APPLICATION OF FUZZY-MLP MODEL TO ULTRASONIC LIVER IMAGE CLASSIFICATION

Aborisade, David. O,
Ojo, John. A,
Amole, Abraham. O
Department of Electronic & Electrical Engineering, Ladoke Akintola University of Technology, Ogbomoso, Nigeria

Abstract

In this paper, we propose the application of fuzzy-MLP in the classification of ultrasonic liver images. The four sets of ultrasonic liver images used in the experiment are: normal, liver cysts, alcoholic cirrhosis and carcinoma.

To deal with the sample images efficiently, we extract textural features from the Pathology Bearing Regions (PBRs) of the ultrasound liver images. The selected features for the classification are entropy, energy and maximum probability-based texture features extracted using gray level co-occurrence matrix second-order statistics. The fuzzy-MLP model is constructed for the selected features classify various categories of ultrasonic liver images.

The efficacy of Fuzzy-MLP model and conventional artificial neural network (ANN) has been compared on the basis of the same feature vector. A test with 82 training data and 110 test data for all the four classes shows 92.73% classification accuracy for the proposed fuzzy-MLP model. It is compared with the 81.82% counterpart provided by conventional ANN method.

Keywords: Texture features, fuzzy-MLP, and ANN

Introduction

The most common medical imaging techniques for early detection and diagnosis of liver tumors include Computed Tomography (CT), Ultrasound (US), and Magnetic Resonance (MR) (Gunasundari and Janakiraman, 2013). Ultrasound imaging is widely used technique in the diagnosis of soft tissues, due to its ability to visualize human tissue without deleterious effects (Wen-Li *et al*, 2004). More also ultrasound imaging is safe to handle, non-radiological and non-invasive.

One application of diagnostic ultrasound is liver imaging. Liver is an important organ that plays a vital role in human organ system. However, this organ is prone to diseases such as cyst, alcoholic cirrhosis, and carcinoma (Suganya and Rajaram, 2012). Visual analysis of ultrasound liver images often enables the physician or radiologists to detect part of liver tissue which has been affected by diseases and to assess their size. The decision made by radiologist principally depends on their clinical experience, which might be related to extraction of characteristic structures that assume diverse values in the present of normal and diseased states of tissues.

Texture is considered as one of the important characteristic used in identifying objects or regions of interest in an image (Haralick *et al*,1973*)*. It is often described as a set of statistical measures of spatial distribution of grey levels in an image. Since ultrasonic liver tissues present various granular structures as texture, the analysis of ultrasound image is similar to the problem in texture analysis. The most important aspect for ultrasonic liver tissues classification is the extraction of meaningful feature. The grey level co-occurrence matrix (Haralick *et al*,1973) assumes that the texture information in an image is contained in the overall or 'average' spatial relationships which the grey tones have to one another. This second-order statistics scheme has been used in generating texture features (Haralick *et al*,1973*).*

In this paper, we present ultrasonic liver tissues classification framework. First, our system extracts textural features from the Pathology Bearing Regions (PBRs) of the ultrasound liver images using second-order gray level co-occurrence matrix. Next, an object image classifier is constructed by a Fuzzy-MLP model using these texture features. The proposed system achieves significantly higher performance when it is compared to the conventional artificial neural network classification system.

This paper is organized as follows: Section 2 gives an overview on how to generate the GLCM texture features. Section 3 deals with the fuzzy-neural network classifier. The experimental results and conclusion are given in Section 4 and 5 respectively.

GLCM Texture Feature Generation:

In a statistical texture analysis, texture features were computed on the basis of statistical distribution of pixel intensity at a given position relative to others in a matrix of pixel representing image (Aswini *et al,*2013). Depending on the number of pixels or dots in each combination, we have the first-order statistics, second-order statistics or higher-order statistics. Feature extraction based on grey-level co-occurrence matrix (GLCM) is the second-order statistics that can be use to analysing image as a texture (Haralick, 1979; DaPonte *et al*,1988). GLCM (also called gray tone spatial dependency matrix) is a tabulation of the frequencies or how often a combination of pixel brightness values in an image occurs (Soo Beom *et al,* 2004).

In the following, textural features are computed from a set of angular nearest-neighbor grey tone spatial-dependence matrices (Haralick *et al*,1973). The contextual texture information is specified by matrix of relative frequencies p_{ij} with which two neighboring resolution cells, having grey levels i and j and separated by a distance δ, occur in the image. The unnormalized frequencies are defined by the elements $p(i,j,\delta;\theta)$ of a set of co-occurrence matrices, where θ is $0°, 45°, 90°$ and $135°$ for horizontal, right-diagonal, vertical and left-diagonal neighbor pairs respectively. Formally, for angles quantized to $45°$ intervals the unnormalized frequencies are defined by

$$
\begin{aligned}
p(i,j,\delta,0°) &= \#\{((k,l),(m,n)) \in (L_x \times L_y) \times (L_x \times L_y) \,|\, k-m=0, \\
&\qquad |l-n| = d, I(k,l)=i, I(m,n)=j\} \\
p(i,j,\delta,45°) &= \#\{((k,l),(m,n)) \in (L_x \times L_y) \times (L_x \times L_y) \,|\, (k-m=d, l-n=-d \\
&\qquad \text{or}(k-m=-d, l-n=d), I(k,l)=i, I(m,n)=j\} \\
p(i,j,\delta,90°) &= \#\{((k,l),(m,n)) \in (L_x \times L_y) \times (L_x \times L_y) \,|\, (k-m=d, l-n=0 \\
&\qquad l-n=0, I(k,l)=i, I(m,n)=j\} \\
p(i,j,\delta,135°) &= \#\{((k,l),(m,n)) \in (L_x \times L_y) \times (L_x \times L_y) \,|\, (k-m=d, l-n=d \\
&\qquad \text{or}(k-m=-d, l-n=-d), I(k,l)=i, I(m,n)=j\}
\end{aligned}
\tag{1}
$$

where # denote the number of elements in the set. Note that these matrices are symmetric, i.e. $p(i,j,\delta;\theta) = p(j,i,\delta;\theta)$.

Fig. 1 represents the formation of GLCM of the grey-level (4 levels) image at the distance $\delta = 1$ along the four angular specifications;

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

(a)

$$P_H = \begin{matrix} 4 & 2 & 1 & 0 \\ 2 & 4 & 0 & 0 \\ 1 & 0 & 6 & 1 \\ 0 & 0 & 1 & 2 \end{matrix} \quad \text{for } \theta = 0°; \qquad P_{RD} = \begin{matrix} 4 & 1 & 0 & 0 \\ 1 & 2 & 2 & 0 \\ 0 & 2 & 4 & 1 \\ 0 & 0 & 1 & 0 \end{matrix} \quad \text{for } \theta = 45°$$

$$P_V = \begin{matrix} 6 & 0 & 2 & 0 \\ 0 & 4 & 2 & 0 \\ 2 & 2 & 2 & 2 \\ 0 & 0 & 2 & 0 \end{matrix} \quad \text{for } \theta = 90^\circ \quad \text{and} \quad P_{LD} = \begin{matrix} 2 & 1 & 3 & 0 \\ 1 & 2 & 1 & 0 \\ 3 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 \end{matrix} \quad \text{for } \theta = 135^\circ$$

(b)

Fig. 1: (a) Image with 4 grey level image (b) GLCM for $\delta = 1$ and direction $0^\circ, 45^\circ, 90^\circ, \text{and} 135^\circ$.

The four matrices in Fig. 1(b) are used to generate the textural features corresponding to the four angles $\theta = 0^\circ, 45^\circ, 90^\circ, \text{and } 135^\circ$ respectively. For clarity sake, let the matrix $p(i, j)$ be consider without any specification about the angle θ. The features are contrast, correlation, energy, homogeneity, entropy, dissimilarity, maximum probability, and cluster tendency (Rishi and David, 2006; Clausi, 2002).

- Contrast (CON) provides a measure of the local variations in the texture, and is evaluated as

$$\text{Contrast}(CON) \quad \sum_{n=0}^{\mathrm{N_g}\text{-}1} n^2 \left\{ \sum_{|i-j|=n} \frac{P(i, j)}{R} \right\} \tag{2}$$

- Correlation measures linear dependency of gray levels on neighbouring pixels and is defined as

$$\text{Correlation}(COR) \quad \sum_{i,j=0}^{\mathrm{G}\text{-}1} \frac{(i-\mu_x)(j-\mu_y)P_{i\,j}}{\sigma_x \sigma_y} \tag{3}$$

where

$$\mu_x = \sum_{i}^{G-1} iP(i) \tag{4}$$

$$\mu_y = \sum_{j}^{G-1} jP(j) \tag{5}$$

$$\sigma_x = \sum_{i}^{G-1} (i-\mu_x)^2 P(i) \tag{6}$$

$$\sigma_y = \sum_{j}^{G-1} (i-\mu_y)^2 P(j) \tag{7}$$

- Energy function measures the homogeneity of an image, and provides the sum of squared elements in the GLCM. It is computed as

$$\text{Energy}\,(E) \quad \sum_{i,j} P^2{}_{i,j} \tag{8}$$

- Homogeneity function returns a value that measures the closeness of the distribution of elements in the gray level co-occurrence matrix to the diagonal gray level co-occurrence matrix. It is expressed as

$$\text{Homogeneity (HOM)} \quad \sum_{i,j=0} \frac{P_{i,j}}{1+|(i-j)|} \tag{9}$$

- Entropy Measures the randomness of gray-level distribution, and is computed as

$$\text{Entropy}(Ent) \quad \sum_{i,j} \frac{P(i,j)}{R} \log \frac{P(i,j)}{R} \tag{10}$$

- Dissimilarity measure the distance between pairs of objects (pixels) in the region of interest. It is expressed as

$$\text{Dissimilarity}\,(Diss) \quad \sum_i \sum_j |\, i-j \,| \frac{P(i,j)}{R} \tag{11}$$

- Maximum probability determines the most predominant pixel pair in an image and is defined as

$$\text{Maximum probability}\,(MP) \quad \max\, p(i,j) \;\text{ for all } i,j \tag{12}$$

- Cluster Tendency (CT) Measures the grouping of pixels that have similar gray-level values. It is computed as

$$\text{Cluster shade}\;(Cl\_shade) \quad \sum_i \sum_j (\mathrm{i}+\mathrm{j}-\mu_{\mathrm{x}}-\mu_{\mathrm{y}})^3 \frac{P(i,j)}{R} \tag{13}$$

$$\text{Cluster prominence}\;(Cl\_prominence) \quad \sum_i \sum_j (\mathrm{i}+\mathrm{j}-\mu_{\mathrm{x}}-\mu_{\mathrm{y}})^4 \frac{P(i,j)}{R} \tag{14}$$

Each measure is calculated four times, corresponding to each of the four directional co-occurrence matrices. The mean values *CON*, *COR*, *E*, *HOM*, *Ent*, *Diss*, *MP*, *CT* provide a non-directional texture representation. We have

$$\begin{aligned}
CON_k &= \frac{1}{4}(CON_{0^\circ} + CON_{45^\circ} + CON_{90^\circ} + CON_{135^\circ}) \\
COR_k &= \frac{1}{4}(COR_{0^\circ} + COR_{45^\circ} + COR_{90^\circ} + COR_{135^\circ}) \\
E_k &= \frac{1}{4}(E_{0^\circ} + E_{45^\circ} + E_{90^\circ} + E_{135^\circ}) \\
HOM_k &= \frac{1}{4}(HOM_{0^\circ} + HOM_{45^\circ} + HOM_{90^\circ} + HOM_{135^\circ}) \\
Ent_k &= \frac{1}{4}(Ent_{0^\circ} + Ent_{45^\circ} + Ent_{90^\circ} + Ent_{135^\circ}) \\
Diss_k &= \frac{1}{4}(Diss_{0^\circ} + Diss_{45^\circ} + Diss_{90^\circ} + Diss_{135^\circ}) \\
MP_k &= \frac{1}{4}(MP_{0^\circ} + MP_{45^\circ} + MP_{90^\circ} + MP_{135^\circ}) \\
CT_k &= \frac{1}{4}(CT_{0^\circ} + CT_{45^\circ} + CT_{90^\circ} + CT_{135^\circ})
\end{aligned} \tag{15}$$

In an attempt to select the best feature vector for classification of ultrasonic liver tissues in this work, feature variance for each category of liver diseases is computed as (Aborisade and Ibiyemi, 2007)

$$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \tag{16}$$

where, N is the number of training liver samples ultrasound liver images with diseases, $\bar{x}$ is the mean of the feature measurement and x_i is the *i*th actual measurement. It is hard to determine the discriminatory features from the feature variance as the features that have large variance also have large means. Hence the best discriminating features is determined by computing the distance between means of the classes normalized by variances

$$V_{i\,j} = \frac{\left|\bar{x}_i - \bar{x}_j\right|}{\sqrt{\sigma_i^2 + \sigma_j^2}} \tag{17}$$

The smaller this value is, the nearer the class means and the worse the feature for deciding between the classes. The best discriminating feature that aids classification of liver diseases are entropy, energy and maximum probability and are arranged in form of a pattern vectors of the form

$$P = (P_1, P_2, \cdots, \mathrm{P}_n)^T \tag{18}$$

Classifier:

In this section we describe the fuzzy MLP model employed in classifying ultrasonic liver diseases. The architecture of our MLP network, with a back-propagation training algorithm which incorporates concepts from fuzzy sets is shown in Figure 2. The main idea implemented in this work originated from (Pal and Mitra, 1992) in which input feature value is converted to fuzzy components in the input vector, which therefore consists of the membership values on overlapping partitions of linguistic properties such as *low*, *medium*, and *high* corresponding for each input feature. Class membership values (μ) of patterns are represented at the output layer of the fuzzy MLP.

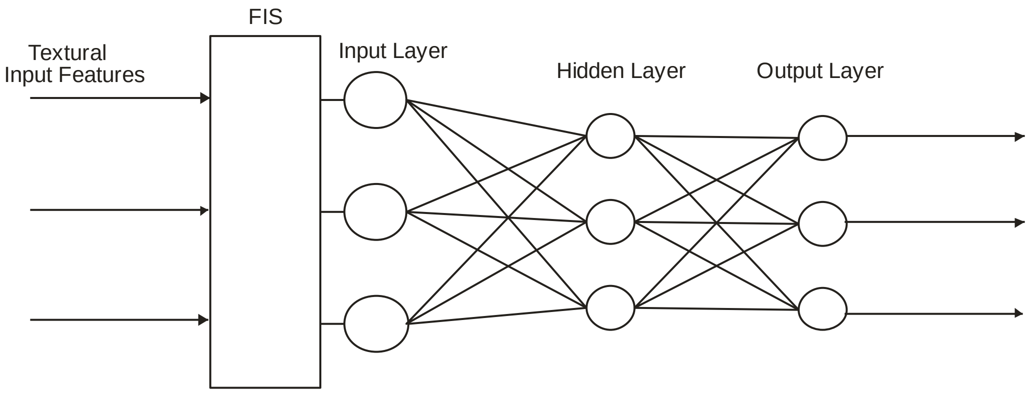


Fig. 2: Multilayer perceptron

In general, the MLP passes through two phases, training and testing. First, a training input pattern is presented to the network input layer. The network propagates the input pattern from layer to layer until the output pattern is generated by the output layer. If this pattern is different from the desired output, an error is calculated and then propagated backward through the network from the output layer to the input layer. The weights are updated by backpropagating errors with respect to these membership values such that the contribution of uncertain vectors is automatically reduced. A three layered feedforward MLP is used. The output of a neuron in any layer other than the input layer as given in (Sushmita *et al*, 1995; Sushmita *et al*, 1997; Sankaret *et al*, 2003) is

$$y_j^{h+1} = \frac{1}{1+\exp(-\sum_i y_i^h \omega_{ij}^h)} \tag{19}$$

where y_i^h is the state of the *i*th neuron in the preceding (*h*)th layer and ω_{ij}^h is the weight of the connection from the *i*th neuron in layer *h* to the *j*th neuron in layer *h*+1. For nodes in the input layer, $y_j^0 = x_j^0$, corresponds to the *j*th component of the input vector. The change in weight vector is being done according to the calculated error in the network, after iterative training. The error and change in weights in the network can be calculated as (Mukul *et al*, 2007)

$$\Delta w_{ho}(n+1) = -\eta \sum_{i=1}^{H} \frac{\partial E}{\partial w_{ho}} + \alpha \Delta w_{ho}(n) \tag{20}$$

$$\Delta w_{ih}(n+1) = -\eta \sum_{i=1}^{N} \frac{\partial E}{\partial w_{ih}} + \alpha \Delta w_{ih}(n) \tag{21}$$

$$E^P = \frac{1}{2} \sum_{m=1}^{M} \left(C_m^P - O_m^{P^O} \right)^2 \tag{22}$$

for $m = 1$ to M output pattern features and $p = 1$ to P presented input patterns and $(C_m^P - O_m^{P^O})^2$ is the squared difference between the actual output value of output layer for pattern P and the target output value.

In back-propagation training algorithm, an input pattern vector ***P*** having ***n*** features as $P_i = [P_{i1}, P_{i2}, P_{i3}, \cdots, P_{in}]$ is represented as a 3*m*-dimensional vector (Pal and Mandal, 1992)

$$\begin{aligned} P_i &= [\mu_{low(P_{i1})}(P_i), \mu_{medium(P_{i1})}(P_i), \mu_{high(P_{i1})}(P_i), \cdots, \mu_{high(P_{im})}(P_i)] \\ &= [y_1^0, y_2^0, \cdots, y_{3m}^0] \end{aligned} \tag{23}$$

where, μ indicate indicates the membership functions of the corresponding linguistic π-sets *low*, *medium*, and *high* along each feature axis and $y_1^0, y_2^0, \cdots, y_{3n}^0$ refer to the activation of the 3*m* neurons in the input layer. Since the input feature is numeric, therefore we use the π-fuzzy sets (in one-dimensional form), with range $[0,1]$, represented as (Pal and Mitra, 1992)

$$\pi(P_j; c, \lambda) = \begin{cases} 2\left(1 - \dfrac{\|P_j - c\|}{\lambda}\right)^2 & \text{for } \dfrac{\lambda}{2} \le \|P_j - c\| \le \lambda \\ 1 - 2\left(\dfrac{\|P_j - c\|}{\lambda}\right)^2 & \text{for } 0 \le \|P_j - c\| \le \dfrac{\lambda}{2} \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

where $\lambda > 0$ is the radius of the π-function with c as the central point.

In the case of *m* -class problem with *n*-dimensional feature space, let the *n*-dimensional vectors O_k and V_k denote the mean and the standard deviation, respectively, of the numerical training data for the *k*th class. The weighted distance, z_{ik}, of the training pattern vector $\vec{P_i}$ from the *k*th class is defined as (Pal and Majumder,1986)

$$z_{ik} = \sqrt{\sum_{j=1}^{n} \left[\frac{P_{i\,j} - O_{k\,j}}{v_{k\,j}} \right]^2} \qquad \text{for } k = 1, \cdots, m \text{ and } j = 1, \cdots, n \tag{25}$$

where $P_{i\,j}$ is the value of the *j*th input feature component of the *i*th pattern point. The weight $1/V_{k\,j}$ is used to take care of the variance of the classes so that a feature with higher variance has less significance in characterizing a class.

The membership of the *i*th pattern to class C_k is defined as (Pal and Majumder,1986)

$$\mu_k(P_i) = \frac{1}{1 + (z_{ik} / f_d)^{f_e}} \qquad \text{for } k = 1, \cdots, m \tag{26}$$

where z_{ik} is the weighted distance from (25) and the positive constants f_d and f_e are the denominational and exponential fuzzy generators controlling the amount of fuzziness in this class-membership set.

Results and Discussion

The image datasets used in the experiment images are provided by St. Gregory's Specialist Clinic and Ultrasound Diagnostic Service, Adeoyo road, Yemetu, Ibadan, LAUTECH Teaching Hospital, Ogbomoso and Several Radiological centre. These were then scanned by AGFA's DUOSCAN scanner with 32-pixel/cm and 8-bit/pixel resolution. The database image contains scans of 200 images of liver diseases; 48 normal liver, 54 liver cysts, 40 alcoholic cirrhosis and 58 carcinoma images. First all images are registered in to the database through intensity based image registration method. Then from the registered images, eight texture features, derived from gray level co-occurrence matrix was extracted from the region of interest (ROI) among the normal and abnormal ultrasound images. By computing the distance between means of the classes normalized by variances; entropy, energy and maximum probability were chosen as the best discriminating input features for the classification of ultrasonic liver disease for every four class cases shown in Fig. 3(a)–(d).

The input data set is divided into 3 different groups of training set, validation set, and testing set to conduct the experiment. Data set size is tabulated in Table 1. With the available data and the experiments conducted, it is found from the result shown in Figure 4 that, conventional artificial neural network with 0.65 as a learning rate for the 98 hidden neuron and 4300 epochs is able to classify the rules with 56% convergence while our proposed fuzzy-MPL network with 1500 epochs classify the rules with 90% convergence.

The performance of the proposed fuzzy-MLP network and convectional artificial neural network were evaluated with 110 test sample of the same feature set namely entropy, energy and maximum probability extracted from the image using the confusion matrix. Results are tabulated in Table 2 and 3. The results are also evaluated by the medical specialists. The human experts examined the performance of the proposed fuzzy-MLP network and convectional artificial neural network for their classification accuracy. In the experiment with 110 test samples for all classes, the proposed fuzzy-MLP network and convectional artificial neural network achieved the classification rate of 92.73% and 81.82% respectively for classification of ultrasound liver images.

Table 1:
Distribution of the available Samples in Training and Testing Sets

Disease Name	Training Set	Validation set	Testing Set
Normal	**20**	**28**	**28**
Liver cysts	**24**	**30**	**30**
Alcoholic cirrhosis	**18**	**22**	**22**
Carcinoma	**20**	**30**	**30**

Table 2:
Confusion Matrix for the Texture Feature using fuzzy-MLP Network

Correct Class	Classified as			
	Normal	**Liver cysts**	**Alcoholic cirrhosis**	**Carcinoma**
Normal	28	0	0	0
Liver cysts	0	27	2	1
Alcoholic cirrhosis	0	1	20	1
Carcinoma	0	2	1	27

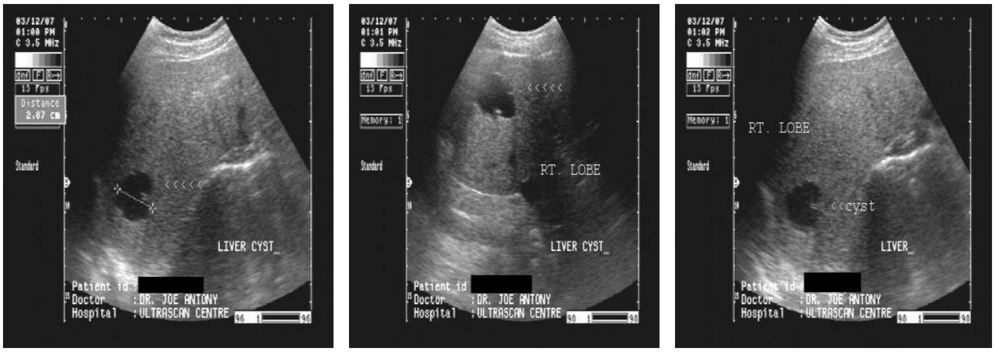
Table 3:
Confusion Matrix for the Texture Feature using convectional artificial neural network

Correct Class	Classified as			
	Normal	Liver cysts	Alcoholic cirrhosis	Carcinoma
Normal	28	0	0	0
Liver cysts	0	24	2	4
Alcoholic cirrhosis	0	4	14	2
Carcinoma	0	5	3	22

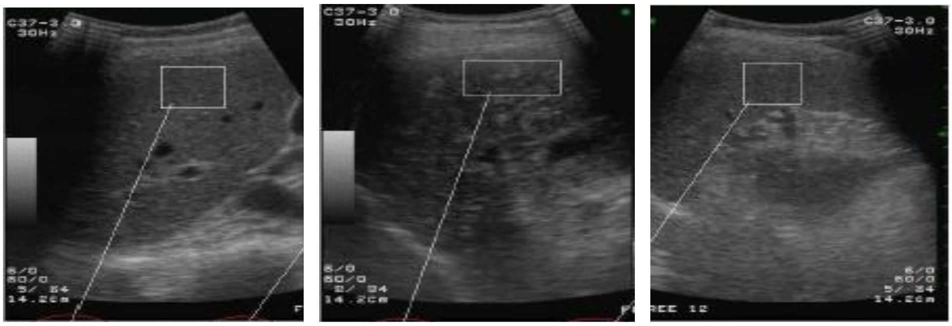
Conclusion

Computer aided classification of ultrasonic liver images using textural features, derived from second order statistics, gray level co-occurrence matrix, obtained from the Pathology Bearing Regions (PBRs) among the normal and abnormal ultrasound images is proposed in this work. Fuzzy MLP neural network classifier was adopted. The model could handle uncertainty both at the input and the output. The input to the network was modeled in terms of linguistic properties *low*, medium, and *high*, using pi-function while the output decision is in terms of class membership values. With the fuzzy MLP neural network, we could build the classification rules very easily by simply training the network with enough number of sample images. Experimental results indicate that the proposed method is effective over the conventional neural network in terms of accuracy, convergence, epochs and has the ability of overcoming the difficulties arising from the problem of overlapping classes.

(a) Normal Liver Image



(b) Liver having cyst



(c) Liver affected by Cirrhosis

(d) Liver affected by Hepatocellular carcinoma (HCC)

Figure 3: Samples of the ultrasonic liver image for the four categories

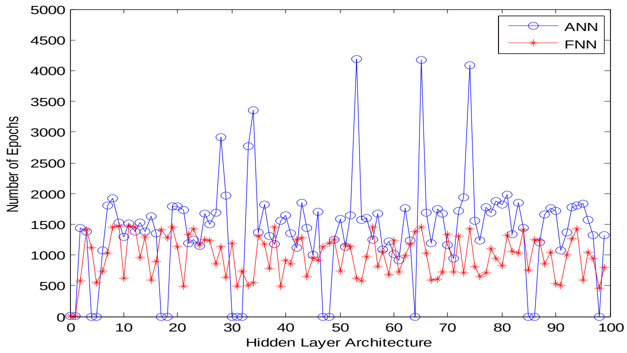


Figure 4: Comparison of fuzzy-MLP and ANN classifiers

References:

Gunasundari S. and Janakiraman S. (2013), "A Study of Textural Analysis Methods for the Diagnosis of Liver Diseases from Abdominal Computed Tomography", International Journal of Computer Applications, Volume 74– No.11.

Wen-Li Lee, Kai-Sheng Hsieh , Yung-Chang Chen, and Ying-Cheng Chen (2004), "A Study of Ultrasonic Liver Images Classification with Artificial Neural Networks Based on Fractal Geometry and Multi-resolution Analysis" Biomedical Engineering- Applications, Basis & Communications,16: 59-67.

Suganya, R. and S. Rajaram (2012),"Content Based Image Retrieval of Ultrasound Liver Diseases Based on Hybrid Approach", American Journal of Applied Sciences 9 (6): 938-945.

R.M. Haralick, K. Shanmugam, I. Dinstein (1973), "Textural features for image classification", IEEE Trans. Syst. Man Cyber.vol. SMC- 3 (6): 610–621.

Aswini, K.M., Swapnasikta, B and Saroj K. L., "Classifying Benign and Malignant Mass using GLCM and GLRLM based Texture Features from Mammogram", International Journal of Engineering Research and Applications, Vol. 1, Issue 3, pp.687-693.

R. Haralick (1979), "Statistical and Structural Approaches to Texture," Proceedings of IEEE 67, pp.786 - 804.

J. S. DaPonte, J. Gelber, and M. D. Fox (1988), "Effect of co occurrence displacement vector on quantization of ultrasonic image texture Bioengineering Conference," in Proceedings of the 1988 Fourteenth Annual Northeast, pp. 298 – 300.

Soo Beom Park, Jae Won Lee, and Sang Kyoon Kim (2004), "Content-based image classification using a neural network", Pattern Recognition Letters, 25: 287–300.

Rishi Jobanputra, David A. Clausi (2006), "Preserving boundaries for image texture segmentation using grey level co-occurring probabilities", Pattern Recognition, 39: 234 – 245.

Clausi, D. A. (2002), "An analysis of co-occurrence texture statistics as a function of grey level quantization", In: Canaadian Journal of Remote Sensing, Vol. 28, No. 1, pp. 45-62.

D.O. Aborisade and T. S. Ibiyemi (2007), "Ceramic Wall Tile Quality Classification Training Algorithms Using Statistical Approach", Research journal of Applied Sciences 2 (12): 1255-1260.

S. K. Pal and S. Mitra (1992), "Multilayer perceptron, fuzzy sets, and classification," IEEE Trans. Neural Networks, vol. 3, pp. 683-697.

Sushmita Mitra, Rajat K. De, and Sankar K. Pal (1997)," Knowledge-Based Fuzzy MLP for Classification and Rule Generation", IEEE Transactions On Neural Networks, Vol. 8, No. 6, pp. 1338- 1350.

Sushmita Mitra and Ludmila I. Kuncheva (1995), "Improving classification performance using fuzzy MLP and two-level selective partitioning of the feature space", Fuzzy Sets and Systems. 70:1-13.

Sankar K. Pal, Sushmita Mitra, and Pabitra Mitra (2003), "Rough-Fuzzy MLP: Modular Evolution, Rule Generation, and Evaluation", IEEE Transactions on Knowledge And Data Engineering, Vol. 15, No. 1, pp. 14-25. January/February 2003.

Mukul Jain, P. K. Butey and Manu Pratap Singh (2007), "Classification of Fuzzy-Based Information using Improved Backpropagation Algorithm of Artificial Neural Networks", International Journal of Computational Intelligence Research.Vol.3, No.3, pp. 267-275.

S. K. Pal and D. P. Mandal (1992), "Linguistic recognition system based on approximate reasoning," *Inform. Sci.*, vol. 61, pp. 135–161.

S. K. Pal and D. D. Majumder (1986), "Fuzzy Mathematical Approach to Pattern Recognition", New York: Wiley, 1986.