ABSTRACT

Breast cancer is one of the most dangerous diseases prevalent among women. Mammography is the most effective imaging modality in the early breast cancer detection. Asymmetry between breasts, the presence of microcalcification, masses and architectural distortion are the most important signs of cancer on a mammogram.

Microcalcifications (MC) are tiny calcium deposits. Tiny microcalcification presence in clusters is generally associated with a high probability of cancer. These microcalcification clusters have been specifically targeted as a reliable early indicator in routine mammography screening. Between 61% and 82% of malignant tumours have microcalcifications, of which approximately 21% have microcalcifications as the only sign of malignancy. On mammography film microcalcifications can vary in size from 0.01 mm$^2$ to 1 mm$^2$. Since microcalcification are tiny calcium deposits and may be easily overlooked by the examining radiologist. The errors of diagnosis actually form the foundation of Computer Aided Diagnosis (CAD) in digital mammography. Many computerized automatic schemes for detecting and diagnosing early stage microcalcification have been developed. Computer Aided Diagnosis integrates diagnostic imaging with the knowledge of computer science, image processing, pattern recognition, and artificial intelligence techniques.

CAD is like a spell checker and by using CAD software the number of errors might decrease. Digital image processing techniques are used by CAD systems to improve the detection performance and efficiency of mammography screening. CAD analyzes the digitized mammogram and identifies areas that are suspicious for cancer and marks those suspicious areas for the radiologist. After detecting the suspicious area, CAD segments it into regions, several features are calculated for each region, and using the relevant features classifies a lesion as benign or malignant. Each method for early detection of breast cancer has its own merits and demerits and is difficult to compare the various techniques, as each one may work efficiently over images taken from a particular database. The performance of any CAD system may employ different figure of merit for their evaluation. The most widely used performance indices are Sensitivity, Specificity and Receiver Operating Characteristic (ROC) curve. The area under the ROC curve Az is considered as the performance index of
the system and an Az value range between 0.75 and 0.95 gives a good figure of merit for any CAD system.

Preprocessing steps are like mammogram orientation, label and artifacts removal, mammogram enhancement and mammogram segmentation which are carried out by any CAD system before they locate and classify the suspicious area in a digital mammography. In order to limit the search for abnormalities and to increase the accuracy and efficiency of processing algorithms extraction of the breast contour is necessary. Extracting the pectoral muscle is also particularly important in an automated mammogram image assessment. Several CAD techniques that support MC detection have been developed with some showing improvement and others showing no improvement. Some of these techniques may tend to overemphasize the sensitivity in their detection ability at the expense of specificity. This may result in increased unnecessary biopsies when using such CAD techniques. For microcalcification detection and classification a large number of algorithms have been developed by several researchers. Several CAD systems that support microcalcification detection and classification have been developed for clinical use in recent years.

An extensive literature survey was carried out and the survey revealed the following facts. The factors that lead to the missed detection of breast cancer include nature of radiographic findings, poor image quality and oversight by the radiologist. 15-25% of biopsy proven cancerous is not detected for various reasons such as technical problems and abnormalities that are not observable. There are many challenges that determine the efficiency of many CAD systems starting from the preprocessing steps to the classification steps.

Literature reports show mixed results on the role of current CAD systems. Most of these systems tend to overemphasize the sensitivity in their detection ability at the expense of specificity. This in many cases result in increased unnecessary biopsies when using such CAD systems. Addition of accurate pre-classifier (classifier) to classify the potential microcalcification into the ‘true microcalcification’ is not done effectively in many CAD systems. This increases in false positive detection. Most of the research work on classification of microcalcification cluster deal with feature extraction and classification using a suitable classifier. Till now no
standard minimum feature set is proposed to accurately classify the microcalcification clusters.

Many classification approaches are developed by assuming the underlying training set is evenly distributed. However, those approaches are faced with a severe bias problem when the training set is a highly imbalanced distribution. There are many real-world problems those are faced with severe problem of learning for imbalanced class. The imbalanced data cause classifiers to perform poorly on the minority class. When the data are highly imbalanced many existing methods tend to misclassify the minority class. In the mammogram data set there is an unbalanced distribution of cases between the malignant class and the benign class, since the number of instances of benign class is much higher than the malignant class. When learning from imbalanced datasets the tendency is that the classifier obtains a high predictive accuracy over the majority class, but predict poorly over the minority class. Many CAD systems have not well addressed these issues.

As the limitations mentioned above provide scope for further research and improvement, they form the main objective of this proposed work. After careful study and research in these areas methods are designed to improve the performance of the CAD system. These methods focus on the improvement in all the phases over the existing model by the following added features. An extensive research has been done on the preprocessing and segmentation. According to these studies, there could be a high failure rate in the proper segmentation of breast area as many methods do not preserve the skin and nipple. A new technique is explored for breast region segmentation using morphological and filtering techniques and able to preserve the skin and nipple. Many of the existing techniques do not preserve the skin and nipple during the processing of mammograms. As a part of pectoral muscle segmentation the region of interest containing pectoral muscle is located and processing area is later refined or reduced using proposed technique so that the processing time for pectoral muscle segmentation is reduced. Several features are needed in order to classify potential microcalcification and microcalcification cluster. A method is used for feature reduction that deals with the “dimensionality curse”. The dimensionality reduction is performed by discarding the irrelevant features.
It is realized that the problem of microcalcification detection should not be simply treated as looking for “blobs” in an inhomogeneous image background. Addition of pre-classifier at the microcalcification level to decide if the found objects are necessary for clusterization should be incorporated in the CAD systems for accurate diagnosis. A new technique is designed for the potential microcalcification detection. This technique detects maximum available potential microcalcification in the digital mammogram.

The issues when dealing with imbalanced data set for classification has been well addressed by proposed methods and incorporated in the CAD system. The results of classification are evaluated and compared in terms of performance using the widely accepted measures. The results obtained from the experiment indicated that proposed techniques generally performs better than other techniques in all test cases of classification and reduce the false positive detection.

The developed CAD techniques help in mitigating the key issues which are affecting the accuracy of most of the CAD systems. By implementing such a system the accuracy in diagnosis very much increases. The CAD system indicates suspicious regions with a strong likelihood of microcalcification cluster presence and classifies clusters as benign or malignant.

This thesis aims at the performance analysis of techniques towards the detection and classification of potential microcalcification and clustered microcalcification in digital mammogram. The proposed techniques have been implemented and their performance was obtained to be better than the certain works in this field.