Abstract

Breast cancer is the most common type of cancer in women worldwide, and the leading cause of death from cancer in women, especially those between 40 and 55 years of age. The early detection and accurate diagnosis of breast cancer is of utmost importance in providing effective treatment in order to increase survival rates. Mammography is the effective, economic and widely used imaging modality for breast cancer detection. Masses are the most common indicators of breast cancer which are predominant on mammograms. Two types of techniques such as screening and diagnostic mammography are routinely performed by radiologists who interpret mammograms by visual inspection. However, manual inspection is a tiring and tedious task prone to errors. Some retrospective studies have revealed that around 10-15% of breast cancer cases are missed by radiologists. In this way, the search for Computer-Aided Detection/Diagnosis (CAD) techniques has been encouraged for effective early detection of breast cancer at most treatable stage followed by successful diagnosis reducing error rate and improving decisions of the radiologists. The fundamental aim of this research work is to design and develop image processing and machine learning algorithms to investigate the potential of mammographic texture to identify breast abnormalities and characterize them as either benign mass or malignant tumors. Contributions are made in the different phases, including: (1) Preprocessing, (2) Detection of suspicious abnormal lesions, (3) Characterization of suspicious lesions and their classification, and (4) Breast cancer diagnosis using ipsilateral views. The goal of this research is to increase the diagnostic accuracy of image processing and machine learning algorithms for optimum classification between malignant and benign abnormalities in digital mammograms by reducing the number of misclassified cancers. All the developed automatic algorithms are thoroughly evaluated using a database of full field digital mammogram images obtained from Tata Memorial Centre, Mumbai, India. A set of few images from publicly available databases namely MIAS and DDSM are also used for comparing the performance of algorithms.

The first part of the research work is image preprocessing which can have a strong impact on the success of subsequent tasks. The preprocessing includes enhancement of mammograms, breast profile separation, air-skin interface (breast border) detection,
and removal of the pectoral muscle region. The methods such as transforming and windowing, fuzzy c-means, scale space approach and RANSAC algorithms are presented in this thesis for aforementioned preprocessing tasks respectively. The breast border detection has yielded promising results with an accuracy of 94.54%, 91.53% and 93.47% on TMC, MIAS and DDSM datasets respectively. The accuracy reported for pectoral muscle extraction is 89.54%, 83.05% and 85.13% on TMC, MIAS and DDSM datasets respectively.

The second part of the work is focused on the identification of the suspicious cancer regions in mammograms. The varying characteristics of breast lumps or masses including contrast, contour, shape and size etc. makes the mass segmentation a difficult task. An algorithm based on adaptive fuzzy region growing is developed for mass segmentation. A novel scheme “Sapate’s Neighboring Pixel Selection Scheme” (SNPSS) is devised to reduce the computational complexity of region growing algorithm. The method developed has been validated using real mammographic images from TMC, Mumbai and publicly available DDSM dataset. Experiments on a wide variety of mammograms show that the methodology presented has achieved better performance than several state-of-the-art and traditional algorithms in terms of both accurate lesion identification and processing time. The segmentation stage has yielded promising results with sensitivity of 91.67% and specificity of 58.33% on 460 FFDM images from TMC dataset.

In the third part of the thesis, the geometric and textural features are extracted from the suspicious lesions. Textural features are extracted from the ROI samples using Gray Level Co-Occurrence Matrices (GLCMs). The features selected are discriminative enough and are used with tuned classifiers for distinguishing the suspicious lesions as either malignant tumor or benign mass. A binary classification step is applied by using $k$-NN and Support Vector Machines to classify the suspicious lesions. The 10-fold cross validation is used to avoid the biasing of the classifiers. The classification performance is assessed using the metrics such as sensitivity, specificity, false positives per image (FPsI), Receiver Operating Characteristic (ROC) curve and the Area Under the ROC curve (AUROC). The geometric and textural features together have achieved the sensitivity of 84.44% and 85.56% with specificity of 91.11% and 91.67% with FPsI of 0.54 and 0.55 using $k$-NN and SVM classifiers respectively.
The radiologists usually analyze two views of the same breast known as ipsilateral views to confirm the abnormalities. Ipsilateral views play a vital role in the diagnosis of the breast anomalies and improve the positive predictions while reducing false predictions. Breast cancer diagnosis using the ipsilateral views is covered as the fourth part of the work. A fusion of single view features with two views correspondence score has improved the case based sensitivity of the cancer detection. The fusion based mechanism using ipsilateral views is promising but needs few modifications to yield better results. The performance of SVM classifier for TMC dataset has shown case based sensitivity of 80% at 0.74 FPsI and area under ROC curve $A_z = 0.82 \pm 0.27$ using fusion based scheme.

All the described techniques were thoroughly evaluated considering that all the previous step are correct. The impact of the conducted research will be reflected in its ability to improve the quality of breast cancer detection, speeding up the time to output a diagnosis with the correspondent useful implications in treatment possibilities. The radiologist will also benefit from the fact that he can better use his time concentrating on more difficult cases.

In conclusion, the algorithms designed and developed for breast cancer detection using FFDM images provide useful information for CAD and have the potential to assist the radiologists in faster and accurate interpretation of mammograms. The experimental results obtained from the system developed in this research prove to be beneficial for the automated detection of breast cancer. The proposed technique will improve the diagnostic accuracy and consistency of the radiologists’ image interpretation in the diagnosis of breast cancer. The resulting computerized breast cancer detection system will subsequently act as a second reader after the manual detection by the radiologist and we believe that this would aid the radiologist in the mammogram screening process.

**Keywords**
Breast Cancer, Computer-Aided Detection (CAD), Mammography, breast abnormalities, segmentation, feature extraction, classification, feature extraction, geometric features, textural features, Support Vector Machine, ipsilateral views