CHAPTER 1

INTRODUCTION

Lung cancer is one of the most common types of cancer in the world, where the mortality rate is high, even after so much of technical and medical advances. Out of all various types of lung cancer, adenocarcinoma is increasing at an alarming rate. The reason is mainly attributed to the increased rate of smoking – both active and passive. Apart from smoking, incidence of adenocarcinoma is also reported due to inhalation of harmful fumes from indoor pollution, as well as various genetic and other factors has been pointed out by Park et al (2008).

Various modalities like light microscopy, X-ray, CT and MRI can be used for the detection and diagnosis of lung cancer. Each of these has its own advantages and disadvantages. Out of various modalities used for prescreening, Sputum cytology images are best suited for prescreening because of noninvasive nature and cost effectiveness. In this work, a low cost and effective pre-screening system which can be deployed on a wide scale was developed. Several previous studies suggest that sputum cytology is best and affordable method for lung cancer detection has been pointed out by Oswald et al (1971), Veena et al (2012), Pa1cic et el (2002) and (Thunnisen 2003). Screiber & Mecrory (2003) studies various modalities for lung cancer diagnosis and found that sputum cytology is well suited for identifying the distinction between various types of lung cancers.
A study on high-risk population using sputum cytology revealed that a 5-year survival rate of 54% was observed in patients who undergone pre-screening as opposed to 13% for those who never had any pre-screening has been pointed out by Petty & Thomas (2000). If a system is developed for pre-screening, capable of detecting lung cancer at early stages, then the survival rate can be substantially increased has been pointed out by Hoda et al (1996).

The present day manual screening system makes it difficult to implement such a system due to lack of expert cytopathologists and the enormous volume of population to be screened. So an automated system which will pre-screen for malignancy (Cancer) for the given sputum cytology images is an urgent requirement of today’s world. In this work, focus was given on adenocarcinoma which is the cancer (malignant cells) affecting glandular epithelium. If the pre-screening using sputum cytology results in positive case, then further confirmatory tests like biopsy can be done for diagnosis.

1.1 MOTIVATION

In developed and developing countries, Lung cancer is a major threat to mankind. Lung cancer, the most common cause of cancer-related death in men and women, is responsible for 1.3 million deaths worldwide annually. Figure 1.1 shows that the mortality rate of Lung Cancer is second highest when compared to the mortality rate due to other diseases. The main reason for the highest mortality due to lung cancer is because of non availability of pre-screening system which can analyze the cancer cells at early stages. Manual scanning of microscopic slide is laborious and susceptible to human errors. And also the task of screening with manual system available currently is difficult to process, involves more cost and is more time consuming. Computer Assisted Diagnostic system can be used for
population screening. This motivated us to develop a new system which will automatically screen for malignancy in an affordable way which will increase the survival rate of lung cancer patients.

![Figure 1.1 Mortality statistics of Various Diseases](Source: http://globocan.iarc.fr/)

**Figure 1.1** Mortality statistics of Various Diseases (Source: [http://globocan.iarc.fr/](http://globocan.iarc.fr/))

### 1.2 NEED OF THE PROPOSED SYSTEM

The need for developing an automated system for detection of Lung cancer arises from the fact that the current method involves screening by
required that result be unique and is needed to be objective. This result can be used for future reference also.

1.3 CHALLENGES OF THE PROPOSED SYSTEM

In the proposed work, Sputum Cytology Images are used for classification of Lung glandular cells as benign and malignant. Sputum samples usually contain less no of cells for a clear diagnosis. Lot of biological noises is also available in the sputum cytology images. The other challenges that are faced while developing the automated system are as follows:

1. The difficulty faced in distinguishing cells from background and the Segmentation of the cells.

2. Getting the domain knowledge of Cancer and especially about Adenocarcinoma Lung Cancer.

3. Very less work is done on classification of Adenocarcinoma Lung Cancer which is very difficult for us to make literature survey and then to compare the results with existing methods.

1.4 OVERVIEW OF THE PROPOSED METHOD

In this work, a Computer Aided Diagnosis (CAD) sputum cytology image analysis system, which classify the cells as benign or malignant for the given Lung glandular cells was proposed. The general block diagram of the proposed System is shown in the Figure 1.2.
1.4.1 Input Microscopic Image

The cell samples collected are placed under a light microscope and digitized using a customized digital camera. The images are properly labeled and stored. The sputum cytology images are so chosen such that the target region contains glandular cells.

1.4.2 Preprocessing Stage

The input digital images may contain noises of various types as physical and biological. The physical noise such as impulse noise or due to power line frequency may possible interferes with the image. These kinds of noises need to be dealt at an earlier stage otherwise can affect the proper functioning of the algorithm. The biological noises are of various types. In this work, focus is given for glandular cells only so anything other than this type of cell may have to be considered as biological noise. These usually include the presence of blood cells, leukocytes, or other non cellular objects.
like mucus, pollen etc. All these types are to be marked as unwanted and have to be removed from further processing.

1.4.3 Segmentation Stage

This stage properly marks out the position of glandular cells in the image. Various image processing algorithms are needed for this purpose. Sometimes a single algorithm may not give satisfactory segmentation and hence different algorithms are in parallel and chose the best output. Since clusters are dealt rather than individual cells it is often not possible to separate the clusters in a proper manner. So it is necessary to make segmentation an approximate one keeping the margin of error at a very low level.

1.4.4 Feature Extraction Stage

The segmentation results are fed into a feature extraction module. There using various image analysis techniques morphological, textural, color and scale based features are extracted. All these features are properly labeled and stored for further analysis.

1.4.5 Classification Stage

Classification is the crucial step in the entire operation as it is in this stage that the decision is taken whether the sample is malignant or benign. To train the system, the initial sample images are used.

1.5 ORGANIZATION OF THE THESIS

Biological details required are collected from the expert doctors to know about lung cancers. The sputum cytology images from are collected from expert cytologist. Then a literature survey required for the implementation of the proposed method was made. Initially, a semi-automated system for classification of glandular cells using multiple color
spaces was developed. Then classification of glandular cells is done using different SVM based kernel techniques. Then to improve the performance, an automated system for detection of Adenocarcinoma lung cancer using novel catastrophe features is proposed. Finally, the experimental results are compared with the methods available in the literature.

In chapter 2, the biological details required to analysis and to detect lung cancer is given. The various methods available for the diagnosis and the screening of lung cancer were studied. A detail study on different modalities available to detect lung cancer was considered. The characteristics of different types of cancer such as squamous carcinoma and adenocarcinoma are studied.

In chapter 3, a thorough literature survey on Different modalities available for detection of cancer was studied, different methods available for delineating nuclear regions of other types of cancers, Cell image segmentation techniques, different features available for cancer cells, how to use multiple color spaces for cell image segmentation, scale space theory, deep structure, catastrophe points for feature extraction and finally the various classification techniques such as Support Vector Machine (SVM), Naïve Bayes Classifier And Artificial Neural Network used for classification of other types of Cancers cells are studied.

In chapter 4, a semi automated system for the classification of Glandular cells using Multiple Color Spaces was developed. In this method, the image is downsized for better processing and segmentation was done using different clustering techniques like K-means clustering and Fuzzy C-means clustering on various color spaces. Then features are extracted and classified using SVM

In chapter 5, classification of glandular cells using different SVM based kernel methods are given.
In chapter 6, automated lung cancer detection by the analysis of glandular cells in sputum cytology images using Scale space features is developed. In this method, cell region detection is done using novel Scale space theory, used Determinant of Hessian based region localization, ROF method for image Denoising, then segmentation is done using K-means clustering, then artifacts such as Red Blood Cells, Histocytes are removed using morphology, then catastrophe point based features are extracted using Scale Space stack. Finally the nuclei are classified as malignant and benign cells.

Chapter 7 gives the conclusion and future work of the proposed method. Flowchart for the overall thesis work is shown in the Figure 1.3.

![Figure 1.3 Flowchart of the overall thesis work](image-url)
1.6 SUMMARY

In this chapter, the following things are discussed:

1. Motivation for the research work.

2. Need of the proposed work.

3. Challenges faced while doing the research work.

4. Given the overview of the proposed system for the classification of the lung glandular cells as benign and malignant.

5. Finally, the overall organization of the thesis is given.