MULTI-SENSOR-BASED FAULT DETECTION AND CLASSIFICATION METHOD FOR RADIAL POWER DISTRIBUTION SYSTEMS

A Thesis by

Nan Wang

Bachelor of Engineering, Information Engineering University, 2011

Submitted to the Department of Electrical Engineering and Computer Science
and the faculty of the Graduate School of
Wichita State University
in partial fulfillment of
the requirements for the degree of
Master of Science

December 2014

MULTI-SENSOR-BASED FAULT DETECTION AND CLASSIFICATION METHOD FOR RADIAL POWER DISTRIBUTION SYSTEMS

The following faculty members have examined the final copy of this thesis for form and content, and recommend that it be accepted in partial fulfillment of the requirement for the degree of Master of Science with a major in Electrical Engineering.

_____
Visvakumar Aravinthan, Committee Co-Chair

_____
Yanwu Ding, Committee Co-Chair

_____
Ward T. Jewell, Committee Member

_____
Gamal Weheba, Committee Member

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my advisers, Dr. Yanwu Ding and Dr. Visvakumar Aravinthan. They patiently guided me through my research project, providing advice on both my academic research and graduate life. When I encountered problems, they always supported me, encouraged me, and helped me resolve them. Without these two people, I could not have achieved what I did.

My sincere gratitude goes to my thesis committee members. Dr. Ward T. Jewell, with his humility and wisdom, was always willing to guide and help me. Dr. Gamal Weheba, with his enthusiasm and patience, helped me develop my statistics skills. Thank you both for your time and support throughout my research and graduate life.

I especially thank all of my friends in the PDS Reliability Research Group at Wichita State University. They were like my family in Wichita and selflessly provided their help and encouragement when I need them the most.

I also want to thank all my friends in Wichita for a most wonderful time. Through happiness or sorrow, they were my company and have become a part of my life.

Last, but not least, I express the most gratitude to my parents for their sustainable and constant support, both materially and mentally. I love them to the depth, breadth, and height that my soul can reach.

ABSTRACT

A novel multi-sensor-based method to detect and classify short circuit faults in radial power distribution systems, including the effects of regulators and distribution transformers, is proposed in this work. This new scheme first calculates the correlation among the data of different sensors and uses principal component analysis (PCA) to reduce the data dimensions. Then, kernel support vector machine (SVM) classifiers are applied to detect faults and identify faulty phases. The proposed method is simulated and tested for normal operations like load switching and different types of faults under different signal-to-noise ratio (SNR) scenarios with various fault durations and fault impedances. The Gaussian mixture noise model is used in the simulations to test the robustness of the algorithm. Two distribution system models (an unbalanced feeder and a balanced feeder) are used in this work to determine the impacts of system configuration on the proposed method. Finally, relationships among the number of sensors, sampling rate of the sensors, and detection performance are discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS/NOMENCLATURE

AWGN	Additive White Gaussian Noise
EHV	Extra-High Voltage
FLS	Fuzzy Logic System
GMD	Gaussian Mixture Density Function
IEEE	Institute of Electrical and Electronics Engineers
KKT	Karush-Kuhn-Tucker
PCA	Principal Component Analysis
RBF	Radial Basis Function
SNR	Signal-to-Noise Ratio
SVM	Support Vector Machine

CHAPTER 1

INTRODUCTION

1.1 Background

The European Union Smart Grid Projects [1], the U.S. Department of Energy Smart Grid Initiative [2], and other smart grid initiatives require self-healing distribution systems that are capable of intelligently detecting, isolating, and classifying faults, and recovering from an abnormal event to minimize any interruption duration. It is important to detect and classify faults in power distribution systems, which will result in faster restoration. Usually, traditional overcurrent relays have a poor performance of detecting high-impedance faults and do not have the capability of classifying the type of fault. In addition, due to complicated system topologies and dynamic loading conditions, current detection practices could take tens of minutes to hours to detect a fault by means of manual searching operations, which is a limitation for the implementation of smart grids [3]. Therefore, an accurate and efficient fault-detection and fault-classification scheme is crucial and essential to power distribution systems.

1.2 Motivation

Signal processing techniques have been applied to solve many kinds of problems in electric power systems, such as harmonics identification, power quality assessment, and fault analysis. With the deployment of new technologies to improve grid operations (smart grid initiatives), a significant number of sensors and pieces of monitoring equipment are expected to be installed in distribution systems, hence making multi-sensor-based detection and classification methods possible and practical. Instead of using only one advanced sensor or monitoring device at the substation, a multi-sensor-based approach could provide better performance and generalization, especially when multiple sensors at different locations of the system are used.

Using only one sensor to measure the data may be insufficient at times, especially when faults occur on a lateral of the system with complicated topology.

In addition, none of the methods above include phase-shifting transformers or regulators in their analysis. Regulators and transformers, which are commonly used in modern distribution systems, affect the data nonlinearly between the primary side and the secondary side of these elements. Figure 1 shows the three-phase current waveforms when a line-line fault occurs at the secondary side of a wye-delta transformer. Figure 1(a) shows primary-side three-phase current waveforms, and Figure 1(b) shows secondary-side waveforms. These three-phase waveforms have different patterns at the primary and secondary sides of the transformer.

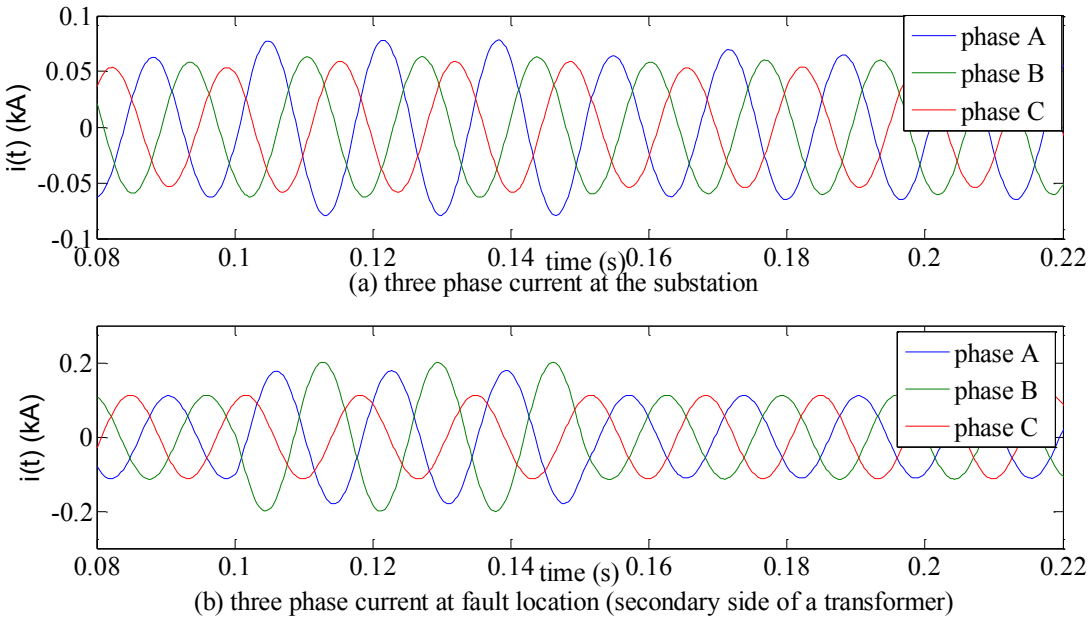

Figure 1: Line-line fault at secondary side of transformer

Multiple sensors could provide more accurate data for fault analysis when there are regulators and transformers in the systems. In addition, using multiple sensors could reduce the effect of the noise disturbance and hence to have a good performance even under low signal-to-noise ratio (SNR) environments.

1.3 Contributions of Thesis

A novel multi-sensor-based scheme to detect and classify short-circuit faults (line-ground, line-line-ground, line-line, three-line) at different locations of radial distribution systems, and to determine the effect of number of sensors and sampling frequency is proposed in this work. All sensors are three-phase current meters and are applied at different locations in the distribution system. Data from these different sensors are sent to a fusion center and integrated as a multi-dimensional data set. Using correlations of the different dimensions/variables, a principal component analysis (PCA) is applied to reduce the dimensions in order to simplify the computation complexity. Finally, a kernel support vector machine (SVM) is used on the preprocessed data to detect faults. Upon detecting a fault, the original signals are divided into four parts, and separate-phase SVM classifiers are used to identify the faulty phases.

The contributions of this work are summarized as follows: (1) The multi-sensor-based scheme detects and classifies possible power distribution-level faults from different locations of the system, including the effects of regulators and distribution transformers [4]. (2) An appropriate noise model for power system communication and associated sensors, a Gaussian mixture noise model, is introduced and simulation results are presented. (3) A relationship between the number of sensors and the detection accuracy, and also between the sampling rate of the sensors and the detection accuracy, is developed in this work. (4) Two distribution system models—an unbalanced IEEE 13-node feeder and a balanced IEEE 34-node test feeder—are used in this work to determine the impacts of system configuration on the proposed technique. (5) The proposed method is simulated and tested for all types of faults under different SNR scenarios with various fault durations and fault impedances. This work includes signals due to normal operating conditions like load switching to check for robustness.

1.4 Outside Scope of Work

The following will not affect the model proposed in this work. However, evaluating them requires additional analysis. Therefore they are not included in this work:

- A methodology to determine the optimal number of sensor numbers.
- A methodology for optimal sensor locations.
- An optimal data sampling rate for the proposed technique.

1.5 Organization of Thesis

The remainder of this thesis is organized as follows: Chapter 2 is a literature review of fault detection and classification methods for power distribution systems. The proposed fault detection and classification method is introduced in detail in Chapter 3, and the simulation and results are presented in Chapter 4. The conclusions and future work are provided in Chapter 5.

CHAPTER 2

LITERATURE REVIEW

Short-circuit faults occur in power systems when equipment insulation fails due to system over-voltages caused by switching surges or lightning, insulation contamination (pollution or salt spray), or other mechanical causes. The resulting short-circuit or ‘‘fault’’ current is determined by internal voltages of synchronous machines and by system impedances between the fault and machine voltages. The magnitude of short-circuit currents may be several orders larger than that of normal operating currents and, if allowed to persist, may cause thermal damage to equipment. Busbars and windings, too, may suffer mechanical damage, due to the high magnetic forces during faults. Therefore, it is necessary to remove faulted parts of a power system from service as soon as possible. Standard protective equipment for extra-high voltage (EHV), such as circuit breakers, fuses, relays, and instrument transformers, are designed to clear faults within three cycles (50 ms at 60 Hz). Protective equipment for lower voltage operates more slowly (for example, 5 to 20 cycles) [5].

Short circuits may cause severe damage when not interrupted promptly. In some scenarios, high-impedance fault currents may not be sufficient to operate blown fuses or protective relays. Standard overcurrent protection schemes utilized on secondary distribution at some large residential, commercial, and industrial buildings, may not detect high-impedance faults, commonly called arcing faults [5].

Short-circuit faults occur in three-phase power systems as follows, in order of frequency of occurrence: single line-ground, line-line, double line-ground, and balanced three-phase faults. The path of the fault current may have either zero impedance, which is called a bolted short

circuit, or nonzero impedance [5]. Figure 2 shows the current waveforms of different types of short-circuit faults .

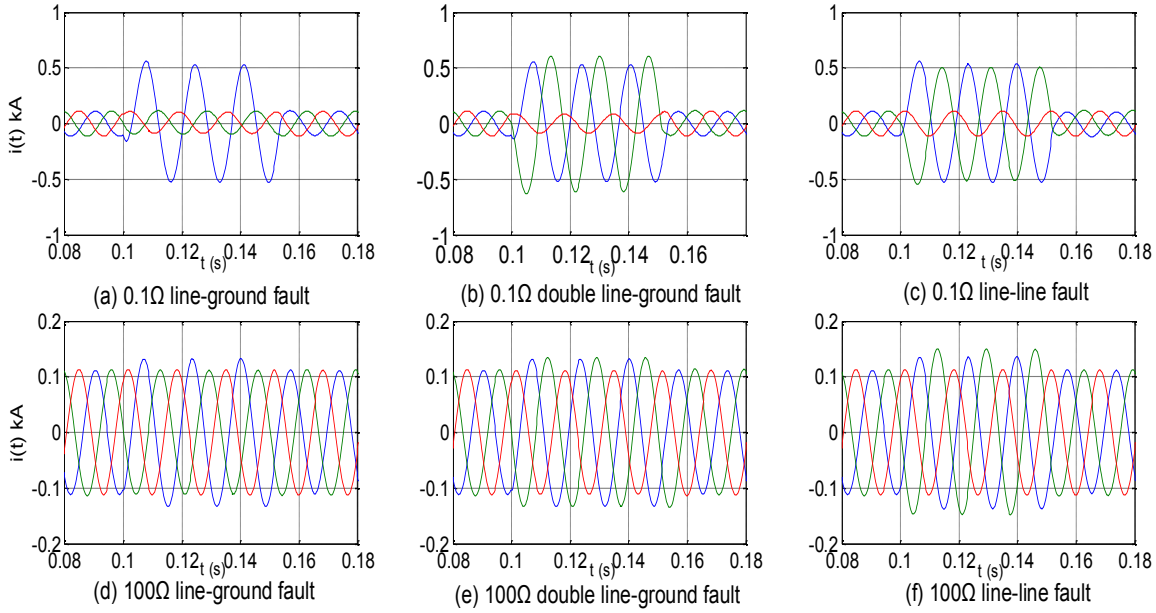

Figure 2: Different fault types and fault impedances of current waveforms

Among the various types of techniques used for fault detection and classification in distribution systems, the most widely used techniques are the following: (1) neural network approach [3],[6]–[9]; (2) fuzzy logic and fuzzy neural network-based methods [10], [11]; and (3) wavelet analysis combined with neural networks or support vector machines [12]–[16].

In the work of Xu and Chow [3], [6], six factors—weather condition, season, time of day, circuit ID, number of phases affected, and protective devices activated—were recorded and then sent to a neural network to classify the faults. Two types of faults were studied in these papers, and the neural network approach was compared with the logistic regression method for fault detection and classification. Results showed that the neural network method has a slightly higher accuracy. In the work of Butler and Momoh [7], a two-stage supervised clustering-based neural network was applied. Compared with the methods of Xu and Chow [3], [6], simpler sensors and measurements were used. First, the three-phase currents were measured, and then statistical

features such as skewness and kurtosis coefficients were computed and sent to a two-layer neural network. The outputs of the neural network were clustered, and could be classified by using the nearest mean decision rule. The performance reached about 0.92 for detection and 0.81 for classification. In the work of Oliveira et al. [8], [9], sequence currents were used instead of using phase currents, and a neural network method was proposed. The instantaneous current was first measured, and then positive, negative, and zero-sequence currents were calculated. Then a factor was calculated from the post-fault negative sequence and pre-fault positive sequence currents using Thevenin equivalents. This factor was defined as the F2 factor. By adding the F2 factor and other sequence components into a resilient back-propagation neural network model, the classification rate was improved to 0.94. Low-impedance cases were simulated (0 to 20 Ω), and only shunt faults and series faults were distinguished.

Usually these methods could achieve a fair accuracy, but they all have obvious limitations. The methods of Xu and Chow [3], [6] need advanced sensors to monitor different kinds of features, which can be very expensive to implement. The scheme of Butler and Momoh [7] was only tested in a three-node feeder, which is not representative of a real distribution system. The methods of Oliveira et al. [8], [9] could classify the shunt fault or series fault, but they do not determine faulty phases or the exact fault type (one-phase-ground fault, two-phase fault, etc.).

Fuzzy logic and fuzzy neural networks are also popular methods to identify faults. Samantaray et al. [10] proposed a fuzzy logic-based fault-classification method based on the angular difference among sequence components of fault currents. Then eight fuzzy variables were introduced, fault identification rules were proposed, and distribution systems with different fault types and fault impedances were simulated. In the work of Xu et al. [11], the E-algorithm,

which was extended form the fuzzy algorithm, was proposed to solve the data imbalance issue. By comparing results with the neural network method, the E-algorithm greatly improved performance when the dataset was imbalanced. However, the fuzzy logic or fuzzy extended algorithms still have some shortcomings. The angular difference-based fuzzy logic method does not work in systems with distribution transformers and regulators, while the E-algorithm is highly computational demanding.

Another widely used method has been to use wavelet analysis combined with neural networks, SVMs, and a rule-based procedure to detect and classify faults [12]–[16]. Usually, transient three-phase voltages and/or currents are measured at a high sampling rate, and then the wavelet analysis is implemented to extract the features of energy at different frequencies. The features are used as the input of neural network models or SVM classifiers to detect and classify faults. Dag and Ucak [12] applied a hybrid wavelet-neural network-based approach to classify faults. The database of line currents and line-to-ground voltages including system faults was computed by simulating the Sagmalcilar-Maltepe distribution system in Istanbul, Turkey. Dwivedi et al. [13] did the first wavelet multi-resolution decomposition of a three-phase current (measured at the substation end), and then operated on wavelet coefficients obtained from the decomposition to follow the rule-based procedures to identify faults. The work of Assef et al. [14] and Samantaray et al. [15] are similar to this. First, a wavelet transform of the three-phase currents was performed to extract the features. Then the neural network was used to distinguish the different faults. The differences of these two works are as afollows: In the work of Samantaray et al. [15], an extension of the wavelet transform, called the S-transform, and its extension, called the TT-transform, were adopted. In the work of Assef et al. [14], the supervised Kohonen network [14] was applied, and in the work of Samantaray et al. [15], the probabilistic

neural network was applied. Both of them performed well. In the work of Livani and Evrenosoglu [16], wavelet analysis was combined with four SVM classifiers to identify whether each phase had a fault or not.

These methods achieve a high accuracy, over 90% in classification, but there still exists some limitations with the wavelet-based methods. In the work of Dag and Ucak [12], ten cycles of three-phase currents and voltages were required for wavelet analysis, which needs a considerable amount of storage and can take too long in practical cases. For wavelet analysis, high sampling rates of 100 kHz [14] and 200 kHz [16] were required, which in practice is critical to the sensors and processors. In addition, the wavelet-based methods are highly prone to noise. When the SNR decreases to 30 dB, the accuracy decreases dramatically [15].

Among these techniques, regardless of the neural networks, SVMs, or rule-based fuzzy logic procedures, all of them do not work well in high-noise situations or in systems with distribution transformers and regulators. The key element here is to extract those features that contain abundant information of the system status and fault properties. The data from the sensor at the substation end is rather monotonous. However, the multi-sensor data can record the information of the entire system, and the correlations of data from different locations of the system can improve the detection and classification performance of these scenarios.

CHAPTER 3

FAULT DETECTION AND CLASSIFICATION MODULES

3.1 Multi-Sensor Distribution Systems

Figure 3 presents a typical radial distribution feeder, where "SS" stands for the substation; "T" is a distribution transformer, "L_i" is the *i*th load, where I = 1,2,…m; and "S_i" is the *i*th sensor, where i = 1,2,…n. For each event, sensor i collects one cycle of the three-phase current data and sends that information to the fusion center.

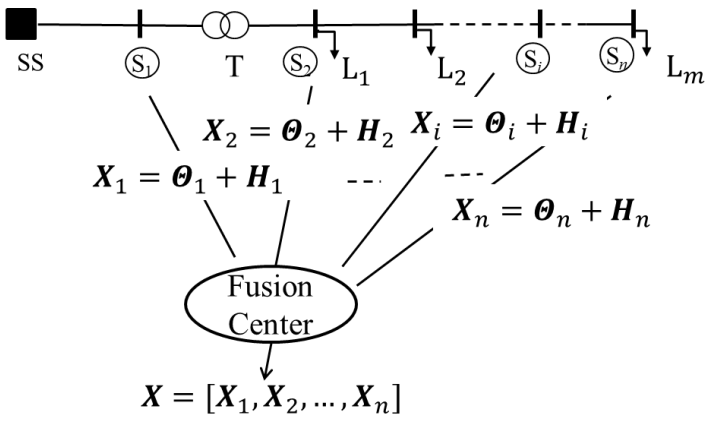

Figure 3: Multi-sensor distribution system

The output of sensor i, $\boldsymbol{X}_i$ is a three-dimensional signal (three dimensions represent three-phase current values), which is modeled as

$$\boldsymbol{X}_i = \boldsymbol{\Theta}_i + \boldsymbol{H}_i, i = 1,2,...,n \tag{1}$$

where $\boldsymbol{\Theta}_i$ is the three-phase current magnitude vector, and $\boldsymbol{H}_i$ is the noise. Previous work has seldom applied noise models in the algorithms, whereas noise does exist in real cases. In the later part of this section, the power line noise modeling is presented.

The fusion center receives m noise-riding observations during one-cycle period; thus, $\boldsymbol{X}_i$ has a dimension of m by 3. All data received at the fusion center are then integrated into a matrix,

$\boldsymbol{X}$, for analysis, where $\boldsymbol{X}$ is represented as

$$\boldsymbol{X} = \begin{bmatrix} x_{1,1\mathrm{A}} & x_{1,1\mathrm{B}} & x_{1,1\mathrm{C}} & \cdots & x_{1,n\mathrm{A}} & x_{1,n\mathrm{B}} & x_{1,n\mathrm{C}} \\ x_{2,1\mathrm{A}} & x_{2,1\mathrm{B}} & x_{2,1\mathrm{C}} & \cdots & x_{2,n\mathrm{A}} & x_{2,n\mathrm{B}} & x_{2,n\mathrm{C}} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{m,1\mathrm{A}} & x_{m,1\mathrm{B}} & x_{m,1\mathrm{C}} & \cdots & x_{m,n\mathrm{A}} & x_{m,n\mathrm{B}} & x_{m,n\mathrm{C}} \end{bmatrix}_{m\times 3n} \tag{2}$$

where the elements are the observation from different sensors and phases (for example, $x_{2,1\mathrm{B}}$ is the second observation from phase B of sensor one.)

In power line communication systems, data are corrupted with noise. Because impulsive noise due to the switching transients is also contributed to systems, the noise source cannot be properly modeled by only the additive white Gaussian noise (AWGN) [17], [18]. In this work, noise in the distribution communication system is modeled as a Gaussian mixture noise, which covers the effects of both thermal noise and impulsive noise [17], [19]. If the variable a is discrete and positive with the probability $p(a = a_i) = p_i$, and vector **u** has a Gaussian density with mean μ and covariance matrix $\sum$, then vector **x** = a**u** has a Gaussian mixture density function (GMD) [19] as

$$p_{\mathbf{x}}(\mathbf{x}) = \sum_{i=1}^{m} p_i N\left(a_i\mu, a_i^2 \sum\right)(\mathbf{x}) \tag{3}$$

In this work, since the sensors are three-phase current meters, for each sensor, the noise is modeled as a Gaussian mixture noise with three equally weighted components ($p_1 = p_2 = p_3 = 1/3$). Each component is a Gaussian distribution with zero mean, and the variance σ^2 can be calculated using

$$SNR_{dB} = 10\lg\left(\frac{A^2{}_{signal}}{\sigma^2}\right) \tag{4}$$

where SNR_{dB} is the given signal-to-noise ratio in dB, and A_{signal}^2 is the signal power. Figure 4 plots a sinusoidal signal waveform, with 15 dB white Gaussian noise and 15 dB Gaussian mixture noise. From this figure, the Gaussian mixture noise has obvious impulses compared with the white Gaussian noise.

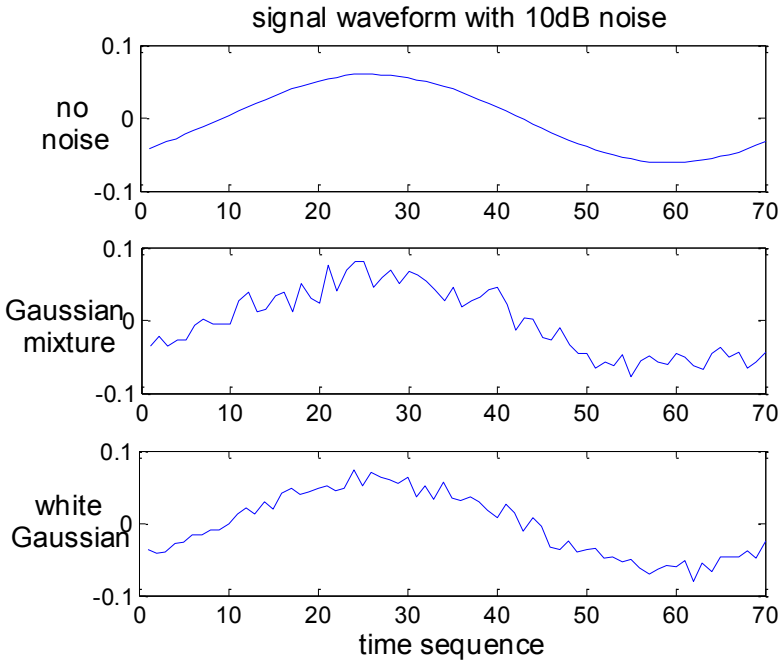

Figure 4: Signal waveform with 10 dB noise

All possible distribution feeder-level short-circuit faults are considered in this work. Figure 5 shows four examples of time domain signal waveforms corrupted with Gaussian mixture noise. The first row illustrates two different fault-current waveforms, and the second row shows the current waveform plots under normal operations. As shown, sometimes the waveform envelope of normal operations, such as load switching, can be similar to that of faults. When the fusion center receives the data from all sensors, the detection module is first applied. In Figure 5, the detection module is expected to determine the first row as fault and the second row as no fault. If a fault is detected, then the classification module is applied to distinguish the different types of faults.

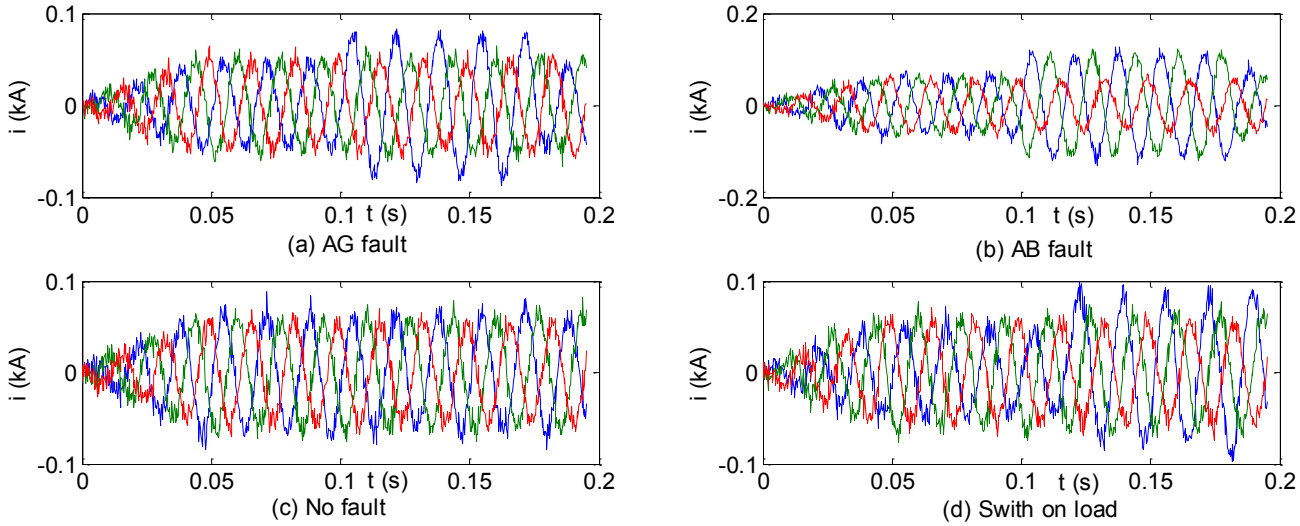

Figure 5: Three-phase current signals with 15 dB Gaussian mixture noise

Both detection and classification modules use a similar signal processing scheme. First, the correlation of the signals among different sensors and phases is calculated, and then the principal component analysis is used to reduce the dimension and extract the features. Finally, the processed features are sent to a kernel SVM to detect the fault or classify the faulty phases.

3.2 Fault-Detection Module

The flow chart of the fault detection module is shown in Figure 6. The original data set is the matrix $\boldsymbol{X}$ in the fusion center, with a size $m \times 3n$.

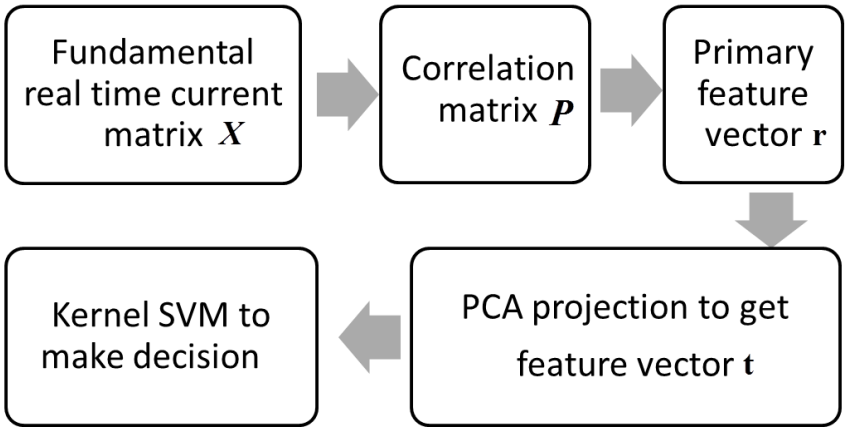

Figure 6: Flow chart of fault detection module

3.2.1 Correlation Matrix and Primary Feature Vector

In matrix equation (2), every column of $\boldsymbol{X}$ can be considered as a one-dimension variable. The correlation matrix $\boldsymbol{P}$ is a $3n \times 3n$ matrix. The entry $\rho_{i,j}$ is the correlation coefficient between the i^{th} and j^{th} columns of matrix $\boldsymbol{X}$, and can be calculated as

$$\rho_{i,j} = \frac{\sum_{k=1}^{m}\left(x_{ki} - \overline{x}_i\right)\left(x_{kj} - \overline{x}_j\right)}{\sqrt{\sum_{k=1}^{m}\left(x_{ki} - \overline{x}_i\right)^2}\sqrt{\sum_{k=1}^{m}\left(x_{kj} - \overline{x}_j\right)^2}} \tag{5}$$

where $\overline{x}_i$ and $\overline{x}_j$ are the mean values of columns *i* and *j*, respectively. The correlation coefficient is the normalized covariance between two variables. Since the correlation matrix is symmetrical, and the diagonal entries are equal to one, only the upper triangular entries (without the diagonal entries) are selected and reshaped to create a primary feature vector $\mathbf{r}$. Let $N = (3n-1)\cdot 3n/2$, then

$$\mathbf{r} = [\rho_{1,2}, \rho_{1,3}, ..., \rho_{1,3n}, \rho_{2,3}, ..., \rho_{2,3n}, ..., \rho_{(3n-1),3n}]_{1\times N} \tag{6}$$

3.2.2 PCA Dimension Reduction

Upon the deployment of a large number of sensors, it is important to reduce the dimension of the data sets, in order to reduce the computational complexity and storage expenses. Since the signals are often correlated, principal component analysis is applied to reduce the dimension of the primary feature vector in equation (6). Wise and Gallagher [20] have suggested that PCA is ideal to reduce the dimension and noise of high-dimensional correlated data. Two steps are involved in the PCA method: first, a training data set is used to calculate the principal components, and then the rest of the data can be projected on the principal components to reduce the dimension.

To use PCA in this scheme, the first step is to obtain M historical data sets, then M primary feature vectors $\mathbf{r}_i$, where $i = 1,\ldots M$ can be calculated using equations (5) and (6). The matrix $\boldsymbol{R}$ is integrated from the vectors as

$$\boldsymbol{R} = \begin{bmatrix} \mathbf{r}_1 \\ \vdots \\ \mathbf{r}_M \end{bmatrix} = \begin{bmatrix} r_{11} & \cdots & r_{1N} \\ \vdots & \ddots & \vdots \\ r_{M1} & \cdots & r_{MN} \end{bmatrix}_{M \times N} \tag{7}$$

Then the covariance matrix $\boldsymbol{C}$ [21] of the primary feature matrix, $\boldsymbol{R}$, is calculated, and $c_{i,j}$, the elements of the covariance matrix $\boldsymbol{C}$, is given as

$$c_{i,j} = \frac{1}{M} \sum_{k=1}^{M} \left(r_{ki} - \bar{r}_i\right)\left(r_{kj} - \bar{r}_j\right) \tag{8}$$

where $\bar{r}_i$ and $\bar{r}_j$ are means of the i^{th} and j^{th} columns of $\boldsymbol{R}$, respectively. By using Eigen decomposition,

$$\lambda \mathbf{v} = \boldsymbol{C}\mathbf{v} \tag{9}$$

which yields eigenvectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_N$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_N$ [21]. By selecting the first p eigenvectors as the principal components to form the principal components matrix,

$$\boldsymbol{V} = [\mathbf{v}_1, \mathbf{v}_2, ... \mathbf{v}_p]_{N \times p} \tag{10}$$

where parameter p is selected based on experiments to have an optimized performance on the training data. After the principal components matrix $\boldsymbol{V}$ is calculated, the dimension of the primary feature vectors can be reduced by projecting them onto the components matrix $\boldsymbol{V}$ as

$$\boldsymbol{T} = [\bar{\boldsymbol{R}}\boldsymbol{V}]_{M \times p} \tag{11}$$

where $\bar{\boldsymbol{R}}$ is the centralized $\boldsymbol{R}$ and is computed using equation (12):

$$\bar{\boldsymbol{R}} = \begin{bmatrix} r_{11} - \bar{r}_1 & \cdots & r_{1N} - \bar{r}_N \\ \vdots & \ddots & \vdots \\ r_{M1} - \bar{r}_1 & \cdots & r_{MN} - \bar{r}_N \end{bmatrix}_{M \times N} \tag{12}$$

and $\bar{r}_i$ is the mean of i^{th} column of $\boldsymbol{R}$. When another event records a data set $\boldsymbol{X}_{new}$ and calculates the primary feature vector $\mathbf{r}_{new}$, then the dimension reduced feature vector $\mathbf{t}$ can be obtained by

$$\mathbf{t} = [\bar{\mathbf{r}}_{new} \boldsymbol{V}]_{1 \times p} \tag{13}$$

where the centralized vector is calculated by $\bar{\mathbf{r}}_{new} = [\mathbf{r}_{new1} - \bar{\mathbf{r}}_1, \mathbf{r}_{new2} - \bar{\mathbf{r}}_2, ..., \mathbf{r}_{newN} - \bar{\mathbf{r}}_N]$. After applying PCA, the dimension of the feature vectors can be dramatically reduced from $N = (3n-1) \cdot 3n/2$ to p.

3.2.3 Kernel SVM Binary Classifier

The support vector machine was originally invented by Cortes and Vapnik as a supervised learning model [22]. It has been developed and applied in many fields, such as regression, estimation, and classification. Cortes and Vapnik were also the first to introduce SVM for binary classification [23]. The main idea behind the binary SVM classifier is to use the training data set of two classes to build an optimal hyperplane in hyperspace and then to classify the new input data. A brief introduction of the SVM classifier is presented in this subsection.

The training data, $\{\mathbf{t}_i, y_i\}$, i=1,2,…, K, are two classes of linearly separable data. Vector $\mathbf{t}_i$, called the data point, is the transposed i^{th} dimension-reduced feature vector in the training data set and can be calculated by using equation (13). The variable $y_i \in \{\pm 1\}$ is the corresponding class with $\mathbf{t}_i$. A hyperplane that separates the points of two classes is given by the following equation [23]:

$$\mathbf{w}^T \mathbf{t}_i + b = 0 \tag{14}$$

where **w** represents the weight vector, and *b* is the bias term. The separation margin between the two classes is defined as [23]

$$M_d = 2 / \|\mathbf{w}\| \tag{15}$$

The SVM will find the values of **w** and *b*, so that the separation margin between the classes is the largest. Once **w** and *b* are determined, the optimal hyperplane for the two classes is determined. In order to maximize M_d , it is necessary to minimize $\|\mathbf{w}\|$. The SVM problem can be summarized as [23]

$$\begin{aligned} &min \ \ \|\mathbf{w}\|^2 / 2 \\ &s.t. \ \ y_i\left(\mathbf{w}^T\mathbf{t}_i + b\right) \geq 1 \ \ , i = 1,2,...,K \end{aligned} \tag{16}$$

This problem can be solved by its dual problem in terms of Lagrange multipliers and can be represented as [23]

$$\begin{aligned} &max \ L(\boldsymbol{\alpha}) = \sum_{i=1}^{K}\alpha_i - \frac{1}{2}\sum_{i=1}^{K}\sum_{j=1}^{K}\alpha_i\alpha_j y_i y_j \left\langle \mathbf{t}_i, \mathbf{t}_j \right\rangle \\ &s.t. \ \ \sum_{i=1}^{K} y_i\alpha_i = 0 \\ &\quad \alpha_i \geq 0 \ \ , i = 1,2,...,K \end{aligned} \tag{17}$$

This quadratic programming problem can be solved effectively by using the sequential minimal optimization algorithm [24]. According to the Karush-Kuhn-Tucker (KKT) complementary conditions, the optimal solutions $\alpha_i^*, \mathbf{w}^*, b^*$ must satisfy

$$\alpha_i^*\left[y_i\left(\left\langle \mathbf{w}^*, \mathbf{t}_i \right\rangle + b^*\right) - 1\right] = 0 \ \ , i = 1,2,...K \tag{18}$$

which implies that only data points $\mathbf{t}_i$ that lie closest to the optimal hyper-plane have the corresponding non-zero α_i^* [23]. Let all these non-zero $\alpha_i^* \in sv$. Finally, the SVM binary classifier can be represented as [23]

$$f(\mathbf{t}) = sign(\sum_{\alpha_i^* \in sv} \alpha_i^* y_i \langle \mathbf{t}, \mathbf{t}_i \rangle + b^*) \tag{19}$$

$$b^* = \frac{1}{2}\left[max_{y_j=1}\left(\sum_{\alpha_i^* \in sv} \alpha_i^* \langle \mathbf{t}_j, \mathbf{t}_i \rangle\right) - min_{y_j=1}\left(\sum_{\alpha_i^* \in sv} \alpha_i^* \langle \mathbf{t}_j, \mathbf{t}_i \rangle\right)\right] \tag{20}$$

When a feature vector **t** is input into the SVM classifier, the value can be calculated by equation (19). Based on the sign of the value, this can be sorted into the corresponding class. However, data points cannot always be linearly separated. In this situation, the kernel trick can be used to solve the nonlinear classification problems. According to Cover's theorem, the nonlinearly separable data are more likely to be linearly separated after mapping the data points to a high-dimensional space as [25]

$$\Phi: \ R^p \to F^h, \ \mathbf{t} \to \phi(\mathbf{t}), \quad (h \gg p) \tag{21}$$

and equations (19) and (20) can be reformulated as [23]

$$f(\mathbf{t}) = sign(\sum_{\alpha_i^* \in sv} \alpha_i^* y_i \langle \phi(\mathbf{t}), \phi(\mathbf{t}_i) \rangle + b^*) \tag{22}$$

$$b^* = \frac{1}{2}\left[max_{y_j=1}\left(\sum_{\alpha_i^* \in sv} \alpha_i^* \langle \phi(\mathbf{t}_i), \phi(\mathbf{t}_j) \rangle\right) - min_{y_j=1}\left(\sum_{\alpha_i^* \in sv} \alpha_i^* \langle \phi(\mathbf{t}_i), \phi(\mathbf{t}_j) \rangle\right)\right] \tag{23}$$

The dot product of the high-dimensional feature space vectors can be expressed by the kernel function as

$$K(\mathbf{t}_i, \mathbf{t}_j) = \langle \phi(\mathbf{t}_i), \phi(\mathbf{t}_j) \rangle \tag{24}$$

According to Cristianini and Shawe-Taylor [23], different kernel functions, indicate different features spaces. The popular kernel functions include polynomial kernel and Gaussian radial basis function (RBF) kernel. This work uses a Gaussian RBF kernel:

$$K(\mathbf{t}_i, \mathbf{t}_j) = \exp\left(-\frac{\left\| \mathbf{t}_i - \mathbf{t}_j \right\|^2}{2c^2} \right) \tag{25}$$

The parameter c is selected based on experiments to have an optimized performance on the training data. By substituting equation (25) into equations (22) and (23), the kernel SVM binary classifier for fault detection can be obtained.

3.3 Fault-Classification Module

When a fault is detected, the fault-classification module is triggered to identify its type. The flow chart of the fault-classification module is as shown in Figure 7.

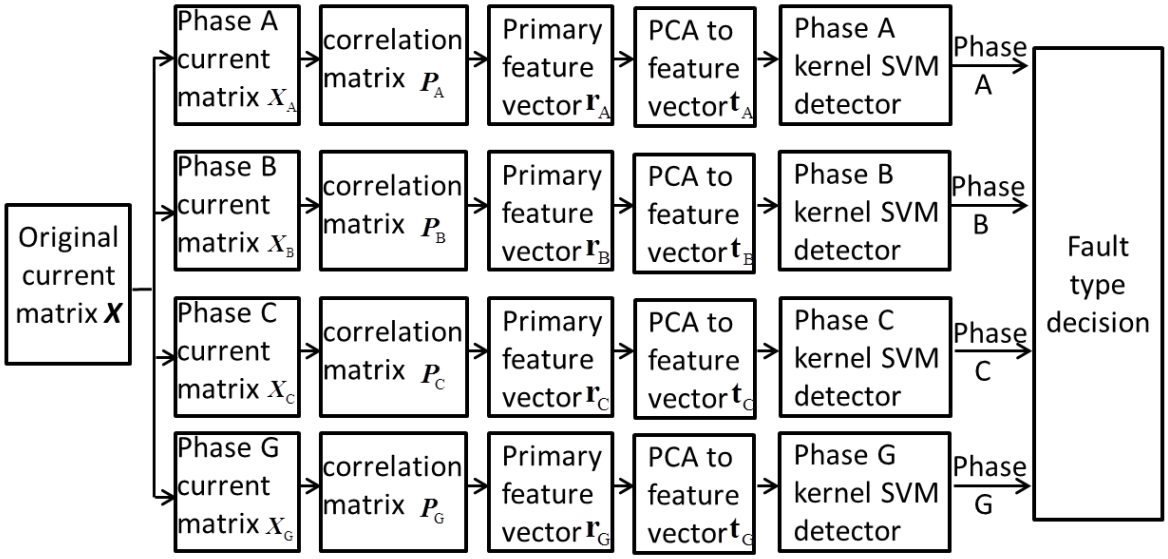

Figure 7: Flow chart of fault-classification module

First, the original data set in equation matrix (2) is divided into $\boldsymbol{X}_A, \boldsymbol{X}_B, \text{and } \boldsymbol{X}_C$. For each phase SVM classifier, the input current matrix $\boldsymbol{X}_k$ (k = A, B, or C) is given by

$$\boldsymbol{X}_k = \begin{bmatrix} x_{1,1k} & x_{1,2k} & \cdots & x_{1,nk} \\ x_{2,1k} & x_{2,2k} & \cdots & x_{2,nk} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m,1k} & x_{m,2k} & \cdots & x_{m,nk} \end{bmatrix}_{m \times n} \tag{26}$$

The phase-neutral current matrix is calculated using $\boldsymbol{X}_A, \boldsymbol{X}_B, \text{and } \boldsymbol{X}_C$ [16] as

$$\boldsymbol{X}_G = \boldsymbol{X}_A + \boldsymbol{X}_B + \boldsymbol{X}_C \tag{27}$$

This classification procedure is the same as with the detection module. Finally, the faulty phases can be identified and the fault type can be classified.

CHAPTER 4

EXPERIMENTAL EVALUATION

The proposed method was tested by IEEE 34-node and 13-node test feeders. First, these two models were built using the PSCAD/EMTDC simulation program. Placements of the three-phase current meters are shown in Figures 8 and 9. Eight thousand faulty cases, indicating all types of faults (line-ground, line-line, line-line-ground, and three-phase faults) at different locations were generated by changing the fault impedance. The circles shown in Figures 8 and 9 represent locations where faults are applied. Triangles represent sensor locations. Crosses are load-switching locations. Of the collected data sets, 50% are used for training the algorithm, 20% are used for validation, and the remaining 30% are used for testing. The Gaussian mixture noise model with different SNRs is applied and simulated in the test feeders.

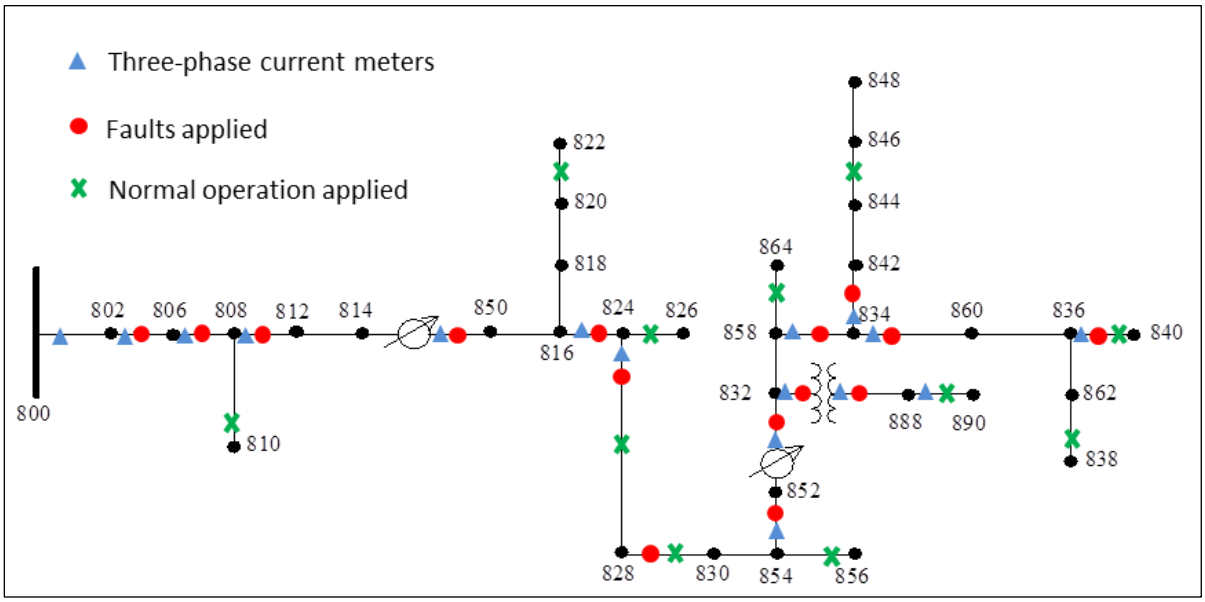

Figure 8: IEEE 34-node test feeder

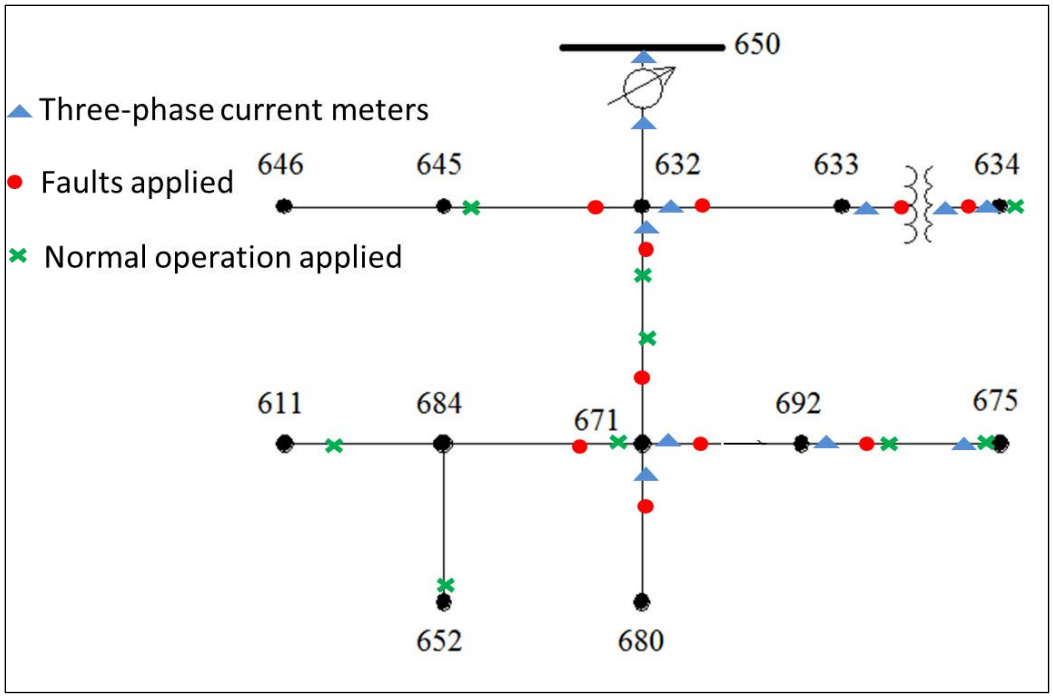

Figure 9: IEEE 13-node test feeder

A high-impedance fault is more difficult to detect. In Figure 10, an 80Ω ground fault is applied at time 0.1 sec.

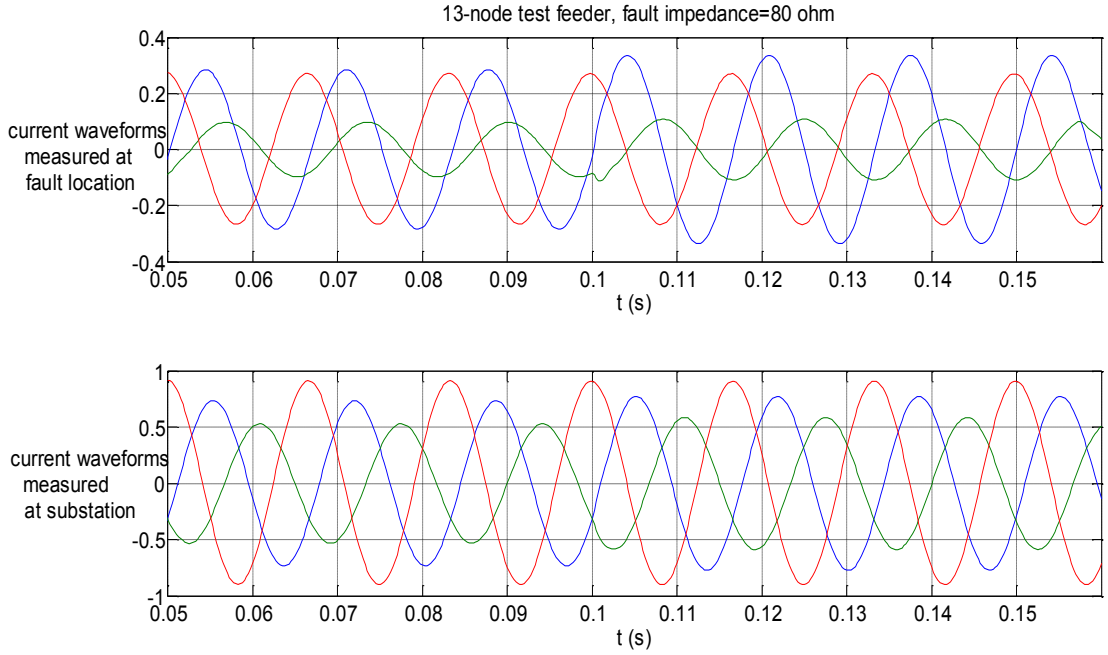

Figure 10: IEEE 13-node system 80Ω fault start at 0.1 sec

The current waveforms at the fault location showed a slight disturbance, but the waveforms at the substation could not detect the abnormality. The fault impedance of the 34-node system was simulated from $0.1\,\Omega$ to $500\,\Omega$, while the 13-node system fault impedance was simulated from $0.1\,\Omega$ to $50\,\Omega$. The fault durations of the two systems were kept the same, from 0.015s to 0.075s. Along with fault data, 8,000 normal cases were generated and consisted of the switching of adding and cutting load, which is 20% of the original load. The sampling rate was kept at 4 kHz.

4.1 Selection of Parameters

In equation (10), parameter p, indicating the proper number of the principal components for PCA projection, must be decided. In equation (25), parameter c is also required to be selected for the SVM feature space projection. Studies on these parameters are limited in the literature but indicate that they are selected based on experiments. In this work, from the simulation, a relationship between these parameters and the fault detection performance was found.

Figure 11 shows the tendency of the fault-detection performance to vary from parameter p. As can be seen, with the increase of p, both the fault-detection rate and the accuracy increase, and the false alarm rate decreases. This means that increasing the number of principal components of PCA will improve the fault-detection performance. However, when p reaches 30 and above, the performance does not change much.

Figure 12 shows the trend of the detection performance in terms of parameter c. The fault detection rate increases with the increase of c and then tends to be stable. However, from $c > 0.4$, the false alarm first decreases and then increases as c grows. Therefore, a proper value of c could be obtained from the minimum value of the false alarm rate.

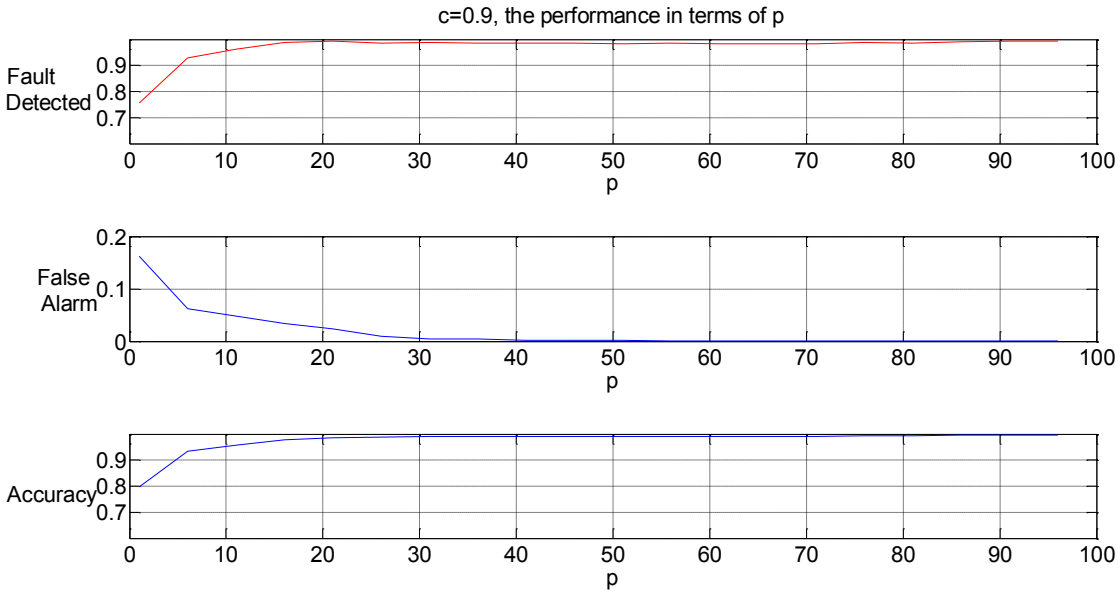

Figure 11: Fault detection performance in terms of p

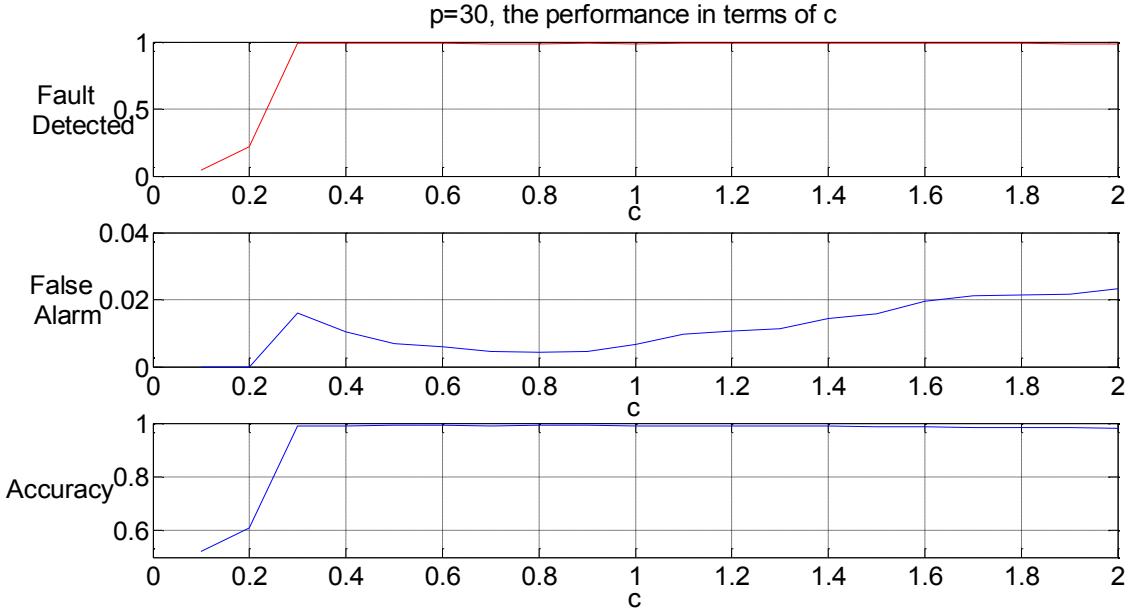

Figure 12: Fault detection performance in terms of c

4.2 Fault Detection Results

Table 1 shows the detection performance of the IEEE 34-node test feeder under different SNR scenarios. The proposed method has a high-detection accuracy for both systems in high SNR situations. When the SNR decreases, the performance degrades, but it still remains 90%

above until the SNR decreases to 10 dB. Under low-SNR scenarios, the fault-detection rate does not decrease much. However, the false alarm increases dramatically.

Table 1: IEEE 34-Node Test Feeder Detection Performance

	34-Node Feeder Detection Performance				
SNR (dB)	**Accuracy**	**Faults Detected**	**Faults Missed**	**No Fault Correct Decision**	**False Alarms**
50	0.9925	0.9895	0.0105	0.9955	0.0045
30	0.9920	0.9882	0.0118	0.9958	0.0042
20	0.9774	0.9805	0.0195	0.9742	0.0258
15	0.9485	0.9755	0.0245	0.9215	0.0785
10	0.8452	0.9463	0.0537	0.7442	0.2558

Table 2 shows the detection performance of the IEEE 13-node test feeder under the same SNR scenarios as the 34-bus case. Comparing these two tables, the performance of the 13-node system is slightly worse than that of the 34-node system because the 13-node system is more unbalanced. As shown in previously in Figure 9, buses 645, 646, 671,684, 611, and 652 are either single-phase or two-phase lines, which have more effect on the fault-detecting decision.

Table 2: IEEE 13-Node Test Feeder Detection Performance

	34-Node Feeder Detection Performance				
SNR (dB)	**Accuracy**	**Faults Detected**	**Faults Missed**	**No Fault Correct Decision**	**False Alarms**
50	0.9712	1	0	0.9425	0.0575
30	0.9654	1	0	0.9308	0.0692
20	0.9554	0.9933	0.0067	0.9175	0.0825
15	0.9458	0.9825	0.0175	0.9092	0.0908
10	0.8867	0.9450	0.055	0.8283	0.1717

4.3 Fault-Classification Results

Table 3 illustrates the fault-classification rate of the 34-node system under SNRs from 50 dB to 15 dB. The proposed method has a high classification rate for all types of faults under high SNR scenarios. When the SNR decreases, the performance also degrades. Compared with the detection algorithm, the classification algorithm tends to be affected by SNRs. Since all results

shown in Table 3 are based on parameters obtained using the no-noise data set, another parameter set was used based on the 15 dB SNR training data set.

Table 3: IEEE 34-Node Test Feeder Classification Performance

Fault	Fault-Classification Accuracy under Different SNRs			
	50 dB	30 dB	20 dB	15 dB
AG	0.9975	1.0000	0.9062	0.8864
AB	0.9852	0.9358	0.8889	0.8222
AC	0.9704	0.9407	0.8173	0.7753
ABG	0.9901	0.9951	0.9062	0.8247
ACG	1.0000	0.9852	0.8889	0.8420
ABC(G)	1.0000	0.9383	0.8173	0.7580
BG	0.9802	0.9877	0.9284	0.7877
CG	1.0000	0.9975	0.8963	0.8568
BC	0.9877	0.9556	0.7877	0.7852
BCG	0.9975	0.9827	0.8296	0.7852

Table 4 shows that given certain SNR scenarios, even the SNR = 15 dB, the classification rate can be above 90%.

Table 4: Classification Rate Given at 15 dB SNR

SNR (dB)	Fault-Classification Accuracy									
	AG	AB	AC	ABG	ACG	ABC	BG	CG	BC	BCG
15	0.9605	0.9728	0.9358	0.9506	0.9704	0.9260	0.9358	0.9481	0.9753	0.9211

Table 5 shows false classification distribution results. As shown, the falsely classified types of faults are randomly distributed at all possible decisions.

Table 5: IEEE 34-Node Test Feeder False Classification Rate under 15 dB SNR

Real Type	False Classification Rate (%)										
	AG	AB	AC	ABG	ACG	ABC	BG	CG	BC	BCG	No Fault
AG	88.64	0	0	5.68	4.69	0.25	0	0.25	0	0	0.49
AB	0	82.22	0	14.32	0	3.21	0	0	0	0	0.25
AC	0	0	77.53	0	18.02	3.95	0	0	0.25	0	0.25
ABG	14.81	0	0	82.47	0.74	1.73	0.25	0	0	0	0
ACG	11.85	0	0	0.74	84.20	1.73	0	1.48	0	0	0
ABC	0.74	4.44	8.40	2.22	3.21	75.80	0	0	0.25	0	4.94

Table 5 (continued)

Real Type	False Classification Rate (%)										
	AG	AB	AC	ABG	ACG	ABC	BG	CG	BC	BCG	No Fault
BG	0	0	0	2.22	0	0	78.77	0.74	0	7.65	10.62
CG	0	0	0	0	0.74	0	1.48	85.68	0	2.47	9.63
BC	0	0	0	0	0	0.99	0	0	78.52	20.49	0
BCG	0	0	0	0	0	0	6.42	10.62	0	78.52	4.44

Table 6 shows the 13-node test feeder classification rate and the false classification rate under the 15 dB SNR case. The algorithm still reaches an 85% classification rate under high SNR cases. However, due to heavy unbalance of the 13-node system, performance is worse than that of the 34-node system. When SNRs decrease, the accuracy of the phase-phase types of faults are seriously degraded.

Table 6: IEEE 13-Node Test Feeder Classification Performance

Fault	Fault-Classification Accuracy under Different SNRs			
	50 dB	30 dB	20 dB	15 dB
AG	0.8750	0.8750	0.8750	0.7639
AB	0.9167	0.8750	0.7361	0.3194
AC	0.9028	0.8056	0.6944	0.4444
ABG	0.8889	0.8750	0.8472	0.8611
ACG	0.8750	0.8611	0.8611	0.8750
ABC(G)	0.8750	0.8611	0.8611	0.8750
BG	0.8889	0.8750	0.8889	0.9167
CG	0.8750	0.8750	0.8750	0.8194
BC	0.8750	0.8750	0.7778	0.5278
BCG	0.8750	0.8750	0.8750	0.8750

Table 7 shows more details. The AB, AC, and BC faults have a large proportion of false classifications as ABG, ACG, and BCG faults, respectively. This is also due to the heavy unbalance of the system. However, this method still provides knowledge of the healthy phase(s).

Table 7: IEEE 13-Node Test Feeder False Classification Rate under 15 dB SNR

Real Type	False Classification Rate (%)										
	AG	AB	AC	ABG	ACG	ABC	BG	CG	BC	BCG	No Fault
AG	76.39	0	0	8.33	1.39	0	1.39	0	0	0	12.50
AB	2.78	31.94	0	54.17	0	0	0	0	0	0	11.11
AC	0	0	44.44	0.00	34.72	5.56	0	1.39	0	0	13.89
ABG	0	0	0	86.11	0	0	4.17	0	0	0	9.72
ACG	0	0	0	0	87.50	0	0	0	0	0	12.50
ABC	0	0	0	0	0	87.50	2.78	0	0	0	9.72
BG	0	0	0	0	0	0	91.67	0	0	0	8.33
CG	0	0	0	0	0	0	2.78	81.94	0	5.56	9.72
BC	0	0	0	0	0	0	1.39	0	52.78	31.94	13.89
BCG	0	0	0	0	0	0	6.94	0	0.00	87.50	5.56

4.4 Sensor Number and Sampling Rate Discussion

Since the multi-sensor-based methods are very new, there is neither the optimization theory about the number of sensors and sensor locations nor the lowest sampling rate. Some simulation results are presented about the performance related to the number of sensors and the sampling rate. The sensors are basically eliminated from left to right in Figure 8. However, the sensor at substation is always kept, and the sensors at the secondary side of the transformer are eliminated. Figure 13 is the plot of the accuracy varying with the number of sensors applied in the 34-node test feeder. As shown, if the sensors at the secondary side of the transformer are eliminated first, then the accuracy at 30 dB and 15 dB SNR is diminished below 0.96 and 0.9, respectively. Furthermore, when the number of sensors is above five, the accuracy does not change much.

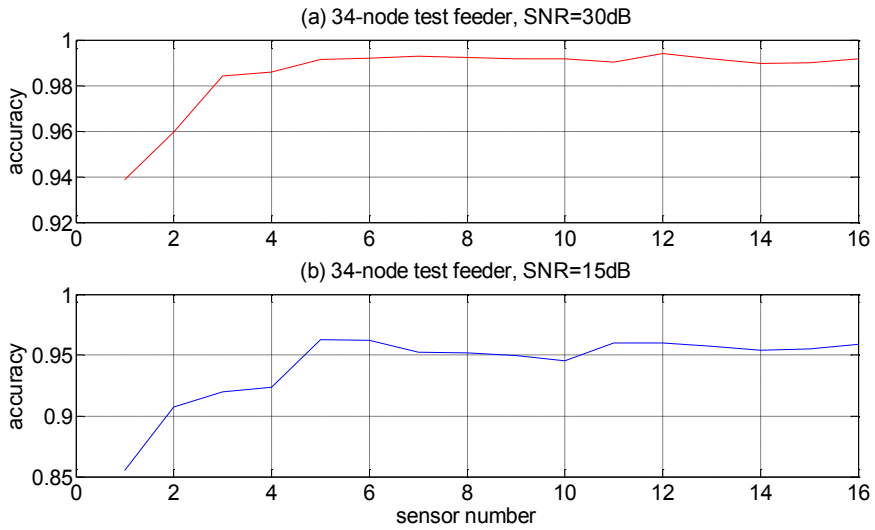

Figure 13: Fault detection performance in terms of sensor number

Figure 14 shows the 34-node system detection accuracy varying with the sensor sampling rate. Different SNRs and numbers of sensors are also considered. As shown, for high SNRs, even though the sampling rate is diminished to 500 Hz, the algorithm accuracy still remains high. However, for low SNRs, there is a remarkable decrease in accuracy as the sampling rate decreases. Comparing the four curves of the 30 dB SNR scenarios, it is obvious that all the three multiple-sensor curves have better performance than the one-sensor case. Meanwhile, the performance has no prominent difference among the 3-, 10-, and 16-sensor cases.

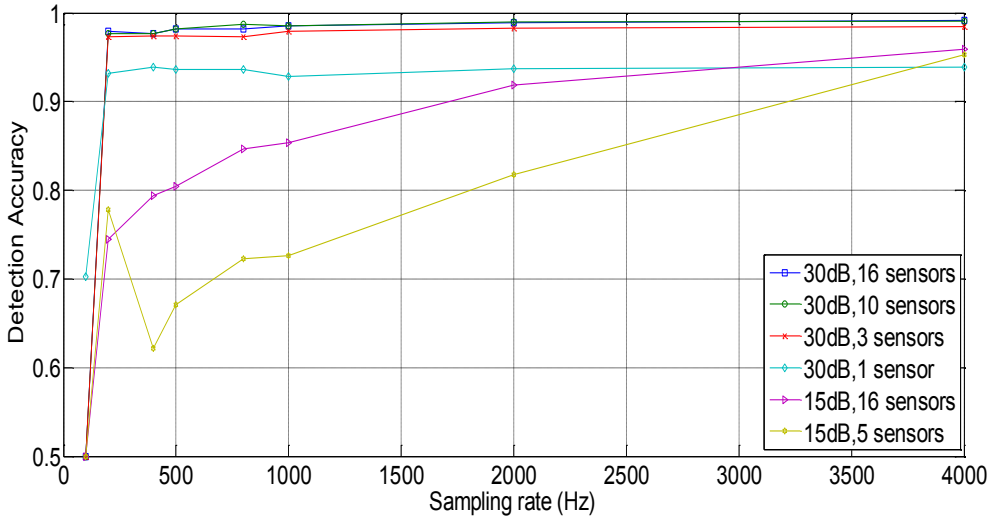

Figure 14: Fault detection performance in terms of sampling rate

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

In this work, a novel multi-sensor-based fault detection and classification scheme was proposed to identify all types of faults in distribution systems. All sensors measured three-phase current magnitudes and sent them to a fusion center. The signal processing techniques, such as correlation matrix, PCA, and kernel SVM were used to realize the fault detection and classification. The proposed method was tested under IEEE 34-node and 13-node test feeders, and the Gaussian mixture noise model was applied in the simulations to test the robustness of the algorithm. Different fault impedances, fault durations, fault types, fault locations, SNR scenarios, and load switching were simulated. The simulation results verified the universality and robustness of the method. In addition, the number of sensors, limit of the sampling rate, and parameter settings were discussed in this work.

5.2 Future Work

From the simulation results, the proposed method has some limitations when the system is unbalanced. The ground-phase classifier has a high probability to make wrong decisions under low-SNR situations. Therefore, the algorithm for the ground-phase classifier needs to be modified to improve performance for the unbalanced systems. One intuitive way to resolve this is to add some features into the ground-phase classifier.

In this work, the effects of the number of sensors and sampling frequency on the detection performance were discussed. However, the methodologies to determine the optimal sensor locations and number of sensors were not developed.

Future work is summarized below:

- "Sensor-weights" could be developed for sensors at different locations to illustrate the importance of the sensor to the system.
- Adaptively updating the PCA and SVM training parts could also be added to the algorithm to improve the performance when the noise or SNR changes quickly.

REFERENCES

REFERENCES

[1] V. Giordano, F. Gangale, G. Fulli, and M. Sánchez-Jiménez, "Smart Grid Projects in Europe: Lessons Learned and Current Developments," JRC Reference Report, ISBN 978-92-79-20487-6, 2011.

[2] "What is the Smart Grid: Distribution Intelligence," URL: http://www.smartgrid.gov/the_smart_grid [cited June 14, 2014].

[3] L. Xu and M. Chow "A Classification Approach for Power Distribution Systems Fault Cause Identification," *IEEE Trans. Power Systems*, Vol. 21, Iss. 1, pp. 53–60, 2006.

[4] N. Wang, V. Aravinthan, and Y. Ding, "Feeder-Level Fault Detection and Classification with Multiple Sensors: A Smart Grid Scenario," in *Proc. IEEE Workshop Statistical Signal Processing* , 2014 .

[5] J. D. Glover, M. Sarma, and T. Overbye, *Power System Analysis & Design*, SI Version, Cengage Learning, 2011, pp. 379–472.

[6] L. Xu and M. Chow. "Power Distribution Systems Fault Cause Identification using Logistic Regression and Artificial Neural Metwork," in *Proc. 13th International Conference on Intelligent Systems Application to Power Systems*, 2005.

[7] K. L. Butler, and J. A. Momoh "A Neural Net Based Approach for Fault Diagnosis in Distribution Networks." in *Proc. Power Engineering Society Winter Meeting*, Vol. 2., 2000.

[8] A. R. Oliveira, P. A. N. Garcia, L. W. Oliveira, J. L. R. Pereira, and H. A. Silva "Distribution System to Fault Classification Using Negative Sequence and Intelligent System," in *Proc.* 16th *International Conference on Intelligent System Application to* Power Systems, 2011.

[9] A. R.Oliveira, P. A. N.Garcia, , L. W. Oliveira, H. A. Silva, J. L. R. Pereira, and M. Tomim, "New Algorithms for Fault Classification in Electrical Distribution Systems," in *Proc. IEEE Power and Energy Society General Meeting*, 2010.

[10] B. Das, "Fuzzy Logic-Based Fault-Type Identification in Unbalanced Radial Power Distribution System," *IEEE Trans. Power Delivery*, Vol. 21, Iss. 1, pp. 278–285, 2006.

[11] L. Xu, M. Chow, and L. S. Taylor, "Power Distribution Fault Cause Identification with Imbalanced Data Using the Data Mining-Based Fuzzy Classification E-Algorithm," *IEEE Trans. Power Systems*, Vol. 22, No. 1, pp. 164–171 2007.

[12] O. Dag and C. Ucak, "Fault Classification for Power Distribution Systems Via a Combined Wavelet-Neural Approach," in *Proc. IEEE International Conference on Power System Technology*, Vol. 2, 2004.

[13] U. D. Dwivedi, S. N. Singh, and S. C. Srivastava "A Wavelet Based Approach for Classification and Location of Faults in Distribution Systems," in *Proc. Annual IEEE India Conference (INDICON)*, Vol. 2., 2008.

[14] Y. Assef, O. Chaari, and M. Meunier "Classification of Power Distribution System Fault Currents Using Wavelets Associated to Artificial Neural Networks," in *Proc. IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, 1996.

[15] S. R. Samantaray, B. K. Panigrahi, and P. K. Dash. "High Impedance Fault Detection in Power Distribution Networks using Time-Frequency Transform and Probabilistic Neural Network," *IET Journal Generation*, *Transmission & Distribution*, Vol. 2, Iss. 2 pp. 261–270, 2008.

[16] H. Livani, and C. Y. Evrenosoglu, "A Fault Classification Method in Power Systems Using DWT and SVM Classifier," in *Proc. IEEE PES Transmission and Distribution Conference and Exposition (T&D)*, 2012.

[17] A. Dubey, R. K. Mallik, and R. Schober, "Performance Analysis of a Power Line Communication System Employing Selection Combining in Correlated Log-Normal Channels and Impulsive Noise," *IEEE Journal on Communications*, Vol. 8, Iss. 7, 2014.

[18] H. Meng, Y. L. Guan, and S. Chen, "Modeling and Analysis of Noise Effects on Broadband Power-Line Communications," *IEEE Trans. Power Delivery*, Vol. 20 Iss. 2, pp. 630–637, 2005.

[19] D. W. J. Stein, "Detection of Random Signals in Gaussian Mixture Noise," *IEEE Trans. Information Theory*, Vol. 41, No. 6, pp. 1788–1801, 1995.

[20] M. B. Wise and N. B. Gallagher, "The Process Chemometrics Approach to Process Monitoring and Fault Detection," *Journal of Process Control*, Vol. 6, Iss. 6, pp. 329–348, 1996.

[21] J. Lee, C. Yoo, S. W. Choi, P. A. Vanrolleghem, and I. Lee, "Nonlinear Process Monitoring using Kernel Principal Component Analysis," *Chemical Engineering Science*, Vol. 59, Iss. 1, pp. 223–234, 2004.

[22] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, Vol. 20, Iss. 3, pp. 273–297, 1995.

[23] N. Cristianini, and J. Shawe-Taylor, *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press, 2000.

[24] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, 2012.

[25] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall Publication, 2010.