

Comparison and Evaluation of BPN and SVM Classifier to Diagnose Liver Lesion Using CT Image

A.Mohanapriya¹, S.Malathi²

¹Assistant Professor, Department of Biomedical Engineering, Dr N.G.P Institute of Technology, Coimbatore, India

²Assistant Professor, Department of Biomedical Engineering, Dr N.G.P Institute of Technology, Coimbatore, India

Abstract— Liver cancer is caused due to chronic hepatitis B or C infections, obesity, alcohol, diabetes and hemochromatosis. Computed Tomography (CT) helps the physician to visualize the structures of internal organs of human body. Pre-processing the image using median filter in order to reduce noise in CT images. Segmentation of tumor alone from CT images is difficult. It is essential to segment the liver from CT image first and then segment the tumour from it. For liver segmentation morphological operations is used and tumor is segmented by using fuzzy c means algorithm. From the segmented lesion the features are extracted by using gray level co occurrence matrix (GLCM). The obtained feature was used to train neural network such as Back propagation Neural Network (BPN), support vector machine (SVM) to classify the disease whether hematoma, hemangioma. Comparison of classifiers is done and found out the SVM classifier gives better accuracy than BPN classifier.

Index Terms — *computed tomography, fuzzy c means, gray level co occurrence matrix, back propagation network, support vector machine.*

I. INTRODUCTION

Liver is the organ in the upper abdomen that helps in digestion and removes waste products from the blood. The liver is the largest solid organ and largest gland in the body. Tumor is a condition affected to a person such that their cells would deform due to various changes in the body and thus lead to cancer. Several types of tumor can develop in the liver because the liver is made of various types of cell. Mainly tumors are of two types benign and malignant. Benign tumors (Hemangioma) is a non cancerous and malignant tumors (Hepatoma) is a cancerous Hepatocellular carcinoma is the primary liver cancer that arises from liver due to liver made up of different types of cells types(for example bile ducts, blood vessels, digestion).the liver cells make up 70% of the liver tissue. Thus the majority of primary liver cancer (over90% to 95%) arises from liver cells and it is called Hepatocellular carcinoma or primary liver cancer [2].

The most commonly used medical imaging technique for detection and diagnosis of liver tumor include ultrasonography (US), computed tomography (CT), magnetic resonance imaging (MRI), X-ray. CT is more preferred than other methods because it gives more detailed information and it helps the radiologist to analyze and evaluate the tumor in a short time.

CT images are widely used in medical diagnosis for its non-intervention, non-invasive, multi-parameter imaging, however, the complexity of human anatomy structure, scrambling of tissues and organs, individual differences, together with the image quality by kinds of factors, make image segmentation become a difficulty [9].

The most commonly used classifier for liver tumor diagnosis is artificial neural networks. Artificial neural networks are a computational tool, based on the properties of biological neural systems. BPN is a feed forward neural network with supervised learning process and it is used for liver tumor classification. SVM is the supervised learning technique. SVM used to classify the benign and malignant tumors.

Digital image data are copied in digital imaging and communications in medicine (DICOM) format and it can be analyzed in various steps such as preprocessing, segmentation, feature extraction, classification. DICOM is a standard for handling, storing, printing, and transmitting information in medical imaging.

II. METHODOLOGY

Liver segmentation based on CT image is a challenging task due to the presence of similar intensity objects in the abdomen with no clear delineation between these objects and the liver [3]. The input CT images which taken from the database. The image is preprocessed by using median filter, the liver segmentation consist of two steps. First the liver is segmented by using morphological operations and the tumor is segmented from segmented liver by using fuzzy c means (FCM) algorithm. The segmented tumor is fed to the feature extraction the features are extracted by using gray level co occurrence matrix (GLCM). The extracted features are used to train various neural networks such as BPN, SVM. By comparisons we can identify best feature and neural network combination for classification. The comparison can be done by evaluating accuracy, sensitivity, specificity, PPV (positive prediction value), NPV (negative prediction value).

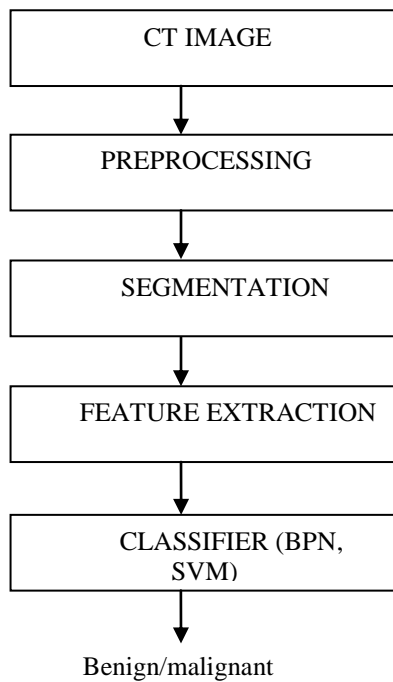


Fig 1 Liver Tumor Prediction System

A. Median filter

Linear filters are generally used to reduce noise in CT images. Here, the neighboring pixels represent the additional Samples of the same value as that of reference pixel. In linear filters the convolution process is used for implementing the neighboring kernels as neighborhood function. But this may lead to the blurring of edges. To overcome such a problem non-linear filters are used for noise reduction. These filters help to preserve edges. Median filter is an example of non-linear filters. In median filter, the ranking of the neighboring pixels is done according to the intensity or brightness level and value of the pixel under evaluation is replaced by the median value of surrounding pixel values [6].

The images distorted or blurred by shot or impulse noise can excellently denoise using this filter. Median filters have many advantages over smoothing filters [7]:

- In median filter the output values consist of only those present in the neighborhood (median value) so there is no reduction in contrast across the steps.
- The boundaries are also not shifted when median filter is used.
- The edges are minimum degraded and hence median filter can be repeatedly applied.

B. Segmentation of liver

Normally, the liver is located on the upper left side of the image and takes up the largest area among the various organs included in the abdominal image. Even though, the liver region maintains a constant intensity throughout, a fixed threshold is not possible because the intensity differs according to the patient slice and the CT machine [4].

Therefore, a system has to be developed to extract the liver automatically with morphological operations.

C. Segmentation of lesion

For lesion segmentation fuzzy c means algorithm (FCM) is used. FCM is used to segment the tumor from segmented liver. The pixels of the input image are divided into 3 clusters. The first cluster includes pixels in the background. The second cluster includes pixels in the liver other than lesion and the third cluster includes pixels in the lesion [1].

D. Feature extraction

Features are extracted from the segmented areas of tumor. Features are individual measurable properties. When the input data to the algorithm is too large to be processed then the input data is transformed to the set of features. Transforming data into set of features called feature extraction [2]. The extracted feature gives the detailed information from the input data in order to perform desired output. Features are extracted by using Gray level co occurrence matrix. The Gray level co occurrence matrix is also called grey tone dependencies matrix. A co occurrence matrix or co occurrence distribution is a matrix or distribution that defined over a image to be a distribution of co occurrence matrix at a given offset [2]. A pair of matrix is used to generate co occurrence matrix through the texture image measure. GLCM method is used to extract second order texture features. The texture characteristics of each segmented areas corresponds to second order statistics. Various features such as autocorrelation, contrast, correlation matrix, cluster performance, cluster shade, energy, entropy, homogeneity, sum average are calculated from segmented images. the features are extracted and fed to the classifier for correct classification.

E. Support vector machine

Support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis[5]. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. The extracted features of images are given as inputs to SVM. SVM is used to classify the liver tumour images as either benign or malignant depending on the feature values. Support vector machines (SVMs) are a set of related supervised learning methods used for classification and regression. It is described in detail by (Vapnik 1998), an SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.

F. Back propagation network

BPN consists of three layers: input layer, hidden layer and output layer. The hidden layer and the output layer have biases and weight. The weight is denoted as w and the bias as b^1 and b^2 as shown in Fig 1. The transfer function used

in hidden layer is transig function and for output layer is purelin function. The output of bias is always 1. The input is feed forward into the network and the error value is computed. From the output layer, the error is back propagated. The weight and bias values are updated according to the back propagated error value.

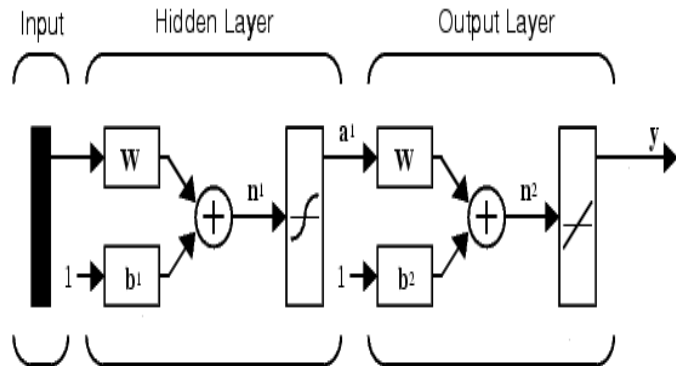


Fig 2 Back Propagation Network

G. Training Phase

All feature vectors that have been generated from the feature extraction process will be trained repeatedly in the network to get the best weight value. When the Mean Square Error (MSE) value was smaller than the value of tolerance. These weights then will be used in the validation and testing phase. The configuration of neural network parameters which have given in the training process in this process is three layers (one input layer, one hidden layer, and one output layer). Each weight will be stored for use at validation and final testing phase.

H. Validation Phase

The validation phase will be used to determine the level of success and error recognition of the ANN based tool for measuring the severity of tumour prior to testing. The network process consists of feed forward only step, which will provide a direct output of data in the form of predictive learning and non-learning are included.

I. Testing Phase

The testing phase is same as validation phase, testing phase also just run the feed forward of testing data using the weights from the previous training process [11].

III. PERFORMANCE EVALUTION

A Performance of networks for various features is evaluated by calculating Accuracy, Sensitivity, Specificity, Positive predictive value (PPV) and Negative predictive value (NPV). Formulas are listed in table I [1]. When combining each neural network such as BPN, SVM different grouping of texture features such as auto correlation, entropy, energy, cluster shade, cluster prominence, homogeneity matrix,

homogeneity parameter, sum average, correlation matrix, correlation parameter, and contrast is united for GLCM. Best combination of suitable texture features for GLCM selected by using following formulas. The performance of the SVM is compared with the performance of the back propagation neural network (BPN) Neural Network [9] to prove the choice of the classifier.

TABLE 1

FORMULAS FOR EVALUATING THE PERFORMANCE OF CLASSIFIER

| METRICES | FORMULA |
|-------------|-------------------------------------|
| ACCURACY | $\frac{TP + TN}{TP + FP + FN + TN}$ |
| SENSITIVITY | $TP / (TP + FN)$ |
| SPECIFICITY | $TN / (TN + FP)$ |
| PPV | $TP / (TP + FP)$ |
| NPV | $TN / (TN + FN)$ |

True Positive (TP): lesions called hepatoma and prove to be hepatoma.

False Positive (FP): lesions called hepatoma that proves to be hemangioma.

False Negative (FN): lesions that are called hemangioma and prove to be hepatoma.

True Negative (TN): lesions that are called hemangioma and prove to be hemangioma.

PPV - Total number of positive instances.

NPV - Total number of negative instances.

IV. RESULTS AND DISCUSSION

Images acquired were in DICOM (Digital Imaging and Communications in Medicine) format. Hemangioma, hematoma images are considered for the work. The system was implemented using the functions available in MATLAB. Fig. 3 is a CT scan image of a patient with a liver tumor. The liver is located along the left half of the image and is light gray in color. The white tubular structures are the normal liver blood vessels. The Liver region after the morphological operations and Segmented Tumor region after Fuzzy C Means Clustering Technique are given in Fig. 3 and Fig. 4

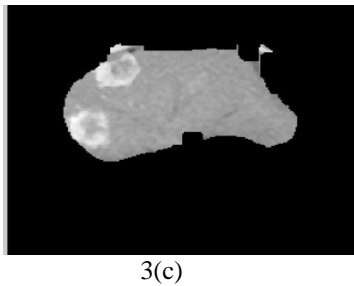
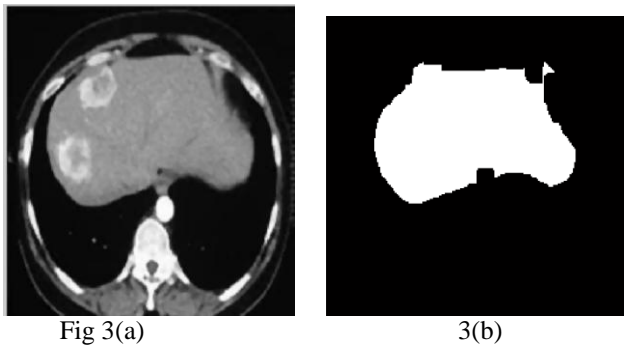


Fig.3 segmented liver region. (a)Original Image, (b)morphological image, (c)Liver region

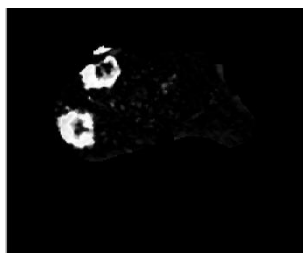


Fig.4 segmented lesion image

After extracting the tumor, the features were extracted as given in the feature extraction algorithm and the optimized feature set was selected by GLCM based on the classification performance of BPN, SVM. From the experiments conducted for feature selection, it is found that the optimal feature set which gives good classification performance are Contrast, Entropy, Homogeneity, autocorrelation, cluster shade, cluster performance, energy, sum average. Hence for this application domain, the above 10 features are enough for classifying the tumor as hepatoma or hemangioma.

TABLE 2
Features for selection

| FEATURES | BENIGN | MALIGNANT |
|--------------------|---------|-----------|
| Autocorrelation | 11.9338 | 18.0091 |
| Contrast | 0.0453 | 0.0603 |
| Correlation matrix | 0.9648 | 0.9825 |
| Cluster shade | 5.8082 | -12.6273 |
| Energy | 0.6985 | 0.9749 |
| Homogeneity | 0.9885 | 0.9884 |
| Sum average | 6.7268 | 8.0801 |

The classifier offers the best results in identifying the malignant liver tumor and benign liver tumor. In some cases the benign tumors are misclassified as normal liver[10]. The intensity of benign and malignant liver tumor is same .

The performance measures are used to access the accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV). The accuracy of a measurement system is the degree of closeness of measurements of a quantity to that quantity's actual (true) value. Sensitivity and specificity are statistical measures of the performance of a binary classification test, also known in statistics as classification function. Sensitivity (also called the true positive rate) measures the proportion of actual positives which are correctly identified as such (e.g. the percentage of sick people who are correctly identified as having the condition). Specificity (sometimes called the true negative rate) measures the proportion of negatives which are correctly identified as such (e.g. the percentage of healthy people who are correctly identified as not having the condition).The performance measures are given in table 2 and table 3.

Training data set is used to adjust the weights on neural networks. Validation data set is used to minimize over fitting. The validation sets are used during the training. For each validation data instance accuracy can be calculated over the validation data. The validation performance for benign tumor is 0.4954 at epoch 2 which shows in Fig 5 and the validation performance for malignant tumor is 0.25008 at epoch 2 which shows in Fig 6.

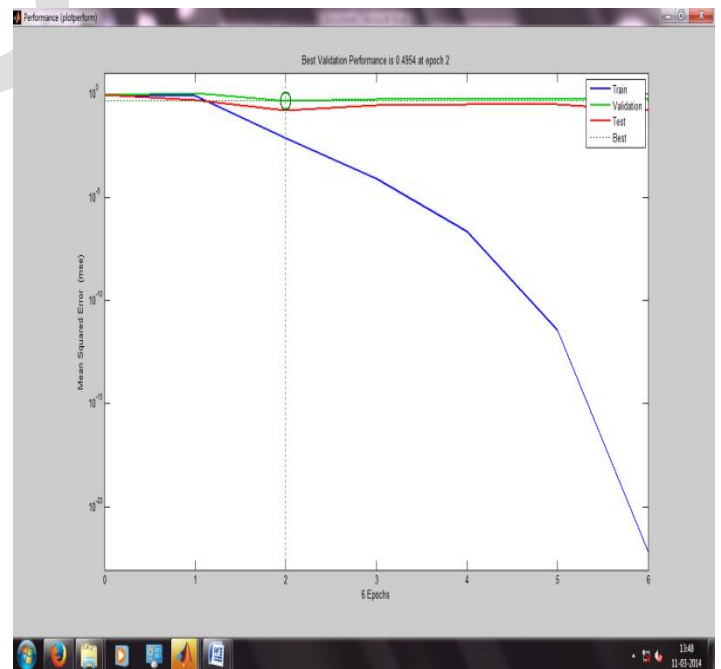
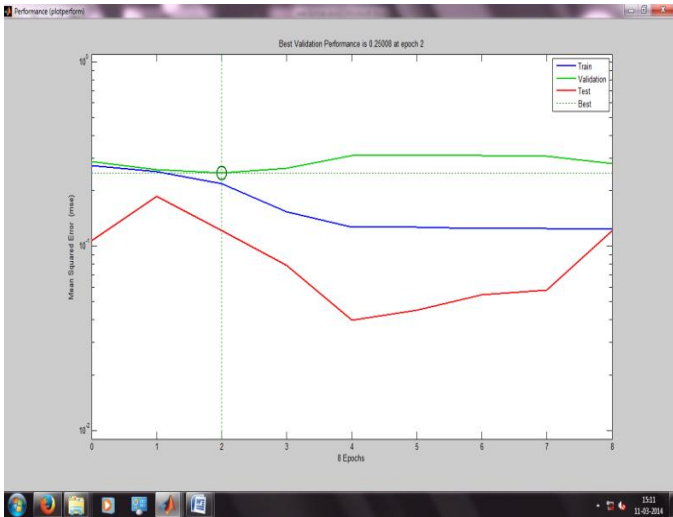


Fig 5 Best Validation Performance for Benign tumor



VI. REFERENCES

- [1] S.Gunasundari, M. Suganya Ananthi, M.: Comparison and Evaluation of Methods for Liver Tumor Classification from CT Datasets. International Journal of Computer Applications (0975 – 8887), Volume 39– No.18, February 2012
- [2] Rahmath.S. Hameed, S.S.kumar.: Assessment of neural network based classifier to diagnose focal liver lesion using CT images 38 elesiver(2012) 4048 – 4056.
- [3] K. Geetha & S. Poonguzhali.: Diagnosis of Liver Tumor Using 3D Segmentation Method for Selective Internal Radiation Therapy, International Journal of Advanced Electrical and Electronics Engineering (IJAEEE), ISSN (Print): 2278-8948, Volume-2, Issue-2, 2013
- [4] Mala. K., Sadasivam V, and Alagappan. S: Neural Network based Texture Analysis of Liver Tumor from Computed Tomography Images. In: International Journal of Biological and Life Sciences 2:1 2006
- [5] Swati P. Tidke, Vrishali A., Chakkarwar.: Classification of Lung Tumor Using SVM, International Journal Of Computational Engineering Research (ijceronline.com) Vol. 2 Issue. 5
- [6] Tarandeep Chhabra, Geetika Dua, Tripti Malhotra, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 2, Issue 7, July 2013
- [7] Sin Hoong Teoh, "exploration of current trend on median filtering methods utilized in digital grayscale image processing", IJMMM 2013, vol 1 (1):50-54 ISSN: 1793-8198.
- [8] Berrichi Fatima Zohra, Benyettou Mohamed, "Automated diagnosis of retinal images using the Support Vector Machine(SVM)", International Journal of Computer Applications(2012).
- [9] Jianhua Liu, Zhongyi Wang, Rui Zhang, "Liver Cancer CT Image Segmentation Methods based on Watershed Algorithm", IEEE(2009).
- [10] V.Ulagamuthalvi, D.Sridharan, "Automatic Identification of Ultrasound Liver Cancer Tumor Using Support Vector Machine", International Journal on Emerging Trends in Computer and Electronics Engineering(ICETCEE)
- [11] Dian Pratiwi, Diaz D. Santika, and Bens Pardamean, "An Application Of Backpropagation Artificial Neural Network Method for Measuring The Severity of Osteoarthritis", International Journal of Engineering & Technology IJET-IJENS Vol: 11 No: 03.

Fig 6 Best Validation Performance for Malignant Tumor

TABLE 3

PERFORMANCE MEASURES FOR SVM

| METRICS | SVM |
|-------------|-----|
| ACCURACY | 98% |
| SENSITIVITY | 98% |
| SPECIFICITY | 98% |
| PPV | 98% |
| NPV | 98% |

TABLE 4

PERFORMANCE MEASURES FOR BPN

| METRICS | BPN |
|-------------|-----|
| ACCURACY | 96% |
| SENSITIVITY | 98% |
| SPECIFICITY | 94% |
| PPV | 94% |
| NPV | 98% |

The total percentage of correctly classified is 98% by SVM and 96% by BPN. SVM classifier gives better accuracy.

V. CONCLUSION

Liver tumor is identified from CT images using GLCM and neural network. Two kinds of liver disease are identified like hematoma and hemangioma. The above results shows that the classification accuracy of SVM classifier higher than BPN classifier. The performance measures can be increased by no of samples. The work can be extended to diagnosis of other types of liver diseases.