Chapter 1: Introduction

1.1 Introduction:

Globally, in the medical domain, cancer has emerged as the scourge of humankind and acts as the foremost pathological situations leading to mortality. Currently, human populations are largely affected by cancer related ailments. Worldwide, cancer is acquiring the shape of endemic leading to mortality. Immediate global awareness and preventive measures are essential for counter measuring this fast spreading disease (Ferlay *et al.,* 2013). In 2015, globally overall cancer patient count tolls up to 1,658,370. Out of that, 589,430 cancer patients attain mortality predominantly due to ovarian, breast, skin and oral cancers. As per latest data of American Cancer Society, there is a higher prevalence of breast cancer in females and prostate cancer in male in 2016. Among the developing countries, Indian population suffers most from the breast, ovarian, prostate and cervix cancer related mortality rates (Jemal *et al.*, 2011).

The occurrence of cancer has piled up more in developing nations, largely owing to western lifestyle adoptions, urbanizations and improvement in survival rates (Boyle *et al.*, 2008). However, the mortality rate of cancer has been declined in the last few decades. The reductions in cancer patients attribute to the advancement in medical sciences through enhanced diagnostic techniques, medical treatment procedures and most importantly, the increase in general awareness (Boyle *et al.,* 2008; Thune *et al.*, 1997).

Cancer is a disease of uncontrolled growth of abnormal cells. These cells used to proliferate in other parts of the body and consecutively inhibit the growth of healthy cells. Cancerous cells replicate much faster than normal cells. Cells divide and multiply to form a tumour that may be cancerous and non-cancerous (Rubin *et al.*, 2008).

Cancer detection has always been a major issue for the pathologists and medical practitioners for diagnosis and treatment planning. The chances of curing cancer are primarily in its detection and diagnosis (Chen *et al.,* 2015). However, its correct diagnosis and treatment is an important factor for the better survival rates of the cancer patients.

Currently, several imaging techniques such as X-Ray, Mammography, Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), Computer Aided Tomography (CAT) and Ultrasound Imaging, have been used for the screening and detection of cancer. Different biochemical tests such as blood test, serum tumour marker test and lipid peroxidation test (LPO) are used.

The above mentioned techniques have the advantage of furnishing effectual diagnostic tools for screening and early stage detection of cancer. However, these have the limitations of not being very useful in determination of the malignancy level of the cancer. Further, these techniques cannot be employed as gold standard and it deserves the biopsy examination, the images of which provide complete information about the tissue at cellular level, and thereby, aids to arrive at the final conclusion.

Histopathological images provide an informative view of abnormalities present in the cellular structure of the underlying tissue since the structure of the tissue is preserved in the preparation process. It deals with the minute structure of the tissue and provides information about abnormalities present in the cellular structure of the tissue and cells. For accurate information about the malignancy levels, the tissue biopsy samples are considered as gold standards as it provides typical cellular details regarding the tissue under laboratory investigations. At present histopathological examination are regularly utilized for screening biopsy samples for the diagnosis of cancer (Fischer *et al.*, 2008).

Pathologists visualize and analyze the histopathology images under microscope for the examination of abnormalities or some of the important signs present in the cells based on various characteristics such as distribution of cells, regularities of cell shape, size, appearance, cluster density and architectural distortions across the tissue which helps pathologist to make decision of cells whether it is normal and cancerous (Chen *et al.*, 2012). The manual evaluation of cancer detection from histopathological images is subjective in nature, time consuming and varies with perception level of expert pathologists (Veta *et al.*, 2014). To overcome this problem, there is a need of automation of image analysis. Therefore, computer aided diagnosis techniques are being introduced for fast, reliable and accurate diagnosis of cancer (Gurcan *et al.*, 2009; Madabhushi *et al.*, 2009).

The primary goal of this thesis is to develop a computer aided diagnosis (CAD) system of histopathological images for classification of cells into benign and malignant using morphological, intensity and texture based features. All the benign and malignant cells based images are obtained after staining with haematoxylin and eosin based histopathological processes (He *et al.*, 2012). The CAD system is based upon four major steps: (a) pre-processing, (b) segmentation, (c) feature extraction and (d) classification of histopathological images (Dougherty, 2009; Arif *et al*., 2007; Price *et al*., 2003; He *et al*., 2012; Gonzales and Woods, 2007). The proposed CAD system aims to detect benign and malignant cells in more quantitative, consistent, and repeatable manner for producing accurate diagnostic decisions (Dermir *et al*., 2005).

In image acquisition technique, various steps are followed. The segmentation technique concentrates on the identification of Region of Interest (ROI) areas from the cancer image slides. The typical characteristics that derived from the ROI images are extracted in the successive steps. The quantitative analysis includes classification of abnormal and normal tissues based on the classifier features. Figure 1.1 depicts the basic steps involved in the automated image analysis for the detection of cancer cells.

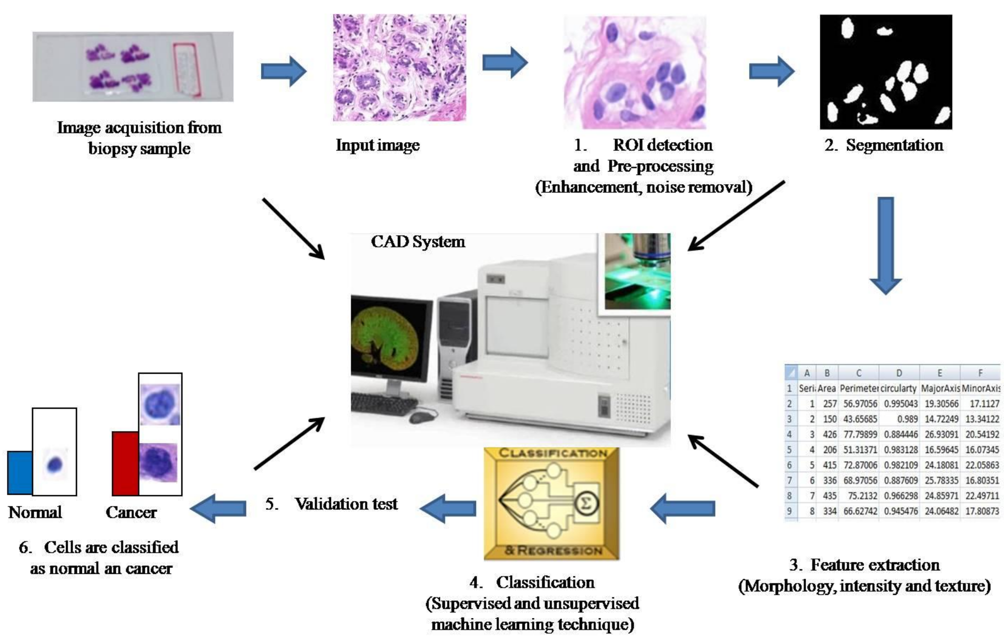

Figure 1.1: The general procedure of computer aided diagnosis.

Numerous investigations have provided computational approaches for yielding typical information regarding tissues. These approaches are applied to designing computer based diagnostic tools that detect the malignancy levels and aids in the enhancement of overall diagnostic pathological approaches. Current investigation revealed that CAD systems have a significant impact on cancer detection and have speed up the process of accurately analyzing more number of cases. Conversely, CAD systems for analysis of images are in developmental stages and new approaches are essential for the further enhancement of the approaches.

1.2 Problem statement and motivation:

The CAD based cancer detection approaches are in developmental phases, so standard examination and biopsy based techniques are still required. This confront is more aggravated by the scarcity of new screening algorithms and insufficient experimental spaces for screening biopsy histology (Mukherjee *et al.*, 2007). Furthermore, an aspect that adds up to these grim situation includes a shortage of knowledge and determining skills of the health professionals those who screens the histological samples for diagnosing cancer WHO (2002).

In cancer investigation processes, the laboratory experts screen the histopathological slides and test the viability of the abnormal cells. This approach forms the basis for cancer diagnosis. Conversely, the observation of the pathologists may differ from subject to subject as their degree of expertise varies in cross examination of the slides (Wang *et al*., 2007). Generally, the laboratory investigation depends on the practical skill and proficiency of the laboratory experts (Thiran and Macq. 1996).

The CAD based system is essentially a computer dependent technique that amalgamates different research areas such as physics, mathematics, statistics, digital image processing, and computer vision. The CAD system aids the medical experts in cancer diagnosis and further interpretations. The improvement in imaging techniques has significantly enhanced the medical imaging qualities and that have paved the way towards better perceptive and image analysis.

Henceforth, it is needed to develop a system which assists the pathologists in the diagnostic process and reduces the subjectivity associated with the diagnosis. This presents work motivated behind CAD systems to detect cancer by using quantitative image analysis of histopathological slides.

Recently, CAD system has become a part of routine clinical detection methods for cancer diagnosis using histopathological images at various screening units and hospitals. It stands for the most important key explorative area for histopathological imaging and diagnosis. Further, there is an imperative need for CAD system to minimize the human error. Henceforth, a considerable amount of effort in CAD has been applied in the field of cancer diagnosis and detection (Chen *et al.*, 2015). A successful quantitative image analysis approach will have to overcome these issues in a robust way, in order to maintain a high level of the quality and accuracy of cells detection, segmentation, and classification.

1.3 Aim and objective of the thesis:

The aim of this research work is to develop efficient algorithms in CAD system. This work uses different classifiers to identify benign and malignant cells of histopathological images of cancer cells on the basis of morphological, intensity and texture features. The developed CAD system will assist in analyzing histopathological images with more accuracy levels. Further, development of Graphical User Interface **(**GUI) tools will assists the utilization of these techniques in the diagnosis of cancer cells.

The objective of the present work is to develop the CAD system for cancer detection using histopathological images. This CAD system includes enhancement, segmentation, feature extraction and classification approaches. The following are the specific objectives:

- To propose efficient and robust approaches for the various steps of an automatic CAD system for cancer detection from histopathology images free from human error analysis.

- To develop the Dynamic Stochastic Resonance (DSR) based approach for enhancement and segmentation. The algorithm helps to identify and separate the ROI i.e. cells from the complex background of histopathological images of cancer corrupted with noise.
- To identify cells as benign and malignant according to the quantitative diagnostic features of cancer cells based on morphology, intensity, and texture.
- To evaluate the effectiveness of the classifiers while minimizing the classification error, for the breast cells to distinguish benign and malignant.
- To develop efficient and robust GUI based tool for differentiating benign and malignant cells.
- Comparative study of different classifiers for histopathological image of breast cancer detection and classification using morphological features based on all cells in the image has been proposed. Ranking of the important features were evaluated by applying Relief-F algorithms
- To evaluate the morphological changes in histopathological images of ovarian and breast cancer tissues and its correlation with their biochemical parameters (Malondialdehyde level).

1.4 Contribution of the thesis:

The main contributions of this research work consist of the development of CAD system for cancer detection using histopathology images. The major steps are depicted below:

- A proposed Dynamic Stochastic Resonance (DSR) based method successfully enhance the images and helps in better segmentation of the cells from the complex background of histopathological images of cancer. That method enables the detection of cells accurately.
- Morphological, intensity and texture based features are extracted from the segmented images. Further, these features provide the significant difference in the quantitative measure of shape, size and circularity, intensity etc. and have the potential for classification of benign and malignant cells.

- Based on extracted features, the best-suited classification approaches are tested on cancerous tissues of histopathology images. The three classifiers Artificial Neural Network (ANN), K-nearest neighbour classifier (k-NN) and Support Vector Machine (SVM) have been used to classify benign and malignant cells in histopathological images in single cells dataset and group cells dataset. Further, evaluation of the performance parameters is used in this research to investigate the efficacy of CAD for cancer: True Negative (TN), True Positive (TP), False Negative (FN) and False Positive (FP). By utilizing the count of samples lying into these classes, the performance parameters accuracy, sensitivity and specificity, balanced classification rate (BCR), F-measure and Matthews' Correlation Coefficient (MCC) are calculated. Further, for a better measure of classifier performance derived from a confusion matrix, relative mean square error, and Receiver Operating Characteristic (ROC) curve are predicted to investigate the effective assessment tool for ANN, k-NN and SVM classifier. The statistical measure regarding the specificity and sensitivity has been found by calculating the area under the ROC curve.
- Graphical User Interface (GUI) has been developed named as CELL CHECK (v1.0) using MATLAB–GUI with ANN classifier to make the process easier, fast, user-friendly and robust identification of benign and malignant cells with the good level of accuracy. It is used in morphological operations for supporting medical activity in recognition of cancer. Further, it speeds up the analysis, enhances the visualization and assists in the exploration of a large number of images and has the ability to detect benign and malignant cells accurately.
- Comparative study of different classifiers using morphological features based on all cells in the histopathological image of breast cancer for classification approaches are presented. For comparing segmentation methods, Trainable Weka Segmentation (TWS) approach is tested on breast cancer histopathological data set and compared with ROI segmented ground truth images using other segmentation methods *viz.* Mixture Modeling Thresholding (MMT), Simple Interactive Object Extraction (SIOX), and Robust Automatic Threshold Selection (RATS).

From this TWS based segmented image important morphology feature is extracted. The comparison of the proposed TWS method has been performed with other methods in terms of correlation, Global Consistency Error (GCE), Normalized Probabilistic Rand (NPR) index and Variation of Information (VI) for enhancement and segmentation. Further, these features are utilized for the classification by using eight popular classifiers like Multilayer Perceptron (MLP), logistic modal tree (LMT), Random forest, Rotation forest, Sequential Minimal Optimization (SMO), Naïve Bayes, J-Rip, and PART was used. Ranking of the feature was done by applying Relief-F algorithms. The efficacy of the feature and classification methods based on the feature of benign and malignant cells have been examined and presented.

- Evaluation of the changes in patterns of morphological, intensity and texture based features of malignant cells as compared to benign cells in ovarian and breast cancer of histopathological images and its correlation with the several biochemical parameters. Further, the morphological observation of these features disclose an improved difference in area, perimeter, major axis, circularity and max intensity values between benign and malignant cells. It is found that significant difference in size, shape and hyperchromatic nuclei in benign and malignant cells exists.

This thesis has been aimed to aid researchers and medical practitioners in the area of cancer detection to go through the state of the art methods for the recent developments in cancer detection and segmentation from histopathology images.

1.5 Organization of the thesis:

The entire thesis is divided into eight chapters as follows:

Chapter-1: This chapter describes a brief introduction on the present scenario regarding the examination of the tissue structure by the pathologists for the detection of the abnormalities leading to cancer. The essential need for automation of the procedure through a CAD system is addressed. In perspective of addressing the requirements for this necessity, the present chapter briefly represents the problem statement and motivation, aims and objectives and research contributions of the present thesis work. At the end of this chapter, the thesis organization section concisely highlights all the chapters present in this thesis.

Chapter-2: This chapter includes a detailed literature survey of the background research work done in the area of cancer detection from histopathological images using CAD system. It starts with an explanation of preprocessing techniques in histopathology image modalities. Segmentation methods are then discussed. The chapter continues with an explanation of feature extraction based on the various attributes including morphological, textural and intensity based features. A discussion of the classification methods and disease identification is also represented in this chapter. Finally, the existing researches on different cancer using image analysis of histopathological images are discussed.

Chapter-3: This chapter deals with the histopathological procedures for the obtaining images from the samples and then image acquisition techniques are discussed. After that, brief description of the image enhancement techniques required to process the acquired images. The performance of the proposed enhancement approach has been compared with the other commonly used enhancement methods such as CHAHE, ϒ correction, Single Scale Retinex (SSC). Further, Otsu's thresholding, k-mean, fuzzy c-mean and proposed DSR based approach has been discussed for the segmentation of histopathological images. Segmentation performance matrices in terms of correlation, GCE, NPR, and VI have also been explained.

Chapter-4: This chapter provides the CAD system for the quantitative measurement of features of haematoxylin and eosin stained breast cancer histopathological images. A brief introduction about breast cancer dataset available online University of California, Santabarbara (UCSB) that used in this research for feature extraction of single and group cells dataset. After that, image pre-processing, segmentation and feature extraction steps have been carried out. The definitions and mathematical descriptions of various morphological, intensity and texture based features have been described which are used throughout the thesis.

Chapter-5: This chapter explains the requirements of classification steps in CAD system, types of classification and training and validation steps have been discussed. Further, definitions and mathematical descriptions of various performance evaluation methods used throughout the thesis for examining the efficacy and suitability of the proposed approaches associated with the existing algorithms have been investigated. In the present investigation three classifiers (1) Artificial Neural Network (ANN), (2) k-Nearest Neighbour (k-NN) and (3) Support Vector Machine (SVM)) has been trained using breast cancer single cell and group cell dataset. The performance was evaluated in terms of classifier's performance measures such as accuracy, sensitivity, and specificity. Finally, a brief description and working of Graphical User Interface (GUI), named as CELL CHECK (v1.0) has been explained for identification of a benign and malignant cell in breast cancer of histopathological images.

Chapter-6: This chapter presents the comparative study of different classifiers for histopathological image of breast cancer classification and detection using morphological features based on all cells present in the image. A computer aided diagnosis approach consisting of various steps, *viz.* pre-processing, segmentation, features extraction and classification has been proposed. Contrast Limited Adapted Histogram Equalization (CHAHE) approach is used for pre-processing and Trainable Weka Segmentation (TWS) approach is used for segmentation purpose. Ranking of the features is accomplished using Relief-F algorithms.

Chapter-7: This chapter evaluates the changes in patterns of morphological, intensity and texture based features of malignant cells and the comparative features of benign cells in ovarian and breast cancer of histopathological images and its correlation with the several biochemical parameters. Further, morphological observation of these features disclose an improved a difference in area, perimeter, major axis length, circularity and max intensity values between benign and malignant cells. Further, it presents an overview of the biochemical tests such as RBC count, WBC count, Haemoglobin level, platelets count, CA-125, and malondialdehyde (MDA) of normal, breast and ovarian cancerous patients.

Chapter-8: This chapter describes the outcomes of the major findings of this research work with brief conclusions and suggestions for the future work.

References: This includes the references as a source of information to carry out the entire research work.