Power System Fault Detection, Classification and Location Using the K- Nearest Neighbors

Abdelmadjid Recioui, Brahim Benseghier, Hamza Khalfallah

Laboratory signals and systems, Institute of Electrical and Electronic Engineering, University of Boumerdes
Avenue de l'indépendance, Boumerdes, 35000, Algeria.

recmad2006@yahoo.fr

***Abstract*— Power systems are frequently subjected to faults. These faults can cause the destruction of expensive power system components such as motors, generators, and transformers, explosions due to over-voltages, high currents, outage and death. A power protection system is required in order to detect, classify, and locate the fault to clear it rapidly and to minimize the effects mentioned before. This paper deals with the detection, classification and location of power system faults on the IEEE 14 Bus Test Network using the K-Nearest Neighbors based on MATLAB/SIMULINK.**

***Keywords*— Power systems, fault detection, fault classification, fault location, K-nearest Neighbour algorithm.**

I. INTRODUCTION

The rapid growth of electric power systems over the past few decades has resulted in a large increase of the number of lines in operation and their total length. These lines are exposed to faults as a result of lightning, short circuits, faulty equipments, mis-operation, human errors, overload, and aging [1].

When a fault occurs on an electrical transmission line, it is very important to detect, classify it, and to find its location in order to make necessary repairs and to restore power as soon as possible. As the time needed to determine the fault point along the line will reflect the quality of the power delivery. Therefore, a sophisticated fault detection technique and an accurate location on the line is an important requirement for a permanent fault. Pointing to a weak spot, it is also helpful for a transient fault, which may result from a marginally contaminated insulator, or a swaying or growing tree under the line [2-3].

Most faults can cause large currents or voltages changing, and they are often detected by traditional protective relay. Whereas, some faults, such as high impedance faults, grounding faults of ineffectively earthed distribution systems, cause small currents and voltages changing and they are difficult to be detected using traditional protective relay. Research efforts are focused on developing efficient fault detection, classification and location methods, intended for application to more and more complex networks varieties of faults [4].

Various fault classification and location methods have been proposed and they can be categorized as analytical methods, artificial intelligence based methods, travelling wave methods and software based methods [5].

Fault classification identifies the type of fault on the fault lines in the system. There are different methods proposed for the fault classification. The wavelet based approach for fault classification is implemented using multi resolution analysis (MRA) of voltage and current signals [6-7]. Adaptive Neuro-Fuzzy Inference System (ANFIS) is based on fuzzy logic modeling and uses artificial neural network as the learning algorithm. The fuzzy logic can map inputs to outputs based on the predefined "if then" rules. Fuzzy logic methods can model the uncertainty in the available data and can help in estimating the fault classification and location [8 –10].

After fault classification, the next step is to determine the fault location. The fault location identifies the physical position of fault in the power system. This information is very useful for isolating the fault and restoring the power immediately. The circuit theory based method locates the fault using voltage and current values and impedance changes [11-13]. The traveling wave theory based method uses the information of voltage and current traveling waves for locating the fault [14]. Artificial Neural Networks (ANN) based method; Fuzzy Logic based method and Expert System based method belong to the class of artificial intelligence methods.

In this project faults generated signals characteristics are investigated, special faults detection, classification and location methods are developed, and their applications in power systems are presented. In the first chapter we have talked about smart grid a new technology in the electrical grid, we have mentioned its advantages and its solutions for all the problems faced in the classic electrical grid. We've briefly talked about the Phasor Measurement Unit and its Calculation of voltage and current phasors based on digital sampling of alternating current (AC) waveforms and a precise time signal provided by a GPS clock .We also introduced small description of the power system faults their causes, types, and their effects on the system .

II. POWER SYSTEM FAULTS

A fault is an event occurring on an electric system such as a short circuit, a broken wire, or an intermittent connection. In an electric power system, a fault is any abnormal electric current such as a short circuit [1]. There are mainly two types of faults in the electrical power system. Those are symmetrical and unsymmetrical faults.

978-1-4673-6673-1/15/$31.00 ©2015 IEEE

A. *Symmetrical faults*

These are very severe faults and occur infrequently in the power systems. These are also called as balanced faults and are of two types namely line to line to line to ground (L-L-L-G) and line to line to line (L-L-L) [1]. Only 2-5 percent of system faults are symmetrical faults. If these faults occur, system remains balanced but results in severe damage to the electrical power system equipments. Fig. 1 shows two types of three phase symmetrical faults. Analysis of these faults is easy and usually carried by per phase basis. Three phase fault analysis or information is required for selecting set-phase relays, rupturing capacity of the circuit breakers and rating of the protective switchgear.

B. *Non-symmetrical faults*

These are very common and less severe than symmetrical faults. There are mainly three types namely line to ground (L-G), line to line (L-L) and double line to ground (LL-G) faults [1]. Line to ground fault (L-G) is most common fault and 65-70 percent of faults are of this type. It causes the conductor to make contact with earth or ground. 15 to 20 percent of faults are double line to ground and causes the two conductors to make contact with ground. Line to line faults occur when two conductors make contact with each other mainly while swinging of lines due to winds and 5- 10 percent of the faults are of this type. These are also called unbalanced faults since their occurrence causes unbalance in the system. Unbalance of the system means that that impedance values are different in each phase causing unbalance current to flow in the phases. These are more difficult to analyze and are carried by per phase basis similar to three phase balanced faults.

III. FAULT DETECTION AND CLASSIFICATION TECHNQIUES

Fault detection and classification techniques used in electric power system is vital for secure operation of power systems and have to be accurate to facilitate quick repair of the system, the most useful techniques are mentioned below:

A. *Discrete Wavelet Transform*

Discrete Wavelet Transform is found to be useful in analyzing transient phenomenon such as that associated with faults on the transmission lines. This technique is well suited to wide band signals that may not be periodic and may contain both sinusoidal and non-sinusoidal components. The wavelet transform has the ability to focus on short time intervals for high frequency components and longtime intervals for low frequency components. Multi-Resolution Analysis (MRA) is one of the tools of Discrete Wavelet Transform (D.W.T), which decomposes original, typically non-stationary signal into low frequency signals called approximations and high frequency signals called details, with different levels or scales of resolution. It decomposes signal into different scales and resolutions. Detail coefficients contain information about the fault, which is required for fault detection. It uses filters with different cut off frequencies to analyze a signal at different resolutions. The signal is passed through a series of high-pass filters, also known as wavelet functions, to analyze the high frequencies and it is passed through a series of low-pass filters, also known as scaling functions, to analyze the low frequencies [15].

B. *Artificial Neural Networks*

Artificial neural networks (ANN's) are inspired by biological nervous systems and they were first introduced as early as 1960 [16]. A neural network is a massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use. Artificial Neural Networks simulate the natural systems behavior by means of the interconnection of basic processing units called neurons. Neurons are highly related with each other by means of links. The neurons can receive external signals or signals coming from other neurons. The output of the neuron is the result of applying a specific function. ANNs have a high degree of robustness and ability to learn. They are prepared to work with incomplete and unforeseen input data [4].

C. *Support Vector Machines*

The support vector machine (SVM) is a supervised learning method that generates input-output mapping functions from a set of labeled training data. The mapping function can be either a classification function, i.e. the category of the input data, or a regression function. For classification, nonlinear kernel functions are often used to transform input data to a high-dimensional feature space in which the input data become more separable compared to the original input space. Maximum-margin hyper planes are then created. The model thus produced depends on only a subset of the training data near the class boundaries. Similarly, the model produced by Support Vector Regression ignores any training data that is sufficiently close to the model prediction. SVMs are also said to belong to kernel methods.

Training an SVM involves solving a constrained quadratic programming problem, which requires large memory and enormous amounts of training time for large scale problems. In contrast, the SVM decision function is fully determined by a small subset of the training data, called support vectors. Therefore, it is desirable to remove from the training and test sets the data that is irrelevant to the final decision function. In the proposed technique the margin data sets are used for SVM training and test. Using margin data in SVM training and test makes the decision on the proposed method more accurate and efficient [17].

D. *The K-Nearest Neighbor Algorithm*

The purpose of the k Nearest Neighbors (KNN) algorithm is to use a database in which the data points are separated into several separate classes to predict the classification of a new sample point. Depending upon its threshold value as compared with test signals by standard deviation mathematical formulae we can detect and classify which type of fault will occur in transmission line system. The difference actual value and the test signal will give us the nearest threshold value about exact fault information. The proposed logic uses wavelet transform for extracting the hidden information in the current wave forms when a fault occurs, which is then suitably transformed to extract fault signatures and characterize the faults. The way in which the algorithm decides which of the points from the training set are similar enough to be considered when choosing the class to predict for a new observation is to pick the k closest data points to the new observation, and to take the most common class among these. This is why it is called k nearest neighbor's algorithm. [18]

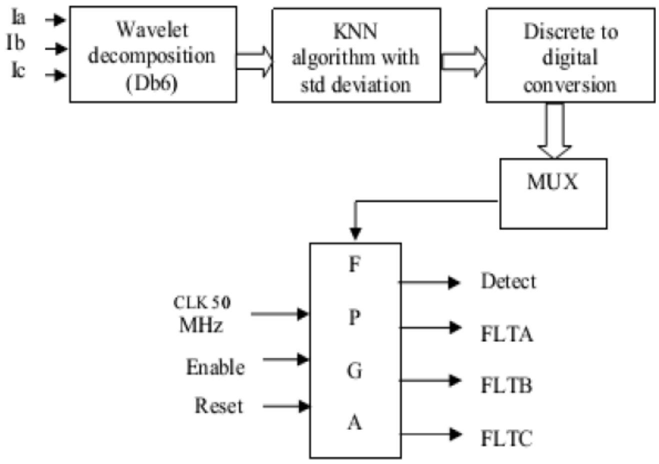

Fig. 1 Block diagram of fault classifier. [17]

E. Fault Location Techniques

Generally speaking, fault location methods can be classified into two basic groups, traveling wave based schemes and impedance measurement based ones as shown in Fig. 2. Traveling wave schemes can be used either with injecting a certain traveling wave from the locator position or with analyzing the generated transients due to the fault occurrence. Impedance measurement schemes are classified whether they depend on the data from one or both line ends. [19]

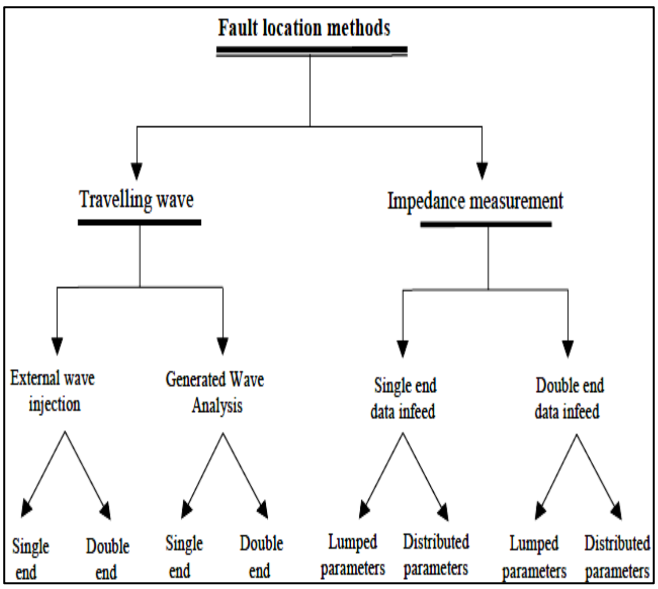

Fig. 2 Classification of fault location methods. [19]

a) Travelling Wave Based Fault Locators

Traveling wave phenomena for fault location is classified into different four types A, B, C and D according to their modes of operation using the traveling voltage waves. Types A and D depend on analyzing the resulting transients from the fault itself needing no further pulse generating circuitry. Type A is a single end one capturing the transients only at one end. It relies on the generated transients from the arcing flashover during the fault. However the assumption of getting generated transients at the line end is not always satisfied. Moreover, the arc itself may extinguish rapidly. They rely on measuring the required time for the injected pulses to go and to be captured after reflection from the fault point. This time can be directly interpreted as a fault distance. [2,20]

b) Impedance Measurement Based Fault Locators

These schemes provide another alternative for the fault location estimation problem. A line to ground fault occurred on phase A at point F through a resistance RF at a distance x from the locator position. The fault current IF is comprised from two components IFs and IFr flowing from sending and receiving ends respectively. The essential task of the fault location algorithm is to estimate the fault distance x as a function of the total line impedance ZL using the sending end measurements(for single end algorithms) or both end measurements(for double end algorithms)with the most possible accuracy. [21]

c) Non-Conventional Fault Locators

Instead of the normal mathematical derivation, non-conventional fault location algorithms were introduced depending on other processing platforms such as Wavelet Transform (WT), ANN or GA. These methods have their own problems that result from the line modeling accuracy, data availability and the method essences [3].

IV. SIMULATION RESULTS AND DISCUSSIONS

The single line diagram of the IEEE 14-bus system is shown in fig. 3. It consists of five synchronous machines, here of which are synchronous compensators used only for reactive power support. There are eleven loads in the system.

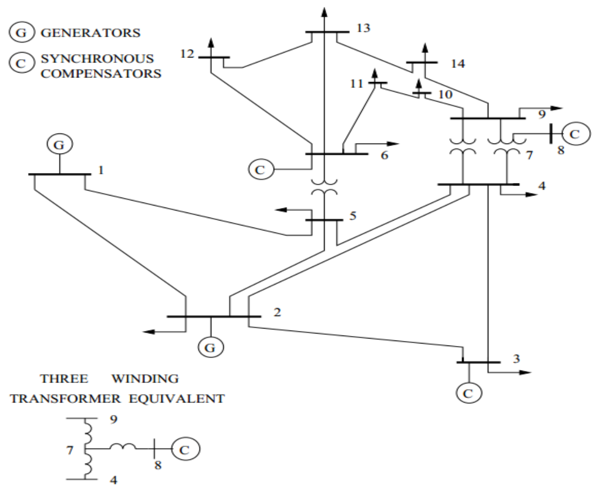

Fig. 3 Single line diagram of the IEEE 14-bus system

A. *The k-Nearest Neighbors simulation*

The *k*-Nearest Neighbors algorithm (or *k*-NN for short) is a non-parametric method used for classification and regression In both cases, the input consists of the *k* closest training examples in the feature space. In *k-NN classification*, the output is a class membership. An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its *k* nearest neighbors (*k* is a positive integer, typically small). If $k = 1$, then the object is simply assigned to the class of that single nearest neighbor. The neighbors are taken from a set of objects for which the class (for *k*-NN classification) or the object property value (for *k*-NN regression) is known. This can be thought of as the training set for the algorithm, though no explicit training step is required.In our simulation we used the code CLASS = KNNCLASSIFY (SAMPLE, TRAINING, GROUP) it classifies each row of the data in SAMPLE into one of the groups in TRAINING using the nearest-neighbors method. SAMPLE and TRAINING must be matrices with the same number of columns (in our case is 35322 columns). GROUP is a grouping variable for TRAINING. Its unique values define groups, and each element defines the group to which the corresponding row of TRAINING belongs. GROUP in our simulation is a numeric vector from 1 to 155. TRAINING and GROUP must have the same number of rows (155 for our case). CLASS indicates which group each row of SAMPLE has been assigned to, and is of the same type as GROUP.

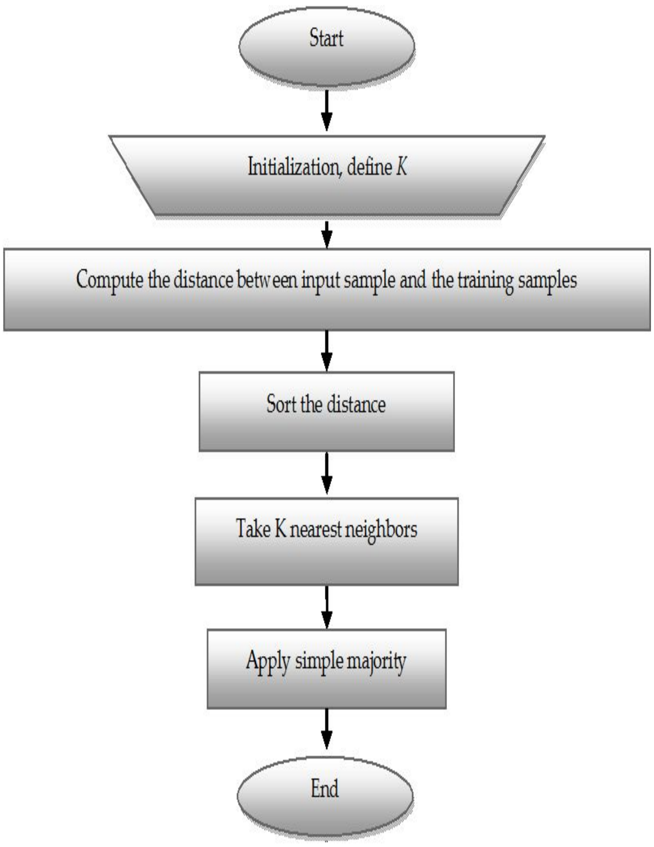

Fig. 4 Flowchart for the K- nearest neighbors

B. *Single line to ground fault*

A fault is created at bus 6. The simulation results of voltage at the point of fault are given in fig. 4. Att=0.018 seconds the current in phase A starts increasing from a peak value of 42.5 KA before the fault, it reaches a peak value of 540 KA after the fault. This is a very high value. The current flowing in the non-shorted phases are negligible in comparison with the short circuit current. This was predicted by using the equation: **$+I_b=0$ and $I_c=0$.**

C. *Double line to ground fault*

A fault is created at bus 6. The simulation results of voltage at the point of fault are given in figure 4. We can see that current flowing in phase C is negligible compared to that flowing in phase A and B when the fault starts, theoretically $I_c=0$. When we look at the voltage results it is clear that $V_a=V_b\approx 0$ and it is theoretically the

D. *Line to line fault*

This fault is created at bus 8. The simulation results of voltage at the point of fault are given in figure 5. We can easily note that $\sum(I_a + I_b) = 0$ and it is theoretically the same . The current flowing in phase C is negligible compared with phases A and B.As we can see in figure 3.10V_a and V_b are perfectly equal just at the beginning of the fault at time t=0.017s.

E. *Three phase balanced fault*

This type of fault is created at bus 8. The simulation results of voltage at the point of fault are given in figure 3.12. During a three phase balanced fault, the system remains balanced, though faulted. Theoretically $\sum(I_a + I_b + I_c)=0$ to verify this equation we pick up the current phases at different time and check the sum of them the table below shows the results.

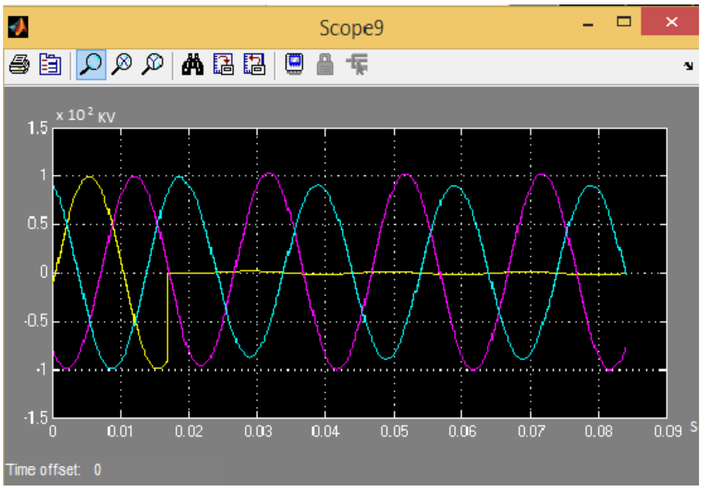

Fig. 5 The voltage for single line to ground fault

Table 1 Phase A, B and C currents at different times

T(s)	0.0273	0.0321	0.0545	0.0756
I_a (KA)	-3.755	-2.265	0.6	1.765
I_b (KA)	1.88	-2.273	-3.378	-3.535
I_c (KA)	1.88	4.545	2.77	1.765
$\sum(I_a + I_b + I_c)$	0.005	0.0069	-0.0079	-0.005

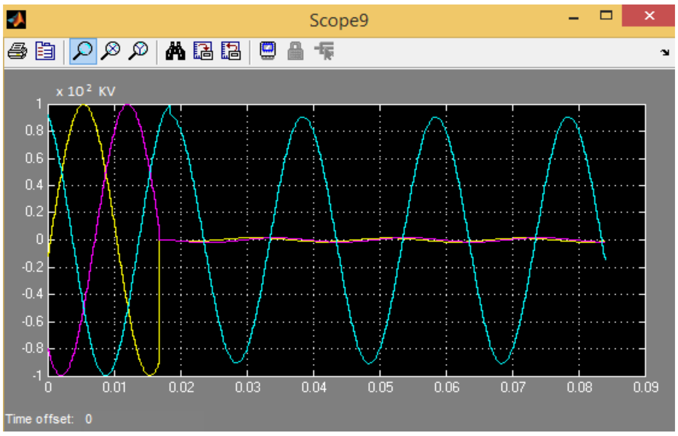

Fig. 6 The voltage for double line to ground fault

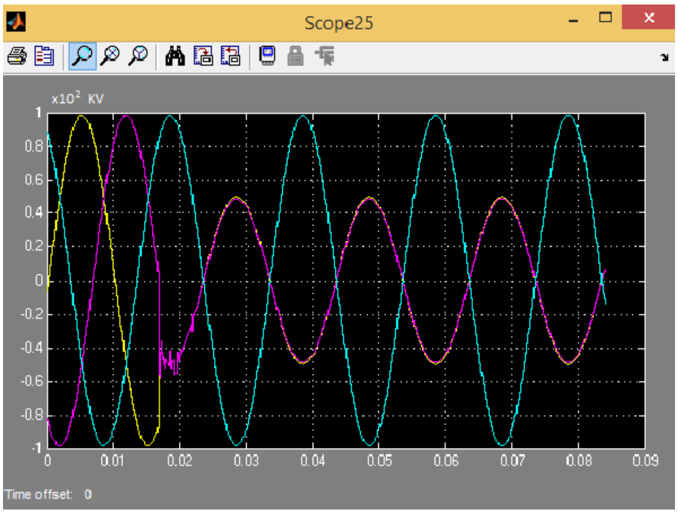

Fig. 7 The voltage for line to line fault

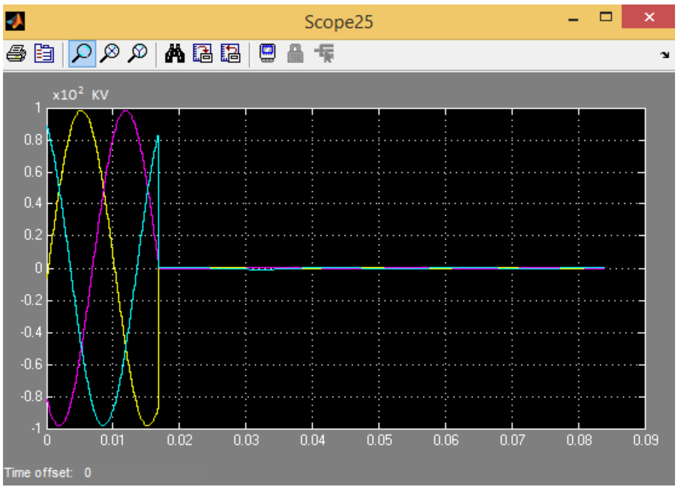

Fig. 8 the voltage for three phase balanced fault

F. Detection, classification and location simulation

First we need to transfer our data from the scope to workspace [scope-parameters-data history-save data to workspace]. Next step is to collect all the training data in one matrix including the normal operation data in the first row and the next eleven row is for the eleven faults data (AG,BG,CG,ABG,BCG,ACG,AB,BC,AC,ABC,ABCG) injected in the first bus, the next eleven row is for the same faults injected in the second bus, until the last eleven row which is for the same faults injected in the 14th bus ending with a 155 rows matrix. After collecting all the needed data we inject any type of fault in any busthen we put the obtained data in one row (test data). Finally we compare the row of the test data with all the rows of training data using KNN classifier or SVM classifier or any other classifier, the expected result is:

- If there is no fault injected, the classifier will output normal operation as in the first row of the training data.
- If there is a fault, the classifier will give us the most similar row from the training data matrix, as a result we will obtain the type and the location of the fault.

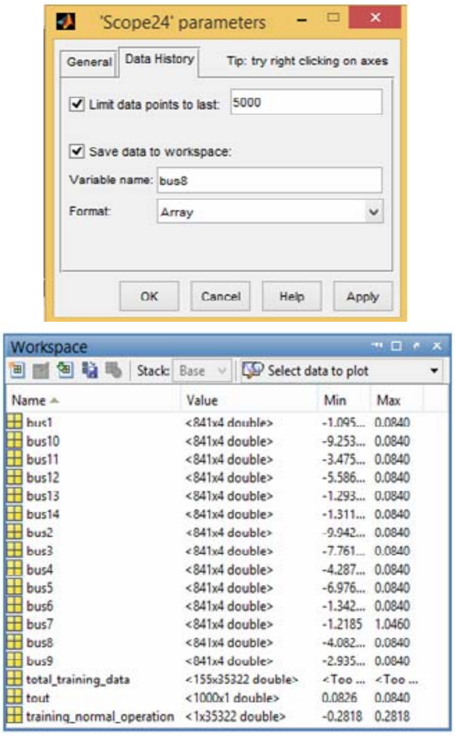

Fig. 9 Demonstration of the fault detection classification and location process

V. CONCLUSIONS

A KNN based fault classification for power system faults is simulated in this paper using MATLAB/SIMULINK. KNN-based technique is not based on measurements of voltages and currents as in conventional relays but rather is dependent on patterns of voltage and current. It is an adequate tool for the development of fault diagnosis system. KNN algorithm extracts the hidden information in the current waveforms when a fault occurs, which is then suitably transformed to extract fault signatures and characterize the faults. It can be also improved by using some condensing KNN techniques which are "The Condensed Nearest Neighbor Rule", "Reduced Nearest Neighbor rule", "Fast Condensed Nearest Neighbor rule", "Improved K-Nearest Neighbor Classification" and "The Class Boundary Preserving Algorithm" to reduce the training set to gain on speed and space efficiencies.

REFERENCES

[1] Ed Richard C. Dorf Boca Roton, The electrical engineering handbook, *CRC Press LLC*, 2000.

[2]. IEEE guide for determining fault location on ac transmission and distribution lines. IEEE Std C37.114-2004, 2005, pp: 1–36.

[3]. M.M Saha, J. Izykowski, E.Rosolowski, Fault location on power networks, Springer, 2010.

[4] Kejariwal, et al. Fault Detection and Location System for Power Transmission and Distribution Lines. The Research and Development Institute, Inc. at Montana State University. Patent 5,343,155. 30 August 1994

[5] Shimaa Hassan Sayed Barakat, "Fault detection, classification and location in Underground Cables", Master thesis, Faculty of Engineering - Fayoum University– Egypt, 2014.

[6] M. Jaya Bharata Reddy, D. Venkata Rajesh, D.K. Mohanta, "Robust transmission line fault classification using wavelet multi-resolution analysis.", Computers & Electrical Engineering, Vol. 39, Issue 4, pp:1219-1247. , May 2013

[7] Lei Ye, Dahai You, Xianggen Yin, Ke Wang, Junchun Wu,"An improved fault-location method for distribution system using wavelets and support vector regression", International Journal of Electrical Power & Energy Systems, Vol. 55, pp:467-472, February 2014.

[8]Ali Rafinia, Jamal Moshtagh, " A new approach to fault location in three-phase underground distribution system using combination of wavelet analysis with ANN and FLS ", International Journal of Electrical Power & Energy Systems, Vol. 55, pp: 261-274 , February 2014.

[9] Tamer S. Kamel M. A. Moustafa Hassan, "Adaptive Neuro Fuzzy Inference System (ANFIS) For Fault Classification in the Transmission Lines", the Online Journal on Electronics and Electrical Engineering (OJEEE), Vol. 2, No. 1, January 2010.

[10] Das, R., Sachdev, M.S., Sidhu, T.S., "A fault locator for radial subtransmission and distribution lines", IEEE Power Engineering Society Summer Meeting, Vol.1, pp: 443-448, 2000.

[11] M.Mirzaei, M.Z. A Ab Kadir, E.Moazami, H.Hizam, "Review of fault location methods for distribution power system", Australian Journal of Basic and Applied Sciences, 3(3), page(s): 2670-2676, 2009.

[12] Myeon-Song Choi, Seung-Jae Lee, Duck-Su Lee, Bo-Gun Jin, "A new fault location algorithm using direct circuit analysis for distribution systems", IEEE Transactions on Power Delivery, Vol.19, No.1, Page(s):35-41, 2004.

[13] A.H.A. Bakar, M.S. Ali, ChiaKwang Tan, H. Mokhlis, H. Arof, H.A. Illias, " High impedance fault location in 11 kV underground distribution systems using wavelet transforms", International Journal of Electrical Power & Energy Systems, Vol. 55, Pages 723-730, February 2014.

[14] Thomas, D.W.P., Christopoulos, C., Carvalho, R.J.O., Pereira, E.T., "Single and double ended travelling-wave fault location on a MV system", Eighth IEE International Conference on Developments in Power System Protection, Vol.1, Page(s): 200-203, 2004.

[15] Valsan SP, Swarup KS. Fault detection and classification logic for transmission line using multi resolution wavelet analysis. Electric Power Comp Syst 2009; 36(4):321–44.

[16] S. Haykin, Neural Networks, IEEE Press, New York, 1994.

[17] P.Srinivasa Rao et al, "Pattern Recognition Approach for Fault Identification in Power Transmission Lines" Int. Journal of Engineering Research and Application ISSN: 2248-9622, Vol. 3, Issue 5, Sep-Oct 2013, pp.1051-1056.

[18] Mohammad Abdul Baseer. Travelling Waves for Finding the Fault Location in Transmission Lines, Journal Electrical and Electronic Engineering. Vol. 1, No. 1, 2013.

[19] A.A. Girgis, C.M. Fallon, D.L. Lubkeman, A fault location technique for rural distribution feeders. IEEE Trans. Ind. Appl. 29(6), 1170–1175,1993.

[20]. H.W. Dommel, J.M. Michels, High speed relaying using traveling wave transient analysis, IEEE PES Winter Power Meeting (1978)

[21]. R.H. Salim, K.C.O. Salim, A.S. Bretas, Further improvements on impedance-based fault location for power distribution systems. IET Gener. Trans. Distrib. 5(4), 467–478 (2011)