ORIGINAL PAPER

Classification and localization of transmission line faults using curve fitting technique with Principal component analysis features

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

Abstract

This paper investigates the application of Principal Component Analysis (PCA) for the development of a simple method of classification and localization of power system faults for a 150 km long transmission line. The proposed work uses only a quarter cycle pre-fault and half cycle post-fault receiving end line current signals for fast identification and isolation of the faulty line. This work analyzes current data of ten different fault classes. The fault signals are recorded at fourteen intermediate geometric locations, out of which, three locations are used for developing the PCA-ratio based classifier and a total of six locations are used for developing the localizer model. PCA is applied here to develop PC score indices, based on which, fault signature curve is developed using best fit analogy. This curve works as the key signature of localizer for each class and phase. The work is made more practically suited by incorporating noise in the signals. Thus an effort has been made in the proposed work for developing a complete practical fault diagnosis algorithm with an aim to achieve high level of accuracy both to classify and localize fault. The proposed classifier is found to produces 100% accuracy, and the localizer is found to achieve an average localization error of only 0.1189% for 40 dB SNR and 0.3965% error at a further higher noise level of 25 dB SNR, with less than 4% of maximum error.

Keywords Principal component analysis (PCA) · Principal component indices (PCI) · Ratio indices (RI) · Best fit curve analysis · Percentage error

1 Introduction

Fault identification and classification of faults are the most important aspects of power system stability, reliability and uninterrupted service. Prediction of fault location in a high power long transmission line also possesses very high importance in the field of power system protection and analysis. Large power transmission networks and grids are the most spatially extended technical systems and fairly often are most vulnerable to minor as well as severe faults since they are mostly exposed to the different atmospheric hazards. It is of utmost importance to identify the faulty phase at the earliest possible time in order to remove the same and bring an immediate stoppage to the outage of huge power through the faulted line. This also helps in preventing damage to the equipments, and most importantly, preventing damage to the persons in vicinity. Retention of fault for long may lead to the development of instability in the system. Hence, identification of faulted phase is very important to restore system stability. Most often, the transmission lines run over different terrains and are often experience short circuit between lines or between lines and ground. Very often these faults are permanent in nature and require manual intervention for its removal. An accurate prediction of fault location is very important to identify the cause of fault by the people at work and hence locate it easily in less time. This helps in quick removal of the fault causing element and restoration of normal power flow. Besides, presence of power line noise makes this works more challenging by introducing harmonics in the system. Advent of digital relays has made the whole protection system much more advanced, accurate and reliable. The different measuring devices, connected to the system, provide real-time data which are processed by different protection algorithms. These primarily extract vital information by continuous condition monitoring of voltage, current,

✉ Arabinda Das
  adas_ee_ju@yahoo.com

1 Government College of Engineering and Ceramic
  Technology, Kolkata, India

2 Electrical Engineering Department, Jadavpur University,
  Kolkata, India

frequency, power factor, etc. to identify any abnormality. The goal of this work is to develop a protection algorithm to detect, classify and most importantly, localize any fault at the earliest in order to remove the abnormality and restore the normal operation as quickly as possible.

Numerous methods have been employed by researchers for developing useful tools for transmission line fault analysis [1, 2]. The proposed work is about fault analysis using multivariate statistical method like Principal Component Analysis (PCA). PCA is effective in reducing the dimension of a multivariate data set; at the same time it is capable of extracting the most significant directions of variability in the descending order of importance. This helps in identifying the key directions of variation; thus allowing for faster numerical simulations with minimal loss of significant information [3]. PCA is used extensively in power system research, especially in fault detection, classification and distance prediction where multi-dimensional data are obtained regarding voltage, current, power, frequency, etc. and/or a combination of these parameters. In this regard, PCA helps in accurate identification of key features by reducing data dimension and enabling easier, faster and accurate computation. Thus, computational burden is significantly smaller for PCA compared to several other methodologies. This statistical method of covariance analysis is lighter in computation compared to wavelet analysis, which uses intricate mathematical analysis as it behaves as a transform based operation. On the other hand, supervised learning approaches like neural network and its variants require large training data for successful updating of internal weights and also associate heavier computational analysis than PCA.

In this work, a simple and direct technique has been discussed for faulty phase identification and localization in a 150 km long single end fed transmission line. This method uses PCA based fault location prediction algorithm. The proposed work is initiated with the development of a 150 km long overhead transmission line in ATP draw simulation. Further, ten different types of power system faults have been simulated at equal intervals of 10 km each. Only quarter cycle pre fault and half cycle post-fault receiving end three phase voltage signals are collected for the proposed PCA based analyzer. Further, the line currents are contaminated externally using power system noise, alike the real-time system, with an average Signal-to-Noise Ratio (SNR) of 40 dB. This noise level is further increased to 25 dB SNR to validate its practical acceptability under more adverse practical situations; although, we have kept fault resistance at a constant low value considering the fact that high resistive faults are a minority in transmission networks. The proposed scheme is found to work satisfactorily well, showing the robustness of the algorithm under practical circumstances. Detection of fault is carried out earlier, followed by classifying the fault to isolate the faulty phase from the system by operating the destined circuit breaker(s). This ensures fast removal of fault and restoration of stability at the earliest

possible time. This is achieved using PCA-ratios of the three phases. Use of the PCA-ratios for the development of the fault classifier algorithm has been carried out as an extension of the work developed in [28–30]. Finally, detection of location of the fault is practiced in order to remove the cause of fault with the least effort. Three locations are used as the training point for the purpose of classification and three more locations are used in addition for developing the fault localizer. The simulation is carried out in ATP draw and analyzer and the analysis is done in MATLAB environment. Ten different classes of faults and healthy condition of line are tested with varying geometric fault locations. PCA is effectively used here to build Principal Component Indices (PCI), which are the representations of the fault signatures. Six training location PCIs for each fault class are used to develop the best fit training curve following the minimum Root Mean Square Error (RMSE) criteria and with the best goodness-of-fit values. The unknown set of PCI is fitted into the different curves to find the best fit model, and finally, predict the geometric fault distance; thus localizing the fault. The entire fault diagnosis method is described using SLG-AG fault as a prototype fault case study.

In the first phase of this paper, the simulation details are given followed by brief description of the Ratio Analysis based fault classification. The proposed fault localization algorithm is also described in connection with these. The next stage describes the detailed result analysis for all the different types of faults at all the different locations. Finally, we have concluded on the usefulness of the proposed PCA based fault localizer scheme and the utility of the component analysis in power system fault diagnosis in relation to the results obtained.

1.1 Background

Power system protection scheme is intended to identify and localize fault during abnormality and eliminate fault at the earliest possible time using fault sensors like relays, current transformer and potential transformer and actuators like circuit breakers. Hence, prompt detection and classification of faults, as well as, precise fault location identification have been practiced by scientists in order to ensure system safety and stability [1, 2]. Researchers have developed many mathematical and computational tools for the detection, classification and localization of faults. Artificial Intelligence (AI), nowadays, is applied extensively by researchers in the area of fault analysis and power system research. Artificial Neural Network (ANN) and its different variants have been one of the most used and fundamental methods used in the research of power system protection of transmission lines and used in many researches effectively [4, 5]. Probabilistic neural network (PNN) has been extremely effective, especially in fault classification analysis, for its well known feature of pattern recognition; hence used in abundance [6]. ANN based works have progressed miles with the recent

development of machine learning and deep learning based analysis. Extreme learning machine (ELM) based analysis using neural networks has been among the recent advancements in this field [7, 8]. Wavelet transform (WT) has been instrumental in several researches of fault analysis as a traditional method of fault signal analysis, even incorporating modern compensating devices [9]. Neural network based methods, although are very accurate, suffer from the requirement of diverse training and WT is computationally more demanding, especially with higher levels of decomposition. Despite the respective disadvantages, one of the most common ways of using WT lies in the use of wavelet coefficients and entropy features, which are obtained as fault features, with supervised learning model like neural network. This method has been instrumental in several researches [10–15], as it possesses the accuracy of transform based signal analysis method like WT as well as supervised learning approach to develop accurate fault analyzers. The authors of [13] have also introduced Parseval's theorem in addition to WT and neural network. Fault features of discrete wavelet transform (DWT) have been analyzed using Chebyshev neural Network (ChNN) for a thyristor controlled series compensated line in [14] for analysis of faults, whereas, wavelet features have been used to model a 2-Tier multilayer perceptron (MLP) network to develop a robust fault classification method in [15]. Fuzzy inference system has been another effective tool for fault analysis, often used as a major standalone method of analysis [17], as well as used as hybrid model in combination with wavelet analysis [18] and neural network, which is used as adaptive neuro fuzzy inference system or ANFIS model. This hybrid model has been often aided by wavelet analysis to develop wavelet based ANFIS models and used as an effective method of complete analysis [19–21]. Support Vector Machine (SVM) too has been used in a wide number of researches related to power system protection algorithm, as a major standalone method of analysis [22], or using features from other analysis methods like neural network [23], wavelet analysis [24], discrete orthogonal S-transform (DOST) [25, 26] and others. Signal entropy has also been analyzed which uses the randomness of a fault signal for each phase and classifies fault accordingly [27].

Traveling wave based methodologies have also been applied effectively in several researches, especially for localization of transmission line faults [28, 29]. Phasor Measurement Units or PMU is one of the relatively new methods that been investigated immensely in modern analysis [30, 31], although often requires additional hardware support for synchronous measurement at both ends of a transmission line. Time domain analysis as well as frequency spectrums are among the other common applications in this field [32, 33]. The frequency of the transient fault oscillations bear major features, especially regarding fault location, which are interpreted using these time or frequency domain analysis. Spectral energy is also combined with DOST and CUSUM algorithm for detection, classification, and localization of faults in power transmission system [34].

PCA has also been applied in transmission line fault analysis for its major advantage to identifying principal directions of variation in the descending order, which, in turn, reduces the dimensionality of the data set. A PCA is often applied as a single feature extraction tool for developing methods of fault analysis [35–37], as well as in combination with PNN is shown in [38, 39]. Multiple linear regression [36] or curve fitting tool [37] has also been used to model fault localizer using PCA extracted fault features. Fault features from PCA have often been combined with other methodologies to develop hybrid fault analyzer models. PCA has also been used in combination with other methodologies like traveling wave and wavelet analysis [40], SVM [41] and others to develop accurate hybrid fault diagnosis techniques. The proposed method fault classification, in this work, has been derived extending the ratio based method analyzed in [35, 36] in the form of a direct threshold based classification method and the localization scheme has been followed using the concepts described in [37], but using the receiving end fault signals only.

Classical methods of methods of fault analysis include mostly the sequence component based techniques, which are mostly used in practical power transmission-distribution systems [42–45]. Advancement of soft computation methods have helped researchers to develop more accurate classical positive and negative sequence network based fault analysis models which are able to produce accurate output [46, 47]. The development of microprocessor, combined with soft computation has helped to develop and implement digital relays, embedding the fast analysis techniques. Sometimes, these sequence components are analyzed directly, or combined with other tools to develop directional effective relaying schemes [48–50]. Sequence voltages and currents are very accurate, as well as sensitive in fault detection. However, many of the works have pointed out that load imbalance is one of the major causes of failure of the sequence based protection systems [43, 44]. Load unbalancing often occurs in real-life in a three-phase system; which, in turn, generates the symmetrical components in lines. It often causes the relays maloperation, even if there is no fault. Other errors include measurement errors, especially CT saturation and CCVT sub-transient errors. These also introduce spurious sequence components, e.g., negative and zero sequence components; which sometimes causes maloperation in relaying by introducing sequence components.

2 Transmission system design and simulation

A single side fed 270 kV 150 km long, single circuit, three phase, radial, overhead transposed AC transmission line model has been considered for the proposed work. The simulation model of the said line is designed in Electro Magnetic Transient Programming (EMTP). Receiving end current waveforms for three phase have been taken and used as the only data and the same is obtained for ten different types of faults, e.g.,

(a) Single Line to Ground (SLG) fault for lines A, B and C (SLG-AG, SLG-BG, SLG-CG, respectively),
(b) Double Line (DL) fault in between lines AB, BC and CA (DL-AB, DL-BC, DL-CA, respectively),
(c) Double Line to Ground (DLG) fault in between lines AB, BC and CA and ground in each case (DLG-ABG, DLG-BCG, DLG-CAG, respectively) and
(d) Three phase fault (LLL-ABC)

The above faults are conducted at different locations 10 km apart throughout the entire length of 150 km line, along with the healthy condition and waveform are collected for further analysis using the proposed algorithm. Fifteen three phase Line Cable Constants (LCC) blocks, each of 10 km, are connected in cascade to develop the 150 km long overhead transmission line. The frequency dependent 'JMarti' model has been adopted for the given purpose. Sampling frequency has been taken as 2000 samples/cycle of the sinusoidal current waveform, hence, giving a sampling frequency of 100 kHz.

3 Development of fault detection and classification algorithm

Power system fault identification and proper classification of the fault type are the first and foremost step in power system protection scheme. Unless the faulty line(s) is identified and immediately isolated, the fatal risk of damage to people and the working personnel exists, apart from the possibility of damage to the different protective instruments and sophisticated devices. Besides, unnecessary drainage of electric power from the transmission network aids in hampering synchronous system stability, which may causes severe damage to the system. Hence, the cause of fault should be removed at the earliest possible measure, followed by system restoration. Hence the proposed work initiates the analysis using the fault identification and classification before proceeding to the localization of the fault.

3.1 Training data preparation

Application of PCA on normalized, quarter cycle pre fault and half cycle post fault, three phase, receiving end current signals yields a set of Principal Component Indices (PCI). This is repeated for ten prototype fault classes, conducted at different intermediate locations of the line, each 10 km apart. Fault identification and classification have been carried out using fault ratio signatures generated using fault signals, for faults conducted at three intermediate locations: at 30 km, 80 km and 140 km from sending end. Fault waveforms are recorded for ten fault classes and one healthy condition. Sampling frequency of 2000 samples/cycle produced a total of 1500 sample points for the above duration of waveform. Hence, the input variable matrix, using the signals of the three phases, is constructed as follows:

$$X_i = \begin{bmatrix} IRa_{i1} & IRb_{i1} & IRc_{i1} \\ IRa_{i1} & IRb_{i1} & IRc_{i1} \\ \dots & \dots & \dots \\ IRa_{i1} & IRb_{i1} & IRc_{i1\ 1500} \end{bmatrix}_{1500 \times 3} \tag{1}$$

where, i denotes the fault class; hence, i = 1, 2, …, 11, represents ten fault class and healthy condition. The suffixes a, b and c represent three different phases. Thus, combining above 11 fault classes, the training variable takes the following form:

$$X = \begin{bmatrix} X_1 X_2 X_3 \dots X_{11} \end{bmatrix}_{1500 \times 33} \tag{2}$$

Further simplification has been carried out by phase separating the training matrix X to construct Xa, Xb and Xc separately, e.g., Xa is given by:

$$X_i = \begin{bmatrix} IRa_{1\ 1} & IRa_{2\ 1} & IRa_{11\ 1} \\ IRa_{1\ 2} & IRa_{2\ 2} & IRa_{11\ 2} \\ \dots & \dots & \dots \\ IRa_{1\ 1500} & IRa_{2\ 1500} & IRa_{11\ 1500} \end{bmatrix}_{1500 \times 11} \tag{3}$$

Xb and Xc are also constructed similarly using B and C phase signals for the same 11 different conditions. Thus each phase produces a data matrix of the dimension 1500 × 11. Hence the modified training matrix, denoted by Xm, is given as:

$$Xm = \begin{bmatrix} Xa & Xb & Xc \end{bmatrix}_{1500 \times 33} \tag{4}$$

3.2 Test data preparation

This is done similarly to that of the training data preparation. Three phase current data of the receiving end for the unknown class of fault are taken as the experimental or test data. Thus the test data matrix (T) takes the form given by:

$$T = \begin{bmatrix} IRa_{i1} & IRb_{i1} & IRc_{i1}; \\ IRa_{i1} & IRb_{i1} & IRc_{i1}; \\ \dots & \dots & \dots; \\ IRa_{i1} & IRb_{i1} & IRc_{i1\ 1500} \end{bmatrix}_{1500\times3} \quad (5)$$

T is again a 1500×3 matrix. Further, this test data matrix (T) is represented with symbols of individual phases to produce the modified test data matrix (Tm) as:

$$Tm = [\ Ta \quad Tb \quad Tc]_{1500\times3} \quad (6)$$

Each of Ta, Tb and Tc is of the dimension 1500×1. Finally, Xm and Tm become the two matrices under consideration which are processed through the proposed fault classifier algorithm discussed in the next section.

3.3 Fault classifier algorithm

As discussed earlier, Principal Component Analysis (PCA) has been used to design the proposed fault classifier algorithm. It is evident that PCA serves a very good purpose in reducing the dimension of any multivariate data set and identify the direction of the most variability from a large set of widespread data. In the proposed work, PCA has been applied on the receiving end current data as discussed before. The training and test set matrices i.e., Xm and Tm, respectively, are processed using the proposed algorithm to find out PCA indices of each phases, corresponding to each of the eleven training and test conditions. The steps of the proposed work for faulty phase identification are discussed as follows:

3.4 PCA algorithm

The training and test data are analyzed using PCA based algorithm with phases combined together independently. Hence, the phase separated matrix of the training and test set are merged together to obtain the combined matrix C as:

$[Ca]_{1500\times12}=[Xa\ Ta]; [Cb]_{1500\times12}=[Xb\ Tb]; [Cc]_{1500\times12}=[Xc\ Tc]$. Hence, the proposed algorithm is developed as follows:

$[Ca]_{1500\times12} = [Xa\ Ta]; [Cb]_{1500\times12} = [Xb\ Tb]; [Cc]_{1500\times12} = [Xc\ Tc]$. Hence, the proposed algorithm is developed as follows:

// Declaration of variables

Input: $[Ca]_{1500\times12}$, $[Cb]_{1500\times12}$, $[Ca]_{1500\times12}$,

Output: Identified class in terms of numerical value

// Beginning of the program

Formation of combined phase split matrix using training and test data as follows:

$[Ci]_{1500\times12} = [Xi, Ti]$, where *i* stands for a, b or c phase

call PCA algorithm to find out the Combined Principal Component Index matrices (PCI)

Combined PCI matrices formed for each phase:

$[PCIi]_{12\times1}$, where *i* stands for a, b or c phase

Form total Combined PCI matrix including all phases: $[PCI]_{12\times3} = ([PCIa]_{12\times1}\ [PCIb]_{12\times1}\ [PCIc]_{12\times1})$

Split [PCI] to form training and test PCIs

$[PCI_{TRAINING}]_{11\times3}$ = Sort ([PCI] all columns, rows 1 to 11)

$[PCI_{TEST}]_{1\times3}$ = Sort ([PCI] all columns, row 12)

for *k* = 1 to 12;

Ratio 1 (*k*) = PCI (*k*, 1) / PCI (*k*, 2);

Ratio 2 (*k*) = PCI (*k*, 2) / PCI (*k*, 3);

Ratio 3 (*k*) = PCI (*k*, 3) / PCI (*k*, 1);

end

Form Ratio Indices matrix as:

$[RI]_{12\times3}$ = ($[Ratio\ 1]_{12\times1}$ $[Ratio\ 1]_{12\times1}$ $[Ratio\ 1]_{12\times1}$);

Split [RI] to form to form training and test fault signatures separately as:

$[RI_{TRAINING}]_{11\times3}$ = Sort ([RI] all columns, rows 1 to 11)

$[RI_{TEST}]_{1\times3}$ = Sort ([RI] all columns, row 12)

end

Table 1 Ratio Index values for ten fault classes at three fault locations

Fault Class	Ratio 1			Ratio 2			Ratio 3		
	30 km	80 km	140 km	30 km	80 km	140 km	30 km	80 km	140 km
AG	4.163	4.53316	4.42338	1.00357	1.00971	1.00258	0.21128	0.21847	0.22549
BG	0.26176	0.25805	0.25262	3.8591	3.86606	3.95582	0.98995	1.00237	1.00067
CG	0.98989	1.00599	1.01139	0.2553	0.29573	0.30004	3.957	3.36132	3.29536
AB	1.00253	1.00156	0.99669	530.09	750.993	951.911	0.00188	0.00133	0.00105
BC	0.00179	0.00143	0.00053	1.00081	1.00058	1.00081	557.512	699.533	1902.74
CA	1156.46	1074.5	791.995	0.00087	0.00093	0.00126	0.99904	1.00102	1.00113
ABG	1.01419	0.92114	0.95606	4.69408	4.63943	4.37384	0.21005	0.234	0.23914
BCG	0.32631	0.22305	0.21176	0.62006	0.84327	0.93262	4.94236	5.31667	5.06338
CAG	6.43071	5.57782	5.1779	0.25314	0.23208	0.22546	0.61431	0.77251	0.85658
ABC	1.48366	1.16395	1.0801	0.85009	1.01346	1.03683	0.79286	0.84773	0.89295

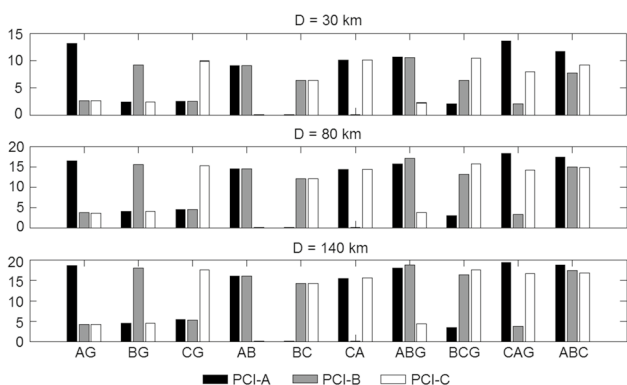**Fig. 1** Three phase PCI for ten different fault classes obtained at three different fault locations (D): 30 km, 80 km and 140 km from sending end

Fig. 2 Ratio Indices values for SLG, DLG and LLL faults at three different fault locations (D): 30 km, 80 km and 140 km from sending end

These two matrices: [PCI] and [RI] are used to develop the proposed fault classifier model. The PCI matrices thus formed are basically an approximate estimation of the extent of disturbance of each fault current waveform from the healthy condition. The direction of each component is given by the eigenvectors obtained from the covariance matrix of the transformed data points. Magnitude of maximum deviation from the origin (which is assigned to the no fault condition) corresponds to the respective eigenvalues.

3.5 Numerical analysis of three phase PCI values: ratio based logic development

The three phase receiving end current signals are analyzed for ten different fault classes as mentioned before and the PCI values are recorded. These PCI are further analyzed to obtain the [RI $_{TRAINING}$] and [RI $_{TEST}$] matrices following the above algorithm. Three intermediate locations: 30 km, 80 km and 140 km are chosen as the training points for the development of the proposed scheme, and hence, constructs
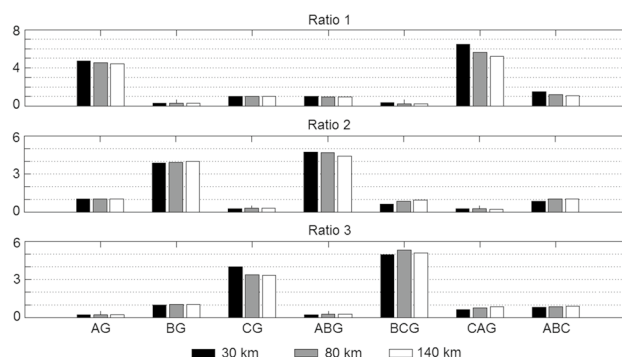
[PCI $_{TRAINING}$]. These values are further analyzed to develop [RI $_{TRAINING}$] and are shown in Table 1. The above two matrices are also described graphically in Figs. 1 and 2, respectively.

3.6 Fault detection

The proposed ratio based algorithm is first used to identify a fault in line, followed by classification of the same. It is readily observed from Table 1 and the associated Figs. 1 and 2 that for all the classes of faults, except the symmetrical ABC fault, at least one of the ratio values easily exceeds 2 by magnitude. Hence, the fault detection algorithm is designed in such a way that the ratio values are tested first and a fault is detected if one of these is found to exceed this limiting magnitude of 2. Hence, a fault detection threshold ϕ_1 is chosen for the ratio index, which is assigned magnitude 2, i.e., $\phi_1 = 2$. But the ABC fault escapes this loop since the symmetrical three phase fault affects all the three phases almost equally, which results in similar PCI magnitudes for the three phases, and in
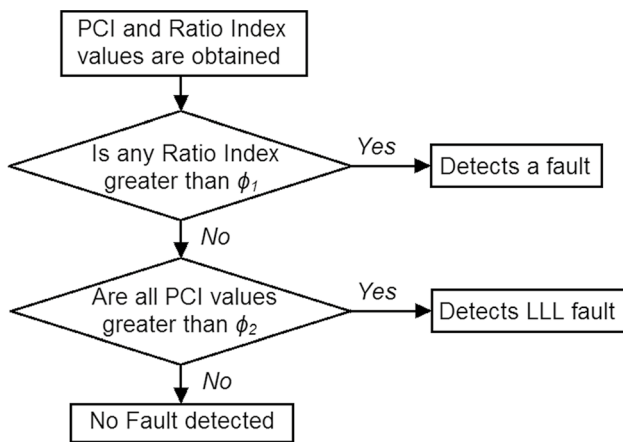
Fig. 3 Proposed fault detection algorithm

turn, produces similar ratio indices values close to unity. Figure 2 illustrates the above discussion. Hence, in order to identify the three phase fault, the three phase PCI values are also investigated. A LLL fault, which is the ABC fault, is identified on detection of the condition when all the three phase PCI values are found to exceed a certain level simultaneously. Observation of Fig. 1 shows that the all the three phase PCI levels exceed magnitude 6 for all LLL faults. Hence, a second threshold ϕ_2 is identified which compares the PCI values of all phases and detects a LLL fault upon observing all the three phase PCIs to exceed this threshold ϕ_2. Hence, as per the above discussion, magnitude of this second threshold ϕ_2 is selected as 6, i.e., $\phi_2 = 6$. The proposed fault detection technique is also shown graphically in the form of a flowchart in Fig. 3.

The proposed algorithm is further tested with fault signals including variation of load and it is found that the algorithm does not detect it as a fault. This is primarily because the three phase signals are found to be affected almost equally with a load change. This is identified by almost equal increase or decrease in all the three signals simultaneously. Further, the proposed analysis is carried out by converting the fault signals into per unit system. This again reduces the effect of variation in load. This helps in analyzing the proposed design in two ways:

1. The change in load does not take the PCI values beyond the threshold level which is ϕ_1, considering even the instantaneous transients arising immediately after a load change, hence, no fault is detected,
2. Since the load change affects the three phases almost equally, the mutual ratios of the PCI values remain almost near unity; hence does not exceed the PCI threshold, i.e., ϕ_2.

These two factors, when analyzed simultaneously, are found to detect only the true faults in all the tested cases. Hence, the proposed work is found to work well even for load changing conditions.

Table 3 Fault Classifier rule base

Ratio 1	Ratio 2	Ratio 3	Fault Class
$\theta_2 < RI_1 < \theta_1$	$\theta_3 < RI_2 < \theta_2$	$RI_3 < \theta_3$	AG
$RI_1 < \theta_3$	$\theta_2 < RI_2 < \theta_1$	$\theta_3 < RI_3 < \theta_2$	BG
$\theta_3 < RI_1 < \theta_2$	$RI_2 < \theta_3$	$\theta_2 < RI_3 < \theta_1$	CG
–	$RI_2 > \theta_1$	–	AB
–	–	$RI_3 > \theta_1$	BC
$RI_1 > \theta_1$	–	–	CA
$\theta_3 < RI_1 < \theta_2$	$\theta_2 < RI_2 < \theta_1$	$RI_3 < \theta_3$	ABG
$RI_1 < \theta_3$	$\theta_3 < RI_2 < \theta_2$	$\theta_2 < RI_3 < \theta_1$	BCG
$\theta_2 < RI_1 < \theta_1$	$RI_2 < \theta_3$	$\theta_3 < RI_3 < \theta_2$	CAG
$\theta_3 < RI_1 < \theta_2$	$\theta_3 < RI_2 < \theta_2$	$\theta_3 < RI_3 < \theta_2$	ABC

Table 2 Classification of ratio indices using three threshold values for ten fault classes at three training locations

Fault Class	Ratio 1			Ratio 2			Ratio 3		
	30 km	80 km	140 km	30 km	80 km	140 km	30 km	80 km	140 km
AG	$\theta_2 < RI < \theta_1$	$\theta_2 < RI < \theta_1$	$\theta_2 < RI < \theta_1$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$RI < \theta_3$	$RI < \theta_3$	$RI < \theta_3$
BG	$RI < \theta_3$	$RI < \theta_3$	$RI < \theta_3$	$\theta_2 < RI < \theta_1$	$\theta_2 < RI < \theta_1$	$\theta_2 < RI < \theta_1$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$
CG	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$RI < \theta_3$	$RI < \theta_3$	$RI < \theta_3$	$\theta_2 < RI < \theta_1$	$\theta_2 < RI < \theta_1$	$\theta_2 < RI < \theta_1$
AB	–	–	–	$RI > \theta_1$	$RI > \theta_1$	$RI > \theta_1$	–	–	–
BC	–	–	–	–	–	–	$RI > \theta_1$	$RI > \theta_1$	$RI > \theta_1$
CA	$RI > \theta_1$	$RI > \theta_1$	$RI > \theta_1$	–	–	–	–	–	–
ABG	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_2 < RI < \theta_1$	$\theta_2 < RI < \theta_1$	$\theta_2 < RI < \theta_1$	$RI < \theta_3$	$RI < \theta_3$	$RI < \theta_3$
BCG	$RI < \theta_3$	$RI < \theta_3$	$RI < \theta_3$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_2 < RI < \theta_1$	$\theta_2 < RI < \theta_1$	$\theta_2 < RI < \theta_1$
CAG	$\theta_2 < RI < \theta_1$	$\theta_2 < RI < \theta_1$	$\theta_2 < RI < \theta_1$	$RI < \theta_3$	$RI < \theta_3$	$RI < \theta_3$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$
ABC	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$	$\theta_3 < RI < \theta_2$

3.7 Fault classification

It is observed from Figs. 1 and 2 that the PCI and the RI values have specific features for definite class of fault which are investigated in this work to obtain the classification rule bases. The following observations were found prominent from the [RI $_{TRAINING}$] matrix, i.e., Table 1 and also, from Fig. 2:

(a) For any DL faults, any one of the ratio indices, i,e,, either ratio 1 or ratio 2 or ratio 3 becomes extremely high. This is clearly observed from Table 1 that, e.g., ratio 2 for AB fault, ratio 3 for BC fault and ratio 1 and for CA fault becomes very high. This high value is more than 100, even considering the three different training fault locations.

(b) The ratio indices corresponding to the rest of the faults (SLG, DLG and LLL) follow some common ranges of values, irrespective of the fault location:

Class 1: Some of the ratio indices values are in the range of 0.2 to 0.35,
Class 2: Some of the ratio indices values ranges within 0.6 to 1.5, and
Class 3: Some other ratio indices values ranges within 3 to 7.

Depending on the above observations, three thresholds are developed as θ_1, θ_2, and θ_3.

θ_1 is the highest threshold limit designed for identification of DL faults. If any one of the three ratios is found to exceed this threshold θ_1, the fault is identified as DL. Detailed examination of the ratio index, i.e., identifying if it is ratio 1 or ratio 2 or ratio 3, the exact fault is classified. This θ_1 is safely assigned the value of 100, as observed from Table 1. This threshold distinguishes the DL class from the ratio indices values belonging to class 3, since the upper limit of class 3 is found as 7, which is much less than this θ_1.

In order to distinguish ratio indices values of class 3 from the set of ratio indices values of class 2 as mentioned above, the second threshold θ_2 is assigned. This is selected to lie in between the upper threshold of class 2 i.e., 1.5 and lower threshold of class 3, i.e., 3. Hence, θ_2 is selected as 2.5.

Separation of ratio indices values of class 2 from the set of ratio indices values of class 1 is done similarly by selecting the third threshold θ_3 as 0.45, which is almost in the middle of the upper threshold of class 1 i.e., 0.35 and lower threshold of class 2, i.e., 0.6.

Hence, the threshold values are written as: $\theta_1 = 100$, $\theta_2 = 2.5$ and $\theta_3 = 0.4$. Using these values, the RI values from Table 1 could be written in terms of the threshold values as shown in Table 2. It is important to note that only the relevant and useful thresholds are written the table and

irrelevant fields in terms of classification are left blank. Depending on this Table 2, the final fault classifier rule bases are obtained as shown in Table 3.

The unknown fault is analyzed using the PCA based algorithm described before to obtain the [RI $_{TEST}$] which is compared with this fault classifier rule base to obtain the predicted fault class. This algorithm is tested using fault signals corresponding to ten different fault classes and the results obtained are described later under the result section.

4 Fault distance estimation

Determination of the fault location is another vital part of the proposed work. Best fit analysis has been applied with the PCA features [38] obtained with the receiving end fault current signals. The three phase current waveforms corresponding to six training locations, as mentioned earlier, have been used to develop the localizer algorithm. Close observation of the PCI values reveal a monotonic variation with fault location. This variation is also observed to be different for each fault class. These PCI values when plotted against the corresponding fault locations reveal mostly curvilinear trend, which have been approximated by different curves in MATLAB environment. The best fit curve so obtained among the various combinations is termed as the fault location signature curve. The PCI of the test data corresponding to unknown fault location is used with this best fit curve to predict the unknown fault distance.

4.1 Processing of training data: case study

The six PCI values so obtained are taken as the training input to the algorithm. The proposed analysis is illustrated using SLG-AG fault only as an example case of study. It is observed that the maximum disturbance is caused for the line directly under fault for each class. It is readily

Table 4 A sample case of three phase PCI for SLG-AG fault with variation in fault location; SNR = 40 dB

Fault location (km)	PCI-A	PCI-B	PCI-C
10	7.653655	1.755881	1.776019
30	13.25115	2.910645	2.949314
50	15.06687	3.320619	3.373498
80	16.44998	3.684799	3.696578
110	17.4887	3.953036	4.00362
140	18.45396	4.213816	4.255525
Test	15.60633	3.464064	3.482196
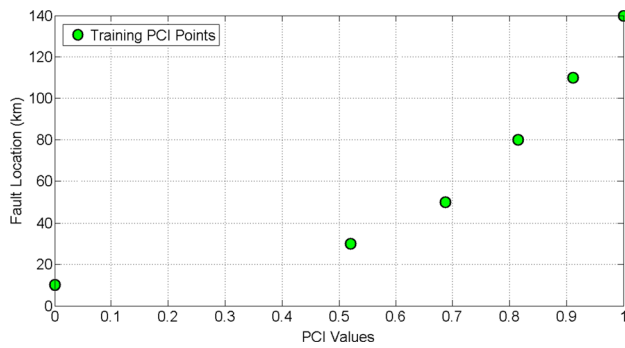
Fig. 4 PCI-A vs. fault location plot of six training points and the test fault for SLG-A fault

Fig. 5 Best fit Shape-preserving interpolant curve fitted to training PCI points: fault location signature curve

found that for SLG faults, only one line is directly affected and the two other lines remain less disturbed; for DL or DLG fault, two lines are directly under fault and are produce maximum disturbance; and finally, all three lines are disturbed for LLL fault. Hence, on PCA, the directly faulted phase(s) gives out highest magnitude of PCI and the indirectly faulted phase(s) produce less significant PCI values. Hence, for SLG fault, PCI of the single directly affected phase is considered; for DL or DLG faults, PCI values obtained from the two directly affected phases are considered; and for LLL fault, all the three PCI values are considered for analysis. Each phase signal is analyzed independently and the mean predicted values is considered as finally predicted location.

Post fault signals are analyzed initially for classification of faults, followed by application of the fault localizer scheme. The primary training input is a 1500×6 matrix for each phase; the six columns of the same denote the six training location points. Hence, for SLG fault, only one; for DL and DLG faults, two; and for LLL faults, three such sets are obtained. PCI value primarily is a measure of the extent of disturbance caused in each phase during fault, with respect to the healthy condition. Hence, phase A, being the most disturbed phase in case for AG fault, has the highest impact of fault. More so, the fault signals are contaminated additionally with a noise level of 40 dB SNR to introduce more practically simulated signals. White Gaussian noise is generated in MATLAB environment and added with all the signals for this purpose. Further attempts have been made to identify and relate the varying pattern of the PCI values computed from these noise contaminated fault signals, to develop fault location curve. The variation of the PCI values for the studied case of SLG-AG fault is shown in Table 4 for all the six training locations.

This is also observed from the same table that PCI-B and PCI-C are very less in magnitude compared to PCI-A. Hence, only the most significant phase A current signals is taken for
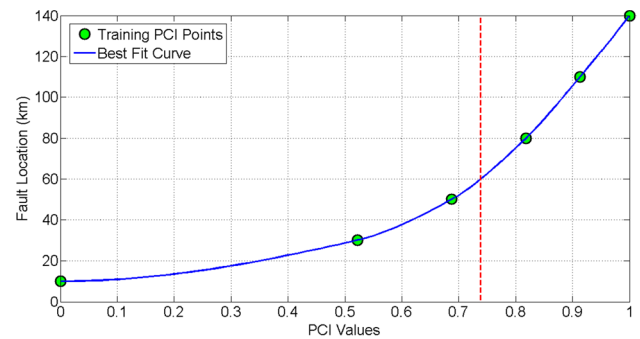
consideration for developing the fault location curve. It is well observed from Table 4 that as the fault distance increases from the source end, the deviation of the phase current increases gradually from no-fault condition, which is interpreted from the PCI values. It is also observed that PCI-A shows a fair monotonic increasing variation with chronological variation of the geometric fault location, which is mathematically interpreted using best fit analysis. The test fault PCI is also shown in the same table at the final row. Apparent observation and consideration of linear interpolation of the test value of PCI-A given in Table 4 show that test fault lies in between 50 and 80 km, closing more toward 50 km.

4.2 Best fit model design

Close observation of Table 4 reveals that the PCI-A vary monotonically for variation of fault location. The input training column vector has the dimension of a 6×1; the six elements represent the PCI corresponding to the six training locations: 10, 30, 50, 80, 110 and 140 km of the 150 km long line. These values are obtained from the PCI-A values of Table 4, which are further scaled in the range [0, 1] and finally, plotted with the actual fault location as the dependent variable in Fig. 4.

A curvilinear nature of the PCI points is evident from Fig. 4. This is approximated using best fit curve analysis. Different curve fit models are applied over each set of these training PCI points for different fault prototypes. Different characteristic curves like linear polynomial, exponential, interpolant, different smoothing spline piecewise polynomial and Gaussian distribution has been investigated in this work. Only six of the minimum error methods among the several curve fitting methods have been chosen to find the estimates of fault locations. The test PCI has been investigated using the best fit method to obtain the predicted test fault location. The following fit models have been investigated for the initial evaluation:

Table 5 Result of the proposed fault classifier algorithm

True fault class	Predicted fault class										
	No Fault	A G	B G	C G	A B	B C	C A	A B G	B C G	C A G	A B C
No Fault	**14**	0	0	0	0	0	0	0	0	0	0
AG	0	**14**	0	0	0	0	0	0	0	0	0
BG	0	0	**14**	0	0	0	0	0	0	0	0
CG	0	0	0	**14**	0	0	0	0	0	0	0
AB	0	0	0	0	**14**	0	0	0	0	0	0
BC	0	0	0	0	0	**14**	0	0	0	0	0
CA	0	0	0	0	0	0	**14**	0	0	0	0
ABG	0	0	0	0	0	0	0	**14**	0	0	0
BCG	0	0	0	0	0	0	0	0	**14**	0	0
CAG	0	0	0	0	0	0	0	0	0	**14**	0
ABC	0	0	0	0	0	0	0	0	0	0	**14**

Table 6 Predicted fault locations using different fitness models for SLG-AG fault

Fit model	Actual distance (km)							
	20	40	60	70	90	100	120	130
	Predicted fault locations (km)							
Fit 1	19.7886	40.1549	59.7361	70.2907	89.8992	100.0793	119.8859	130.1105
Fit 2	17.3982	42.2551	61.1024	69.8468	87.7094	97.12397	117.8839	129.4838
Fit 3	17.4963	40.6237	61.3343	71.0332	90.6601	100.2405	119.4554	129.2313
Fit 4	23.4522	39.8213	60.146	69.8426	89.9415	100.2124	120.0913	130.0966
Fit 5	21.9737	39.7744	60.0808	70.0835	89.9806	100.3726	120.1659	129.96
Fit 6	17.5741	41.1447	61.5818	70.7479	89.0913	98.71731	118.4105	129.868

Table 7 Error of prediction fault locations using different fitness models for SLG-AG fault

Fit model	Actual distance (km)								Average prediction error (km)
	20	40	60	70	90	100	120	130	
	Predicted fault locations (km)								
Fit 1	0.21142	0.15488	0.26386	0.29066	0.10079	0.079283	0.114111	0.110456	0.165682
Fit 2	2.60178	2.25514	1.10241	0.15323	2.29061	2.876026	2.116095	0.516248	1.738943
Fit 3	2.50372	0.62371	1.33429	1.03318	0.66014	0.240511	0.544567	0.768651	0.963595
Fit 4	3.45218	0.17873	0.14603	0.15745	0.05846	0.212395	0.091322	0.096562	0.549141
Fit 5	1.97367	0.22558	0.08076	0.08351	0.01942	0.37258	0.165859	0.04003	0.370176
Fit 6	2.42588	1.14465	1.58185	0.74793	0.90875	1.282691	1.589472	0.131969	1.226648

Fit 1: Shape-preserving interpolant
Fit 2: Exponential method 2nd-order
Fit 3: Smoothing spline piecewise polynomial
Fit 4: Cubic interpolating spline piecewise polynomial
Fit 5: Gaussian distribution 2nd-order
Fit 6: Linear model Polynomial 3rd-order

A curvilinear trend of the training PCA points is clearly visible from the PCI values of Fig. 4. Best fit analysis is carried out over these values to obtain the smooth curve joining these six training PCI points. A comparative analysis using the goodness of fit values, regarding the minimum root mean square error (RMSE) criteria has been adopted here to obtain the best suited one among the six fitness models proposed here. This method is performed for the three

Table 8 Sample observations of the predicted fault locations using Shape-preserving interpolant (fit 1) model

Fault Type	Actual Distance (km)	Average Predicted Fault Distance (km)			Average Predicted Output (km)	Absolute Error (km)	PE (%)
		Phase A	Phase B	Phase C			
AG	20	19.2886	–	–	19.2886	0.7114	0.4743
BG	70	–	70.5954	–	70.5954	0.5954	0.3969
CG	90	–	–	88.9882	88.9882	1.0118	0.6745
AB	100	99.2748	99.6547	–	99.4648	0.5352	0.3568
BC	60	–	60.4125	58.7457	59.5791	0.4209	0.2806
CA	130	130.2413		130.8125	130.5269	0.5269	0.3513
ABG	120	120.3124	120.7478	–	120.5301	0.5301	0.3534
BCG	130	–	130.2315	129.5514	129.8915	0.1085	0.0723
CAG	20	19.5907	–	19.8718	19.7313	0.2687	0.1791
ABC	40	40.3149	40.9346	39.9854	40.4116	0.4116	0.2744

phases individually for each fault class. Further, the curves obtained are tested with PCI data of some test faults with unknown geometric distances. It is observed that shape-preserving interpolant model produces the minimum level of error of prediction; hence, adopted in this work. This best fit curve is also shown in Fig. 5, which is also denoted as the fault location signature curve. The test fault PCI is also marked with a red vertical dotted line in the same plot. This test fault line intersects the curve at a certain point; projection of the same point on the vertical axis predicts the fault location. In this example test case, the predicted test fault location is found nearly 60 km. Hence, the inference drawn in the earlier section from Table 4 regarding the location of the test fault remains valid from this best fit fault location signature curve of Fig. 5. This again confirms the location of the test fault in between 50 and 80 km, and further, its proximity toward the 50 km end.

5 Results and discussion

5.1 Results of the fault classifier

Table 5 shows the results of the ratio analysis based fault classifier algorithm based on 14 sets of test current signals. It is found that the proposed classifier produces 100% accurate result, irrespective of the addition of noise. This shows the effectiveness of the proposed scheme even in practical like situations.

5.2 Results of the proposed fault localizer

In this work, six locations along the 150 km line, viz., 10, 30, 50, 80, 110 and 140 km are used for the development of the proposed scheme, hence termed as training locations. The rest eight locations, viz., 20, 40, 60, 70, 90, 100, 120 and 130 km are used for testing the same. The results of the different curve fit models are shown in Table 6. These results

Table 9 Summary of fault location predictor algorithm considering all faults and using Shape-preserving interpolant (fit 1) fitness model, SNR = 40 dB

Type of Fault	Average error from target (km)	Average PE (%)	Maximum error (km)	Maximum PE (%)
AG	0.1657	0.1105	1.2046	0.8031
BG	0.1812	0.1208	1.6721	1.1147
CG	0.1753	0.1169	1.3047	0.8698
AB	0.1808	0.1205	1.6412	1.0941
BC	0.1614	0.1076	1.2042	0.8028
CA	0.1146	0.0764	1.2745	0.8497
ABG	0.1852	0.1235	1.8157	1.2105
BCG	0.1383	0.0922	1.3407	0.8938
CAG	0.2186	0.1457	1.7051	1.1367
ABC	0.2625	0.175	1.7408	1.1605
Overall	0.17836	0.1189	1.8157	1.2105

Table 10 Summary of fault location predictor algorithm considering all faults and using Shape-preserving interpolant (fit 1) fitness model at higher noise level of SNR = 25 dB

Type of Fault	Average error from target (km)	Average PE (%)	Maximum error (km)	Maximum PE (%)
AG	0.5197	0.3465	5.2145	3.4763
BG	0.6784	0.4523	4.9578	3.3052
CG	0.6627	0.4418	3.5404	2.3603
AB	0.5349	0.3566	4.3475	2.8983
BC	0.4212	0.2808	2.9845	1.9897
CA	0.3185	0.2123	3.6842	2.4561
ABG	0.7108	0.4739	5.8142	3.8761
BCG	0.5745	0.383	5.0412	3.3608
CAG	0.7419	0.4946	4.7451	3.1634
ABC	0.7854	0.5236	5.1547	3.4365
Overall	0.5948	0.3965	5.8142	3.8761

are obtained using the PCI-A values for a prototype SLG-AG fault, which is used here as an example case. The predicted fault locations are shown in Tables 6, and 7 show the corresponding errors of prediction. Finally the choice of the best fit is decided on the minimum error of prediction criteria.

It is observed that Shape-preserving interpolant (Fit 1) produces the best results, followed by Gauss 2nd-order (fit 5) and Cubic interpolating spline piecewise polynomial model (Fit 4); although, Shape-preserving interpolant model is superior by fair margin compared to the others. This model is again tested for other classes of faults with one or multiple set of PCI values corresponding to one or more phases, where the same fitness model is found to produce appreciable results; hence, considered as the global fitness model in this work for fault location prediction. Table 8 shows a few samples of the simulation results by the proposed fault location predictor algorithm using the same shape-preserving interpolant model for ten different prototype classes, conducted at different locations of the line. Two accuracy parameters: Absolute Error and the Percentage Error (PE) are calculated according to the below-mentioned formulae:

The average location error, as is observed from Table 9, is found out to be 0.1784 km which is about 0.1189% as computed from the above expression. The average deviation between the prediction and the target fault locations is found minimum for DL-CA fault which is in the range of 0.1146 km and is worst for LLL-ABC fault which is about 0.2625 km for the designed 150 km long overhead transmission line. The worst prediction percentage error was also found as 1.8157 km i.e., 1.2105% of PE for SLG-ABG fault. It is further found that DL faults produce marginally better average location prediction compared to ground faults like SLG or DLG faults. The performance of the model is again investigated for further higher level of noise of 25 dB SNR. The performance of the model using the same fitness model is also found quite accurate and these results are described in Table 10.

It is still observed from Table 10 that the proposed fault localizer is capable of producing an accurate result, even with this elevated noise level of 25 dB SNR. The average percentage prediction error is found 0.3965% and the maximum percentage error is found to be 3.8761% for a sample SLG-ABG fault. Thus, the overall performance of the location predictor model

$$\text{Absolute Error(AE in km)} = \text{Actual Fault Distance} - \text{Estimated Fault Distance} \tag{7}$$

$$\text{Percentage Error (PE)} = (\text{AE/Line Length}) \times 100 \tag{8}$$

Since noise contaminated current signals are used as working data in our work; the PCIs are prone to a minor variation due to the randomness of power system noise. Hence, each prediction is carried out three times for each signal in order to obtain an average outcome of fault location; thereby, reduce the effect of random noise. Table 9 further shows the summary of the maximum location error for different classes of faults. The expressions of fitness models are described in tabular form in "Appendix" in Table 11.

is again found high satisfactory, even at this high noise level.

We have further tested the proposed model for parametric variation of the line. We have designed a new line with different line parameters and applied the method. We found that the classifier method works well directly on the new line, but the localization method looses accuracy to some extent, as expected. But we have conducted faults on the new line and trained the best fit model using the PCA features from fault signals of the new line when we found that the model regains comparable accuracy. Hence we could confirm that the model could be applied to any other line with variation in line parameters, post training. Similarly, if the line is split into multiple segments with

different line parameters and connected in cascade; it would behave similarly. Since, the measurement is taken only at the receiving end only, even if the line is made up with multiple segments having different impedance levels, the proposed algorithm would work, provided the model is training with the fault signals of the new line. Thus the concept of variation in upstream impedance could be satisfied this way, even for the single end fed, radial transmission line. So the model would work with other upstream impedance as well, provided, the model is recalibrated each time when a variation of line impedance occurs; otherwise a unique model would fail to deliver the claimed accuracy.

5.3 Discussion

The results obtained so far are studied carefully and the following outcomes are highlighted as the key findings of this research work:

- Operation of the proposed scheme is faster due to the requirement of less than one cycle of data for analysis.
- This scheme requires lesser memory compared to other schemes like neural network of wavelet analysis. This is achieved in a sense that PCA extracts key features in terms of the principal components in the descending order of importance. Hence, consideration of a few most important directions only reduces the entire data set to a few sets of data, simultaneously retaining the most significant information with very low loss. Hence, the memory requirement reduces from storing a large data set to a very low one.
- Absolutely accurate ratio analysis process for faulty phase identification yielding a 100% correct result
- The proposed method is tested with two different levels of power line noise. The robustness of the scheme is verified even for an increased noise level of 25 dB SNR, at which the method is still found to work accurately.
- Accurate fault localization with average localization error of 0.1189% and maximum localization error of even less than 1.25% at SNR of 40 dB. The model is further tested at higher power line noise level of 25 dB SNR and average localization error of 0.3965% and maximum localization error of less than 4% is achieved.
- The proposed analysis is simple as it does not involve either supervised learning approaches like neural network or transform methods with intricate mathematical analysis like wavelet or Fourier transform, etc.
- The method requires only a single end data, which is another advantage of the scheme as it discards the requirement of synchronized data acquisition from both

ends, which involves additional hardware support, and hence, cost.
- The proposed algorithm is less sensitive toward unbalancing of load. Since we are converting the system to per unit model, any unbalance in load is automatically scaled, and more importantly in all three lines; although, their effects will be reflected in the per unit magnitude. Most importantly, PCA identifies the principal directions of variations only. The major effect during fault is the drastic and large disturbance of line current from the healthy condition. Hence, the minor effect of unbalancing of load is minimized to good extent using PCA. Since we have modeled the system using balanced condition, this unbalancing of load is found to introduce some error in the localization algorithm; although, it is much less sensitive toward the classifier model.
- The proposed model is also applicable to other lines with different line parameters, but recalibration of the model is required in each case for each separate line.

A comparative analysis of the different existing schemes would show that the proposed method is well justified as an effective fault analysis scheme. The proposed classifier scheme produces 100% of classification accuracy, which is the highest possible level of accuracy to be achieved. This accuracy level achieved in this work is marginally better than [5, 8, 14] which mostly uses supervised learning schemes and its advanced forms. The proposed method of classification also performs better than SVM-WT based methods adopted in [24]. The present output is also marginally better than [26] which uses SVM aided by discrete orthogonal S-transform (DOST); although the above researches compared here consider variable fault resistance, which is not followed in this work. We have simulated the faults with fixed fault resistance, rather than considering partial or high resistance faults; considering that faults on transmission lines does not usually occur due to high resistance. Other research works like [10, 35] have produced 100% classifier accuracy, as well as considered fixed fault resistance similar to the present study; hence, is found very much comparable to the present work in terms of the outcomes. The half cycle post fault cycle of fault signal required in this present study is also comparable to few of the other existing schemes [10, 14, 26, 35]. But most of the methods discussed here are high in computation where the present scheme takes an upper hand with low computational burden as it uses PCA as the only computational tool.

Accuracy of fault localization achieved using this scheme is also high with an error level of 1.25% only at SNR of 40 dB. This level of accuracy is better than many other methods like neural network based approaches

adopted in [4, 5, 8]. The accuracy of the present scheme is also found mostly higher compared to sequence network based schemes like [46, 47] or wavelet-neural network based approach like [10]. Many of the methods mostly use wide variation of fault resistance; especially the works carried out in [8, 46, 47] analyze high resistance faults, where it is considered more than 100Ω. But, this is not practiced in this present analysis, as also mentioned earlier. The hybrid WT-ANN based approach described in [10] has rather used fixed fault resistance; yet the proposed method is found to achieve higher accuracy compared to [10]. Hence, it can be stated that the proposed fault diagnosis scheme is simple in computation using PCA as the only method for feature extraction; as well as efficient both for classification and localization of transmission line faults, especially considering practical constraints like power line noise.

6 Conclusion

An efficient transmission line fault detection, classification and localization scheme has been developed in this work for a single end fed 150 km long overhead transmission line. Principal component Analysis (PCA) has been applied here to realize, design and implement the proposed protection algorithm in MATLAB environment. Fault current waveforms are measured at the receiving end for quarter cycle pre-fault and half cycle post-fault duration to design the algorithm. PC indices (PCI) have been computed from the PCA scores, which are used to develop a threshold based algorithm to identify and classify faults. The results show that the classifier produces 100% accurate classification using only three sets set of training data at intermediate locations. The method is simple and has less computational complexity, especially compared to different supervised learning schemes like neural network or other transform based methods possessing high complexity mathematical analysis. The proposed algorithm is further extended to develop a fault location prediction scheme. The PCI values corresponding to six intermediate locations are used to develop a best fit analysis. The average error of localization is only about 0.1784 km, i.e., 0.1189% with a maximum error of 1.2105% at 40 dB SNR level. The same algorithm, when tested with higher noise level of 25 dB SNR, produced an average error of 0.3965% with a maximum PE of 3.8761%. This is quite appreciable considering this high level of noise. Accurate distance prediction helps the personnel to identify the fault location at the nearly exact locations; thus, requires less effort to find the fault. Hence, the proposed algorithm has considerable contribution to actuate prompt and accurate circuit breaker operation and fast restoration of system stability.

Appendix

The details of the fitness models for a certain noise pattern are described in a tabular form as shown below (Table 11).

Table 11 Details of the fitness models

Fitness model	Parameters
Fit 1 model	Shape-preserving (pchip) interpolant:
	$fit1(x) =$ piecewise polynomial computed from p
	Coefficients:
	p = coefficient structure
Fit 2 model	General model Exp2:
	$fit2(x) = a*exp(b*x) + c*exp(d*x)$
Fit 3 model	Smoothing spline:
	$fit3(x) =$ piecewise polynomial computed from p
	Coefficients:
	p = coefficient structure
Fit 4 model	Cubic interpolating spline:
	$fit4(x) =$ piecewise polynomial computed from p
	Coefficients:
	p = coefficient structure
Fit 5 model	General model Gauss2:
	$fit5(x) = a1*exp(-((x-b1)/c1)^2) + a2*exp(-((x-b2)/c2)^2)$
Fit 6 model	Linear model Poly3:
	$fit7(x) = p1*x^3 + p2*x^2 + p3*x + p4$

References

1. Mishra DP, Ray P (2017) Fault detection, location and classification of a transmission line. Neural ComputAppl 30(5):1377–1424
2. Mukherjee A, Kundu PK, Das A (2021) Transmission line faults in power system and the different algorithms for identification, classification and localization: a brief review of methods. J InstitutEngSer B 1:23. https://doi.org/10.1007/s40031-020-00530-0
3. Smith LI (2002) A tutorial on principal components analysis
4. Jain A, Thoke AS, Patel RN (2009) Double circuit transmission line fault distance location using artificial neural network. World Congress on Nature & Biologically Inspired Computing (NaBIC), IEEE, Coimbatore, India
5. Roy N, Bhattacharya K (2015) Detection, classification, and estimation of fault location on an overhead transmission line using S-transform and neural network. Electric Power Components Syst 43(4):461–472
6. Ngaopitakkul A, Leelajindakrairerk M (2018) Application of probabilistic neural network with transmission and distribution protection schemes for classification of fault types on radial, loop, and underground structures. ElectrEng 100(2):461–479
7. Akmaz D, Mamiş MS, Arkan M, Tağluk ME (2018) Transmission line fault location using traveling wave frequencies and extreme learning machine. Electric Power Syst Res 155:1–7
8. Chen YQ, Fink O, Sansavini G (2017) Combined fault location and classification for power transmission lines fault diagnosis

with integrated feature extraction. IEEE Trans Industr Electron 65(1):561–569

9. Rao HG, Prabhu N, Mala RC (2020) Wavelet transform-based protection of transmission line incorporating SSSC with energy storage device. Electr Eng, pp 1–12

10. Dasgupta A, Nath S, Das A (2012) Transmission line fault classification and location using wavelet entropy and neural network. Electric Power Components Syst 40(15):1676–1689

11. Gowrishankar M, Nagaveni P, Balakrishnan P (2016) Transmission line fault detection and classification using discrete wavelet transform and artificial neural network. Middle-East J Sci Res 24(4):1112–1121

12. Abdullah A (2017) Ultrafast transmission line fault detection using a DWT-based ANN. IEEE Trans IndAppl 54(2):1182–1193

13. Silveira EG, Paula HR, Rocha SA, Pereira CS (2018) Hybrid fault diagnosis algorithms for transmission lines. ElectrEng 100(3):1689–1699

14. Vyas BY, Maheshwari RP, Das B (2014) Improved fault analysis technique for protection of Thyristor controlled series compensated transmission line. Int J Electr Power Energy Syst 55:321–330

15. Mahmud MN, Ibrahim MN, Osman MK, Hussain Z (2018) A robust transmission line fault classification scheme using class-dependent feature and 2-Tier multilayer perceptron network. ElectrEng 100(2):607–623

16. Swetapadma A, Yadav A (2015) All shunt fault location including cross-country and evolving faults in transmission lines without fault type classification. Electric Power Syst Res 123:1–12

17. Yadav A, Swetapadma A (2015) Enhancing the performance of transmission line directional relaying, fault classification and fault location schemes using fuzzy inference system. IET GenerTransmDistrib 9(6):580–591

18. Reddy MJ, Mohanta DK (2007) A wavelet-fuzzy combined approach for classification and location of transmission line faults. Int J Electr Power Energy Syst 29(9):669–678

19. Eristi H (2013) Fault diagnosis system for series compensated transmission line based on wavelet transform and adaptive neuro-fuzzy inference system. Measurement 46(1):393–401

20. Meyur R, Pal D, Sundaravaradan NA, Rajaraman P, Srinivas KVVS, Reddy MJB, Mohanta DK (2016) A wavelet-adaptive network based fuzzy inference system for location of faults in parallel transmission lines. In: International Conference on Power Electronics, Drives and Energy Systems (PEDES), IEEE, Trivandrum, India

21. Reddy MJ, Mohanta DK (2007) A wavelet-neuro-fuzzy combined approach for digital relaying of transmission line faults. Electric Power Components Syst 35(12):1385–1407

22. Salat R, Osowski S (2004) Accurate fault location in the power transmission line using support vector machine approach. IEEE Trans Power Syst 19(2):979–986

23. Samantaray SR, Dash PK, Panda G (2007) 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

24. Bhalja B, Maheshwari RP (2008) Wavelet-based fault classification scheme for a transmission line using a support vector machine. Electric Power Components Syst 36(10):1017–1030

25. Reddy MJB, Gopakumar P, Mohanta DK (2016) A novel transmission line protection using DOST and SVM. EngSciTechnolInt J 19(2):1027–1039

26. Patel B (2018) A new FDOST entropy based intelligent digital relaying for detection, classification and localization of faults on the hybrid transmission line. Electric Power Syst Res 157:39–47

27. Mukherjee A, Kundu PK, Das A (2021) Classification and fast detection of transmission line faults using signal entropy. J InstitutEngSer B 1:16. https://doi.org/10.1007/s40031-020-00526-w

28. Lopes FV, Dantas KM, Silva KM, Costa FB (2017) Accurate two-terminal transmission line fault location using traveling waves. IEEE Trans Power Delivery 33(2):873–880

29. Ma G, Jiang L, Zhou K, Xu G (2016) A Method of line fault location based on traveling wave theory. Int J Control Autom 9(2):261–270

30. Barman S, Roy BKS (2018) Detection and location of faults in large transmission networks using minimum number of phasor measurement units. IET GenerTransmDistrib 12(8):1941–1950

31. Devi MM, Geethanjali M, Devi AR (2018) Fault localization for transmission lines with optimal Phasor Measurement Units. ComputElectrEng 70:163–178

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

33. Taheri B, Salehimehr S, Razavi F, Parpaei M (2020) Detection of power swing and fault occurring simultaneously with power swing using instantaneous frequency. Energy Syst 11(2):491–514

34. Krishnanand KR, Dash PK, Naeem MH (2015) Detection, classification, and location of faults in power transmission lines. Int J Electr Power Energy Syst 67:76–86

35. Mukherjee A, Kundu P, Das A (2014) Identification and classification of power system faults using ratio analysis of principal component distances. Indonesian J ElectrEngComputSci 12(11):7603–7612

36. Mukherjee A, Kundu PK, Das A (2020) Power system fault identification and localization using multiple linear regression of principal component distance indices. Int J Appl Power Eng 9(2):113–126

37. Mukherjee A, Kundu PK, Das A (2020) Transmission line fault location using PCA-based best-fit curve analysis. J InstitutEng India Ser B. https://doi.org/10.1007/s40031-020-00515-z

38. Sinha AK, Chowdoju KK (2011) Power system fault detection classification based on PCA and PNN. In: 2011 international conference on emerging trends in electrical and computer technology, IEEE

39. Mukherjee A, Kundu PK, Das A (2020) Application of principal component analysis for fault classification in transmission line with ratio-based method and probabilistic neural network: a comparative analysis. J InstEng India Ser B 101(4):321–333

40. Jafarian P, Sanaye-Pasand M (2010) A traveling-wave-based protection technique using wavelet/PCA analysis. IEEE Trans Power Delivery 25(2):588–599

41. Guo Y, Li K, Liu X (2012) Fault diagnosis for power system transmission line based on PCA and SVMs. In: International conference on intelligent computing for sustainable energy and environment, Springer, Berlin, Heidelberg, pp 524–532

42. IEEE Guide for Determining Fault Location on AC Transmission and Distribution Lines, in IEEE Std C37.114–2014 (Revision of IEEE Std C37.114–2004), Jan. 2015, doi: https://doi.org/10.1109/IEEESTD.2015.7024095

43. Kasztenny B, Mynam MV, Fischer N (2019) Sequence component applications in protective relays–advantages, limitations, and solutions. In: 2019 72nd annual conference for protective relay engineers (CPRE), (College Station, TX, USA), pp 1–23

44. Garcia PAN, Pereira LR, Vinagre MP, Oliveira EJ (2004) Fault analysis using continuation power flow and phase coordinates. In: IEEE power engineering society general meeting, IEEE, pp 872–874

45. Mora-Florez J, Melendez J, Carrillo-Caicedo G (2008) Comparison of impedance based fault location methods for power distribution systems. Electric Power Syst Res 78(4):657–666

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

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

48. Biswal M, Biswal S (2017) A positive-sequence current based directional relaying approach for CCVT subsidence transient condition. Protect Control Modern Power Syst 2(1):1–11

49. Adly AR, Ali ZM, Abdel-hamed AM, Kotb SA, Mageed HMA, Aleem SHA (2019) Enhancing the performance of directional relay using a positive-sequence superimposed component. Electr Eng, pp 1–19

50. Camacho A, Castilla M, Miret J, de Vicuña LG, Guzman R (2017) Positive and negative sequence control strategies to maximize the voltage support in resistive–inductive grids during grid faults. IEEE Trans Power Electron 33(6):5362–5373

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.