

Detecting Cancerous Nodules From Chest X-rays Using Deep Learning Techniques

CSE 499 report submitted in partial fulfillment of the requirements for the degree

of

Bachelor of Science in Computer Science and Engineering

Md. Tareq Mahmud
1811174042

Nafis Iqbal
1811250642

Shayam Imtiaz
1811710642

Done Under the Supervision of

Dr. Sifat Momen

Associate Professor



May 16, 2022

DECLARATION

Project Title Detecting Cancerous Nodules From Chest X-rays Using Deep Learning Techniques
Author *Md. Tareq Mahmud, Nafis Iqbal and Shayam Imtiaz*
Student IDs 1811174042, 1811250642 and 1811710642
Supervisor Dr. Sifat Momen

This report is prepared as a requirement of the Capstone Design Project CSE499 A & B which is a two semester long senior design course. We declare that this **cse499** report entitled **Detecting Cancerous Nodules From Chest X-rays Using Deep Learning Techniques**, has not been accepted for any degree and is not currently submitted in candidature of any other degree. We would like to request you to accept this report as a partial fulfillment of Bachelor of Science degree under Electrical and Computer Engineering Department of North South University.

Md. Tareq Mahmud
1811174042

Department of Electrical Computer Engineering
North South University

Nafis Iqbal
1811250642

Department of Electrical Computer Engineering
North South University

Shayam Imtiaz
1811710642

Department of Electrical Computer Engineering
North South University

Date: May 16, 2022



Department of Electrical & Computer Engineering
North South University
Bashundhara, Dhaka - 1229, Bangladesh

APPROVAL

This is to certify that the cse499 report entitled **Detecting Cancerous Nodules From Chest X-rays Using Deep Learning Techniques**, submitted by **Md. Tareq Mahmud** (Student ID: 1811174042), **Nafis Iqbal** (Student ID: 1811520642) and **Shayam Imtiaz** (Student ID: 1811710642) are undergraduate students of the **Department of Electrical Computer Engineering**. This report partially fulfills the requirements for the degree of Bachelor of Science in Computer Science and Engineering on May 16, 2022, and has been accepted as satisfactory.

Dr. Sifat Momen
Associate Professor

Department of Electrical Computer Engineering
North South University

Dr. Mohammad Rezaul Bari
Associate Professor & Chairman

Department of Electrical Computer Engineering
North South University

Place: Dhaka, Bangladesh

Date: May 16, 2022

ACKNOWLEDGEMENTS

- We acknowledge the continuous guidance we received from our supervisor **Dr. Sifat Momen.**

Md. Tareq Mahmud, Nafis Iqbal, Shayam Imtiaz

North South University

May 16, 2022

Contents

1	Introduction	3
1.1	Problem Statement	3
1.2	Motivation	3
1.3	Goals and Objectives	4
1.4	Importance in countries like Bangladesh	4
1.5	Detection Approaches	4
1.6	Report Layout	5
2	Literature Review	6
2.1	Summary	7
3	Proposed Methodology	8
3.1	Faster RCNN with Resnet 50	8
3.2	YOLOv5	10
3.3	EfficientDet	11
4	Experimental Setup	13
4.1	Dataset	13
4.2	Training environment and Machine configuration	13
4.3	Training on Faster-RCNNResnet50	13
4.4	Training on Yolov5l	14
4.5	Training on EfficientDet	14
5	Evaluation Metrics	15
5.1	Precision	15
5.2	Recall	15
5.3	F1 score	16
5.4	mean Average Precision	16
6	Results and analysis	17
6.1	Results	17
6.2	Result Analysis and Comparison	17
6.3	Discussion	19
7	Deploying the model	21
8	Conclusion and Future Work	23

List of Figures

3.1	Faster RCNN with Resnet 50 model architecture	8
3.2	YOLOv5 model architecture	11
3.3	EfficientDet model Architecture	12
6.1	Result comparison of the 3 models	18
6.2	Results after training yolov5l for 62 epochs	18
6.3	Precision Recall and F1 score for Yolov5l	18
6.4	Yolov5l detection on validation set	19
7.1	Deploying the model on a web server	21
7.2	User interface of the webapp	21
7.3	Showing detection results on webapp	22

Chapter 1

Introduction

Cancer is the leading cause of death worldwide, accounting for nearly 10 million deaths in 2020, or almost one in six deaths(37). The most common cancers are breast, lung, colon, rectum, and prostate cancers. Lung cancer is a type of cancer that develops in the tissues of the lungs, most commonly in the cells lining the airways(6). It is the leading cause of cancer death among men and the second leading cause of cancer death among women worldwide(32). More than 2.2 million new lung cancer cases were reported in 2020 alone(3). According to the National Cancer Institute of America (15), lung cancer is responsible for 51.6 percent of cancer-related deaths among males and 34.4 percent of cancer-related deaths among women in the United States. It demonstrates that lung cancer patients had the highest mortality rate.

1.1 Problem Statement

Pulmonary nodules are the first signs and symptoms of lung cancer. Pulmonary nodules are frequently discovered due to regular examination or CXR imaging for reasons unrelated to lung cancer. Nevertheless, long before clinical symptoms or any indicators appear(4), a chest radiograph (CXR) can reveal them, and chest radiography is the most common radiological exam globally. Although CXR can detect early-stage lung cancer, radiologists face a challenging task due to the various sizes of lung nodules(7). Furthermore, the density of pulmonary nodules varies significantly across the human body. As a result, CXR is crucial in correctly identifying nodules in the quest for early lung cancer diagnosis(4). Additionally, bones and organs frequently conceal lung nodules(4). Simultaneously, as the number of new cases grows, The danger of inaccuracy in detecting malignant nodules is a very complex undertaking. Thus CXR interpretation can quickly become complicated for radiologists(4). To detect the characteristics of lung nodules on CXRs, researchers devised a computer-aided detection (CADe) technique.

1.2 Motivation

Accurate early discovery of a malignant lung nodule can considerably improve the patient's chances of survival; yet, detection of early-stage lung cancer has been a serious concern for