# UNIVERSITI TEKNOLOGI MARA

# A MODIFIED WATERSHED SEGMENTATION ALGORTIHM FOR AUTOMATED LUNG CANCER LESION DETECTION IN CT SCAN IMAGES

**NUR NAJIHAH SOFIA BINTI MOHD MARZUKI**

Thesis submitted in fulfillment
of the requirements for the degree of
**Master of Science**
**(Electrical Engineering)**

**College of Engineering**

**March 2024**

# ABSTRACT

Lung cancer is one of the leading cancers among both men and women and ranks among the most dangerous and life-threatening diseases worldwide. Computed Tomography (CT) imaging is a common method for detecting lung cancer. In the era of advanced computer technology, Computed Aided Diagnosis (CAD) has gained prominence particularly in medical applications such as diagnosing lung cancer. In medical applications, such as diagnosing lung cancer, CAD systems are adopted, utilizing lung CT images as input and various algorithms to assist doctors in image analysis and decision-making. A key problem addressed in this study is the challenge of accurately distinguishing between lesion and non-lesion areas in lung CT images. This difficulty underscores the necessity for an effective segmentation method, a critical aspect highlighted throughout this research. Therefore, the main aim of this research is to establish a segmentation method suitable for the automated detection of lung cancer in CT scan images. To achieve this goal, the research is divided into three parts: image pre-processing, segmentation to detect lung cancer lesions, and feature extraction for lung cancer identification. The dataset utilized comprises images from The Cancer Imaging Archive (TCIA Images) as a benchmark and the Imaging Department at the Advanced Medical and Dental Institute (AMDI), USM, Bertam, Pulau Pinang, Malaysia, involving 50 subjects. The image processing technique is implemented using MATLAB software, with the input CT scan images in grayscale. Initially, the images undergo thresholding, image filtering, and enhancement processes to obtain clearer lung area images. This pre-processing stage is essential to improve image quality, removing unwanted information that may obscure important features and reducing distortion or noise. The next crucial stage is the segmentation, where modified watershed is used to demarcate the lung region in the CT scan images. The performance of the segmentation process in conventional and modified watershed segmentation technique for detecting lung lesion, which produce average F-Score is 97.80% and 99.09%, respectively. The outcome of this research is highly valuable for doctors in determining appropriate treatments for patients and diagnosing lung cancer and nodules from the images. Implementing CAD on CT scan images can potentially aid doctors in diagnosing lung cancer and nodules effectively.

# ACKNOWLEDGEMENT

In the name of Allah S.W.T, Most Gracious, Most Merciful. Being grateful for His bless that giving me the opportunity to embark on my MSc and for completing this long and challenging journey successfully. My gratitude and thanks go to my supervisor Assoc. Prof Ts. Dr. Siti Noraini Sulaiman, and co-supervisors, Dr. Noor Khairiah A. Karim and Prof. Dr. Nooritawati Md. Tahir. Thank you for the support, patience and ideas in assisting me with this project. I also would like to express my gratitude to the staff of the Advanced Medical and Dental Institute (AMDI), Bertam Pulau Pinang especially Prof. Dr Ibrahim Lutfi Shuaib and Suzana Ismail for providing the facilities, knowledge and assistance.

My appreciation goes to all supportive staff from Advanced Control System and Computing research group (ACSCRG) and all members of Centre for Electrical Engineering Studies, UiTM Pulau Pinang who were both directly and indirectly involved. Without their cooperation and help I would not have been able to complete this project.

Thanks to my dear husband Sahal bin Mohd Radzuan who was willing to engage and support my daily struggles. Loving thanks goes to my sons Muhammad Yousuf Imaan and Muhammad Noah Imaan who have been the sources of encouragement and inspiration to me throughout my life.

Finally, this thesis is dedicated to my family members especially my parent, Mohd Marzuki and for their support and encouragements in the completion of this project. They were always together with me when I faced problems in this project. A special appreciation goes to Dr. Iza Sazanita for her support in my journey.

Thank you for your support, encouragement, ideas and comments for the success of this project. This victory is dedicated to all of you.

Alhamdulillah.

# TABLE OF CONTENTS

# CHAPTER 1
# INTRODUCTION

## 1.1 Introduction

Lung cancer is one of the main causes of cancer death in the U.S and worldwide [1]. According to the estimated number in 2023 by the American Cancer Society [2], there are about 14% or 234,340 new lung cancer cases and 127,070 deaths because of lung cancer. According to P. Rajadurai et al., in Malaysia lung cancer is the highest incident of cancer caused accounting for approximately 10% of all malignancies [3]. The survival rate of lung cancer patients in Malaysia at 1 and 5 years is one of the lowest compared to other cancer types. The lungs are the organ located in the chest on either side of the heart within the rib cage. There are a pair of spongy, air-filled organs responsible for breathing. There are two major types of lung cancer; Small Cell Lung Cancer (SCLC) and Non- Small Cell Lung Cancer (NSCLC) [4]. The human's body is composed of cells, and each cell is frequently divides based on its DNA (according to genetic orders) to create tissue. When a cell starts to divide uncontrollably and out of orders in the lung, the tumor is created.

Lung cancer is a disease characterized by uncontrolled growth and the formation of abnormal cells in the lungs. It is a fast growing tumor with the capability to spreads or invade other organs [5]. The addiction of smoking cigarettes, exposure to carcinogenic environment like radioactive gas and air pollution, are the main factors contributing lung cancer. Additionally, genetic factors also have the contribution to cause lung cancer [6]. According to the Khin and Aung [7], the types of lung cancer disease can be divided into four stages. In stage I, the cancer is confined to the lung. Stage II and III involve larger and more invasive tumors, but remain within the chest. Stage IV is characterized by cancer spreading from the chest to other parts of the body.

Medical image processing techniques are widely used by many researchers, mainly driven by the need to develop improved methods for clinical analysis and diagnosis. Early diagnosis of lung cancer is crucial for effective treatment and reducing number of deaths.