Electric Power Systems Research 148 (2017) 254–263

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

ICA feature extraction for the location and classification of faults in high-voltage transmission lines

A.R. Almeida [a,*], O.M. Almeida [a], B.F.S. Junior [a], L.H.S.C. Barreto [b], A.K. Barros [c]

[a] Federal University of Piauí – UFPI, Department of Electrical Engineering, Campus Ininga, Terezina, Brazil
[b] Federal University of Ceará – UFC, Department of Electrical Engineering, Campus Pici, Fortaleza, Brazil
[c] Federal University of Maranhão – UFMA, Department of Electrical Engineering, Campus Bacanga, São Luiz, Brazil

ARTICLE INFO

Article history:
Received 30 September 2016
Received in revised form 24 February 2017
Accepted 25 March 2017
Available online 18 April 2017

Keywords:
Independent component analysis
Fault location
Fault classification
Support vector machine
Multidimensional scaling

ABSTRACT

Several methods for the location and classification of faults in power transmission lines using computational intelligence and digital signal processing techniques have been described in literature. Artificial neural networks (ANNs) and wavelet transform (WT) have drawn significant attention lately, but they present some drawbacks when dealing with power systems faults where data are often contaminated by noise. This paper proposes an approach by combining independent component analysis (ICA) with travelling wave (TW) theory and support vector machine (SVM). The approach is adequate to locate and recognize faults in high-voltage (HV) transmission lines, while the acquired signals are noisy. Experiments performed for distinct types and locations of faults in a real transmission line model have shown that the proposed combined methods are able to provide excellent performance in fault location. The obtained errors are lower than 1% and accuracy is 100% for the classification of fault signals with noise. It can be stated that this method presents better performance than those regarding the main conventional techniques such as wavelets and neural networks in the presence of noise.

Contents

1. Introduction

The rapid expansion of electric power systems over the last decades has also led to the increased need for additional transmission lines. Such lines are exposed to distinct faults as a result of lightning, eventual short circuits, faulty equipment, human errors, overload, and aging. Fast detection, isolation, location and restoration of systems under fault are critical issues while maintaining reliable operation.

When a fault occurs in a given transmission line, the voltage across the point of incidence is suddenly reduced to a low level. As a consequence, a high-frequency electromagnetic impulse called TW is generated [1]. The location and classification of faults with adequate accuracy must be performed as fast as possible to decrease the occurrence time interval and protect the system against harmful and undesirable resulting effects. Significant effort has been dedicated to the development of methods based on the direct use

* Corresponding author.
E-mail address: aryfrance@ufpi.edu.br (A.R. Almeida).

http://dx.doi.org/10.1016/j.epsr.2017.03.030
0378-7796/

of the TW theory or signals involved with TWs. Interesting methods for the location and classification of faults in transmission lines based on the TW theory have been proposed so far [2–5]. For instance, methods using wavelet transforms (WT) combined with SVM have been adopted to decompose the current and voltage signals associated to the fault in various input levels to classify the faults properly [6]. The errors when estimating the fault location in this case vary from 1% to 6%. Other methods based on wavelet packet transform (WPT) combined with ANN to extract fault characteristics using the energy and entropy criterion have also been applied [7]. Such strategy typically presents errors lower than 1%. In order to obtain improved results, wavelet entropy has been adopted to reduce the vector size regarding the detail coefficient as a preparation stage for the fault classification [8]. Generalized neural network (GNN), WT and support vector regression (SVR) have also been considered in this case providing errors lower than 2% [9,10]. It is worth to mention that errors rated at about 1% are acceptable since the practical effects are measured when dealing with distances measured in kilometers.

Although many works on the location and classification of faults in transmission lines based on WT were introduced in the last decade [11], only a few ones effectively present the detailed analysis of noise contaminated data. It is known that the parameters of the WT are very sensitive to noise [12], which is generated due to many transmission line related phenomena i.e. electromagnetic fields, leakage currents in isolators, and environmental conditions.

In order to overcome noise issues, ICA can be used. It is well known that this is a powerful computational method for separating a multivariate signal in to additive subcomponents so that a given desired signal can be extracted. ICA is a particular case of blind source separation and differs from other methods since the subcomponents of an original signal are separated in statistically-independent signals in the form of a Gaussian distribution. This technique has been widely applied to separate Gaussian signals from nonGaussian ones in many fields e.g. biomedical signal processing, radar and mobile communication, and feature extraction (image, audio, video representation, among others) [13].

The main contributions of the paper can be stated as follows:

1. Introduction of an approach that combines ICA, TW theory, and SVM to locate and recognize faults in HV transmission lines when the existing signals are significantly noisy.
2. Presentation of original concepts regarding the application of ICA and SVM for the separation of noise and data in fault signals. On the other hand, ICA is combined with SVM as a classification stage, while ICA associated with TW theory is applied to the fault location. Considering the aforementioned approaches, very noisy data can be properly treated.
3. Application of Multidimensional scaling (MDS) theory to the graphical representation and demonstration of SVM effectiveness regarding the classification of faults with and without the presence of noise for distinct types of events.

This paper is organized as follows: Section 2 presents the TW theory for the location of faults. Section 3 describes the real transmission system model implemented in software ATP. Section 4 is concerned with ICA applied to feature extraction using the basis function for the location and classification of faults. Section 5 presents the combination of library support vector machine (LibSVM) and MDS as applied to the classification of faults (single line to ground fault – SLG; double line to ground fault – LLG; double line fault – LL; and three-phase fault ––LLL). Section 6 discusses the results obtained with the proposed method, as relevant conclusions are also given.

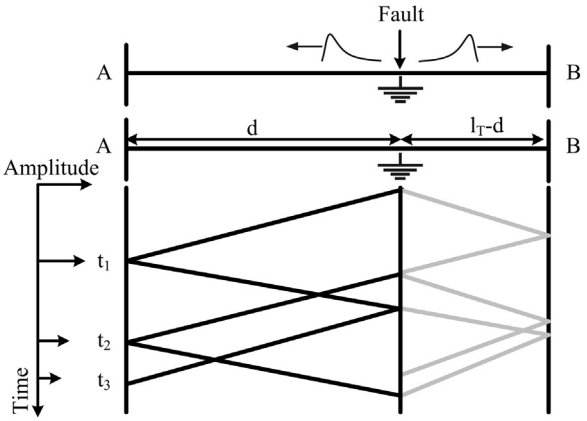

Fig. 1. Lattice diagram.

2. TW theory

A TW in a transmission line is a transient disturbance that moves along the line with constant speed and shape. When a fault occurs, a voltage TW is generated and propagates in both directions from the fault incidence point. When the TW comes to an eventual discontinuity in the line, it is reflected back to the fault origin, while part of the signal is refracted. In order to track all the components that exist in the former signal, the Lattice diagram has been used as shown in Fig. 1.

The path of each wave component can easily tracked along its zig–zag trajectory, giving rise to reflections and refractions at each junction. Thus the TW characteristics can be defined at any point and any time instant [14]. If the fault occurs in the first half of the line, the distance from the fault point to the measurement one is given by (1).

$$d = \frac{v(t_2 - t_1)}{2} \tag{1}$$

where l_t is the transmission line total length, while t_1 and t_2 represent the travelling time intervals from the fault location to the measurement point and v is the wave propagation speed in the air. The TW speed in this propagation mode is very close to the light speed in the air i.e. 3×10^8 m/s, which is considered for calculation purposes. The propagation modes can be separated by using Clarke or Whedepohl transform [15]. If the fault occurs in the second half of the line, the distance from the fault point to the measurement one is given by (2).

$$d = l_t - \frac{v(t_2 - t_1)}{2} \tag{2}$$

3. Transmission system model

The voltage data have been acquired from an oscilographic recorder installed at a transmission line. The line is part of the Brazilian power system managed by São Francisco's Hydroelectric Company (CHESF) and this scenario has been used to test and evaluate the proposed strategy. Besides, it is responsible for connecting Presidente Dutra (PDT) and Boa Esperança (BEA) substations (SE). The total length of the transmission line is 200 km, whose topology consists in a two terminal arrangement i.e. measurement terminal (PDT) and remote terminal (BEA) as discussed in Section 2. Fig. 2 represents the schematic diagram of the model adopted in the analysis of faults in the first half (F1) and second half of the line (F2) using ATP [16]. The positive sequence reactances for the generators (G1 and G2) that exist in SE-PDT and SE-BEA are 5.65 mH/m. On the other hand, the positive sequence parameters of the line are

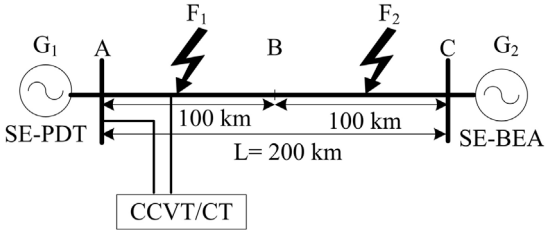

Fig. 2. 500-kV, 200-km transmission system model.

Table 1
Data for 500 kV, 60 Hz, 200 km three phase line.

Phase arrangement	Horizontal tower configuration
Phase condutors	
Height at tower	24 m
Height at midspan	12 m
Phase spacing	12 m
Number of bundle conductors	2
Radius of subcondutors	1.257 cm
Spacing between subcondutors	40 cm
DC resistance	0.08998 Ω/km
Ground wires	
Height at tower	33 m
Height at midspan	26 m
Spacing	16 m
Radius	0.4572 cm
DC resistance	4.188 Ω/km

Table 2
Case studies defined for the 500 kV, 200 km transmission line.

Fault type	SLG, LLG, LL and LLL
Real distance (km)	10, 16, 45, 84, 90, 128, 147, 155
Resistance of fault (Ω)	20, 50, 80, 120, 150, 180, 200, 240
Angle of incidence (θ)	0°, 45°, 90°, 135°, 180°, 225°, 270°, 315°
SNR (dB)	30, 35, 40, 45, 50, 55, 60, 65, 70, 75

L_1 = 1.075 mH/km and C_1 = 10.805 nF/km, while others parameters are given in Table 1.

The 500 kV, 60-Hz transmission line model was derived using Marti [17] and the LCC (line and cable constants) block in ATP software. A set compromising four types of faults (SLG, LLG, LL, and LLL) under 7680 distinct situations was used to test the performance of proposed approach. The employed voltage data were sampled at a frequency of 200 kHz. Distinct values for the distance from the incidence point and PDT terminal were used, as well as different angles of incidence and resistances for the fault as shown in Table 2. Noise was added to the source side signal in order to verify the immunity of the technique to deal with noise contaminated data. Since current and potential transformers (CCVT/CT) affect signal data, noise was added to the source-side signal after the measurement. The signal-to-noise ratio (SNR) of the signal is considered to be between 30 dB and 75 dB given by.

$$\mathrm{SNR\,(dB)} = 10\log\frac{(P_s)}{(P_n)} \tag{3}$$

where P_s is the signal power variance and P_n is noise power. In order to obtain the fault database, the conditions described in Table 2 were used.

3.1. Transients in power systems

Transmission lines are considered the most important components in power systems, since they connect generating stations to distribution systems through a huge network. During the fault, the relay related equipment enables the associated circuit breakers to isolate the sector under fault, as the adoption of accurate schemes

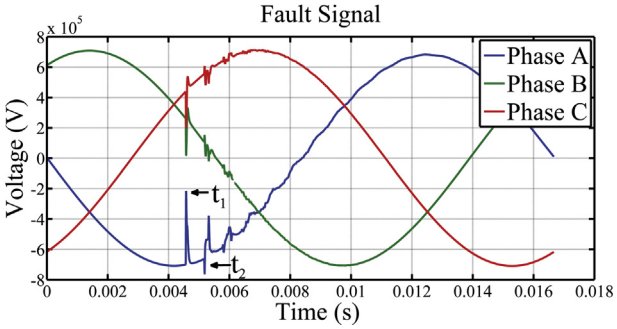

Fig. 3. Fault signal.

for the location of faults is quite important. Such strategies are based on checking some line parameters that change during the fault. Three quantities can be distinguished in this case: angle of incidence of the fault, the fault resistance, and the travelling wave shape [18].

Fig. 3 shows the voltage signal for a fault that occurred 90 km far from PDT terminal. This particular case reports a SLG fault with an angle of incidence of 90° and fault resistance of 120 Ω. Data representation of the signal window corresponding to one cycle (16.67 ms) is shown in detail, where the peak time instants are evidenced. By applying Eq. (1), it is possible to determine the peaks in the voltage TW corresponding to t_1 and t_2 to estimate the fault distance to the measurement terminal.

$$d = \frac{3\times10^8\times\left(5.185\times10^{-3} - 4.4575\times10^{-3}\right)}{2}$$
$$d = 91.5\mathrm{km}$$

This simple example shows that the TW theory can be used to determine the fault location with relative accuracy and absence of noise. However, the problem generally becomes more complex as the TW data is corrupted by uncorrelated noise. In order to overcome this issue, the combination of some techniques using ICA, SVM, and MDS are proposed in this work.

4. Independent component analysis

Let us consider a given mixed signal x_i with N observations, which are modeled as linear combinations of s_{ji} whose basis functions are a_m.

$$x_i = a_1 s_{1i} + a_2 s_{2i} + a_3 s_{3i+} a_m s_{ji} \tag{4}$$

where j = 1. . .*m* and s_{ji} is a random vector known as source vector. Vector s_{ji} is applied to a mixer, whose input–output characterization is defined by a nonsingular matrix A called mixture matrix. The linear system comprising the source vector S and the mixture matrix A is completely unknown to the observer. By using such matrix definition notation, Eq. (4) can be rewritten as follows:

$$X = AS \tag{5}$$

This model describes how data are generated from the mixing process with the independent components. The simplified model aims at estimating A, as well as the array of independent components S by simply observing X. The ICA approach is based on the blind source separation problem [19]. The block diagram in Fig. 4 describes ICA computation, where an unknown environment contains a set of source signals S as subjected to the action of a mixture system A. From A, only the observation vector X is provided and N corresponds to the noise added to the process, while the main goal is to estimate the demixture matrix W. In this case, the system output Y should be able to provide the signal sources with minimal

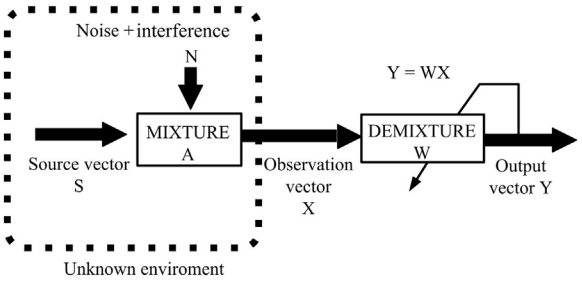

Fig. 4. Block diagram representation of the processor to solve the blind source separation problem.

errors. The problem is said to be blind due to the lack of information on the mixture and the sources.

The estimation of independent components is based on some assumptions:

1. The components are inherently statistically independent;
2. The components present a nonGaussian distribution;
3. The signals regarding additive noise N are statistically independent of each other and independent of sources S;
4. The number of sensors is greater than or equal to the number of sources.

On the other hand, the features regarding the ICA model present some ambiguity regarding the independent components i.e.

1. Their respective variances cannot be determined;
2. The orders cannot be determined.

Such ambiguities exist since both *A* and *S* are unknown. As a result, it is not possible to determine the energy associated to the signals i.e. their amplitudes or even estimate the order for *S*. In order to use ICA methods accurately, Eq. (5) must be solved. They are based on the statistical distribution model for the standard set, being used to determine the existing characteristics i.e. the ICA basis function.

The FastICA algorithm has been chosen in this work since it is one of the most robust, fastest, and easiest approaches for practical implementation. In this case, the demixture matrix W can be determined by maximizing the negentropy of signals s_i so that $W \approx A^{-1}$. The aforementioned code was developed using Matlab software according to the following assumptions:

1: Center data *x* so that their respective mean is zero;

2: Whiten data *x* so that *z* is obtained through z = Vx, where V is the whitening matrix;

3: Choose an initial vector *W* of unit norm randomly;

4: Let $W \longleftarrow E\left\{zg\left(W^T z\right)\right\} - E\left\{g'\left(W^T z\right)\right\} W$, where *g* is defined as $g_1(y) = \tanh(a1y)$, $g2(y) = y\exp(y^2/2)$, $g3(y) = y3$;

5: Let: $W \longleftarrow W/\|W\|$;

6: If convergence is not achieved, return to step 4.

The negentropy defined as *J*(*y*) is a differential entropy measure that gives the amount of information for random variables. It can be used to separate Gaussian signals from nonGaussianones obtained in step 4 of FastICA algorithm, being expressed as follows:

$$J(y) \approx [E\{G(y)\} - E\{G(v)\}]^2 \quad (6)$$

where *v* is a zero mean Gaussian variable with unit variance and *G* is some nonquadratic function, generally used for super-Gaussian distribution g(y) = tanh(y) and for sub-Gaussian distribution g(y) = y − tanh(y). It is worth to mention that the FastICA algorithm must be run repeatedly (i.e. up to the dimensionality of the collected data) to estimate the desired number of independent components.

4.1. ICA applied to feature extraction in a transmission line

Feature extraction using high-order statistics has been influenced by neural information processing model. Neuroscience studies suggest that the neural processing stimulates information according to the concept of efficient codding [20]. By applying the aforementioned theory to data regarding faults in transmission lines, the demixture matrix W can be treated as a set of linear filters in the form:

$$S = WX \quad (7)$$

If the demixture matrix W is assumed to be invertible, ICA estimation can be obtained by $X = W^{-1}S$, which is equivalent to X = AS. In order to finish the execution of the FastICA algorithm, the system output must be determined as Y = WX.

Fig. 5 shows the graphical application of FastICA to achieve feature extraction associated to faults. The procedure can be summarized as follows:

Vector X is initially used as an input of FastICA algorithm. It corresponds to a SLG fault in phase A, with distance of 84 km, angle of incidence of 135°, and fault resistance of 200 Ω. A noise whose SNR is 30 dB is then added according to the conditions listed in Table 1. According to Fig. 5(a), the propagation times are $t_1 = 0.00655$ s and $t_2 = 0.00683$ s. By using Eq. (1), the fault location could be estimated at 42 km, as poor accuracy results if compared to the real value.

The application of FastICA algorithm provides two basis functions for the detection and extraction of information regarding the magnitudes of the first and second peak amplitudes At_1 and At_2 at t_1 and t_2, respectively.

According to basis function 1 shown in Fig. 5(b), where $t_1 = 0.00655$ s and $t_2 = 0.007115$ s, Eq. (1) provides the estimated fault location of 84.75 km, as excellent accuracy results in this case.

According to basis function 2 shown in Fig. 5(c), where $t_1 = 0.00655$ s and $t_2 = 0.007115$ s, Eq. (1) provides the estimated fault location of 84 km, which is also close to the real distance.

It can be stated that the ICA basis function is able to provide the magnitude of the transient voltage signal, which is due to the fact that negentropy is used as a criterion for the estimation of non-Gaussiansignals. In this case, they act as filters used for separated signals.

The basis functions are represented by the mixture matrix A determined by the ICA algorithm, composed by five feature attributes that have been applied to the fault location and classification with and without noise. In this work, the ICA basis functions have also been used to reduce redundancy in the fault pattern from 7680 for 250. Therefore it is possible to simplify the design and improve performance regarding the classifier, whose matrix *A* with dimension of 250 × 5 is given by:

$$A = [t_{1n} t_{2n} A_{t1} A_{t2} SNR] \quad (8)$$

where t_{1n} is the time instant for the first peak, t_{2n} is the time instant for the second reverse peak, At_{1n} is the first peak amplitude, At_{2n} is the second reverse peak amplitude, and SNR represents the existing noise.

Fig. 6 shows a typical characteristic pattern vector representing a SLG fault signal. Details are also provided on important issues e.g. amplitude, wave propagation time, and presence of noise in terms of t_1, t_2, A_{t1}, A_{t2}, and SNR, respectively.

Fig. 7 corresponds to the graphical representation of one column of matrix X in an expanded ICA feature space. Therefore each column in matrix A corresponds to a given feature of the fault and each element of *S* is a coefficient that evidences the importance of *X*. This is the main advantage of ICA regarding the achievement of

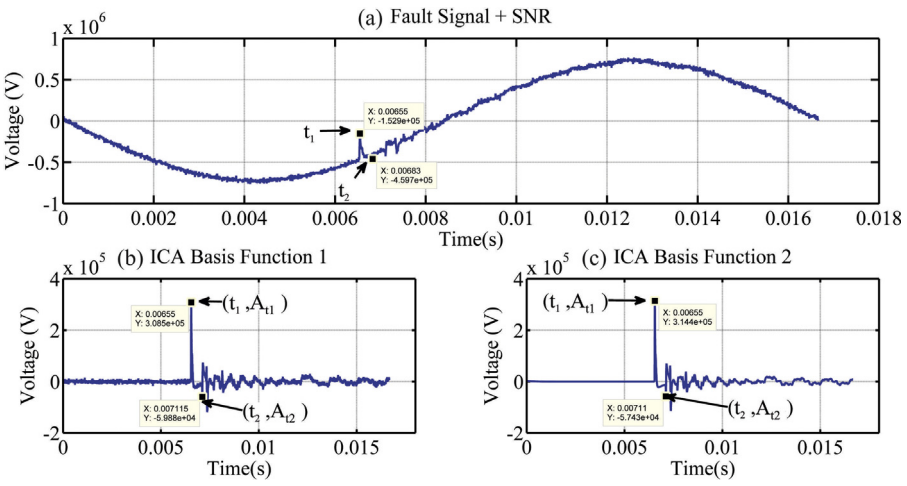

Fig. 5. ICA feature extraction for fault location.

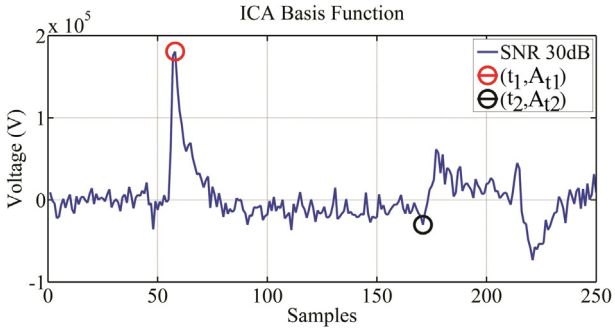

Fig. 6. Typical fault signal in a transmission line where the first and the second peak amplitudes (A_{t1}, A_{t2}) are represented.

efficient coding. Here, let X be a set of observations taken from the same noisy signal database. By using X = AS as a training input, ICA learns the basis functions from the columns of matrix A for feature extraction purposes so that the variables comprising vector S are mutually and statically independent.

The ICA basis function represents the combinations of a set of functions in the frequency domain and time domain describing every source features. From the mixture signal, ICA aims to obtain the independent component or the source signals. Thus a single basis function is able to extract the information required for the source separation and identification of the fault location. In order to determine if the fault occurred in the first half or second half of the line, the time delay method is used applied to the modal components [21].

According to the previous section, ICA is related to the efficient coding that reduces redundancy in the patterns with minimal loss of information. Therefore, it ensures effectiveness associated to the feature extraction stage and provides adequate data representation, thus making classification easier. Before the fault classifier is executed with noise contaminated data, the projection of the basis function in subspaces is carried out, since it is known that the sources of the mixture signal are independent, what implies independent subspace features. An independent subspace is generated from a set of basis functions that describes the sources as returned by the ICA algorithm. In other words, the components extracted by the ICA basis function can be grouped so that each one of them generates different subspaces.

Fig. 8 represents an example where independent subspaces are used as inputs for the classifier through a simple projection of the inner product between the observation vectors X (Xtrain and Xtest) and the basis function.

This provides new values for the training and testing vectors $\tilde{X}$train and $\tilde{X}$test which are given by Eqs. (9) and (10), respectively.

$$\tilde{X}\text{train} = \text{Axtrain} \tag{9}$$

$$\tilde{X}\text{test} = \text{Axtest} \tag{10}$$

The new parameters are:

$$\tilde{X}\text{train} = (x_{t1n}, x_{t2n}, x_{At1}, x_{At2}, x_{SNR}, \ldots x_M),\ x_M \in R^M \tag{11}$$

$$\tilde{X}\text{test} = (x_{t1n}, x_{t2n}, x_{At1}, x_{At2}, x_{SNR}, \ldots x_N),\ x_N \in R^N \tag{12}$$

where M and N are features describing a particular signature for samples that belonging to source 1 (class 1) or source 2 (class 2).

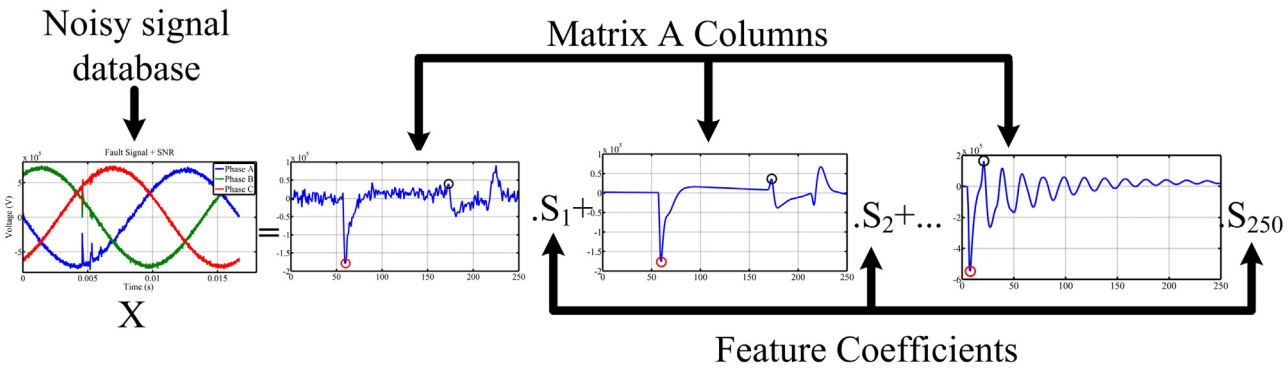

Fig. 7. Representation of faults signal in ICA model, where each column of A is a feature and each s_i is a feature coefficient.

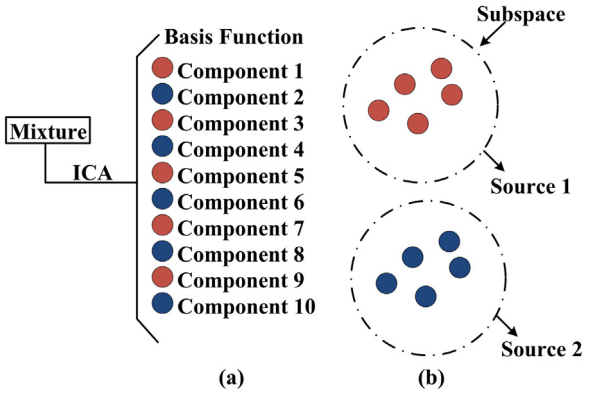

Fig. 8. Conception of independent subspace (a) 10 ICA basis function returned from the mixture signal; (b) two sources subspaces with independent components.

The vector x_{t1n}, x_{t2n}, x_{At1}, x_{At2},x_{SNR} are related to first peak of TW, second peak of TW, amplitude of first peak and second peak, the noise and the number of samples, respectively. Besides, the samples correspond to normalized vectors (Xnorm) comprising [0,1], being this a feature pattern for the fault type adopted in the training and testing of the input regarding the LibSVM classifier [22]. Finally, the output is represented by Yout = (+1, −1), where +1 or −1 indicates if the fault signal contains noise or not, respectively. The normalization process is carried out by using (13).

$$\text{Xnorm} = \frac{X_a - X_{min}}{X_{max} - X_{min}} \tag{13}$$

where Xa is a given variable, while X_{max} and X_{min} are the maximum and minimum values for feature attributes, respectively.

The fault classification task is performed according to the aforementioned values provided to LibSVM algorithm.

5. Support vector machines

SVM is a very useful technique for data classification and regression problems. It was firstly proposed by Vladimir Vapnik and has been successfully used in many areas related to pattern recognition and regression estimation problems regarding dependency estimation, forecasting, and construction of intelligent machines.

The SVM approach is able to handle with very large feature spaces if training is carried out with this purpose. The dimension of classified vectors does not influence the SVM performance as it occurs in conventional classifiers. Besides, SVM based classifiers are claimed to have good generalization properties if compared to their conventional counterparts because the so called structural misclassification risk in the training phase is minimized. On the other hand, traditional approaches are usually trained to minimize eventual empirical risks. Detailed information on SVM can be found in [23].

Nonlinear problems are solved by using SVM in terms of a kernel function. For this purpose, classified data are mapped into a high dimensional feature space where the linear classification is possible. Let the nonlinear transform from the input space x to future space $\Phi(x)$ be given by:

$$\phi(x) = \phi_1(x), \phi_2(x), \ldots \phi_m(x) \tag{14}$$

while the decision surface be obtained as:

$$f(x) = \sum_{j=1}^{m} \alpha_j y_j \phi^T(x_i)\phi(x) \tag{15}$$

Where α_j is the Lagrange multiplier and y_j is the output vector using a given component j. It is worth to mention that inner products are used in (15). A function that returns the dot product of the feature space mapping regarding original data points is called a kernel function $K(x, z) = \phi^T(x)\phi(x)$.

Learning in the feature space does not require the existence of inner products where a kernel function is applied. By using this concept, the decision function can be written as:

$$f(x) = \sum_{j=1}^{m} \alpha_j y_j K(x_i, x) \tag{16}$$

There are different kernel functions proposed in literature.Mercer's theorem states that any symmetric positive definite matrix can be regarded as a kernel matrix. It is important to mention that several kernel function models have been developed in this work.

Common kernel functions include:

1. linear: $K(x_i, x) = x_i^T x$
2. polynomial: $K(x_i, x) = (\gamma x_i^T x + r)^d$, $\gamma > 0$
3. radial basis (RBF) functions: $K(x_i, x) = \exp\left(-\frac{x_i - x^2}{2\sigma^2}\right)$
4. sigmoid: $K(x_i, x) = \tanh\left(\gamma x_i^T x + r\right)$

Where σ is the width of the Gaussian function and γ, r, and d are kernel parameters. The classification accuracy is measured by the classification rate (C.R) of the SVM model, given by:

$$C.R = \frac{\text{accurately predicted data}}{\text{total testing data}} \cdot 100\% \tag{17}$$

Literature shows that many SVM algorithms have been developed e.g. LibSVM, LightSVM, Ls-SVM, among others. LibSVM is used in this work, which has been inspired in the algorithm proposed by Chang and Lin. It is considered one of the most efficient approaches for practical applications due to its excellent properties as described in [24].

LibSVM is employed in this case to train and test each new subspace $\tilde{X}$train and $\tilde{X}$test obtained by the ICA basis function. In order to improve the computation speed of SVM, an efficient strategy for feature extraction in high dimensional spaces must be adopted. An important issue to be considered in this scenario is computational cost, which tends to increase as the number of input variables also does. Many feature extraction methods have been introduced so far to overcome this problem e.g. Principal Component Analysis (PCA), Self Organization Map (SOM) network, among others [25,26].

Since the problem studied in this work is also related to high dimensional spaces, it is useful to adopt a graphical approach to deal with involved data, as MDS is chosen for this purpose since it is adequate for multivariate data [27]. It is widely used in research and application due to its representation capacity, where a set of observation is represented by a set of points in a low dimensional real Euclidean vector space. In order to perform this task, similar observations are represented by points that are close to each other [28].

This work combines LibSVM with MDS using Matlab software for the better graphical representation of training and test data obtained by the ICA basis function. Classification of faults with and without the presence of noise is addressed for several types of faults. A proper flowchart of the proposed method is shown in Fig. 9.

6. Results and discussion

The results presented and discussed in this section are divided in two parts according to the flowchart in Fig. 9. The first one is dedicated to the fault location, where the TW theory and ICA are applied to the fault voltage database in order to separate the noise signal from the fault. The second one comprises SVM and MDS strategies

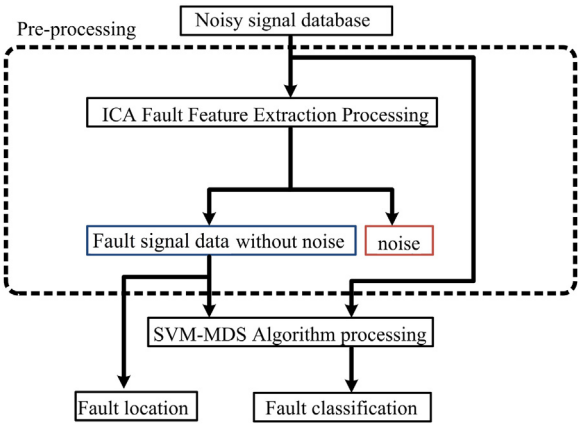

Fig. 9. Flowchart representing data processing.

Table 3
Location of a SLG fault for a fault distance of 10 km using 50 ICA basis functions.

SNR (dB)	Estimated distance (km)	Error (km)	Error (%)
30	9.75	0.25	0.125
35	9.75	0.25	0.125
40	9.75	0.25	0.125
45	9.75	0.25	0.125
50	9.75	0.25	0.125
55	9.75	0.25	0.125
60	9.75	0.25	0.125
65	9.75	0.25	0.125
70	9.75	0.25	0.125
75	9.75	0.25	0.125

Table 4
Location of a SLG fault for a fault distance of 90 km using 50 ICA basis functions.

Angles of incidence (θ)	Estimated distance (km)	Error (km)	Error (%)
0°	90.15	0.15	0.075
45°	90.15	0.15	0.075
90°	90.15	0.15	0.075
135°	90.15	0.15	0.075
180°	90.15	0.15	0.075
225°	90.15	0.15	0.075
270°	90.15	0.15	0.075
315°	90.15	0.15	0.075

as applied to the database in the absence of noise, considering that it was previously treated by the ICA algorithm.

The case studies presented in Table 2 are used to test the performance of the adopted algorithms. Before the strategies can be properly evaluated, it is necessary to define performance indexes capable of generalizing the results. Two parameters i.e. the relative error Error (%) and the average error AE (%) are defined as follows:

$$\text{Error}\,(\%) = \frac{|D_E - D_R|}{l_t} \times 100\% \tag{18}$$

$$\text{A.E}(\%) = \frac{\Sigma_1^N \text{Error}(\%)}{N_s} \tag{19}$$

where D_E and D_R are the estimated distance and real distance, respectively; and N_s is the number of samples in a given test case.

Tables 3–5 shown the performance of the ICA basis function on the location of fault distance of under the influence of distinct values of SNR, angle of incidence, and fault resistance. Several case studies considering distinct values assumed by the aforementioned parameters are evaluated in order to test the performance regarding the proposed approach, as relevant results are summarized in Table 6. The ICA technique using noisy data presents excellent performance even if compared with traditional processing ones as it will be seen in the forthcoming analysis. It is worth to mention that such typical methods do not consider the presence of noise in the employed data, what can be clearly stated from the values calculated for AE (%) listed in Table 7.

Table 5
Location of a SLG fault for a fault distance of 150 km using 50 ICA basis functions.

Resistance of fault (Ω)	Estimated distance (km)	Error (km)	Error (%)
20	155.33	0.33	0.165
50	155.33	0.33	0.165
80	155.33	0.33	0.165
120	155.33	0.33	0.165
150	155.33	0.33	0.165
180	155.33	0.33	0.165
200	155.33	0.33	0.165
240	155.33	0.33	0.165

Table 6
Location of faults using ICA basis function for all cases.

Real distance (km)	Estimated distance (km)	ICA basis function	Error (km)	Error (%)
10	9.75	50	0.25	0.125
16	15.95	50	0.005	0.025
45	45.33	70	0.33	0.165
84	84.17	115	0.17	0.085
90	90.15	115	0.15	0.075
128	128.56	115	0.56	0.280
147	147.67	250	0.67	0.335
155	155.33	250	0.33	0.165

Table 7
AE (%) for distinct fault types.

Fault type	10 km	16 km	84 km	90 km	128 km	147 km	155 km
SLG	0.01	0.01	0.01	0.02	0.05	0.09	0.10
LLG	0.01	0.01	0.02	0.02	0.05	0.08	0.10
LL	0.01	0.01	0.02	0.02	0.05	0.08	0.10
LLL	0.01	0.01	0.01	0.02	0.05	0.09	0.11

It can be seen that accuracy when determining the fault location is not influenced by the SNR, high values of fault resistance, and small angles of incidence. This is due to the fact that the ICA basis function splits the data space in two new subspaces containing data with the presence and absence of noise. Thus data without noise can be used for the accurate determination of the fault incidence point in the transmission line. Critical issues regarding the detection of the second peak in TW can be overcome when using ICA approach since noise is filtered. This is one of the most important properties of ICA.

In order to evaluate the performance of the proposed method, a proper comparison can be established with the other works studied in literature, which do not consider the effect of noise on the fault location. The wavelet energy associated to the coefficients can be defined as:

$$E_{jk} = |D_{(jk)}|^2 \tag{20}$$

where E_{jk} is the wavelet energy spectrum at scale j and instant k. The wavelet energy corresponds to the sum of the squared detailed coefficients obtained from the wavelet transform. The wavelet coefficient energy varies over different scales depending on the input signals. Then it is possible to achieve feature extraction regarding the fault location by applying such concept.

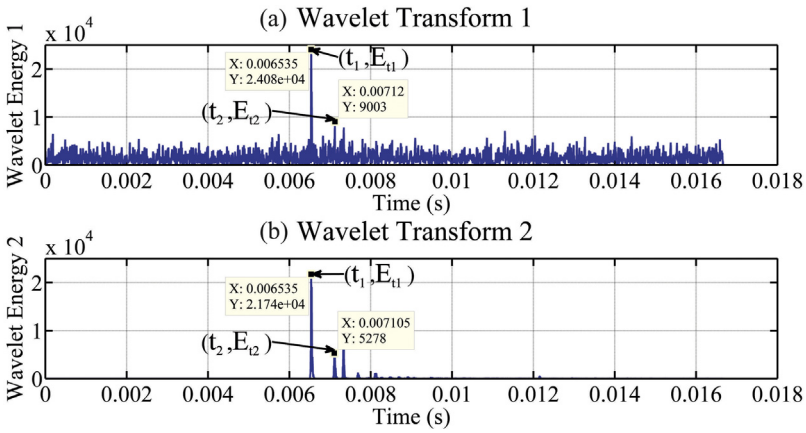

Fig. 10. Wavelet energy spectrum.

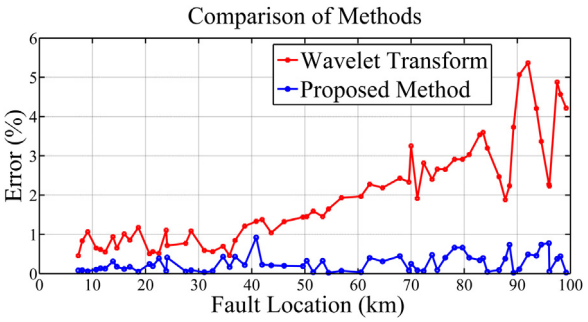

Fig. 11. Comparison between two methods used in fault location.

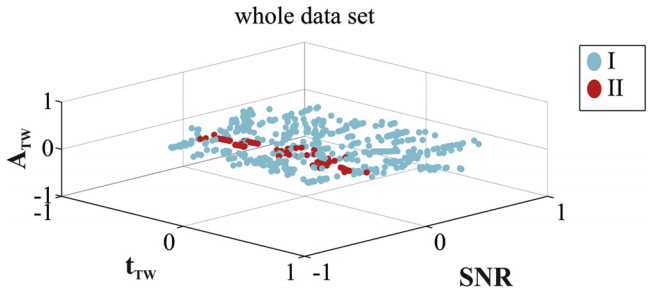

Fig. 12. Dimensional map for the data training in LibSVM with MDS.(For interpretation of the references to color in the text, the reader is referred to the web version of this article.)

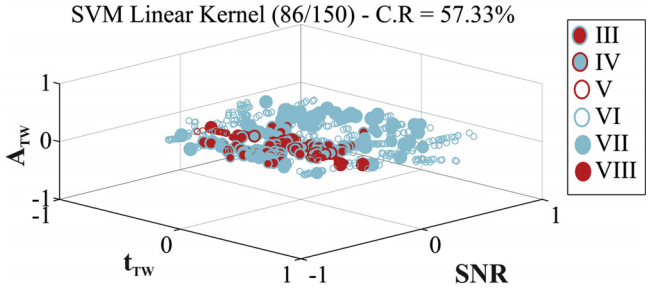

Fig. 13. SVM using a linear kernel.

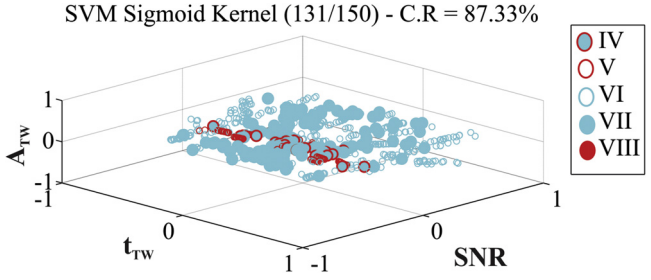

Fig. 14. SVM using a Sigmoid kernel.

Fig. 10 (a) and (b) presents the wavelet energy used in the detection and fault location under the same conditions assumed by the fault signal in Fig. 5 using daubechies 4 wavelet transform (db4) with two-level scale.

Fig. 10(a) shows that the wavelet energy 1 employs the first scale of db4 to detect and determine the magnitudes of the first peak E_{t1} and the second peak E_{t2}, while $t_1 = 0.006535$ s and $t_2 = 0.00712$ s are obtained by applying Eq. (1). The estimated distance for the fault location is 87.75 km, whose percent error is higher than 1%. Besides, both estimated values for E_{t2} and t_2 are corrupted due to the influence of noise if compared to the case where the ICA basis function is used.

Fig. 10(b) evidences that the wavelet energy 2 adopts the second scale of db4, where $t_1 = 0.006535$ s and $t_2 = 0.007105$ s as obtained from Eq. (1). The estimated distance for the fault location is 85.5 km and good accuracy results in this case.

For comparison purposes, simulation tests involving the method previously described in this work and the wavelet transform using wavelet energy are performed in order to verify their performance for SLG faults that occur until the first half of the line (F1). In this case, the fault distance is between 7.34 km and 99.2 km, the fault resistance is between 20 Ω and 240 Ω, and the angle of incidence varies from 0° to 360°. Simulation results are shown in Fig. 11, where SNR ranges from 30 dB to 75 dB.

The behavior of the curves is in accordance with the theoretical analysis. It can be seen that the relative error regarding the fault distance estimation using the proposed method is always lower than 1% under the adopted test conditions. On the other hand, the wavelet transform-based one assumes relative errors higher 1% for fault distances longer than 40 km, thus demonstrating that the technique presents superior performance.

According to the flowchart represented in Fig. 9, after the location of the fault point using TW and ICA, the next step consists in identifying the fault type, which is determined by employing a combination of SVM and MDS strategies. Although data and noise have been separated when using ICA, the original database corrupted by noise is considered for computation purposes in order to validate the effectiveness of SVM and MDS when dealing with noisy data. Besides, it is worth to mention that MDS is adequate as a graphical representation tool for the analysis of results.

Table 8
Test results regarding the classification of faults using LibSVM.

SNR (dB)	SLG		LLG		LL		LLL		Total		C.R (%)	N° SVs
	C	I	C	I	C	I	C	I	C	I		
30	150	0	147	3	148	2	149	1	594	6	99	40
35	150	0	148	2	148	2	149	1	595	5	99.1	20
40	150	0	149	1	150	0	149	1	598	2	99.6	20
45	150	0	150	0	150	0	149	1	599	1	99.8	20
50	150	0	150	0	150	0	150	0	600	0	100	17
55	150	0	150	0	150	0	150	0	600	0	100	17
60	150	0	150	0	150	0	150	0	600	0	100	15
65	150	0	150	0	150	0	150	0	600	0	100	15
70	150	0	150	0	150	0	150	0	600	0	100	15
75	150	0	150	0	150	0	150	0	600	0	100	15

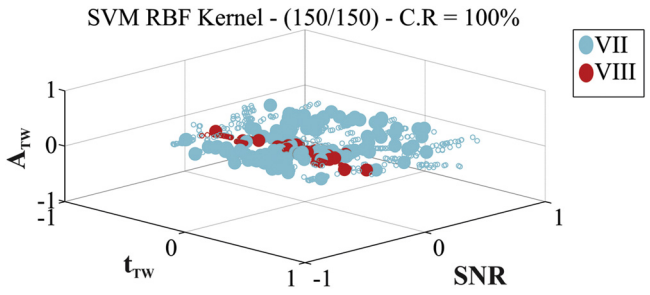

Fig. 15. SVM using a RBF kernel.

Table 9
Classification of faults using: OAO and OAA strategies.

Fault type	Accuracy
OAO strategies	
SLG ($C = 2000$ and $\sigma = 0.09$)	99.1%
LLG ($C = 100$ and $\sigma = 0.7$)	98%
LL ($C = 2000$ and $\sigma = 0.01$)	97.7%
LLL ($C = 100$ and $\sigma = 0.6$)	96.3%
OAA strategies	
SLG ($C = 100$ and $\sigma = 0.05$)	93.1%
LLG ($C = 10$ and $\sigma = 0.1$)	86.19%
LL ($C = 100$ and $\sigma = 0.04$)	80.1%
LLL ($C = 10$ and $\sigma = 1.2$)	87.56%

The voltage database used in this work is composed by four types of faults, which are divided in two sets labeled as $\tilde{X}$train and $\tilde{X}$test with 600 and 150 components each, respectively, corresponding to the input data for LibSVM. On the other hand, the output Yout = (+1; −1) is represented graphically by MDS with and without the presence of noise. By using such database, the fault classification process based on LibSVM can be summarized as follows:

1. Pre-processing of sample data. This step consists in the data minimization process while mapping a reduced dimensional space. The MDS representation is very important to the analysis of the fault classification process as well as the visualization of results.
2. Classification method for multi-classification problems. One-against-one (OAO) and one-against-all (OAA) are the strategies used to determine the best kernel parameters, where C is the regularization constant and σ is the Gaussian function width. Therefore LibSVM is responsible for performing the cross-validation used in fault classification in this stage.

In order to apply multi-class SVM with N types of faults, N(N-1)/2 binary classifiers are necessary. Some binary classifiers can be used for this purpose e.g. OAO and OAA, although OAO has provided better results when using LibSVM. Firstly the binary classifier distinguishes the noise signal, while OAO is then adopted to classify the corresponding fault type.

The results obtained with LibSVM are presented and discussed as follows, where the MDS graphical representation is used for this purpose. In order to explain the fault classification process, only a SLG fault is represented when data is corrupted by noise, whose SNR is 30 dB. The obtained results are properly classified from I to VIII according to the degree of success i.e. I – fault signals without noise; II – fault signals with noise; III – fault signals without noise and classified incorrectly; IV – fault signals with noise and classified incorrectly; V – fault signals with noise used in the training process; VI – fault signals without noise used in training process; VII – fault signals with noise and classified correctly and VIII – fault signals without noise and classified correctly.

After the voltage database is treated by ICA algorithm, two output subspaces are generated, thus representing data with and without noise. The aforementioned subspaces are shown in Fig. 12 by using MDS approach for graphical representation. This graph shows a dimensional map corresponding to propagation time (t_{TW}) and amplitude of the TW (A_{TW}), as well as SNR regarding the representation of the ICA output results valid for SLG fault. The fault data without noise (blue cluster – I) and with noise (red cluster – II) are also presented in Fig. 12. In order to establish a proper performance comparison, the input of the SVM algorithm is defined after the pre-processing stage according to Fig. 9. Data without noise and noisy data are used to train and test $\tilde{X}$train and $\tilde{X}$test defined as inputs for the SVM algorithm. Three types of SVM kernels are used in this case: i.e.linear, sigmoid, and RBF kernels under different noisy conditions, where SNR varies from 30 dB to 75 dB.

Fig. 13 shows the obtained results when employing the proposed SVM algorithm with linear kernel to classify 150 cases regarding SLG faults. According to the results, it can be seen that it is quite hard to classify the faults when in this case since only 86 among 150 SLG faults patterns have been correctly classified. Besides, significant problems exist: fault signals without noise are classified incorrectly (III); and fault signals with noise are classified incorrectly (IV). In other words, it can be stated that applying linear kernels to SVM method is unsuccessful when dealing with feature extraction in fault classification. According to Fig. 14, by using the sigmoid kernel only 131 among 150 SLG faults patterns have been correctly classified and one significant problem is been found: faults with noise are classified incorrectly (IV). Fig. 15 shows that the best results are achieved when using RBF kernel associated to SVM, as a classification rate of 100% is obtained when dealing with the faults. It is also worth to mention that both fault signals with and without noise are correctly classified in this case.

In order to apply multi-class SVM with N types of faults, N(N – 1)/2 binary classifiers are once again necessary. LibSVM is able to provide the best results when using OAO as a binary multi-class pattern. The accurate execution of SVM requires the proper turning of the involved parameters (C and σ), which are defined according to the cross-validation test. After three fold cross

Table 10
Performance comparison regarding several techniques applied to fault location and classification problems.

Proposed by	Reference number	Proposed method	Error (%)	Accuracy (%)
V. Malathi et al.	[6]	Wavelet and SVM	1%	99.11%
S. Eikici et al.	[7]	Wavelet and ANN	2.05%	–
S. Eikici	[8]	Wavelet and SVM	1%	99%
M. Jamil et al.	[9]	Wavelet and GNN	2%	–
A.A. Yusuff et al.	[10]	Wavelet and SVR	1%	100%
S.R. Samantary	[12]	Wavelet, S and Fuzzy	–	98%

validation, the parameters converge to $C = 1000$ and $\sigma = 1$, which correspond to the best outputs for the RBF kernel as shown in Table 8.

Table 8 represents the test results for LibSVM applied to fault classification. Columns C and I indicate that the classification has been performed correctly and incorrectly, respectively. Considering the first row in Table 8 where SNR is 30 dB, it can be stated that the algorithm presents a classification rate 100% when dealing with 150 cases regarding a SLG fault. The best classification results are achieved with SNR = 50 dB for an accuracy of 100% considering all types of faults.

On the other hand, the worst results correspond to 30 dB. It is worth to mention that data are strongly corrupted by noise in this case even when C.R = 99%, where accuracy is considered to be satisfactory. Therefore it is reasonable to conclude that combining ICA with SVM and MDS provides an adequate tool for the classification of faults when dealing with noisy voltage signal database. Table 9 presents supplementary values with the variation of C and σ for two distinct multi-classification problems: OAO and OAA strategies.

According to this study, the proposed approach combining ICA with SVM associated with OAO presents accuracy values up to 100% regarding the classification of faults for several values of SNR (from 30 dB to 75 dB). Besides, it presents superior performance if compared with the techniques describes in Table 10.

- The works dealing with combined strategies such as wavelet and SVM and wavelet with ANN have only addressed the classification of fault signals without noise, even though good accuracy has been achieved.
- The combination of wavelet with fuzzy is only satisfactory for SNRs up to 20 dB, thus evidencing that wavelets are sensitive to noise. Besides, the fuzzy system demands a large set of logical rules specifically chosen for the data set.
- The combination of ICA with TW theory and ICA with SVM has provided more relevant results considering the presence of noise. This is due to the application of basis functions using the concept of efficient codding so that minimization of data redundancy with better simplification is achieved in the modeling process. Consequently, faster processing time and improved accuracy in the location and classification of faults result if compared to traditional techniques such as wavelets and ANNs.

7. Conclusions

The analysis of many case studies regarding transmission line faults has shown that the ICA approach proposed in this work can be used for the successful extraction of noise in databases. Besides, the TW theory can be used to determine the fault location in a simple way. On the other hand, a powerful technique results when ICA is combined with SVM. By using MDS for the graphic representation of data, the advantages associated to ICA can be clearly demonstrated.

Generally, the analysis carried out in this work has shown that the proposed combination of methods provides excellent results for fault location purposes since the involved errors are lower than 1%, with good classification of faults noise levels higher with an accuracy of 100% considering all types of faults.

The obtained results have demonstrated that the introduced approach presents better performance than those regarding some conventional techniques e.g. wavelets and ANNs as summarized in Table 10, since noise issues can be readily overcome.

Acknowledgement

The authors would like to thank FAPEPI (Fundação de Amparo a Pesquisa do Estado do Piauí) for their assistance and financial support.

References

[1] A. Abur, F.H. Magnago, Fault location using wavelets, IEEE Trans. Power Deliv. 13 (October) (1998) 1475–1480.
[2] S. Hasheminejaed, S. Seifossadat, M. Razaz, M. Joorabian, Ultra-high speed protection of transmission lines using travelling wave theory, Electr. Power Syst. Res. 132 (2016) 94–103.
[3] Z. He, X. Li, R. Mai, A novel travelling wave directional relay based on apparent surge impedance, IEEE Trans. Power Deliv. 30 (2015) 1153–1161.
[4] Z. He, X. Li, S. Chen, A travelling wave natural frequency based single-ended fault location method with unknown equivalent system impedance, Int. Trans. Electr. Energy Syst. (2015), http://dx.doi.org/10.1002/etep.2090 (Online publication).
[5] G. Zhang, H. Shu, Y. Liao, Automated double-ended travelling wave record correlation for transmission line disturbance analysis, Electr. Power Syst. Res. 136 (2016) 242–250.
[6] V. Malathi, N.S. Marimuthu, S. Baskar, Intelligent approaches using support vector machine and extreme learning machine for transmission line protection, Neurocomputing 73 (2010) 2160–2167.
[7] S. Ekici, S. Yildirim, M. Poyraz, Energy and entropy-based feature extraction for locating fault on transmission lines by using neural network and wavelet packet decomposition, Expert Syst. Appl. 34 (4) (2008) 2937–2944.
[8] S. Ekici, Support vector machine for classification and locating faults on transmission lines, Appl. Soft Comput. 12 (2012) 1650–1658.
[9] M. Jamil, A. Kalam, A.Q. Ansari, M. Rizwan, Generalized neural network and wavelet transform based approach for fault location estimation of a transmission line, Appl. Soft Comput. 19 (2014) 322–332.
[10] A.A. Yusu_, A.A. Jimoh, J.L. Munda, Fault location in transmission lines based on stationary wavelet transform, determinant function feature and support vector regression, Electr. Power Syst. Res. 110 (2014) 73–83.
[11] V.H. Ferreira, R. Zanghi, M.Z. Fortes, G.G. Sotelo, R.B.M. Silva, J.C.S. Souza, A survey on intelligent system application to fault diagnosis in electric power system transmission lines, Electr. Power Syst. Res. 136 (2016) 135–153.
[12] S.R. Samantary, A systematic fuzzy rule based approach for fault classification in transmission lines, Appl. Soft Comput. 13 (2013) 928–938.
[13] A. Hyvarinen, J. Karhunen, E. Oja, IndependentComponentAnalysis, Wiley-Interscience, 2001.
[14] E.E. Ngu, K. Ramar, A combined impedance and travelling wave based fault location method for multi-terminal transmission lines, Electr. Power Energy Syst. 33 (2011) 1767–1775.
[15] E. Clarke, Circuit Analysis of A-C Power Systems, JohnWiley, New York, 1993.
[16] Hans Kristian HØidalen, ATPDraw version 5.6p5 for Windows 9x/NT/2000/XP/Vista/7, 201.
[17] J.R. Marti, Accurate modelling of frequency dependent transmission lines in electromagnetic transient simulations, IEEE Trans. Power Appar. Syst. PAS-101 N1 (1982) 147–155.
[18] M.M. Saha, J. Izykowski, E. Rosolowski, Fault Location on Power Networks, first ed., Springer, New York, 2010, pp. 8.
[19] A. Cichocki, S. Amari, Adaptive Blind Signal and Image Processing, John Wiley-Sons, Ltd., 2003.
[20] A.C. Ribeiro, A.K. Barros, E. Santana, J.C. Principe, Diabetes classification using a redundancy reduction preprocessor, Res. Biomed. Eng. 31 (2015) 97–106.
[21] A. Abur, F.H. Magnago, Use of time delays between modal components in wavelet based fault location, Electr. Power Energy Syst. 22 (2000) 397–403.
[22] C.C. Chang, C.J. Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 27 (2) (2011) 1–27.
[23] V. Vapnik, The support vector method of function estimation, in: J. Suykens, J. Vandewalle (Eds.), Nonlinear Modeling: Advanced Black-Box Techiniques, Kluwer Academic Publishers, 1998, pp. 55–56.
[24] C.C. Chang, C.J. Lin, LIBSVM: a Library for Support Vector Machines, 2001, Software available at https://www.csie.ntu.edu.tw/_cjlin/libsvm.
[25] I.T. Jolliffe, Principal Component Analysis, Springer, New York, 2002.
[26] G.K. Matsopoulos, Self-Organizing Maps, INTECH, 2010.
[27] I. Borg, P.J.F. Groenen, Modern Multidimensional Scaling: Theory and Applications, Springer, New York, 2005, pp. 207–212.
[28] S.H. Bae, J. Qiu, G. Fox, Adaptive interpolation of multidimensional scaling, in: International Conference on Computational Science ICCS, Omaha, Nebraska, June 4–6, 2012.