ABSTRACT

Lung cancer is the leading cause of cancer deaths in both women and men. It is estimated that 1.2 million people are diagnosed with this disease every year and about 1.1 million people are dying of this disease yearly. The survival rate is higher if the cancer is detected at early stages. The early detection of lung cancer is not an easy task. About 80% patients are diagnosed correctly at the middle or advanced stage of cancer. Computer aided diagnosis system is very helpful for radiologist in detection and diagnosing abnormalities earlier and faster. The computer aided diagnosis is a second opinion for radiologists before suggesting a biopsy test.

In this thesis, image processing techniques are proposed and discussed to detect and locate cancer nodules present in Computed Tomography (CT) images of lungs. The main objectives of the research work is to segment the lung lobe from CT image and classify as cancerous or non-cancerous, to detect the location of the cancer nodule and to classify the nodule as benign or malignant with better sensitivity, specificity and accuracy compared to those of the existing techniques. In this thesis, four classification techniques are proposed for classification of CT image as cancerous or non-cancerous and two detection methods to detect and locate the nodule from the image are proposed. The lung lobes are segmented from the CT image using morphological operations. The lung lobes are classified as cancerous or non-cancerous by Neural Network, Fuzzy Logic, Content Based Image Retrieval and Clustering Techniques. The parameters used for classification are statistical parameters and Gray Level Co-Occurrence Matrix (GLCM) parameters. The location of cancer nodule is detected by Radon, Ridgelet transforms and Genetic Algorithm. The nodules are segmented from the

image using Curvelet transform. The nodules are classified as benign or malignant based on shape features and GLCM features. The classification is done by neural network.

The images are collected from a database of Lung Image Database Consortium (LIDC) and also from reputed hospital. CT images of 200 patients are collected including both men and women. In the dataset 115 images are with cancer nodules and 85 without cancer nodules. Both primary and secondary stage cancer nodules (classified by two radiologists depending on the size of the nodules) with three kinds of nodules; well-circumscribed nodule, juxta pleural nodule and vascularized nodule are considered in the work.

Neural Networks are effective tools in the field of pattern classification, using training and testing data to build a model. The neural network derives the power for classification due to their massively parallel structure and also the ability to learn from experience. The segmented lung lobes are classification by neural networks. The statistical parameters are extracted from the images. The parameters used are mean, standard deviation, skewness, kurtosis, fifth central moment and sixth central moment are used for classification. The classification process is done by feed forward and back propagation neural networks. Compared to feed forward networks the back propagation network gives better classification. Compared to other parameters, skewness gives the maximum classification accuracy. Thirteen training algorithms are used for training the dataset in back propagation network. Among the thirteen training algorithms, the Traingdx algorithm gives the maximum classification accuracy. A new training algorithm with modified momentum factor and learning rate is proposed. The proposed training algorithm gives better classification accuracy compared to the exiting training algorithms.

Content-Based Image Retrieval (CBIR) is a technique uses visual contents to search images from large scale image databases according to user's interests. Texture features based on gray level co-occurrence matrix (GLCM) are used for retrieval of images. The parameters which are more relevant for retrieval process have to be found out. The parameters are given to different classifiers to find the classification accuracy. The classifiers used are K-Nearest Neighbor (KNN), Support Vector Machine (SVM) and Artificial Neural Network (ANN). From the results obtained by the classifiers, ten relevant parameters are selected for retrieval. The similarity comparison is done by different distance measures. The performance of the system is evaluated by Precision Rate (PR). The maximum retrieval rate is obtained by Bray-Curtis distance. An attempt is made to increase the retrieval rate by modifying Bray-Curtis distance.

A computer aided classification method in Computed Tomography (CT) images of lungs using fuzzy logic is performed. Statistical and GLCM parameters are calculated from the image. The parameters for classification are selected by principal component analysis. Compared to Fuzzy Inference System (FIS), Adaptive Neuro Fuzzy Inference System (ANFIS) gives better classification. The back propagation neural network is used in ANFIS. The training function of the neural network is modified and the classification accuracy is increased.

Clustering techniques are used for computer aided classification. Statistical and GLCM parameters are calculated from the image. The classification is performed by K means, K medoid, Fuzzy c means, GK fuzzy and FMLE clustering techniques. Cluster validity indexes are used to evaluate the fitness of data partitions produced by a clustering algorithm. The cluster validity indices used are partition coefficient, classification entropy, partition index and separation index. The maximum classification accuracy is obtained

for Fuzzy c means (FCM) clustering technique. The objective function of FCM clustering is modified by changing the distance matrix. The modified FCM gives better classification accuracy compared to FCM.

A computer aided detection of cancer nodules in Computed Tomography (CT) images of lungs using Radon transform, Ridgelet transform and Curvelet transform is performed. The lung lobes are segmented from the CT images using morphological operations. The transforms are applied to the segmented lung lobes. From the transformed image the presence of nodules in the lung lobes can be detected and located. The Radon transform is the projection of the intensity of an image along a radial line which is oriented at an angle. The Radon transform indicates the presence of nodule in the image. The correct location of the nodule is given by the Ridgelet transform output. The exact shape of the nodule is given by Curvelet transform. From the Curvelet transformed image, the nodule can be segmented correctly. The segmented tumor has to be classified as malignant or benign. The shape and texture features are used to classify the tumor as malignant and benign. A back propagation neural network is used as the classifier.

Genetic algorithm is used to detect the location of cancer nodule in CT images of lungs. The lung lobe with cancer nodule is segmented and genetic algorithm is applied to find the location of nodule. The intensity of the pixel values is taken as the fitness value of the respective chromosomes. During each iteration, the chromosomes which have fitness values 240 or above are traced. The chromosomes that are traced through all the iterations, till the convergence is reached, represent the cancer area of the lungs. With the population size of 1600, the accuracy obtained is 94.2%. With this population size, cancer nodules of size less than 4 mm are not detected. When the population size is increased, small nodules are also detected.