Chapter 1

INTRODUCTION

1.1 Medical Image Processing

1.1.1 Introduction to Digital Image Processing

“Image speaks sound than thousand words.” Vision is one of the main progressive senses of our human body, so digital images play an amazing role in our fast moving world. Images can be created from various devices such as cameras, x-rays, ultrasound, microscope and so on and this can be used for different fields such as medical, newspaper, industry, entertainment, military, education, business, scientific, and civil purpose. The main aim, in all these cases, is to obtain useful information from the images through the human observer or through the machine. In most of the cases, the raw image does not contain the necessary information. So, the processing of image is necessary for further study.

Digital Image Processing (DIP) is the term used to process digital images through digital computers. DIP is one of the prominent areas of electronic domain in which an image can be represented as an array of integers called “pixels” or “pels,” and it represents the features of a digital image. Interesting facts about digital image processing based on the application area include enhancement of pictorial definition and handling of the image data for storing, retrieving and transmission purposes.

The necessary steps in DIP are:

1
1. **Image Acquisition**: Before processing, the image is acquired by means of any one of the input devices. Generally, image acquisition stage needs pre-processing before they are processed.

2. **Image Enhancement**: Once the image is pre-processed, then it undergoes enhancement. Ultimately, the goal of enhancement is to stimulate certain region of interest for better visualization. The best example for enhancement is to increase the contrast of the image to ‘sharp’ the image visually.

3. **Image Restoration**: Another area that expands the detailed information of the image is ‘Image Restoration.’ Unlike enhancement, which is biased, image restoration is unbiased. That is, restoration deals with the mathematical models, whereas enhancement deals with human interpretation for ‘better’ visualization.

4. **Segmentation**: Segmentation process divides the image into various segments based on the algorithm. Obviously, the automatic segmentation process is a tedious task, as the machine does not know what exactly the human observer wants. An ‘uneven’ segmentation makes the process more complicated; on the other hand, an ‘erratic’ segmentation creates the system failure. To build the best segmentation, the algorithm must be error-free, simple to implement and understand and give best result. As a general rule, better the efficiency of segmentation, the recognition rate is also improved.

5. **Representation and Feature Selection**: The next stage of process always following the segmentation is ‘Representation and Feature Selection.’ The object may be represented as a discrete or continuous points based on the result obtained from the segmentation. The descriptors may be at boundary or at regional. The boundary descriptor is the one in which the objects boundary such as corners
and edges are characterized. In the regional descriptors, it is suitable only when the representation is internal, namely texture-based or shape-based. In both the cases, the data are converted into a procedure which is, compulsorily, understandable by a computer.

Feature selection deals with extracting the useful features from the pertained image, which represents either assessable information or basic for segregating one group of information to the other.

6. Subject Recognition: Once the useful features are extracted, in order to recognise the subject, assign a label to each subject based on the descriptors. For instance, consider the class of vehicle, in which the subject ‘bicycle’ comes under the label - ‘two-wheeler’.

1.1.2 Introduction to Medical Image Processing

Without knowing how the organs look like and which organs present inside the body, the physician could not know the perfect way to diagnose it. The purpose of examining the inner organs of the body clearly focuses the need for medical imaging. Medical Imaging is the procedure to construct the optical illustration of human organs without really opening them. Through medical imaging, the anatomy and the physiology of the normal organs are known; if there is any abnormality found, then that will be identified with the help of medical imaging.

In the normal context for viewing the content of a still picture or video there is no need of special equipment or persons’ are needed. In the clinical perspective, for viewing the medical imaging pictures, it needs a special person, namely ‘radiologist’ and the special equipment CT scanner or MRI scanner.

Benefits of Medical Applications Based on Digital Image Processing

3
• When the images are taken, the images are displayed instantaneously.

• Computations can be altered over time.

• Digital data could not be altered once it was taken and it retains their originality.

• Medical imaging is a dominant tool, as it guides the physician to take the right decision.

• It delivers a set of images to make easier to understand and interpret them to teach diseases and diagnosis.

• It is the easier way to compare other images.

Challenges in Medical Image Processing [4]

• Large amount of medical data from a different or single system ranging from kilobytes to terabytes, available in the database make the processing difficult.

• Specific software technique must be required for processing 3D [5] or 4D [6] data.

1.1.3 Cancer Identification

In normal cell division, a cell, in normal human body undergoes cell division and forms two new cells. In these two cells, one cell again undergoes cell division and the remaining one gets destroyed no longer survives. In this manner, the number of dividing and non-dividing cells are kept in balance. ‘Proliferation’ refers to the evolution and replica of the cells. ‘Apoptosis’ or ‘cell self-death,’ is the mechanism by which old or damaged cells, normally, destroy themselves. If these balanced processes get disturbed and the cells proliferate uncontrollably or fail to die off at the appropriate time, or both actions happen, it may cause cancer. Fig. 1.1 shows the difference between normal cell division and the cancerous cell division.
Our body already has a feedback system; if there is any deviation from the normal process, it recovers naturally. The feedback may be either positive or negative. Negative feedback is the one in which a negative result is obtained, once the system is triggered. For example, if we have a high fever, our body will eliminate sweat to reduce the warmth of the body. On the other hand, the positive feedback changes, in the system, will create the deviations in the same way. In the normal cell growth, the growth initiating factors will trigger the negative feedback in order to maintain a balance between the newly formed cells and the destroyed old cells. Conversely, in the anomalous cell reproduction, the growth initiating factors trigger the positive feedback.

Not all the anomalous cells are harmful. For instance, ‘hyperplasia’ refers to uncontrollable growth of cells that will not produce any harm. Hyperplasia may be the sudden changes in the body because of an external stimulus. For example, a lump formed on the hand when we first attempt to play tennis. ‘Dysplasia’ is another type of non-cancerous anomalous cell growth. In dysplasia, the cell loses their normal structure that can be easily realized under the microscope. An untreated dysplasia
may lead to cancer. The most severe form of dysplasia is ‘carcinoma in-situ’ (‘in-situ’ in Latin means ‘in place’). In this form of tumor, the anomalous cell growth sticks in the same place. It does not invade the neighboring organs. Different categories of anomalous cell growths are depicted in fig. 1.2.

![Diagram of Different Categories of Anomalous Cell Growths](image)

**Fig. 1.2: Different Categories of Anomalous Cell Growths**

### 1.1.4 Types of Cancer [1]

Though many types of cancers spread worldwide, the most important forms of cancers are discussed here. In general, the origin of the tumor generated is termed as ‘primary tumor’ and the tumor generates in one organ and spreads through blood and goes to other organs are termed as ‘secondary tumor.’

- **Brain tumor:** As the name implies, the origin of this tumor is in brain for primary brain tumor or the tumor generated somewhere and transported to brain via blood (metastasis) for secondary brain tumor. Based on the signs and
symptoms of the tumor, benign and malignant tumors are treated individually. By considering the tumor stages, it may vary based on the age of the patient, tumor size, location, how much the brain affected and how fast it is growing and also the treatment method may vary based on the above criterion. The most common types of brain tumor occur in children is ‘astrocytomas.’ But in adults, almost one-third of the brain tumor is caused by ‘glioblastomas’ and ‘anaplastic astrocytomas.’ The prognosis of brain tumor may be caused due to the risk factors such as tumor type, tumor size, age, and where the tumor is present.

- **Breast cancer:** Each breast contains lobules which create milk and ducts which carry the secreted milk to nipple. Blood tissues, in the breast, have fat and connective tissues, blood vessels and lymph nodes. The most casual form of breast cancer is ‘ductal carcinoma’, which creates in the duct cells. Breast cancer can also form owing to the misplacement of the cells in the lobules or other surrounding tissues in the ducts. Worldwide, breast cancer is the second most leading cancer occurrence. Breast cancer mostly occurs among women than men.

- **Bone cancer:** Bone cancer is an uncommon one. Mostly, all the bone cancers are harmless. One of the most common bone cancer is ‘osteosarcoma.’ It, generally, starts at the end of long bones; on the other hand it can form in any other bone. It, usually, occurs among elderly persons and teenagers. Unlike osteosarcoma, ‘ewing sarcoma’ forms in the ribs, in the middle of long bones, or hip bones. It, usually, occurs among young adults and teenagers. If a person takes the treatment of radiation, it will increase the possibility of osteosarcoma. Swelling in the bones, lump, and pain in the bone are the main signs and symptoms of bone cancer.
• *Cervical cancer:* The origin of cervix starts from the uterus to the vagina. It is found in the lower narrow end of the uterus. The main types of cervical cancers are ‘adenocarcinoma’ and ‘squamous cell carcinoma.’ The best way to reduce the risk of these types of cancer is vaccination - an ultimate remedy. If there is any abnormal findings in the Pap test or in HPV test, the observation from the cells shows the type of cells which may form the cervical cancer. Cervical cancer may be healed completely, if it is detected early.

• *Colon cancer:* Colon and rectum are found in the large intestine - the last stage of digestive system. The food we eat starts its journey from the stomach through the small intestine and finally reaches the colon. Inside the colon, it absorbs the nutrients from the food and disposes the waste through rectum. The common type of colon cancer is ‘adenocarcinoma’ - the cancer starts in the cells of the colon. Colon cancer threatens, worldwide, as it is the third most prevailing cancer among men and women. Blood tests and the use of colonoscopies may reduce the amount of deaths.

• *Liver cancer:* Liver is one of the important organs as it cleans all the toxins in the body, initiate to digest fat, to assist blood clot, besides storing and discharging sugar whenever necessary. Liver cancer is occasional in teenagers and children, but some types of liver cancer, namely ‘hepatoblastoma’ found among children and ‘hepatocellular carcinoma’ found among elder children and teenagers. Cancer can also arise inside and outside the liver of the bile ducts. Cancer found outside liver bile duct is more occasional than inside liver bile cancer.

• *Prostate cancer:* Prostate cancer is the secondmost common cancer next to skin cancer among men. Prostate is found in front of the rectum and below
the bladder. Mostly, all the prostate cancers are adenocarcinomas. In general, prostate cancer has no primary symptoms. In the advanced stage men may frequently or less frequently urinate - the result of a condition caused by benign prostate cancer. Unlike other types of cancers, the maturity of prostate cancer is very low. Deaths due to prostate cancer are rare and the cancer arises mostly among men over 65 years. Treatment for prostate cancer does not improve the health condition of the patient. So, preliminary steps are to be taken to avoid prostate cancer.

1.2 Role of Lung Cancer in Medical Image Processing

Until now, nobody knows the cause of cancer. Modernized life style may be the reason for the occurrence of cancer, besides the changes in the environment and history of cancer (i.e) their parents or their relatives having cancer. We cannot fix the boundary for cancer. In the recent years, the diagnosed rate of the cancer increases rapidly. Almost 17% of the deaths, worldwide, indicate cancer of all types as the reason. According to the worldwide reviews [7], the morbidity and the mortality of lung cancer is very high when compared to other cancers like breast, liver or colon cancer. ‘Cancer’ or literally ‘tumor’ is termed as an uncontrollable form of anomalous cells [8]. The anomalous cells divide rapidly and forms tumor. In general, the proportion of the old cells destroyed must be equal to the new cells formed. If this proportion delineates, then they produce anomalous cells.

In our day-to-day life, the number of people die of cancer is more than other sorts of diseases. Today, cancer is one among the most challenging health problems the
human beings find it difficult to tackle. Unlike other diseases, we neither predict the cause of cancer nor prevent it; but after diagnosing it, we undergo the treatment for further control.

1.2.1 Kinds of Lung Cancer

There are two kinds of tumors:

1. Benign or non-cancerous tumor
2. Malignant or cancerous tumor

1. Benign tumor: In contrast to cancerous tumors, benign tumors do not invade other parts of the body, because of their slow growth. This tumor can be found at any place in the body. Indeed, many tumors all over the body are benign. As it stick to a particular place, it is easy to remove this type of tumor by means of surgery, based on their site.

2. Malignant tumor: The word ‘cancer’ literally means the malignant tumors. This tumor is different from benign as they invade the surrounding organ, which is termed as ‘metastasis.’ This type of cancer can spread through the blood or through lymph nodes. It is one of the main causes where surgery may not be possible for this type of cancer. As surgery is unimaginable, the only remedy is opting different therapies such as chemotherapy, radiotherapy and even laser treatment applied over a wider range.

1.2.2 Indicators/Symptoms of Lung Cancer

If a tumor is present externally, namely skin tumor, we have to identify it as a tumor or cancer by seeing through our naked eyes. Instead, if the tumor is present inside
the internal organs, like lungs, brain or prostate cancer, the tumor type can not be identified easily. In such cases, the patients have to visit radiologists to detect the type and location of the cancer at an early stage and make survival possible.

Mostly, lung cancer may not have symptoms, until otherwise they spread widely. In some cases, the symptoms may arise in the early stages. If any of the following symptoms arises, one has to go to the physician to get diagnosed of the cancer in its early stage. The most frequent symptoms of lung cancer are:

- A cough that does not quit or get poorer.
- Chest pain is repeatedly worst with deep breathing, coughing or laughing.
- Hoarseness
- Weight loss or loss of appetite
- Coughing up blood
• Shortness of breath
• Feeling tired or weak
• Infections such as bronchitis and pneumonia that do not leave or retain back
• New commencement of wheezing

1.2.3 Categories of Lung Cancer [2]

Lung carcinoma (another name for ‘cancer’) is broadly classified as: 1) small cell lung cancer (SCLC) and 2) non-small cell lung cancer (NSCLC).

Of the total lung cancers, around 10%-15% consists of SCLC. This category of lung cancer spreads quickly and compactly. The SCLC is highly akin to cigarette smoking; and it occurs only 1% in non-smokers. Metastasis of SCLC is more likely to compare to the other categories and it is frequently revealed after they have spreaded widely.

Of all the cases, the most common lung cancer is NSCLC ranging from 85% to 90%. The NSCLC has 3 main types labeled by the types of cells present in the tumor. They are:

1. **Adenocarcinoma** is the most common type of NSCLC and it comprises 40% of the NSCLC cases. While all other types of lung cancer are related to smoking, this type is most common in non-smokers, especially, women. The growth of adenocarcinoma is primarily in the outer or peripheral areas of lungs. They also have a trend to blowout to the lymph nodes and outside.

2. **Squamous cell carcinoma** at present encompasses 25% of the NSCLC cases. This type of cancer arises more frequently in the central chest area in the bronchi. Unlike adenocarcinoma, this type sticks within the lung, rarely spread to lymph nodes.
3. Large cell carcinoma is the uncommon type of the NSCLC ranges from 10% - 15% of all the NSCLC cases and it is occasionally denoted as the indistinguishable carcinomas. This category of cancer has a high preference to spread to distance places and lymph nodes.

Other types of lung cancer comprise only about 5% - 10% of all lung cancer cases. One of them is Bronchial carcinoids which encompass 5% of lung cancers. This type of cancers are generally small (3-4 cm or less), when detected, and occur most frequently among individuals under 40 years.

1.2.4 Specific Attributes of Lung Cancer [3]

Initially, malignant tumors are found inside the main bronchi and may be in the peripheral system of the lungs. During differentiation, the peripheral lung tumor is, approximately, circular in projection and it is termed as nodules. As the size of the nodule in this area is very small, the normal radiographs can not view these type of nodules. Nodules, less than 1 cm in diameter, cannot be examined through normal radiographs. While considering the small size nodule, it should not be confused with scars, mimic swellings and vascular shadows. Radiologists can not differentiate the true and false nodules [9]. In order to overcome this situation cross-sectional imaging will help the radiologists to discover the nodules as small as 3mm in diameter. Observing the slices sequentially will help the radiologists to differentiate the blood vessels from the nodules [10]. As the malignant tumors grow very fast, they need large amount of blood supply; make the tumor irregular in shape and produce a fuzzy and speculated boundary. The main attribute of benign lung tumor is calcification. Benign calcification is opaque and sometimes asymmetrical and it is located in central or concentric.
1.2.5 Prevalence of Lung Cancer

Lung cancer stayed out as a disease till 1761; the first person who informed about the relation between lung cancer and smoking [11] was Fritz Lickint [12] from Germany. Nearly, one-in-ten adult deaths, world wide, are mainly due to tobacco usage. Lung cancer mortality rate is estimated to cross 10 million by 2030 world wide [13].

![Mortality of Lung cancer due to smoking per year](image)

**Fig. 1.4:** Mortality of Lung cancer due to smoking per year

The Indian Council of Medical Research (ICMR) had reviewed lung cancer statistics during the period 1982-2005 and revealed that the screening of lung cancer in one lakh male residents found an increase of lung cancer incidence rate in Chennai with 160%, Bangalore with 100%, Delhi with 40% and Mumbai having the fallen rate of 60%.

Smoking is responsible for about 80% of all lung cancers. The reason for the non-smokers being victims to lung cancer may be heredity, regular use of chemicals (if a person is working in a chemical company or a farmer), asbestos, and living under severe air pollution.

Table 1.1 predicts the point that although Adenocarcinoma is more common
among non-smokers, the Squamous cell carcinoma is the most common.

Table 1.1 depicts that the ratio of adenocarcinoma among smokers and non-smokers is approximately 2.2:1. Eventually, the ratio of squamous cell carcinoma is 1.2:1. When both the smokers and non-smokers are considered in squamous cell carcinoma and others types of cancer, the ratio is expected to be 4:1. In case of non-smokers, the affected percentage is same for both adenocarcinoma and squamous cell carcinoma. The risk factor of Small Cell Lung Cancer (SCLC) and other types of cancers is same for both the smokers and the non-smokers.
1.3 Various methods in Lung cancer identification

1.3.1 Lung Cancer Modalities

Cancer can be identified by using different modalities such as computed tomography (CT), magnetic resonance imaging (MRI), or positron emission tomography (PET) scan images. The CT scanning is suitable to catch in-depth three-dimensional (3D) images of certain parts such as lungs, brain, bones, abdomen, blood vessels and pelvis. The CT scan is the most likely desired process of diagnosing many cancers such as liver, lung [14], [15], and pancreatic cancers. The resultant image is confirmed by the radiologist for the existence of tumor. From the image, the physicians will predict the size of the tumor, its precise location and assess its intensity and determine how much it will affect the neighboring tissues.

The 3D image of the inside organ is generated by means of digital geometry projection of the CT scanner. The 3D image is generated by taking the same area in different angles and combines the 2-dimensional (2D) areas into 3D. Even though the CT is a suitable tool for assisting the radiologists to diagnose the treatment; the radiation emitting from the scanning unit may sometime cause cancer. Unless there is a necessity the doctors must not advise the patients to take CT scan.

In contrast to the X-ray machine, which drives only one radiation beam, the CT scanner releases a sequence of thin rays to the human body as it travels through an arc. The resultant output is extremely more detailed image when compared to X-rays. This information is transferred to the computer, which builds 3D cross-sectional image of the portion of the body and shows it on the monitor. Occasionally, a contrast dye is injected to show up the internal organs clearly on the output image. The patients having a history of allergic to some medicine must inform the physician before the
Fig. 1.6: Three Different Views of a CT image and their 3D View

test. The exactness and effectiveness of the CT scans may be enhanced by means of spiral CT. For a spiral CT scan of chest, for instance, the subject will be asked to hold their breath for a few seconds.

1.3.2 Clinical Dataset

The preliminary step in tumor detection is to analyse the existing system in terms of clinical perspective. To authorize the system is harmless and effective; it must be
tested on the platform for which it is intended. The designated use of a recognition
the system must be anticipated before specifying the suitable clinical test design.

Computer Aided Diagnosis (CAD) system is a completely automatic system having
the skill to determine the survival of an abnormality on its own. Infact the CAD system
is a competitor for the experts in the diagnostic field in terms of accuracy. Another
prospective application of this system is that it can constantly eliminate the normal
cases from the abnormal ones and it is useful for further review. The system is an
extremely sensitive and is able to differentiate the normal from the abnormal cases
with a higher degree of assurance. In the current medical image analysis, the CAD is
the most convincing use of a tumor detection system which acts as an assistance to
radiologists confidently. The system acts as a second-reader in which the performance
of the radiologists after using the system need be improved when compared to that of
the solo performance.

1.3.3 Sputum Testing

Specifically, several samples of sputum in cytological check-up is a supportive tool for
detecting tumors arising, especially, around larger bronchi such as small cell and squa-
mous cell carcinoma. Adenocarcinoma is a kind of peripheral tumor arising, especially,
in smaller airways and other carcinomas having a diameter of less than 2cm can be
spotted rarely through sputum samples. As the consumption of cigarette increases,
now-a-days, smoking related cancer such as adenocarcinoma get superior importance
[16], [17]. Sputum screening will yield only 20% to 30% sensitivity for primary lung
cancer and this yield will be improved by proper collection, analysing and processing
of data [18], [19].
1.4 Objective of the Research

The main objective of the research is to establish an Auto-Detective Computed Aided Diagnostic (AD-CAD) system to discover the lung nodules in Computed Tomography (CT) images in the following ways:

- To improve the quality of the image and to remove the noise and other artifacts through preprocessing.
- To segment the pre-processed image, various segmentation algorithms are used and an optimal segmentation algorithm is identified using similar measures.
- To select the features for classification, create a subset that contains both features gathered from the patients and also the features obtained from the CT images.
- To classify the features into benign (non-cancerous) or malignant (cancerous) tumors, highly effective classifiers used and the optimal one is discovered based on high accuracy and low error rate.

1.5 Scope of the Research

- Owing to the alarming rate of increase in the smoking habits among the adults, the incidence and the mortality of lung cancer increased statistically. Lung cancer is one among the leading causes of cancer deaths, worldwide and more particularly in India. As the symptoms of the lung cancer cannot exhibit explicitly, early detection of lung cancer is a bliss; the survival rate of lung cancer is more, if it is found early. To promote the early diagnosis of lung cancer, the patient must undergo screening immediately after the symptoms are noticed.
CAD system helps the physicians to take decisions and acts as a ‘second reader’ in the diagnosis of diseases. Of the existing imaging modalities, Computed Tomography (CT) is a well-defined one based on accuracy. It deliberately detects the pathological findings; even the smaller nodules. The misclassification error rate could be minimized as there is a possibility to find the small nodule using the CT. A large number of slices per image makes the modality highly complex and exhaust the radiologists to find the specific slice containing the tumor. As there is enormous number of slices, there is a chance to misunderstand that the blood vessels as nodules; because blood vessels also look bright in CT images like the nodules.

1.6 Methodology

The overall methodology aims to develop an Auto-Detective CAD (AD - CAD) system for the effective detection of lung nodules in CT images. Fig. 1.7 demonstrates the overall flow diagram of this study.

![Fig. 1.7: Overall Flow Diagram for Auto-Detective CAD](image)
The input is obtained from the CT scan images collected from the various scan centres and research hospitals. The first phase of the proposed Auto-Detective CAD is to preprocess the input image. Before preprocessing, the specified slice is taken into account by means of ABC algorithm. Once the appropriate slice is chosen by ABC algorithm, it undergoes preprocessing in order to eliminate the noise and other artifacts in the images. The second phase is the segmentation phase. Segmentation is a method to group or to segregate the image into meaningful regions. For this research, various segmentation algorithms are evaluated and the optimal one is identified based on the similarity methods. The third phase in the AD-CAD system is feature extraction. As soon as the tumor part is segregated from its original position, the features are extracted for further study. The extracted features that are obtained from the CT images and the attributes obtained from the ‘Query Form’ are taken into consideration for classification of nodules. The final outputs that are obtained from the AD-CAD system are to recognize whether the nodule belongs to benign group or malignant group.

1.7 Organization of the Thesis

Chapter 1 represents the Introduction to the Thesis. The introductory chapter contains the introduction to Medical Image Processing, role of Lung Cancer in Medical Image Processing, various methods in Lung Cancer identification, objective, scope, and contribution of the research and the overall research methodology.

Chapter 2 deals with the Literature surveyed to focus the general background, literature about Lung Cancer, Preprocessing, Segmentation, Feature extraction, and Classification.

Chapter 3 throws light on the Preprocessing of the input images by various filters
and the best one is chosen based on PSNR and MSE values.

**Chapter 4** deals with the Segmentation process in which various segmentation algorithms are discussed, evaluated and the optimal segmentation is obtained.

**Chapter 5** explains the Feature extraction phase covering the features obtained from the CT images and also the details gathered directly from the subjects.

**Chapter 6** signifies the Classification process in which it deals with different classification algorithms and the best algorithm chosen based on the Accuracy and Error rate.

**Chapter 7** provides the Summary and Conclusion of the thesis and the future enhancement of the thesis.