Faculty of Electronics & Computer Technology and Engineering

DESIGN AND ANALYSIS OF AUTOMATED LUNG CANCER DETECTION USING CONVOLUTION NEURAL NETWORK IN MATLAB

NUR FARAH NADIA BINTI RAMLI

Bachelor of Electronics Engineering Technology with Honours

2025

DESIGN AND ANALYSIS OF AUTOMATED LUNG CANCER DETECTION USING CONVOLUTION NEURAL NETWORK

NUR FARAH NADIA BINTI RAMLI

**A project report submitted
in partial fulfillment of the requirements for the degree of Bachelor of Electronics Engineering Technology with Honours**

Faculty of Electronics & Computer Technology and Engineering

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2025

UNIVERSITI TEKNIKAL MALAYSIA MELAKA
FAKULTI TEKNOLOGI KEJUTERAAN ELEKTRIK DAN ELEKTRONIK

BORANG PENGESAHAN STATUS LAPORAN
PROJEK SARJANA MUDA II

Tajuk Projek : Design and Analysis of Automated Lung Cancer Detection using Convolution Neural Network.

Sesi Pengajian : 2024/2025

Saya NUR FARAH NADIA BINTI RAMLI mengaku membenarkan laporan Projek Sarjana Muda ini disimpan di Perpustakaan dengan syarat-syarat kegunaan seperti berikut:

1. Laporan adalah hakmilik Universiti Teknikal Malaysia Melaka.
2. Perpustakaan dibenarkan membuat salinan untuk tujuan pengajian sahaja.
3. Perpustakaan dibenarkan membuat salinan laporan ini sebagai bahan pertukaran antara institusi pengajian tinggi.
4. Sila tandakan (✓):

☐	**SULIT***	(Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub di dalam AKTA RAHSIA RASMI 1972)
☐	**TERHAD***	(Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan dijalankan)
✓	**TIDAK TERHAD**	

(TANDATANGAN PENULIS)

Alamat Tetap:

(COP DAN TANDATANGAN PENYELIA

PROF. MADYA TS. DR. NORHASHIMAH MOHD SAAD
Profesor Madya
Jabatan Teknologi Kejuruteraan Elektronik dan Komputer
Fakulti Teknologi dan Kejuruteraan Elektronik dan Komputer
Universiti Teknikal Malaysia Melaka

Tarikh: 25/01/2025

Tarikh: 25/01/2025

*CATATAN: Jika laporan ini SULIT atau TERHAD, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali tempoh laporan ini perlu dikelaskan sebagai SULIT atau TERHAD.

DECLARATION

I declare that this project report entitled **"DESIGN AND ANALYSIS OF AUTOMATED LUNG CANCER DETECTION USING CONVOLUTION NEURAL NETWORK IN MATLAB"** is the result of my own research except as cited in the references. The project report has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature :

Student Name : NUR FARAH NADIA BINTI RAMLI

Date : 25/01/2025

APPROVAL

I hereby declare that I have checked this project report and, in my opinion, this project report is adequate in terms of scope and quality for the award of the degree of Bachelor of Electronics Engineering Technology with Honours.

Signature	:	....
Supervisor Name	:	PROFESOR MADYA TS. DR. NORHASHIMAH BINTI MOHD SAAD
Date	:	25/01/2025

DEDICATION

I dedicate this bachelor's degree project to my creator, Allah s.w.t the Almighty, my strong pillar, my source of inspiration, wisdom, knowledge, and understanding. He has been the source of my strength to accomplish this project throughout this degree.

To my beloved parents,

Ramli bin Sluiman and Siti Rokayah binti Asit who consistently provide me their moral, spiritual, emotional, and financial support, who have inspired me and given me courage when I felt like giving up.

To my great supervisor,

Assoc. Prof. Ts. Dr. Norhashimah binti Mohd Saad for the direction, advice, and pearls of wisdom but more importantly for putting up with my endless whatsapp and queries while giving me feedback in such a timely manner and encouraging me just when I needed it, without which it would have been very difficult to complete this piece of work.

ABSTRACT

In the modern world, one of the most prevalent and hazardous cancers is the lung cancer disease that causes the most fatalities each year. Accurate lung cancer identification could increase endurance rates. In this research work, a computer aided system for detecting lung cancer using convolution neural network (CNN) is proposed. The proposed model includes preprocessing, image segmentation model training, and tumor classification. The model is based on the Dataset in Kaggle, which contains 287 lung images in which 161 cancer lung images and 126 non-cancer images. The proposed model classify the lung CT images as cancerous or normal image accurately with 97.67% accuracy, 96% of sensitivity and 94%. of specificity.

ABSTRAK

Di dunia moden, salah satu kanser yang paling lazim dan berbahaya adalah penyakit kanser paru-paru yang menyebabkan kematian paling banyak setiap tahun. Pengenalpastian kanser paru-paru yang tepat boleh meningkatkan kadar ketahanan. Dalam kerja penyelidikan ini, sistem bantuan komputer untuk mengesan kanser paru-paru menggunakan rangkaian neural konvolusi (CNN) dicadangkan. Model yang dicadangkan termasuk prapemprosesan, latihan model segmentasi imej, dan klasifikasi tumor. Model ini berdasarkan Dataset dalam Kaggle, yang mengandungi 287 imej paru-paru di mana 161 imej paru-paru kanser dan 126 imej bukan kanser. Model yang dicadangkan mengklasifikasikan imej CT paru-paru sebagai imej kanser atau normal dengan tepat dengan ketepatan 97.67%, 96% sensitiviti dan 94%. kekhususan.

ACKNOWLEDGEMENTS

First and foremost, I am thankful to my creator, Allah s.w.t for His kindness by giving me a good health, patience, determination and blessed me far exceeding my needs. Without Him, I may not be able to complete this thesis within the given time despite many obstacles I went through.

Second, I would like to express my gratitude to my excellent supervisor, Assoc. Prof. Ts. Dr. Norhashimah binti Mohd Saad for his precious guidance, patient, support, and words of wisdom that encourage me to not give up on this project. I am so lucky to have him as my supervisor. Without his guidance, I might not be able to continue this thesis.

My highest appreciation goes to my parents, Ramli bin Sluiman and Siti Rokayah binti Asit and to my family members for their love and prayer during the period of my study. Their support played a very important role for me throughout my journey.

I also indebted to my fellow friends for the willingness of sharing their thoughts and ideas regarding this project. An honourable mention also goes to friends, Amira Shazlyana, and Anis Syahirah for being a pillar of support for me to stay motivated while providing encouragement to each other at times when it seemed impossible to continue.

This gratitude is also extended to other individuals who are not listed here for assisting me greatly in the completion of this thesis. Only God has the power to reward you for your good deeds.

TABLE OF CONTENTS

LIST OF TABLES

TABLE **TITLE** **PAGE**

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

This chapter will set stage for the remainder of the project, which will focus on the Design and Analysis of Automated Lung Cancer Detection Using Convolution Neural Network (CNN) in MATLAB. This project aims to create and advanced diagnostic tool that can accurately identify and classify lung nodules in CT image. By leveraging the power of CNNs, the system will automatically analyze medical images to distinguish between cancer and non-cancer to see nodules at lung, significantly aiding in early detection and diagnosis of lung cancer. This chapter will discuss the project's context, problem description, aim and scope, providing a comprehensive overview of the motivations and objectives behind developing this automated detection system.

1.1 Background

Lung Lung cancer is the most diagnosed cancer worldwide, accounting for approximately 12.4% of all new cancer cases, with an estimated 2.5 million new cases in 2022[1]. It also remains the leading cause of cancer-related deaths globally, responsible for about 1.8 million deaths, which is approximately 19% of all cancer deaths [2]. Early detection is crucial for improving survival rates; however, lung cancer is often diagnosed at advanced stages when treatment options are limited [3].

Advancements in medical imaging and artificial intelligence have facilitated the development of automated systems for disease detection. Convolutional Neural Networks (CNNs), a class of deep learning models, have demonstrated exceptional performance in medical image analysis tasks, including disease classification, localization, and detection [4]. CNNs are particularly effective due to their ability to automatically learn and extract hierarchical features from images, making them suitable for identifying complex patterns associated with diseases like lung cancer [4].

MATLAB, a high-level programming environment, offers comprehensive tools for implementing CNNs and processing medical images. Its robust libraries and user-friendly interface make it a preferred choice for developing and analyzing deep learning models in medical research.

This project aims to design and analyze an automated system for lung cancer detection using Convolutional Neural Networks implemented in MATLAB. The system will focus on processing and classifying medical images to distinguish between cancerous and non-cancerous tissues, thereby aiding in early diagnosis and improving patient outcomes.

1.2 Societal/Global Issue for Design and Analysis of Automated Lung Cancer Detection using Convolution Neural Network

Lung cancer is a major global health problem, causing nearly 1.8 million deaths each year [3]. It is mainly caused by smoking, though non-smokers can also develop the disease [3]. Early detection is key to improving survival, but current methods like CT scans and chest X-rays can be slow and prone to mistakes, leading to late diagnoses. The shortage of trained doctors, especially in low-income areas, makes this issue worse. Using Convolutional Neural Networks (CNNs) for automated detection could improve accuracy, speed, and access to lung cancer diagnosis [5]. By using AI, CNNs can help reduce errors, support healthcare systems, and make early detection more accessible, ultimately saving lives and easing the global burden of lung cancer.

1.3 Problem Statement

Lung cancer is one of the deadliest types of cancer, and most people are diagnosed at an advanced stage when treatment is less effective. Early detection is key to improving survival rates, but current methods like chest X-rays and CT scans rely on doctors to manually examine the images. These methods can be slow and prone to mistakes, as early signs of cancer can sometimes be missed, leading to delayed diagnosis and worse outcomes for patients.

Another problem is the shortage of trained radiologists, especially in poorer or remote areas where there are fewer doctors. This creates an unequal healthcare system, where many patients might not receive the timely diagnosis, they need. Traditional methods also depend on the experience of individual doctors, which can lead to differences in how accurately the disease is diagnosed, increasing the risk of misinterpretation.

To solve these problems, using Artificial Intelligence (AI) and deep learning techniques like Convolutional Neural Networks (CNNs) to automatically detect lung cancer in CT scans could be a powerful solution. AI can help make diagnoses faster, more accurate, and more available to people in areas with fewer doctors. This project aims to create an automated lung cancer detection system using CNNs in MATLAB, helping to improve early detection and reduce the global burden of lung cancer.

1.4 Project Objective

The primary objective of this project is to design and develop an automated system for detecting lung cancer using Convolutional Neural Networks (CNNs) in MATLAB. The specific objectives are:

a) To develop an automated lung cancer detection system using MATLAB, including image preprocessing, segmentation, and deep learning for analyzing CT scan images.

b) To enhance the performance of the system by testing it on a CT scan dataset, calculating accuracy, and generating confusion matrix to improve the assess classification results.

c) To design and implement a user-friendly graphical user interface (GUI) for the lung cancer detection system, allowing users to input CT scan images and visualize the detection result.

1.5 Scope of Project

The scope of this project are as follows:

a) This project will involve improving CT scan images through techniques like noise reduction and enhancement to ensure they are ready for analysis.

b) The system will identify regions and detect any potential cancerous or non-cancerous nodules using CNN algorithm.

c) A CNN model will be trained to classify CT scan images as either cancerous or non-cancerous, with performance evaluated using accuracy and other measures.

1.6 Outline of the project

Chapter 1: Introduction

This chapter introduces the project, detailing the goals importance of detecting lung cancer using Convolutional Neural Networks (CNNs) in MATLAB. It presents the problem statement, project objectives, and the scope of the work, emphasizing the importance of early detection in improving lung cancer survival rates. The chapter also discusses the potential impact of the project in enhancing the accuracy, speed, and accessibility of lung cancer diagnosis, offering a valuable tool for medical professionals and supporting early intervention efforts.

Chapter 2: Literature Review

This chapter reviews current research and technologies related to lung cancer detection using Convolutional Neural Networks (CNNs). It discusses various methods, performance metrics, and datasets that have been used in previous studies for automated lung cancer diagnosis. The literature review highlights the strengths and limitations of existing approaches, including the accuracy of different CNN models in detecting cancerous nodules in CT scan images. By examining these studies, the chapter aims to identify gaps in the current state of the art and potential areas for improvement in lung cancer detection.

Chapter 3: Methodology

This chapter describes the approach and techniques used in the design and analysis of an automated lung cancer detection system using Convolutional Neural Networks (CNNs) in MATLAB. It includes detailed descriptions of the data collection process, image preprocessing steps, segmentation techniques, feature extraction methods, and the CNN model selection and training procedures. The methodology provides a comprehensive guide to the development and implementation of the lung cancer detection system, ensuring accurate and efficient classification of cancerous and non-cancerous nodules in CT scan images.

Chapter 4: Result and Discussion

This chapter presents the results of the automated lung cancer detection system using CNNs in MATLAB, including performance analysis and visualizations of detected nodules and classification results.

Chapter 5: Conclusion and Recommendations

This chapter summarizes the main results of the project and evaluates the overall success in achieving the objectives of developing an automated lung cancer detection system using Convolutional Neural Networks (CNNs) in MATLAB. It provides recommendations for future work and potential improvements to the system. Additionally, the chapter explores the practical applications of the developed system and

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This section discusses and summarize the overall concept and theory behind this project, which is aiming to detect lung cancer. It provides an analysis of previous studies, methodologies, and key findings to establish a foundation for the current research. In this chapter, it discussed about the concept and theory that were used to solve the problem with this project. Journals, articles, and case studies make up the bulk of the information that was gathered, and these formats were chosen for their similarity to the project's scope to ensure accuracy.

2.2 Understanding [Global/Current Issue] in the Literature

Detecting lung cancer late is a big problem. Often, it's not found until it's too late for effective treatment, leading to poor chances of survival. This shows how important it is to find better ways to detect lung cancer earlier. Looking at lung scans is hard because they show many details. It's tough to spot cancer among all those details. This makes it hard for doctors to know for sure if what they're seeing is cancer or not. So, we need smarter ways to look at these scans and find cancer more easily.

Sometimes, doctors might make mistakes when looking at lung scans. They might see something that isn't really cancer, or they might miss something that is. This happens because it's hard to be sure just by looking at a picture. That's why we need tools that can help doctors make more accurate decisions when it comes to finding cancer in lung scans. More people want to find faster ways to spot lung cancer. They want to use machines that can do the job quickly and accurately. These machines could help doctors make better decisions and find cancer sooner, which is important for saving lives.

New technology called Convolutional Neural Networks (CNNs) is making waves in cancer detection. These networks are super smart and can learn to spot cancer in pictures. By using these networks, we might be able to find cancer in lung scans faster and more accurately than ever before. So, understanding the problems we face in finding lung cancer early shows us how important it is to come up with new solutions. By finding better ways to look at lung scans, reducing mistakes, and using smart machines like CNNs, we can improve our chances of finding lung cancer sooner and helping more people survive

2.3 What is lung cancer

Lung cancer happens when cells in the lung grow abnormally and form a tumor. This tumor can appear on images like chest x-rays or CT scans as a small spot or a larger growth. While lung cancer is more common in people who smoke, it can also occur in people who have never smoked. Other risk factors include exposure to radon, asbestos, second-hand smoke, polluted air, and having a family history of lung cancer. Getting diagnosed with lung cancer is serious, but treatments have improved a lot. Early detection usually leads to better outcomes, so it's important to know the symptoms. However, sometimes lung cancer doesn’t show any symptoms, making regular screenings a good idea for those at higher risk.[6]

2.4 What are the types of lung cancer

Lung cancer comes in two main types: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC is more common and has different subtypes like squamous cell carcinoma, adenocarcinoma, and large cell carcinoma. These names describe the kinds of cells found in the cancer when looked at under a microscope. Sometimes, doctors also do tests to check for certain markers and DNA changes in NSCLC to learn more about the cancer. SCLC is faster-growing and tends to spread to other parts of the body more quickly [7].

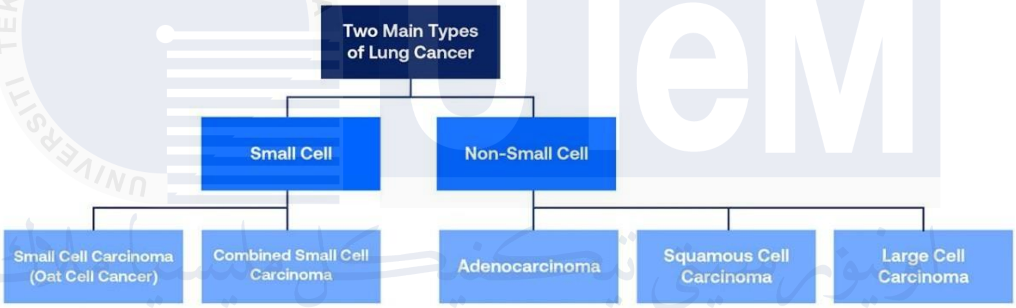


Figure 2.1 Types of Lungs Cancer

2.5 Small Cell Lung Cancer (SCLC)

There are two main types of small cell lung cancer: small cell carcinoma and mixed small cell/large cell cancer, also called combined small cell lung cancer. These types are named based on the cells found in the cancer and how they look under a microscope. Small cell lung cancer is almost always related to smoking cigarettes. Treatment usually involves chemotherapy [7].

2.5.1 Small Cell Carcinoma (Oat Cell Cancer)

Small cell carcinoma, sometimes called oat cell cancer, is a type of lung cancer named after the small, oval-shaped cells that look like oats under a microscope. This cancer grows rapidly and often spreads to other parts of the body early. It is almost always caused by smoking. Treatment usually involves chemotherapy to help stop the cancer from growing and spreading [7].

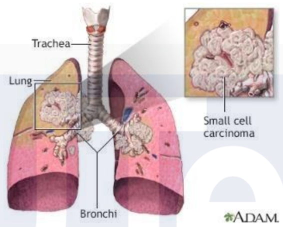


Figure 2.2 Small Cell Carcinoma

2.5.2 Combined Small Cell Carcinoma

Combined small cell carcinoma is a type of lung cancer that includes both small cell carcinoma and other types of lung cancer cells, like large cell carcinoma. This means it has a mix of different cancer cell types. Like other small cell lung cancers, it is usually linked to smoking and tends to grow and spread quickly. Treatment often involves a combination of chemotherapy and other therapies to address the different types of cancer cells present [7].

2.6 Non-Small Cell

Non-small cell lung cancer (NSCLC) is the most common type of lung cancer, making up about 80% of all cases. This cancer usually grows and spreads more slowly than small cell lung cancer. There are three main types of NSCLC:

2.6.1 Adenocarcinoma

One type of non-small cell lung cancer often found in the outer parts of the lung is called adenocarcinoma. This cancer develops in the cells of the epithelial tissues, which are the cells that line the cavities and surfaces of the body and also form glands. These cells help produce mucus and other substances. Adenocarcinoma is common and usually grows more slowly than other types of lung cancer [7].

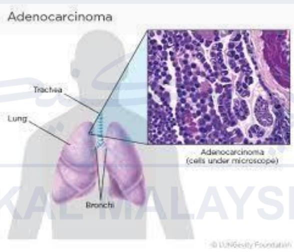


Figure 2.3 Adenocarcinoma

2.6.2 Squamous cell carcinoma

Squamous cell carcinoma is a type of non-small cell lung cancer that typically develops in the center of the lung near one of the main airways, called a bronchus. This cancer starts in the flat, thin cells that line the airways. It is often linked to smoking and can cause symptoms like coughing and breathing problems as it grows near the lung's central passages [7].

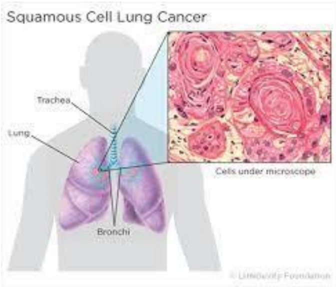


Figure 2.4 Squamous Cell Lung Cancer

2.6.3 Large cell carcinoma

Large cell carcinoma, a form of non-small cell lung cancer, has the potential to arise in any lung region. This type of cancer exhibits rapid growth and spread compared to adenocarcinoma or squamous cell carcinoma. Its name derives from the prominent, round appearance of the cancer cells when viewed under a microscope. Due to its rapid growth, it often requires aggressive treatment [7].

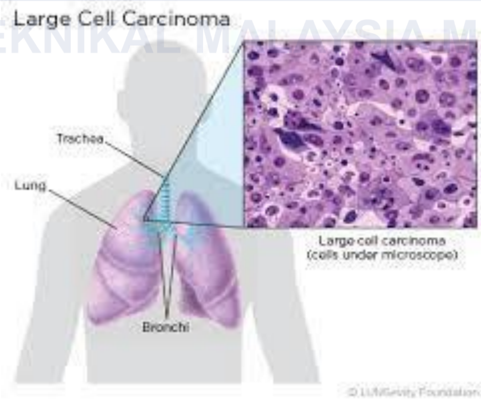


Figure 2.5: Large cell carcinoma

2.7 Symptoms of lung cancer

Symptoms of lung cancer are varied and can include a persistent cough lasting three weeks or more, changes in an existing chronic cough, or recurrent chest infections. Breathlessness without any apparent cause, coughing up blood, and persistent chest or shoulder pain are also warning signs. Hoarseness lasting three weeks or more, along with loss of appetite, unexplained weight loss, and fatigue, are additional symptoms that warrant medical attention. It's essential to consult a GP if experiencing any of these symptoms, although they can also be indicative of other lung conditions or be linked to smoking. Less common symptoms may include finger clubbing or shoulder pain radiating down the arm. Early diagnosis and prompt medical attention are crucial for effective management and treatment of lung cancer [6]

2.8 Literature Review Based on Several Research Paper

In a recent study [8], to categorise lung cancer nodules, researchers created a novel deep learning model that makes use of adaptive swarm intelligence. They employed real-time samples from the PSG Institute of Medical Sciences and Research (PSGIMSR) and data from the Lung Image Database Consortium (LIDC). By improving its search techniques, their program aimed to enhance the process of identifying the best answers. With mean square error rates of 0.0018 for the LIDC data and 0.0027 for the PSGIMSR data, their findings demonstrated improved lung cancer detection.

In another research project cited as [9], the velocity-enhanced whale optimisation strategy is a new approach that scientists have presented. For the identification and diagnosis of several cancer kinds, such as breast, cervical, and lung cancer, this approach uses a hybrid approach that incorporates an artificial neural network. They compared it to four common algorithms: learning vector quantisation, factorised distribution method, C4.5, and linear discriminate analysis (LDA). The suggested method outperformed the benchmark algorithms with remarkable classification accuracies of 97.65% for breast cancer, 94.6% for cervical cancer, and 84% for lung cancer.

In a separate investigation referenced as [10] to predict lung cancer using CT scans from a cancer picture library, researchers used an augmented dense clustering algorithm in conjunction with a strategy employing instantly trained neural networks (DITNN). They preprocessed the data using a number of methods. Their study's results showed that their approach produced a classification error of 0.038 and an accuracy of 98.42%.

In a study referenced as [11], for feature selection and target gene identification, researchers created a model that combines the Bhattacharya coefficient with a genetic algorithm (GA). Using ensemble outputs from several classifiers and deep learning approaches, they used this integrated approach to choose features and assess fitness. Their approach outperformed alternative methods for identifying and forecasting neuromuscular diseases.

In a different study cited as [12], researchers presented R2MNet, a malignancy evaluation network that combines radiological analysis. They used the LIDC-IDRI dataset to analyse radiological characteristics in order to determine the malignancy of lung nodules. Their results showed that the suggested approach obtained an AUC of 97.52% for evaluating nodule malignancy and 96.27% for nodule radiology analysis.

In another study referenced as [13], a deep learning method was presented by researchers to investigate lung cancer and pneumonia. To improve classification accuracy in lung cancer assessment, they suggested two models: a modified AlexNet (MAN) and a combination of manually created and learnt features within the MAN. According to their results, the accuracy of the suggested model was 97.27%.

In another study referenced as [14] to forecast lung cancer, researchers used an automated meta-heuristic optimisation method called the crow search algorithm. Their findings outperformed alternative techniques with a sensitivity of 99.12%. They also compared the efficacy of deep learning structures for automatically extracting features for the diagnosis of lung nodule CT images with that of conventional computer-aided diagnosis (CADx) systems, which use hand-crafted features. Using a 10-fold cross-validation approach, performance was evaluated, and convolutional neural networks (CNN) outperformed other models with an AUC of 0.899.

In a subsequent study cited as [15], a new metaheuristic algorithm was presented by researchers, who drew inspiration from the Ebola virus's and its associated disease's propagation model. The Ebola optimisation algorithm (EOSA) is the term given to this innovative method. Using two different kinds of benchmark functions—47 classical functions and more than 44 constrained IEEE CEC-benchmark functions—the researchers assessed the efficacy of the suggested paradigm. Based on scalability, convergence, and sensitivity assessments, their results showed that EOSA may rival cutting-edge techniques like genetic algorithms (GA), artificial bee colony (ABC), and particle swarm optimisation (PSO) algorithms.

Another study by [16] suggested a two-stage strategy for detecting lung cancer that involves segmenting and extracting features from image samples as well as putting the image through preprocessing processes. Their proposed approach detected early-stage lung cancer. Image acquisition, preprocessing, binarization, thresholding, segmentation, feature extraction, and neural network detection make up their suggested system. According to their findings, the accuracy of the suggested system was 96.67%.

Next, another study by [17] offered a method for determining the stage of lung cancer that was based on the GA algorithm and k-Nearest Neighbours. The GA method is used to extract the features. On the lung cancer database, the suggested approach received a 100% accuracy rating.

Another study by [18] proposed a deep residual learning method for identifying lung cancer from CT data. To prepare the data for additional analysis, preprocessing and feature extraction were done. They employed the Random Forest and XGBoost techniques, after which the ensemble strategy was put into practice using the individual classifiers. With an 84% success rate on the LIDC-IRDI dataset, their results surpassed those of earlier research.

Next, another study by [19] suggested a novel CNN design for the best possible lung cancer diagnosis. For better network accuracy and optimal design, they employed a metaheuristic technique known as marine predators. Additionally, the RIDER dataset has been used to test the methodology. Pre-trained deep networks like AlexNet, CNN ResNet-18, VGG-19, and GoogLeNet have been used to compare their outcomes. The accuracy, sensitivity, and specificity of their suggested MPA-based method were 93.4%, 98.4%, and 97.1%, respectively.

Another study by [20] to extract the tumour from the lung cancer image, five optimisation strategies were studied and put into practice: k-median clustering, guaranteed convergence particle swarm optimisation (GCPSO), k-means clustering, inertia-weighted particle swarm optimisation, and particle swarm optimisation. They also get the data ready for additional analysis. The GCPSO has the maximum accuracy of 95.89%, according to the data.

Next, another study by [21] suggested a computer-aided detection method for lung cancer that is both automatic and optimised. Improved thermal exchange optimisation (ITEO) served as the basis for feature selection. According to their findings, the accuracy of the suggested approach was 92.27%.

Another study by [22] applied the feature selection method for dimension reduction in clinical X-ray imaging datasets by optimising the rebalancing of the unbalanced class distribution using metaheuristic techniques. Their findings demonstrated that, when applied to the lung X-ray dataset, the Bat algorithm's suggested method produced 94.6% classification accuracy with 88.3% Kappa.

2.9 The Comparison of Selected Literature Review

Table 2.1 Comparison Between Previous Researchers

Author	Preprocessing	Methods	Findings
Revathi et al. (2020)	Noise removal Gabor filter	Multi-swarm particle swarm optimizer	Accuracy: 98%
Das et al. (2020)		Velocity-Enhanced Whale Optimization Algorithm	Accuracy: 84%
Shakeel et al. (2019)	Noise removal Histogram of images Quality enhances images	Deep learning with instantaneously trained neutral networks (DITNN)	Accuracy: 98.42%
Khamparia et al. (2020)		Deep network ensemble method	Accuracy: 87.35%
Zheng et al. (2021)	Normalization Extraction Resample Crop	Combination of radiology analysis and malignancy evaluation network (R2MNet)	Accuracy: 89.90% AUC: 96.27%
Bhandary et al. (2020)	Image separation Nodule segmentation	Deep Learning approach	Accuracy: 97.27%
Alagarsamy et al. (2021)		Metaheuristic based optimization	Sensitivity: 99.2%

Oyelade & Ezugwu (2022)		New metaheuristic algorithm (EOSA)	Friedman test for EOSA is 1.60 which is ranked as the first compared
Miah & Yousuf (2019)	Gray Scale Conversion Normalization Noise Reduction Binary Image Removed Unwanted	Neural network	Accuracy: 96.67%
Maleki et al. (2021)	Remove missing values fill the missing values clean the dataset	Approach based on k Nearest Neighbours and GA algorithm	Accuracy: 100%
Bhatia et.al (2019)	Applications of region growing and morphological operation	XGBoost Random Forest	Accuracy: 84%
Lu et al. (2021)	Noise Removal Image Level Balancing	MPA	Accuracy: 93.4% Sensitivity: 98.4% Specificity:97.1%
Kumar et al. (2019)	Median, Average, Adaptive median, Adaptive histogram Equalization	Guaranteed convergence particle swarm optimization (GCPSO)	Accuracy: 95.89%

Shan & Rezaei (2021)	normalizing and denoising the input images	Improved Thermal Exchange Optimization (ITEO)	Accuracy: 92.27%
Li et al (2019)	Segmentation	Bat algorithm's Kappa	Accuracy: 94.6% Accuracy: 88.3%

2.10 Summary

Finally, this project benefits from a review of various research papers related to the topic. This review gives an overview of different methods and technologies used in previous studies. While examining these studies for this project, several gaps were found in using Deep Learning models for lung cancer detection. One major issue is the long training time required for Deep Neural Network (DNN) models. Additionally, DNN models have many parameters and use high-dimensional data, making training inefficient. To solve these problems, it is necessary to optimize the CNN model using specific optimization methods. Various metaheuristic algorithms have been suggested and used to address these issues. However, effectively classifying lung cancer using deep learning models remains a significant challenge that needs to be resolved.

CHAPTER 3

METHODOLOGY

3.1 Introduction

The methodology for this project focuses on the systematic design and implementation of an automated lung cancer detection system using Convolutional Neural Networks (CNNs) in MATLAB. This section outlines the step-by-step approach, from data collection to the final system design, ensuring accuracy and efficiency in detecting and classifying lung nodules. The methodology integrates key processes such as image preprocessing, segmentation, feature extraction, and classification to analyze CT scan images effectively. Each stage is designed to address the project's objectives and provide reliable results for early lung cancer detection.

3.2 Methodology

The methodology for this project involves designing an automated lung cancer detection system using Convolutional Neural Networks (CNNs) in MATLAB. The process begins with collecting a dataset of lung CT scan images in JPG format, followed by image preprocessing steps such as resizing, noise reduction, and contrast enhancement to improve image quality. Segmentation is performed to isolate lung regions and identify potential nodules. Features like shape, texture, and intensity are then extracted from the segmented images to provide input for the CNN, which classifies the images as cancerous or non-cancerous. A graphical user interface (GUI) is developed in MATLAB to enable users to input CT scan images and view detection results. Finally, the system's performance is evaluated using metrics such as accuracy, sensitivity, and specificity to ensure reliable detection and classification of lung cancer.

3.3 System Block Diagram

Based on Figure 3.1 the system block diagram shows the steps of the automated lung cancer detection system using CNNs in MATLAB. It starts with uploading CT scan images in JPG format, which are improved through preprocessing steps like resizing, noise removal, and enhancing quality. The images are then segmented to focus on the lung area and detect possible nodules. Important details like texture, shape, and intensity are extracted and sent to a CNN, which classifies the images as cancerous or non-cancerous. Finally, the results, including detected nodules and their classification, are shown through a simple and easy-to-use interface.

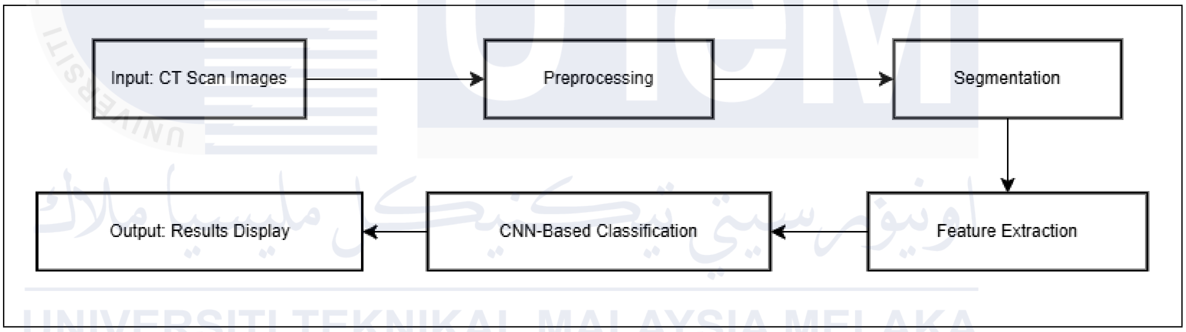


Figure 3.1 Block Diagram

3.4 Flowchart

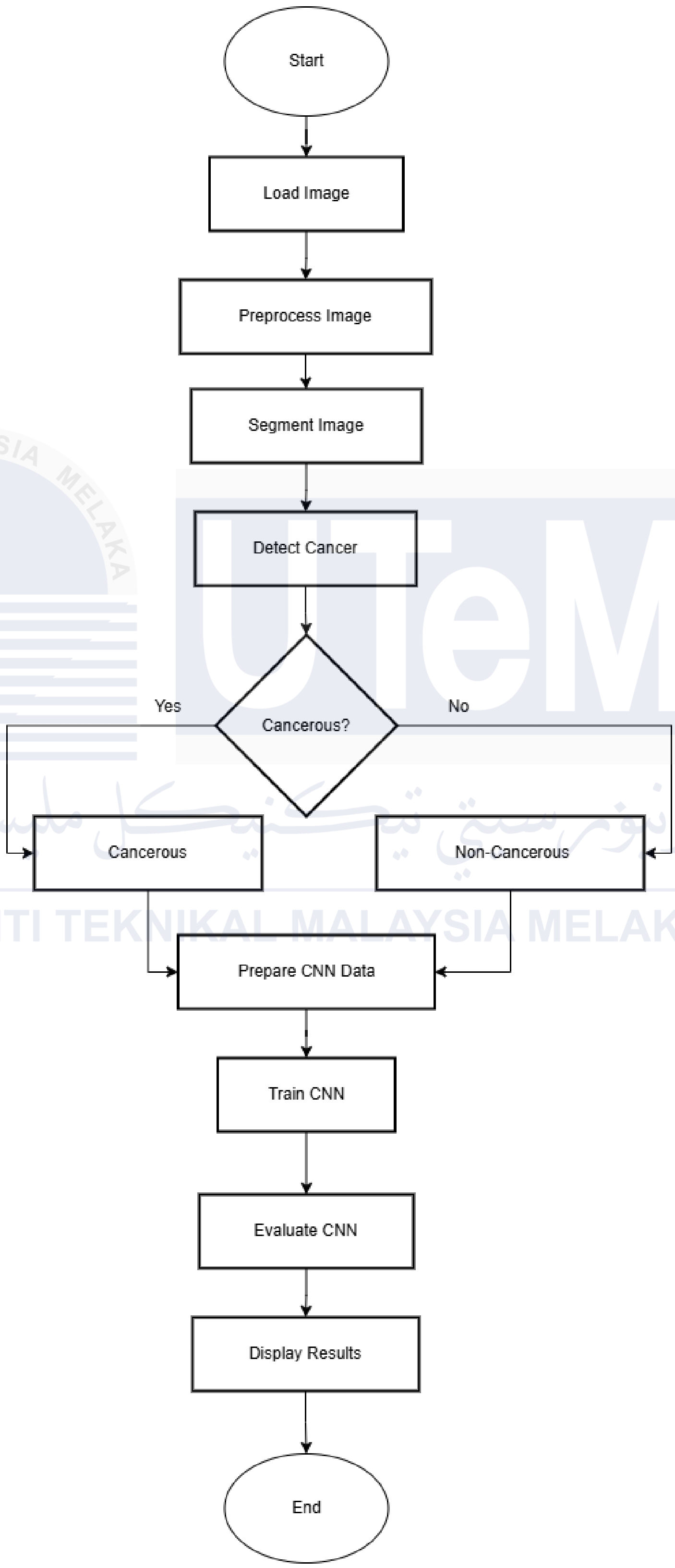
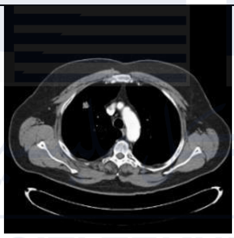
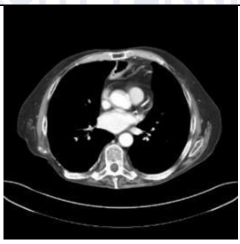
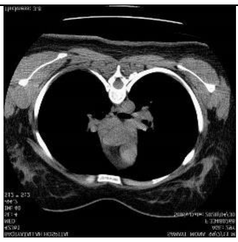


Figure 3.2 Flowchart of Lung Cancer Detection

3.5 Dataset

Images were collected from the hospital situated in Iran. Part of this CT-scan images of lungs belonged to lung cancer patients and classified as cancerous images, and the rest of them belonged to other lung diseases, for instance patients who caught COVID-19, and classified as non-cancerous images. The total number of CT-scan images, which were used in this paper is equal to 287 that 161 of them belong to cancerous images and 126 of the rest belong to non-cancerous images [23]. All each these images were collected with the help of a pulmonologist to skip any probable error in classifying these images. Table 3.1 depicts some sample photos of the classes in the dataset.

Table 3.1 Image of Dataset

Image	Classes Dataset
	Cancerous
	Cancerous
	Non-Cancerous
	Non-Cancerous

3.6 Preprocessing

Preprocessing is an important step in preparing CT scan images for lung cancer detection using machine learning. First, the images are converted to grayscale to simplify the data, focusing on the intensity of the pixels instead of color. Then, the pixel values are normalized, meaning they are scaled to a range between 0 and 1. This helps make the images consistent and improves the performance of machine learning models, like Convolutional Neural Networks (CNNs). Additionally, the intensity distribution of the image is analyzed using a histogram, which helps to identify areas of interest, such as potential cancerous regions, making the next steps like segmentation more effective. These preprocessing steps help ensure the images are in the best possible format for accurate detection and analysis of lung cancer [24].

3.7 Segmentation

Segmentation is the process of finding and separating areas in a CT scan image that might contain cancerous or non-cancerous nodules. The aim is to focus on the parts of the image that are important, such as potential tumors, and ignore irrelevant background or other tissues. This is done by looking at the brightness (intensity) of pixels in the image. Areas with higher intensity could represent denser tissues like nodules. Thresholding techniques are used to separate these areas, while other methods like morphological operations may be used to improve the segmentation and make sure the correct regions are identified. Accurate segmentation helps to isolate the important parts of the image for further analysis, which is essential for improving the accuracy of lung cancer detection [25].

3.8 Feature Extraction

Feature extraction is the process of identifying important details or patterns from the segmented regions of a CT scan image to help decide whether the lung tissue is cancerous or not. After the image is divided into areas of interest, such as potential nodules, the system looks for key characteristics, like the shape, texture, and size of these nodules. For instance, cancerous nodules often have irregular shapes and rough surfaces, while non-cancerous nodules tend to be rounder and smoother. These features are then used to describe the nodules in a way that makes it easier for the system to tell the difference between cancerous and non-cancerous tissues. The extracted features are fed into machine learning models, such as Convolutional Neural Networks (CNNs), to make an accurate classification [26]. The better the features are extracted, the more accurate the lung cancer detection system will be.

3.9 Classification

CNN-based classification is the process of using a Convolutional Neural Network (CNN) to automatically categorize CT scan images into "Cancerous" and "Non-Cancerous" groups. A CNN is a type of deep learning model specifically designed for image analysis, and it can automatically learn to recognize patterns in images without the need for manual feature extraction [26]. After preprocessing, segmentation, and feature extraction, the processed images are fed into the CNN, which consists of multiple layers that progressively extract higher-level features, such as edges, shapes, and textures. During training, the CNN learns to associate these features with the correct labels (cancerous or non-cancerous) by being exposed to a large dataset of labeled CT scan images. Once trained, the CNN can then classify new, unseen CT scan images based on the patterns it has learned. This method is highly effective because it can capture complex relationships and subtle features that might be difficult for traditional machine learning methods to identify. By using CNNs, the system can make accurate and efficient predictions, which significantly enhances the detection of lung cancer in medical images.

3.10 Accuracy

Accuracy is one of the measures for a classification model. It involves calculating the proportion of correctly predicted cases, including true positives and true negatives—among all forecasts in relation to the actual output. in other words, this demonstrates the accuracy with which the model recognizes or categories the incoming data. The formula for precision is given by [27]:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

3.11 Evaluation Matrix

Evaluation metrics are essential for assessing the performance of machine learning models in lung cancer detection. They quantify the model's ability to classify CT scan images as "cancerous" or "non-cancerous" by comparing predictions with actual labels. Key metrics include accuracy, which measures overall correctness, and the confusion matrix, which provides a detailed breakdown of correct and incorrect classifications. Precision and recall are crucial in medical applications, focusing on minimizing false positives and false negatives, while the F1-score balances precision and recall into a single value. These metrics ensure the model is effective, reliable, and suitable for real-world use, especially in critical healthcare contexts where errors can significantly impact outcomes [28].

3.11.1 Confusion Matrix

The confusion matrix is a critical evaluation tool used to measure the performance of classification models, including those for lung cancer detection. It provides a detailed summary of the model's predictions by categorizing them into four outcomes: true positives (TP), where cancerous nodules are correctly identified; true negatives (TN), where non-cancerous cases are accurately classified; false positives (FP), where non-cancerous cases are incorrectly labeled as cancerous; and false negatives (FN), where cancerous nodules are missed. This matrix helps analyze the model's strengths and weaknesses, guiding improvements. Metrics such as accuracy, precision, recall, and F1-score are derived from the confusion matrix, making it a cornerstone for evaluating classification performance in critical applications like medical diagnostics [29].

3.11.2 Precision

Precision is measuring how accurate positive predictions are. It’s defined as the ratio of true positive predictions to the sum of positive predictions (true positives + false positives). A high precision means a model has a low false positive rate. The formula for precision is given by [29]:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

3.11.3 Recall

Recall, also can be referred to as sensitivity or true positive rate, measures a model's ability to find all the relevant instances. It is a ratio of true positive predictions and the total number of actual positive (true positive + false negative). High recall means that a model is being bad at predicting false negative. The formula for recall is given by [29]:

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

3.11.4 F1 Score

The F1 Score is a harmonic mean of both precision and recall, a single metric which comes with a balance between both concerns. However, it's very helpful when working with unbalanced datasets. The value is the F1 score, which ranges from 0, with perfect precision and recall, to 1. The formula for F1 score is given by [29]:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

3.11.5 Specificity

Specificity, also known as the True Negative Rate, is a metric used to measure a classification model's ability to correctly identify negative cases. The formula for Specificity is given by [29]:

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positive}$$

3.11.6 Sensitivity

Sensitivity, also known as the True Positive Rate or Recall, measures a classification model's ability to correctly identify positive cases. The formula for sensitivity is [30]:

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

3.12 User-friendly Graphical User Interface (GUI)

A GUI (Graphical User Interface) is a user-friendly screen with buttons, menus, and visuals that makes it easy to use a system without needing to type commands. It simplifies how people interact with software.

3.13 Summary

This project involves preprocessing CT scan images, segmenting potential cancerous regions, and extracting features like area, mean intensity, and standard deviation. A Convolutional Neural Network (CNN) is used for classifying images as cancerous or non-cancerous. A user-friendly GUI allows users to upload and process images, visualize results, and evaluate the model using accuracy and confusion matrices for lung cancer detection.

CHAPTER 4

RESULTS AND DISCUSSIONS

4.1 Introduction

In this section, the results of the automated lung cancer detection system will be presented. The system uses image processing techniques such as preprocessing, segmentation, and feature extraction, followed by training a Convolutional Neural Network (CNN) to classify CT scan images as either cancerous or non-cancerous. We will evaluate the performance of the system using measures like accuracy, sensitivity, specificity, and confusion matrices to see how well it performs in detecting lung cancer. This part will provide an overview of the results and assess how well the system works in practice.

4.2 Result and Analysis

4.2.1 Automated Lung Cancer Detection in CT scans.

The process begins with the original raw CT scan of the chest. This image shows everything inside, such as the lungs, heart, ribs, and other tissues. However, it also includes unnecessary details that can make it harder to analyze. To make the image clearer, the system adjusts the brightness and contrast, so the lungs become more noticeable, and all the scans look more similar, even if they come from different machines.

Next, the image is turned into a black-and-white version. The lungs are shown as white, and the rest of the chest is black. This makes it easier for the system to focus on the lungs and ignore everything else, like the heart and ribs.

Then comes the segmentation step, where the system separates the lungs from the rest of the chest. It removes things like the heart and ribs, leaving only the lungs for closer examination. This step is important because it helps the system focus on the areas that matter for further analysis.

After that, the system looks closely at the lungs to find anything that might be cancer. It looks for spots that are brighter or have unusual shapes, which could indicate cancer. Once it finds something that seems unusual, it isolates that area to look at it more closely.

Finally, the system highlights the area where cancer might be by drawing a colored line around it, usually yellow This outline is added on top of the original image, making it easy for doctors to see exactly where the cancer is located. This step helps doctors quickly spot the problem and make a diagnosis.

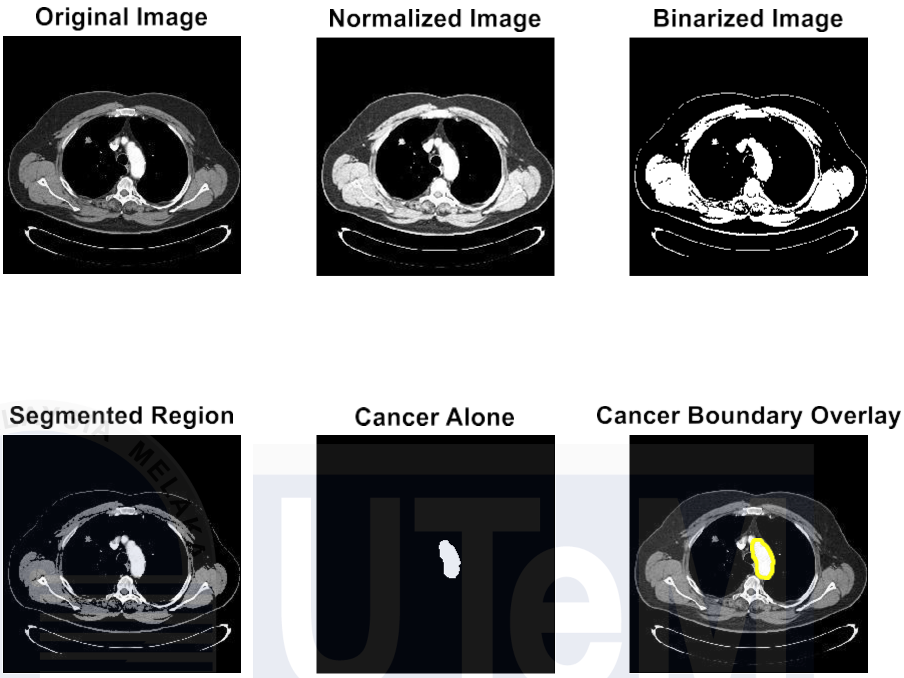


Figure 4.1 Automated Lung Cancer Detection in CT scan

Table 4.1 Feature Calculation for Normalized Image

Area	**Mean**	**Standard Deviation**
40656	0.602044	0.193963

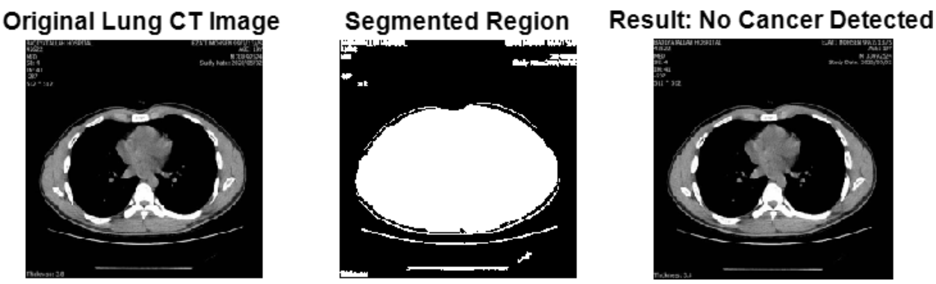


Figure 4.2 Non-cancerous

4.2.2 Histogram of Normalizes Image

The histogram of a normalized image shows the distribution of pixel intensities (brightness levels) after scaling them to a range from 0 to 1, helping to understand how bright or dark different parts of the image are, and how evenly the intensity values are spread.

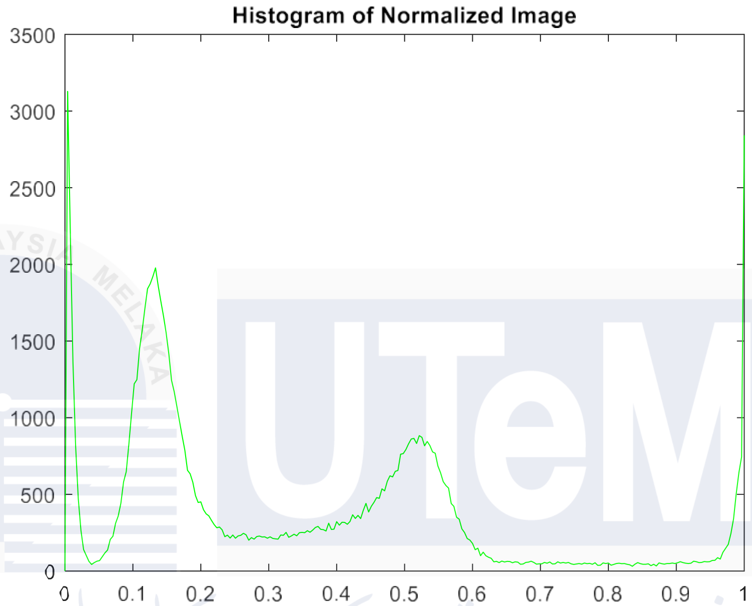


Figure 4.3 Histogram of Normalized Image

4.3 Training Model

The training progress shown in the image highlights how well a Convolutional Neural Network (CNN) is performing for automated lung cancer detection. The top chart shows accuracy, where the training accuracy (blue line) keeps improving and reaches almost 100% by the end, meaning the model is learning well. The validation accuracy (dashed black line), which tests how well the model works on new data, stabilizes at 97.67%, showing it can generalize effectively. The bottom chart shows the loss, or error, getting smaller over time for both training and validation, which means the model is improving and not overfitting. The summary on the side shows the model was trained for 20 cycles (epochs) in about 6 minutes, with a final validation accuracy of 97.67% and a steady learning rate of 0.0001. These results show the model is accurate and reliable. To make it even better, you could check if the dataset covers all types of lung cancer, add more data, or adjust the CNN design to improve its performance further.

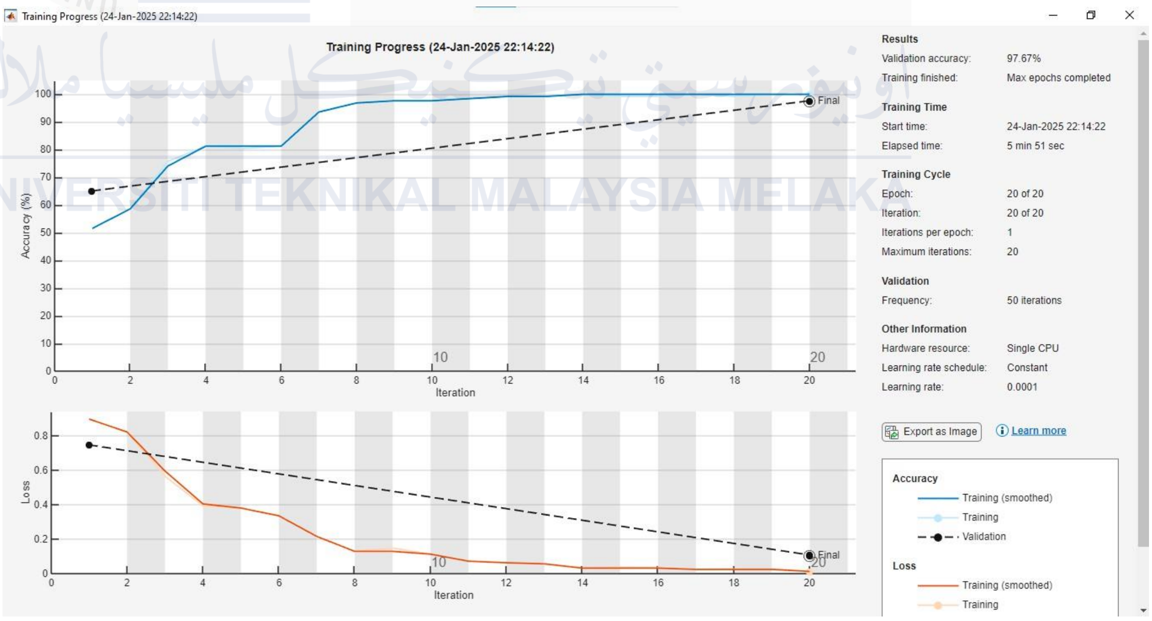


Figure 4.4 Training Progress

Network	Accuracy	Specificity	F1 Score	Sensitivity
CNN	97.67%	0.94	0.97	0.96

4.4 Confusion Matrix

This confusion matrix evaluates the performance of your lung cancer detection model. It shows that the model correctly identified 24 cancerous cases (true positives) and 18 non-cancerous cases (true negatives). It made one mistake by classifying a non-cancerous case as cancerous (false positive) and made no errors in missing actual cancerous cases (false negatives).

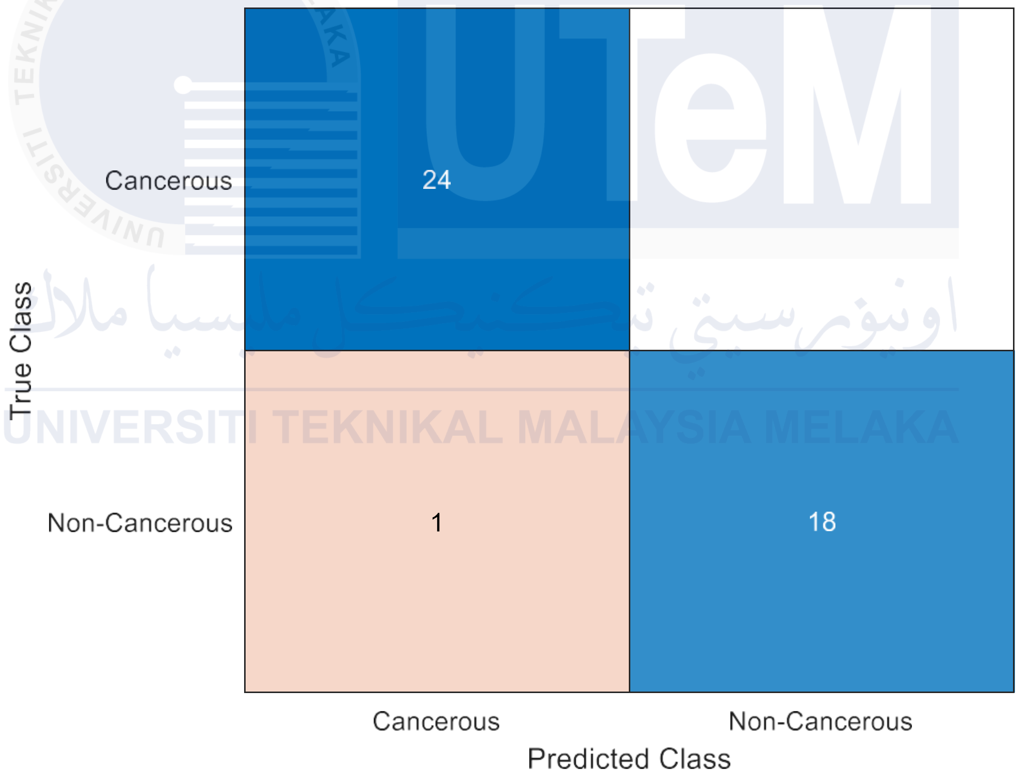


Figure 4.4 Confusion Matrix

4.5 Graphical User Interface (GUI) design

This GUI is designed for a lung cancer detection system using CT images. At the top, the title "LUNG CANCER DETECTION" is displayed in purple, and the interface is divided into sections for different processing steps. The first section, "Lung CT," allows the user to load a CT image using the "Browser CT Image" button and displays the original scan. The "Med Filter" section applies median filtering to reduce noise in the image while preserving important details. In the "Edge Detection" section, edges in the CT scan are highlighted to make abnormalities more visible. The "Segmentation" section isolates the lung region from the image for better focus. Finally, the "Cancer Detection" section identifies and highlights cancerous regions in the scan. If cancer is detected, the result is displayed with the message "Cancer Detected!!!" and highlighted in yellow. This GUI provides a clear and user-friendly workflow for analyzing CT scans and detecting lung cancer.

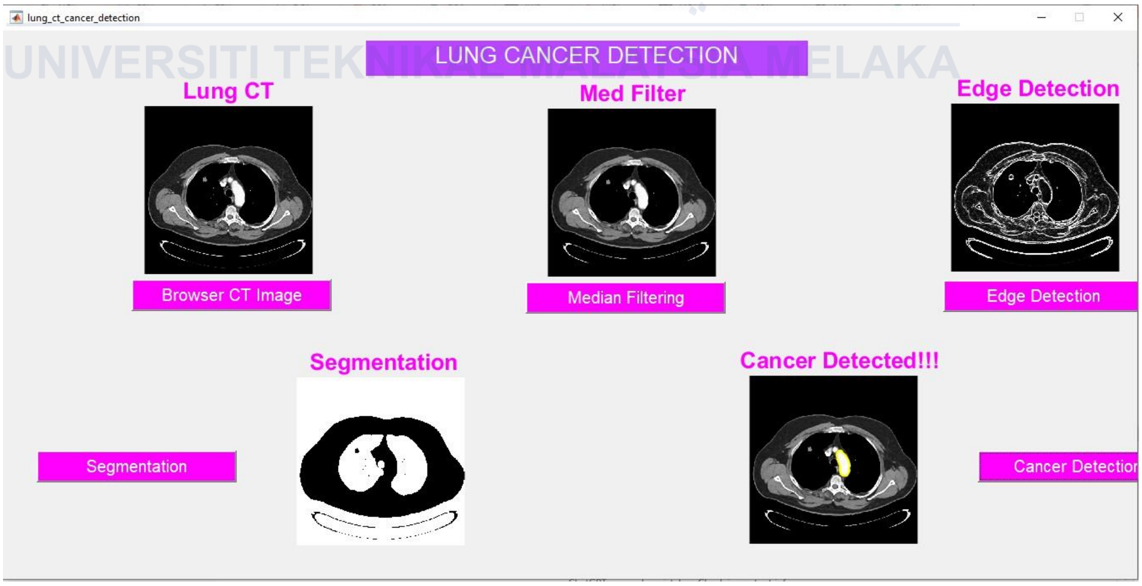


Figure 4.6 GUI design

4.6 Summary

The automated lung cancer detection system uses advanced image processing techniques to analyze CT scan images. It starts by adjusting the brightness and contrast of the raw scans to make the lungs more visible and converts the images to black and white, helping focus on the lungs. The system then isolates the lung region by removing unnecessary details like the heart and ribs, making it easier to examine. It looks for any unusual spots or shapes in the lungs that could indicate cancer. If it finds something suspicious, it highlights the area with a colored outline, helping doctors spot potential cancer quickly.

The system was trained using a Convolutional Neural Network (CNN) and showed excellent results, with the accuracy of the training process reaching nearly 100% and a validation accuracy of 97.67%. The model kept improving, and the error decreased over time, indicating it was learning well and not making mistakes. The confusion matrix showed that the system correctly identified 24 cancerous and 18 non-cancerous cases, with only one false positive and no false negatives, proving it is reliable.

A simple and easy-to-use Graphical User Interface (GUI) was also created, which allows users to load CT images, apply filters to reduce noise, highlight edges, isolate lung areas, and detect cancer. If cancer is found, the system shows a message saying "Cancer Detected!!!" to make it clear for doctors. Overall, the system is accurate, efficient, and reliable, offering a useful tool for doctors to detect lung cancer early from CT scans.

CHAPTER 5

CONCLUSION AND RECOMMENDATIONS

5.1 Conclusion

The automated lung cancer detection system using Convolutional Neural Networks (CNNs) in MATLAB successfully achieved its objectives. By leveraging advanced image processing and deep learning techniques, the system demonstrated an accuracy of 97.67%, with a sensitivity of 96% and specificity of 94%. These results highlight the reliability of the model in identifying cancerous and non-cancerous lung nodules in CT scan images. The proposed system simplifies the diagnostic process through a user-friendly GUI, enabling seamless interaction for medical professionals. Overall, this project significantly contributes to the early detection of lung cancer, potentially improving patient outcomes and reducing the global burden of this disease.

5.2 Potential for Commercialization

This lung cancer detection system has great potential to be used in real-world healthcare settings. One way it could be used is by adding it to online healthcare platforms where doctors and patients can connect remotely. With more people using telemedicine, especially in areas where specialists are not available, this system could let patients or local clinics upload CT scan images to be analyzed. The system would give results quickly, helping doctors make faster and more accurate decisions.

Another possibility is to work with companies that make CT scan machines. The system could be built directly into these machines so that the scans are analyzed immediately after they are taken. This would save time, as doctors wouldn't have to wait for separate analysis, and could start treatment sooner if needed.

The system could also become part of a smartphone or computer app for general health use. This app could help people check for early signs of lung cancer by analyzing CT scans and giving them advice on whether they need to see a doctor. It could also remind users about regular check-ups or provide health tips, making it a tool for early detection and prevention.

5.3 Future Work

For future improvements, the performance and estimation results of lung cancer detection can be enhanced as follows:

I. Using more data

- Even though the system works well, there are ways to make it better. One way is to use more data to train it. Right now, the system uses images from a specific group of people, but adding more images from different groups or rare types of lung cancer could help it work better for a wider range of patients. This would make the system more reliable and useful in different countries and hospitals

II. Incorporating other health information

- Another way to improve the system is by combining it with other health information, like blood test results or personal medical history. This extra information could help the system make even better predictions about whether someone has lung cancer. It could give doctors a fuller picture of the patient's health and help with more accurate diagnoses.

III. Enhancing system speed for real-time diagnosis

- Finally,optimizing the system to deliver instant lung cancer detection results would significantly reduce the waiting time for diagnoses. This would allow doctors to make faster, more confident decisions, improving the efficiency of medical consultations. For patients, it means quicker treatment initiation, leading to better prognosis and reduced anxiety during the diagnostic process.

REFERENCES

[1] [1] F. Bray *et al.*, "Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A cancer journal for clinicians*, vol. 74, no. 3, pp. 229–263, Apr. 2024, doi: https://doi.org/10.3322/caac.21834.

[2] Sandy McDowel, "Lung Cancer Kills More People Worldwide Than Other Cancer Types," *Cancer.org*, Jul. 29, 2024. https://www.cancer.org/research/acs-research-news/lung-cancer-kills-more-people-worldwide-than-other-cancers.html?utm_source=chatgpt.com (accessed Jan. 25, 2025).

[3] World, "Lung cancer," *Who.int*, Jun. 26, 2023. https://www.who.int/news-room/fact-sheets/detail/lung-cancer?utm_source=chatgpt.com

[4] H. Jia, J. Zhang, K. Ma, X. Qiao, L. Ren, and X. Shi, "Application of convolutional neural networks in medical images: a bibliometric analysis," *Quantitative Imaging in Medicine and Surgery*, vol. 14, no. 5, pp. 3501–3518, May 2024, doi: https://doi.org/10.21037/qims-23-1600.

[5] C. Zhang *et al.*, "Toward an Expert Level of Lung Cancer Detection and Classification Using a Deep Convolutional Neural Network," *The Oncologist*, vol. 24, no. 9, pp. 1159–1165, Apr. 2019, doi: https://doi.org/10.1634/theoncologist.2018-0908.

[6] National Cancer Institute, "Cancer of the Lung and Bronchus - Cancer Stat Facts," *National Cancer Institute*, 2024. https://seer.cancer.gov/statfacts/html/lungb.html

[7] American Lung Association, “Types of Lung Cancer,” *www.lung.org*, Jul. 25, 2023. https://www.lung.org/lung-health-diseases/lung-disease-lookup/lung-cancer/basics/lung-cancer-types

[8] M. Revathi, I. J. S. Jeya, and S. N. Deepa, “Deep learning-based soft computing model for image classification application,” *Soft Computing*, May 2020, doi: https://doi.org/10.1007/s00500-020-05048-7.

[9] S. Das, S. Mishra, and Manas Ranjan Senapati, “New Approaches in Metaheuristic to Classify Medical Data Using Artificial Neural Network,” *Arabian Journal for Science and Engineering*, vol. 45, no. 3, Jul. 2019, doi: https://doi.org/10.1007/s13369-019-04026-y.

[10] P. M. Shakeel, M. A. Burhanuddin, and M. I. Desa, “Lung cancer detection from CT image using improved profuse clustering and deep learning instantaneously trained neural networks,” *Measurement*, vol. 145, pp. 702–712, Oct. 2019, doi: https://doi.org/10.1016/j.measurement.2019.05.027.

[11] A. Khamparia *et al.*, “A novel deep learning-based multi-model ensemble method for the prediction of neuromuscular disorders,” *Neural Computing and Applications*, vol. 32, no. 15, pp. 11083–11095, Nov. 2018, doi: https://doi.org/10.1007/s00521-018-3896-0.

[12] S. Zheng *et al.*, “Interpretative computer-aided lung cancer diagnosis: From radiology analysis to malignancy evaluation,” vol. 210, pp. 106363–106363, Oct. 2021, doi: https://doi.org/10.1016/j.cmpb.2021.106363.

[13] A. Bhandary *et al.*, “Deep-learning framework to detect lung abnormality – A study with chest X-Ray and lung CT scan images,” *Pattern Recognition Letters*, vol. 129, pp. 271–278, Jan. 2020, doi: https://doi.org/10.1016/j.patrec.2019.11.013.

[14] Saravanan Alagarsamy, R. R. Subramanian, Theepika Shree, S. kannan, and Vishnuvarthanan Govindaraj, “Prediction of Lung Cancer using Meta-Heuristic based Optimization Technique: Crow Search Technique,” *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, pp. 186–191, Feb. 2021, doi: https://doi.org/10.1109/ICCCIS51004.2021.9397199.

[15] Oyelade, Olaide N and Ezugwu, Absalom E, “Ebola Optimization Search Algorithm (EOSA): A new metaheuristic algorithm based on the propagation model of Ebola virus disease,” *arXiv (Cornell University)*, Jan. 2021, doi: https://doi.org/10.48550/arxiv.2106.01416.

[16] Md. B. A. Miah and M. A. Yousuf, “Detection of lung cancer from CT image using image processing and neural network,” *IEEE Xplore*, May 01, 2015. https://ieeexplore.ieee.org/abstract/document/7307530?casa_token=gMBPzVWVjK4AAAAA:wcr4hG1PwLPiPQpOvcM-OYFNx1NkN3-3b1H7CK8SgWTu1A7pTvB6Bz7PMj1gDCLsLameBW2ueq0

[17] “Maleki, N., Zeinali, Y. and Niaki, S.T.A. (2021) A k-NN Method for Lung Cancer Prognosis with the Use of a Genetic Algorithm for Feature Selection. Expert Systems with Applications, 164, Article ID 113981. - References - Scientific Research Publishing,” *Scirp.org*, 2021.

https://www.scirp.org/reference/referencespapers?referenceid=3476571 (accessed Jan. 25, 2025).

[18] S. Bhatia, Y. Sinha, and L. Goel, “Lung Cancer Detection: A Deep Learning Approach,” *Advances in Intelligent Systems and Computing*, pp. 699–705, Oct. 2018, doi: https://doi.org/10.1007/978-981-13-1595-4_55.

[19] X. Lu, Y. A. Nanehkaran, and M. Karimi Fard, “A Method for Optimal Detection of Lung Cancer Based on Deep Learning Optimized by Marine Predators Algorithm,” *Computational Intelligence and Neuroscience*, vol. 2021, p. e3694723, Aug. 2021, doi: https://doi.org/10.1155/2021/3694723.

[20] K. Senthil Kumar, K. Venkatalakshmi, and K. Karthikeyan, “Lung Cancer Detection Using Image Segmentation by means of Various Evolutionary Algorithms,” *Computational and Mathematical Methods in Medicine*, vol. 2019, pp. 1–16, Jan. 2019, doi: https://doi.org/10.1155/2019/4909846.

[21] R. Shan and T. Rezaei, “Lung Cancer Diagnosis Based on an ANN Optimized by Improved TEO Algorithm,” *Computational intelligence and neuroscience*, vol. 2021, p. 6078524, Winter 2021, doi: https://doi.org/10.1155/2021/6078524.

[22] J. Li, S. Fong, L. Liu, N. Dey, A. S. Ashour, and L. Moraru, “Dual feature selection and rebalancing strategy using metaheuristic optimization algorithms in X-ray image datasets,” *Multimedia Tools and Applications*, vol. 78, no. 15, pp. 20913–20933, Mar. 2019, doi: https://doi.org/10.1007/s11042-019-7354-5.

[23] N. Maleki, “CT-Scan images,” *data.mendeley.com*, vol. 1, Jun. 2020, doi: https://doi.org/10.17632/p2r42nm2ty.1.

[24] A. Nagarajan, “Image processing techniques for analyzing CT scan images towards the early detection of lung cancer,” *Bioinformation*, vol. 15, no. 8, pp. 596–599, Aug. 2019, doi: https://doi.org/10.6026/97320630015596.

[25] Y. Said, A. A. Alsheikhy, Tawfeeq Shawly, and Husam Lahza, “Medical Images Segmentation for Lung Cancer Diagnosis Based on Deep Learning Architectures,” *Diagnostics*, vol. 13, no. 3, pp. 546–546, Feb. 2023, doi: https://doi.org/10.3390/diagnostics13030546.

[26] Simie Endalew and Kaur Mandeep, “Lung cancer detection using Convolutional Neural Network (CNN),” Nov. 18, 2019. https://www.researchgate.net/publication/385036583_Lung_cancer_detection_using_Convolutional_Neural_Network_CNN

[27] L. Wang, “Deep Learning Techniques to Diagnose Lung Cancer,” *Cancers*, vol. 14, no. 22, p. 5569, Jan. 2022, doi: https://doi.org/10.3390/cancers14225569.

[28] Vishnuprasanth Vijaya Kumar, “Evaluating machine learning models-metrics and techniques,” *AI Accelerator Institute*, Feb. 05, 2024. https://www.aiacceleratorinstitute.com/evaluating-machine-learning-models-metrics-and-techniques/

[29] GeeksForGeeks, “Confusion Matrix in Machine Learning - GeeksforGeeks,” *GeeksForGeeks*, Feb. 07, 2018. https://www.geeksforgeeks.org/confusion-matrix-machine-learning/

[30] D. G. Diaz, “Sensitivity is the proportion of truly diseased persons ...,” *GrepMed*, Apr. 03, 2018. https://www.grepmed.com/images/1480/negativepredictivevalue-

positivepredictivevalue-epidemiology-sensitivity-calculation (accessed Jan. 25, 2025).

APPENDICES

Appendix A Gantt Chart PSM 1

ACTIVITY	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14
Confirmation project's title	PSM Briefing and The Registration						Mid Term Break							
Introduction (Chapter 1)														
Project progress														
Update Logbook														
Research journals (Literature review)														
Methodology (Chapter 3)														
Preliminary result analysis														
Full report progress														
Presentation PSM 1														

Appendix B Gantt Chart PSM 2

ACTIVITY	W1	W2	W3	W4	W5	W6	W7	W8	W9	W10	W11	W12	W13	W14
Development Feature extraction	■	■						Mid Term Break						
Development GUI		■	■	■			■							
Performance evaluation						■	■			■	■	■		
Update Logbook					■	■				■				
Result and Discussion (Chapter 4)					■	■	■		■	■	■	■		
Testing and Validation							■		■	■				
Result analysis							■			■	■	■	■	
Presentation PSM 2													■	
Submission of final report														■