CHAPTER 1

INTRODUCTION

1.1 BACKGROUND

Worldwide, cancer has become the leading cause of mortality. The increasing global burden due to cancer can be attributed to the adoption of cancer-inducing lifestyle by the growing population of the world. According to the World Health Organization’s (WHO) GLOBOCAN 2008 cancer fact sheets, about 12.7 million cancer cases and 7.6 million cancer deaths have been reported in the year 2008. The WHO has also estimated that breast cancer is the highest incident cancer among women, accounting for 23% of total cases. It is also the primary cause of death in women. Nearly 14% of female deaths due to cancer are caused by breast cancer. Early detection and subsequent treatment of breast cancer can save the life of a woman without the need for mastectomy (Jemal et al 2011).

A mammogram, which is a low-dose X-ray image of the breast, can depict the earliest sign of breast cancer even in asymptomatic women. It is a two-dimensional (2D) projection of the compressed three-dimensional (3D) breasts. The two types of mammography are film mammography and digital mammography. In film mammography, the image is created on a film, whereas in digital mammography, an electronic image of the breast can be stored directly on a computer and interpreted using a high-definition monitor (Tang et al 2009).

The breast imaging-reporting and data system (BI-RADS) is a mammography lexicon developed by the American college of radiology (ACR). BI-RADS has categorized the breast tissue into one of the following
four categories: (1) almost entirely fatty (2) scattered fibroglandular (3) heterogeneously dense and (4) extremely dense (Bovik 2005). In mammography, the absorption of X-rays and hence the image formation depends on density. Indicators of breast cancer are radiodense, i.e., they appear white on mammograms. Hence, fatty tissue, which is radiolucent and appearing dark gray-to-black on mammograms, provides a good background to visualize cancer. As the density of breast tissue increases, interpretation of mammograms becomes difficult (Prasad & Houserková 2007).

Screening mammography is performed on women who do not exhibit any signs or symptoms of breast abnormality, in an attempt to detect breast cancer in the very early stages. On the other hand, diagnostic mammography, intended for post-screening, aims at evaluating the clinically determined abnormalities or those found during screening mammography. Decisions regarding requirement of additional imaging or biopsy are made in diagnostic mammography. The standard practice in mammography is to acquire breast images from two different angles. These include the mediolateral (MLO) and craniocaudal (CC) views, also called ipsilateral mammograms. The MLO projection is a side-to-side view of the breast, taken at a 45 degree angle and shows part of the pectoral muscle. This view shows a better perspective of the glandular portion. The CC projection is a top-down view of the breast. The focus of this view is on the central and inner portions of the breast. In screening mammography, both these views are acquired for both breasts, whereas in diagnostic mammography they are acquired for the breast that is to be examined (Tang et al 2009).

Radiologists usually search for visual indicators on a mammogram for detection and diagnosis of breast cancer. These indicators can be classified into two categories, i.e., space-occupying lesions and microcalcification clusters (MCCs). The former include masses, architectural distortion and
asymmetry. Masses are seen in more than one projection and they appear as dense regions on mammograms. MCCs are small clusters of calcium deposits that are seen as small, bright spots on a mammogram. Architectural distortion results from the distortion of the normal breast architecture without any definite mass. It produces spiculations radiating from a point and causes focal retraction or distortion at the parenchymal edge. Asymmetry is defined as larger volume or density in one breast as compared to the other breast. It might also be characterized by more prominent ducts in one breast. As a result of asymmetry, the overall appearances of the left and right breasts differ in the corresponding mammographic images (Rangayyan et al 2007).

Normally, while analysing mammograms, radiologists compare bilateral mammograms, i.e., mammograms of right and left breasts, to evaluate suspicious asymmetric densities. In addition, radiologists also compare temporal mammograms to check for growing densities. Temporal mammograms are mammograms comprising current mammograms and prior mammograms (mammograms acquired in previous screening rounds). Further, the suspicious regions in different mammographic projections, i.e., ipsilateral mammograms, are analysed in combination so as to improve the detection and diagnostic performance (Van Engeland & Karssemeijer 2007).

Currently, mammography is the only widely accepted imaging modality for routine screening of breast cancer. It has been shown that screening mammography can reduce breast cancer mortality rates. However, the radiation effects of mammography can be hazardous to patients as well as radiologists. Mammographic abnormalities are subtle, especially for dense breasts, which limits the detection performance. It has been reported that radiologists fail to detect 10-30% of cancers, using mammography. Failure to detect early cancer results in missing the best period for treatment. Also, misdiagnosis of benign cases as malignant is common in mammography and
this leads to biopsies that cause unnecessary physical, emotional and financial discomfort to the patients. Biopsy is an invasive procedure and is considered to be the gold standard to determine whether a tumor is malignant. In mammography, about 65% to 85% of biopsy operations are reported to be unnecessary (Bovik 2005).

Sonography (ultrasound imaging) is an import adjunct to mammography for interpreting breast abnormalities, especially in the diagnostic phase. Unlike mammograms, image formation depends on the echogenicity of the tissue/organ being scanned. Transmission of ultrasound waves by the tissue results in an echo-free area that appears black on the image. In contrast, the reflected ultrasound creates an echogenic area that appears gray or white on the film. Since ultrasound does not involve radiation, it is safer than mammography. Sonography is also useful in detecting breast lesions in women with dense breasts and those who are less than 35 years old, for whom mammogram is not effective. Statistics show that the use of sonography can enhance detection rate by 17%. Sonography has an important role in differentiating cysts (fluid-filled lesions that are almost benign) from solid lesions, whereas this is not possible with mammograms. Some studies have also reported that the accuracy of classifying solid benign lesions from malignant lesions is higher for sonograms (images created using sonography) than mammograms. It has been shown that sonography can reduce the number of unnecessary biopsies by 30% (Cheng et al 2010, Liu et al 2010).

Despite these advantages, sonography is not considered as a standard screening test and is used primarily in combination with mammography in the diagnostic phase. The reason is that sonography involves a hand-held probe for scanning and hence it is much more operator-dependent than mammography. The nature of image acquisition might also result in some areas of the breast not being scanned. Sonography has poor ability to capture
deep lesions and is more useful for diagnosing clinically palpable and/or mammographically detected lesions only. Further, sonography cannot always detect microcalcifications. Nevertheless, it is finding increased use as a complementary imaging technique for mammography and can add value to breast cancer detection and diagnosis (Prasad & Houserova 2007).

Irrespective of the imaging modality used, the level of expertise of the radiologists and the quality of the image limit the accuracy of the breast cancer detection and diagnosis. Double reading by two different radiologists has been sometimes opted to overcome this problem. It has been reported that double reading can improve the detection rate by 5% to 15%. However, the time, cost and work load involved is high and the resulting diagnostic decisions might be ambiguous. Computer aided detection/diagnosis (CAD) systems assess breast images objectively as opposed to the subjective analysis made by the radiologists. Many studies show that the use of a CAD system as a second reader has the potential to improve the accuracy of breast cancer detection and diagnosis (Mencattini et al 2010).

Computer aided detection (CADe) systems locate suspicious regions called regions of interest (ROIs) in the breast images. A CADe system consists of two stages; detection of suspicious regions and false positive reduction. In the latter step, false positives, i.e., normal regions deemed suspicious, are ruled out by classifying the detected ROIs as normal regions or true lesions. A CADe classification problem is binary in nature, wherein normal findings are considered as negative cases and lesions are considered as positive cases. Thus, true negatives are lesions classified correctly and true positives are normal findings classified correctly. On the contrary, wrong classification of normal cases as lesions results in false positives and lesions classified as normal results in false negatives. Computer aided diagnosis systems (CADx) systems classify lesions as benign or malignant. A CADx classification problem is also binary
in nature, where benign lesions are considered to be negative whereas malignant lesions are considered to be positive. Thus, true positives and true negatives are malignant and benign findings, respectively, classified as malignant. On the other hand, true negatives and false negatives are benign cases that are classified as benign and malignant, respectively (Jalalian et al 2013).

Motivated by the potential benefits of CAD systems, some mammogram-based commercial systems including the ImageChecker (R2 Technology), SecondLook (CADx Medical Systems) and MammoReader (Intelligent System Software Inc.) have been developed and are available on the market in the United States (Rangayyan et al 2007). To the best knowledge of the research scholar, there exists one commercial sonographic CAD system, i.e., the B-CAD, developed by Medipattern, Canada (Shen et al 2007). All these commercial CAD systems have been approved by Food and Drug Administration (FDA). However, all the above mentioned commercial CAD systems can only detect suspicious lesions and are not useful for classifying the detected regions as normal regions or abnormal.

Accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) are common performance metrics used to measure the performance of a CAD system (Cheng et al 2010). Accuracy, defined in Equation (1.1), is the fraction of cases (both positive and negative) that are correctly diagnosed.

\[
\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP}
\]  

(1.1)

In the above equation, \(TP\) and \(TN\) indicate the number of true positives and the number of true negatives, respectively; \(FN\) and \(FP\) indicate the number of false negatives and false positives, respectively.
Sensitivity is defined as the fraction of diseased cases that are correctly diagnosed. It is also called the true positive rate (TPR) and is calculated according to Equation (1.2).

\[
Sensitivity = TPR = \frac{TP}{TP + FN} \tag{1.2}
\]

Specificity given by Equation (1.3) is defined as the fraction of non-diseased cases that are correctly diagnosed. It is also called the true negative rate (TNR).

\[
Specificity = TNR = \frac{TN}{TN + FP} \tag{1.3}
\]

PPV is the proportion of positively diagnosed cases that are actually diseased and is defined in Equation (1.4).

\[
PPV = \frac{TP}{TP + FP} \tag{1.4}
\]

NPV defined in Equation (1.5) is the proportion of negatively diagnosed cases that are actually non-diseased.

\[
NPV = \frac{TN}{TN + FN} \tag{1.5}
\]

However, a major drawback associated with the above simple measures is that they depend on a single decision threshold used for classifying the cases as positives or negatives. Though a lower threshold yields a high sensitivity, the associated cost is a large number of false positives, thus reducing specificity and PPV. On the other hand, a higher cut-off that yields a high specificity generally results in high false negatives and reduces the sensitivity and NPV. Although high sensitivity is generally preferred, a high specificity may be of higher priority in screening programs where a large population is screened for detecting early cancer. Generally, the decision threshold is optimized to obtain the desired performance (Llado et al 2009).

The receiver operating characteristic (ROC) curve allows for a detailed evaluation of diagnostic accuracy at varying levels of sensitivity and
specificity. It is a plot of TPR along the y-axis and false positive rate (FPR) along the x-axis for different decision thresholds in a binary decision problem, where TPR and FPR are given by Equation (1.2) and Equation (1.6), respectively.

\[
FPR = 1 - \text{Specificity} = \frac{FP}{TN + FP} \tag{1.6}
\]

Thus, rather than using a single sensitivity-specificity pair, the ROC curve is a more useful means of evaluating the performance of a diagnostic test. Also, unlike measures like accuracy, PPV and NPV, the ROC analysis is not influenced by the prevalence of disease in the sample; prevalence is defined as the proportion of a population found to have a certain diseased condition (Nagarajan et al 2014).

An ideal ROC curve would start at (0,0), move vertically upward to (0,1) and then horizontally to (1,1). The area under the curve (AUC) derived from the ROC plot is considered to be a good measure that summarizes the test accuracy. It can assume a value that ranges from 0 to 1. The closer its value is to 1, the better is the diagnostic performance of the test. AUC is not influenced by the prevalence of a disease in the sample. Further, it is not calculated on the basis of only one cut-off point, i.e., it does not consider false positive and false negative results as equally undesirable (Bewick et al 2004). Yet another diagnostic measure that can be obtained from the ROC plot is the equal error rate (EER). This is the rate at which FPR and false negative rate (FNR) are equal, where FPR and FNR are defined in Equation (1.6) and Equation (1.7), respectively. In general, a test with the low value for EER is better (Wu et al 2011).

\[
FNR = 1 - TPR = \frac{FN}{TP + FN} \tag{1.7}
\]
Recently, integration of information from multiple sources is gaining wide popularity in data analysis. This usually involves integration of features or intermediate decisions from multiple complementary sources or a combination of the two. These methods for fusion are called feature fusion, decision fusion and hybrid fusion, respectively (Atrey et al 2010). In developing CAD systems for breast cancer, information fusion would serve to mimic the radiologist’s practice of combining information from multiple mammographic views (bilateral, temporal and ipsilateral) and from multiple imaging modalities like sonography and mammography.

1.2 FOCUS AND ORGANIZATION OF THE THESIS

The work presented in this thesis focuses on multiview and multimodal information fusion based on correlation analysis, which would aid better false positive reduction in the detection of masses as well as better classification of masses. To be more specific, canonical correlation analysis (CCA) is used to explore the correlation among multiple sources that are fused. Among the many indicators of breast cancer, mass is the focus of this work as it is the most common indicator of breast cancer and yet difficult to diagnose. (Eltonsy et al 2007). To be more specific, false positive reduction based on combining unilateral and bilateral information, mass classification based on combining ipsilateral information and mass classification based on combining sonographic and mammographic information are the research topics of interest in this thesis. Chapter 2 presents a detailed literature review pertaining to these subjects.

Section 2.1.1 reviews different algorithms based on unilateral and bilateral analysis of mammograms for false positive reduction in CADe systems as well as fusion techniques that have been employed in the past for combining the two analyses. Unilateral analysis deals with classification of an
ROI as mass or normal, based on the characteristics of that ROI only. On the contrary, bilateral analysis derives discriminatory information from the ROI in the candidate mammogram (mammogram under consideration) and the symmetric ROI in the contralateral mammogram (bilateral counterpart of the candidate mammogram).

In Section 2.1.2, a survey of algorithms that use information from a single mammographic projection for mass classification in CADx systems is presented initially. This is followed by a review of different fusion techniques employed to combine ipsilateral mammographic information in CADx systems meant for classification of masses as benign or malignant.

In Section 2.1.3, a survey of algorithms that employ sonography for mass classification in CADx systems is presented. Further, a review of fusion schemes that combine information from sonograms and mammograms in CADx systems intended to classify masses has also been presented.

Following literature review in Chapter 2, the proposed fusion schemes based on CCA have been presented in Chapter 3, Chapter 4 and Chapter 5. In particular, Chapter 3 validates the use of CCA for combining unilateral and bilateral information to aid false positive reduction in the detection of masses in multiview mammographic CADe systems. Chapter 4 discusses the application of CCA for combining ipsilateral mammographic information for mass classification in a multiview mammographic CADx system, while Chapter 5 explores the benefits of using CCA for fusing sonographic and mammographic information in a multimodal CADx framework intended for the same task. While the focus of the applications considered in Chapter 3 and Chapter 4 is the screening phase, Chapter 5 targets the diagnostic phase. Finally, in Chapter 6, major inferences based on the results of various experiments have been discussed. Also, conclusion and future scope have been presented in this chapter.