Accurate Fault Classification of Transmission Line Using Wavelet Transform and Probabilistic Neural Network

M. Mollanezhad Heydar-Abadi* and A. Akbari Foroud*

Abstract: Fault classification in distance protection of transmission lines, with considering the wide variation in the fault operating conditions, has been very challenging task. This paper presents a Probabilistic Neural Network (PNN) and new feature selection technique for fault classification in transmission lines. Initially, wavelet transform is used for feature extraction from half cycle of post-fault three phase currents at one end of line. In the proposed method three classifiers corresponding with three phases are used which fed by normalized particular features as Wavelet Energy Ratio (WER) and Ground Index (GI). The PNNs are trained to provide faulted phase selection in different ten fault types. Finally, logic outputs of classifiers and GI identify the fault type. The feasibility of the proposed algorithm is tested on transmission line using PSCAD/EMTDC software. Variation of operating conditions in train cases is limited, but it is wide for test cases. Also, quantity of the test data sets is larger than the train data sets. The results indicate that the proposed technique is high speed, accurate and robust for a wide variation in operating conditions and noisy environments.

Keywords: Distance Protection, Fault Classification, Neural Network, Probabilistic, Wavelet Energy Ratio.

1 Introduction

Basically, fast detecting, isolating, locating and repairing of the different faults are critical in maintaining a reliable power system operation [1]. On the other hand, classification of the different fault types plays very important role in digital distance protection of the transmission lines [2]. During fault, the amplitude and frequency of the current waveforms change from normal state to transient state. Due to the electromagnetic coupling between three phase lines of the transmission line, the transient disturbance is generated in the healthy phase currents. Therefore, distinguishing of the faulted phase and identifying of the fault type are complex task. The oscillation of the three phase currents in different faults are affected by operating conditions such as, Source impedance (Z_s), power angle (δ), fault inception angle (FIA), fault location (L_f) and fault resistance (R_f). In the case of the EHV transmission line fault study, the response of the C-phase current to c-g (faulty phase) and ab-g (healthy phase) for two different operating conditions has been shown in Fig. 1. Thus, fault classification in digital distance protection, with considering the wide variation in the fault conditions, has been very challenging task and modern pattern recognition techniques can used for this task.

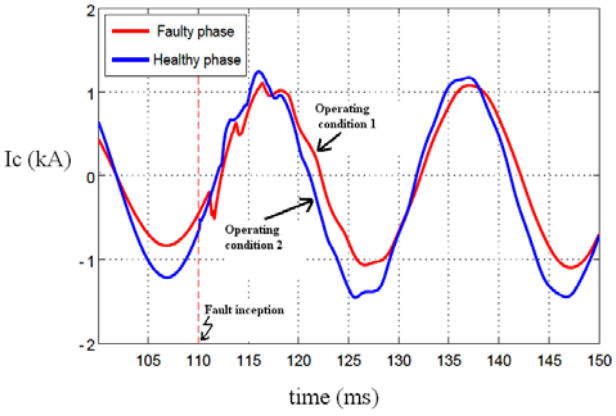

Fig. 1 Phase current at two different operating conditions. Operating condition 1: $\delta = 20^o$, $L_f = 0.95^{p.u}$, $R_f = 200\ \Omega$. Operating condition 2: $\delta = 30^o$, $L_f = 0.05^{p.u}$, $R_f = 0\ \Omega$.

Iranian Journal of Electrical & Electronic Engineering, 2013.
Paper first received 22 Mar. 2013 and in revised form 6 May 2013.
* The Authors are with the Department of Electrical Engineering, Semnan University, Semnan.
E-mails: mollanezhad_mehrdad@yahoo.com and aakbari@semnan.ac.ir.

The increasing complexities of the modern power transmission systems have greatly raised the importance of the fault classification research studies in recent years. The fault classification techniques for designing digital distance relay are used in many researches. Transmission line fault classification techniques can be classified into some main categories: wavelet transform [3, 4], fuzzy logic system [5, 6], neural network [7-10] and support vector machine (SVM) [11]. Chanda et al. [3] have proposed an algorithm with Daubechies eight (Db8) wavelet for fault classification in transmission line using 3rd level output of multi-resolution analysis (MRA) detail signals of currents. But in that algorithm the phases involved in the fault does not explicitly determined and time duration considered in the analysis comes out to 40.96 ms, which is about two cycles after post fault. Das et al. [5] have presented a fault classification technique based on fuzzy logic for ten types of fault in transmission line on half cycle post-fault current sample, but that approach is valid for limited variable conditions.

Silva et al. [9] have proposed wavelet energy and artificial neural network (ANN) for fault detection and fault classification, respectively. But validation of their method has been tested just by same fault inception angle variables (FIA=60, 150 for both training and test data sets).While FIA is an effective variable in fault current analysis. Also, that approach uses both post-fault voltage and current samples. Samantaray et al. [10] have presented a technique based on combination of S-transform and PNN for fault classification in thyristor controlled series compensated (TCSC) transmission line. In this method, features extracted by S transform are used as inputs to PNN for fault classification. Moreover, in that work, 300 data sets have been used for training whereas only 200 data sets have been used for testing the performance of the proposed technique. Parikh et al. [11] have proposed SVM technique for fault classification in series compensated transmission line. Classification accuracy has been reported over 98% for 25,200 test cases. But validation of the proposed approach investigated for limited variation in fault resistance (Rf) up to 50Ω and variation in fault distance (Lf) up to 0.8 (in p.u. of line length).

Recently, Upendar et al. [12, 13] have presented an algorithm based on the wavelet transform of three phase currents and the PNN and the classification and regression tree (CART) methods. The accuracy of the fault classification method has been reported over 99% upon 1,209,960 test cases. But, the structure of the studied power system has been composed of the one generator and one fixed load with the transmission between them. Thus, the authors have considered only three operating condition variables (R_f, L_f and FIA). Also, the fault identification algorithm is low speed. Because, the 7rd level output of MRA detail signals of two cycles post fault currents have been used for feature extraction technique.

The previous fault classification techniques [3-12] haven't considered wide variation of all five operating conditions for test cases while training cases included limited variation of operating conditions. Also, some of the above schemes are affected by noise signal. However, in this work, an improved approach has been proposed for solving above drawbacks. For example, our proposed technique is tested for fault resistance up to 200 Ω, fault location up to 0.95 of line length, fault inception angle up to 160°, power angle up to 40° and source impedance up to125% of Z_s, while the algorithm is trained for fault resistance up to 100 Ω, fault location up to 0.8 of line length, fault inception angle up to 90°, power angle up to 30° and source impedance 100% of Z_s. In addition, the accuracy of this technique is investigated in noisy environments.

PNN classifier and Wavelet Energy (WE-PNN) combination technique has been proposed for transient signal classification [14, 15].The wavelet transform is a powerful tool in the analysis of the transient phenomena in power system due to its ability to extract information from the transient signals simultaneously in both of time and frequency domain, rather than conventional Fourier transform which can only give the information in the frequency domain. ANNs have some drawbacks, including the determination of network architecture and network parameters assignment. When networks are applied in dynamic environments, especially for online applications, for example protection purpose, traditional networks can become the bottleneck in adaptive applications. Considering these limitations, the PNN is suitable for faults identification. The PNN classifiers are recognized as having expandable or reducible network structure, fast learning speed, and promising results. PNN can function as a classifier, and it has the advantage of a fast learning process, requiring only a single-pass network training stage without any iteration for adjusting weights, and it can adapt itself to architectural changes.

In this paper, a new fault classification scheme for transmission line is presented based on combination of Wavelet Energy Ratio and Probabilistic Neural Network (WER-PNN). Three PNNs and one ground detector are proposed for fault type identification. Faulted phase and ground fault are determined by PNN classifiers and ground detector, respectively. Three inputs are used to train each classifier in order to classify the faulty phase and healthy phase. For selecting the mentioned features, initially, wavelet transform is used for feature extraction from half cycle post-fault three phase currents at the one of two ends (relay position). The 4rd level wavelet Multi-Resolution Analysis (MRA) detail and approximation signals are found to be most suitable and are used for the analysis. The Wavelet Energy Ratio (WER) and Ground Index (GI) are proposed as features for classifiers. WER is equal to ratio of wavelet coefficient energy at each phase to summation wavelet energy coefficients of all phases. The GI is determined

based on comparison of the maximum absolute of ground current with a threshold value. Thus, the proposed feature selection method is a new technique. Each individual classifier with such particular three features (two WERs and one GI) is able to recognize the healthy phase from faulted phase, correctly. Finally, identification of ten types of fault is based on logic output of three PNN classifiers and ground detector. The proposed method is tested under different fault conditions such as different fault locations, different fault inception angles, different fault resistances, different load angles and different sources impedances. The investigation results confirm the validity and high accuracy of the proposed technique.

The most important advantages of the proposed method are as follows:

- Half cycle post-fault currents are required at one of the two ends of the transmission line.
- Low dimension of classifier normalized input vector (three features 0~1) is required, thus the memory requirement and computation time will be reduced.
- Training and testing process of PNN classifier are fast.
- The PNN classifier needs to determine only one parameter (smoothing parameter).
- The proposed method is applicable for a wider variation in the operating conditions in testing stage rather than training process.

For evaluation of proposed algorithm, the extra high voltage (EHV) transmission line has been simulated in PSCAD/EMTDC software [16]. It is powerful software that generates transient signals. Classification operation has been carried out in MATLAB environment [17]. The structure of this paper is as follows:

In Section 2, the proposed technique is explained. Section 3 describes simulation consideration. Section 4 performance results in details. Section 5 includes further studies such as, results comparison and noise effect. Finally, the paper closes with conclusions in Section 6.

2 WER-PNN Technique for Fault Classification

2.1 Architecture of the Proposed Fault Classification Algorithm

The structure of the proposed fault classification system is shown in Fig. 2. The algorithm consists of three stages, including feature extraction, feature selection and fault classification. The system takes half cycle of the post-fault from the three phase currents at the current recorder (relay location).The ground current is calculated from the samples of the three line currents ($I_g=I_a+I_b+I_c$). In this algorithm, three PNNs and one ground detector (GD) have been used for fault classification method. Each of the three PNNs (PNN_a, PNN_b, PNN_c) is used to identify the faulted phase(s) and the ground detector (GD) is used to determine the involvement of the ground in the fault. At the output of each PNN, the value '1' and '0' denotes the presence or absence of the fault, respectively**.** Table 1 shows the fault classification format in the proposed method while Fig. 2 depicts the overall fault classification scheme.

Table 1 Fault classification format.

Sr. no.	Output of PNN_a for phase Ia	Output of PNN_b for phase Ib	Output of PNN_c for phase Ic	Output of GI for ground	Type of fault
1	1	0	0	1	a-g
2	0	1	0	1	b-g
3	0	0	1	1	c-g
4	1	1	0	1	ab-g
5	1	0	1	1	ca-g
6	0	1	1	1	bc-g
7	1	1	0	0	ab
8	1	0	1	0	ca
9	0	1	1	0	bc
10	1	1	1	0	abc

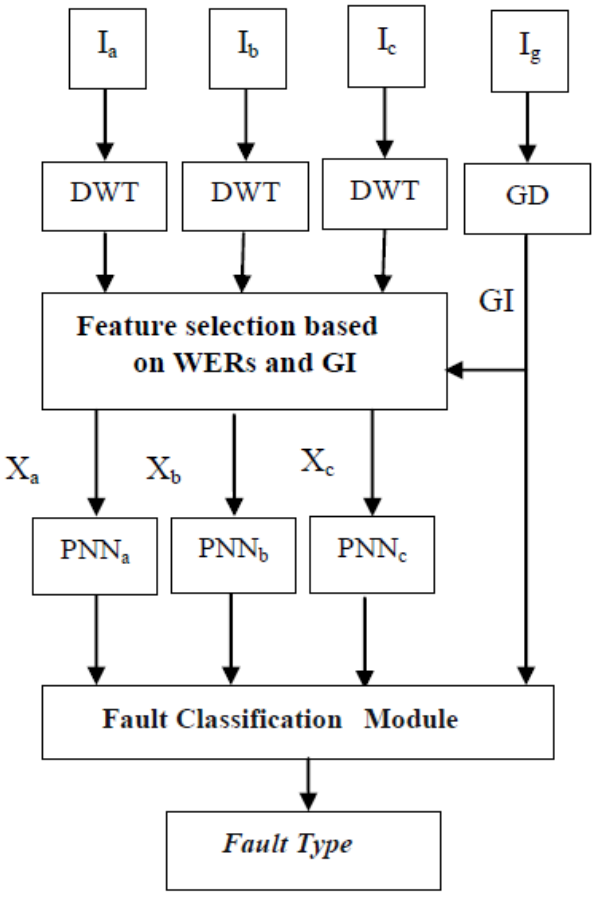

Fig. 2 Structure of the fault classifier.

2.2 New Feature Selection from the DWT for Fault Classification

A wavelet transform is a powerful tool for feature extraction of the transient signals. The transmission line faults are common transient phenomena in power system. Faults are low amplitude, short duration, fast decaying and oscillating type of high frequency current signals. There are many types of mother wavelets, such as Harr, Daubichies, Coiflet and Symmlet wavelets. The choice of the mother wavelet plays an important role in the feature extraction from transient signals. One of the most popular mother wavelets suitable for a wide range of applications is Daubichies's wavelet [14].

The wavelet technique has been applied in many literatures for feature extraction of transient fault signals in power system. The differentiations between these methods are: different frequency sampling rates, different mother wavelet types, different number of decomposition levels and state of calculating the energy or entropy features. Most of the previous studies have used the detail coefficients as feature vectors in the fault transient analysis [3, 12-13]. The above works find limitations arisen of highly susceptible of wavelet detail coefficients to noise so that provides erroneous results even with noise of SNR 20 dB. Also, time duration considered in those methods comes out to two cycles after post fault. However, these algorithms have low speed. In our work, a new feature selection using wavelet transform has been proposed for solving above drawbacks.

In the proposed paper, we introduced a new feature selection technique based wavelet energy ratio (WER). The WER is normalized feature (0 ~1) and defined as a ratio of corresponding phase energy to sum of the three phase energies. Half cycle from post fault current samples are required in our algorithm, hence our method is fast. In other word, due to using of 4th level approximation coefficients, our technique provides successful results even with noise of SNR 30 dB and 20 dB. Detail explanation of the proposed feature selection method is presented in below.

In this work, the db8 is used as mother wavelet transform, particularly. Fig. 3 illustrates the implementation procedure of a Discrete Wavelet Transform (DWT), in which X(n) is the original signal, h(n) and g(n) are high-pass and low-pass filters, respectively. At the first stage, an original signal is divided into two halves of the frequency bandwidth, and sent to both high-pass filter and low-pass filter.

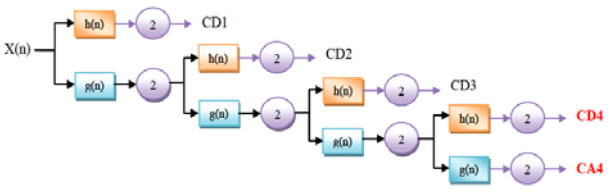

Fig. 3 Implementation of DWT.

Then the output of low-pass filter is further cut in half of the frequency bandwidth, and sent to the second stage. This procedure is repeated until the signal is decomposed to a pre-defined certain level. The set of attained signals represent the same original signal, but all corresponding to different frequency bands. Sampling frequency in this work is 20 kHz. Therefore frequency domain for detail coefficients (CD1, CD2, CD3, CD4) are 10~20 kHz, 5~10 kHz, 2.5~5 kHz, 1.25~2.5 kHz, respectively, and for approximation coefficient (CA4) is 0~1.25 kHz.

The robustness against noise effect is one of the proposed algorithm purposes. Thus, for feature selection task, considering the noise effect is essential. The cases of pure signal fault current and added noisy signal are shown in Figs. 4 and 5. With comparison to pure signal, it is found that noisy signal has little effect on fourth level of wavelet coefficients (CD4, CA4). Also, In order to reduce the effect of the dc offset in the fault current waveform, the first level detail coefficient (CD1) is neglected. Also, it is observed that, the detail coefficients at second and third levels (CD2 and CD3) are affected by signal noise. However, mentioned coefficients (CD1, CD2 and CD3) are neglected and CD4 and CA4 are selected as suitable features for feature selection analysis.

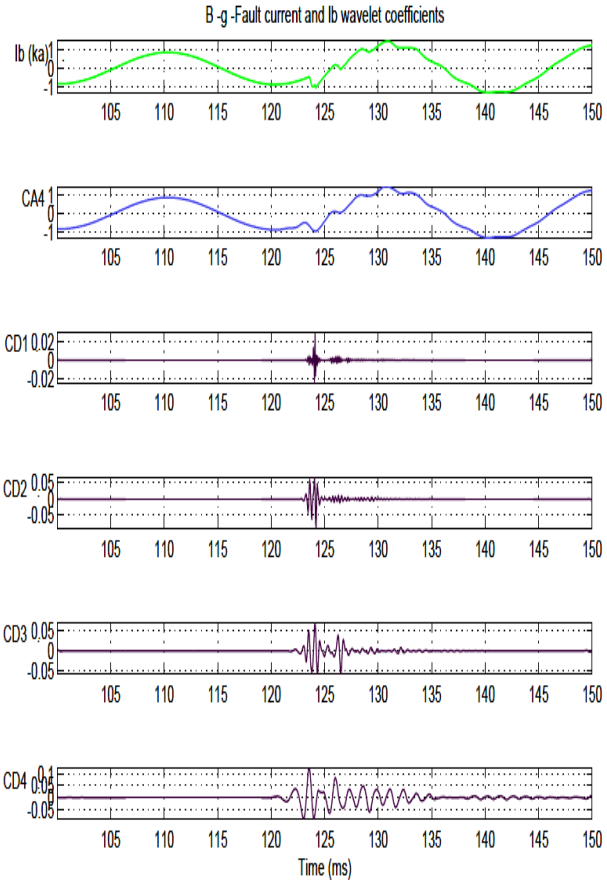

Fig. 4 Fault current and wavelet coefficients for pure signal.

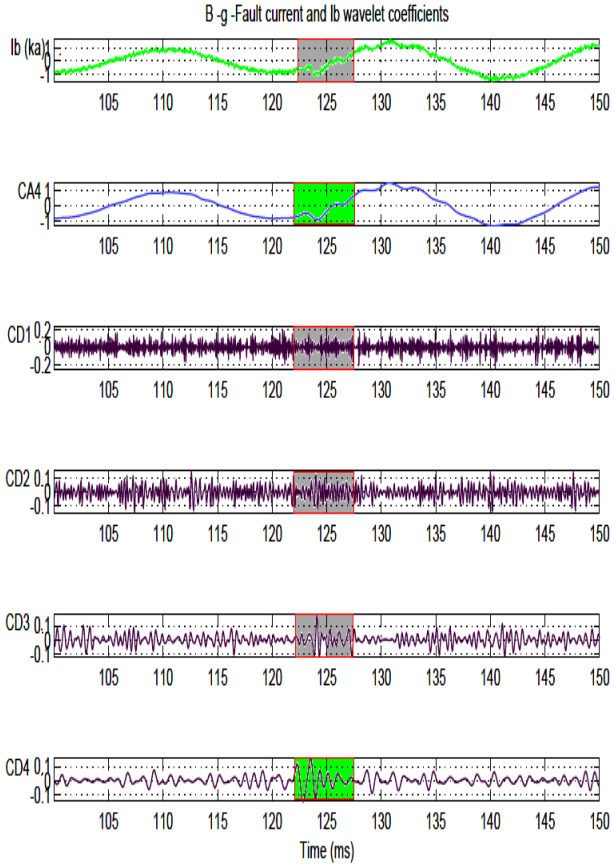

Fig. 5 Fault current and wavelet coefficients for signal with 20 dB noise.

The energy of the approximate wavelet coefficient (EA4) and the energy of the detailed wavelet coefficient (ED4) are calculated by using Eqs. (1) and (2), respectively:

$$ED4 = \sum_{i=1}^{N} |CD4_i|^2 \quad (1)$$

$$EA4 = \sum_{i=1}^{N} |CA4_i|^2 \quad (2)$$

Here, N denotes the length of the window in terms of the number of sample points. In this work all current waveforms are sampled in a frequency of 20 kHz and the window length considered for energy calculation is 10 ms post fault inception time. As mentioned above, the energy of the three currents ($ED4_a$, $EA4_a$, $ED4_b$, $EA4_b$, $ED4_c$, $EA4_c$) are calculated.

Proposed fault classification scheme involves three classifiers for faulted phase selection in various fault types. The classifiers (PNNa, PNNb, PNNc) are fed by feature vectors as X_a, X_b and X_c, respectively. The feature vectors are three dimension vectors and presented below:

$$\begin{aligned} X_a &= [\,WER_{1,a}\ WER_{2,a}\ GI] \\ X_b &= [\,WER_{1,b}\ WER_{2,b}\ GI] \\ X_c &= [\,WER_{1,c}\ WER_{2,c}\ GI] \end{aligned} \quad (3)$$

In characteristic features, first two features are WER for detailed and approximate coefficient in fourth level of MRA ($ED4_n$, $EA4_n$). The WER is ratio of wavelet coefficient energy in each phase to summation wavelet coefficient energies of all phases. WERs have been computed as:

$$WER_{1,k} = ED4_{n,k} = \frac{ED4_k}{ED4_t},\ WER_{2,k} = EA4_{n,k} = \frac{EA4_k}{EA4_t}$$

$$\text{for } k = a, b, c \quad (4)$$

where:

$$\begin{aligned} ED4_t &= ED4_a + ED4_b + ED4_c \\ EA4_t &= EA4_a + EA4_b + EA4_c \end{aligned} \quad (5)$$

The third feature is GI that it is same for three classifiers in each case and calculated based on maximum absolute of ground current in half cycle after post-fault. This technique has been applied for detection ground faults in [6]. The GI is calculated as follows:

$$\begin{aligned} &\text{If } \max(\text{abs}(Ig = Ia + Ib + Ic)) > 0.001(kA), GI = 1 \\ &\text{Else } GI = 0 \end{aligned} \quad (6)$$

2.3 PNN for Automatic Faulted Phase Selection

The PNN at first proposed by Donald Specht in 1990 [18]. This is an Artificial Neural Network (ANN) for nonlinear computing which approaches the Bayes optimal decision boundaries. The ANNs, which have gained prominence in the area of pattern recognition, have several properties that make those attractive for transient signal recognition. These include a relatively simple implementation, inherently parallel algorithm (making parallel implementation for natural progression), robustness to noise and self-learning ability. The PNN is a 3-layer, feed-forward, one-pass training algorithm used for classification and mapping of data. Unlike other ANNs, like the back-propagation neural network, it is based on well-established statistical principles derived from Bayes' decision strategy and non-parametric Kernel based estimators of probability density functions. Advantages of the PNNs is that it is easy to add new categories, or new training inputs, into the already running structure, which is good for the on-line applications.

The PNN operates using spherical Gaussian radial basis functions centered at each training vector. Fig. 6 shows the architecture of a PNN model that is composed of the radial basis layer and the competitive layer. In the signal classification application, the training examples are classified according to their distribution values of probabilistic density function (PDF), which is the basic principle of the PNN. A simple PDF is defined as follow:

$$f_k(X) = \frac{1}{N_k} \sum_{j=1}^{N_k} \exp\left(-\frac{\|X - X_{kj}\|}{2\delta^2}\right) \quad (7)$$

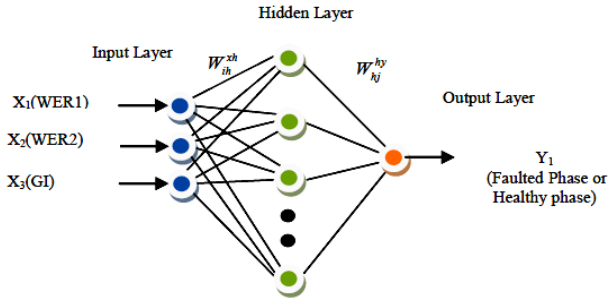

Fig. 6 Architecture of the proposed PNN.

Modifying and applying Eq. (8) to the output vector H of the hidden layer in the PNN is as below:

$$H_h = \exp\left(\frac{-\sum_i (X_i - W_{ih}^{xh})^2}{2\delta^2}\right) \quad (8)$$

The algorithm of the inference output vector H in the PNN is as follows:

$$net_j = \frac{1}{N_j}\sum_h W_{hj}^{hy} \times H_h .. and .. N_j = \sum_h W_{hj}^{hy}, \quad (9)$$

$$If .. net_j = \max_k(net_k) .. Then .. Y_j = 1 .. Else .. Y_j = 0$$

where i is number of input layers, h is number of hidden layers; j is number of output layers; k is number of training examples; N_k is number of classifications (clusters); δ is smoothing parameter (standard deviation), $0.1 < \delta < 1$, The choice of the δ parameter has an effect on PNN's classification accuracy. In this paper, particular classifier parameters have been appropriately chosen. Smoothing parameters only determined in loading the training data set, experimentally. Also, $\|X - X_{kj}\|$ is Euclidean distance between the vectors X and X_{kj}; i.e., $\|X - X_{kj}\|$; $\|X - X_{kj}\| = \sum_i (X_i - X_{kj})^2$; W_{ih}^{xh} connection weight between the input layer X and the hidden layer H; W_{hi}^{hy} is the connection weight between the hidden layer H and the output layer Y.

3 Case Studies

3.1 Transmission Line Model

A transmission line model has been simulated using PSCAD/EMTDC software [16]. Fig. 7 shows the simulated network. For each and every fault case, the duration of the fault has been assumed to be 0.1 s (5 cycles). The 400 kV, 50Hz EHV power system consists of two sources connected by transmission line is used for the distance protection study. The location of the relay is bus no. 1.

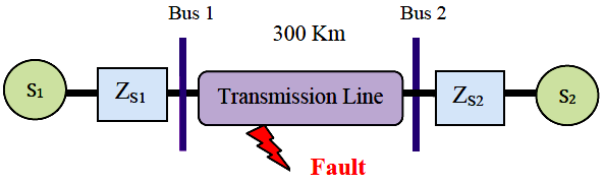

Fig. 7 Transmission line.

The transmission line has been represented by the 'Bergeron' line model, which is recommended for relay studies [5]. The parameters of the power system model (Fig. 7) are as follows [5]:

- line length = 300 km
- source voltages:
  - source 1: v1 = $400\angle\delta$ kV
  - source 2: v2 = 400 kV
    where δ is the load angle.
- source impedance (both sources):
  - positive sequence impedance = 1.31 + j15.0 Ω
  - zero sequence impedance = 2.33 + j26.6 Ω
- frequency = 50 Hz
- transmission line impedance:
  - positive sequence impedance = 8.25 + j94.5 Ω
  - zero sequence impedance = 82.5 + j308 Ω
  - positive sequence capacitance = 13 nF/km
  - zero sequence capacitance = 8.5 nF/km

3.2 Generated Different Train and Test Data Sets

To verify the robustness of the proposed algorithm, extensive fault simulations under various conditions are performed. To illustrate the performance of the proposed method, two faults simulations are studied. First simulation is performed under limited operating conditions and the generated data sets are used to train the classifiers. But second simulation is carried out under wide variation of operating conditions and the generated data sets are used to test the classifiers. There are only very few previous studies that address the above issue of classification fault in transmission line.

In Table 2, all five simulation variables including source impedance (Z_s), power angle (δ), fault inception angle (FIA), fault location (L_f) and fault resistance (R_f) are presented. As shown in Table 2, the training cases include limited variation of variables, while the test cases involve wide variation of variables. For example, for train cases, fault resistances and fault locations are set up to100 Ω and $0.8^{p.u}$ respectively. But, the test cases include fault resistances and fault locations up to 200 Ω and $0.95^{p.u}$ respectively. In the training case, value of two impedance source is equal 100% (impedance equal to BSLV). But the different five combinations of Z_{s1} and Z_{s2} values have been investigated for test cases are shown in Table 3. For each combination of different variables, all the ten types of faults have been considered.

Table 2 Simulation variables data set used in digital simulation.

Simulation variables	Training	Test
Source impedance (of base source impedance value)	100	75-100- 125
Power angle (δ)	10 –30	20-40
Fault inception angle (°)	0 – 90	45-160
Fault location (percent of line)	0.05 – 0.8	0.5 – 0.95
Fault resistance (Ω)	5 – 50 – 100	25 -75 -200
Fault type	a-g , b-g , c-g, ab-g, ca-g bc-g , ab, ca, bc, abc	
Total cases	Train cases: 1×2×2×2×3×10=240 Test cases: 4×2×2×2×3×10=1200 Total cases: 240+1200=1440	

Table 3 Combinations of Z_S for test data.

Case no.	Z_{S1} (%)	Z_{S1} (%)
1	100	100
2	100	75
3	75	100
4	100	125
5	125	100

A number of 240 training cases and 1200 test cases are generated by simulation of transmission line. Thus, 1440 simulation cases are generated for all system conditions. All the fault simulation studies are carried out using PSCAD/EMTDC software [16], which is an extremely versatile, industry standard simulation tool used for studying the transient behavior of electric power networks. The wavelet toolbox in MATLAB is used to perform the wavelet transformation to extract the feature vectors. All three classifiers are trained offline using the feature vectors generated using the training vector (3×240). The PNN classifiers have been implemented in the MALAB environment.

The performance of the trained PNN is evaluated by the testing pattern sets that are different from training pattern sets. The size of testing pattern sets are1200 patterns for each classifier. Therefore, the total size of classifier input vector testing stage is 3 × 1200 (three features per phase and 1200 patterns). After the PNNs are trained, their performance is tested with the test vectors. The resultant output of individual PNN denotes whether the corresponding phase is involved in the fault or not. Subsequently, the accuracy of the fault classification algorithm is computed as:

$$\eta = \frac{\text{Number of correct fault classification cases}}{\text{Total number of test cases}} \times 100 \quad (10)$$

4 Results and Discussion

After executing the training procedure, the PNNs are tested using simulated fault patterns not presented during the training process. To test the effectiveness of the proposed scheme, a number of simulation studies have been carried out on the test system. The train and test data sets are composed of over 1440 cases including different fault resistances, fault inception angles, source impedances, load angles and fault distances. The feature vectors X_a, X_b and X_c are extracted and fed to the each classifier.

As an example, values of feature vectors for ten types of fault under limited variation of operating condition (Z_{s1} = Z_{s2}=100%, δ = 10°, FIA = 0°, L_f = $0.05^{p.u}$, R_f = 5Ω) are shown in Table 4. Also, the values of feature vectors for ten types of fault under wide variation of operating condition (Z_{s1} =125%, Z_{s2} =100%, δ = 40°, FIA = 160°, L_f = $0.95^{p.u}$, R_f = 200Ω) are shown in Table 5. As seen from the Tables 4 and 5, the values of the feature 2 of three feature vectors is higher for the faulty phase in the fault process compared to healthy phase. For example, in a-g fault shown in Table 4, the feature 2 is 0.9949. But for b-phase and c-phase the values are 0.0041 and 0.0010, respectively. Considering the numerical results of Tables 4 and 5, it seems that the feature1 for faulted phase in comparison with healthy phase has untidy behavior. But in section 5.2 (Comparison with other classification scheme), it is illustrated that the behavior of first feature depends on the value of the feature 2 for two classes and this relationship is nonlinear and tidy. Lead to nonlinear behavior of the features, the PNN classifier is used as artificial intellect in this paper.

It has been observed that the values of max(abs(Ig)) are high (greater than 100A) for faults involving ground and low (less than 1A) for faults not involving ground [6]. Tables 4 and 5 confirm that the involvement of ground in a fault can be easily detected based on the value of max(abs(Ig)).

After training each classifier with their particular training feature vectors (3×240), individually, proposed algorithm is tested with test feature vectors (3×1200). Upon testing of test cases, an overall fault classification accuracy of 99.15% has been obtained by the proposed algorithm. Table 6 depicts the performance of the proposed fault classification technique for different types of faults. It is observed that the proposed technique gives highly accurate results for all types of faults. In Table 6, the proposed algorithm accuracy for each type fault and total types is indicated. It is observed that the ground faults have the lowest accuracy of classification among fault categories.

Table 4 Feature vectors for different faults under limited variation of operating conditions.

Fault Type		X_a for PNN_a	X_b for PNN_b	X_c for PNN_c	Max(abs(Ig)) (kA)
L - g	a-g	[0.1633 - 0.9949 - 1]	[0.4153 - 0.0041 - 1]	[0.4215 - 0.0010 - 1]	9.9113
	b-g	[0.4554 - 0.0031 - 1]	[0.0893 - 0.9890 - 1]	[0.4553 - 0.0080 - 1]	7.5481
	c-g	[0.4442 - 0.0115 - 1]	[0.4438 - 0.0074 - 1]	[0.1120 - 0.9811 - 1]	7.6370
L – L - g	ab-g	[0.0523 - 0.5805 - 1]	[0.1310 - 0.4187 - 1]	[0.8167 - 0.0008 - 1]	5.8145
	ca-g	[0.0332 - 0.6736 - 1]	[0.8387 - 0.0015 - 1]	[0.1282 - 0.3250 - 1]	5.6356
	bc-g	[0.0560 - 0.0021 - 1]	[0.4587 - 0.6841 - 1]	[0.4853 - 0.3138 - 1]	7.4037
L – L	ab	[0.5000 - 0.5166 - 0]	[0.4999 - 0.4828 - 0]	[0.0000 - 0.0006 - 0]	0.0000
	ca	[0.4994 - 0.4842 - 0]	[0.0000 - 0.0010 - 0]	[0.5005 - 0.5148 - 0]	0.0000
	bc	[0.0000 - 0.0028 - 0]	[0.5004 - 0.5345 - 0]	[0.4996 - 0.4627 - 0]	0.0000
L – L – L	abc	[0.0273 - 0.5495 - 0]	[0.4715 - 0.3355 - 0]	[0.5012 - 0.1150 - 0]	0.0000

Table 5 Feature vectors for different faults under wide variation of operating condition.

Fault Type		X_a for PNN_a	X_b for PNN_b	X_c for PNN_c	Max(abs(Ig)) (kA)
L - g	a-g	[0.5023 - 0.3849 - 1]	[0.2491 - 0.3004 - 1]	[0.2486 - 0.3147 - 1]	0.1995
	b-g	[0.1656 - 0.3149 - 1]	[0.6688 - 0.3821 - 1]	[0.1656 - 0.3030 - 1]	0.3786
	c-g	[0.1786 - 0.3029 - 1]	[0.1782 - 0.3118 - 1]	[0.6432 - 0.3853 - 1]	0.4664
L – L - g	ab-g	[0.1627 - 0.3635 - 1]	[0.5100 - 0.3500 - 1]	[0.3272 - 0.2865 - 1]	0.4094
	ca-g	[0.3421 - 0.3556 - 1]	[0.1335 - 0.2832 - 1]	[0.5244 - 0.3613 - 1]	0.3328
	bc-g	[0.0080 - 0.2864 - 1]	[0.4667 - 0.3590 - 1]	[0.5252 - 0.3546 - 1]	0.1807
L – L	ab	[0.5005 - 0.4184 - 0]	[0.4995 - 0.3279 - 0]	[0.0000 - 0.2537 - 0]	0.0000
	ca	[0.5004 - 0.3389 - 0]	[0.0000 - 0.2511 - 0]	[0.4996 - 0.4101 - 0]	0.0000
	bc	[0.0000 - 0.2591 - 0]	[0.4999 - 0.4099 - 0]	[0.5001 - 0.3310 - 0]	0.0000
L – L – L	abc	[0.0211 - 0.3530 - 0]	[0.4004 - 0.3237 - 0]	[0.5785 - 0.3233 - 0]	0.0000

The results show that the highest fault accuracy for phase fault category (L-L, L-L-L). The minimum fault accuracy is for L-L-g with nine errors, that seven of them are belong to fault on condition $L_f = 0.95^{p.u}$ and R_f = 200 Ω. While, the train data cases involve the maximum of variable condition for above patterns up to $L_f = 0.80^{p.u}$ and R_f = 100 Ω. The L-L-g fault classification accuracy is lower than L-g, because disturbance in healthy phase current is greater in it. Thus, PNN classifiers have largest mismatch in L-L-g fault types.

Based on above description, it is found that even with a small training data set (which is roughly 20% of the testing data set), the achievable accuracy of the proposed method is quite high. The performance of the proposed technique for varying source impedance on five cases has also been analyzed. The results are shown in Table 7. Results show the breakup of the results for source impedance variations in both ends of the transmission line. The table shows that the developed technique is fast (10 ms after FIA) and quite effective for variations of parameters (Z_{S1}, Z_{S2}) except for one case, where fault classification accuracy goes 100%.

Table 6 Fault classification accuracy for different fault types.

Fault Type	No. of Test Cases	No. of Erroneous Classification	Accuracy (%)
L – g	360	1	99.72
L – L – g	360	9	97.50
L – L	360	0	100
L – L – L	120	0	100
Total	1200	10	99.17

Table 7 Performance of proposed algorithm with different source impedance.

Case No.	Zs1 (%)	Zs2 (%)	No . of Test Cases	Accuracy (%)
1	100	100	240	99.17
2	75	100	240	98.75
3	100	75	240	98.33
4	125	100	240	99.58
5	100	125	240	100
Total			1200	99.17

5 Further Studies

5.1 Effect of Noise Signal

Generally, the current waveforms taken of the CT outputs involve added noise signal in actual power system. In power quality analysis, the performance of algorithms is investigated in noisy environment. In digital protection, the effected to noise signal has been reported as main drawback of wavelet techniques. But in this work, the mentioned limitation has been removed by selecting of the fourth level of wavelet coefficients by MRA. Therefore, the 20 dB and 30 dB Gaussian white noises are added to pure current samples in MATLAB software. The new feature vectors are calculated by added noise signals. The fault classifier is tested by new feature vectors. Although, it had been trained by test data cases achieved of preprocessed the pure current signals. The 97.5% and 92.08% fault classification accuracies are found out for signals with 30 dB and 20 dB noises, respectively. The results show that, an increased noise level in the input signals reduces the classification accuracy. From this result, it can be observed that, the WER-PNN technique is robust at noisy environment. Thus, this approach doesn't require to de-noising equipment. However, the proposed technique has removed delay time for de-nosing process.

5.2 Comparison with other Classification Schemes

In recent years, Das has been proposed a reasonably accurate fuzzy logic based fault classification scheme [5] in which a high classification accuracy for half cycle (10ms) after fault inception had been reported for EHV transmission line. It has been found that the 97.416% overall classification accuracy by the fault classification scheme of [5], which is lower than the accuracy in this paper (99.15%) obtained by the proposed WER-PNN scheme. On the other hand, the proposed method is applicable for a wider variation in the operating conditions in comparison to the method proposed by Das. Where as Das's method is valid for variation in R_f up to 50 Ω, variation in L_f up to 0.8 (in p.u. of line length) and variation in δ up to 30°; our proposed method is valid for variation in R_f up to 200 Ω,variation in L_f up to $0.95^{p.u}$ and variation in δ being up to 40°. In this wok, variation of the source impedance has been considered for both sources, while Das has neglected this variable from the operating conditions.

In order to achieve valid comparison between our proposed technique and technique of [5], the different conditions for ten types of fault have been simulated based on generated data cases according to Table 8. The above simulation had been considered in [5]. Beforehand, the same transmission line configuration has been considered for simulation study of both of the methods. As shown in Table 8, a total of $3 \times 5 \times 4 \times 4 \times 10 = 2400$ cases have been generated. In this study, 400 cases are peaked up as train cases and 2000 cases are selected for test cases, randomly.

Finally, after fault cases generating, upon testing over 2000 test cases, It has been found that the classification accuracy by the proposed scheme is

Table 8 Simulation variables in reference [5].

Simulation Variables	Training & Test
Power Angle (δ)	10 – 20 – 30
Fault Inception Angle (°)	10 – 30 – 60 – 90 – 190
Fault Location (percent of line)	0.2 – 0.4 – 0.6 – 0.8
Fault Resistance (Ω)	0 – 5 – 25 - 50
Fault Type	a-g , b-g , c-g, ab-g, ca-g bc-g , ab, ca, bc, abc

Table 9 Classification accuracy with selecting 400 random train cases.

Random Case No.	No. of Test Cases	No. of Erroneous Classification	Accuracy (%)
1	2000	9	99.55
2	2000	19	99.05
3	2000	1	99.95
4	2000	4	99.80
5	2000	6	99.70
6	2000	5	99.75
7	2000	8	99.60
8	2000	14	99.30
Average	2000	8.25	99.58

99.55%, which is higher than the accuracy reported by method of [5] (97.416%) based on the Half Cycle Discrete Fourier Transform (HCDFT) and fuzzy logic system. In order to validate of the proposed method, random selecting of the 400 train cases has been repeated eight fold. Table 9 shows classification results for each random selected case. From the results presented in this table it can be observed that the proposed method has high accuracy (over 99%).

Recently, feature extraction technique based on wavelet transform packages have been presented for transmission fault classification [3, 12-13]. In the above methods, authors have presented a feature selection techniques using sum of the absolute values of detail coefficients of two cycles post fault line currents. The above works find limitations as wavelet detail coefficients is highly susceptible to noise and provides erroneous results even with noise of SNR 20 dB. Also, time duration considered in the analysis comes out to two cycles after post fault. However, these algorithms have low speed. But, in our proposed technique, the energy values of detail and approximation coefficients of half cycles post fault line currents have been used and above drawbacks have been removed.

In this paper, a new feature selection scheme for transmission line fault classification is presented based on GI and WERs using 4th level detail and approximation coefficient of half cycle post fault line currents. The technique is very fast and simple in comparison to the [12, 13] approaches. Feature vectors are normalized and suitable for other power system with different voltage and frequency. Fig. 8 shows Three-dimensional feature vectors (X_a, X_b, X_c) as input for three PNN classifiers corresponding data cases at Table 8 [5]. As shown in Fig. 8, the faulty and healthy phases have been classified into two groups (ground faults (GI=0) and phase fault types (GI=1)). It is observed that the feature 3 (GI) plays basic rule in discriminating of the healthy phase cases (o) (belong to ground fault

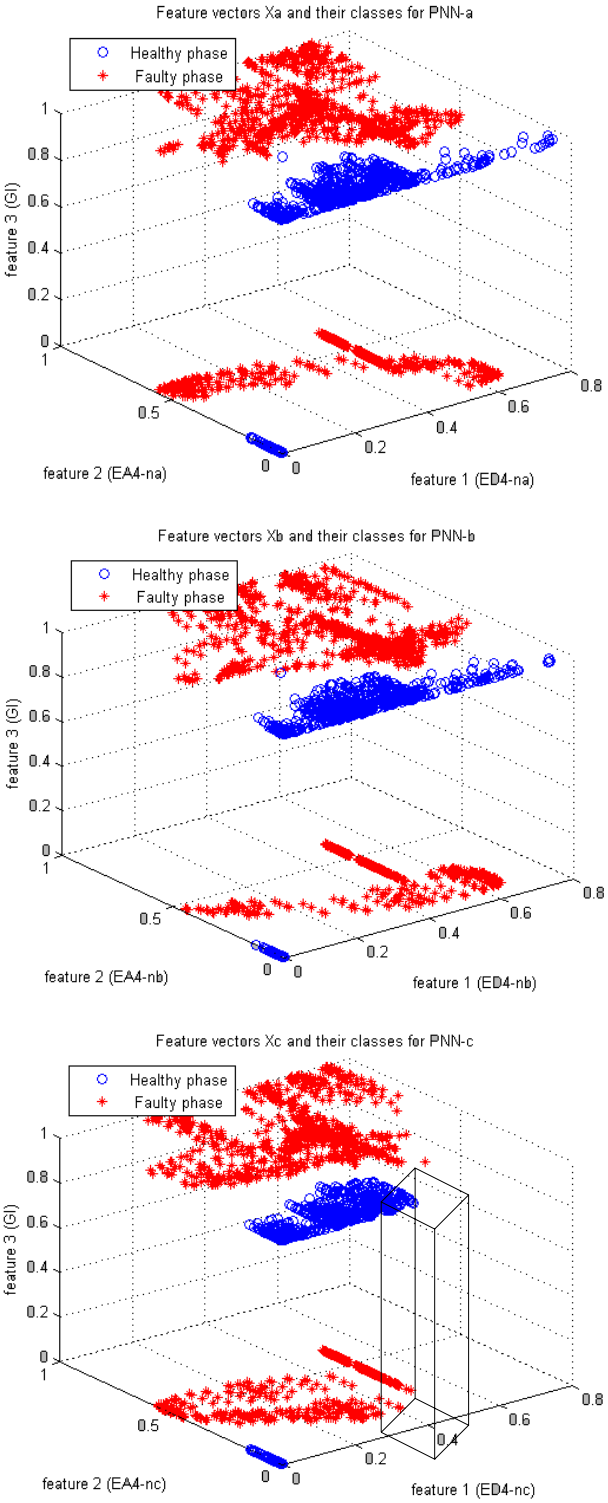

Fig. 8 Three-dimensional features vectors for three classifiers.

types) from the faulty phase cases (*) (belong to phase fault types). For example, in the C-phase features (X_c) case in Fig. 8, it has been found that if the GI hadn't been applied in this algorithm as third feature, the WER nods of healthy phase for ground fault types and faulty phase for phase fault types have covered each other and the phase classifiers hadn't ability to recognition the faulted phase. Also, the relationship between feature 1 and feature 2 corresponding with each phase for fault phase category (GI=1) is nonlinear. However, the PNN classifier as intelligent tool is able to distinguish the transients originating of the faulted phase from those originating of other healthy phases.

6 Conclusion

In this paper, high speed and accurate fault classification technique of transmission line based on new feature selection of wavelet transform and PNN has been presented. Scheme of proposed approach consist of three PNNs and one ground detector. The PNNs distinguish healthy phase from faulted phase in different fault types. Each classifier fed by particular three features. Two first features are WER for detail and approximate coefficient at fourth level decomposition and three features for three classifiers is GI. The proposed technique uses ground current for detecting ground fault and calculating GI. The required data sets are obtained through simulation in PSCAD/EMTDC considering different conditions. The wavelet transform is found to be very powerful tool for feature extraction of transient signal.

The feasibility of the developed technique has been tested on an extensive data set of 1200 test cases covering a wider range of operating conditions rather than to the training data cases. The proposed method is tested with parameters that are not included as part of training. Hence, it is observed that the proposed method is robust to parameter variations. From these test studies, the accuracy of the proposed classification technique has been found to be at least 99% by using only half cycle post fault currents. Thus, this technique is fast and accurate method. Also, the effect of noise on the classification performance is investigated. Therefore, the proposed fault classification technique can be considered quite suitable for digital distance protection scheme of a transmission line. The fault analysis in the transmission line with TCSC and other FACTS devices and those performances on distance protection can be investigated for future works.

References

[1] Elhaffar A. M., *Power transmission line fault location based on current traveling waves*, Helsinki university of technology 2008.

[2] Phadke A. G., *Computer relaying for power systems*, John Wiley & Sons Ltd., New York 1988.

[3] Chanda D., Kishore N. K. and Sinha A. K., "Application of wavelet multi-resolution analysis for identification and classification of faults on transmission lines", *Int. Journal of Electric Power Systems Research*, Vol. 73, No. 3, pp. 323-333, 2005.

[4] El Safty S. and El-Zonkoly A., "Applying wavelet Entropy principle in fault classification" *Int. Journal of Electrical Power and Energy System*, Vol. 31, No. 10, pp. 604-607, 2009.

[5] Das B. and Reddy J. V., "Fuzzy-logic-based fault classification scheme for digital distance protection", *IEEE Trans. of Power Deliver*, Vol. 20, No. 2, pp. 609-616, 2005.

[6] Mahanty R. N. and Gupta P. B. D., "A fuzzy logic based fault classification approach using current samples only", *Int. Journal of Electric Power Systems Research*, Vol. 77, pp. 501-507, 2007.

[7] Sanaye-Pasand M. and Khorashadi-Zadeh H., "An extended ANN-based high speed accurate distance protection algorithm", *Int. Journal Electrical Power and Energy Systems*, Vol. 28, pp. 387-395, 2006.

[8] Martin F. and Aguado A., "Wavelet-based ANN approach for transmission line protection", *IEEE Trans. of Power Deliver*, Vol. 18, No. 2, pp. 1572-1574, 2003.

[9] Silva K. M., Souza B. A. and Brito N. S. D., "Fault detection and classification in transmission lines based on wavelet transform and ANN", *IEEE Trans. of Power Deliver*, Vol. 21, No. 4, pp. 2058-2063, 2006.

[10] Samantaray S. R. and Dash P. K., "Pattern recognition based digital relaying for advanced series compensated line", *Int. Journal of Electr Power and Energy System*, Vol. 30, No. 1, pp. 102-112, 2008.

[11] Parikh U. B., Das B. and Maheshwari R., "Fault classification technique for series compensated transmission line using support vector machine", *Int. Journal of Elect. Power Energy System*, Vol. 32, pp. 629-636, 2010.

[12] Upendar J., Gupta C. P. and Singh G. K., "Discrete wavelet transform and probabilistic neural network based algorithm for classification of fault on transmission systems", *India Conference INDICON 2008*, pp. 206-211, 2008.

[13] Upendar J., Gupta C. P. and Singh G. K., "Statistical decision-tree based fault classification scheme for protection of power transmission lines", *Int. Journal of Electrical Power and Energy Systems*, Vol. 36, pp. 1-12, 2012.

[14] Gaing Z. L., "Wavelet-based neural network for power disturbance recognition and classification", *IEEE Trans. on Power Delivery*, Vol. 19, No. 4, pp. 1560-1568, 2004.

[15] Perera N. and Rajapakse A. D., "Recognition of fault transients using a probabilistic neural-network classifier", *IEEE Trans, on Power Delivery*, Vol. 26, No. 1, pp. 410-419, 2011.
[16] PSCAD/EMTDC *Power systems simulation Manual*, Winnipeg, MB, Canada, 1997.
[17] MATLAB 7.1 version, Math Works Company.
[18] Specht D. F., "Probabilistic neural networks", *Neural Networks*, Vol. 3, No. 1, pp. 109-118, 1990.

Mehrdad Mollanezhad Heydar-Abadi was born in Tehran, Iran, in 1984. He received the B.Sc degree in Electrical engineering from the Birjand University in 2007 and M.Sc. degree from Semnan University in 2013. His areas of interest include power system protection, transient simulations and digital relays.

Asghar Akbari Foroud was born in Hamadan, Iran, in 1972. He received B.Sc. degree from Tehran University and M.Sc.and Ph.D degrees from Tarbiat-modares University, Tehran, Iran. He is now with Semnan University. His research interests include power system dynamics & operation and restructuring.