ABSTRACT

Breast cancer is the most frequently diagnosed cancer in women. It is the second leading cause of cancer death in women after lung cancer. Early detection of breast cancer is believed to enhance the chance of survival. Mammography is the best available breast imaging technique at present which uses low-dose x-rays for detecting the breast cancer early before the symptoms are experienced.The most commonly present abnormalities in mammograms that may indicate the breast malignancy are *masses* and *micro calcifications*.

The prime objective of this research is to increase the diagnostic accuracy of the detection of breast cancer malignancy in Computer Aided Diagnosis (CAD) systems by developing image processing algorithms and to categorize the women into different risk groups. The first step in any CAD is preprocessing of the images to enhance the visual appearance of the images. The typical diagnostic signs in mammograms are difficult to be detected because they are low contrast images. Hence, the fundamental need in mammographic image processing is the contrast enhancement. In this thesis, a comparative study of the contrast enhancement capability indirect contrast enhancement techniques such as Histogram Equalization (HE), Contrast Limited Adaptive Histogram Equalization (CLAHE), Brightness Preserving Bi-Histogram Equalization (BBHE), Minimum Mean Brightness Error Bi-Histogram Equalization (MMBEBHE), and Recursive Mean Separate Histogram Equalization (RMSHE) was addressed. The performance of the methods was analyzed using the metrics like Effective Measure of Enhancement (EME), Peak Signal to Noise Ratio (PSNR) and Absolute Mean Brightness Error (AMBE). The simulation results prove that RMSHE

technique achieves better contrast enhancement as well as brightness preservation.

After contrast enhancement, the next step in CAD is extraction of Region of Interest (RoI). In the mammograms, masses are the Region of interest which varies in size and shape. In this thesis, an Automated Histon based Integrated Clustering Algorithm is proposed for segmentation of masses. Initial centroids play a great role in the segmentation using clustering algorithms. According to roughset theory, Histon is the upper approximation which is a contour plotted on the top of the histogram by considering a similar intensity sphere of a predefined radius around a pixel. The main idea is to find the peaks in the histon and choose the corresponding gray levels for initial cluster centers of the K-means clustering algorithm automatically. The new cluster centers obtained from the K-means are given to fuzzy c-means algorithm for segmentation. By integrating the K-means algorithm with fuzzy c-means algorithm, the proposed method gets the benefits of the both the techniques especially with respect to of minimal computation time and accuracy. The accuracy of segmentation is measured using area overlap measure and the mammographic risk classification.

From the segmented mass (RoI) of the mammogram margin, shape and texture based features are then statistically extracted. A soft computing based fuzzy cognitive maps (FCM) method is proposed to classify the features of the screening mammogram into normal, benign and malignant cases. The Data Driven Nonlinear Hebbian Learning Algorithm (DDNHL) is used to train this model in order to predict the BC risk grade based on these mammographic image features. With respect to tumor grading, the overall classification accuracy of DDNHL-FCM using 70 mammogram screening images is found to be 94.3%. The testing accuracy of the proposed model using 10-fold cross validation technique outperforms other standard machine learning based inference engines.

There is a growing demand for women to be classified into different risk groups of developing breast cancer (BC). The focus of the next work is on the development of an integrated risk prediction model using a two-level fuzzy cognitive map (FCM) model. The Level-1 FCM models the demographic risk profile. A nonlinear Hebbian learning algorithm is used to train this model and thus to help on predicting the BC risk grade based on demographic risk factors identified by domain experts. The risk grades estimated by the proposed model are validated using two standard BC risk assessment models viz. Gail and Tyrer–Cuzick (Bell cross, 2009). The predictions of the proposed Level-1 FCM model comply with the Tyrer–Cuzick model for 36 out of 40 patient cases. The Level-2 FCM is the DDNHL- FCM which models the features of the screening mammogram concerning normal, benign and malignant cases. An overall risk grade is calculated by combining the outcomes of these two FCMs. The Gail model and Tyrer–Cuzick model based only on the demographic details and do not take into account the findings of the screening mammogram. In the perspective of clinical oncologists, this is a comprehensive front-end medical decision support system that assists them in efficiently assessing the expected post-screening BC risk level of the given individual and hence prescribing individualized preventive interventions and more intensive surveillance for high risk women.

In short, the intriguing problem of detecting the breast cancer as early as possible is attacked in this thesis in terms of the following issues:

i. Enhancing the early detection of breast malignancy by

   a. identifying the appropriate algorithm for improving the contrast of the mammographic images.

   b. employing the cluster based image segmentation algorithm for *mass* segmentation.

   c. employing soft computing based decision support system namely FCM to categorize the mammograms into *benign*, *malignant* and *normal.*

ii. Categorizing women into different risk groups by

   a. categorizing the demographic risk profile into different risk grades using FCM.

   b. integrating the findings of the mammogram study with demographic risk factors of an individual to assess the overall risk of developing breast cancer.