Fault Detection and Classification in Transmission Lines Using Wavelet Transform and Support Vector Machines

Detección y Clasificación de Fallas en Líneas de Transmisión Utilizando Transformada Wavelet y Máquinas de Soporte Vectorial

F.I. Lozada F.L. Quilumba F.E. Pérez

Escuela Politécnica Nacional, Quito, Ecuador
E-mail: fabricio.lozada@epn.edu.ec; franklin.quilumba@epn.edu.ec; fabian.perez@epn.edu.ec

Abstract

This paper proposes a method using Discrete Wavelet Transform and Support Vector Machines (SVM) to detect and classify faults along a transmission line. Discrete Wavelet Transform is used to detect the high frequency components contained in a fault signal spectrum. Different faults (short circuits) on different positions of a transmission line are simulated using a link between MATLAB and ATP/EMTP programs. The algorithm implemented in MATLAB is tested in various case studies based on the largest transmission line of the IEEE 9 bus system.

***Index terms*− ATP/EMTP, fault detection, Fourier transform, machine learning, discrete wavelet transform (DWT), support vector machine (SVM).**

Resumen

Este documento propone un método utilizando transformada Wavelet discreta (DWT) y Máquinas de Soporte Vectorial (SVM) para detectar y clasificar fallas a lo largo de una línea de transmisión. La transformada Wavelet discreta es utilizada para detectar componentes de alta frecuencia contenidas en el espectro de una señal con falla. Diferentes fallas (cortocircuitos) son simulados en diferentes posiciones a lo largo de una línea de transmisión, para ello se utiliza un enlace entre los programas MATLAB y ATP/EMTP. El algoritmo implementado en MATLAB es probado en varios casos de estudio a lo largo de la línea de transmisión más larga del sistema de 9 barras de la IEEE.

***Palabras clave*−ATP/EMTP, detección de fallas, transformada de Fourier aprendizaje automático, transformada wavelet discreta (DWT), máquinas de soporte vectorial (SVM).**

Recibido: 29-06-2017, Aprobado tras revisión: 15-11-2017
Forma sugerida de citación: Cruz, C.; Mier, A.; Quilumba, F. (2018). "Fault Detection and Classification in Transmission Lines Using Wavelet Transform and Support Vector Machines ". Revista Técnica ***"energía"***. No. 14, Pp.151-158
ISSN 1390-5074.

1. INTRODUCTION

Considering the continuous growth of energy demand at world level, it is necessary to fulfill two requirements of utmost importance: 1) continuity and 2) quality of the electric power supply, these requirements will guarantee the proper functioning of the electric power system and thus society [1]. The correct performance of the protection system ensures compliance with these two requirements. They are conceived to protect assets in the event of a failure in the system, and ensure continued supply of electricity.

It is practically impossible to discard the existence of faults in any of the elements that constitute a power system. Thus, it is imperative to have a well-designed protective relay system that has as objective to minimize the effects of the different faults that can be present in the system.

On one hand, there is electrical equipment such as generators and power transformers that are not exposed to the inclement weather, and are generally located within power plants, and substations, making it very unlikely that any faults due to short circuits affect them [1]. On the other hand, transmission lines being the element that covers most of the system's trajectory and being subjected to any type of weather condition, are more vulnerable and therefore considered the most critical part of the power system. Transmission lines must be protected as best as possible to keep them in excellent operating condition. When a fast detection of a fault event is achieved, it is possible to improve restoration times, thus ensuring the reliability of the power system.

The Fourier transform is an excellent tool used to perform the analysis of stationary signals. It presents certain deficiencies, especially when performing analysis and detection of sudden changes that may be contained in those signals considered non-stationary [2], [3]. At present, there are a variety of sophisticated tools, which are very useful when performing fault analysis on transmission lines. One of them is the Wavelet transform, which analyzes signals that do not behave in a stationary way like signals after a fault event in transmission lines. A methodology that is capable of detecting and identifying a fault is developed in this work.

2. THEORICAL FRAMEWORK

2.1. **Wavelet Transform**

The Wavelet transform is a mathematical tool used for signal analysis. It appears to overcome deficiencies presented by the Fourier analysis when treating nonstationary signals [3]. A wavelet is the name given to a "small wave" whose energy is concentrated in a given period of time, and it is characterized by being a signal of non-stationary nature with a window function of finite length [4]. It presents a good performance when working with transient and high frequency phenomena, since once the width of the window is defined, it analyzes all frequencies with the same resolution, both in time and frequency [2], [5].

Considering the discretization of the values for both, scale and time, the Discrete Wavelet Transform (1) is used, thus achieving a better computational time in terms of data processing.

$$DWT\ (m\ ,n)=\frac{1}{\sqrt{2^m}}\sum_k f(k)\ \varphi\left[\frac{n-k\cdot 2^m}{2^m}\right] \tag{1}$$

where $\varphi\left[\frac{n-k\cdot 2^m}{2^m}\right]$ is the mother wavelet.

According to [6-8] regarding to the application of the Wavelet transform in power systems, specifically for the study of transients caused by faults, the Daubechies family is used, concretely, the Daubechies with 4 vanishing moments "db4".

2.2. Machine Learning

Machine learning is a branch of Artificial Intelligence that allows a machine or computer to learn (or improve its performance) based upon algorithms that have the ability to generalize behaviors and recognize patterns from an available data set. Supervised and unsupervised learning are two categories based on the way machines learn from data [6]which will be used in various applications. Specifically, it explains data mining and the tools used in discovering knowledge from the collected data. This book is referred as the knowledge discovery from data (KDD. In the supervised learning methods, there are several techniques, one known as Support Vector Machine (SVM), which possesses a good theoretical foundation and a good generalization capability. SVM will be used as a classification tool in this work.

2.3. Support Vector Machines

Due to the great performance in the different applications including surpassing in some cases the performance of traditional neural networks, they have been introduced as powerful tools to solve classification problems [7].

The idea is based on the classical concept of the optimal separation hyperplane, where the separation direction vector is expressed as a function of the training examples [6]which will be used in various applications. Specifically, it explains data mining and the tools used in discovering knowledge from the collected data. This book is referred as the knowledge discovery from data (KDD as shown in Fig. 1.

The hyperplane, H, thus obtained is called the optimal separation hyperplane, for which the margin is maximum. The margin is obtained as the distance between the two hyperplanes, H1 and H2, which contain at least one point in each class, these points are known as "support vectors". Within the SVMs there are 3 cases to be considered:

1) Linearly separable data.

2) Soft-margin hyperplanes.

3) Non-linearly separable data.

The first case is the least applicable to real problems, but it turns out to be an illustrative method, considering that the other two methods are a generalization of this one. To solve cases 2 and 3, one can use the so-called Kernel functions, which are mathematical functions used to project the data from a characteristic space to a space represented by a higher dimension [6]which will be used in various applications. Specifically, it explains data mining and the tools used in discovering knowledge from the collected data. This book is referred as the knowledge discovery from data (KDD, [8]. In other words, they are functions that convert a non-linear classification problem in the original dimensional space, into a simple linear classification problem in a larger dimensional space (Fig. 2).

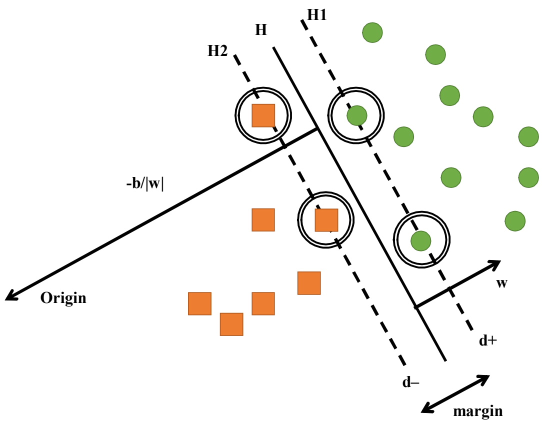

Figure 1: Optimal Separating Hyperplane

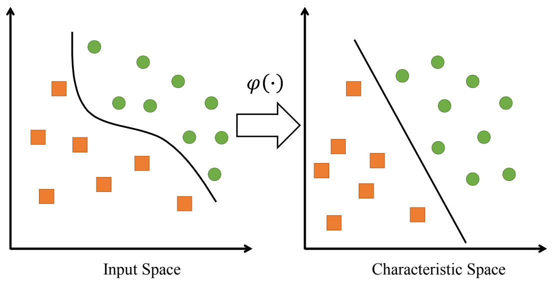

Figure 2: Data Transformation using Kernels

3. PROPOSED METHODOLOGY

Within the proposed methodology are used two computer programs. The first program, ATP/EMTP is a tool mostly used to performing simulations whose nature is directly related to electromagnetic and electromechanical transient's phenomena in electric power systems. The second program is MATLAB that is a mathematical software tool that offers an integrated development environment with its own programming language.

A link between the two programs are used in order to obtain a database. The database consists of the signals obtained from the different short circuits simulations along the transmission line. Discrete Wavelet Transform is used to decompose different signals to obtain detail coefficients. Finally detail coefficients are used to detect and classify the different type of faults (Short Circuits).

3.1. Automatic Fault Simulation

As a test system, the IEEE 9-Bus System is used. This test system consists of 3 synchronous machines, 9 buses, 6 transmission lines, 3 transformers and 3 constant impedance loads [9]. The test system is constructed in ATP/EMTP. Fig. 3 shows the single-line diagram.

As previously mentioned, once the entire test system has been modeled, the study focuses on the longest transmission line of the test system, in this case the L2 transmission line. For this, in [8] a tool is developed that can drive ATP from MATLAB. This tool is an efficient alternative to automatically simulate different fault conditions along the transmission line.

Fig. 4 shows the transmission line L2. The transmission line has a length of 180 km, has three transpositions and has been divided into four sections of 1/6, 1/3, 1/3 and 1/6 of the total length of the line. Each section of the line has been subdivided into two sections and four internal nodes named NF1, NF2, NF3 and NF4 that are used to connect different fault schemes. The schemes for simulation of the different types of faults are shown in Fig. 5.

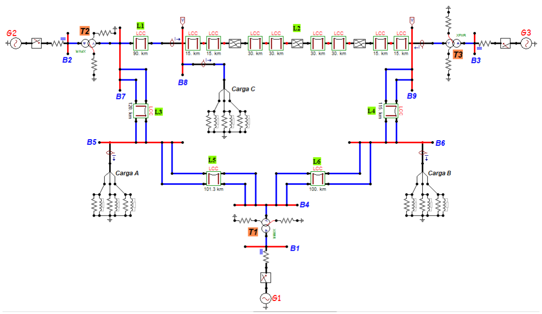

Figure 3: IEEE 9-Bus System in ATPDraw

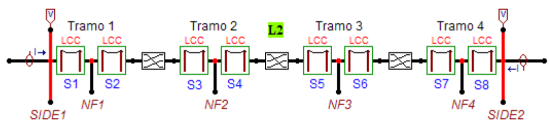

Figure 4: Modeling of the Longest Transmission Line in ATPDraw for Different Fault Conditions

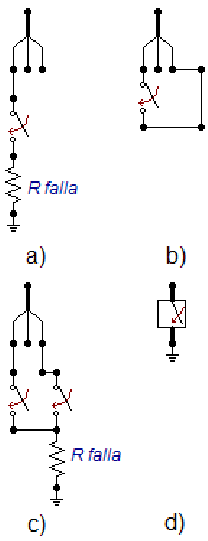

Figure 5: Modeling of a) Single Line-to-Ground b) Line-to-Line c) Double Line-to-Ground d) and Three-Phase Fault in ATPDraw

To carry out the simulation process, within the program developed in [10] the following should be considered:

- L: total length of transmission line 180 [km].
- f: System Frequency 60 [Hz].
- Rrfault: Fault resistance value. For single line-to-ground (SLG) faults, a fault resistance of 20 Ω has been selected, in the case of double line-to-ground (DLG) faults a value of 10 Ω, and for line-to-line (LL) and three-phase (TP) faults a fault resistance of 0 Ω.
- phi: Fault inception angle in degrees. For all types of fault, a fault inception angle of 90 ° is selected.
- x0: Initial distance at which the fault occurs. An initial distance of 1 km is selected.
- step_x: Length at which the fault travels along the transmission line. Steps of 1 km are defined so that the fault moves from one end to the other.

3.2. Data Preparation

To each of the signals, both current and voltage measured at Bus 1, the data reduction technique known as discrete wavelet transform is applied. The mother wavelet used is the Daubechies 4 “db4”, since it is a type of Wavelet that fits very well when it comes to identifying high frequency transients.

The decomposition of the signals of interest is performed using MATLAB with the help of the “wavedec” command. It performs a multilevel Wavelet decomposition in a single dimension (1-D), for which it uses a specific wave type “wname” or a specific set of Wavelet decomposition filters [11].

[C, L] = wavedec (X, N, ‘wname’) returns the Wavelet decomposition of the signal X, for a decomposition level N, where N is a strictly positive integer value and a specific wavelet type “wname” [11].

3.3. Fault Detection Algorithm

The fault detection is on the basis of sudden change in the current signal by comparing between the pre-fault current and fault current. If the change in the current signal is analyzed in the frequency domain, it´s clearly found that the change is detected in both high frequency components and the fundamental component [12]”page”:”2246-2250vol.3”,”volume”:”3”,”source”:”IEEE Xplore”,”event”:”Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES”,”*abstract*”:”This paper presents a new method to diagnose faults in a transmission system. This is based on detecting high frequency components contained in a fault signal spectrum. The discrete wavelet transform (DWT. It is noted that the change in the high frequency components occurs more quickly than that in the fundamental component, and such a change also depends on fault characteristics.

The following steps are essential to be considered to perform the fault detection algorithm:

1) Calculate the positive sequence current from the phase currents obtained in the different simulations.

2) Decompose the positive sequence current signal using the DWT to obtain the detail coefficients at the N = 1 level.

3) Calculate the "spectral energy" from the coefficients obtained performing the decomposition of the signal using DWT.

For fault detection, coefficients obtained using the Wavelet Transform of the positive sequence current are squared so that the abrupt change in the spectra can be clearly found. The condition for judging the fault current is that there is a change in the coefficients around five times of those before an occurrence of the fault as stated in (2):

$$SE_1 > 5 \cdot SE_{CN} \quad (2)$$

Where SECN is the spectral energy at normal conditions, and SE1 is the spectral energy of the fault signal. The spectral energy of a fault signal is bigger than spectral energy of a signal in normal conditions [12].To stablish the threshold, it´s necessary to determine the minimum short circuit current (of all set of simulated faults) and of that signal obtain their spectral energy. It is to be expected that even though the signal is the minimum short circuit current, her energy is greater than 5 times the pre-fault energy. For this reason the value of 5 times is selected. This value is adequate because, it is not so low (say 2 times) as to produce confusions between fault signals with signals of overloads or external faults. The flow chart for fault detection is illustrated in Fig. 6.

Positive sequence current is decompose using Discrete Wavelet Transform to obtain detail coefficients. The spectral energy exceeds the established threshold. In this case the algorithm detect a fault condition. Results are shown in Fig. 7.

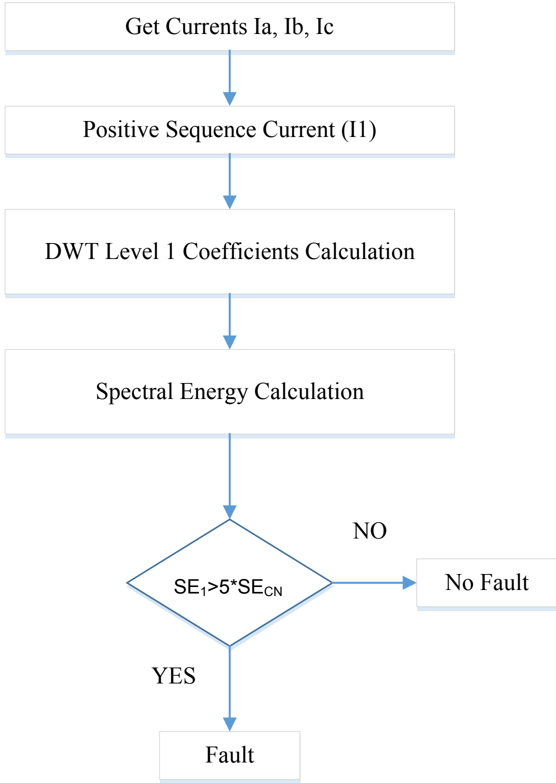

Figure 6: Flow Chart for Fault Detection

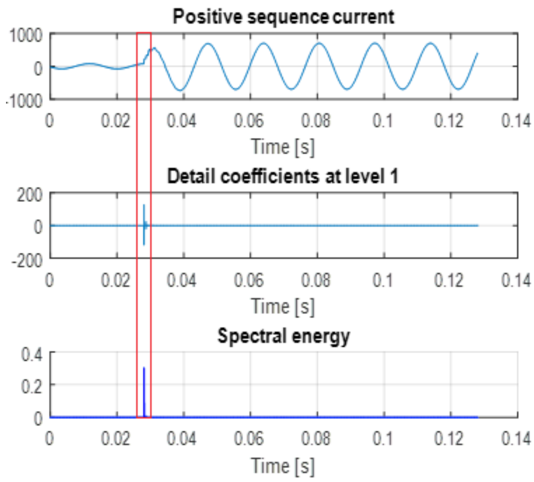

Figure 7: Technique used in Fault Detection

On the other hand in Fig. 8, phase fault currents (IA, IB and IC) are decomposed using Discrete Wavelet Transform to obtain detail coefficients. As in the previous case this coefficients are squared so that abrupt change in the spectra can be clearly found. In this case the technique used to detect faults works properly, additionally this method could be used as a classifier, since as it is observed in Fig. 8 the variation of the spectral energy occurs in the three phase currents and it could be considered as a three-phase fault.

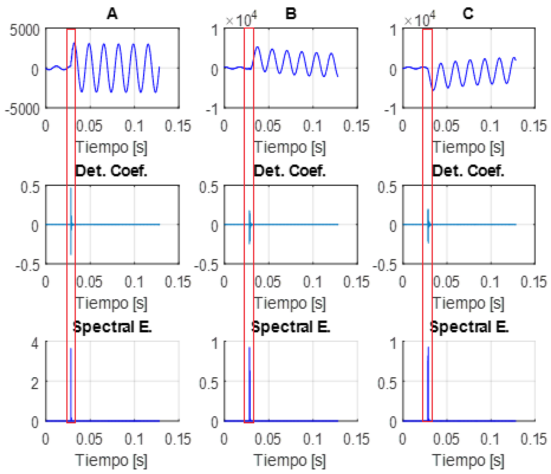

Figure 8: Fault Detection using Phase Currents

In Fig. 9 we can observe an abrupt change in the spectral energy that occurs in two phases. In this case the algorithm detect a fault between A and C phase currents.

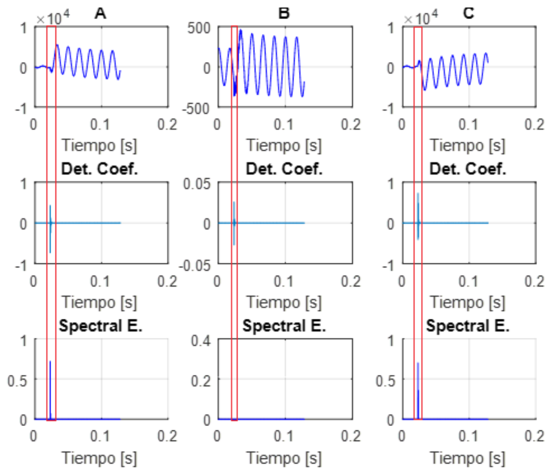

Figure 9: Fault Detection using Phase Currents

3.4. Fault Classification Algorithm

To carry out the tasks of fault classification, it is necessary to form two models. One model formed by patterns obtained through the Wavelet decomposition of current fault signals, and the second model formed by patterns obtained through the Wavelet decomposition of voltage and current fault signals. Both models has associated only one output with eleven different categories as shown in Table I.

Table I: SVM Output

Fault Type	SVM Output
	Y
Single Line-to-Ground A	SLG A
Single Line-to-Ground B	SLG B
Single Line-to-Ground C	SLG C
Line-to-line AB	LL AB
Line-to-line BC	LL BC
Line-to-line CA	LL CA
Double line-to-ground ABT	DLG ABT
Double line-to-ground BCT	DLG BCT
Double line-to-ground CAT	DLG CAT
Three-Phase Fault	TP
No Fault	NF

The decomposition of the signals is done to obtain the detail coefficients for a level 7 using the Daubechies 4 “db4” Wavelet.

The first input matrix, on now called matrix Q1 consists of 1980 rows corresponding to the 180 simulations for the 11 output classes and the 318 columns corresponding to the 106 coefficients for three phase currents. On the other hand, the second input matrix, on now called matrix Q2 is formed by the coefficients obtained from Wavelet decomposition of current and voltage signals, it has 1980 rows and 636 columns corresponding to the 106 coefficients for three phase currents and 106 coefficients for three phase voltages.

The output matrix, now called matrix Y has the correct responses of the matrix Q1 and matrix Q2. Once the set of samples has been obtained, it’s necessary to divide it into three data sets:

- Training: data used for model training.
- Validation: select the best of the trained models.
- Test: Indicates how well the model classifies. In this case it’s necessary use different samples from those in the main database.

Before starting the training model, the ‘Normalization’ concept is applied in order to obtain data with the same weights.

The selection of training and validation sets is made based on percentages, whose typical values are 60% and 40% respectively. Other typical values used are 70% and 30%. In this case 60% of the total data is extracted randomly, thus forming the training set. The remaining 40% constitutes the validation set.

The ‘fitcecoc’ command fit multiclass models for support vector machines or other classifiers.

Mdl = fitcecoc(Qtraining, Ytraining) returns a full, trained ECOC model using the predictors X and the class labels Y.

- Qtraining: matrix that contains predictor variables, specifically contains numeric values.
- Ytraining: is the matrix that contains responses.

Once the training set is defined, SVM’s are used by the ‘fitcecoc’ command in MATLAB.

Mdl = fitcecoc(Q, Y)

Mdl is the variable that contains the classifier or training model.

4. FINAL RESULTS

The technique used to detect faults along a transmission line in the selected test system has been used successfully. It is a good idea to work only with the detail coefficients obtained from the decomposition of phase fault currents because we also use this method as a classifier.

In the other hand, to know the performance of the classifier, the confusion matrix is obtained as shown in Fig. 10. All classes except Class 8 are classified correctly. From a total of 80 signals corresponding to DLG faults between the BC-T (class 8), 70 of them are correctly classified as DLG faults between BC-T phases, while 10 of them are classified as LL faults between BC phases. It would not represent a major problem, because in case any of the two faults (biphasic BC-T or biphasic BC) are present in the transmission line, the relay associated with the protection system would receive the order of open the phases involved and open BC phases independently if it were an isolated biphasic fault or a biphasic ground fault between the phases in question. Overall, 98.7% of the predictions are correct and 1.3% are wrong classifications.

Output Class

80 10.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100 0.0%
0 0.0%	61 7.7%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100 0.0%
0 0.0%	0 0.0%	75 9.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100 0.0%
0 0.0%	0 0.0%	0 0.0%	72 9.1%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100 0.0%
0 0.0%	0 0.0%	0 0.0%	0 0.0%	61 7.7%	0 0.0%	0 0.0%	10 1.3%	0 0.0%	0 0.0%	0 0.0%	35.9 14.1%
0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	76 9.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100 0.0%
0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	75 9.5%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	100 0.0%
0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	70 8.8%	0 0.0%	0 0.0%	0 0.0%	100 0.0%
0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	68 8.6%	0 0.0%	0 0.0%	100 0.0%
0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	75 9.5%	0 0.0%	100 0.0%
0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	69 8.7%	100 0.0%
100 0.0%	100 0.0%	100 0.0%	100 0.0%	100 0.0%	100 0.0%	100 0.0%	37.5 12.5%	100 0.0%	100 0.0%	100 0.0%	98.7 1.3%

Target Class

Figure 10: Confusion Matrix

5. CONCLUSIONS

This work presents a practical application of Discrete Wavelet transform for the fault diagnosis in transmission lines. Coefficients of positive sequence current signals are calculated and employed in fault detection decision algorithm.

Automatic simulation tool allows to obtain different signals automatically and systematically, considerably reducing the simulation time in the different cases studied.

Wavelet transform has proven to be very efficient and effective in analyzing a wide class of signals and phenomena. In this case, fault signals corresponding to currents and voltages.

The Wavelet transform with Daubechies 4 ‘Db4’ as a mother Wavelet has been used. It has been found that the proposed technique give satisfactory results in detecting and classifying faults along of the transmission line in the test system used.

The average accuracy of fault classification using DWT and SVM´s from the decision algorithm proposed in this paper is highly satisfactory.

Support vector machines are sophisticated tools that can be used to implement new methodologies in the area of power protection systems to develop a great variety of algorithms that allow to detect, classify and locate faults.

REFERENCES

[1] R. A. Margil S., “Protección de Sistemas Eléctricos de Potencia,” Universidad Autónoma de Nuevo León, San Nicolás de los Garza, Nuevo León, 2005.

[2] S. Kouro and R. Musalem, “Tutorial Introductorio a la Teoría Wavelet.” [Online]. Available: http://www2.elo.utfsm.cl/~elo377/documentos/Wavelet.pdf. [Accessed: 26-Jul-2016].

[3] J. Cortés, F. Medina, and J. Chávez, “Del Análisis de Fourier a las Wavelets.” [Online]. Available: https://dialnet.unirioja.es/descarga/articulo/4807129.pdf. [Accessed: 26-Jul-2016].

[4] P. S. Addison, The Illustrated Wavelet Transform Handbook: Introductory Theory and Applications in Science, Engineering, Medicine and Finance. CRC Press, 2002.

[5] J. Martínez Malo and R. Castro Fernández, “Análisis de la teoría de ondículas orientadas a las aplicaciones en Ingeniería Eléctrica.” [Online]. Available: http://www.uta.cl/hdiaz/Documents/Fundamentos_Wavelets.pdf. [Accessed: 26-Jul-2016].

[6] J. Han, J. Pei, and M. Kamber, Data Mining: Concepts and Techniques. Elsevier, 2011.

[7] C. J. C. Burges, “A Tutorial on Support Vector Machines for Pattern Recognition,” Data Min. Knowl. Discov., vol. 2, Jan. 1998.

[8] R. Solera Ureña, “Máquinas de vectores soporte para reconocimiento robusto de habla,” 2011.

[9] J. Grainger and W. Stevenson, Análisis de Sistemas de Potencia. .

[10] F. E. Pérez Yauli, “Validación del sistema de protección de las líneas que conforman el anillo de 230 KV del Sistema Nacional de Transmisión - SNT, con el programa Power Factory de la empresa DigSILENT,” Thesis, QUITO/EPN/2004, 2004.

[11] “Multilevel 1-D wavelet decomposition - MATLAB wavedec.” [Online]. Available: https://www.mathworks.com/help/wavelet/ref/wavedec.html?s_tid=srchtitle.

[12] P. Makming, S. Bunjongjit, A. Kunakorn, S. Jiriwibhakorn, and M. Kando, “Fault diagnosis in transmission lines using wavelet transform analysis,” in Transmission and Distribution Conference and Exhibition 2002: Asia Pacific. IEEE/PES, 2002, vol. 3, pp. 2246–2250 vol.3.

Fabricio I. Lozada Jordán.- was born in Ambato, Ecuador, in 1991. He received his B. S. degree in Electrical Engineer from National Polytechnic School (Escuela Politécnica Nacional EPN), Quito, Ecuador, in 2017. His main research area is in application of computer technology to power system monitoring, operation, protection and control. His other research interests includes smart grids, artificial intelligence and signal processing.

Fabián E. Pérez Yauli.- (M’2013) was born in Ambato, Ecuador. He graduated as Electrical Engineer (B.E.) from National Polytechnic School (Escuela Politécnica Nacional EPN), Quito, Ecuador in 2004. In 2012, he obtained the Doctor degree at the Universidad Nacional de San Juan, San Juan, Argentina, with a scholarship awarded by the German Academic Exchange Service (DAAD). Currently is a teacher in the Escuela Politécnica Nacional. His major interest included power system protection and signal processing.

Franklin L. Quilumba.- (S’10-M’14) received the Diploma degree in electrical engineering from the National Polytechnic School (Escuela Politécnica Nacional EPN), Quito, Ecuador, in 2008, and the M.S. and Ph.D. degrees in electrical engineering from the University of Texas at Arlington (UTA), Arlington, TX, in 2014. In 2014, he joined the faculty of National Polytechnic School, where he currently holds the rank of Associate Professor. His areas of interest are power systems analysis, operation, stability and control; computer simulation of electric power systems; power load modeling; generation and transmission planning; demand response; and load forecasting.