

Comparative Study on Cancer Image Diagnosis using Soft Computing Techniques

V.Sivakrithika
M.Tech
SASTRA University
Thanjavur- India

B.Shanthi
Professor, School of Computing
SASTRA University
Thanjavur- India

ABSTRACT

Computer-aided diagnosis system (CAD) can be very helpful for radiologist in detection and diagnosing abnormalities earlier and faster than traditional screening programs. CAD as such employs several techniques to accomplish this task. In this paper, we propose to make a comparative study of two classification methods: One in which we utilize the texture features extracted from the images by directly feeding to the Neural Network based classifier stage to classify the images into benign or malign and in the other hybrid method, those texture features are made to undergo fuzzy discretization before feeding to the Neural Network classifier for the classification. The studies so far conducted using both the systems show that the hybrid system is far superior to the first method in its accuracy. Backward Propagation Network (BPN) algorithm is used in the training stage.

General Terms

Medical Image Mining, Data Mining, Cancer Diagnosis

Keywords

Feature Extraction, Fuzzy, Neural Network

1. INTRODUCTION

The field of medicine has its own computer aided, manual as well as automated, tools for various activities. Diagnosis is the most important aspect of treatment for diseases. Though diagnosis is easy and simple for many diseases there are few diseases that includes cancer which requires much caution because, the fatal diseases are required to be detected and confirmed in real time or at very early stages, since the available treatment methods call for it. Computer Aided Diagnosis[1] is an automated system, which utilizes techniques available in the areas of data mining, digital image processing and radiology. CAD as on today employs different Classification techniques available in different gadgets in data mining for cancer detection. However, there is a wild search for a high tech methodology in data mining [2] for improving the accuracy and speed in classifying the images.

Image classifiers established using neural network architecture achieved accuracy to a great extent. A Back Propagation Feed forward network is an interconnected network in which computing elements are arranged in multilayer. The weight associated with each connection modifies the input before they are fed into threshold element. In training the neural network structure, weights are adjusted using the Back Propagation algorithm. Once the learning phase is over, the network is used to perform the image classification. The proposed neural

network classifiers make use of spatial information of the image known as features for input.

Fuzzy logic deals with uncertainty and impreciseness in various domains. A hybrid neuro-fuzzy system improves the accuracy and speed of the system. When fuzzified data set is given to neural structure the classification accuracy is improved. Fuzzy discretization is a process that characterizes sub ranges of a continuous variable. Fuzzy version of the crisp data set is defined by the degree of membership of crisp attribute to a fuzzy set by a membership function. In the proposed work trapezoidal membership function is used for fuzzifying the feature data set of images before classifying.

In this paper classification method based on neural model and fuzzy-neural model is compared. Extraction of texture features from the images efficiently represents the information about the textural characteristics of the image and the variations in the intensity or gray level. Using this fuzzified feature set and non fuzzified feature set classification is done using BPN algorithm and their performance is evaluated.

The paper is organized as follows: Related work is represented in section 2. Details of the proposed method are described in section 3. Section 4 contains details of experimental procedure and results. Conclusion and future work is presented in section 5.

2. RELATED WORKS

Jesmin Nahar et.al [3] discussed how image data classification plays a vital role in detecting cancer. They proposed a hybrid approach combining both microarray data and image data for early detection of cancer.

Existing CAD system makes use of number of techniques for classifying the medical images. Mining information from the images is the most important fact in classifying the images. Statistical features play a key role in diagnosing the cancer images. Haralick et.al [4] discussed about classification of images using texture features. Texture feature shows the difference in the intensity level which could easily identify the cancer images.

M.Vasantha et.al [5] proposed a method in which texture features are extracted and classification is done using decision tree ID3 algorithm. Since the feature set was not discretized it may affect the accuracy of the classifier. BPN based classifier can provide more accuracy than decision tree algorithm.

Qurat-ul-ain et.al [6] proposed a work that combines both first order and second order features. Features are directly used for classification that leads to inaccuracy. Number of features may increase the computation complexity and minimizes prediction accuracy.

Jenn- Lung Su et.al [7] compared various data analysis techniques and discussed that BPN network classify the images with high accuracy.

Brijesh Verma and John Zakos [8] proposed neural network based classifier and the input feature vector is the combination of selected features. Classifier is tested with different set of features for accuracy. With all features only 72.2% of accuracy is obtained. Number of iterations needed in training phase is also large. Fuzzification could improve the accuracy and reduce the number of iterations.

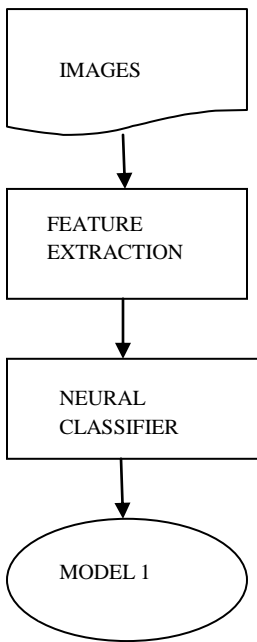
The proposed work contains the following phases:

- Collection of images
- Feature Extraction
- Fuzzy discretization
- Classification using BPN Network

3.1 Feature Extraction

Texture features are effectively estimated from the co-occurrence matrices, the gray-level spatial dependence matrix with good discriminating power. The proposed work makes use of the gray-level co-occurrence matrix (GLCM) that characterize the texture of an image by considering the distribution of intensities and relative positions of pixels in an image. GLCM matrix is created, where each element (i,j) in the resultant matrix is the number of times that the pixel pairs with intensities i and j occur in the specified spatial relationship position in the input image. Holalu Seenappa et.al [9] used statistical features for classifying the breast tissues. Statistical features are calculated using GLCM matrix. These statistics provide information about the texture of an image. Table 1 lists the statistics. The feature values are calculated by applying the statistics formula to the GLCM matrix. The initial model with feature set is extracted at this stage.

MODEL 1



MODEL 2

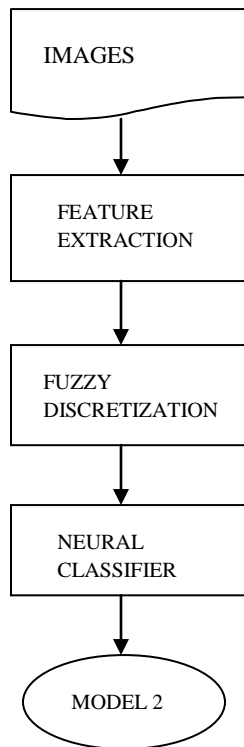


Figure 1 Proposed Models

3. PROPOSED WORK

In this paper we, analyzed two models. Features are extracted from the images. In Model 1, the texture features are directly fed into the neural classifier. In Model 2 the feature set undergo fuzzification before feeding into the network in order to improve the classification capability. Combination of fuzzy with neural reduces the computational complexity. Fuzzy could speed up the system and where as neural improves the accuracy. Performances of the models are evaluated in terms of accuracy.

Table 1 Texture Features

Feature	Equation	Description
Step (f ₁)	$\sum_i \sum_j P_{ij}$	Distribution
Variance (f ₂)	$\sum_i \sum_j (i-j)^2 P(i,j)$	Contrast
Entropy (f ₃)	$\sum_i \sum_j P(i,j) \log(P(i,j))$	Randomness
Energy (f ₄)	$\sum_i \sum_j (P_{ij})^2$	Uniformity
Homogeneity (f ₅)	$\sum_i \sum_j \frac{P(i,j)}{(1+ i-j)}$	Homogeneity
3 rd Moment (f ₆)	$\sum_i \sum_j (i-j)^3 P(i,j)$	Distortion
Inverse Difference (f ₇)	$\sum_i \sum_j \frac{P(i,j)}{(i-j)^2}$	Inverse Contrast

3.2 Fuzzy Discretization

The crisp data set leads to uncertainty and loss of information at the boundaries of ranges. Karla Taboada et.al [10] in their proposed work adopted Fuzzy set theory. The proposed work makes use of combination of trapezoidal and triangular fuzzy membership functions. For each feature fuzzy membership function is calculated using the formula

$$\alpha = 2 \beta - \gamma$$

Where α = lower end range
 β = average value of the feature f_i in the database
 Where f is the feature and $i=1$ to 8
 γ = the largest value of the feature f_i in the database

The values of the continuous feature is discretized into three linguistic terms LOW,MEDIUM,HIGH .The probability of the crisp dataset value with respect to the membership function is known as fuzzy categorical value of the features. Highest membership value is assigned to the feature. Fig.1 shows the fuzzy membership function. More over fuzzy discretization technique provides high accuracy while classifying the test images.

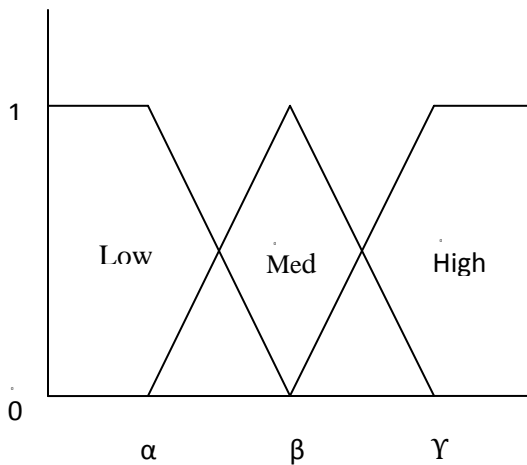


Figure 2 Fuzzy Membership Functions

3.3 Neural Network Classifier

The proposed method used a feed forward BPN for neural network based classification. Fig.3 shows the architecture of BPN. The BPN architecture has only one hidden layer. The number of input nodes depends upon the feature set extracted. The weight is initially assigned and continually adjusted in an attempt to achieve optimum classification rates. Activation function maps the output of the summing junction into the final output. A value of less than 0.5 is labeled as 0 and the network classify the input image features as benign images. A value of more than 0.5 is labeled as 1 and the neural network classify the input image features as malign images. In training the BPN, the desired output was specified as 0 for benign and 1 for malignant.

4. EXPERIMENTAL RESULTS

Image database consists of 15 images including benign and malign images. Out of 15 images 6 are malign and 9 are benign images.BPN network is trained with all the 15 images .Features

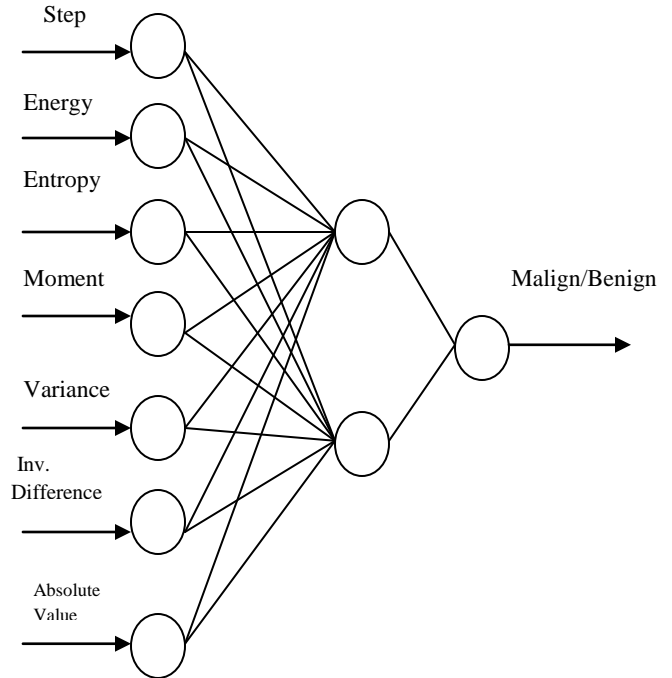


Figure 3 BPN Structure

are extracted from the images and fuzzified. Table 2 shows the fuzzy version of the features extracted for 5 images only. All the eight features form the feature vector for all the 15 images. The extracted features are directly fed into the BPN network and the classifier performance is shown in Fig 4(a).

Table 2 Fuzzy Version of Extracted Feature

IMG No.	Energy	Entropy	Contrast	Moment	Homogeneity	Absolute value	Inverse Difference	Step
1	1	1	.7	.6	1	1	1	1
2	.6	.6	1	.6	.9	1	.7	.7
3	1	1	.9	.6	.5	1	1	1
4	.8	.8	.8	.7	.9	1	.7	.7
5	.1	1	.7	.7	.9	.7	1	1

In the other model the extracted features are fuzzified and the fuzzy version of the feature set is fed into the network. The performance of BPN network for fuzzy version is shown in Fig 4(b). As a first experimental setup, non fuzzified data set is fed into the BPN structure and 47% of accuracy is obtained, and on other experimental setup fuzzified feature vector is fed into network and 73.6% of accuracy is obtained.

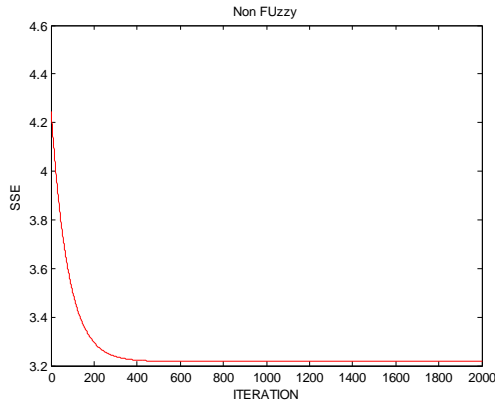


Figure 4 (a) Learning Rate of BPN for Non Fuzzy

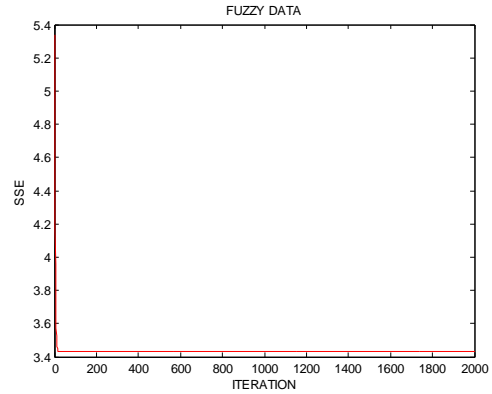


Figure 4 (b) Learning Rate of BPN for Fuzzy

5. CONCLUSION

5.1 Performance Analysis

Performance of classifier is measured by analyzing how it recognize the different classes. Confusion matrix is a tool that is used to measure the performance of the classifier .Confusion matrix represents how far the classifier label correctly and how far it mislabel the classes. It is a table with two rows and two columns in binary classification that reports the number of True Negative (TN), False Positive (FP), False Negative (FN), and True Positive (TP).In the proposed work, system recognizing the malign images correctly is considered as true positive. Confusion matrix is created using the experimental results as shown in Table 3(a) for Non Fuzzy and table 3(b) for Fuzzy.

Table 3 (a) Confusion Matrix For Non Fuzzy

ACTUAL CLASS	PREDICTED CLASS	
	Malign	Benign
Malign (pos)	6 (TP)	3 (FN)
Benign (neg)	4 (FP)	2 (TN)

Table 3(b) Confusion Matrix For Fuzzy

ACTUAL CLASS	PREDICTED CLASS	
	Malign	Benign
Malign (pos)	5 (TP)	4 (FN)
Benign (neg)	0 (FP)	6 (TN)

Sensitivity measures the proportion of malign images correctly identified as malign. Specificity measures the proportion of negatives which are correctly identified that is correct identification of benign images. Accuracy measures the probability or proportion of images correctly classified. Accuracy, Sensitivity, Specificity are calculated using the formula stated in the Table 4, taking values from the confusion matrix. $N(TN)$ is the number of True negatives and $N(TP)$ is the number of True Positives. Positives (pos) are the total number of malign images. Negatives (neg) are the total number of benign images in the database

Table 4 Performance Measures

Performance Measure	Formula
Specificity	$\frac{N(TN)}{N(TN) + N(TP)}$
Sensitivity	$\frac{N(TP)}{N(TN) + N(TP)}$
Accuracy	Sensitivity $\frac{pos}{pos+neg}$ +Specificity $\frac{neg}{pos+neg}$

Performance of the classifier is analyzed in terms of accuracy, sensitivity and specificity and the results are shown in Table 3(c).

Table 3(c) Performance of Fuzzy and Non Fuzzy

Method	Iterations	Accuracy	Sensitivity	Specificity
Fuzzy	2000	73.6%	57%	100%
Non Fuzzy	2000	47%	67%	33%

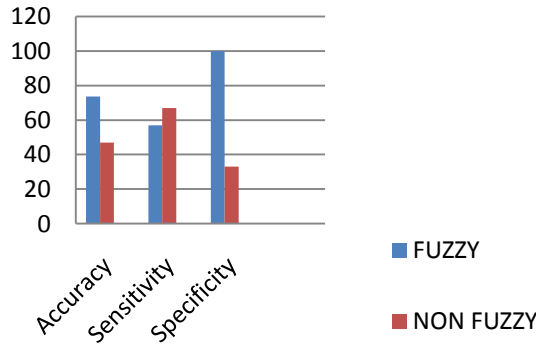


Figure 5 Graph Shows Performance of Classifier

Performance on the basis of confusion matrix is analyzed and the results are shown in Fig 5. Desired Output (DO) is compared with Fuzzy output and Non Fuzzy output in Fig.6. Fuzzy output matches with desired output mostly than Non Fuzzy in less number of iterations.

5.2 Future Work

In the comparative study, we achieved good accuracy for the hybrid model. The accuracy could be improved by including more number of sample images in database. Performance of the BPN network shows better results by making use of 70% of the sample images for training and 30% of images for testing .The proposed work consists of only statistical texture feature set. If spectral and structural features are also included then the accuracy could be increased. But when the number of features is increased it may lead to computation complexity. Therefore we decided to make use of a feature selection algorithm to select only most relevant feature set. As a future work, Genetic Network Programming (GNP) can be used for optimization and to improve learning accuracy adaptive neural network could be used.

6. REFERENCES

- [1] Jinshan Tang, M. Rangayyan,, Jun Xu, Issam El Naqa, Yongyi Yang, "Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances", IEEE Transactions On Information Technology In Biomedicine, Vol. 13, No. 2, March 2009.
- [2] J.Han,M.Kamber , Data Mining : Concepts and Techniques , Morgan Kaufmann Publishers, San Francisco ,2000.
- [3] Jesmin Nahar, Kevin S. Tickle, A B M Shawkat Ali, Yi-Ping Phoebe Chen", Early Breast Cancer

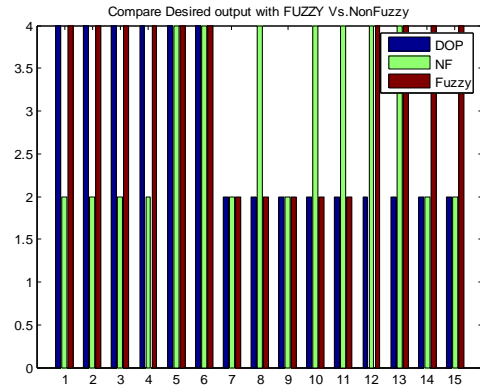


Figure 6 Comparison of DO Vs Fuzzy Vs Non Fuzzy

Identification: Which Way to Go? Microarray or Image Based Computer Aided Diagnosis! ", 2009 Third International Conference on Network and System Security, IEEE.

- [4] Robert .M, Haralick .K "Textural Features For Image Classification"IEEE Trans. On Systems, Man and Cybernetics, Vol. SMC-3, No.6, November 1973.
- [5] M.Vasantha,R.Dhamodharan, "Medical Image Feature, Extraction, Selection And Classification", International Journal Of Engineering Science And Technology Vol. 2(6), 2010, 2071-2076.
- [6] Qurat-UI-Ain, Ghazanfar Latif, Sidra Batool Kazmi, M. Arfan Jaffar, Anwar M. Mirza "Classification And Segmentation Of Brain Tumor Using Texture Analysis", Recent Advances In Artificial Intelligence, Knowledge Engineering And Data Bases.
- [7] Jenn-Lung Su, Guo-Zhen Wu, I-Pin Chao," The Approach Of Data Mining Methods For Medical Database", 2001 Proceedings of the 23rd Annual EMBS International Conference.
- [8] Brijesh Verma and John Zakos "A Computer-Aided Diagnosis System for Digital Mammograms Based on Fuzzy-Neural and Feature Extraction Techniques"IEEE Transactions On Information Technology In Biomedicine, Vol. 5,No.1, March 2001.
- [9] Holalu Seenappa Sheshadri and Arumugam Kandaswamy,"Breast Tissue Classification Using Statistical Feature Extraction Of Mammograms", Vol.23 No.3,2006.
- [10] KarlaTaboada, EloyGonzales, KaoruShimada ,ShingoMabu and KotaroHirasawa ,"Discovering Fuzzy Classification Rule Using Genetic Network Programming", SICE Annual Conference 2008.