the above analysis as well as from the diagram of Fig. 3. ΔH, the higher threshold or the direct fault threshold, indicates that if a line has PCI more than ΔH, the line may be treated as directly faulted. ΔL, the lower threshold or the no-fault threshold, indicates that if a line has PCI below ΔL, that line may be treated as no-fault line or the indirectly affected line of DL fault. Finally, a line with PCI in between ΔL and ΔH is treated as the un-faulted line of any ground fault. In case of SLG faults, only a single line is affected directly which produces maximum disturbance, and the other two lines are affected in minor amount, which lies in the intermediate level of disturbance. For DLG faults, similar conditions prevail, only the number of directly affected lines becomes 2 and the intermediate disturbance class becomes 1. For DL faults, only difference from DLG faults arises in the fact that instead of 1 number of intermediate disturbance class, DL faults have a single almost un-faulted line, and there are 2 directly affected lines similar to DLG faults. Finally, for LLL fault, all the lines are affected directly. Based on the above propositions, Table 5 is constructed which defines some rule bases for classifying the faults. This is denoted as fault classifier

412

J. Inst. Eng. India Ser. B (June 2021) 102(3):403–414

table. This shows the inroads to the development of the complete fault classifier with broader classes. The subsequent narrow classes for each broad class are found using the flowchart of Fig. 4, which describes the complete analysis of the three-phase PCI values used for the development of the proposed transmission line fault classifier model.

Results and Discussion

The proposed PCA-based classifier is tested for all the ten classes of faults at fourteen intermediate locations. Faults have been carried out at locations starting from the 10 km and extending up to 140 km, each 10 km apart. Noise often causes randomness to the results; hence, the classifier is tested twice at each location for each set of fault signal. As mentioned before, each class of fault is tested at all the 14 locations, with 2 repetitions at each fault location. Besides, at each intermediate location, each fault class is conducted with three random set of fault inception angle ranging in between 0° and 90°. Thus, the set of test data comprises of (14 location × 3 FIA × 2 repetitions), i.e., 84 fault signals for each fault class. Thus, total test data set contains 840 set of signals considering all ten fault classes. The accuracy of the proposed classifier is given in Table 6 categorically.

Only 5 observations among the 840 results are found wrong, producing an overall efficiency of 99.41%. Three observations from DL class and two observations from DLG class of fault are found to be predicted wrongly. The detailed distribution of the classifier results as per the fault classes is shown in Table 7.

It is observed from Table 7 that in case of each of the wrong results, the DL fault has been wrongly classified as the corresponding DLG fault and vice versa. For example, a BC fault is misclassified as BCG and a ABG fault is misclassified as AB, as shown in red in Table 5. But most importantly, in both of these cases, the same sets of circuit breakers are required to be operated instantly, irrespective of the DLG or DL fault. The ground involvement doesn't majorly affect the lines to be isolated. Hence, this misclassification is less significant in respect to classification error.

Conclusion

An effective fault classifier model is developed in this work using Principal Component Analysis. The sending end three-phase line currents are analyzed using PCA to develop the Principal Component Indices (PCIs) of each line and each fault class. Fault locations are varied along the whole span along with the fault inception angle. These PCIs of individual phases are analyzed to develop a fault classifier table, from which the unknown fault is classified directly.

This proposed classifier uses only PCA to develop the algorithm. PCA is very effective in reducing the dimension of the large data set associated with fault. This PCA brings out only the major directions, which are the fault features for the proposed work. Supervised learning methods require large time for training and wavelet transform becomes progressively complicated with an increase in the level of decomposition. Hence, PCA becomes an effective choice for the purpose with less complexity of analysis. The highlights of our work are described as follows:

a) The proposed method is very simple, computationally light, yet very effective for fault classification.
b) Two major parameters of fault: fault location and fault inception angle, are varied.
c) Fault location is varied from 10 to 140 km of the line span at an interval of 10 km.
d) Fault inception angle (FIA) is also varied in the range 0°–90°, which constitute one symmetrical quarter of a sinusoidal current signal.
e) Practical power line noise of 30 dB is incorporated in simulated signals to introduce practicality to the analysis and develops robustness in the algorithm.
f) The present work requires only half-cycles of post-fault current transient signal, which is comparable or even better than many of the existing methods.
g) The proposed method uses differential-based signal modulation. This method is found to highlight the transient fault features more prominently, which on analysis using PCA produces PCIs which are more distinctly identifiable, especially for segmenting the same into major three-level disturbance classes: directly affected, indirectly affected for ground faults and indirectly affected for line faults. Based on it, a simple classifier topology is developed; hence, reduced analysis is achieved with enhanced distinct fault features.
h) In this work, the proposed PCA algorithm considers only the two most important directions of variation, i.e., principal components (PC), and thus reduces the dimension of the fault current data to unified PCI values, thus further reducing the memory requirement.
i) The overall classifier accuracy achieved is 99.41% considering all the above factors, which is considerably high in accuracy.

Besides, the method is very effective to suit practical environment by extracting fault features from the noisy fault signals post-filtering, which indicates the practical acceptability of the method. Thence, the proposed method

shows a simple as well as practical and effective way for transmission line fault classification.

References

1. A. Prasad, J.B. Edward, K. Ravi, A review on fault classification methodologies in power transmission systems: part—I. J. Electr. Syst. Inf. Technol. **5**(1), 48–60 (2018)
2. L.I. Smith, A tutorial on principal components analysis (2002)
3. A. Mukherjee, P. Kundu, A. Das, Identification and classification of power system faults using ratio analysis of principal component distances. Indones. J. Electr. Eng. Comput. Sci. **12**(11), 7603–7612 (2014)
4. A. Mukherjee, P.K. Kundu, A. Das, Power system fault identification and localization using multiple linear regression of principal component distance indices. Int. J. Appl. Power Eng. **9**(2), 113–126 (2020)
5. Q. Alsafasfeh, I. Abdel-Qader, A. Harb, Symmetrical pattern and PCA based framework for fault detection and classification in power systems, in *2010 IEEE International Conference on Electro/Information Technology* (pp. 1–5). IEEE (2010)
6. J.A. Jiang, C.L. Chuang, Y.C. Wang, C.H. Hung, J.Y. Wang, C.H. Lee, Y.T. Hsiao, A hybrid framework for fault detection, classification, and location—part I: concept, structure, and methodology. IEEE Trans. Power Deliv. **26**(3), 1988–1998 (2011)
7. J.A. Jiang, C.L. Chuang, Y.C. Wang, C.H. Hung, J.Y. Wang, C.H. Lee, Y.T. Hsiao, A hybrid framework for fault detection, classification, and location—part II: implementation and test results. IEEE Trans. Power Deliv. **26**(3), 1999–2008 (2011)
8. A.K. Sinha, K.K. Chowdoju, Power system fault detection classification based on PCA and PNN, in *2011 International Conference on Emerging Trends in Electrical and Computer Technology* (pp. 111–115). IEEE (2011)
9. A. Mukherjee, P.K. Kundu, 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. Instit. Eng. (India) Ser. B **101**(4), 321–333 (2020)
10. Y. Guo, K. Li, X.Liu, Fault diagnosis for power system transmission line based on PCA and SVMs, in *International Conference on Intelligent Computing for Sustainable Energy and Environment* (pp. 524–532). Springer, Berlin, Heidelberg (2012)
11. E. Vázquez-Martínez, A travelling wave distance protection using principal component analysis. Int. J. Electr. Power Energy Syst. **25**(6), 471–479 (2003)
12. P. Jafarian, M. Sanaye-Pasand, A traveling-wave-based protection technique using wavelet/PCA analysis. IEEE Trans. Power Deliv. **25**(2), 588–599 (2010)
13. N. Roy, K. Bhattacharya, Detection, classification, and estimation of fault location on an overhead transmission line using S-transform and neural network. Electr. Power Compon. Syst. **43**(4), 461–472 (2015)
14. S.C. Shekar, G. Kumar, S.V.N.L. Lalitha, A transient current based micro-grid connected power system protection scheme using wavelet approach. Int. J. Electr. Comput. Eng. **9**(1), 14 (2019)
15. S.P. Valsan, K.S. Swarup, Wavelet transform based digital protection for transmission lines. Int. J. Electr. Power Energy Syst. **31**(7–8), 379–388 (2009)
16. P.S. Bhowmik, P. Purkait, K. Bhattacharya, A novel wavelet transform aided neural network based transmission line fault analysis method. Int. J. Electr. Power Energy Syst. **31**(5), 213–219 (2009)
17. A. Swetapadma, A. Yadav, All shunt fault location including cross-country and evolving faults in transmission lines without fault type classification. Electr. Power Syst. Res. **123**, 1–12 (2015)
18. A. Yadav, A. Swetapadma, A single ended directional fault section identifier and fault locator for double circuit transmission lines using combined wavelet and ANN approach. Int. J. Electr. Power Energy Syst. **69**, 27–33 (2015)
19. A. Dasgupta, S. Nath, A. Das, Transmission line fault classification and location using wavelet entropy and neural network. Electr. Power Compon. Syst. **40**(15), 1676–1689 (2012)
20. Z. Jiao, R. Wu, A new method to improve fault location accuracy in transmission line based on fuzzy multi-sensor data fusion. IEEE Trans. Smart Grid **10**(4), 4211–4220 (2018)
21. A. Yadav, A. Swetapadma, Enhancing the performance of transmission line directional relaying, fault classification and fault location schemes using fuzzy inference system. IET Gener. Transm. Distrib. **9**(6), 580–591 (2015)
22. M.J. Reddy, D.K. Mohanta, A wavelet-fuzzy combined approach for classification and location of transmission line faults. Int. J. Electr. Power Energy Syst. **29**(9), 669–678 (2007)
23. R. Goli, A.G. Shaik, S.S.T. Ram, Fuzzy-wavelet based double line transmission system protection scheme in the presence of SVC. J. Inst. Eng. (India) Ser. B **96**(2), 131–140 (2015)
24. H. Eristi, Fault diagnosis system for series compensated transmission line based on wavelet transform and adaptive neuro-fuzzy inference system. Measurement **46**(1), 393–401 (2013)
25. F.V. Lopes, K.M. Dantas, K.M. Silva, F.B. Costa, Accurate two-terminal transmission line fault location using traveling waves. IEEE Trans. Power Deliv. **33**(2), 873–880 (2017)
26. S. Hasheminejad, S.G. Seifossadat, M. Razaz, M. Joorabian, Traveling-wave-based protection of parallel transmission lines using Teager energy operator and fuzzy systems. IET Gener. Transm. Distrib. **10**(4), 1067–1074 (2016)
27. S. Ekici, Support Vector Machines for classification and locating faults on transmission lines. Appl. Soft Comput. **12**(6), 1650–1658 (2012)
28. B.Y. Vyas, R.P. Maheshwari, B. Das, Pattern recognition application of support vector machine for fault classification of thyristor controlled series compensated transmission lines. J. Inst. Eng. (India) Ser. B **97**(2), 175–183 (2016)
29. P. Jafarian, M. Sanaye-Pasand, High-frequency transients-based protection of multiterminal transmission lines using the SVM technique. IEEE Trans. Power Deliv. **28**(1), 188–196 (2012)
30. M.J.B. Reddy, P. Gopakumar, D.K. Mohanta, A novel transmission line protection using DOST and SVM. Eng. Sci. Technol. Int. J. **19**(2), 1027–1039 (2016)
31. B. Bhalja, R.P. Maheshwari, Wavelet-based fault classification scheme for a transmission line using a support vector machine. Electr. Power Compon. Syst. **36**(10), 1017–1030 (2008)
32. S.R. Samantaray, P.K. Dash, G. Panda, Distance relaying for transmission line using support vector machine and radial basis function neural network. Int. J. Electr. Power Energy Syst. **29**(7), 551–556 (2007)
33. Z. Moravej, J.D. Ashkezari, M. Pazoki, An effective combined method for symmetrical faults identification during power swing. Int. J. Electr. Power Energy Syst. **64**, 24–34 (2015)
34. P. Gopakumar, M.J.B. Reddy, D.K. Mohanta, Fault detection and localization methodology for self-healing in smart power grids incorporating phasor measurement units. Electr. Power Compon. Syst. **43**(6), 695–710 (2015)
35. S. Barman, B.K.S. Roy, Detection and location of faults in large transmission networks using minimum number of phasor

measurement units. IET Gener. Transm. Distrib. **12**(8), 1941–1950 (2018)

36. R. Fan, Y. Liu, R. Huang, R. Diao, S. Wang, Precise fault location on transmission lines using ensemble Kalman filter. IEEE Trans. Power Deliv. **33**(6), 3252–3255 (2018)

37. A. Ghorbani, H. Mehrjerdi, Negative-sequence network based fault location scheme for double-circuit multi-terminal transmission lines. IEEE Trans. Power Deliv. **34**(3), 1109–1117 (2019)

38. L. Ji, X. Tao, Y. Fu, Y. Fu, Y. Mi, Z. Li, A new single ended fault location method for transmission line based on positive sequence superimposed network during auto-reclosing. IEEE Trans. Power Deliv. **34**(3), 1019–1029 (2019)

39. Y.Q. Chen, O. Fink, G. Sansavini, Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction. IEEE Trans. Industr. Electron. **65**(1), 561–569 (2017)

40. S.K. Sharma, GA-GNN (Genetic Algorithm-Generalized Neural Network)-based fault classification system for three-phase transmission system. J. Inst. Eng. (India) Ser. B **100**(5), 435–445 (2019)

41. A. Swetapadma, A. Yadav, Data-mining-based fault during power swing identification in power transmission system. IET Sci. Meas. Technol. **10**(2), 130–139 (2016)

42. M.S. Mamiş, M. Arkan, C. Keleş, Transmission lines fault location using transient signal spectrum. Int. J. Electr. Power Energy Syst. **53**, 714–718 (2013)

43. R. Godse, S. Bhat, Mathematical morphology-based feature-extraction technique for detection and classification of faults on power transmission line. IEEE Access **8**, 38459–38471 (2020)