ARI The Bulletin of the Istanbul Technical University VOLUME 53, NUMBER 1
Communicated by Ahmet Hamdi Kayran

Combined Wavelet-Neural Fault Classifier for Power Distribution Systems

Oben Dağ
Department of Electrical Engineering, Istanbul Technical University, 34469 Maslak, Istanbul, Turkey
Canbolat Uçak*
Department of Electrical Engineering, Istanbul Technical University, 34469 Maslak, Istanbul, Turkey
(Received 29 August 2002)

This paper presents an integrated design of a fault classifier for distribution systems using a hybrid Wavelet-artificial neural network (ANN) based approach. Data for the fault classifier is produced by PSCAD/EMTDC simulation program for 34.5 kV Sagmalcılar-Maltepe distribution system in Istanbul, Turkey. It is aimed to design a classifier capable of recognizing ten classes of three-phase distribution system faults. A database of line currents and line-to-ground voltages is built up including system faults at different fault inception angles and fault locations. The characteristic information over six-channel of current and voltage samples is extracted by the wavelet multi-resolution analysis technique. Afterwards, an ANN-based tool was employed for classification task. The main idea in this approach is solving the complex fault (three-phase short-circuit) classification problem under various system and fault conditions. A self-organizing map, with Kohonen's learning algorithm and type-one learning vector quantization technique is implemented into the fault classification study. The performance of the wavelet-neural fault classifier is presented and the results are analyzed in the paper. It is shown that the technique presented correctly recognizes and discriminates the fault type and faulted phases(s) with a high degree of accuracy for different location and time of occurrence in the simulated model distribution system.

Keywords: Distribution systems, fault classification, wavelet multi-resolution analysis, artificial neural networks, self-organizing feature map, learning vector quantization.

1. Introduction

The quality of electric power has become an important issue for electric utilities and their customers. Customers, in particular, have become less tolerant of power quality disturbances and faults due to the fact that these phenomena degrade the performance and efficiency of customer loads. To improve the quality of power, electric utilities continuously monitor power delivered at customer sites. Disturbance waveforms are captured and recorded continuously using power monitoring instruments. However, these monitoring instruments lack the ability to distinguish between events (without the intelligence). Therefore, existing methods to analyze and identify power disturbances are delicate and laborious since the primary methods are based on visual inspection of the waveforms; it is the duty of the power quality engineers to categorize the enormous amount of data for further analysis manually [1], [2].

It would be desirable if the data collection process could be further automated and the monitoring device not only monitors and records the disturbances, but also classifies them according to appropriate criteria. This would help to immediately detect a disturbance or fault and then make the appropriate decision to eliminate the fault. Consequently, this would minimize the customer displeasure and provide obtaining optimum efficiency from the power system.

There is usually a high volume of recorded event data to be processed and classified when dealing with power quality analysis. This makes it very difficult and time-consuming to interpret the data and provide useful operations. Large dimensionality of the data is one general problem that exists. Moreover, a major concern arising from the classification of a large data set is the complexity of the discrimination process. Due to changes in the disturbance type, duration and its frequency components (which may overlap in time), the parameters in the discriminant model become highly variable. This leads to a considerable deterioration in classification performance

*Current adress: Department of Electrical and Electronics Engineering, Yeditepe University, 81120 Kayışdağı, Istanbul, Turkey

of the classifier. To overcome this problem, it is often necessary to extract the key features from the raw data to obtain a manageable size of data.

Generally, the Fourier techniques are being used to transform the raw data into the frequency domain. These methods are very familiar to the engineering community. However, these methods are not without limitations. For example, the Fourier series requires periodicity of all the time functions, and Fourier transforms are inefficient in capturing short-term transients. Additionally, traditional Fourier analysis does not consider frequencies that evolve with time (non-stationary signals). These limitations often make it difficult to analyze the transient events that are included in the disturbance waveforms. One technique that solves the problems to some level is the windowed Fourier transform (short-time Fourier transform-stft). However, it has the limitation of a fixed window width that means the trade-off between frequency resolution and time resolution (depends on the signal analyzed), must be fixed in order to capture a particular case. A wide window, for instance, gives good frequency resolution but poor time resolution, whereas a narrow window gives good time resolution but poor frequency resolution. A different method is the Wavelet analysis, which provides greater resolution in time for high frequency components of a signal and greater resolution in frequency for the low frequency components of a signal. Wavelet analysis techniques have been applied with success in a wide variety of research areas such as signal analysis, image processing, data compression, de-noising and numerical solution of differential equations. The wavelet analysis techniques have been proposed extensively in the literature as an approach for fault detection, localization and classification of different power system transients [3], [4].

When a fault occurs in a distribution system, disturbance signals like transients would be present in the voltage and current signals. These high-frequency parts of the signals carry essential information that could be used in classifying the fault types. By careful observation of current and voltage waveforms and frequency spectra, some characteristics may be identified for each fault type. Wavelet transform provides the task of extracting the information in the current and voltage waveforms.

In the literature, a great deal of work has been done dealing with the fault classification problem by the artificial neural network (ANN) training techniques [5], [6], [7], [8], [9]. ANN training techniques can be categorized as supervised, unsupervised, and reinforced learning algorithms. For example, one of the typical supervised learning algorithm is back-propagation (BP), which employs a nonlinear regression technique to achieve minimum error goal. Even though it has been reported that BP is adequate for pattern classification problems owing to its high discrimination power and excellent generalization ability [10], the number of classes is limited in practise to apply it directly to large-set classification problems without preclassification. In other words, as the number of classes increases, the computational complexity of the learning problem quickly reaches unmanageable proportions. Furthermore, it is very difficult to determine the structure and size of the network for the classification of large-sets and complex patterns. It is presented in [11] that for a fault classification study the classification rate is only about 79%. However, when a fault classifier is based on an unsupervised (i.e., Kohonen ANN) or combined supervised and unsupervised training technique, such as self-organizing mapping (SOM), radial-basis function (RBF), counter propagation network (CPN), the classification rate can reach a high level of about 95%. In [12], Mladen Kezunovic and Igor Rikalo designed a system to detect and classify faults using a combined supervised-unsupervised neural network based on self-organizing for a twelve-class faults at transmission lines. They used a neural-network algorithm with ISODATA clustering algorithm. For their system, they obtained a classification rate of 95.93%. This result shows how a combined supervised-unsupervised scheme based on self-organizing is appropriate for fault classification studies. Another ANN training technique is reinforced learning algorithm, i.e., a genetic algorithm that is used to search the weight space of a multilayer feed-forward ANN without the use of any information. The basic concept behind this technique is that complete sets of weights are coded into a string, which has an associated fitness finding attribute for the optimal weight. Although the reinforced learning performs a global search and therefore minimizes the possibility of getting stuck in local minima, the training is very time-consuming; classification rate is around 85%.

In this paper, a new approach to the fault classification studies is proposed to further improve the classification performance. The main idea is to design a classifier capable of recognizing ten classes of three-phase system faults. It is aimed to design an intelligent structure that will be able to distinguish the fault type using the captured and recorded six-channel current and voltage signals from the power system. Therefore, this will provide automation in the power disturbance analysis and identification process. As the type of the fault is identified by the intelligent structure, possible solutions can be quickly determined to solve the problem. So this scheme will gain speed in the power quality study process. For this, first the 34,5 kV Sagmalcılar-Maltepe distribution system in Istanbul, Turkey with a 12-bus configuration is simulated by PSCAD/EMTDC and a data-set of six channels current and voltage signals for each of the fault categories is generated. Then the important characteristics that distinguish each class from another are obtained with the wavelet multiresolution analysis technique. After the uniqueness of each signal is found, an ANN-based classification tool was employed for the classification task. The proposed method utilizes the concept of supervised learning and unsupervised learning based on self-organizing maps. The ANN technique provides the ability to classify each fault classes by identifying different patterns of the associated voltages and currents. The results of the study are summarized in the conclusion part of the paper.

This paper is organized as follows. An introduction to feature detection and extraction technique is presented in Section 2. Section 3 presents the SOM based adaptive neural classifier. Application of the proposed fault classifier on the 34.5 kV Sagmalcılar-Maltepe distribution system in Istanbul, Turkey is presented in Section 4. Finally, the conclusion is presented in Section 5.

2. Feature detection and extraction

The neural network approach to the detection and classification of system faults consists of three general tasks; generating sets of line current and line-to-ground voltages, using these sets to train a neural network, and testing the network on separate sets of line currents and line-to-ground voltages. The preprocessor is an internal part of this scheme. Training cases were generated using an electro-magnetic transient simulation program on a distribution system. To enhance the competence of the classifier system, it is necessary to pre-process the event signals to extract characteristic information. Besides, it is impractical to use the raw waveforms directly as input for a neural network. Thus, certain characteristics of the waveforms must be identified and reduced to quantitative form to make it possible for the network to distinguish between faulty conditions.

Generally, the primary objectives of feature extraction are; (i) To represent the typical characteristics of categories, and (ii) to extract the discriminative information between categories. In the case of this study, feature extraction is performed as a pre-processing operation that transforms a pattern from its original form to a new form suitable for further analysis. So, the operation in this study reduces the high dimensionality of the initial system description. The feature extraction method proposed in this paper is based on the wavelet MRA technique, which maps the raw data generated by EMTDC into a small size of interpretable features. The MRA is a tool that utilizes the Discrete Wavelet Transform (DWT) to represent a time-varying signal in terms of its frequency components [13]. It essentially maps a one-dimensional signal of time into a two-dimensional signal of time and scale. Wavelet analysis involves representing signals in terms of simpler, fixed building blocks (wavelets) at different scales and positions. The main idea is to develop representations of a complicated signal $f(t)$ in terms of its orthonormal basis, which are the scaling and wavelet functions. These two functions are translated and scaled to produce wavelets at different locations (positions) and on different scales (durations). Fine-scale wavelets are narrow and brief; coarse-scale wavelets are wide and long lasting. The wavelet functions represent the high frequencies corresponding to the detailed parts of a given signal, and scaling functions represent low frequencies or smooth parts of the signals. These functions can be scaled and translated to decompose $f(t)$ and represent it at different resolutions or scales. This decomposition technique is called multi-resolution signal decomposition (MSD). In the next paragraphs, a brief description of the wavelet transform technique in the context of fault classification problem in this paper will be given. A complete description of the DWT and MSD theory can be found

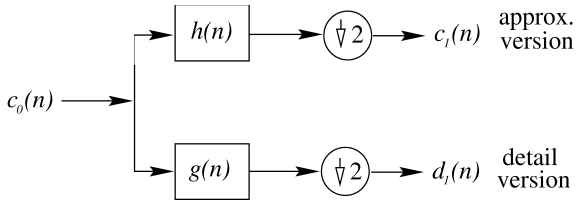

Figure 1. One stage MSD using convolution and decimation by factor 2.

in [14].

Figure 1 shows the block diagram of the Discrete Wavelet Transform implemented with discrete-time filters and subsampling by two. It illustrates the decomposition of $c_0(n)$, a recorded digitized time signal (which is a sampled version of $f(t)$), into its detailed $d_1(n)$ and smoothed $c_1(n)$ signals using filters $g(n)$ and $h(n)$, respectively. Filter $g(n)$ has a high-pass filter response. Thus, the filtered signal $d_1(n)$ is a detailed version of $c_0(n)$ and contains higher frequency components than the smoothed signal $c_1(n)$, because filter $h(n)$ has a low-pass frequency filter response. The digitized time signal $c_0(n)$ can be written as

$$c_0(n) = \sum_k h(k-2n)c_0(k) + \sum_k g(k-2n)c_0(k). \quad (1)$$

One iteration of the scheme in Figure 1 on the first low-band creates a new low-band that corresponds to the lower quarter of the frequency spectrum. Each further iteration halves the width of the low-band, but due to sub-sampling by two, its time resolution is halved as well. The decomposition of $c_0(n)$ in Figure 1 is a first-scale decomposition. Higher-order decompositions are then performed in a similar manner. At each iteration, the current high band portion correspond to the difference between the previous low-band portion and the current one, which is a pass-band.

The purpose of feature extraction task in the fault classification system in this study is to identify specific signatures of the fault types in the system. The wavelet transform breaks down the signal into different time-frequency scales. Each scale represents the signal in the corresponding sub-band. By using wavelet analysis, the sub-band information can be extracted from the simulated waveforms, which contain useful fault features. Some bands are intensive to some types of fault. The energy content of the scale signals relative to the given signal changes depending upon the type of disturbance. By analyzing these features of the detail signals, different types of faults can be detected and classified. Parsevals theorem relates the energy of the signal to the energy in each of the expansion components and their wavelet coefficients if the selected scaling function and the wavelet function form an orthonormal basis. This means that the energy of the signal $f(t)$ can be partitioned in terms of the expansion coefficients as in [3]:

$$\int |f(t)|^2 dt = \sum_{k=-\infty}^{\infty} |c_0(k)|^2 + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} |d_j(k)|^2, \quad (2)$$

where $d_j(k)$ is the detail coefficients at scale j, and $c_0(k)$ presents the last approximate coefficients. The energy of the analyzed signal is portioned at five resolution levels according to Equation 2. The standard deviation can be considered as a measure of the energy for a signal with zero mean [3]. For this reason, the standard deviations at different resolution levels of the signal will be used to form the feature vector to classify different fault classes.

3. Adaptive pattern classification

A literature search in [15] shows that most of the ANN studies for fault classification are based on multi-layer, feed- forward nets. In the case of the typical supervised back-propagation (BP) network, sets of associated input-output pairs are presented to the ANN that learns a model of the mapping between input and output. However, training of a BP network is very time consuming, needs very large training sets, and easily gets stuck on local minima. Furthermore, it can be difficult to retrain the ANN with new training data. As a result, it may not be sufficient for the task of fault classification. Another approach for the ANN application for fault classification is using data self-organization obtained through the use of unsupervised learning. Here, the task of the classification network is to cluster the faults into separate classes. So, it can be viewed as a pattern recognition problem. Self-organization refers to the specific learning method without external examples. This is also called unsupervised

learning. Given a set of input patterns, neighboring processing units (neurons) in a self-organizing net develop into detectors of specific categories of patterns. So, each local neuron-group acts as a decoder for the inputs [16]. After the learning phase through unsupervised learning, the ANN is ready for the classification task, where the features selected from the input data are assigned to individual classes. Although a self-organizing map is equipped to perform the role of a classifier, it is recommended in literature ([17], [18]) that for best performance it should be accompanied with a supervised learning scheme. Computation of the self-organizing map (feature map) may be viewed as the first of two stages for adaptively solving a pattern classification problem. The second stage is provided by learning vector quantization, which performs a mechanism for the final fine-tuning of a feature map. The combination of a self- organizing map and a supervised learning scheme forms an adaptive pattern classification that is hybrid in nature [16].

In this study, a self-organizing map (SOM), with Kohonens learning algorithm and learning vector quantization (LVQ) technique [17], is implemented into the fault classification study. The SOM is intended to discover significant patterns or features from a set of feature vectors obtained by the data preprocessor. SOM obtains the information hidden in high dimensional data that is otherwise difficult to interpret. The SOM converts the complex nonlinear relationship between high-dimensional data into a simple geometric relationship on a low-dimensional display. So it is a vector quantization technique: it compresses the information, while preserving the most important topological relationship of the primary data elements.

The SOM consists of neurons organized on a regular low-dimensional grid. Each neuron represents a d-dimensional weight vector (prototype vector, codebook vector, model vector) where d is equal to the dimension of the input vectors. The neurons are connected to adjacent neurons by neighborhood relation that dictates the topology or structure of the map. The SOM is especially suitable for data analysis because it has important visualization properties. It creates a set of prototype vectors representing the data set and carries out a topology preserving projection of the prototypes from d-dimensional input space onto a low-dimensional grid. This ordered grid can be used as a convenient visualization surface for showing different features of the SOM and the data, i.e. the cluster structure.

The SOM reflects variations in the statistics of the input distribution. Regions in the input space from which sample vectors are drawn with a high probability of occurrence are mapped onto larger domains of the output space, and therefore with better resolution than regions in the input space from which sample vectors are drawn with a low probability of occurrence. This is the density matching property of the feature map [16]. So, this property implies that if a particular region of the input space contains frequently occurring stimuli, it will be represented by a larger area in the feature map than a region of the input space where the stimuli occur less frequently.

The topology preserving mapping algorithm of Kohonen is an iterative process for training a class of neural networks [16]. The learning procedure is unsupervised or self organizing and is used to train a network of units or neurons that are arranged in a low-dimensional sheet-like structure. A two-dimensional structure for the network is shown in Figure 2-a; however, it is also possible to use one or more dimensional structures. The local topology type of the map can be selected to be either rectangular or hexagonal as shown in Figure 2-b and Figure 2-c. The difference between rectangular and hexagonal lattices is that in the former all 8 neighbors of a neuron are at the same distance and in the latter 6 neighbors of a neuron are at the same distance.

The training procedure of SOM is iterative. At each training step, a sample vector x is randomly chosen from the input data set. Distances between x and all the prototype vectors are computed. The best matching unit (BMU), which is denoted by b, is the map unit with prototype closest to x,

$$||x - m_b|| = \min_i \{||x - m_i||\}. \tag{3}$$

Then, the prototype vectors are updated. The BMU and its topological neighbors are moved closer to the input vector in the input space. The update rule for the prototype vector of unit i, according to [17] is

$$m_i(t+1) = m_i(t) + \eta(t) h_{bi}(t)[x - m_i(t)] \tag{4}$$

where t is discrete time, $0 < \eta(t) < 1$ is time dependent learning rate parameter, $h_{bi}(t)$ is the neighborhood kernel function centered on the

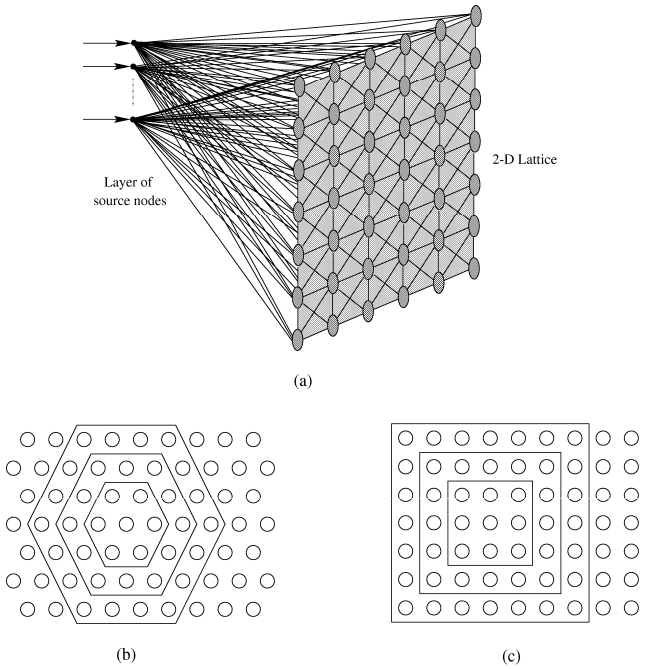

Figure 2. (a) Two-dimensional lattice of neurons, (b) Neighborhood (size 1, 2, and 3) of the unit marked with black dot for hexagonal lattice, (c) Neighborhood (size 1, 2, and 3) of the unit marked with back dot for rectangular lattice.

winner unit. A typical choice of $h_{bi}(t)$ is the Gaussian function [16]

$$h_{bi}(t) = \exp(-\frac{||r_b - r_i||^2}{2\sigma^2(t)}), \tag{5}$$

where r_b and r_i are positions of neurons b and i on the SOM grid, and $\sigma(t)$ is the width of the topological neighborhood function [16]. It is denoted as,

$$\sigma(t) = \sigma_0 \exp(-\frac{t}{\tau_1}), \tag{6}$$

where σ_0 is the value of $\sigma(t)$ at the initiation of the SOM algorithm, and τ_1 is a time constant. In Equation 4, $\eta(t)$ is

$$\eta(t) = \eta_0 \exp(-\frac{t}{\tau_2}), \tag{7}$$

where τ_2 is another time constant and η_0 is the initial state of learning rate. Both $\eta(t)$ and $\sigma(t)$ decrease monotonically with time. In the case of a discrete data set and fixed neighborhood kernel, the error function of SOM can be shown to be

$$E = \sum_{i=1}^{N}\sum_{j=1}^{M} h_{bi}||x_i - m_j||^2, \tag{8}$$

where N is number of training samples in the input data set, and M is the number of map units. Neighborhood kernel is centered at unit b, which is the BMU of vector x_i, and evaluated for unit j [18].

In an attempt to accelerate the computation of the SOM, the batch algorithm is used in this paper, [19]. In batch map principle, the whole training set is gone through before the map is updated. Actually, updating is done by replacing the prototype vector with a weighted average over the samples, where the weighting factors are the neighborhood function values. In each training step, the data set is partitioned according to the Voronoi regions of the map weight vectors, i.e., each data vector belongs to the data set of the map unit to which it is closest; this set is called the Voronoi set. After this, the new weight vectors are calculated as:

$$m_i(t+1) = \frac{\sum_{j=1}^{n} h_{bi}(t)x_j}{\sum_{j=1}^{n} h_{bi}(t)}, \tag{9}$$

where

$$b = \arg\min_k\{||x_j - m_k||\} \tag{10}$$

is the index of the BMU of data sample x_j. The new weight vector is a weighted average of the data samples, where the weight of each data sample is the neighborhood function value $h_{bi}(t)$ at its BMU b [16]. This is the way batch algorithm has been implemented in this paper. In batch version of the SOM algorithm the order in which the input patterns are presented to the network has no effect on the final form of the feature map, and there is no need for a learning-rate schedule. But the algorithm still requires the use of a neighborhood function [16].

According to Kohonen in [17], if the SOM is to be used as a pattern classifier the map units are grouped into subsets, each of which correspond to a class of patterns, then the problem becomes a decision process. One should not use the maps as such for pattern recognition or decision processes because it is possible to increase the recognition

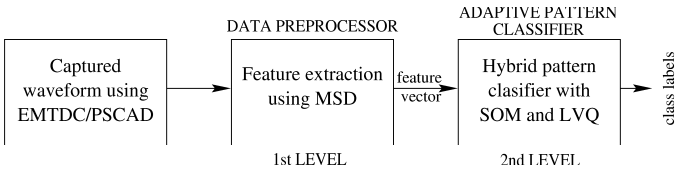

Figure 3. Block diagram of fault classification system.

accuracy by a significant amount if the maps are fine tuned according to a supervised learning algorithm. The SOM algorithm is intended to approximate input signal values, or their probability density function, by codebook vectors. If the signal sets are to be classified into a number of categories, then several codebook vectors usually represent each class, and decisions made at class borders will be important, but the identity of the codebook vectors within the classes is no longer important. It is possible to define values for the codebook vectors that directly define the decision borders between the classes. In this paper, Type One-Learning Vector Quantization (LVQ1) algorithm [17], is used as a supervised classifier that uses class information to move the Voronoi vectors slightly to improve the quality of the classifier decision regions. After the hidden patterns in a set of feature vectors are discovered and initial classification is performed by the SOM, the LVQ1 technique is applied to improve the quality of the classifier. The LVQ1 is called a binary output pattern classifier since its output is either zero or one. It is a supervised version of the self-organizing map network, suitable particularly for pattern recognition problems. The purpose of LVQ1 is to group a set of related input signals into a finite number of categories based on similarity of the input signals, thereby fine-tuning the initial map, with the number of such categories predetermined by the SOM. Fine-tuning the map is achieved by pulling the codebook vectors away from the decision surfaces to demarcate the class borders more accurately. Let m_c be the codebook vector closest to x in the Euclidean metric; this then also defines the classification of x. Then apply training vectors x, the classification of which is known. Update the $m_i = m_i(t)$ as follows:

$$m_c(t+1) = m_c(t) + \alpha(t)[x(t) - m_c(t)]$$

if x is classified correctly; (11)

$$m_c(t+1) = m_c(t) - \alpha(t)[x(t) - m_c(t)]$$

if x is classified incorrectly; (12)

$$m_i(t+1) = m_i(t) \quad \text{for} \quad i \neq c. \tag{13}$$

Here $\alpha(t)$ is a scalar adaptation gain (learning rate) $(0 < \alpha(t) < 1)$, which is decreasing monotonically in time. Since this is a fine-tuning method, its initial value should be selected a small value, i.e. 0.01 or 0.02 for 100,000 steps [17].

The whole fault classification scheme with the hybrid neural network structure is demonstrated in Figure 3. The hybrid neural network forms a two-level classification approach. The whole feature extraction-classification system forms a pyramid, where the number of patterns and connections decrease. The primary benefit of the two-level approach is the reduction of the computation cost.

The major parts of the system in Figure 3 are as follows:

i. Data preprocessor that extracts the feature vectors from the raw six channel signals generated by the PSCAD/EMTDC on the model distribution system.

ii. Self-Organizing Map: Unsupervised layer that clusters the data vectors taken from data preprocessor to separate clusters.

iii. Learning Vector Quantization: Supervised layer carrying out the fault classification.

4. Simulation results

The fault classification scheme in this study [15] is defined as a multi-class problem with ten types of short-circuit system faults which are AG, BG, CG, AB, AC, BC, ABG, ACG, BCG, and ABCG. The letters A, B, C, and G, stands for phase A,

phase B, phase C, and Ground, respectively. For example, BG means that single phase to ground fault occurred in phase B and ABCG means that three phase to ground fault occurred in the distribution system.

In this study, a distribution system in Istanbul, Turkey is simulated by PSCAD/EMTDC simulation program. The network parameter of the test system illustrated in Figure 4 is provided in Table 1. The test system is a 12-bus distribution system with a base power of 100 MVA and a base voltage of 34.5 kV. All of the test data were monitored from the secondary of the main step-down transformer (TRF 1) located at the Sagmalcılar substation as shown in Figure 4, node 2. The objective is to monitor the voltage and the current at only the 34.5 kV bus and identify the fault classes. Each sample of data contained six channels, a set of three line currents and three line-to-ground voltages, which is typically what a recording device would measure in a real system.

4.1. Data generation using PSCAD/EMTDC

Various fault conditions were simulated using the PSCAD / EMTDC software, which is an ideal tool for fault simulation and transient analysis. Specific events simulated included system faults at different fault inception angles and fault locations. The fault inception angles (FIA) used in the simulations were in the range of 0°∼180°. Since the waves were periodic, it was sufficient to study angles in the range of 0°∼180°. Short-circuit faults are simulated at various locations of the test system. Four main types of system faults were generated: single-phase to ground, two-phase fault, two-phase to ground, and three-phase to ground fault.

A database of three channel line currents and three channel line-to-ground voltages was built up for various types of faults at different locations and fault inception angles. Then, a data set of six channel waveforms was created for further processing by the data pre-processor. One channel consisted of a 10 cycle signal generated at an equivalent sampling rate of 5 KHz (fundamental frequency of 50 Hz) with 1001 sample points lasting 200 msec. Thus, the six-channel data set was a data matrix of (1001,6).

Figure 5 shows the three-phase line currents and the line-to-ground voltage signals for a phase A-phase B to ground fault occurred at bus 4 on

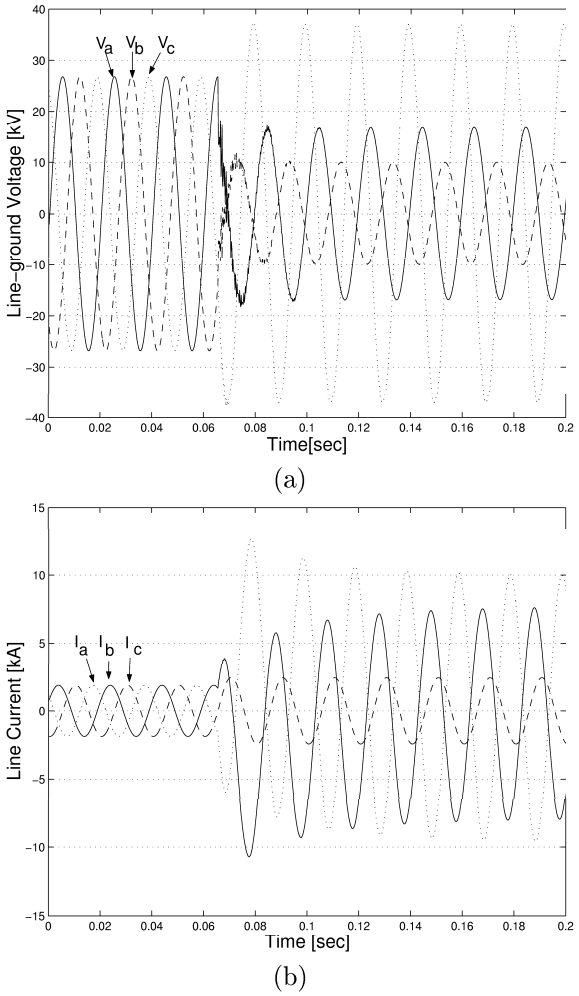

Figure 5. Typical measured voltage and current patterns. (a) Voltage waveforms, (b) current waveforms.

0.0636 sec. with a fault inception angle of 90°. It is possible to investigate various types of disturbances in this example. An abrupt change can be observed in both of the current and voltage signals at the time 0.0636 sec. Figure 5-a shows a 15% and 50% sag disturbance on the faulted phase voltages $V_a(t)$ and $V_b(t)$ respectively, and a 30% swell disturbance on the unfaulted phase $V_c(t)$. There is some high-frequency (HF) distortion on the voltage waveforms, in particular on the faulted a and b phases at fault occurrence time and this is so by virtue of the fact that there is a large step change in the a phase

Table 1
The parameters of the 34.5 kV Sagmalcılar-Maltepe distribution system.

Source	154 kV, 10000 MVA (short circuit), 50 Hz
TRF1	100 MVA, 154 kV/34.5 kV, 50 Hz, Leakage inductance=10.0%
TRF2, TRF3, TRF4, TRF5, TRF6	20 MVA, 34.5 kV/10.5 kV, 50 Hz, Leakage inductance=10.0%
LOAD 1, LOAD 2, LOAD 3, LOAD 4	19 MVA, $\cos\phi$=0.906, Delta connected
LOAD 5	19 MVA, $\cos\phi$=0.707, Delta connected
CBANK	7.2 MVA, Delta connected
CABLE 1, CABLE 2, CABLE 3, CABLE 4	3 km, 3x(1x240 mm^2) XLPE, 20.3/34.5 kV
CABLE 5	3.5 km, 3x(1x240 mm^2) XLPE, 20.3/34.5 kV

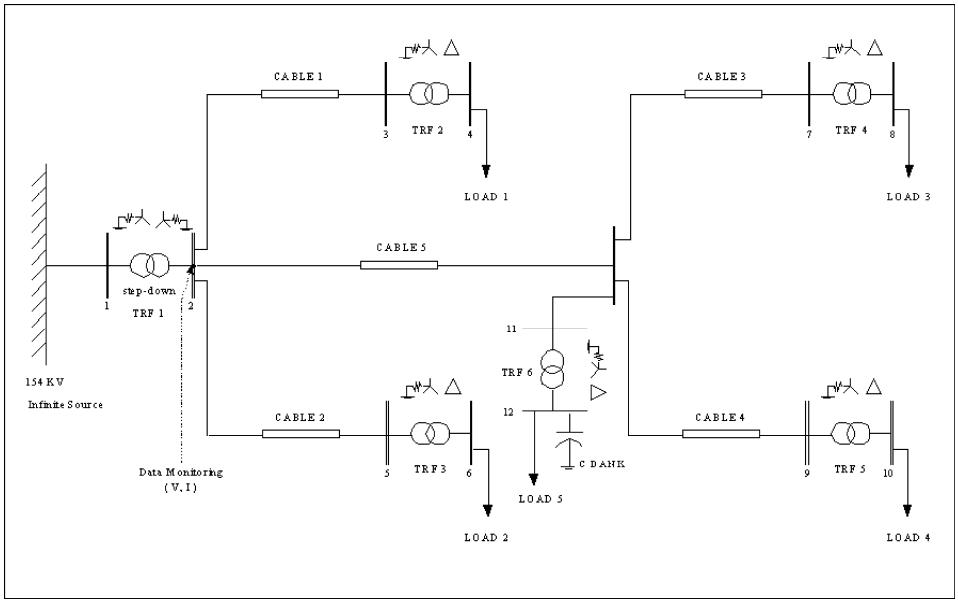

Figure 4. One-line diagram of the 34.5 kV Sagmalcılar-Maltepe Distribution System.

voltage and the b phase voltage when the fault occurs. In addition to voltage signals, the faulted line ($I_a(t)$ and $I_b(t)$) currents increase greater than the current ($I_c(t)$) in healthy line as it can be seen from Figure 5-b. In the case of double-line-ground fault, only the faulty line currents increases greatly: the increase in the magnitude of the faulted a phase current and b phase current is larger than c phase current, as expected. The abrupt change observed in the current and voltage signal at 0.0636 sec. enables the classification scheme to identify the fault phase and fault type.

4.2. Feature extraction using MSD technique

Since the waveforms have certain distinct characteristics, any successful classification tool would be able to pick out these relevant features and associate the waveforms with those of a certain fault class. Here, fault classification is defined as a multi-class problem. The ten types of faults produce a ten-class classification problem. Having chosen the classes, the next step in developing a classifier is the selection and extraction of desired features. This refers to the preprocessing of raw data into a smaller set of features that would be the input to the classifier. This is probably the most critical step in the analysis. The criterion used in feature selection was to represent the important characteristics that distinguish each class from another. Here the wavelet multi-resolution analysis technique is used as a preprocessing unit to obtain a smaller set of data to represent each of the class. The key idea in using wavelet transform analysis (for classifying fault types) is based on the uniqueness of the wavelet transform coefficients (WTCs) for each type of signals. The second technique is descriptive statistics that summarize the data into a small set of numbers that contain most of the relevant information.

As shown in Figure 6, five-scale MSD analysis based on a dyadic-orthonormal wavelet transform analysis with Daubechies ten-coefficient filter is performed to ensure that all disturbance features in both high and low frequency are extracted. The given input signal in Figure 6-a and Figure 6-b is the phase-A line current and phase-C line current respectively, for a phase A-phase B to ground fault occurred at bus 4 on 0.0636 sec. with a fault inception angle of 90°. The input signal has 1001 sample points lasting 0.2 sec. The signal consists of 10 cycle data, each

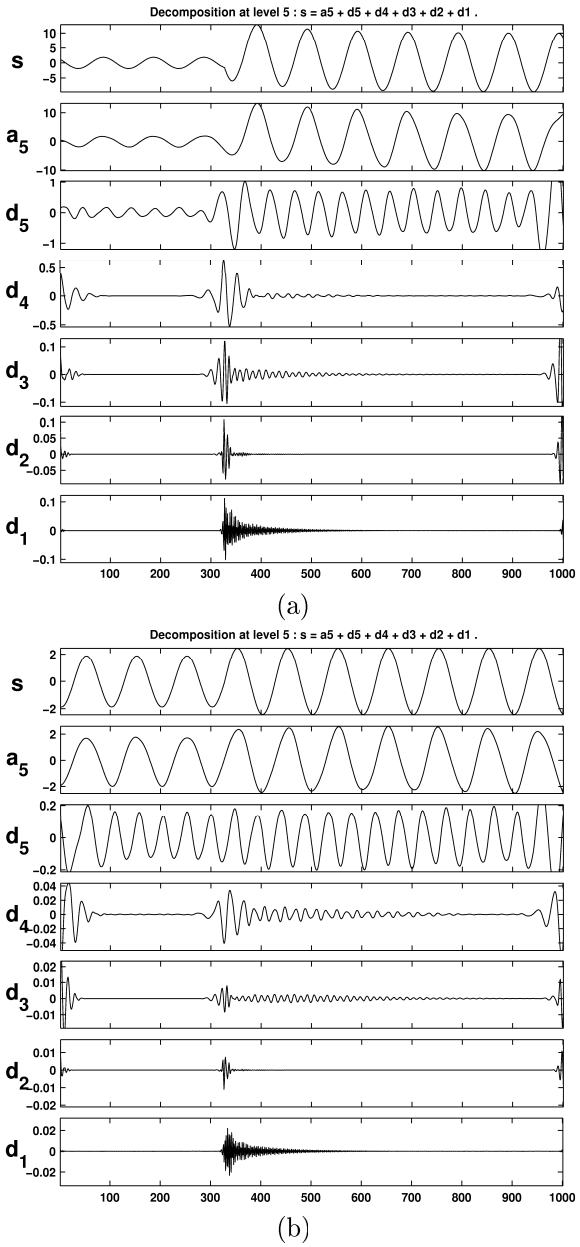

Figure 6. MSD of current waveforms. (a) Phase-a current waveforms, (b) Phase-c current waveforms.

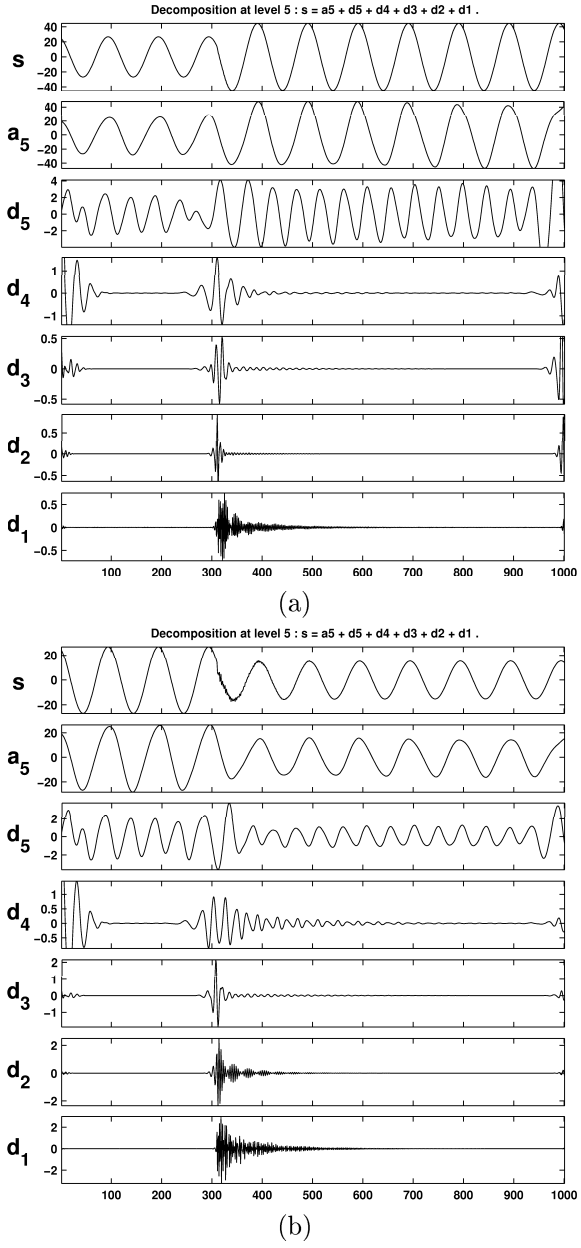

Figure 7. MSD of current waveforms. (a) The voltage swell disturbance signal. (b) The voltage sag disturbance signal.

cycle lasting 0.02 sec. The input signal has been sampled at 5 KHz. Thus, a five-scale decomposition of a signal yields 5 detailed signals having a frequency band of 2.5-1.25 KHz at scale 1; 1.25-0.625 KHz at scale 2; 625-312.5 Hz at scale 3; 312.5-156.25 Hz at scale 4; 156.25-78.125 Hz at scale 5 and one smooth signal contains frequency band 78.125 Hz to DC level. As it is seen from Figure 6, the MSD technique is adequate in detecting the change in the current waveform at the fault point. In Figure 6-a and Figure 6-b, the first finer decomposition levels of the distorted signal may be adequate to detect and localize the disturbance. However, the other coarser resolution levels are also used to extract more features that can help in the classification scheme. (The spikes, which are seen at the initial and final edges of Figure 6, are caused by discontinuities or edge effects inherent in the analysis process that can be easily eliminated by using proper window scheme.)

The voltage waveforms for A-g fault and ABC-g fault at bus-4 with a fault inception angle of 30°are given in Figure 7-a and Figure 7-b, respectively. The two distorted signals belong to $V_c(t)$ channel. Figure 7-a shows a 50% swell disturbance and Figure 7-b shows a 50% sag disturbance. These are slow varying disturbances. For both of the disturbances, the abrupt change in the magnitude of signals when fault occurs is seen at scales 1, 2, 3, and 4. The rapid oscillation disturbances (high frequency) in voltage sag in Figure 7-b are seen in scales 1 and 2. The change in the magnitude of the signals is best seen in scales 4 and 5. Although these two waveforms belong to two different disturbance classes, their detail components have similar characteristics. And it is difficult to separate single-phase to ground faults and three-phase to ground faults, since the difference between the two fault classes are the sag and swell disturbances. Therefore the MSD technique is not adequate in detecting slowly varying events like voltage sags, swells, and outages owing to the poor time resolution at low frequency [20]. This limitation of the MSD can be overcome by tracking the voltage signal power, which is its mean square value [20].

4.2.1. The Feature Vector

The characteristic information over six-channel current and voltage samples is extracted by the data pre-processor shown in Figure 3. The six-channel data set is then reduced to a feature vec-

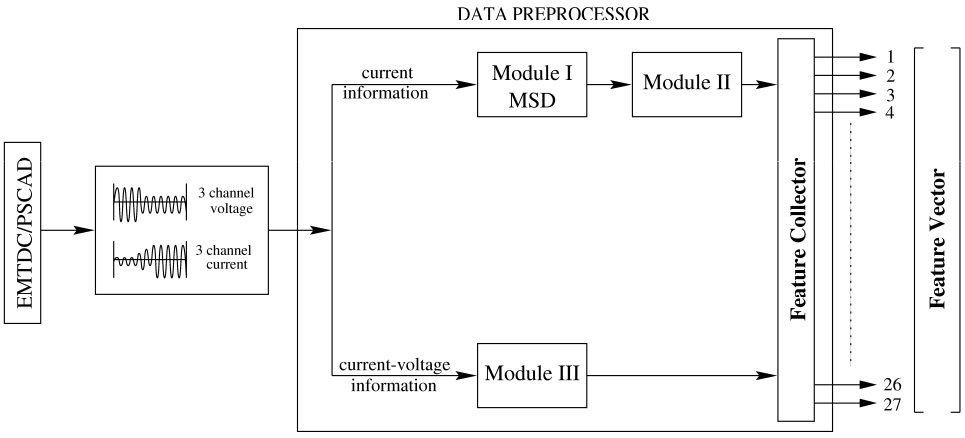

Figure 8. Data processing and feature extraction architecture.

tor of a small set with 27 parameters. Accordingly, the event feature vector parameters are as follows:

i. The maximum modulus current for each three channel before the fault occurs.

ii. The maximum modulus current for each three channel immediately after the fault occurs.

iii. Energy of each current channel over the 0.078-0.156 kHz band range.

iv. Energy of each current channel over the 0.156-0.312 kHz band range.

v. Energy of each current channel over the 0.312-0.625 kHz band range.

vi. Energy of each current channel over the 0.625-1.250 kHz band range.

vii. Energy of each current channel over the 1.250-2.500 kHz band range.

viii. The signal power for each three channel voltage waveforms.

ix. The energy for each three channel voltage waveforms.

The definition of the above properties is as follows: The energy of the current signals are partitioned at different resolution levels by the property of Parsevals Theorem [3], [21]. The standard deviation of the DWT coefficients at the resolution levels of five-level MSD serves as a representative of the current signal energy partitioning and hence to aid in classification [3], [21]. The standard deviation of five-level detail coefficients of each current signal channel is computed, yielding fifteen parameters. The magnitude change in the current waveforms is used as a feature to aid in classification of fault classes. This yields six parameters for three channel current waveforms. To detect and separate slowly varying events like voltage sags and swells, the voltage signal power, which is the mean square value, is used [20]. The calculated mean square value for three channel voltage waveforms yields three parameters for feature vector. Furthermore, to enhance the competence of the classifier system and to extract more relevant information the standard deviation of voltage channels is computed, which yields three parameters for the feature vector. Thus, a set of 27 variables are obtained.

The data pre-processing scheme in Figure 3 is illustrated in more detail in Figure 8. The preprocessor extracts pertinent information over 10 cycles of operation. The voltage and current signals monitored from the secondary of the main step- down transformer (TRF 1) in Figure 4 are fed into the signal-processing unit. These modules extract the features required by the fault detection and classification network. Figure 8 shows the different modules belonging to the signal-processing unit. Module I extracts the five-level MSD detail coefficients of three current

channels. Here, the MSD approach is based on a dyadic-orthonormal wavelet transform analysis with Daubechies wavelet with a ten-coefficient filter. The statistical information and other average quantities are extracted by the Module II and Module III. Module II extracts the distribution of the energy of the sub-bands obtained by the Module I. Module III extracts the statistical information and average quantities of the three channel voltage and three channel current information. The feature collector module collects the information produced by the previous three modules. The information is then processed by the feature collector module and finally the feature collector module produced a feature vector of 27 parameters, which will be the data for the input layer of the ANN.

4.3. ANN classifier and simulation results

The whole fault classification architecture is shown in Figure 9. By changing fault type, fault inception angle, and fault location, a set of 350 cases were generated using the PSCAD/EMTDC simulation software. For each fault condition the raw data was preprocessed by the data preprocessor in Figure 8 and then 350 input vectors with 27 variables were generated. Thus, there were a total of 27 input units, which included the invariant parameters obtained by the data preprocessor. All of the different cases were then divided into two sets, one to be used for neural-network training and the other for testing. The training set consisted of 250 training examples (25 examples per class) with ten fault classes, five fault locations, and five inception angles. The test set consisted of 100 examples (10 examples per class) with ten fault classes, five fault locations, and two inception angles. Verification of training results was performed so that the ANN was first tested with training patterns, which were used in training, then with samples, which were not used in training.

Several simulations with varying parameters (self-organizing map size, map structure, etc.), related to the hybrid pattern classifier illustrated in Figure 9, were performed to obtain the best performance. The simulations included six cases, which were obtained with the best performance overall the simulations. The Hybrid neural network classification results are shown in Table 2. Feature extraction with the wavelet multiresolution analysis technique and classification of ten-fault categories with ANN approach was performed on the Matlab platform running on a Pentium III processor at 1 GHz workstation. The hybrid pattern classifier in Figure 3 was trained using 250 different fault patterns. During the testing, the pattern classifier was presented with 100 new fault patterns.

The common properties of the simulations are as follows. The simulations included several map sizes between 80 map units to 110 map units. A sheet shaped map with a local topology of rectangular and hexagonal structure was used for the simulations. The map sizes were (8,10), (9,10), and (11,10). In the first phase, the initial values of the prototype vectors of feature map were given according to random initialization and linear initialization. In random initialization the map weight vectors were initialized with small random values. And in linear initialization the map weight vectors were initialized in an orderly fashion along the two greatest eigenvectors of the covariance matrix of the training data. The neighborhood function for SOM algorithm was chosen to be an Epanechicov function; because the performance obtained by this function was better then the Gaussian function given by Equation 5. The Epanechicov neighborhood function is denoted as [19],

$$h_{bi}(t) = \max\{0, 1 - sigma(t) - ||x - m_i||\}^2. \quad (14)$$

The SOM algorithm was implemented in batch training algorithm given by Equation 9. The number of training steps for the SOM algorithm was between 4000 and 5500 iterations. After adaptation of the map with the SOM algorithm, it was calibrated according to majority voting [17]. In the second phase, the classification task was performed with LVQ1 algorithm to fine-tune the feature map and to better approximate the data cloud. For the LVQ1 algorithm, the initialization of the codebook vectors was performed according to the final state of the SOFM corresponding to each simulation. The learning rate function for LVQ1 algorithm was selected to be the inverse-of-time function. The initial learning rate was 0.01 for all cases. The number of training steps for the LVQ1 algorithm was between 2800 and 3850 iterations. First the hybrid classifier (shown in Figure 9) was trained with the training set that consists of 250 training examples (25 examples per class). Afterwards, it was tested with the test set that consists of 100 training examples

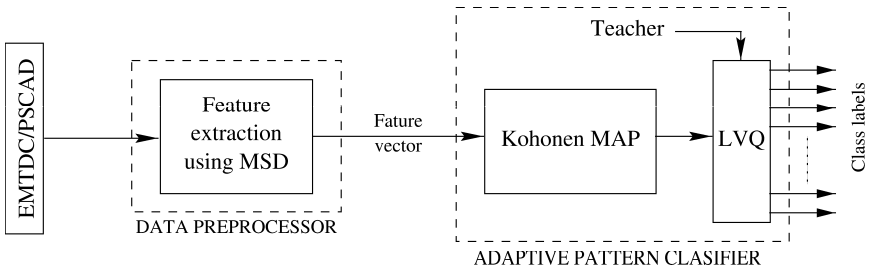

Figure 9. The adaptive pattern classification.

Table 2
Hybrid neural network classification results.

		Classification rate (%)	
Simulation	Description	Training data set	Test data set
SIM-1	Random initialized hexagonal structure of (8,10) map	100	89
SIM-2	Linear initialized rectangular structure of (8,10) map	99	92
SIM-3	Random initialized hexagonal structure of (9,10) map	100	89
SIM-4	Linear initialized rectangular structure of (9,10) map	99	91
SIM-5	Linear initialized hexagonal structure of (11,10) map	99	85
SIM-6	Random initialized rectangular structure of (11,10) map	100	88

(10 examples per class). (The data set was normalized according to logarithmic transformation to improve numerical accuracy before being used in adaptive pattern classifier). The logarithmic transformation is denoted as [19],

$$X(t) = \ln\{x(t) + 1 - min(x(t))\}. \quad (15)$$

The rated performance of the test data was between 85% and 92% as shown in Table 2. Among all the simulation cases, the best classification value was obtained using the empiric formula of Kohonen [19]. The number of map units was 80 ($5\sqrt{n} = 80$) for SIM-2. Here, n denotes the number of training samples which is equal to 250. Using rectangular map structure generally increases the classification rate of the classifier. For example, in the case of a (8,10) map the result obtained by a rectangular lattice (92%) is better than the hexagonal lattice (89%). Also in the case of (9,10) map, rectangular lattice (91%) is better than the hexagonal lattice (89%). Finally in the case of (11,10) map, rectangular lattice (88%) is better than the hexagonal lattice (85%). This may be because the eight neighbors of a neuron are at the same distance in comparison to the six neighbors in a hexagonal lattice (as seen in Figure 2-b and Figure 2-c). Therefore, the map may organize itself better.

Table 3 shows the computational times for neural processes. As the number of map unit increase the computation takes longer time. Note that the training of the neural structure is a batch computation which is, especially in Matlab environment, is significantly faster than sequential computation scheme.

A visualization of the SOM after training with the type-one LVQ is demonstrated in Figure 10-a to give an idea about how a map looks like at the end of the simulation. This map illustrates the SIM-2 case in Table 2. It shows the unified distance matrix (U-mat) of the SOM for the test data set. The U-mat is a visualizing technique to analyze the cluster structure of SOM. It shows the distance between prototype vectors of neighboring map units and thus shows the cluster structure of the map. Dark color indicates large distance between neighboring map units (indicates cluster borders). Clusters are typically uniform areas with bright color. Generally, neurons with the same color belong to the same category [18]. It is possible to distinguish the ten types of fault categories from the U-matrix.

Another visualization method is hit histograms [19]. The hit histograms are formed by taking a data set, finding the BMU of each data sample

Table 3
Computational times for neural processes.

	SOM [min]	LVQ [min]	Test [sec]
(8,10) map	5	15	$\ll 1$
(9,10) map	7.3	17.4	$\ll 1$
(11,10) map	10	20	$\ll 1$

from the map, and increasing the counter in the map unit whenever it is a BMU. The hit histogram shows the distribution of the data set on the map. In Figure 10-b multiple hit histograms are shown simultaneously to investigate the test data set of SIM-2 in Table 2. Here the hit histograms of ten data sets are illustrated according to the following match: fault type AG is represented by number 1, BG is represented by 2, CG is represented by 3, AB is represented by 4, BC is represented by 5, AC is represented by 6, ABG is represented by 7, BCG is represented by 8, ACG is represented by 9,and finally fault type ABCG is represented by number 0. Here U-mat uses interpolated shading of colors. Interpolation is a process for estimating values that lie between known data points. The size of the hit histogram determines how many times the corresponding map unit is selected as BMU. The clusters are better represented on the U-mat with the hit-histograms. Since there is some correlation between the cluster labeled with 4 and the cluster labeled with 7, the separation between them is not very clear on the U-mat, but from hit-histograms, it seems that they correspond to two different clusters. The same argument can be done for clusters 6 and 9.

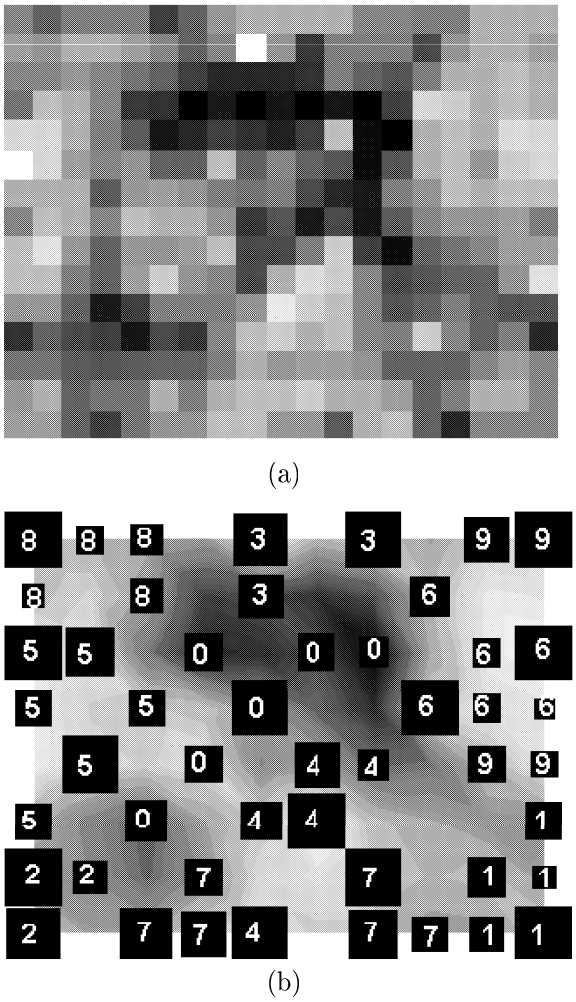

Figure 10. U-matrix (a) and hit-histogram (b) of the SOM for the test data set.

5. Conclusions

A Combined Wavelet-ANN based fault classifier has been investigated for electrical distribution systems in this paper. Ten fault categories have been selected to be identified using the proposed approach. It is shown that the technique presented correctly recognizes and discriminates the fault type and faulted phases(s) with a high degree of accuracy for different location and time of occurrence in the simulated model distribution system.

The underlying approach of the proposed classifier is to carry out (preprocessed) waveform recognition in the self- organizing feature map. The SOM is intended to discover significant patterns or features from a set of feature vectors obtained by the data preprocessor. SOM obtains the information hidden in high dimensional data that is otherwise difficult to interpret. The test results show that the decision regions between different fault classes are quite clearly defined. Combining the information extracted by SOM with a supervised learning algorithm: type-one learning vector quantization makes a final decision about the fault type.

The performance of the proposed fault classification technique is comparable and close to the classifiers in the literature. In [12], Mladen Kezunovic and Igor Rikalo reported that a neu-

ral network based on supervised and unsupervised learning can achieve a high rate of 95.93 % for a twelve-class fault classification problem at transmission lines. Moreover, in [20], a wavelet-neurofuzzy approach for power quality violations detection performed 95% for training set and 92% for test set. The proposed fault classifier approach demonstrated in Figure 9 is trained using 250 different fault patterns for the distribution system. During the testing, the pattern classifier was presented with 100 new fault patterns. The rated performance of the test data is between 85% and 92%. The classification architecture used in this paper correctly classified 92 % of the patterns, and it could not classify 8% out of 100 new fault patterns. All the test results presented show that the proposed fault classification technique based on SOM is well suited for fault classification problems. This hybrid method is easy and very promising for fault classification problem.

This research has showed that combined Wavelet-ANN technique can be used for the classification of power system short-circuit faults. More work is needed to further explore the other aspects of cluster characteristics and to better classify each of them. Future work can involve the use of actual recorded field data to verify initial results obtained in this study.

6. Acknowledgements

This study is sponsored by the Institute of Science and Technology, Istanbul Technical University, Turkey. Grant Number: 3049/2001.

We thank chief engineer Mehmet Gonen and engineer Nazim Kenc of 'Boğaziçi Dağıtım A.Ş.' for providing Sağmalcılar-Maltepe distribution system data.

References

[1] S. Santoso, E. J. Powers, W. M. Grady, and A. J. Parsons, IEEE Transactions on Power Delivery **15**, 222 (2000).

[2] S. Santoso, E. J. Powers, W. M. Grady, and A. J. Parsons, IEEE Transactions on Power Delivery **15**, 229 (2000).

[3] A. M. Gaouda, E. F. El-Saadany, M. M. A. Salama, V. K. Sood, and A. Y. Chikhani, IEEE Transactions on Power Systems **16**, 662 (2001).

[4] W. Zhao, Y. H. Song, and Y. Min, Electric Power System Research **53**, 23 (2000).

[5] A. K. Ghosh and D. L. Lubkeman, IEEE Transactions on Power Delivery **10**, 109 (1995).

[6] R. K. Aggarwal, Q. Y. Xuan, R. W. Dunn, A. T. Johns, and A. Bennett, IEEE Transactions on Power Delivery **14**, 1250 (1999).

[7] N. C. Fahmida, J. L. Arevena, W. M. Grady, and A. J. Parsons, IEEE Transactions on Control Systems Technology **6**, 623 (1998).

[8] N. Kandil, V. K. Sood, K. Khorasani, and R. V. Patel, IEEE Transactions on Power Systems **7**, 812 (1992).

[9] K. L. Frick and S. K. Starret, Classification of Disturbance Characteristics Using a Kohonen Neural Network, Large Engineering Systems Conference on Power Engineering, Halifax, Canada, 124 (2000).

[10] H. Song and S. Lee, IEEE Transactions on Neural Networks **9**, 369 (1998).

[11] R. K. Aggarwal, Q. Y. Xuan, A. T. Johns, F. Li, and A. Bennett, IEEE Transactions on Neural Networks **10**, 1214 (1999).

[12] M. Kezunovic and I. Rikalo, IEEE Computer Applications in Power System **9**, 43 (1996).

[13] M. Vetterli and C. Herley, IEEE Transactions on Signal Processing **40**, 2207 (1992).

[14] S. Mallat, A Wavelet Tour Of Signal Processing (Academic Press, Cambridge, 1999).

[15] O. Dağ, MSc thesis, Istanbul Technical University, (2002).

[16] S. Haykin, Neural Networks: (A Comprehensive Foundation, Prentice Hall International. Inc., New Jersey, 1999).

[17] J. A. Kangas, T. Kohonen, and J. T. Laaksonen, IEEE Transactions on Neural Networks **1**, 93 (1990).

[18] B. D. Ripley, Pattern Recognition and Neural Networks (Cambridge University Press, Cambridge, 1996).

[19] J. Vesanto, J. Himberg, E. Alhoniemi, and J. Parkankangas, SOM Toolbox for Matlab 5, Helsinki University of Technology, URL: http://www.cis.hut.fi/projects/somtoolbox/, (2000)

[20] A. Elmitwally, S. Farghal, M. Kandil, S. Abdelkader, and M. Elkateb, IEE Proc.- Gener. Transm. Distrib **148**, 662 (2001).

[21] A. M.Gaouda, M. M. A. Salama, M. R. Sultan, and A. Y. Chikhani, IEEE Transactions on Power Delivery **14**, 16 (1999).