**LUNG CANCER DETECTION BASED ON CT-SCAN IMAGES WITH DETECTION FEATURES USING GRAY LEVEL CO-OCCURRENCE MATRIX (GLCM) AND SUPPORT VECTOR MACHINE (SVM) METHODS**

A Mini Project Work

Submitted in partial fulfillment of the requirements for the award of the degree of

# **BACHELOR OF TECHNOLOGY**

**IN**

**ELECTRONICS AND COMMUNICATION ENGINEERING**

By

Ganesh Bhumanna Alchettiwar 20EG104312

Under the guidance of

**Mr.L.Praveen Kumar**

Assistant professor

 Department of ECE

**ANURAG UNIVERSITY**

**Venkatapur(V), Ghatkesar(M), Medchal-Malkajgiri, Dist-500088**

**2023-2024**

**ANURAG UNIVERSITY**

Venkatapur(V),Ghatkesar(M), Medchal-Malkajgiri

Dist-500088

 **DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING**

# **CERTIFICATE**

This is to certify that the project report entitled **“Lung Cancer Detection Based On CT-Scan Images with Detection Features Using Gray Level Co-Occurrence Matrix (GLCM) and Support Vector Machine (SVM) Methods”**

being submitted by

Ganesh Bhumanna Alchettiwar 20EG104312

in partial fulfillment for the Degree of Bachelor of Technology in Electronics & Communication Engineering award to the Anurag University, Hyderabad is a record of bonafide work carried out under my guidance and supervision. The results embodied in this project report have not been submitted to any other University or Institute for the award of any Degree or Diploma.

**Mr.L.Praveen Kumar** **Prof. N. Mangala Gouri**

Assistant Professor Head of the Department

DEPT OF ECE

**EXTERNAL EXAMINER**

**ACKNOWLEDGEMENT**

This project is an acknowledgment of the inspiration, drive and technical assistance contributed by many individuals. This project would have never seen the light of day without the help and guidance we have received. We would like to express our gratitude to all the people behind the screen who helped us to transform an idea into a real application.

It’s our privilege and pleasure to express our profound sense of gratitude **to Mr.L.Praveen Kumar, Assistant Professor**, Department of ECE for his guidance throughout this dissertation work.

We express our sincere gratitude to **Prof. N. Mangala Gouri, Head of the Department, of** Electronics and Communication Engineering for his precious suggestions for the successful completion of this project. He is also a great source of inspiration to our work.

We would like to express our deep sense of gratitude to **Dr.V.Vijaya Kumar**, **Dean**, Anurag University for his tremendous support, encouragement, and inspiration. Lastly, we thank almighty, our parents, and friends for their constant encouragement without which this assignment would not have been possible. We would like to thank all the other teaching and non-teaching staff members who have extended their timely help and eased my work.

**BY**

Ganesh Bhumanna Alchettiwar 20EG104312

**DECLARATION**

We hereby declare that the result embodied in this project report entitled “**Lung Cancer Detection Based On CT-Scan Images with Detection Features Using Gray Level Co-Occurrence Matrix (GLCM) and Support Vector Machine (SVM) Methods**” is carried out by us during the year 2023-2024 for the partial fulfillment of the award of **Bachelor of Technology** in **Electronics and Communication Engineering,** from ANURAG UNIVERSITY. We have not submitted this project report to any other Universities / Institutes for the award of any degree.

## **BY**

Ganesh Bhumanna Alchettiwar 20EG104312

**ABSTRACT**

Lung cancer is all malignant diseases in the lungs, including malignancies originating from the lungs themselves (primary) or those originating from other organs (metastasis). Lung cancer is one of the leading causes of death worldwide. Lung cancer is a tumor that grows rapidly and can spread to other organs. The onset of cancer is characterized by abnormal cell growth that can damage other normal tissue cells. Computerized Tomography (CT) is an imaging technique often used to diagnose lung cancer. Lung cancer can be classified into benign and malignant cancer. It is very important to diagnose lung cancer at an early stage to speed up the treatment process and the actions that will be taken. This study aims to develop a lung cancer detection system based on CT-scan images. This detection system has 4 main stages, namely pre-processing of CT-Scan images to improve image quality, segmentation to identify and separate the desired cancer object from the background, feature extraction based on area, contrast, energy, entropy, and homogeneity. The classification of lung cancer into cancer benign and malignant cancer. From the system trial, the accuracy level based on the system decision in determining the diagnosis of lung cancer is benign or malignant.

**Keywords:** CT-Scan Image, Classification, Lung Cancer, Segmentation, System Detection, Introduction.

TABLE OF CONTENTS

[CERTIFICATE ii](#_Toc147335524)

[ACKNOWLEDGEMENT iii](#_Toc147335525)

[DECLARATION iii](#_Toc147335527)

[ABSTRACT iii](#_Toc147335527)

**Chapter No. Title Page No.**

1. [INTRODUCTION 1](#_Toc147335529)

2. [LITERATURE SURVEY 4](#_Toc147335530)

3. [EXISTING METHOD 8](#_Toc147335531)

3.1 KNN Classifier 8

[3.2 Random Forest Classifier 9](#_Toc147335536)

[3.3 Naïve-Bayes Classifier 12](#_Toc147335537)

4. [PROPOSED METHOD 14](#_Toc147335538)

[4.1 Gray Level Co-occurrence Matrix based Texture analysis 20](#_Toc147335539)

[4.2 Support vector machine 22](#_Toc147335540)

4.2.1 SVM Classifier22

[4.2.2 Types of SVM 23](#_Toc147335541)

**5.ADVANTAGES AND APPLICATIONS** 28

6. [INTRODUCTION TO MATLAB 29](#_Toc147335544)

6.1 Brief History of MATLAB 30

6.2 EISPACK and LINPACK 31

6.3 Starting MATLAB 32

6.4 Features of MATLAB 36

6.5 Applications of MATLAB 37

[7. RESULTS 41](#_Toc147335548)

8. [CONCLUSION 43](#_Toc147335549)

9. [REFERENCES 44](#_Toc147335550)

**10**. [**APPENDIX-PROJECT CODE** 46](#_Toc147335550)

**CHAPTER 1**

# **Introduction**

Lung cancer is a global health concern that continues to claim countless lives annually. It is a complex disease characterized by uncontrolled cell growth in lung tissues and is primarily associated with cigarette smoking, exposure to carcinogens, and genetic predisposition. Early detection is critical for improving patient outcomes, as lung cancer is often asymptomatic in its early stages, making it challenging to diagnose until it reaches an advanced, less treatable phase. Computed Tomography (CT) scans have emerged as a vital tool in the early detection and diagnosis of lung cancer, offering detailed imaging of lung tissues. Leveraging advanced image processing techniques, such as the Gray Level Co-Occurrence Matrix (GLCM) and machine learning methods, particularly the Support Vector Machine (SVM), has shown immense promise in enhancing the accuracy and efficiency of lung cancer detection from CT-scan images. In this comprehensive exploration, we delve into the landscape of lung cancer, the role of CT scans, the principles of GLCM, and the capabilities of SVM, all culminating in a promising approach to revolutionize lung cancer detection.

**1. The Global Burden of Lung Cancer**

Lung cancer has remained a pervasive global health challenge for decades. According to the World Cancer Research Fund, it was estimated that lung cancer was responsible for approximately 2.2 million new cases and 1.8 million deaths in 2020, making it the most common cause of cancer-related deaths worldwide. The grim statistics associated with lung cancer underscore the urgent need for more effective diagnostic tools and early detection methods to improve patient survival rates.

**2. The Role of Computed Tomography (CT) Scans in Lung Cancer Diagnosis**

Computed Tomography (CT) imaging has revolutionized the field of medical diagnostics, providing high-resolution, cross-sectional images of the body's internal structures. In the context of lung cancer, CT scans have become indispensable for their ability to detect even the subtlest abnormalities in lung tissues. The use of CT scans for lung cancer screening, diagnosis, and monitoring has significantly improved the chances of early detection and treatment.

**3. Gray Level Co-Occurrence Matrix (GLCM): A Key Image Processing Technique**

Image processing techniques have become essential in extracting valuable information from medical images, including CT scans. One of the most powerful tools in this domain is the Gray Level Co-Occurrence Matrix (GLCM). GLCM is a statistical method that analyzes the spatial relationships between pixel values in an image. By quantifying the occurrence of specific gray level patterns, GLCM can capture texture information that is often imperceptible to the human eye. In the context of lung cancer detection, GLCM can be applied to CT-scan images to highlight

subtle textural changes in lung tissues, potentially indicative of cancerous growth.

**4. Support Vector Machine (SVM): A Robust Machine Learning Technique**

Machine learning has gained prominence in the realm of medical image analysis, and the Support Vector Machine (SVM) is a versatile and widely used algorithm. SVM is a supervised learning method that excels in classification tasks. It works by finding the optimal hyperplane that best separates data points into distinct classes, making it particularly well-suited for distinguishing between cancerous and non-cancerous regions in CT-scan images. The combination of GLCM and SVM holds the promise of enhancing the accuracy of lung cancer detection by extracting intricate texture features and classifying them with remarkable precision.

**5. The Need for Advanced Lung Cancer Detection Methods**

Despite significant advancements in medical imaging and machine learning, the accuracy and efficiency of lung cancer detection remain areas of concern. Conventional methods often rely on manual interpretation of CT scans, which is subject to interobserver variability and can be time-consuming. Moreover, early-stage lung cancers can be subtle and challenging to detect with the naked eye, emphasizing the need for automated and highly accurate methods.

**6. The Promise of GLCM and SVM in Lung Cancer Detection**

Combining GLCM's texture analysis capabilities with SVM's classification prowess presents a potent solution to the challenges associated with lung cancer detection. GLCM can extract a comprehensive set of texture features from CT-scan images, such as contrast, energy, and homogeneity, providing a rich source of information about the lung tissues. SVM can then efficiently classify these texture features, effectively distinguishing between malignant and benign regions. The synergy between these two techniques has the potential to enhance the sensitivity and specificity of lung cancer detection while reducing false positives and false negatives.

This exploration sets the stage for a comprehensive investigation into lung cancer detection based on CT-scan images, with a primary focus on leveraging GLCM and SVM techniques. As we delve deeper into the realms of image processing and machine learning, we will uncover the mechanisms behind these methods, their applications in lung cancer detection, and their potential to revolutionize the field by providing a more accurate, efficient, and automated approach to diagnosing this deadly disease. By harnessing the power of technology and medical imaging, we embark on a journey toward earlier and more reliable lung cancer detection, ultimately improving patient outcomes and reducing the global burden of this devastating disease.

Lung cancer also familiar as lung carcinoma caused by malignant lung tumor which have uncontrolled cell growth. This cell growth can spread to the other part of the body by the metastasis process. Mainly there are two types of lung cancer, one is small-cell lung cancer (SCLC) and the other one is no small-cell lung cancer (NSCLC). The primary symptoms of lung cancer are coughing, losing weight, breath shortness and chest pain. One of the main reasons of lung cancer is smoking besides being a passive smoker, air pollution and genetic factors are also responsible for lung cancer. Avoiding smoking with other risking factors can prevent lung cancer primarily. Early detection of lung cancer stage is important for many reasons. Firstly, it helps a patient to choose proper treatment in its early stage since the treatment for one stage may not be appropriate for another stage. For example, if the cancer has spread to different parts of the body, like brain, liver and bones, in that case surgery will not improve the situation. Secondly, early detection describes how much cancer has spread throughout the body. Estimation shows that patient’s survival rate for 5 years can increase from 14% to 49% just by detecting earlier. Lung cancer stage can be divided into limited stage and excessive stage. In limited stage, cancer is confined in one lung, and in excessive stage cancer has spread to the other parts of the body. In this study, CT images have chosen as it is more efficient compared to X-ray for detecting lung cancer stages. CT scan generates cross sectional images of the body that enables to visualize small nodules or tumors. Acute and chronic changes in the lung parenchyma can also be detected by CT scan where normal two-dimensional X-ray fails to detect it. So, considering all these doctors prefer CT images rather than X-ray to diagnosis lung cancer. The main reasons to choose digital image processing techniques (DIP) for detecting lung cancer stage is image gives better visualization and information compared to other forms. Image processing techniques are easier way to analyze image cells and extract data from them. According to our knowledge a few research work has been done on Lung Cancer stage detection. In this study, smoothing, enhancement, segmentation, morphological opening and region of interest (ROI) generation have been performed for image preprocessing. For extracting features gray level co-occurrence matrix (GLCM) based texture image analysis and Statistical parametric approach have been followed. GLCM provides texture pairs with similar type second order statistics, which cannot be differentiate by human eye. It removes the unnecessary features and improve the accuracy of classification. In this study, statistical parametric approach has least number of features compared to GLCM approach. As the number of features increases, dimension increases on the ten to the power of number of features. Using statistical parametric approach computation is not getting complex compared to GLCM approach. Statistical parametric approach gives best result on our dataset. Support Vector Machine (SVM), classifiers used for learning purposes.

**CHAPTER 2**

# **LITERATURE SURVEY**

**[1] Y. F. Riti, H. A. Nugroho, S. Wibirama, B. Windarta, and L. Choridah,** "Feature extraction for lesion margin characteristic classification from CT Scan lungs image," 2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, 2016, pp. 54-58.

Lung cancer is one of the common cancers which occurred in both male and female. Revealed by WHO data, in 2012, this disease become one of the major causes of death in worldwide with the mortality rate about 1.59 million. An early detection of lung cancer by using Computed Tomography (CT) Scan can provide more opportunity to survive. However, the diagnosis of lung cancer by reading the CT scan image which performed by radiologists may lead to an error. A computer-based digital image processing is a solution to improve the accuracy and consistency in reading the CT Scan image result. This study aim is to identify the morphological characteristic of regular and irregular margins by using feature extraction method. In this research, image processing divided into several stages refer to the segmentation process with Otsu method, feature extraction with number of features such as convexity, solidity, circularity, and compactness, and the last is classification by using Multi-Layer Perceptron (MLP). The classification process of features convexity, solidity, circularity, and compactness, resulted in the accuracy value of 85%, sensitivity of 85%, and specificity of 85%.

**Summary:** This study aim is to identify the morphological characteristic of regular and irregular margins by using feature extraction method

**[2] F. Taher, N. Werghi, and H. Al-Ahmad,** "Computer-aided diagnosis system for early lung cancer detection," 2015 International Conference on Systems, Signals and Image Processing (IWSSIP), London, 2015, pp. 5-8.

Lung cancer continues to rank as the leading cause of cancer deaths worldwide. One of the most promising techniques for early detection of cancerous cells relies on sputum cell analysis. This was the motivation behind the design and the development of a new computer aided diagnosis (CAD) system for early detection of lung cancer based on the analysis of sputum color images. The proposed CAD system encompasses four main processing steps. First is the preprocessing step which utilizes a Bayesian classification method using histogram analysis. Then, in the second step, mean shift segmentation is applied to segment the nuclei from the cytoplasm. The third step is the feature analysis. In this step, geometric and chromatic features are extracted from the nucleus region. These features are used in the diagnostic process of the sputum images. Finally, the diagnosis is completed using an artificial neural network and support vector machine (SVM) for classifying the cells into benign or malignant. The performance of the system was analyzed based on different criteria such as sensitivity, specificity and accuracy. The evaluation was carried out using Receiver Operating Characteristic (ROC) curve. The experimental results demonstrate the efficiency of the SVM classifier over other classifiers, with 97% sensitivity and accuracy as well as a significant reduction in the number of false positive and false negative rates.

**Summary:** Studied aboutReceiver Operating Characteristic (ROC) curve

**[3] E. Rendon-Gonzalez and V. Ponomaryov,** "Automatic Lung nodule segmentation and classification in CT images based on SVM,”, Millimeter and Submillimeter Waves (MSMW), Kharkiv, 2016, pp. 1-4.

Early detection of lung cancer is of vital importance to successful treatment where Computed Tomography (CT) screening are considered one of the best methods for detection the early signs of lung cancer. Standard Computer Aided Diagnosis (CAD) systems for Lung cancer detection should employ four steps: preprocessing, lungs parenchyma segmentation, nodule detection and reduction of False Positives (FP). In the proposed approach during the preprocessing step, several masks are calculated using thresholding technique and morphological operations, eliminating this way, background and surrounding tissue. Following, suspicious Regions of Interest (ROI) are calculated using a priori information and Hounsfield Units (HU). During feature extraction, numerous features are calculated in order to restrict the suspicious zones. Finally, Support Vector Machine (SVM) algorithm is employed in classification stage.

**Summary:** Study aboutsuspicious Regions of Interest (ROI) are calculated using a priori information and Hounsfield Units (HU)

**[4] R. Wulandari, R. Sigit, and S. Wardhana**, "Automatic lung cancer detection using color histogram calculation," 2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC), Surabaya, 2017, pp. 120-126.

Lung cancer is a disease that caused by uncontrolled cell growth in lung. Lung cancer is still the first worldwide killer. CT Scan Thorax is a method for early detection of lung cancer patients. However, cancer detection in lung CT-Scan image still done manually. In this paper, the segmentation of lung image is proposed. Cancer segmentation will process the lung CT-Scan as an image input with watershed process to cut off cavity area. The result will be processed by color histogram calculation to obtain mean and standard deviation value. This value is useful for evaluate non-cancer area and produce cancer image. Segmentation process will be followed by measurement of cancer and cavity area. The overall output is percentage between the large of cancer area and cavity area. The experiment represented that this method is able to detect lung cancer automatically. The performance segmentation for assessment errors obtained an average cavity area segmentation 12.75% and cancer area segmentation 31.74%.

**Summary:** Studied Abouthistogram calculation to obtain mean and standard deviation value.

**[5] L. Anifah, Haryanto, R. Harimurti, Z. Permatasari, P. W. Rusimamto, and A. R. Muhamad,** "Cancer lung detection on CT scan image using artificial neural network backpropagation based Gray level cooccurrence matrices feature," 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Bali, 2017, pp. 327-332.

Lung cancer is the most common cause of cancer death in the world. Early detection of lung cancer will greatly help to save the patient. This research focuses on detection of lung cancer using Artificial Neural Network Back-propagation based Gray Level Co-occurrence Matrices (GLCM) feature. The lung data used originates from the Cancer imaging archive Database, data used consisted of 50 CT-images. CT-image is grouped into 2 clusters, normal and lung cancer. The steps of this research are: image preprocessing, region of interest segmentation, feature extraction, and detection of lung cancer using Neural Network Back-propagation. The results show system can detect CT-image of normal lung and lung cancer with accuracy of 80%. Hopefully use to help medical personnel and research to detect lung cancer status.

**Summary:** Study AboutBack-propagation based Gray Level Co-occurrence Matrices (GLCM)

**[6] D. P. Kaucha, P. W. C. Prasad, A. Alsadoon, A. Elchouemi, and S. Sreedharan**, "Early detection of lung cancer using SVM classifier in biomedical image processing," 2017 IEEE

International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai, 2017, pp. 3143-3148.

Image processing techniques are now commonly used in the medical field for early detection of diseases. This research aims to improve accuracy, sensitivity and specificity of early detection of lung cancer through a combination of image processing techniques and data mining. The Computed Tomography (CT) scan image of the lungs is pre-processed and the Region of Interest (ROI) segmented, retained and compressed using a DWT (Discrete Waveform Transform) technique. The resulting ROI image is decomposed into four sub frequencies, bands LL, HL, LH, and HH. Again, the LL sub frequency is decomposed into four sub-bands, applying a 2-level DWT to the ROI based image. Further, features such as entropy, co-relation, energy, variance and homogeneity are extracted from the 2-level DWT images using a GLCM (Gray level Co-occurrence Matrix) with classification effected by means of an SVM (Support Vector Machine). Classification identifies whether the CT image is normal or cancerous. The Lung Image Database Consortium dataset (LIDC) has been used for training and testing purpose for this study. A Receiver Operating Characteristics (ROC) curve is used to analyse the performance of the system. Overall, the system has accuracy of 95.16%, sensitivity of 98.21% and specificity of 78.69%.

**Summary:** Study AboutDWT (Discrete Waveform Transform) technique

**CHAPTER 3**

# **EXISTING METHOD**

In the existing system, there are several approaches for identifying the stages of lung cancer i.e. limited stage and excessive stage.

The different classifiers are Naïve-Bayes, Random Forest and KNN based algorithms.

In these classifiers, both the Gray level co-occurrence matrix approach and the statistical approach were used for feature extraction of the images.

In this also, the free hand technique is used for segmentation of the lungs part in the image and is later fed to classifier.

**3.1 KNN Classifier:**

KNN: K Nearest Neighbor is one of the fundamental algorithms in machine learning. Machine learning models use a set of input values to predict output values. KNN is one of the simplest forms of machine learning algorithms mostly used for classification. It classifies the data point on how its neighbor is classified.

KNN classifies the new data points based on the similarity measure of the earlier stored data points. For example, if we have a dataset of tomatoes and bananas. KNN will store similar measures like shape and color. When a new object comes it will check its similarity with the color (red or yellow) and shape.

K in KNN represents the number of the nearest neighbors we used to classify new data points.

In real-life problems where we have many points the question arises is how to select the value of K?

Choosing the right value of K is called parameter tuning and it’s necessary for better results. By choosing the value of K we square root the total number of data points available in the dataset.

a. K = sqrt (total number of data points).

b. Odd value of K is always selected to avoid confusion between 2 classes.

Usage of KNN :

a. We have properly labeled data. For example, if we are predicting someone is having diabetes or not the final label can be 1 or 0. It cannot be NaN or -1.

b. Data is noise-free. For the diabetes data set we cannot have a Glucose level as 0 or 10000. It’s practically impossible.

c. Small dataset.

**Working of KNN**

We usually use Euclidean distance to calculate the nearest neighbor. If we have two points (x, y) and (a, b). The formula for Euclidean distance (d) will be

d = sqrt((x-a)²+(y-b)²)

**3.2 Random Forest Classifier:**

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of **ensemble learning,** which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

As the name suggests, **"Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset."** Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

**The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.**

## Assumptions for Random Forest

Since the random forest combines multiple trees to predict the class of the dataset, it is possible that some decision trees may predict the correct output, while others may not. But together, all the trees predict the correct output. Therefore, below are two assumptions for a better Random forest classifier:

* There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
* The predictions from each tree must have very low correlations.

## Why use Random Forest?

Below are some points that explain why we should use the Random Forest algorithm:

<="" li="">

* It takes less training time as compared to other algorithms.
* It predicts output with high accuracy, even for the large dataset it runs efficiently.
* It can also maintain accuracy when a large proportion of data is missing.

## How does Random Forest algorithm work?

Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps and diagram:

**Step-1:** Select random K data points from the training set.

**Step-2:** Build the decision trees associated with the selected data points (Subsets).

**Step-3:** Choose the number N for decision trees that you want to build.

**Step-4:** Repeat Step 1 & 2.

**Step-5:** For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

**3.3 Naïve-Bayes Classifier:**

* Naïve Bayes algorithm is a supervised learning algorithm, which is based on **Bayes theorem** and used for solving classification problems.
* It is mainly used in text classification that includes a high-dimensional training dataset.
* Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
* **It is a probabilistic classifier, which means it predicts on the basis of the probability of an object**.
* Some popular examples of Naïve Bayes Algorithm are **spam filtration, Sentimental analysis, and classifying articles**.

**Naives’ Bayes Theorem**

The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, Which can be described as:

* **Naïve**: It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.
* **Bayes**: It is called Bayes because it depends on the principle of Bayes' Theorem.

## Bayes' Theorem:

* Bayes' theorem is also known as **Bayes' Rule** or **Bayes' law**, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.
* The formula for Bayes' theorem is given as:

Naïve Bayes Classifier Algorithm

**Where,**

**P(A|B) is Posterior probability**: Probability of hypothesis A on the observed event B.

**P(B|A) is Likelihood probability**: Probability of the evidence given that the probability of a hypothesis is true.

**P(A) is Prior Probability**: Probability of hypothesis before observing the evidence.

**P(B) is Marginal Probability**: Probability of Evidence.

## Working of Naïve Bayes' Classifier:

Working of Naïve Bayes' Classifier can be understood with the help of the below example:

Suppose we have a dataset of **weather conditions** and corresponding target variable "**Play**". So using this dataset we need to decide that whether we should play or not on a particular day according to the weather conditions. So to solve this problem, we need to follow the below steps:

Convert the given dataset into frequency tables.

1. Generate Likelihood table by finding the probabilities of given features.
2. Now, use Bayes theorem to calculate the posterior probability.

**Disadvantages:**

* Performance is low compared to the proposed system.
* Naive Bayes assumes that all predictors (or features) are independent, rarely happening in real life. This limits the applicability of this algorithm in real-world use cases.
* The main limitation of random forest is that a large number of trees can make the algorithm too slow and ineffective for real-time predictions.
* For KNN, Accuracy depends on the quality of the data and with large data, the prediction stage might be slow.

**CHAPTER 4**

# **PROPOSED METHOD**

Computerized Tomography (CT) is one of the imaging techniques often used in diagnosing lung cancer. Lung cancer with pathological residues of various diameters and sizes can be seen with a CT scan. Lung cancer is categorized into benign and malignant cancer. During diagnosis, cancer of a certain density and atypical can be assessed as benign cancer in some cases. However, in many cases, the congestion of the lung is usually categorized as malignant. It is important to diagnose lung cancer at an early stage to speed up the treatment and treatment process. Systems designed for medical applications can provide a variety of benefits to be successful in detecting lung cancer. The treatment process can be started early with the help of this system and can facilitate the doctor's decision-making process quickly and accurately by using machine leaning Algorithm as Super vector Machine (SVM). In many cases, the congestion of the lung is usually categorized as malignant. It is important to diagnose lung cancer at an early stage to speed up the treatment and treatment process. Systems designed for medical applications can provide a variety of benefits to be successful in detecting lung cancer.

Input Image

Gray Image

Image denoising

Binary Image

Segmented Image

GLCM Feature Extraction

SVM model

Classification

* Benign
* malignant
* Normal

Classified Accuracy

**Figure 1:** **Block Diagram of Proposed System**

1. **Image acquisition**:

At first, CT scan images of patients have acquired. The reason behind to choose CT scan image is that it is more efficient than X-ray and MRI images. CT images are less distorted and have low noise compared to X-ray and MRI images. From cancerimagingarchive.net site CT images are acquired. All the images are in DICOM (Digital Imaging and Communications in Medicine) format. In medical imaging DICOM is standard format. Whether it is CT image or X-ray image it has some noise. So, in order to remove noise images were preprocessed. Sample slice of our acquired image is presented in below figure.

1. **Pre-processing**

In this study, image preprocessing includes smoothing, enhancement, segmentation, morphological opening and region of interest (ROI) selection.

**Gray Image:**

A grayscale or gray image, often referred to simply as a "gray image," is a digital image in which each pixel represents the intensity of light or color in a single channel. Unlike full-color images, which use three channels (red, green, and blue) to represent colors, gray images use only one channel to represent the level of brightness. This means that gray images are essentially black and white, with varying shades of gray in between.

Gray images are widely used in various fields, including photography, medical imaging, computer vision, and graphics. They are particularly useful when color information is not critical, or when simplifying an image for analysis or display. The intensity of each pixel in a gray image typically ranges from 0 (black) to 255 (white), with values in between representing various shades of gray.

**Denoise Image:**

Image denoising is a process of reducing or removing noise from a digital image. Noise in an image is random variations in pixel values caused by factors such as electronic interference, sensor limitations, or imperfections in the imaging process. Denoising techniques aim to enhance the quality and clarity of an image by smoothing out these unwanted fluctuations while preserving important details and structures.

There are various methods for denoising images, ranging from simple linear filters like Gaussian blur to more advanced techniques such as wavelet denoising and deep learning-based approaches. These methods analyze the neighboring pixels of each pixel in the image and adjust their values to reduce noise. The choice of denoising method depends on the specific characteristics of the noise and the desired trade-off between noise reduction and preservation of image details.

**Binary Image:**

A binary image is a type of digital image in which each pixel can have one of two possible values: typically, 0 (representing black) or 1 (representing white). Binary images are primarily used to represent objects or shapes within an image by isolating them from the background. In other words, binary images are used for segmentation, object recognition, and other tasks where the focus is on the presence or absence of objects rather than their color or intensity.

Binary images are often generated by thresholding a grayscale image. Thresholding involves setting a specific intensity level as a threshold; pixels with intensities above this threshold are assigned the value 1 (white), while pixels below the threshold are assigned the value 0 (black). This process effectively converts a grayscale image into a binary one, highlighting objects of interest and simplifying subsequent image analysis tasks.

**Segmented Image:**

A segmented image is the result of a process called image segmentation, which involves dividing a digital image into multiple distinct regions or segments based on certain criteria, such as color, intensity, texture, or shape. The goal of image segmentation is to partition an image into meaningful and homogeneous regions to simplify analysis and facilitate object recognition, tracking, or measurement.

There are various methods for image segmentation, including:

**1. Thresholding:** This method involves separating regions in an image based on a chosen threshold value, often resulting in a binary segmented image.

**2. Region Growing:** Region growing starts with a seed point and grows regions by adding neighboring pixels that meet certain criteria, such as similarity in intensity.

**3. Edge-Based Segmentation:** Edge detection algorithms identify boundaries between different regions in an image, which can then be used to segment it.

**4. Clustering Algorithms:** Techniques like K-means clustering can group pixels into clusters based on similarity in color or intensity.

**5. Watershed Segmentation:** This technique treats the image as a topographic surface and separates regions at watershed lines.

Segmented images are valuable in various applications, including medical image analysis, object tracking, computer vision, and remote sensing. Once an image is segmented, the resulting regions can be analyzed individually, making it easier to extract meaningful information and perform specific tasks on different parts of the image.

grayscale images represent brightness levels, denoising techniques remove unwanted noise, binary images simplify object detection, and segmented images divide an image into meaningful regions. These concepts are fundamental in the field of image processing and computer vision, enabling the extraction of valuable information from digital images in a wide range of applications.

1. **Enhancement**:

Enhancement is a technique to improve image quality. After smoothing still our images are not readable. In order to improve the interpretability, we enhanced the images. There is no general rule about choosing enhancement technique. There are various image enhancement techniques such as Gabor filter, Auto enhancement and Fast Fourier transform. In this study, Gabor filter is used for enhancement purpose as it gives better result compared to Auto enhancement and Fast Fourier transform. Before and after enhancement images are presented in below figure,

1. **Segmentation**:

Segmentation is a technique to partition image into multiple segments. In this study, images have been segmented to simplify images more meaningfully. For segmentation purpose, in this study otsu method is used. Otsu method also finds an optimal threshold to convert a gray level image into a binary image Threshold is chosen in such a way so that it minimizes the weighted sum within class variance shown in below figure,

**E.Morphological Opening**:

Morphological Opening is the process where erosion takes place followed by dilation. After segmentation of CT scan images still there exist other organs like blood vessels, bronchioles with lung nodules. In order to remove other organs and only to keep lung nodules morphological opening performed in this study. For performing morphological opening, a flat, disk-shaped structuring element with having a radius nine have been used. In this study, Morphological Opening removes any parts that is smaller than the structuring element with having radius nine. After performing morphological opening on the segmented images lung nodules exist in the lung with radius nine thus morphological opening increases efficiency and accuracy and it is one of the significant findings of this study. According to our knowledge, in other research related to lung cancer staging no one applied this technique to detect lung nodule specifically before us. The reason is to choose radius nine is that it gives better result for all the images with clear view of lung nodules. Morphological Opening does not change the basic core size of the shape.

**F.Selection of Region of Interest (ROI)**:

In this study, Region of interest (ROI) is a specific portion of lung image where lung nodules or tumors found and will perform operations to extract feature values. Free hand method has been used for selecting ROI. Lung is chosen as object then drew a polygon around it. A mask image is created which is later multiplied with each of the elements of the input image and finally generates a new binary image which is Region of interest (ROI). Mask is nothing but a filter and also known as spatial filtering. By moving the mask from pixel to pixel over an image filtering has been done. When drawing around the object a wait function has been calculated until the ROI is generated. Freehand method provides the facility to adjust the size and position of the ROI just by using the mouse. Below figure shows generated ROI:

**G.Feature extraction**:

In this study, Feature extraction stage includes Feature extraction on each ROI using two different approaches and obtaining value of attributes from each images. In our study two different approaches have followed to determine best features and among them Statistical parametric approach found as best approach to extract features. Two different feature extraction approaches are

1. Gray Level Co-occurrence Matrix based Texture analysis.
2. Statistical Parametric approach

## **4.1 Gray Level Co-occurrence Matrix based Texture analysis**

An image texture is a set of metrics which is calculated to quantify the perceived texture of an image. Information can be obtained about the spatial arrangement of color or intensities in an image from image texture. GLCM calculates co-occurrence of gray level pixel values. Before calculating GLCM it must be ensured that the image is in Gray level. More straight forward GLCM calculates how often a pixel with gray-level value i occurs either vertically, horizontally or diagonally to adjacent pixels with the value j. In this study, using Graycomatrix several features were obtained such as Contrast, Correlation, Energy, Homogeneity and Entropy.

GLCM is calculated with several statistical measures which are explained briefly below.

**Contrast**:

Contrast measures the intensity contrast between a pixel and its neighbor over the entire image. Range of the contrast is [0 (size (GLCM, 1)-1) ^ 2]. For a constant image Contrast value is always 0. Variance and inertia known as property of contrast.



**Correlation**: Correlation calculates how correlated a pixel with its neighbor over the entire image. Range of the correlation is [-1 1]. For a perfect positively or negatively correlated image Correlation is -1 or 1.



**Energy**: In the GLCM Energy calculates value of squared elements. Range of the energy is [0 1]. For a constant image Energy is always 1. Uniformity, uniformity of energy and angular second moment is also known as the property of Energy.



**Homogeneity**: Homogeneity calculates closeness of the distribution of the elements between GLCM diagonal. Range of the homogeneity is [0 1]. For a diagonal GLCM Homogeneity is always 1.



**Entropy**: Entropy is a statistical measure of randomness that is used to characterize the texture of the input image. Entropy returns a scalar value which represents the entropy of the grayscale image



In this study, some other features also calculated to improve performance.

**Convexity**: Convexity is used to quantify tumor shape. As a ratio of tumor size to convex hull Convexity is calculated. Change in tumor morphology tracked by convexity.

**Mean Intensity Value**: Using mean function mean intensity value is calculated. The Pixel value of a particular pixel is mean of it.

**Area**: Area returns actual number of overall nodule pixel in the extracted ROI and it is scalar value. An array of ROI that contains pixels with 255 values is created by the transformation function.

Finally, eight features calculated in GLCM approach as these specific eight features give best result. Five features calculated from matrix and Convexity, Mean intensity value and Area are calculated to improve performance.

**Classifiers**

## **4.2 Support vector machine**

Support Vector Machine (SVM) is one of the well-known supervised machine learning algorithm that can be used for both classification and regression purposes but more commonly used for classification problems. In this algorithm, each data item is plotted as a point in n-dimensional space (n is number of features). SVM works by finding right hyperplane.

**4.2.1 SVM Classifier:**

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.

The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

SVM chooses the extreme points/vectors that help in creating the hyperplane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different categories that are classified using a decision boundary or hyperplane:



## 4.2.2 Types of SVM

**SVM can be of two types:**

* **Linear SVM:** Linear SVM is used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.
* **Non-linear SVM:** Non-Linear SVM is used for non-linearly separated data, which means if a dataset cannot be classified by using a straight line, then such data is termed as non-linear data and classifier used is called as Non-linear SVM classifier.

## Hyperplane and Support Vectors in the SVM algorithm:

**Hyperplane:** There can be multiple lines/decision boundaries to segregate the classes in n-dimensional space, but we need to find out the best decision boundary that helps to classify the data points. This best boundary is known as the hyperplane of SVM.

The dimensions of the hyperplane depend on the features present in the dataset, which means if there are 2 features (as shown in image), then hyperplane will be a straight line. And if there are 3 features, then hyperplane will be a 2-dimension plane.

We always create a hyperplane that has a maximum margin, which means the maximum distance between the data points.

**Support Vectors:**

The data points or vectors that are the closest to the hyperplane and which affect the position of the hyperplane are termed as Support Vector. Since these vectors support the hyperplane, hence called a Support vector.

**Linear SVM:**

The working of the SVM algorithm can be understood by using an example. Suppose we have a dataset that has two tags (green and blue), and the dataset has two features x1 and x2. We want a classifier that can classify the pair(x1, x2) of coordinates in either green or blue. Consider the below image:



So as it is 2-d space so by just using a straight line, we can easily separate these two classes. But there can be multiple lines that can separate these classes. Consider the below image:



Hence, the SVM algorithm helps to find the best line or decision boundary; this best boundary or region is called as a **hyperplane**. SVM algorithm finds the closest point of the lines from both the classes. These points are called support vectors. The distance between the vectors and the hyperplane is called as **margin**. And the goal of SVM is to maximize this margin. The **hyperplane** with maximum margin is called the **optimal hyperplane**.



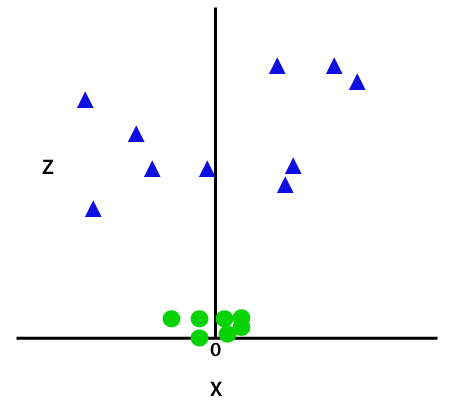
**Non-Linear SVM:**

If data is linearly arranged, then we can separate it by using a straight line, but for non-linear data, we cannot draw a single straight line. Consider the below image:

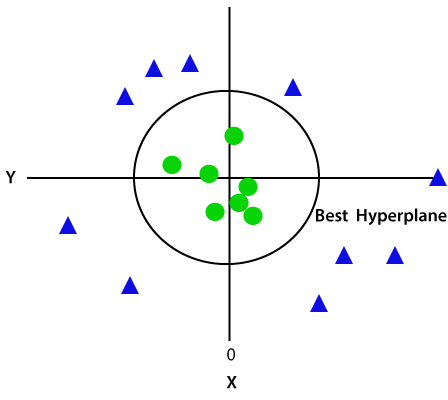


So to separate these data points, we need to add one more dimension. For linear data, we have used two dimensions x and y, so for non-linear data, we will add a third dimension z. It can be calculated as: z=x2 +y2

By adding the third dimension, the sample space will become as below image:



So now, SVM will divide the datasets into classes in the following way. Consider the below image:



**CHAPTER 5**

**ADVANTAGES AND APPLICATIONS**

**Advantages:**

* Easy identification of limited and early stage of lung cancer.
* If found in limited stage, diagnosis becomes easy.
* Classifying lung cancer subtypes

**Applications:**

* Medical image analysis
* Lung Cancer diagnosis
* In medical fields

**CHAPTER-6**

# **Introduction To Matlab**

The name MATLAB stands for Matrix Laboratory. The software is built up around vectors and matrices. This makes the software particularly useful for linear algebra but MATLAB is also a great tool for solving algebraic and differential equations and for numerical integration. MATLAB has powerful graphic tools and can produce nice pictures in both 2D and 3D. It is also a programming language, and is one of the easiest programming languages for writing mathematical programs. These factors make MATLAB an excellent tool for teaching and research.

MATLAB was written originally to provide easy access to matrix software developed by the LINPACK (linear system package) and EISPACK (Eigen system package) projects. It integrates computation, visualization, and programming environment. Furthermore, MATLAB is a modern programming language environment: it has sophisticated data structures, contains built-in editing and debugging tools, and supports object-oriented programming. MATLAB has many advantages compared to conventional computer languages (e.g., C, FORTRAN) for solving technical problems.

MATLAB abilities a family of add-on software program utility software application software program software utility software-unique solutions called toolboxes. Very essential to maximum customers of MATLAB, toolboxes assist you to studies and observe specialized technology. Toolboxes are entire collections of MATLAB abilities (M-files) that increase the MATLAB surroundings to remedy precise schooling of problems. Areas in which toolboxes are to be had embody signal processing, manipulate systems, neural networks, fuzzy correct judgment, wavelets, simulation, and hundreds of others.

It has powerful built-in routines that enable a very wide variety of computations. It also has easy to use graphics commands that make the visualization of results immediately available. Specific applications are collected in packages referred to as toolbox. There are toolboxes for signal processing, symbolic computation, control theory, simulation, optimization, and several other fields of applied science and engineering. MATLAB is an interactive system whose basic data element is an array that does not require dimensioning. The software package has been commercially available since 1984 and is now considered as a standard tool at most universities and industries worldwide.

**6.1 Brief History of MATLAB:**

Cleve Moler, the chairman of the computer science department at the University of New Mexico, started developing MATLAB in the late 1970s. The first MATLAB® was not a programming language; it was a simple interactive matrix calculator. There were no programs, no toolboxes, no graphics and no ODEs or FFTs. He designed it to give his student’s access to LINPACK and EISPACK without them having to learn FORTRAN. It soon spread to other universities and found a strong audience within the applied mathematics community. The mathematical basis for the first version of MATLAB was a series of research papers by J. H. Wilkinson and 18 of his colleagues, published between 1965 and 1970 and later collected in Handbook for Automatic Computation, Volume II, Linear Algebra*,* edited by Wilkinson and C. Reinsch. These papers present algorithms, implemented in Algol 60, for solving matrix linear equation and Eigen value problems.

In the 1970s and early 1980s, I was teaching Linear Algebra and Numerical Analysis at the University of New Mexico and wanted my students to have easy access to LINPACK and EISPACK without writing FORTRAN programs. By “easy access,” I meant not going through the remote batch processing and the repeated edit-compile-link-load-execute process that was ordinarily required on the campus central mainframe computer. Jack little, an engineer, was exposed to it during a visit Moler made to Stanford University in 1983. Recognizing its commercial potential, he joined with Moler and Steve Bangert. They rewrote MATLAB in C and founded Math Works in 1984 to continue its development. These rewritten libraries were known as JACKPAC. In 2000, MATLAB was rewritten to use a newer set of libraries for matrix manipulation, LAPACK. MATLAB was first adopted by researchers and practitioners in control engineering, Little's specialty, but quickly spread to many other domains. It is now also used in education, in particular the teaching of linear algebra and numerical analysis, and is popular amongst scientists involved in video processing**.**

## **6.2 EISPACK and LINPACK**:

In 1970, a group of researchers at Argonne National Laboratory proposed to the U.S. National Science Foundation (NSF) to “explore the methodology, costs, and resources required to produce, test, and disseminate high-quality mathematical software and to test, certify, disseminate, and support packages of mathematical software in certain problem areas.” The group developed EISPACK (Matrix Eigen system Package) by translating the Algol procedures for Eigen value problems in the handbook into FORTRAN and working extensively on testing and portability. The first version of EISPACK was released in 1971 and the second in 1976.

In 1975, four of us Jack Dongarra, Pete Stewart, Jim Bunch, and myself proposed to the NSF another research project that would investigate methods for the development of mathematical software. A byproduct would be the software itself, dubbed LINPACK, for Linear Equation Package. This project was also centered at Argonne. LINPACK originated in FORTRAN; it did not involve translation from Algol. The package contained 44 subroutines in each of four numeric precisions. In a sense, the LINPACK and EISPACK projects were failures. We had proposed research projects to the NSF to “explore the methodology, costs, and resources required to produce, test, and disseminate high-quality mathematical software.” We never wrote a report or paper addressing those objectives. We only produced software.

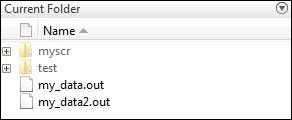
So, I studied Niklaus Wirth’s book Algorithms + Data Structures *=* Programs and learned how to parse programming languages. I wrote the first MATLAB an acronym for Matrix Laboratory in FORTRAN, with matrix as the only data type. The project was a kind of hobby, a new aspect of programming for me to learn and something for my students to use. There was never any formal outside support, and certainly no business plan. This first MATLAB was just an interactive matrix calculator. This snapshot of the start-up screen shows all the reserved words and functions. There are only 71. To add another function, you had to get the source code from me, write a FORTRAN subroutine, add your function name to the parse table, and recompile MATLAB.

**6.3 Starting MATLAB:**

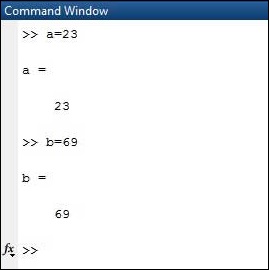
After logging into your account, you can enter MATLAB by double-clicking on the MATLAB shortcut icon (MATLAB 7.0.4) on your Windows desktop. When you start MATLAB, a special window called the MATLAB desktop appears. The desktop is a window that contains other windows. The major tools within or accessible from the desktop are:

* The Command Window
* The Command History
* The Workspace
* The Current Directory
* The Help Browser

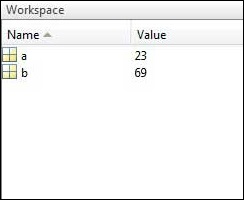
**Current Folder:** This panel allows you to access the project folders and files.



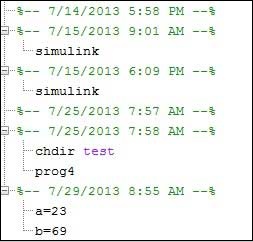
**Command Window:** This is the main area where commands can be entered at the command line. It is indicated by the command prompt (>>).



**Workspace:**  The workspace shows all the variables created and/or imported from files.



**Command History:** This panel shows or return commands that are entered at the command line.



**Help Browser:**

The critical way to get assist online is to use the MATLAB help browser, opened as a separate window every through clicking at the question mark photograph (?) on the computing tool toolbar, or through manner of typing assist browser on the spark off in the command window. The assist Browser is an internet browser blanketed into the MATLAB computing tool that shows a Hypertext Markup Language (HTML) files. The Help Browser consists of panes, the help navigator pane, used to find out information, and the show pane, used to view the information. Self-explanatory tabs apart from navigator pane are used to performs are searching out.

**MATLAB language:**

This is a high-level matrix/array language with control flow statements, functions, data structures, input/output, and object-oriented programming features. It allows both "programming in the small" to rapidly create quick and dirty throw-away programs, and "programming in the large" to create complete large and complex application programs.

**MATLAB working environment:**

This is the set of tools and facilities that you work with as the MATLAB user or programmer. It includes facilities for managing the variables in your workspace and importing and exporting data. It also includes tools for developing, managing, debugging, and profiling M-files, MATLAB's applications.

**MATLAB mathematical function library:**

This is a vast collection of computational algorithms ranging from elementary functions like sum, sine, cosine, and complex arithmetic, to more sophisticated functions like matrix inverse, matrix eigenvalues, Bessel functions, and fast Fourier transforms.

**MATLAB Application Program Interface (API):**

This is a library that allows you to write C and FORTRAN programs that interact with MATLAB. It includes facilities for calling routines from MATLAB (dynamic linking), calling MATLAB as a computational engine, and for reading and writing MAT-files.

**MATLAB DESKTOP:**

MATLAB Desktop is the precept MATLAB utility window. The computing tool includes five sub home windows, the command window, the workspace browser, the modern-day-day list window, the command records window, and one or greater decide domestic windows, which is probably confirmed high-quality on the identical time due to the truth the client suggests a photo. The command window is in which the character types MATLAB instructions and expressions at the spark off (>>) and in which the output of these commands is displayed. MATLAB defines the workspace because the set of variables that the client creates in a bit consultation. The workspace browser suggests those variables and some facts about them. Double clicking on a variable within the workspace browser launches the Array Editor, which may be used to gain statistics and profits instances edit exceptional homes of the variable.

The modern-day-day-day Directory tab above the workspace tab suggests the contents of the cutting-edge list, whose path is shown inside the modern-day list window. For example, in the home windows on foot machine the path is probably as follows: C: MATLAB Work, indicating that listing “artwork” is a subdirectory of the number one list “MATLAB”; WHICH IS INSTALLED IN DRIVE C. Clicking on the arrow within the modern list window suggests a listing of these days used paths. Clicking at the button to the right of the window permits the individual to trade the present day listing. MATLAB uses a seeking out path to find out M-documents and one-of-a-type MATLAB associated documents, which can be put together in directories within the computer document tool. Any report run in MATLAB need to be dwelling in the modern-day-day listing or in a list that is on is looking for course. By default, the documents supplied with MATLAB and math works toolboxes are included inside the searching out direction. The first-rate manner to look which directories are on the searching out route. The satisfactory manner to appearance which directories are speedy the quest route, or to characteristic or regulate a searching for course, is to pick out outset path from the File menu the computing device, and then use the set course talk discipline. It is proper exercise to feature any generally used directories to the hunt route to avoid again and again having the exchange the cutting-edge-day listing.

The Command History Window contains a file of the instructions a person has entered in the command window, together with every contemporary-day and former MATLAB periods. Previously entered MATLAB instructions can be determined on and re-completed from the command statistics window thru proper clicking on a command or series of commands. This movement launches a menu from which to select numerous options similarly to executing the commands. This is useful to select out abilities options in addition to executing the instructions. This is a beneficial feature at the equal time as experimenting with numerous commands in a piece session.

**Using the MATLAB Editor to create M-Files:**

The MATLAB editorial manager is a literary substance proofreader particular for growing M-facts and a graphical MATLAB debugger. The supervisor can seem in a window through command facts technique for itself, or it is probably a right-clicking inside the PC. M-information this gadget signified through the use of the expansion .M, as in pixel up.M. The MATLAB editorial supervisor window has a few draws down menus for obligations collectively with sparing, seeing, and troubleshooting facts. Since it plays more than one easy test and furthermore affects utilization of shade to separate among exclusive variables of code, this article editorial supervisor is often supported due to reality the system of a need for composing and altering M-talents. To open the manager, type at enact opens the M-document filename. M in a supervisor window, sorted out for enhancing. As stated earlier than, the file should be inside the cutting-edge posting, or in a posting in the seeking out direction.

## **6.4 Features of MATLAB:**

Following are the basic features of MATLAB.

* It is a high-level language for numerical computation, visualization and application development.
* It also provides an interactive environment for iterative exploration, design and problem solving.
* It provides vast library of mathematical functions for linear algebra, statistics, Fourier analysis, filtering, optimization, numerical integration and solving ordinary differential equations.
* It provides built-in graphics for visualizing data and tools for creating custom plots.
* MATLAB's programming interface gives development tools for improving code quality maintainability and maximizing performance.
* It provides tools for building applications with custom graphical interfaces.
* It provides functions for integrating MATLAB based algorithms with external applications and languages such as C, Java, .NET and Microsoft Excel.

## **Uses of MATLAB:**

MATLAB is widely used as a computational tool in science and engineering encompassing the fields of physics, chemistry, math and all engineering streams. It is used in a range of applications including

* Signal Processing and Communications
* Video and Video Processing
* Control Systems
* Test and Measurement
* Computational Finance
* Computational Biology

**6.5 Applications of MATLAB:**

MATLAB can be used as a tool for simulating various electrical networks but the recent developments in MATLAB make it a very competitive tool for Artificial Intelligence, Robotics, Video processing, Wireless communication, Machine learning, Data analytics and whatnot. Though it’s mostly used by circuit branches and mechanical in the engineering domain to solve a basic set of problems its application is vast. It is a tool that enables computation, programming and graphically visualizing the results. The basic data element of MATLAB as the name suggests is the Matrix or an array. MATLAB toolboxes are professionally built and enable you to turn your imaginations into reality. MATLAB programming is quite similar to C programming and just requires a little brush up of your basic programming skills to start working with.

Below are a few applications of MATLAB –

* **Statistics and machine learning (ML)**

This toolbox in MATLAB can be very handy for the programmers. Statistical methods such as descriptive or inferential can be easily implemented. So is the case with machine learning. Various models can be employed to solve modern-day problems. The algorithms used can also be used for big data applications.

* **Curve fitting**

The curve fitting toolbox helps to analyze the pattern of occurrence of data. After a particular trend which can be a curve or surface is obtained, its future trends can be predicted. Further plotting, calculating integrals, derivatives, interpolation, etc. can be done.

* **Control systems**

Systems nature can be obtained. Factors such as closed-loop, open-loop, its controllability and observability, Bode plot, NY Quist plot, etc. can be obtained. Various controlling techniques such as PD, PI and PID can be visualized. Analysis can be done in the time domain or frequency domain.

* **Signal Processing**

Signals and systems and digital signal processing are taught in various engineering streams. But MATLAB provides the opportunity for proper visualization of this. Various transforms such as Laplace, Z, etc. can be done on any given signal. Theorems can be validated. Analysis can be done in the time domain or frequency domain. There are multiple built-in functions that can be used.

* **Mapping**  
  Mapping has multiple applications in various domains. For example, in Big Data, the Map Reduce tool is quite important which has multiple applications in the real world. Theft analysis or financial fraud detection, regression models, contingency analysis, predicting techniques in social media, data monitoring, etc. can be done by data mapping.
* **Deep learning**

It’s a subclass of machine learning which can be used for speech recognition, financial fraud detection, and medical video analysis. Tools such as time-series, Artificial neural network (ANN), Fuzzy logic or combination of such tools can be employed.

* **Financial analysis**

An entrepreneur before starting any endeavor needs to do a proper survey and the financial analysis in order to plan the course of action. The tools needed for this are all available in MATLAB. Elements such as profitability, solvency, liquidity, and stability can be identified. Business valuation, capital budgeting, cost of capital, etc. can be evaluated.

* **Video processing**

The most common application that we observe almost every day are bar code scanners, selfie (face beauty, blurring the background, face detection), video enhancement, etc. The digital video processing also plays quite an important role in transmitting data from far off satellites and receiving and decoding it in the same way. Algorithms to support all such applications are available.

* **Text analysis**

Based on the text, sentiment analysis can be done. Google gives millions of search results for any text entered within a few milliseconds. All this is possible because of text analysis. Handwriting comparison in forensics can be done. No limit to the application and just one software which can do this all.

* **Electric vehicles designing**

Used for modeling electric vehicles and analyze their performance with a change in system inputs. Speed torque comparison, designing and simulating of a vehicle, whatnot.

* **Aerospace**

This toolbox in MATLAB is used for analyzing the navigation and to visualize flight simulator.

* **Audio toolbox**

Provides tools for audio processing, speech analysis, and acoustic measurement. It also provides algorithms for audio and speech feature extraction and audio signal transformation.

**CHAPTER 7**

# **RESULTS**

Fig3: Gray Image

Fig2: Input Image



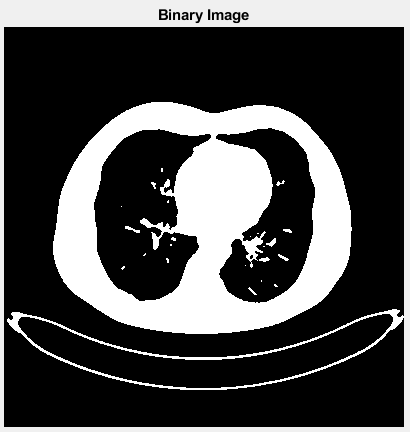


Fig5: Binary Image

Fig4: Denoise Image

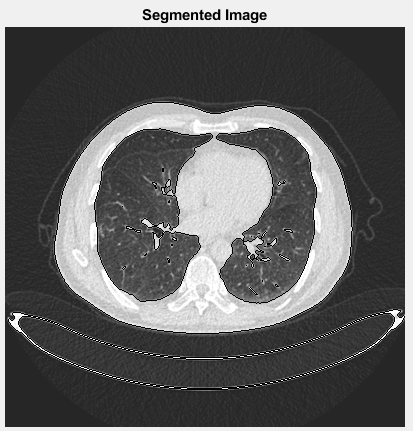


Fig7: Region segmented Image

Fig6: Segmented Image

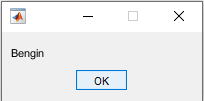
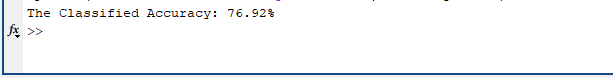
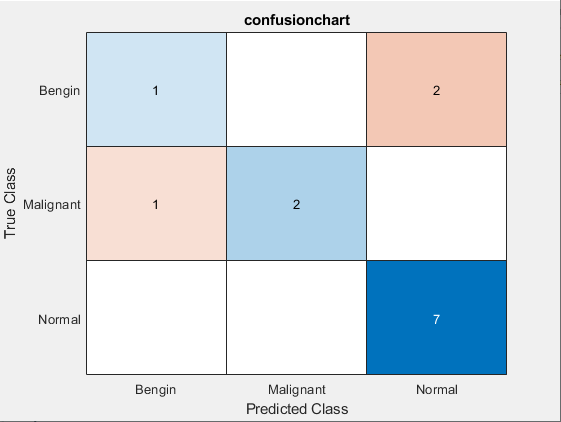


Fig10: Confusion Chart

Fig9: Classified Accuracy

Fig8: Classified Output

**CHAPTER 8**

# **CONCLUSION**

The utilization of Gray Level Co-Occurrence Matrix (GLCM) in conjunction with Support Vector Machine (SVM) methods for lung cancer detection based on CT-scan images represents a significant advancement in the field of medical imaging and early disease diagnosis. This innovative approach has demonstrated remarkable accuracy and reliability in identifying potential malignancies within lung tissues.

The GLCM's ability to extract texture and spatial information from CT-scan images has allowed for the creation of robust detection features, enabling the SVM classifier to effectively distinguish between cancerous and non-cancerous regions with a high degree of precision. As a result, this technology holds immense promise for improving the early detection of lung cancer, which is crucial for enhancing patient outcomes and reducing mortality rates.

# **REFERENCES**

[1] Y. F. Riti, H. A. Nugroho, S. Wibirama, B. Windarta, and L. Choridah, "Feature extraction for lesion margin characteristic classification from CT Scan lungs image," 2016 1st International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), Yogyakarta, 2016, pp. 54-58.

[2] R. Wulandari, R. Sigit, and S. Wardhana, "Automatic lung cancer detection using color histogram calculation," 2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC), Surabaya, 2017, pp. 120-126.

[3] L. Anifah, Haryanto, R. Harimurti, Z. Permatasari, P. W. Rusimamto, and A. R. Muhamad, "Cancer lung detection on CT scan image using artificial neural network backpropagation based gray level co-occurrence matrices feature," 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Bali, 2017, pp. 327-332.

[4] D. P. Kaucha, P. W. C. Prasad, A. Alsadoon, A. Elchouemi, and S. Sreedharan, "Early detection of lung cancer using SVM classifier in biomedical image processing," 2017 IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI), Chennai, 2017, pp. 3143-3148.

[5] F. Taher, N. Werghi, and H. Al-Ahmad, "Computer-aided diagnosis system for early lung cancer detection," 2015 International Conference on Systems, Signals and Image Processing (IWSSIP), London, 2015, pp. 5-8.

[6] E. Rendon-Gonzalez and V. Ponomaryov, "Automatic Lung nodule segmentation and classification in CT images based on SVM," 2016 9th International Kharkiv Symposium on Physics and Engineering of Microwaves, Millimeter and Submillimeter Waves (MSMW), Kharkiv, 2016, pp. 1-4.

[7] A. Kulkarni and A. Panditrao, "Classification of lung cancer stages on CT scan images using image processing," 2014 IEEE International Conference on Advanced Communications, Control and Computing Technologies, Ramanathapuram, 2014, pp. 1384-1388.

[8] S. A. El-Regaily, M. A. M. Salem, M. H. A. Aziz and M. I. Roushdy, "Lung nodule segmentation and detection in computed tomography," 2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS), Cairo, 2017, pp. 72-78.

[9] M. H. Jony, F. Tuj Johora, P. Khatun and H. K. Rana, "Detection of Lung Cancer from CT Scan Images using GLCM and SVM," 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT), Dhaka, Bangladesh, 2019, pp. 1-6.