International Journal of Computational Intelligence and Informatics, Vol. 3: No. 3, October - December 2013

Mammogram Classification using Fuzzy Neural Network

M Velmurugan
Department of ComputerScience
Periyar University, Salem
Tamilnadu, India
velmca23@gmail.com

K Thangavel
Department of ComputerScience
Periyar University, Salem
Tamilnadu, India
drktvelu@yahoo.com

R Subash Chandra Boss
Department of Computer Science
Periyar University, Salem
Tamilnadu, India
rmsubash_18@yahoo.com

***Abstract* - Breast cancer is one of the major causes for the increased mortality among women especially in developed countries. It is second most common cancer in women. The World Health Organization's International estimated that more than 1, 50,000 women worldwide die of breast cancer in year. In India, breast cancer accounts for 23% of all the female cancer death followed by cervical cancer which accounts to 17.5% in India. Early detection of cancer leads to significant improvements in conservation treatment. However, recent studies have shown that the sensitivity of these systems is significantly decreased as the density of the breast increased while the specificity of the systems remained relatively constant. Mammography is a medical imaging technique that combines, low-dose radiation and high-contrast, high resolution film for examination of the breast and screening for breast cancer. Another disadvantage is false-positive result. This research proposes a fuzzy neural network for classifying mammograms. Results of screening the mammograms are organized by classification and finally grouped into three categories i.e., Normal, malignant and Benign. Experimental results show that this method performs well with the classification accuracy reaching nearly 82% in comparison with the already existing algorithms. The fuzzy neural network has provided high accuracy in the early diagnosis of Mammography, which can provide quantitative indicators for early clinical diagnosis and serve as a convenient diagnostic tool for physicians.**

Keywords **–** Neural networks, Learning algorithm, Fuzzy logic, Mammogram

I. INTRODUCTION

Mammography associated with clinical breast examination and breast self-examination are the only viable and efficient methods at present for mass screening to detect breast cancer. Breast cancer is the second most deadly form of cancer in women. It appears in women in the form of tumors. The diagnosis of breast cancer in its early stage of development has become important in the prevention of breast cancer. To avoid a surgical procedure such as a biopsy at its initial stage, women widely depend on mammography. Mammography is the standard approach 70 for preliminary examination of breast cancer abnormalities. This research proposes a fuzzy neural network for classifying mammograms. Results of screening the mammograms are organized by classification and finally grouped into three categories i.e., Normal, malignant and Benign. Experimental results show that this method performs well with the classification accuracy reaching nearly 90% in comparison with the already existing algorithms. The fuzzy neural network may provide high accuracy in the early diagnosis of Mammography, which can provide quantitative indicators for early clinical diagnosis and serve as a convenient diagnostic tool for physicians.

II. RELATED WORK

Several researchers have introduced different approaches for classifying the mammogram images. A histogram intersection based image classification was proposed in [4]. Initially they used the bag-of-words model for image classification for capturing the texture information. A normalized histogram intersection with the K-Nearest Neighborhood classifier was applied. The classification accuracy depends on the normalization of the histogram. [5] presents mammogram image classification based on rough set theory in conjunction with statistical feature extraction techniques. The features were derived from the gray level co-occurrence matrix, these features were normalized and the rough set dependency rules are generated from the attribute vector. The generated rules were passed to the classifier for the classification purpose. Classification of mammograms with benign, malignant and normal tissues using independent component was proposed by the authors in [6] with a classification accuracy of 97.3%.The face recognition methods such as AdaBost and Support vector machines for the analysis of digital mammograms was presented in [7]. The AdaBost classifier achieved 76% for all lesions and 90% for the masses. A fractal approach was proposed in [8] to model the mammographic parenchymal, ductal patterns and enhance the micro calcifications. The results proved that fractal modeling is an efficient approach for detection and classification of micro calcification in a computer aided diagnosis systems.

The most commonly used algorithms for classification purpose are based on neural networks, like genetic algorithm, rule based classifier and fuzzy theory. In this work we have used the neural network approach for the classification of the digital mammogram images. The images considered in the present work are listed in the Digital Database for Screening Mammography [16] and MIAS [17]. Recently BIRADS (Breast Imaging Reporting and Data System) [18] is becoming the most common acceptable standard for mammography images. Based on the tissue density, they are classified into four categories. Fig. 1 shows a typical example of mammogram images with different BIRADS standards. Fig. 2 shows the different phases involved in the proposed method. BIRADS I: the breast is almost entirely fatty. BIRADS II: there is some fibro glandular tissue. BIRADS III: the breast is heterogeneously dense. BIRADS IV: the breast is extremely dense.

III. PRE-PROCESSING

Pre-processing is an important issue in low-level image processing. It is possible to filter out the noise present in image using filtering. A high pass filter passes the frequent changes in the gray level and a low pass filter reduces the frequent changes in the gray level of an image. That is; the low pass filter smoothes and often removes the sharp edges. A special type of low pass filter is the Median filter. The Median filter takes an area of image (3 x 3, 5 x 5, 7 x 7 etc), observes all pixel values in that area and puts it into the array called element array. Then, the element array is sorted and the median value of the element array is found out. We have achieved this by sorting the element array in the ascending order using bubble sort and returning the middle elements of the sorted array as the median value. The output image array is the set of all the median values of the element arrays obtained for all the pixels. Median filter goes into a series of loops which cover the entire image array. Some of the important features of the Median filter are: It is a non-linear digital filtering technique. It works on a monochrome color image. It reduces “speckle” and “salt and paper” noise. It is easy to change the size of the Median filter. It removes noise in image, but adds small changes in noise-free parts of image. It does not require convolution. Its edge preserving nature makes it useful in many cases. The selected median value will be exactly equal to one of the existing brightness value, so that no round-off error is involved when we take independently with integer brightness values comparing to the other filters.

A.Median filter

Pre-processing is an important issue in low-level image processing. Using filtering it is possible to filter out the noise present in image. A high pass filter passes the frequent changes in the gray level and a low pass filter reduces the frequent changes in the gray level of an image. That is; the low pass filter smoothes and often removes the sharp edges. A special type of low pass filter is the Median filter. The Median filter takes an area of image (3 x 3, 5 x 5, 7 x 7, etc), observes all pixel values in that area and puts it into the array called element array. Then, the element array is sorted and the median value of the element array is found out. We have achieved this by sorting the element array in the ascending order using bubble sort and returning the middle elements of the sorted array as the median value. The output image array is the set of all the median values of the element arrays obtained for all the pixels [13]. Median filter goes into a series of loops which cover the entire image array. Following are some of the important features of the Medi-an filter: It is a non-linear digital filtering technique. It works on a monochrome color image. It reduces “speckle” and “salt and paper” noise. It is easy to change the size of the Median filter. It removes noise in image, but adds small changes in noise-free parts of image. It does not require convolution. Its edge preserving nature makes it useful in many cases. The median value selected will be exactly equal to one of the existing brightness value so that no round-off error is in-volved when we work independently with integer brightness values comparing to the other filters .

IV. Image Segmentation

Image segmentation refers to the process of partitioning a digital image to multiple segments or set of pixels. The goal of segmentation is to simplify the representation of an image into different segments that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries in images. It is also the process of assigning a label to every pixel in an image such that pixels with the same label share certain visual characteristics. The result of image segmentation is a set of segments that collectively cover the entire image. Image segmentation is nothing but the process of dividing an image into disjoint homogenous regions. These regions usually contain similar objects of interest. The homogeneity of the segmented regions can be measured using pixel intensity. Image segmentation techniques can be broadly classified as into five main classes threshold based, Cluster based, Edge based, Region based, and Watershed based segmentation.

The median filter is used for pre-processing of image. It is normally used to reduce noise in an image. The 14 Haralick features are extracted from mammogram image using Gray Level Co-occurrence Matrix (GLCM) for different angles. The features are clustered by K-Means and FCM algorithms inorder to segment the region of interests for further classification**.**

4.1. Fuzzy C Means Theorem:

A constrained fuzzy partition $\{C_1, C_2, \ldots, C_k\}$ can be a local minimum of objective function J_m only if the following conditions are satisfied:

$$\mu_{C_i}(x) = \frac{1}{\sum_{j=1}^{k}\left(\frac{\| x - v_i \|^2}{\| x - v_j \|^2}\right)^{\frac{1}{m-1}}}, \ \forall\ 1 \le i \le k, x \in X \qquad (1)$$

$$v_i = \frac{\sum_{x \in X} (\mu_{C_i}(x))^m x}{\sum_{x \in X}^{n} (\mu_{C_i}(x))^m}, \ \forall\ 1 \le i \le k, x \in X \qquad (2)$$

Based on this theorem, FCM updates the prototypes and the membership function iteratively using equations (1) and (2) until a convergence criterion is reached. We describe the algorithm below.

The Fuzzy c-means algorithm is given below

Fuzzy C-Means Algorithm

Algorithm: FCM(X, c, m, ε)

X: an unlabeled data set.

c: the number the clusters.

m: the parameter in the objective function.

ε : a threshold for the convergence criteria.

Step 1: Initialize prototype $V=\{v^1, v^2, \ldots, v_c\}$
Step 2:Repeat
Step 3:$V^{Previous} \leftarrow V$
Step 4: Compute membership functions using.

Step 5:Update the prototype, v_i in V

Step 6: Until $\sum_{i=1}^{c} \| v_i^{\text{Previous}} - v_i \| \leq \varepsilon$

V. Feature Extraction

The idea is to calculate the co-occurrence matrix for small regions of the image and then use this matrix to find statistic values, for instance Contrast, Correlation, Uniformity, Homogeneity, Probability, Inverse and Entropy. The distance and angle is converted to a vertical and a horizontal offset in pixels according to the following list of angle offset. Gray-Level Co-occurrence Matrix (GLCM) is one of the texture descriptors most used in the literature. Starting to summarize different researches we can find works by Bovis and Singh [7] studying how to detect masses in mammograms on the basis of textural features using five co-occurrence matrices statistics extracted from four spatial orientations, horizontal, left diagonal, vertical and right diagonal corresponding to (00, 450, 900 and 1350) and four pixel distance (d = 1, 3, 6 and 9). Hence, a classification is performed using each feature vector and linear discriminate analysis. According to Marti et al. [8], GLCMs are frequently used in computer vision obtaining satisfactory results as texture classifiers in different applications. Their approach uses mutual information with the purpose to calculate the amount of mutual information between images using histograms distributions obtained by grey-level co-occurrence matrices. Blot and Zwiggelaar [9] [10] proposed two approaches based in detection and enhancement of structures in images using GLCM. The purpose is to compare the difference between these two matrices obtaining a probability estimate of the abnormal image structures in the ROI. Finally, another study based on background texture extraction for classification of Blot and Zwiggelaar [11] presented their work where there is a statistical difference between GLCM for image regions that include image structures and regions that only contain background texture which is provided by a classification in mammograms. In 2003, different approaches based on co-occurrences matrix as a feature descriptors extraction were developed. Houssay et al. [12] presented a neuro-fuzzy model for fast detection of candidate circumscribed masses in mammograms and texture features are estimated using co-occurrence matrices which are used to train the neuro-fuzzy model. On the other hand, Marti et al. [13] proposed a supervised method for the segmentation of masses in mammographic images using texture features which present a homogeneous behaviour inside the selected region. Jirari [14] proposes an intelligent Computer-Aided

Detection system (CAD) by constructing five co-occurrence matrices at different distances for each suspicious region. A different number of statistical features are used to train and test the Radial Basis neural network. Another work is presented by Lena et al. [15] with the study of a multi resolution texture feature of second order statistics were extracted from spatial GLCM using different orientations and distances. Recent studies, Karahaliou et al. [16] investigate whether texture properties of the tissue surrounding micro calcifications using a wavelet transform. Thirteen textural features were calculated from four GLCMs. Finally, Lyra et al. [17] study how to identify breast tissue quality data quantification using a CAD system, where images categorized using the BIRADS breast density index. The texture features were derived for each sub-region from an averaged gray- level co-occurrence matrix (GLCM). Karahaliou et al. [18] Graylevel texture and wavelet coefficient texture features at three decomposition levels are extracted from surrounding tissue regions of interest

5.1 Grey-Level Co-Occurrence Matrices:

The Statistics of grey-level histograms give parameters for each processed region but do not provide any information about the repeating nature of texture. According to Beichel and Sonka [19], the occurrence of gray-level configuration may be described by matrices of relative frequencies, called co-occurrence matrices. Hence, the GLCM is a tabulation of how often different combinations of pixel brightness values (grey levels) occur in an image. GLCM are constructed by observing pairs of image cells distance *d* from each other and incrementing the matrix position corresponding to the grey level of both cells. This allows us to derive four matrices for each given distance: 0, *P (450, d)*, *P (900, d)*, and *P (1350, d)*. For instance, *P (00; d)* is defined as follows:

$$P\ ((00), d\ (a, b)) = |\ \{((k, l), (m, n)) \in D:$$

$$k - m = 0,\ |l - n| = d,\ f(k, l) = a,\ f(m, n) = b\}|$$

where each P value is the number of times that: $f(x1, y1) = i, f(x2, y2) = j$, $|x1 - x2| = d$ and $y1 = y2$ append simultaneously in the image. *P (450, d), P (900, d), P (1350, d)* are defined similarly:

$$P\ ((450), d\ (a, b)) = |f\ ((k, l); (m, n)) \in D:$$

$$(k - m = d, l - n = - d)\ OR\ (k - m = d; l - n = d), f\ (k, l) = a, f\ (m, n) = b\}|$$

$$P\ ((900), d\ (a, b)) = |\ \{((k, l), (m, n)) \in D:$$

$$|k - m| = d,\ l - n = 0, f\ (k, l) = a, f\ (m, n) = b\}|$$

$$P\ ((1350), d\ (a, b)) = |\ f\ ((k, l); (m, n)) \in D:$$

$$(k - m = d, l - n = d)\ OR\ (k - m = - d, l - n = -d), f\ (k, l) = a, f\ (m, n) = b\}|$$

A co-occurrence matrix contains the frequency of a certainpair of pixels repetition in an image. According to the previous formulas the parameters needed are the follows:

Number of grey levels: Normally, it is used a grayscale image of 256 grey levels, which means a high computational cost because all possible pixel pairs must be taken into account. The solution is to generate the matrix reducing the number of greyscales, and so the number of possible pixel combinations. The co-occurrence matrix is always square with the same dimensionality as the number of grey-levels chosen. Distance between pixels (*d*): the co-occurrence matrix stores the number of times that a certain pair of pixels is found in the image. Normally the pair of pixels are just neighbours, but the matrix could also be computed analyzing the relation between non consecutive pixels. Thus a distance between pixels must be previously defined. Angle (θ): Similarly to the distance it is necessary to define the direction of the pair of pixels. The most common directions are 00, 450, 900, 1350 and its symmetric equivalents. Figure 1. shows an example of how we can construct a cooccurrence matrix with eight grey levels, computed using one for distance between pixels and zero degrees for the direction. In this case, the element (1, 1) of *C* matrix is equivalent to 1 because it has been found only one occurrence in the original image f. Another example is shown in the Figure 1**.** On the element (6, 2), where there are three occurrences because a pixel with a value of 6 has a pixel valued 2 immediately to its right. The other elements of *C* are computed in the same way. Figure 2. How to generate a co-occurrence matrix

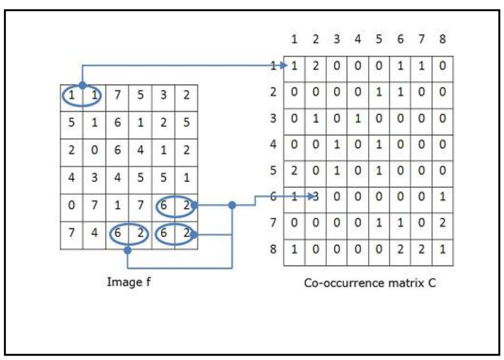

Figure 2. How to generate a co-occurrence matrix

The co-occurrence matrix has some properties about the spatial distribution of the gray levels in the texture image. Haralick [20] proposed descriptors used for characterizing cooccurrence matrices of size *K x K*. The term

VI. Feature Selection

In Image Processing and computer vision, feature selection process refers to choose a subset of attributes from the set of original attributes. The purpose of the feature selection is to identify the significant features, eliminate the irrelevant of dispensable features to the learning task, and build a good learning model such as web categorization discussed. The benefits of feature selection are twofold: it considerably decreased the computation time of the induction algorithm and increased the accuracy of the resulting mode.

Feature selection and feature extraction are two kinds of methods of dimensionality reduction for classification. Feature extraction creates new features by irreversibly transforming the original features such that the created features contain most useful information for the target concept. In contrast, feature selection only removes the features that are necessary or unimportant to the target concept and the remaining features are kept intact. The feature selection problem is essentially a combinatorial optimization problem which is computationally expensive.

Definition 1:

Let I = {U,A} is an information system ,where U is a non-empty set of objects and A is a non-empty finite set of attributes such that a:u→ v_a for every a∈ $v_a$is the set of values.

The information table assigns a value a(x) from $v_a$to each attribute aand object x in the universe U.

$$IND(P) = \{(X, Y) \in U^2 \mid \forall a \in P, a(x) = a(y)\} \quad (3)$$

The relation $IND(P)$ is called a Pindiscernibility relation. The partition of U is a family of all equivalence classes of $IND(P)$and is denoted by $U/IND(P)$ or U/P.

If$(x, y) \in IND(P)$, then x and y are indiscernible by the attributes from P[5].

Lower and upper approximations

Definition 2:

Let $\mathrm{I} = (\mathrm{U}, \mathrm{A})$ be an information system. Where U is the universe with a non-empty set of finite objects, A is a non-empty finite set of condition attributes. $\forall a \in A$, There is a corresponding function $\mathrm{f_a}: \mathrm{U} \rightarrow \mathrm{V_{a,}}$ where $\mathrm{V_{a,}}$ is the set of values of a. Let$X \subseteq U$, the P-lower approximation $\underline{P}X$ and P-upper approximation $\overline{P}X$ of set X can be defined as:

$$\mathrm{X} = \{\mathrm{x} \in \mathrm{U} \mid [\mathrm{x}]_\mathrm{p} \subseteq \mathrm{X}\} \quad (4)$$

$$\overline{\mathrm{P}}\,\mathrm{X} = \{\mathrm{x} \in \mathrm{U} \mid [\mathrm{x}]_\mathrm{p} \cap \mathrm{X} \neq \emptyset\} \quad (5)$$

In Rough set theory, The Quick Reduct algorithm searches for a minimal subset without exhaustively generating all possible subsets. The search begin with an empty subset; attributes which result in the greatest increase in the roughest dependency value are added iteratively. This process continues until the search produces its maximum possible dependency value for that dataset (γ_c(D))

Algorithm:QUICKREDUCT(C,D)

C, the set of all conditional features;

D, the set of decision features.

(1) $R \leftarrow \{\}$
(2) do
(3) $T \leftarrow R$
(4) $\forall_{x\in}$(C-R)
(5) if $\gamma_{R \cup \{x\}}(D) > \gamma_T(D)$
(6) $T \leftarrow R \cup \{x\}$
(7) $R \leftarrow T$
(8) Until$\gamma_R(D) = \gamma_C(D)$
(9) Return R

VII. CLASSIFICATION

A.Neural Network and Fuzzy

Recently, the interest in neural networks has grown dramatically: it is expected that neural network models will be useful both as models of real brain functions and as computational devices. One of the most popular neural networks is the layered feed forward neural network with a Back Propagation (BP) least mean- square learning algorithm. Its topology is shown in Fig. 1. The network edges connect the processing units called neurons. With each neuron input there is associated a weight, representing its relative importance in the set of the neuron's inputs. The inputs' values to each neuron are accumulated through the net function to yield the net value: the net value is a weighted linear combination of the neuron's inputs' values. For the purpose of multicriteria analysis a hierarchy of criteria is used to determine an overall pattern evaluation. The hierarchy can be encoded into a hierarchical neural network where each neuron corresponds to a criterion. The input neurons of the network correspond to single criterion. The hidden and output neurons correspond to complex criteria. As evaluation function it can be used as the net function of the neurons. However, the criteria can be combined linearly when it is assumed that they are independent.

But in practice the criteria are correlated to some degree. The linear evaluation function is unable to capture relationship between the criteria. In order to overcome this drawback of Standard Back Propagation (SBP) algorithm

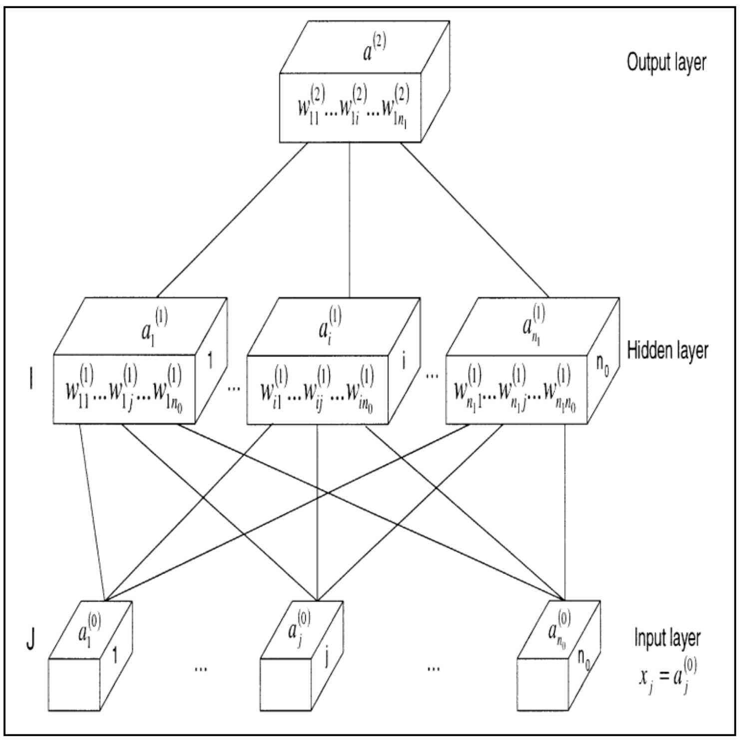

Fig1. Example neural network with single-output neuron and a hidden layer.

In this paper propose a fuzzy extension called Fuzzy Backpropagation (FBP) algorithm. It determinesthe net value through the LR type fuzzy number and thus does not assume independence between the criteria. Another advantage of FBP algorithm is that it reaches always forward to the target value without oscillations and there is no possibility to fall into local minimum. Necessary and su_cient conditions for convergence of FBP algorithm for single-output networks in case of single- and multiple-training patterns are proved. The results of computer simulation are reported and analysed: FBP algorithm shows considerably greater convergence rate compared to SBP algorithm.

A. Standard back propagation algorithm (SBP algorithm)

First we describe the standard back propagation algorithm [17,18] because FBP algorithm proposed by us can be viewed as fuzzy-logic-based extension of the standard one. SBP algorithm is an integrative gradient algorithm designed to minimise the mean-squared error between the actual output and the desired output by modifying network weights. In the following we use for simplicity a network with one hidden layer but the results are valid also for networks with more than one hidden layers.

Let us consider a layered feed forward neural network with n0 inputs, n1 hidden neurons and single-output neuron (cf. Fig. 1). The input {output relation of each neuron of the neural network is defined as follows:

Hidden neurons:

$$net_i^{(1)} = \sum_{j=1}^{n0} w_{ij}^{(1)} x_j \quad (6)$$

$$a_i^{(1)} = f(net_i^{(1)}), i = 1,2, \ldots., n-1 \quad (7)$$

Output neurons:

$$net^{(2)} = \sum_{i=1}^{n1} w_{1i}^{(2)} a_i^{(1)} \quad (8)$$

$$a^{(2)} = f(net^{(2)}) \quad (9)$$

Where X =(x1; x2;: : ; xn0) is the vector of pattern's

inputs.

The activation functions f (net) can be linear ones

or Fermi functions of type

$$f(net) = \frac{1}{1+e^{-4\sigma(net-\sigma)}} \quad (10)$$

Analogous to the writing and reading phases, there are also two phases in the supervised learning BP network. There is a learning (training) phase when a training data set is used to determine the weights that define the neural model. So the task of the BP algorithm is to find the optimal weights to minimise the error between the target value and the actual response. Then the trained neural model will be used later in the retrieving phase to process and evaluate real patterns. Let the pattern's output corresponding to the vector of pattern's inputs X be called the target value t. Then the learning of the neural network for training pattern (X; t) is performed in order to minimize the squared-error between the target and the actual response

$$E = \frac{1}{2}(t - a^{(2)})^2 \quad (11)$$

The weights are changed according to the following formula:

W new

$$w_{ij}^{new} = w_{ij}^{old} + \Delta w_{ij} \quad (12)$$

Where

$$\Delta w_{ij=n\delta_i^{(l)}} a_j^{(l)}$$

$n\varepsilon[0,1]$ denotes the learning rate, _(l) i is the error signal relevant to the ith neuron, and a(l) j the signal at jth neuron input. The error signal is obtained recursively by back propagation.

Output layer

$$\delta^{(2)} = f'(net^{(2)})(t - a^2) \quad (13)$$

Where the derivation of the activation function

$$f'(net) = 4net(1 - net) \quad (14)$$

Hidden layer

$$\delta_i^{(1)} = f'(net_i^{(1)}\delta^{(2)}w_{1i}^{(2)} \quad (15)$$

where (2) is determined by formula (1).

When there are more than one training pattern

(Xm; tm); m=1; 2; : : : ; M, the sum-squared error is defined as

$$E = \frac{1}{2}\sum_{m=1}^{M}(t^m - a_m^{(2)}) \quad (16)$$

B. Fuzzy Backpropagation algorithm (FBP algorithm)

Recently, many neuro-fuzzy models have been introduced [2, 6, 8, 10, and 11]. The following extension of the standard BP algorithm to fuzzy BP algorithm is proposed. For yielding the net value net i the inputs values of the ith neuron are aggregated. The mapping is mathematically described by the fuzzy integral of Sugeno that relies on psychological background

Step1: Randomly generate the initial weight sets w for the input hidden layer where each $w_{ji}=(w_{mji},w_{\alpha ji},w_{\beta ji})$ is an LR-type fuzzy number. Also generate the weight set w' for the hidden output layer

Where $w'_{kj} = (w'_{mkj},w'_{\alpha kj},w'_{\beta kj})$

$w_{ji} = (w_{mji},w_{\alpha ji},w_{\beta ji})$

Where $w'_{kj} = (w'_{mkj}, w'_{\alpha kj},w'_{\beta kj})$

Step2: Let (I_p, D_p) p=1, 20…N input-output pattern set, that fuzzy back propagation needs to be trained with. Here $I_p = (I_{p0}, I_{p1}, I_{pl})$ where each I_{pi} is an LR-type fuzzy number

Step3: Assign values for α and η

Alpha=0.1

Neta =0.9

Step4: Get next pattern set (I_p,D_p).Assign $O_{pi}=I_{pi,}$, i=1,2,3..1

Step5: Compute the input to hidden neurons

$O'_{pj}=f\ (NET_{pj})$,j=1,2,..m; $O'_{p0}=1$.

Where $NET_{pj}=CE\ (\sum w_{ji}o_{pi})$

Step6: compute the hidden to output neurons

$O''_{pk}=f\ (NET'_{pk})$,k=1,2,.. n;

Where $NET'_{pk}= CE\ (\sum w_{kj}o'_{pj})$

Step7: compute change of weights Δ w' (t) for the hidden output layer as follows

Compute

$$\Delta E_p\ (t)=(\partial E_p/\partial w'_{mkj}\ ,\ \partial E_p/\partial w\alpha_{kj}\ ,\ \partial E_p/\partial w'_{\beta kj}\ )$$

Compute

$$\Delta w'\ (t) =-\eta\Delta E_p\ (t)+\alpha\Delta w'(t-1)$$

The update weight iof hidden to output neuron is

$$W'\ (t) =W'\ (t-1) +\Delta W'\ (t)$$

Step 8: compute change of weights Δw (t) for the input hidden layer as follows

Let

$$\delta_{pmk} =-(D_{pk}-O''_{pk})O''_{pk}(1-O''_{pk}).1$$

$$\delta_{pmk} =-(D_{pk}-O''_{pk})O''_{pk}(1-O''_{pk}).(-1/3)$$

$$\delta_{pmk} =-(D_{pk}-O''_{pk})O''_{pk}(1-O''_{pk}).(1/3)$$

Compute

$$\Delta E_p(t)=(\partial E_p/\partial w'_{mji}\ ,\ \partial E_p/\partial w\alpha_{ji}\ ,\ \partial E_p/\partial w'_{\beta ji}\ ).$$

Compute

$$\Delta w'(t)=-\eta\Delta E_p(t)+\alpha\Delta w'(t-1)$$

Compute

$$\Delta w'(t)=-\eta\Delta E_p(t)+\alpha\Delta w'(t-1)$$

Step9: update weight for the input-hidden and hidden-output layer as

$$W\ (t) = W\ (t-1) + \Delta W\ (t)$$

$$W'\ (t) = W'\ (t-1) + \Delta W'\ (t)$$

Step10: p=p+1;

If (p<=N) go to step 5

Step11: COUNT_OF_ITRNS = COUNT_OF_ITRNS +1:

```
If COUNT_OF_ITRNS < ITRNS
{
        Reset Pointer to first pattern in the training set;
        P=1;
        Goto step 5;
}
```

Step12: output w' and w'' the final weight sets.

VIII. Experimental Results

This section presents the results of experimental studies using Fuzzy Neural Network Classification algorithm with the BPN, NaviBayes, Multiple Layer Perceptron, JRip, RBFNS algorithms. All the data sets have been obtained from the images used for the experimental analysis are taken from the mammogram image analysis society(MIAS).It consist of 322 images, which belongs to three big categories: normal, benign and , malignant .there are 209 normal images, 61 a benign and 52 malignant .which are considered abnormal.

Table1 : Correctly classified Accuracy for various Classification Algorithms

S.No	Fuzzy NN	Decision Table	BPN	JRip	RBFN	ML Perceptron
1	82.5%	49.1%	60%	54.7%	53%	53%

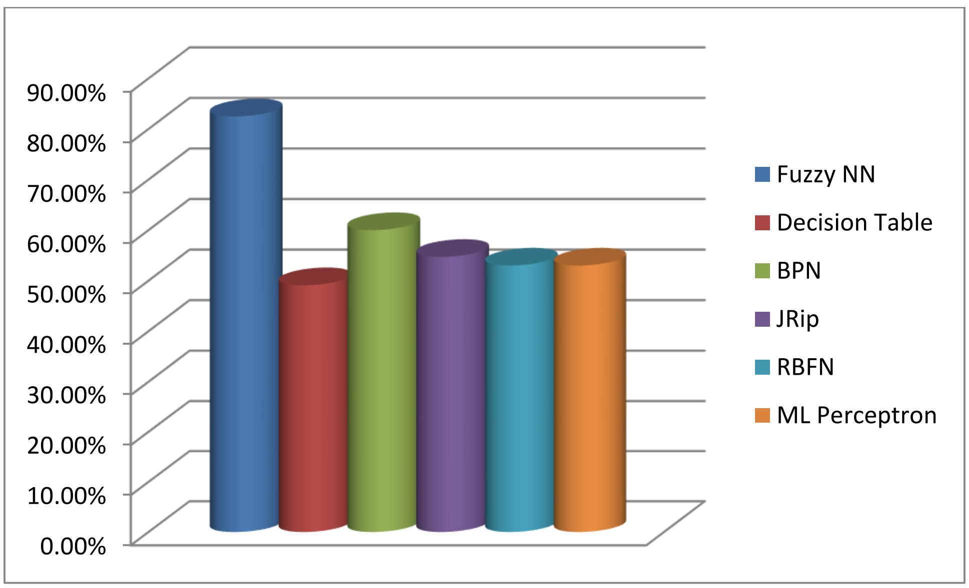

Chart for correctly classified accuracy

IX. Conclusion

In this paper, we addressed the problem of classification using fuzzy neural network. The classification of breast cancer images is based on multi-layered back propagation algorithm. A fuzzy neural network for classifying mammogram are results of screening the mammograms are organized by classification and finally grouped into three categories i.e., Normal, malignant and Benign. Experimental results show that this method performs well with the classification accuracy reaching nearly 90% in comparison with the already existing algorithms .

REFERENCES

[1] J. Angstenberger, M. Fochem, R. Weber, Intelligent dataanalysis { methods and applications, in: H.-J. Zimmermann(Ed.), Proc. 1st Eur. Congr. on Fuzzy and IntelligentTechnologies EUFIT'93, Verlag Augustinius Buchhandlung,Aachen, 1993, pp. 1027 -1030.
[2] M. Brown, C. Harris, Neurofuzzy Adaptive Modelling andControl, Prentice-Hall, New York, 1994.
[3] D. Dubois, H. Prade, Fuzzy Sets and Systems: Theory andApplications, Academic Press, New York, 1980.
[4] G. Grunst, R. Oppermann, C.G. Thomas, Adaptive and adaptable systems, in: P. Hoschka (Ed.), A New Generation of Support Systems, Lawrence Erlbaum Associates, Hillsdale,1996, pp. 29-46.
[5] R. de Hoog et al., The Common KADS model set, ESPRIT Project P5248 KADS-II, University of Amsterdam,Amsterdam, 1993.
[6] E. Ikonen, K. Najim, Fuzzy neural networks and application to the FBC process, IEE Proc. Control Theory Appl.1996, 259-269.
[7] M. Jager, K. Hecktor, W. Laurig, Ergon expert { a knowledge based system for the evaluation and design of manual materials handling, in: Y. Qu_einnec, F. Daniellou (Eds.),Proc. 11th Congr. of IEA Designing for Everyone, Taylor &Francis, London, 1991, pp. 411- 413.
[8] M. Keller, H. Tahani, Backpropagation neural network for fuzzy logic, Inform. Sci. 62 (1992) 205-221.
[9] J.-H. Kirchner, Ergonomic assessment of products: some general considerations, in: Proc. 13th Triennial Congr. International Ergonomics Association, vol. 2, Tampere, Finland, 1997, pp. 56-58.
[10] C.-H. Lin, C.-T. Lin, An ART-based fuzzy adaptive learning control network, IEEE Trans. Fuzzy Systems 5 (4) (1997) 477-496.
[11] D. Nauck, F. Klawonn, R. Kruse, Foundations on Neuro- Fuzzy Systems, Wiley, Chichester, 1997.
[12] S. Neubert, F. Maurer, A tool for model-based knowledge engineering, Proc. 3rd KADS Meeting, Munich, 1993, pp. 97 -112.
[13] A. Nikov, G. Matarazzo, A methodology for human factors analysis of o_ce automation systems, Technol. Forecasting Social Change 44 (2) (1993) 187 -197.
[14] Boss, R. Subash Chandra, K. Thangavel, and D. Arul Pon Daniel. "Mammogram image segmentation using fuzzy clustering." Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on. IEEE, 2012.
[15] A. Nikov, S. Delichev, S. Stoeva, A neuro-fuzzy approach for user interface adaptation implemented in the information system of Bulgarian parliament, in: Proc. 13th Int.Ergonomics Congr. IEA'97, vol. 5, Tampere, Finland, 1997,pp. 100 -112.
[16] R. Oppermann, H. Reiterer, Software evaluation using 9241evaluator, Behaviour Inform. Technol. 16 (4=5) (1997)
[17] S.Rajasekaran,G.A.Vijayalakshmi pai,Neural Networks,Fuzzy Logic,and Genetic Algorithms Synthesis and applications
[18] MIAS database of mammograms