

Electronic Thesis and Dissertation Repository

4-12-2024 10:00 AM

Machine Learning Classifiers for Chronic Obstructive Pulmonary Disease Assessment Using Lung CT Data.

Halimah Alsurayhi,

Supervisor: Abbas Samani, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Doctor of Philosophy degree in Electrical and Computer Engineering

© Halimah Alsurayhi 2024

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Biomedical Commons](#), and the [Other Electrical and Computer Engineering Commons](#)

Recommended Citation

Alsurayhi, Halimah, "Machine Learning Classifiers for Chronic Obstructive Pulmonary Disease Assessment Using Lung CT Data." (2024). *Electronic Thesis and Dissertation Repository*. 10047. <https://ir.lib.uwo.ca/etd/10047>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

Chronic Obstructive Pulmonary Disease (COPD) is a condition characterized by persistent inflammation and airflow blockages in the lungs, contributing to a significant number of deaths globally each year. To guide tailored treatment strategies and mitigate future risks, the Global Initiative for Chronic Obstructive Lung Disease (GOLD) employs a multifaceted assessment system of COPD severity, considering patient's lung function, symptoms, and exacerbation history. COPD staging systems, such as the high-resolution eight-stage COPD system and the GOLD 2023 three staging systems, have been later developed based on these factors. Lung Computed Tomography (CT) is becoming increasingly crucial in investigating COPD as it can detect various COPD phenotypes, such as emphysema, bronchial wall thickening, and gas trapping. Deep learning techniques show promise in leveraging CT imaging to assess the severity of COPD. This thesis uses lung CT data in conjunction with machine learning techniques to classify COPD patients according to these staging systems. For the eight-stage system, both Neural Network and Convolutional Neural Network (CNN) approaches were employed for classification. To develop the Neural Network model, features were extracted from lung CT scans at inspiration and expiration breathing phases, including lung air features and COPD phenotypes features. The CNN model utilized a single lung CT scan at the expiration phase. The GOLD 2023 three staging system involves training separate CNN models using lung CT scans at expiration to predict symptom levels and COPD exacerbation risk. In this thesis, in addition to models trained from scratch, Transfer Learning was also employed to develop models for the eight-stage COPD classification, Symptom level prediction, and exacerbation risk prediction. The developed classifiers demonstrate reasonably high classification performance, indicating their potential for deployment in clinical settings to enhance COPD assessment using image data.

Keywords

Chronic Obstructive Pulmonary Disease, Lung CT, Machine Learning, Deep Learning, Transfer Learning, Lung CT Image Augmentation, Lung Air Map, COPD Staging System

Summary for Lay Audience

Summary Chronic Obstructive Pulmonary Disease (COPD) is a health condition that affects the lungs and can cause inflammation and blockages in their airways. It is a leading cause of death worldwide. To help treat COPD, the Global Initiative for Chronic Obstructive Lung Disease (GOLD) uses three factors to assess the disease's severity: the patient's lung function, symptoms, and history of exacerbation (flare-ups). There are two classifications used to determine the stage of COPD: the eight-stage system and the GOLD 2023 three staging systems. These staging systems use the three factors to determine the severity of the disease. Lung computed tomography (CT) images can be used to investigate COPD because they can detect different types of conditions, including lung tissue destruction, airway wall thickening, and air trapping in the lungs. Machine learning techniques can be used to assess the severity of COPD automatically using CT imaging. In this study, machine learning was used to classify COPD patients according to the eight-stage and GOLD 2023 three staging systems. Two machine learning methods were used to develop the eight-stage classifiers: Neural Networks (NN) and Convolutional Neural networks (CNN). For the NN model, lung imaging algorithms were used to extract features from paired lung CT at the inhalation and exhalation breathing phases. To develop the GOLD2023 classifier, two separate CNN models were also trained to predict symptom levels and the risk of COPD exacerbation. The developed classifiers showed promising results, demonstrating their potential for clinical use in improving COPD assessment using image data.

Dedication

This thesis is dedicated to my parents for their unwavering love, support, and encouragement.

And to all those with COPD. Your strength and patience are truly inspiring. Let us work towards a better understanding of this condition and strive for more effective treatments that can improve the quality of life for everyone affected.

Co-Authorship Statement

For the project presented in Chapter 2, the major contributors are Halimah Alsurayhi and Dr. Abbas Samani; Halimah was the first author to develop all the code and write the initial draft of the manuscript. Dr. Samani was responsible for conceiving ideas and consultation for algorithm development, supervision, and editing the manuscripts. Dr. Najafi contributed to the study design, provided feedback on the clinical contents of the manuscript and edited the manuscript.

The primary authors for the projects outlined in Chapters 3 and 4 are Halimah Alsurayhi and Dr. Abbas Samani. Halimah was the first author, developing the code and writing the initial manuscript draft. Dr. Samani contributed by advising on algorithm development, supervising the project, and editing the manuscripts.

Acknowledgments

I would like to start by expressing my deep appreciation to Dr. Abbas Samani, my supervisor, for his constant support and invaluable guidance during my time as a graduate student. Dr. Samani has been instrumental in shaping me both as a researcher and as an individual. Under his supervision, I learned the importance of setting clear research goals, maintaining persistence, and fostering creativity. During our brainstorming sessions to address various challenges, Dr. Samani nurtured my problem-solving and communication skills, pushing me to excel. His insightful feedback and attention to detail elevated this project to new heights, and his encouragement to strive for more has been invaluable. I am truly fortunate to have had the opportunity to work with Dr. Samani, whose belief in me and compassionate support helped me navigate through the complexities of this project.

I am also grateful to the members of my advisory committee, Drs. Mahdi Najafi, Ali Sadeghi, and Katarina Grolinger, for their valuable insights and contributions to my research. Their expertise enriched my understanding of the subject matter and provided invaluable guidance along the way. Additionally, I extend my appreciation to my lab mates for fostering a productive and supportive environment. Their willingness to assist with coding challenges and refine my presentation skills has been immensely beneficial.

Above all, I am deeply thankful to my parents and my loving family for their unwavering support and encouragement. My husband, Khalid, has been a constant source of inspiration, challenging me to push my limits and strive for excellence. His unwavering support and infectious optimism have been my pillars of strength throughout this journey. I am profoundly grateful for everything he has done to uplift me and help me reach this milestone.

I would like to express my gratitude to Umm Al-Qura University and the Cultural Bureau of Saudi Arabia in Canada for their advice and financial support.

Table of Contents

Abstract.....	ii
Summary for Lay Audience.....	iii
Dedication.....	iv
Co-Authorship Statement.....	v
Acknowledgments.....	vi
Table of Contents.....	vii
List of Tables.....	xii
List of Figures.....	xiii
List of Appendices.....	xv
List of Abbreviations.....	xvi
Chapter 1.....	1
1 « INTRODUCTION ».....	1
1.1 Motivation and Rationale.....	1
1.2 Lung Structure and Function.....	3
1.2.1 Lung Airways.....	4
1.2.2 Lung Parenchyma.....	5
1.3 Pathophysiology of Chronic Obstructive Pulmonary Disease.....	5
1.3.1 Chronic Bronchitis.....	5
1.3.2 Emphysema.....	6
1.4 Clinical Measures of Global Lung Function.....	6
1.4.1 Pulmonary Function Testing.....	6
1.4.2 Symptoms Reporting.....	7
1.4.3 Exacerbation Frequency.....	8
1.4.4 Computed Tomography Imaging.....	9

1.5 COPD Staging Systems	9
1.6 Clinical Management for COPD.....	10
1.7 Lung CT Image Processing and Analysis Methods	10
1.7.1 Lung Volume Segmentation	11
1.7.2 Lung Volumes Registration	11
1.8 Machine Learning	13
1.8.1 Convolutional Neural Network (CNN).....	14
1.8.2 Residual Network (ResNet)	15
1.8.3 Model Hyperparameters.....	16
1.8.4 Data Partitioning and Evaluation Metrics	18
1.8.5 Transfer Learning.....	20
1.9 Thesis Hypotheses and Objectives.....	21
1.10 Thesis Outline	21
1.10.1 Chapter 2: Neural Network for COPD Eight Staging System Based on Features Extracted from Inspiratory/ Expiratory Lung 3D CT.....	22
1.10.2 Chapter 3: CNN with a Single Lung 3D CT Volume for COPD Eight Staging System.....	22
1.10.3 Chapter 4: CNN with a Single Lung 3D CT Volume for COPD Classification Based on The GOLD 2023 Staging System	23
1.10.4 Chapter 5: Summary, Conclusion, and Future Work.....	23
1.11 References.....	23
Chapter 2.....	29
2 « Neural Network for COPD Eight Staging System Based on Features Extracted from Inspiratory/ Expiratory Lung 3D CT»	29
2.1 Introduction.....	29
2.2 Materials & Methods	32

2.2.1	Dataset.....	32
2.2.2	Lung CT data preprocessing	33
2.2.3	Feature Extraction.....	36
2.2.4	NN Model Training.....	40
2.3	Results.....	41
2.3.1	Features Cross-Correlation Coefficients.....	41
2.3.2	Neural Network Models Evaluation	42
2.4	Discussion.....	45
2.5	Conclusions.....	47
2.6	References.....	48
Chapter 3	51
3	« CNN with a Single Lung 3D CT Volume for COPD Eight Staging System»	51
3.1	Introduction.....	51
3.2	Materials & Methods	54
3.2.1	Dataset.....	54
3.2.2	Data Preprocessing.....	55
3.2.3	Transfer Learning for COPD Eight Staging	55
3.2.4	3D-CNN model for COPD Eight Staging.....	56
3.2.5	3D-CNN model with hybrid inputs for COPD Eight Staging	57
3.3	Results.....	58
3.3.1	Transfer Learning for COPD Eight Staging Evaluation	58
3.3.2	3D-CNN model for COPD Eight Staging Evaluation	59
3.3.3	3D-CNN model with hybrid inputs for COPD Eight Staging Evaluation	59
3.4	Discussion.....	60
3.5	Conclusions.....	62

3.6	References.....	63
Chapter 4	66
4	« CNN with a Single Lung 3D CT Volume for COPD Classification Based on The GOLD 2023 Staging System».....	66
4.1	Introduction.....	66
4.2	Materials & Methods	69
4.2.1	Proposed GOLD2023 based COPD classification.....	69
4.2.2	Dataset:	69
4.2.3	Transfer Learning.....	70
4.2.4	Thoracic CT Preprocessing.....	71
4.2.5	Pretrained Models for Transfer Learning	71
4.2.6	Symptoms Detection.....	72
4.2.7	Exacerbation Detection.....	74
4.3	Results.....	75
4.3.1	<i>Symptoms Detection Evaluation</i>	75
4.3.2	<i>Exacerbation Detection Evaluation</i>	78
4.3.3	<i>Evaluation of the GOLD2023 COPD Staging Model</i>	82
4.4	Discussion.....	83
4.5	Conclusions.....	85
4.6	References.....	86
Chapter 5	89
5	« Summary, Conclusion, and Future Work».....	89
5.1	Summary	89
5.2	Conclusions and Future Work	92
Appendices	96

Curriculum Vitae 109

List of Tables

Table 1-1: GOLD cut-off values for diagnosing COPD via pulmonary function tests.	7
Table 2-1: Distribution of COPD subjects among the Eight classes	33
Table 2-2: Correlation coefficients between the lungs' air features and corresponding PFT measurements.....	41
Table 2-3: Correlation coefficients between phenotype features and PFT measurements.	42
Table 2-4: Comparison between our proposed COPD classification model and the other COPD classification techniques.....	45
Table 3-1: Distribution of COPD subjects among the Eight classes	54
Table 4-1: Symptoms' models classification metrics.....	77
Table 4-2: Exacerbation detection models evaluation metrics	80
Table 4-3: GOLD2023 classification accuracy.....	82

List of Figures

Figure 1-1: The frequency of hospitalization for primary causes of hospitalization in Canada [3]..... 2

Figure 1-2: The probability of surviving is decreased with increased exacerbation frequency [4]..... 2

Figure 1-3: Human Airways generation diagram 4

Figure 1-4: Pathophysiology of the Chronic Obstructive Pulmonary Disease. Copyright Illustration 273493892 | Anatomy © Tetiana Pavliuchenko | Dreamstime.com 6

Figure 1-5: Diagram of a simple neural network with an input layer, two hidden layers and an output layer. 13

Figure 1-6: Typical CNN architecture [34] 15

Figure 1-7: Residual learning block with the skip connection [35]..... 16

Figure 1-8: Illustration of Confusion Matrix. 19

Figure 2-1: An example of segmented lung mask from a coronal view 34

Figure 2-2: An example of a lung CT scan at inspiration, expiration, and the registered image at inspiration to expiration. 35

Figure 2-3: Examples of coronal views of original and corresponding lung CT scans..... 36

Figure 2-4: Inspiration and expiration histograms and the initial thresholds A and B in the lung air segmentation technique. 37

Figure 2-5: An example of a lung air map for inspiration and expiration scans. 38

Figure 2-6: The distribution of voxels with varying values at inspiration and expiration for an individual with normal lung function (a), and another with severe lung function, FEV1 = 18 (b) [11]. 40

Figure 2-7: NN-CT Performance in terms of Accuracy, Precision, Recall, and F1 score..... 43

Figure 2-8: NN-Hybrid Performance in terms of Accuracy, Precision, Recall, and F1 score.44

Figure 2-9: Confusion Matrices for NN-Ct and NN-Hybrid 44

Figure 3-1: 3D-CNN architecture for the proposed eight staging schemes.....	57
Figure 3-2: Diagram showing the illustration of the classification model.....	58
Figure 3-3: Confusion matrices for the eight staging models using transfer learning.....	59
Figure 3-4: Confusion matrices for the eight staging models using the proposed 3D CNN model.....	60
Figure 4-1: COPD staging scheme based on GOLD2023 staging system.....	67
Figure 4-2: Proposed GOLD2023 classification scheme.....	70
Figure 4-3: Proposed symptoms-CNN model structure.....	73
Figure 4-4: Exacerbation-CNN Model Structure.....	74
Figure 4-5: Loss learning curve for training and validation sets	75
Figure 4-6: Confusion Matrix and ROC curve for symptoms models: (a) fine-tuned ResNet-18 Confusion Matrix, (b) fine-tuned ResNet-18 ROC curve, (c) fine-tuned ResNet-34 Confusion Matrix, (d) fine-tuned ResNet-18 ROC curve, (e) Symptoms-CNN Confusion Matrix, (f) Symptoms-CNN ROC curve.....	78
Figure 4-7: Confusion Matrix and ROC curve for the exacerbation models: (a) fine-tuned ResNet-18 Confusion Matrix, (b) fine-tuned ResNet-18 ROC curve, (c) fine-tuned ResNet-34 Confusion Matrix, (d) fine-tuned ResNet-18 ROC curve, (e) Exacerbation-CNN Confusion Matrix, (f) Exacerbation-CNN ROC curve.....	81

List of Appendices

Appendix A: St. George's Respiratory Questionnaire (SGRQ)	96
Appendix B: Modified Medical Research Council) Dyspnea Scale.....	103
Appendix C: Permissions for Reproduction of Scientific Articles	104

List of Abbreviations

COPD	Chronic Obstructive Pulmonary Disease
GOLD	Global Initiative for Chronic Obstructive Lung Disease
PFT	Pulmonary Function Test
CT	Computed Tomography.
mMRC	Modified Medical Research Council dyspnea scale
CAT	COPD Assessment Tool
SGRQ	St. George's Respiratory Questionnaire
PRM	Parametric Response Mapping
fSAD	Functional Small Airway Disease
NN	Neural Network
NN-CT	Neural Network trained with image features
NN-Hybrid	Neural Network trained with image features and other two clinical features
CNN	Convolutional Neural Network
CNN-Hybrid	Convolutional Neural Network trained with image and other two clinical features
ResNet	Residual Network
ResNet-18	Residual Network with 18 layers
ResNet-34	Residual Network with 34 layers
GOLD2023	Global Initiative for Chronic Obstructive Lung Disease 2023 updated staging system
Symptom-CNN	Convolutional Neural Network trained to detect Symptom Severity

Exacerbation-CNN Convolutional Neural Network trained to detect Exacerbation Severity

Chapter 1

1 « INTRODUCTION »

Chronic obstructive pulmonary disease (COPD) is a prevalent and progressive lung condition. The management of COPD typically relies on data derived from lung function assessment through spirometry testing whereby symptom severity and risk of exacerbation is evaluated. This thesis explores visible and underlying structural information within CT images using quantitative measurements and deep learning techniques to enhance COPD diagnosis before facilitating treatment planning.

1.1 Motivation and Rationale

COPD is a debilitating and progressive lung disease that results in persistent coughing, respiratory difficulties, diminished physical capacity, and reduced lung performance in those afflicted. The manifestations of COPD, including breathlessness, coughing, and excessive mucus production, can have a significant impact on patients' daily lives, overall health, and functional abilities. Respiratory failure and exacerbations contribute significantly to COPD-related deaths, making it the third leading cause of death worldwide in 2019 [1].

COPD disease significantly burdens individuals, healthcare systems, and societies at large. Approximately 4% of Canadians were diagnosed with COPD in 2020 [2]. This has increased the burden on the healthcare system through intensifying hospitalization frequency, emergency department visits, and healthcare provider consultations. Figure 1-1 displays the frequency of hospitalization for the primary causes of hospitalization in Canada where COPD has the highest rate of hospitalization. Notably, many patients are admitted for multiple hospital visits compared to other diseases responsible for significant hospitalizations [3].

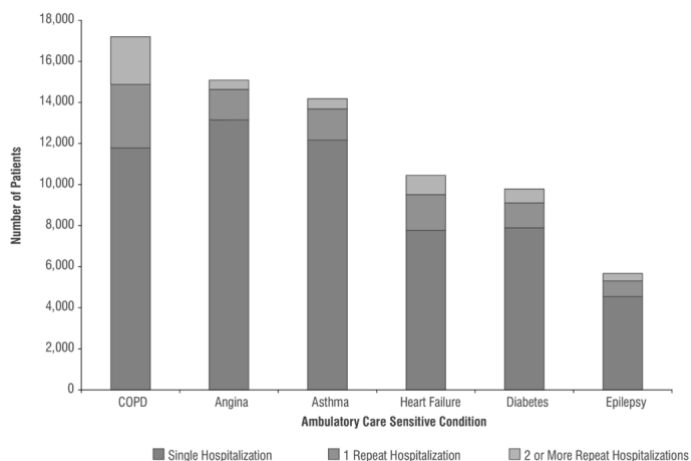


Figure 1-1: The frequency of hospitalization for primary causes of hospitalization in Canada [3].

Experiencing a COPD exacerbation increases the risk of subsequent episodes of exacerbation. The frequency of severe exacerbations is associated with an increased risk of mortality [4]. Figure 1-2 shows that a higher frequency of exacerbations is associated with a lower probability of surviving. In clinical practice, managing symptoms is the primary focus of treatment rather than employing disease-modifying therapies [5].

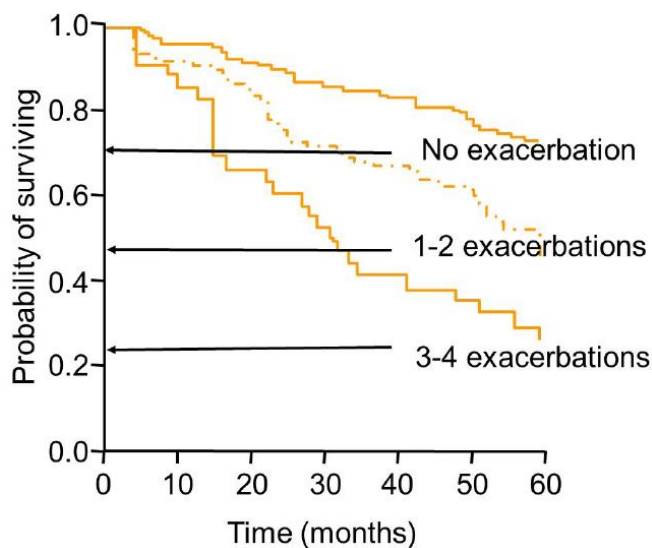


Figure 1-2: The probability of surviving is decreased with increased exacerbation frequency [4].

Addressing the multifaceted burden of COPD requires comprehensive management strategies aimed at improving outcomes and enhancing patients' quality of life. As of now, there is no cure for COPD [6]. Instead, the current management goal is to improve health status, prevent exacerbations and prevent related complications and mortality [6] [5].

Spirometry tests, also known as Pulmonary Function Tests (PFT), are currently used to evaluate lung function. However, these tests only provide a general measure of the diverse nature of Chronic Obstructive Pulmonary Disease (COPD), which manifests in different forms, such as airway obstruction, lung tissue destruction, and physiological changes like vascular abnormalities. Furthermore, PFT is not sensitive enough to detect the disease in its early stages. Therefore, researchers are exploring pulmonary imaging techniques to provide a more localized understanding of the structural abnormalities of the disease and to detect it in its early stages to prevent further progression. It has been reported that the disease can be detected in the early stages through quantitative CT measurements before any noticeable changes are detected with PFT [7].

This chapter provides background information for understanding the motivation and rationale behind the research work detailed in Chapters 2, 3, and 4. Section (1.2) offers an overview of lung structure and function, while Section (1.3) explores the pathophysiology of COPD. Section (1.4) outlines the clinical measures used to assess COPD, and Section (1.5) specifies the different staging systems used for COPD. Additionally, Section (1.6) introduces the current treatment options for managing the disease. The theoretical background for CT imaging and machine learning methods required to develop the proposed methods are provided in Section (1.7) and Section (1.8), respectively. Section (1.9) introduces the hypothesis and objectives of the thesis. Lastly, Section (1.10) presents the thesis outline.

1.2 Lung Structure and Function

The lung structure encompasses the airways, parenchyma, and vasculature. A well-organized network of these components is structured to facilitate the seamless distribution and exchange of oxygen into the bloodstream.

1.2.1 Lung Airways

The human lung has an intricate airway network. The large airways are divided into generations 0 to 7 and kept open by cartilage. The walls of airways have a smooth muscle layer that can dilate or constrict, and the innermost layer has cilia that eliminate mucus and dirt. The clearance of mucus is essential in protecting airway function by preventing and eliminating obstructions. Without proper clearance, obstructions could impede downstream ventilation, leading to regional loss of function. The airway bifurcates into each lung, lobe, and segment, which divides into 19 bronchopulmonary segments [8]. The first 16 generations are conducting airways, and the last 7 are where gas exchange occurs as illustrated in Figure 1-3. At the terminal end, the smallest airways, known as terminal bronchioles, terminate in clusters of air sacs called alveoli, where the crucial process of gas exchange occurs.

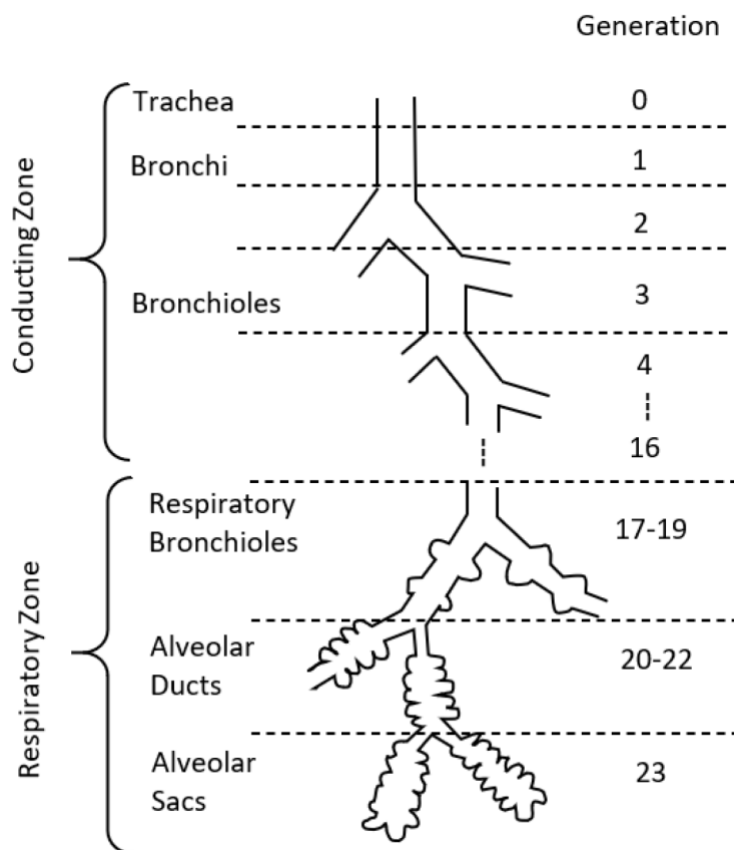


Figure 1-3: Human Airways generation diagram

Adapted from Respiratory Physiology: The Essentials [8], permission provided in Appendix C

1.2.2 Lung Parenchyma

The lung parenchyma comprises the functional tissue of the lung, which performs gas exchange. This tissue includes the alveoli, alveolar ducts, and respiratory bronchioles (Figure 1-3). The parenchyma comprises delicate structures that enable the exchange of oxygen and carbon dioxide between the bloodstream and the surrounding air. This exchange takes place through the thin walls of the alveoli, which are surrounded by a network of capillaries. The capillaries involved in gas exchange are only slightly larger than a single blood cell, and the alveolar-capillary membrane is incredibly thin, around 0.2-0.3 μm , facilitating fast diffusion of gases across it [8].

1.3 Pathophysiology of Chronic Obstructive Pulmonary Disease

The pathophysiology of COPD is complex and multifactorial, involving inflammation, structural changes, and impaired gas exchange, typically caused by significant exposure to noxious particles or gases. There are two main pathologies: lung airway inflammation and/or obstruction which is called Chronic Bronchitis, and parenchymal lung destruction, which is called Emphysema [5].

1.3.1 Chronic Bronchitis

Chronic bronchitis is the inflammation and irritation of the bronchial tubes, resulting in an increase in mucus production and airway narrowing (Figure 1-4). As the inflammation continues, the bronchial glands may produce an excess of mucus, leading to a chronic cough with sputum production. This excess mucus can obstruct the airways, making it difficult for air to flow freely in and out of the lungs. This obstruction can cause wheezing, chest tightness, and shortness of breath, especially during physical activity.

Over time, chronic inflammation and mucus production can cause structural changes in the airways, such as thickening of the bronchial walls, scar tissue formation, and narrowing of the airway lumen. These changes further worsen airflow limitation and reduce lung function [8].

1.3.2 Emphysema

Emphysema is a medical condition that causes the alveoli, where gas exchange occurs, to become enlarged due to damage to the lung tissue. This damage results in the loss of terminal bronchioles, reducing the lungs' ability to exchange gases [8].

The extent of emphysematous lung volume is consistent with the decrease in lung capacity. The primary cause of this condition is the damage to alveolar walls and the loss of elastic recoiling ability in lung tissue, leading to decreased elasticity (Figure 1-4). These changes cause reduced lung compliance, resulting in symptoms such as dyspnea, wheezing, and chronic cough, which are characteristic of emphysema [8].

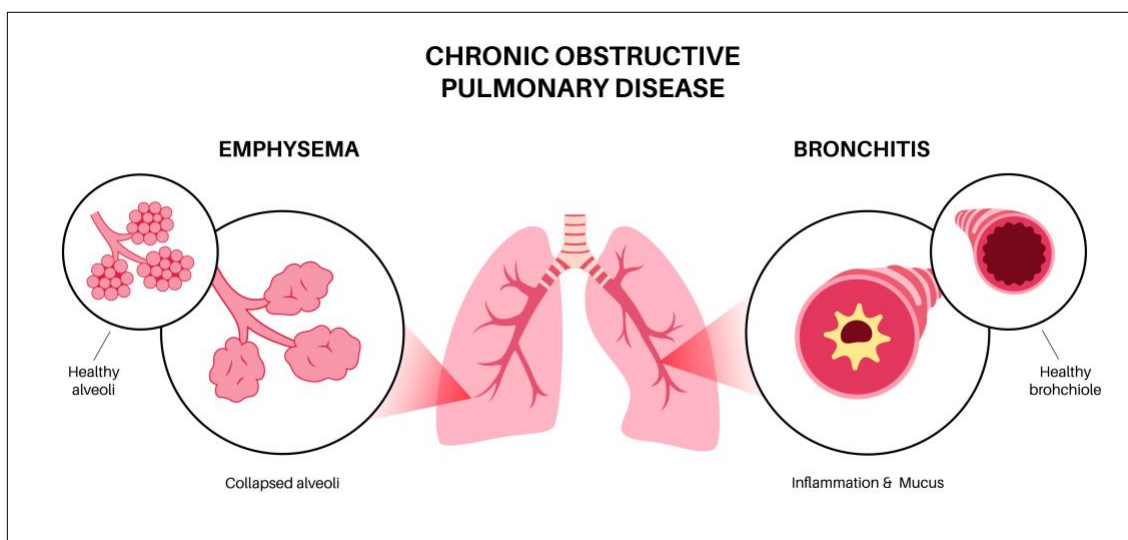


Figure 1-4: Pathophysiology of the Chronic Obstructive Pulmonary Disease. Copyright Illustration 273493892 | Anatomy © Tetiana Pavliuchenko | Dreamstime.com

1.4 Clinical Measures of Global Lung Function

1.4.1 Pulmonary Function Testing

The diagnosis of Chronic Obstructive Pulmonary Disease (COPD) depends primarily on the results of pulmonary function testing. A handheld spirometer is used to measure airflow obstruction and the results are reported as a percentage of predicted values. These predicted values are determined based on the patient's race, sex, age, and height. Forced Expiratory Volume in One Second (FEV1) is a measure of the percentage of total lung volume that a

person can exhale in the first second of expiration, whereas Forced Vital Capacity (FVC) is the total amount of air that a person forcefully exhales. If the ratio of FEV1/FVC is less than 70%, it indicates airflow obstruction. The severity of airflow obstruction is classified based on FEV1 into four stages, with a lower percentage of predicted values indicating increased obstruction [5]. The four stages that represent the severity of COPD are presented in Table 1-1.

Table 1-1: GOLD cut-off values for diagnosing COPD via pulmonary function tests.

GOLD Stages	Pulmonary Function Test FEV1
GOLD Stage 1 (Mild)	FEV1 > 80% predicted
GOLD Stage 2 (Moderate)	50% < FEV1 < 80% predicted
GOLD Stage 3 (Severe)	30% < FEV1 < 50% predicted
GOLD Stage 4 (Very Severe)	FEV1 < 30% predicted

The criteria used in pulmonary function tests (PFTs) are not capable of differentiating among the various causes of obstruction of airway inflammation or emphysema. Additionally, they do not provide any insights into the heterogeneous nature of the disease across different lung regions. Although this testing process is standardized to ensure consistent outcomes, it can be challenging to obtain accurate results in certain populations such as children or patients who struggle with following instructions or completing testing, which may lead to inaccuracies. To address this limitation, imaging technologies have been employed to enhance disease diagnosis.

1.4.2 Symptoms Reporting

Various tests are available to assess the level of symptoms in patients suffering from COPD. These tests are aimed at evaluating symptoms such as shortness of breath, cough, sputum production, and their impact on daily activities. One commonly used test is the

COPD Assessment Test (CAT), which is a simple questionnaire that assesses the severity and impact of symptoms on an individual's life. Another commonly used tool is the Modified Medical Research Council (mMRC) Dyspnea Scale, which measures the severity of breathlessness (Appendix B) [5].

The St. George's Respiratory Questionnaire (SGRQ) is another questionnaire that evaluates the impact of COPD on the quality of life of the patient (Appendix A) [5]. It considers various aspects such as daily activities, symptoms, and the overall impact of the disease on the individual's life. The questionnaire provides scores for symptoms, activity, impact, and an overall score. A lower score indicates a lower burden of the disease.

However, a limitation of using such self-report tests to evaluate chronic disease is that these tests may introduce a bias in the results. Factors such as the patients' sex may affect symptom reporting, leading to inaccurate results [9]. Moreover, patients with chronic disease may have unconsciously modified their activities over time to manage their disease and may not identify specific limitations related to their COPD, such as not taking the stairs due to exercise limitations.

1.4.3 Exacerbation Frequency

Exacerbations in COPD are episodes where symptoms worsen and require medical intervention, significantly impacting patients' quality of life. These episodes are an important measure used to assess the severity and progression of the disease. Evaluating the frequency of exacerbations involves monitoring the number of episodes experienced by a patient over a specific period, such as one year. This information helps healthcare providers determine the appropriate treatment strategies, including medication adjustments and preventive measures, to minimize exacerbation risk and improve patient outcomes.

In clinical settings, patients are considered at high risk based on the number of hospitalizations they required in the prior year. However, this criterion is not always reliable, as not all COPD exacerbations result in hospitalization. Furthermore, the rate of hospitalization for COPD exacerbations has decreased by 53% compared to the pre-COVID pandemic period [10]. This reduction may be due to changes in treatment strategies

caused by avoiding emergency room visits during the pandemic. Therefore, other assessment tools are essential to detect the severity of COPD exacerbation.

1.4.4 Computed Tomography Imaging

Computed Tomography (CT) imaging is crucial in assessing chronic obstructive pulmonary disease (COPD). It provides detailed information about the structure and pathology of the lungs. CT generates an attenuation map of the lungs, which shows the density of different regions. A density of -1000 Hounsfield units (HU) indicates the presence of air, thus voxels with a density near -1000 HU indicate regions of tissue destruction. These measurements can be used to measure emphysema, which has a low density of lung tissue. Therefore, a threshold of -950 HU is commonly used to identify the destruction of tissue.

Other quantitative CT measurements have been used to investigate changes in the structure and volumes of the lungs, including changes in airway and blood vessels [11], [12], [13], [14]. Parametric response mapping is a technique that applies multiple methods of quantitative CT analysis to identify different imaging phenotypes of COPD [11]. By generating voxel-wise maps of emphysema, gas trapping, and functional small airways disease from inspiration-expiration CT scans, researchers can identify relative volumes of each phenotype in participants with COPD.

1.5 COPD Staging Systems

COPD staging systems have undergone several updates to enhance disease management and prevent progression. Initially, a four-stage system based on pulmonary function tests (PFTs) was developed and utilized. In 2011, the GOLD committee introduced the ABCD classification system for COPD, which considers symptom severity and exacerbation history alongside lung function assessment, providing a more comprehensive evaluation of disease severity [2]. However, this system introduced some heterogeneity within groups C and D, as the risk factors were not clearly delineated [16]. Subsequently, a high-resolution eight-stage system was developed by the COPDGene study to address this issue [17], [18]. More recently, in 2023, an ABE three-stage system was proposed, focusing primarily on

symptom severity and COPD risk [5]. This system emerged from studies highlighting the significant contribution of exacerbations to disease progression and COPD mortality [4].

1.6 Clinical Management for COPD

Inhaled bronchodilators are an important part of the management of COPD. The choice of bronchodilator is based on the patient's symptoms, lung function, and risk of exacerbation. Combining bronchodilators with different mechanisms and durations of action can improve bronchodilation and reduce side effects compared to increasing the dose of a single bronchodilator [19] [5]. Studies have shown that combining short-acting β 2-agonists (SABAs) and short-acting muscarinic antagonists (SAMAs) is more effective in improving FEV1 and symptoms than using either medication alone [20]. There are many combinations of long-acting β 2-agonists (LABAs) and long-acting muscarinic antagonists (LAMAs) available in a single inhaler. Clinical trials have shown that using a LABA+LAMA combination improves lung function and symptoms compared to using only one inhaler, especially in patients with a low exacerbation risk who are not receiving inhaled corticosteroids [19]. A lower-dose, twice-daily regimen of LABA+LAMA has also been shown to improve symptoms and health status in COPD patients [21].

While most studies on LABA+LAMA combinations have been conducted in patients with a low exacerbation rate, evidence suggests that this combination may be more effective than monotherapy in preventing exacerbations, particularly in patients with a history of frequent exacerbations [22], [23]. However, studies have produced mixed results regarding the comparative efficacy of LABA+LAMA versus LABA+inhaled corticosteroid (ICS) combinations, with some indicating that LABA+ICS may reduce exacerbations to a greater extent [24].

1.7 Lung CT Image Processing and Analysis Methods

In order to accurately extract lung information from CT images, various image processing methods are utilized, including lung volume segmentation, registration, and thresholding.

1.7.1 Lung Volume Segmentation

Image segmentation involves dividing an image into distinct segments, where each segment represents voxels sharing similar properties. This process finds applications in computer vision, facilitating tasks like object recognition, image database searches, and image compression. In medical image analysis, segmentation is crucial for isolating specific organs or structures from surrounding tissue. For instance, in lung imaging, segmentation is used to delineate the lung volume from the background. Medical image segmentation enables comparison across datasets or modalities, tracks changes in structures over time (e.g., treatment response), and facilitates quantitative analysis of images.

There are several automatic lung segmentation methods to extract lung volumes from CT scans. Rayan's method combines thresholding and region-growing segmentation to extract left and right lung volumes. The lung is detected using Hounsfield unit (HU) thresholding, and large airways are identified by a lower HU region near the mid-line. A region-growing technique is then used to exclude the large airways from the lung [25].

Another method for lung segmentation utilizes deep learning, specifically a network architecture called U-net. U-net is a convolutional neural network (CNN) structure commonly used for image segmentation tasks [26]. It consists of an encoder-decoder architecture, where the encoder captures features from the input image at multiple scales, and the decoder generates the segmentation mask based on these features. U-net is particularly well-suited for medical image segmentation tasks due to its ability to capture detailed spatial information and handle complex structures.

1.7.2 Lung Volumes Registration

Image registration refers to the process of aligning two or more images of the same scene captured under different conditions. This can include images taken from slightly different angles, viewpoints or state, such as lung CT scans at different breathing phases, or using different image modalities. In image registration, one of the images is designated as the fixed, target or reference image, while the other is considered the moving image. Geometric

transformations or local displacement fields are then applied to the moving image to bring it into alignment with the target image [27], [28]. Image registration is usually performed as a preparatory step before applying other image processing algorithms such as image subtraction. For example, in medical imaging, CT images of a patient's lungs taken at different phases of the respiratory cycle may be registered to quantify certain pathophysiological conditions accurately [11] [29].

Image registration process requires a combination of different transformations, depending on the differences between the acquired images. These transformations can include rigid transformations, which involve translation and rotation, affine transformations, which include rigid transformations along with scaling and shearing, or deformable registration, which allows for more complex transformations with 2-3 degrees of freedom for each pixel. By applying these transformations, the images can be accurately aligned, taking into account variations in position, orientation, and deformation between them.

There are several essential components in any image registration algorithm that play crucial roles in achieving accurate alignment between the moving and target images [30]. Firstly, a metric is utilized to assess the similarity or closeness between the two images after applying a transformation. The choice of metric, such as mean squared errors, normalized cross correlation, or normalized mutual information, significantly influences the effectiveness of the registration process. Secondly, an optimizer is employed to optimize the chosen similarity metric, aiming to maximize it to achieve the best alignment. Selecting the appropriate optimizer and adjusting its parameters, such as step size and relaxation factor, is essential for optimizing the registration algorithm's performance. Lastly, interpolators are required to determine the grayscale values of pixels in the transformed image. Interpolators help reconstruct the transformed image accurately by estimating pixel values at non-integer locations because transformed pixels may not align perfectly with the target image grid. Collectively, these components contribute to the successful alignment of images during the registration process.

Due to significant changes in lung volume shape across breathing phases, deformable image registration methods are necessary. Free Form Deformation (FFD)) emerges as a

suitable approach for lung image registration because it can capture complex deformations and variations in lung morphology) [31].

1.8 Machine Learning

Machine learning is a specific area of study within artificial intelligence, which aims to develop and understand methods for computer systems to learn similarly to humans. By utilizing various statistical models and mathematical techniques, machines can extract meaningful information from data to perform complex tasks and recognize patterns.

Deep learning is a type of machine learning that focuses on learning hierarchical representations of data. Deep learning algorithms, particularly artificial neural networks (ANNs), consist of interconnected nodes, called neurons or units, organized into layers. The most basic version of this consists of an input layer, followed by one or more hidden layers, and finally, an output layer (Figure 1-5). Each neuron in a layer is connected to every neuron in the subsequent layer, with each connection having an associated weight that determines the strength of the connection. These weights are adjusted iteratively through optimization algorithms such as the gradient descent [32].

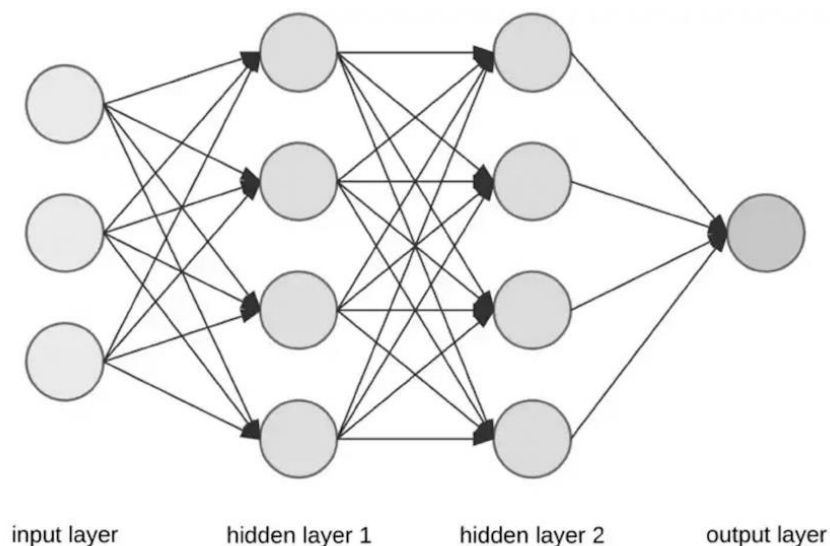


Figure 1-5: Diagram of a simple neural network with an input layer, two hidden layers and an output layer.

Gradient descent is a technique used to minimize the cost or loss function of a model. The cost function represents the relationship between the predicted output of the model and the actual output. By lowering the cost, the prediction becomes more accurate and aligns closely with the expected outcome. In simple convex functions, the goal is to reach the local minimum value of the cost function. To achieve this, an arbitrary starting point is chosen, and the steepest slope of the function is computed before a new point is determined through moving along this slope. However, in machine learning, functions are often more complex, with multiple local minima and maxima. To address this complexity, various optimization techniques have been devised to facilitate greater optimization efficiency, such as momentum, adaptive learning rates (e.g., Adam), and different variants of gradient descent (e.g., stochastic gradient descent). These techniques help accelerate convergence and improve the robustness of optimization in complex and high-dimensional optimization landscapes [32].

1.8.1 Convolutional Neural Network (CNN)

Convolutional Neural Networks, or CNNs, are a type of deep neural network that is exceptionally efficient at processing and analyzing visual data, including images and videos. They are highly effective in tasks such as image classification, object detection, and image segmentation [33]. CNNs have the unique ability to automatically and adaptively learn hierarchical feature representations directly from pixel values and their spatial organization, making them ideal for tasks that require an understanding of the spatial structure of the input data. A typical structure of CNN is illustrated in Figure 1-6. It consists of multiple layers, including convolutional layers, pooling layers, and fully connected layers, which work together to extract meaningful features from raw input data and utilize them to make accurate predictions [32].

Convolving an image at a specific pixel location entails summing up the products of the filter values with the corresponding neighbouring pixel values of the image. Mathematically, this process can be represented as:

$$y(i, j) = \sum_{u=-k}^k \sum_{v=-k}^k F(u, v) \cdot I(i - u, j - v) \quad (1-1)$$

where $y(i, j)$ denotes the output pixel value at location (i, j) in the resulting feature map. The summation iterates over the spatial extent of the filter, with u and v from $-k$ to k , where k represents half of the filter size. At each iteration, the product of the filter coefficient $F(u, v)$ and the corresponding pixel value from the input image $I(i - u, j - v)$ is computed and accumulated. The result of this operation is termed a feature extraction operation, which leads to a feature map. During training, the filters' weights are adjusted to achieve the correct label [32].

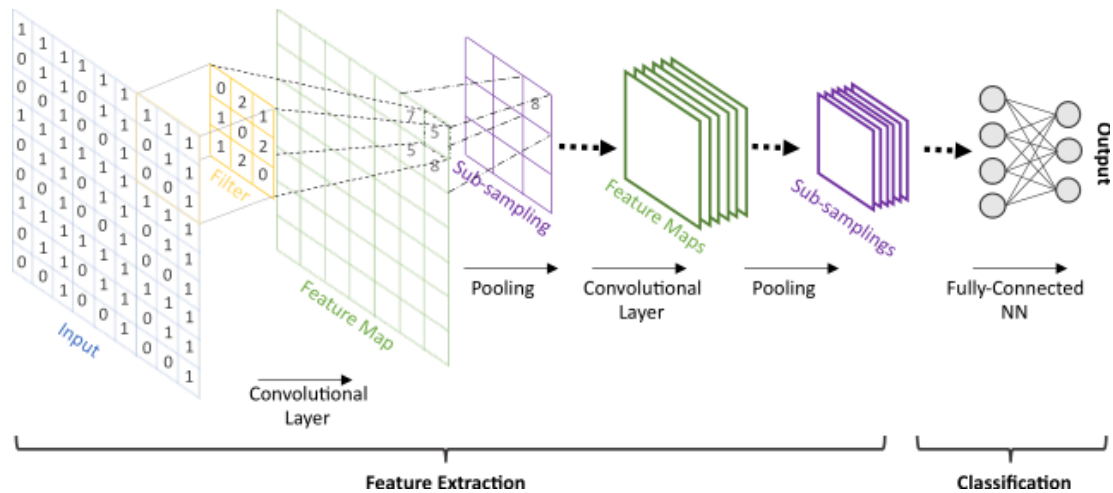


Figure 1-6: Typical CNN architecture [34]

1.8.2 Residual Network (ResNet)

When training deep learning models, gradients are computed for each weight with respect to the loss function to update the weights. However, due to the chain rule for derivative multiplication, gradients may approach zero if associated gradients are small, causing the vanishing gradient problem. This issue is particularly prevalent in the lower layers of deep models.

ResNet is a state-of-the-art architecture designed to tackle the vanishing gradient issues that arise in deep CNN networks [35]. ResNet incorporates skip connections, also known as "identity shortcut connections," which allow gradients to flow more freely during backpropagation without impeding the learning process Figure 1-7. This approach effectively mitigates the vanishing gradient problem, enabling the successful training of very deep networks without sacrificing performance.

ResNet has various depths, including 10, 18, 34, 50, and 110 layers, providing flexibility for different applications and computational resources.

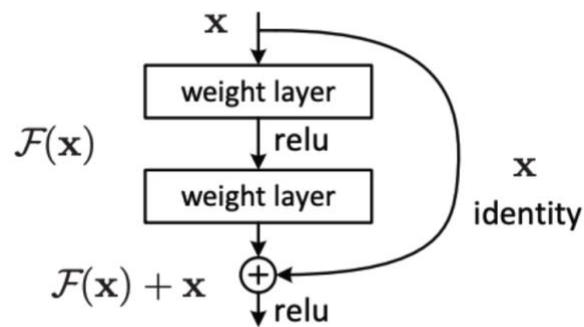


Figure 1-7: Residual learning block with the skip connection [35]

1.8.3 Model Hyperparameters

Hyperparameters play a crucial role in the configuration of machine learning models. They have a significant impact on the performance and behaviour of the model during both training and inference. These parameters are predetermined and not learned from the data.

Some of the key hyperparameters include the batch size, which is the number of training examples processed in each iteration. This parameter affects the speed and stability of training. Another important hyperparameter is the epoch, which refers to the number of passes through the training set during one training cycle [36].

The learning rate is a critical factor in determining the size of each step taken to update a model's parameters during optimization. Its value can be adjusted to control the speed and effectiveness of the gradient descent's movement towards the cost function's minimum. A higher learning rate results in more significant steps being taken, leading to faster

convergence, but it may cause the algorithm to overshoot or miss the optimum [37]. On the other hand, a lower learning rate allows for more precise convergence, but it may significantly increase training time. To achieve better performance, it is essential to tune this parameter by evaluating the model's loss after training and evaluation. Learning rate schedules can further refine the learning rate over time to optimize convergence. Additionally, the choice of optimizer (such as SGD, Adam, or RMSprop), which adjusts the network's weights during training, can also be fine-tuned for improved performance [37].

To prevent overfitting during training and enhance the model's ability to generalize to new data, different regularization techniques can be employed. The dropout rate is one such technique that can be used to regulate the fraction of neurons randomly dropped out during training [32], [36]. This encourages robust feature representations across different samples. Other regularization techniques such as L1 or L2 regularization can be employed to penalize large weights in the neural network. L1 regularization adds an absolute value penalty to the weights, while L2 regularization adds a squared value penalty. Both methods discourage the model from relying too heavily on any single feature, promoting simplicity and stability in the learned model [36]. Ultimately, these regularization techniques lead to better generalization performance of the model.

Architectural design choices are crucial in determining the performance and capabilities of a neural network. These choices include the number of layers, which determines the depth and complexity of the network architecture, and the number of neurons or filters in each layer. The selection of activation functions is also essential as they introduce non-linearity into the model's outputs, enabling the network to learn complex patterns and relationships within the data. Common activation functions are ReLU (Rectified Linear Unit), sigmoid, and tanh, each with its own characteristics and suitability for different types of data and tasks. These architectural design decisions significantly affect the network's capacity to learn and generalize from the data. Careful consideration of these factors is essential to achieve optimal performance in neural network applications.

Tuning these hyperparameters effectively is crucial to achieve optimal performance and generalization across various tasks and datasets in neural network applications.

1.8.3.1 Model Hyperparameters Optimization

Optimizing model hyperparameters can be achieved through various methods with unique strengths and applications. Grid search systematically explores a predefined grid of hyperparameter values, evaluating each combination to identify the best-performing one. On the other hand, random search randomly samples hyperparameter values from specified distributions, providing a simple and efficient option [38]. Bayesian optimization, which uses surrogate models and acquisition functions, selects hyperparameters that maximize performance through an iterative process, making it well-suited for expensive objective functions [39]. Hyperband, a bandit-based algorithm, combines random search with successive halving, effectively allocating resources to promising hyperparameter configurations [40]. Ultimately, the choice of optimization method depends on a range of factors, including the complexity of the search space, available computational resources, and desired optimization level, ensuring that the model achieves the best possible performance within the given constraints.

1.8.4 Data Partitioning and Evaluation Metrics

When working with machine learning models, it is important to divide the dataset into three parts: training, testing, and validation sets. The training set is used to fit the model, the testing set evaluates the model's performance, and the validation set helps tune the model's performance. Cross-validation is an effective technique for evaluating the model when the dataset is limited. It involves dividing the dataset into multiple subsets or folds, training the model on one subset, and evaluating it on the remaining data. The evaluation scores from each iteration are averaged to obtain a more reliable estimate of the model's performance.

When evaluating a machine learning model performance, a variety of metrics are employed to compare its predictions with the ground truth labels in test data. By examining true positives, false positives, true negatives, and false negatives, valuable insights into the model's overall effectiveness can be gained, which are often depicted through a confusion matrix (Figure 1-8).

		True Labels	
		Positive (1)	Negative (0)
Predicted Labels	Positive (1)	TP	FP
	Negative (0)	FN	TN

Figure 1-8: Illustration of Confusion Matrix.

In classification tasks, a True Positive (TP) occurs when the positive label is accurately predicted by the model, while a False Positive (FP) counts when the model predicts the positive class incorrectly. Conversely, a False Negative (FN) indicates that the model fails to predict the positive class, and a True Negative (TN) indicates correct negative predictions. The accuracy score provides a comprehensive assessment of the model's performance across all classes, and it can be calculated as:

$$accuracy = \frac{\text{number of correct predictions}}{\text{total number of predictions}} \quad (1-2)$$

Precision, recall, and F1 Score are other metrics for evaluating a classification task. Precision measures the accuracy of positive predictions by calculating the proportion of true positive predictions among all positive predictions made by the model. This metric emphasizes the quality of positive predictions and is sensitive to false positives, providing insight into the reliability of the model's positive predictions.

$$precision = \frac{TP}{TP + FP} \quad (1-3)$$

On the other hand, recall, also known as sensitivity or true positive rate, assesses the model's ability to capture all positive instances by calculating the rate of true positive predictions among all positive instances. Recall is sensitive to false negatives and highlights the model's ability to avoid missing positive instances.

$$recall = \frac{TP}{TP + FN} \quad (1-4)$$

The F1 Score is a single metric often used to evaluate model performance by combining precision and recall into a harmonic mean. This balanced measure considers both false positives and false negatives.

$$F1 \text{ Score} = \frac{2 * precision * recall}{precision + recall} \quad (1-5)$$

These metrics play crucial roles in assessing the overall effectiveness of classification models and understanding the trade-offs between false positives and false negatives.

1.8.5 Transfer Learning

Transfer learning is a powerful machine learning technique that leverages pre-trained models with access to large amounts of labeled data to expedite the training process for new tasks. The approach transfers knowledge from one task to another, enabling models to improve performance on smaller datasets. Generating large, labeled datasets can be a daunting and time-consuming process, making transfer learning particularly beneficial.

Pretrained models possess knowledge garnered from complex tasks and large datasets. Starting with a pre-trained model makes the training process for a new task faster and more efficient, as the model has already learned valuable features and representations. As a result, transfer learning assists models in better generalizing novel and unseen data.

State-of-the-art architectures, including Alex Net [41], VGG16 [42], ResNet34 [35], often provide robust starting points for transfer learning. Their ability to learn sophisticated representations on large datasets renders them outstanding candidates for fine-tuning on specific tasks, contributing to the success of transfer learning approaches.

1.9 Thesis Hypotheses and Objectives

COPD is a long-term and irreversible disease that is associated with increased morbidity and mortality due to exacerbations. To combat the progression of the disease and the risk of exacerbations, GOLD has developed staging systems based on three severity factors. Implementing these systems in the clinic requires evaluating each severity factor individually. We hypothesize that by applying neural network and convolutional neural network (CNN) algorithms to lung CT scans, we can quantify severity factors associated with COPD. This will result in the development of image-based tools that accurately assess COPD and facilitate treatment planning. Our research aims to achieve the following objectives:

- Overcoming the limitations of PFT in assessing lung function by employing quantitative CT measurements for lung function assessment.
- Improving the detection of COPD exacerbation risk and symptom severity by automatically identifying them from CT scans.
- Making COPD staging systems more accessible and reproducible by providing an automatic and accurate assessment of patient COPD severity based on the high-resolution eight staging system and GOLD's 2023 guidelines.

1.10 Thesis Outline

This thesis is structured into five chapters. The current chapter serves as an introduction, providing relevant background information. The subsequent three chapters delve deeply into the research methodology employed to achieve the outlined objectives. In conclusion,

the final chapter offers a comprehensive summary of the findings, accompanied by conclusive remarks.

1.10.1 Chapter 2: Neural Network for COPD Eight Staging System Based on Features Extracted from Inspiratory/ Expiratory Lung 3D CT

In this chapter, we employed lung CT volumes at the inspiration and expiration breathing phases to extract features representing lung pathologies, lung air volume and distribution within the lung. These extracted features formed the basis for constructing two distinct neural network models tailored to the eight-stage COPD classification system. In the first model, these features were leveraged to predict the three severity factors and subsequently classify COPD subjects into the respective eight stages. Conversely, in the second neural network model, the extracted features primarily contributed to predicting lung function, while symptom level and exacerbation frequency were provided as additional inputs.

1.10.2 Chapter 3: CNN with a Single Lung 3D CT Volume for COPD Eight Staging System

In this chapter, advanced deep convolutional neural network (CNN) techniques were utilized to extract features from a single lung CT volume at the expiration phase for an accurate eight-stage classification of COPD. The approach also utilized transfer learning by incorporating pre-trained ResNet architectures with depths of 18 and 34 for the eight-stage COPD classification task.

Additionally, a separate CNN model was developed for predicting lung function, which used symptom level and exacerbation frequency as supplementary inputs to classify COPD into the eight stages.

1.10.3 Chapter 4: CNN with a Single Lung 3D CT Volume for COPD Classification Based on The GOLD 2023 Staging System

This chapter employed convolutional neural networks (CNNs) with single lung CT scans taken at the expiration phase to classify COPD patients according to the GOLD 2023 staging system. The CNNs were utilized to predict symptom severity and exacerbation risk independently, employing both training from scratch and transfer learning approaches. Subsequently, the models developed for the symptom severity and exacerbation risk were integrated to develop an algorithm for the GOLD 2023 staging system.

1.10.4 Chapter 5: Summary, Conclusion, and Future Work

In this chapter, we summarize the findings presented in the thesis and critically examine the limitations of the proposed algorithms. Additionally, we explore potential future research directions and offer suggestions for further studies in this area. Finally, this section provides a concluding remark to encapsulate the key insights gained from the dissertation.

1.11 References

- [1] World Health Organization WHO, “Chronic obstructive pulmonary disease (COPD).” Accessed: Mar. 07, 2024. [Online]. Available: [https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-\(copd\)](https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd))
- [2] John Elflein, “Percentage of Canadians who reported being diagnosed with chronic obstructive pulmonary disease (COPD) from 2003 to 2020 [Graph],” Feb. 2024.
- [3] Canadian Centre for Health Information., Statistics Canada. Health Statistics Division., and Gibson Library Connections, *Health indicators 2008*. Statistics Canada, Health Statistics Division, 2008.
- [4] J. J. Soler-Cataluña, M. Á. Martínez-García, P. Román Sánchez, E. Salcedo, M. Navarro, and R. Ochando, “Severe acute exacerbations and mortality in patients

- with chronic obstructive pulmonary disease,” *Thorax*, vol. 60, no. 11, pp. 925–931, Nov. 2005, doi: 10.1136/thx.2005.040527.
- [5] “Global Initiative For Chronic Obstructive Lung Disease Global Strategy For The Diagnosis, Management, And Prevention Of Chronic Obstructive Pulmonary Disease (2023 REPORT),” 2022. [Online]. Available: www.goldcopd.org
- [6] C. F. Vogelmeier *et al.*, “Global strategy for the diagnosis, management, and prevention of chronic obstructive lung disease 2017 report,” *American Journal of Respiratory and Critical Care Medicine*, vol. 195, no. 5. American Thoracic Society, pp. 557–582, Mar. 01, 2017. doi: 10.1164/rccm.201701-0218PP.
- [7] E. Pompe *et al.*, “Five-year progression of emphysema and air trapping at ct in smokers with and those without chronic obstructive pulmonary disease: Results from the COPDGene study,” *Radiology*, vol. 295, no. 1, pp. 218–226, 2020, doi: 10.1148/radiol.2020191429.
- [8] John B. West and Andrew Luks, *Respiratory physiology : the essentials*, 10th ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2016.
- [9] D. Raghavan and R. Jain, “Increasing awareness of sex differences in airway diseases,” *Respirology*, vol. 21, no. 3. pp. 449–459, Apr. 01, 2016. doi: 10.1111/resp.12702.
- [10] J. Y. So *et al.*, “Population Decline in COPD Admissions During the COVID-19 Pandemic Associated with Lower Burden of Community Respiratory Viral Infections,” *American Journal of Medicine*, vol. 134, no. 10, pp. 1252-1259.e3, Oct. 2021, doi: 10.1016/j.amjmed.2021.05.008.
- [11] C. J. Galbán *et al.*, “Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression,” *Nat Med*, vol. 18, no. 11, pp. 1711–1715, Nov. 2012, doi: 10.1038/nm.2971.
- [12] P. A. Grenier, “Emphysema at CT in smokers with normal spirometry: Why it is clinically significant,” *Radiology*, vol. 296, no. 3. Radiological Society of North America Inc., pp. 650–651, Sep. 01, 2020. doi: 10.1148/radiol.2020202576.

- [13] M. Kirby *et al.*, “Total airway count on computed tomography and the risk of chronic obstructive pulmonary disease progression,” *Am J Respir Crit Care Med*, vol. 197, no. 1, pp. 56–65, Jan. 2018, doi: 10.1164/rccm.201704-0692OC.
- [14] S. Matsuoka *et al.*, “Quantitative CT Measurement of Cross-sectional Area of Small Pulmonary Vessel in COPD. Correlations with Emphysema and Airflow Limitation,” *Acad Radiol*, vol. 17, no. 1, pp. 93–99, Jan. 2010, doi: 10.1016/j.acra.2009.07.022.
- [15] “Global Initiative for Chronic Obstructive Lung Disease Global Initiative for Chronic Obstructive Lung Disease POCKET GUIDE TO COPD DIAGNOSIS, MANAGEMENT, AND PREVENTION A Guide for Health Care Professionals,” 2017. [Online]. Available: www.goldcopd.org
- [16] A. Agustí *et al.*, “Clinical and prognostic heterogeneity of C and D GOLD groups,” *European Respiratory Journal*, vol. 46, no. 1. European Respiratory Society, pp. 250–254, Jul. 01, 2015. doi: 10.1183/09031936.00012215.
- [17] M. L. K. Han *et al.*, “GOLD 2011 disease severity classification in COPDGene: A prospective cohort study,” *Lancet Respir Med*, vol. 1, no. 1, pp. 43–50, Mar. 2013, doi: 10.1016/S2213-2600(12)70044-9.
- [18] E. A. Regan *et al.*, “Genetic epidemiology of COPD (COPDGene) study design,” *COPD: Journal of Chronic Obstructive Pulmonary Disease*, vol. 7, no. 1, pp. 32–43, 2010, doi: 10.3109/15412550903499522.
- [19] M. Cazzola and M. Molimard, “The scientific rationale for combining long-acting β 2-agonists and muscarinic antagonists in COPD,” *Pulmonary Pharmacology and Therapeutics*, vol. 23, no. 4. pp. 257–267, Aug. 2010. doi: 10.1016/j.pupt.2010.03.003.
- [20] N. Gross, D. Tashkin, R. Miller, J. Oren, W. Coleman, and S. Linberg, “Inhalation by Nebulization of Albuterol-Ipratropium Combination (Dey Combination) Is Superior to Either Agent Alone in the Treatment of Chronic Obstructive Pulmonary Disease,” 1998. [Online]. Available: www.karger.com/http://BioMedNet.com/karger

- [21] D. A. Mahler *et al.*, “FLIGHT1 and FLIGHT2: Efficacy and safety of QVA149 (indacaterol/glycopyrrolate) versus its monocomponents and placebo in patients with chronic obstructive pulmonary disease,” *Am J Respir Crit Care Med*, vol. 192, no. 9, pp. 1068–1079, Nov. 2015, doi: 10.1164/rccm.201505-1048OC.
- [22] J. A. et al. Wedzicha, “Analysis of chronic obstructive pulmonary disease exacerbations with the dual bronchodilator QVA149 compared with glycopyrronium and tiotropium (SPARK): a randomised, double-blind, parallel-group study,” *Lancet Respir Med*, vol. 1, no. 3, pp. 199–209.
- [23] D. A. Lipson *et al.*, “Once-Daily Single-Inhaler Triple versus Dual Therapy in Patients with COPD,” *New England Journal of Medicine*, vol. 378, no. 18, pp. 1671–1680, May 2018, doi: 10.1056/nejmoa1713901.
- [24] S. et al. Suissa, “Comparative Effectiveness and Safety of LABA-LAMA vs LABA-ICS Treatment of COPD in Real-World Clinical Practice,” *CHEST*, Volume 155, Issue 6, 1158 - 1165, vol. 155, no. 6, pp. 1158–1165.
- [25] S. M. Ryan, B. Vestal, L. A. Maier, N. E. Carlson, and J. Muschelli, “Template Creation for High-Resolution Computed Tomography Scans of the Lung in R Software,” *Acad Radiol*, vol. 27, no. 8, pp. e204–e215, Aug. 2020, doi: 10.1016/j.acra.2019.10.030.
- [26] O. and F. P. and B. T. Ronneberger, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and Computer-Assisted Intervention--MICCAI 2015: 18th International Conference*, Munich, German: Springer, Oct. 2015, pp. 234--241.
- [27] B. Zitová and J. Flusser, “Image registration methods: A survey,” *Image Vis Comput*, vol. 21, no. 11, pp. 977–1000, 2003, doi: 10.1016/S0262-8856(03)00137-9.
- [28] J. B. A. Maintz and M. A. Viergever, “A survey of medical image registration,” 1998.

- [29] E. Castillo, R. Castillo, J. Martinez, M. Shenoy, and T. Guerrero, “Four-dimensional deformable image registration using trajectory modeling,” *Phys Med Biol*, vol. 55, no. 1, pp. 305–327, Jan. 2010, doi: 10.1088/0031-9155/55/1/018.
- [30] M. Ibanez. Johnson, *ITK Software Guide: Design and Functionality*, Fourth. Kitware Inc, 2015.
- [31] R. Szeliski and J. Coughlan, “Spline-Based Image Registration,” Kluwer Academic Publishers, 1997.
- [32] Ethem Alpaydin, *Introduction to Machine Learning (4th ed.)*, 4th ed. MIT Press., 2020.
- [33] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, “Object Detection with Deep Learning: A Review,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11. Institute of Electrical and Electronics Engineers Inc., pp. 3212–3232, Nov. 01, 2019. doi: 10.1109/TNNLS.2018.2876865.
- [34] B. Zaparoli Cunha, C. Droz, A. M. Zine, S. Foulard, and M. Ichchou, “A review of machine learning methods applied to structural dynamics and vibroacoustic,” *Mech Syst Signal Process*, vol. 200, p. 110535, Oct. 2023, doi: 10.1016/J.YMSSP.2023.110535.
- [35] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition.” [Online]. Available: <http://image-net.org/challenges/LSVRC/2015/>
- [36] “Deep Learning.” Accessed: Mar. 02, 2024. [Online]. Available: <https://mitpress.mit.edu/9780262035613/deep-learning/>
- [37] Yu. S. Fedorenko, “An overview of gradient descent optimization algorithms,” *Vestnik komp iuternykh i informatsionnykh tekhnologii*, no. 186, pp. 10–17, Dec. 2016, doi: 10.14489/VKIT.2019.12.PP.010-017.
- [38] J. Bergstra and Y. Bengio, “Random Search for Hyper-Parameter Optimization,” *Journal of machine learning research*, 2012, doi: 10.5555/2503308.2188395.
- [39] J. Snoek, H. Larochelle, and R. P. Adams, “Practical Bayesian optimization of machine learning algorithms.” pp. 2951–2959, 2012. Accessed: Mar. 02, 2024.

[Online]. Available: <https://collaborate.princeton.edu/en/publications/practical-bayesian-optimization-of-machine-learning-algorithms>

- [40] L. Li, K. G. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, “Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization,” *Journal of machine learning research*, 2016.
- [41] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks.” [Online]. Available: <http://code.google.com/p/cuda-convnet/>
- [42] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Sep. 2014, [Online]. Available: <http://arxiv.org/abs/1409.1556>

Chapter 2

2 « Neural Network for COPD Eight Staging System Based on Features Extracted from Inspiratory/ Expiratory Lung 3D CT »

A draft of this chapter has been prepared for peer review and potential publication in the Expert Systems with Applications journal.

2.1 Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a chronic lung condition that causes persistent inflammation and obstruction of airflow, resulting in difficulty breathing. COPD contributes to a significant number of deaths globally each year. According to the World Health Organization (WHO), COPD was the third leading cause of death worldwide in 2019, accounting for approximately 3.23 million deaths, or 5.7% of all deaths [1]

There is a growing interest in improving diagnostic techniques for COPD, given its high mortality rate. The primary goals of COPD assessment include determining the extent of airflow limitation, its impact on the patient's health, and the risk of future exacerbations. To assess the severity of airflow limitation, healthcare providers use a spirometry test - also known as PFT - which measures airflow during breathing. The severity of the disease is then categorized based on the results of the PFT, using the guidelines established by the Global Initiative for Chronic Lung Disease (GOLD) [2]. The GOLD staging system classifies the severity of chronic obstructive pulmonary disease (COPD) into four stages based on the ratio of forced expiratory volume in 1 second (FEV1) and forced vital capacity (FVC). If the FEV1/FVC ratio is less than 70%, it indicates airflow limitation, and the four GOLD stages of mild, moderate, severe, and very severe are applicable, with FEV1 values progressively decreasing with each stage. However, this staging system only measures the severity of airflow limitation in the entire lung and doesn't provide the necessary critical information for determining the appropriate treatment course to manage the disease or prevent its progression.

In 2011, the GOLD committee proposed a new classification system called ABCD for COPD [3]. This system is more reliable in assessing disease severity by taking into account the level of symptoms and exacerbation history in addition to the lung function assessment data. The symptom level can be measured by using either the modified Medical Research Council (mMRC) dyspnea score (Appendix B), the COPD Assessment Test (CAT) score or the St. George's Respiratory Questionnaire (SGRQ)(Appendix A) [2]. Exacerbation history is a measure of how often a patient requires assessment in the Emergency Department or hospital admission in the previous year. This information, along with the patient's symptoms, can help clinicians develop a personalized treatment plan. According to Lange et al., using the ABCD GOLD classification is a better predictor of future exacerbations than the PFT assessment only. This suggests that symptoms and exacerbation history play a significant role in the progression of COPD [4].

Utilizing the above-mentioned three factors of COPD assessment, an eight-stage COPD system was developed in the COPDGene study [5], [6], which aims to provide essential information for guiding therapy decisions. In this system, COPD subjects are categorized into eight classes based on symptoms and risk, with risk determined by exacerbation history and airflow limitation measured by PFT ($FEV_1\% < 50$). The eight stages are defined as follows: A, low symptoms/low risk; B, high symptoms/low risk; C, low symptoms/high risk; and D, high symptoms/high risk. Classes C and D are further subdivided based on the cause of high risk: C1 and D1 have high risk due to lung function only ($FEV_1\% < 50$); C2 and D2 meet exacerbation criteria only; and C3 and D3 meet both exacerbation and FEV_1 criteria.

Several studies have investigated the necessity of subgrouping C/D patients according to factors contributing to risk elevation [7], [8]. Augsti et al. reported that different therapeutic options must be considered for the sub-groups within C and D [9]. Depending on the severity of the disease, different types of bronchodilators are utilized for treatment. For Class A individuals with low symptoms, short-acting bronchodilators are recommended, while those with high symptoms require long-acting bronchodilators. Patients at high risk of lung function decline and/or exacerbation require combinations of two or three types of long-acting bronchodilators, depending on the underlying cause of risk.

Quantitative assessment of COPD using lung CT plays a crucial role in various aspects of disease management. These measures aid in diagnosing COPD by providing objective and detailed information about lung structure and function. Additionally, they help assess disease severity by quantifying the extent of emphysema, airway remodelling, and lung parenchymal abnormalities. Moreover, lung CT allows for monitoring disease progression over time, enabling clinicians to track changes in lung morphology and density.

Various quantitative methods that rely on lung CT data have been developed to improve the assessment of lung diseases. These image-based techniques are more reliable in quantifying lung pathologies such as emphysema, airway disease, and air trapping. The presence of emphysema can also predict the progression of both physiological and structural disease [10]. Furthermore, it is possible to detect disease progression through quantitative CT measurements even before any noticeable changes are detected with PFT. In a study by Pompe et al. [11], longitudinal changes in emphysema and air-trapping were investigated by measuring them quantitatively from lung CT and FEV1 in cigarette smoker subjects with and without COPD. The results showed that emphysema and air-trapping significantly increased during the 5-year follow-up in both smokers with and without COPD. The study also found that the progression of emphysema and air-trapping was partly correlated with FEV1. However, the authors concluded that a significant proportion of emphysema and air-trapping progression was not correlated with FEV1. Another study was conducted on smokers who were part of the COPDGene study and had normal spirometry readings, with $FEV1/FVC > 70\%$, which is labeled as GOLD stage 0. The study investigated these subjects after a 5-year follow-up period and found that they had developed progressive airflow obstruction ($FEV1/FVC$, $P = .005$) and showed greater progression in quantitative emphysema [10].

By integrating image data with powerful machine learning (ML) approaches, the quantitative assessment of COPD and its accuracy of diagnosis can be significantly improved. Various ML-based methods have been developed in recent years for disease diagnosis, which support physicians in making accurate diagnoses and devising effective treatment plans. Nimura et al. [12] have developed a classification model that assesses COPD severity based on six features extracted from CT image analysis in addition to PFT features. Recently, a more effective classification model was proposed to assess COPD

severity [13]. The model uses features extracted from the lung air volume variation and distribution throughout the respiration cycle. A highly accurate lung air segmentation technique was used to segment lung air volume using end-inhale and end-exhale phases [14]. Ho et al. developed a model that extracts two 3D volumes from 4D-CT data to create a 3D parametric response map (PRM) [15]. The PRM categorizes the lung voxels into three groups: emphysema, functional small-airway disease (fSAD), and healthy tissue. These PRM maps are then used to train a convolutional neural network to identify the presence of COPD.

Our approach in this investigation utilizes a Neural Network approach to accurately classify COPD, based on the eight-stage COPD system. This was achieved by training our models on features extracted from lung CT data, namely lung air volume features and COPD phenotype features. The efficacy of such COPD classification and staging systems has been demonstrated by their high accuracy in assessing COPD severity using the high-resolution COPD staging scheme. These systems provide valuable diagnostic information that can aid in the planning of appropriate treatments.

2.2 Materials & Methods

2.2.1 Dataset

The NIH COPDGene study is a vast research project that includes a dataset which was used for this project [6]. The study involves 21 clinical centers and using different multi-detector CT scanners. It aims to investigate COPD, which is one of the most common lung diseases. The dataset used in our investigation includes data from 10,192 participants with COPD from both genders and with age ranging from 45 to 80 years. It contains volumetric lung CT image data, captured during both inhalation and exhalation phases, with voxel resolutions ranging from $(512 \times 512 \times 100)$ to $(512 \times 512 \times 736)$.

The COPDGene study labels COPD subjects according to the Eight COPD Staging System. The distribution of subjects among these classes is detailed in Table 2-1. It is worth noting that the dataset has class imbalance, particularly for classes C2 and C3. To address this

issue, we utilized paired lung CT scans (inspiration and expiration) from 320, ensuring a balanced distribution of 40 subjects per class.

To enrich the data in classes C2 and C3, we used a novel augmentation method based on deformable image registration.

Table 2-1: Distribution of COPD subjects among the Eight classes

A	B	C1	C2	C3	D1	D2	D3
1425	1118	174	38	10	1192	254	410

2.2.2 Lung CT data preprocessing

This approach involves image preprocessing steps to extract morphological features from thoracic CT data. It includes image segmentation to segment the lung region, image registration to spatially align the paired inhale and exhale scans, in addition to other image processing techniques to extract the features.

2.2.2.1 Lung Volume Segmentation

To extract lung volumes, we used a fully automated image segmentation method [16]. In this method, the left and right lung volumes are extracted from lung CT using a combination of thresholding and region-growing segmentation methodology. The lung is first detected from the original CT scan by Hounsfield unit (HU) thresholding (lung < -300 HU). Then, large airways are detected by identifying region near their mid-line that has lower HU than normal lung tissue (-500 HU). Last, a region growing is applied to exclude the large airways from the lung. Figure 2-1 shows an example slice of a segmented lung from a coronal view.

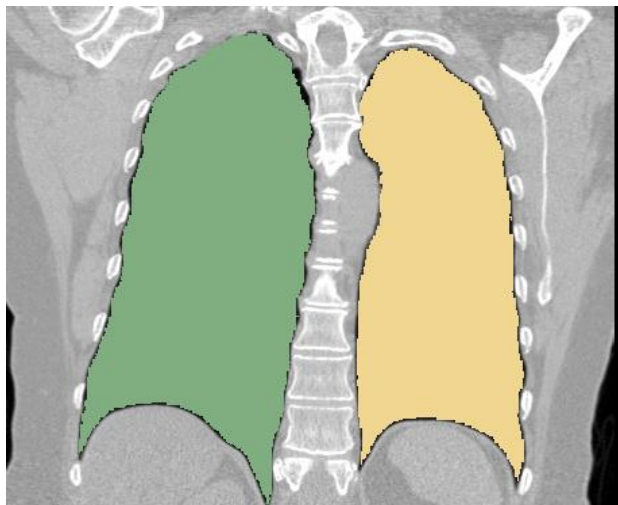


Figure 2-1: An example of segmented lung mask from a coronal view

2.2.2.2 Lung Volume Image Registration

Image registration is the process of aligning two or more images or volumes of the same scene captured at different times/conditions, from different viewpoints or using different modalities. In this project, we used Free Form Deformation (FFD) to apply registration on lung volumes.

In FFD image registration, a deformation field is used to warp or deform one image to match the other image. This deformation field is represented by a grid of control points, and each control point has associated displacement vectors that define how much the point should move in the x, y, and z directions. We optimized the displacement vectors of the control points to minimize the difference between the two images using mutual information as a similarity metric. By iteratively adjusting the displacement vectors, the algorithm deforms one volume until it aligns with the other volume as closely as possible as illustrated in Figure 2-2. ITK, an open-source library with extensive tools for medical image analysis, was used for lung volume registration [17].

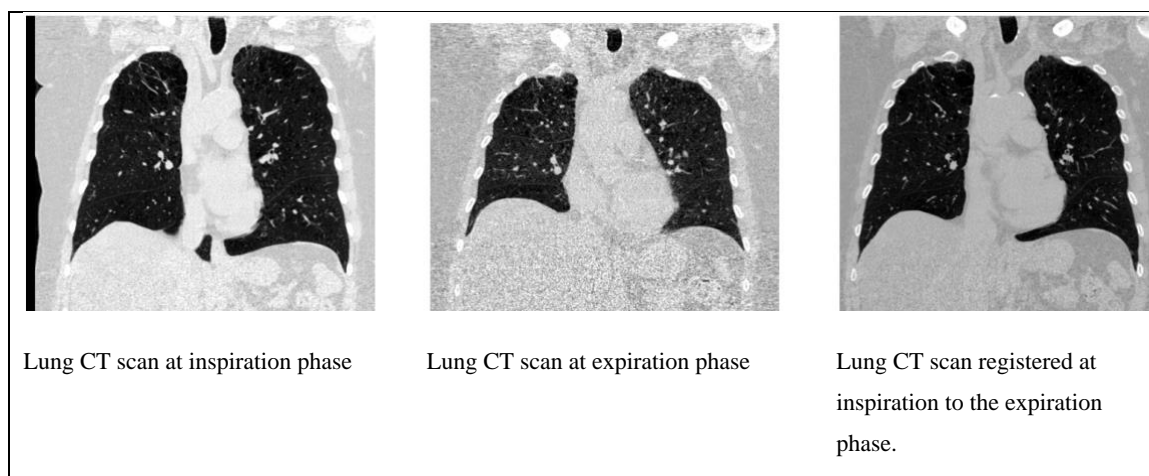


Figure 2-2: An example of a lung CT scan at inspiration, expiration, and the registered image at inspiration to expiration.

2.2.2.3 Lung Volume Image Augmentation

We have developed a new method to augment lung CT images that involves two main steps: image deformable registration and image warping. To align the inspiration scan with the expiration one and vice versa, we used Free Form Deformation (FFD) for image registration. We then extracted the resulting transformation matrices that align the paired scans. Subsequently, three additional scans, representing scans between the two inspiration and expiration phases, were generated from each original scan (moving image) by applying image warping with different fractions of the obtained transformation matrix. For the inspiration scan, we used the transformation matrix that aligns it as a moving image with the expiration scan. We applied image warping at 10%, 15%, and 20% of the transformation matrix to produce three new inhale scans. Similarly, for the exhale scan, we generated three new scans using the transformation matrix that aligns the expiration scan as a moving image to the inspiration counterpart. We only applied this image augmentation step to data samples of classes C2 and C3 to ensure that the total number of samples in these classes reached 40.

Figure 2-3 shows two samples from the augmentation results. Although visually, the difference between the original volume and the corresponding augmented one may not seem substantial, the change in the internal structure of the image is significant. For instance, the ratio of damaged volume to the total volume of the lung show measurable

increase or decrease. Therefore, extracted features exhibit measurable difference between original and augmented volumes.

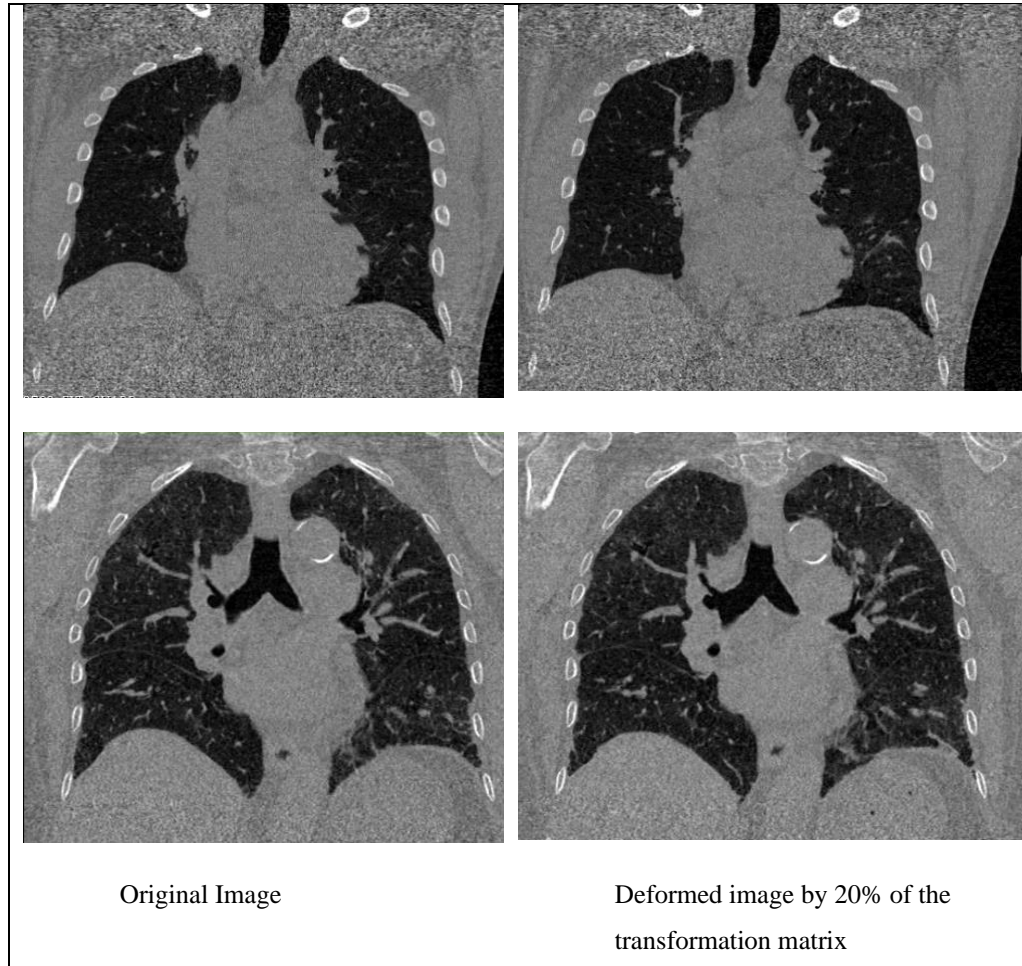


Figure 2-3: Examples of coronal views of original and corresponding lung CT scans.

2.2.3 Feature Extraction

To train the NN model, two sets of imaging features were used. The first set consisted of lung air volume features developed by Moghadas et al. [13]. The second set consisted of a set of COPD phenotypes data which included emphysema, air trapping, and fSAD measured from PRM[15]. These two sets of features were derived and combined to train the NN models.

2.2.3.1 Lung Air Features

Lung air features were extracted by segmenting lung air volumes using a highly accurate lung air segmentation technique [14]. In this technique, first, the image intensity of pure air and lung tissue are calculated for each pair of scans. This is performed by computing the histogram difference between the inspiration and expiration volumes, resulting in the initial lower threshold A and upper threshold B (refer to Figure 2-4). The lower threshold separates background air from lung air/tissue, while the upper threshold distinguishes between lung air/tissue and lung tissue only.

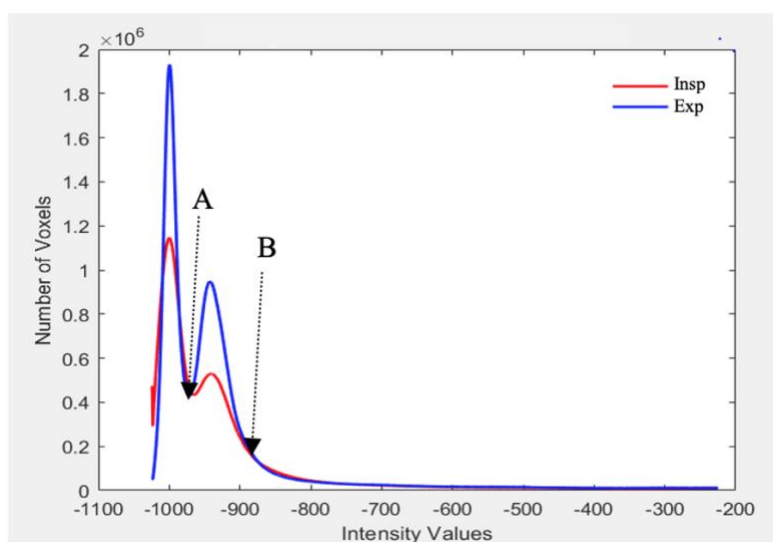


Figure 2-4: Inspiration and expiration histograms and the initial thresholds A and B in the lung air segmentation technique.

The two thresholds are further optimized in order to obtain accurate thresholds. The lower threshold is optimized based on air mass conservation between background air and lung air, while the upper threshold is optimized based on tissue incompressibility. This optimization process ensures that the thresholds delineate different tissue components in the lung scan accurately before obtaining the image intensity of pure air and pure tissue. These air and tissue intensities are then utilized to calculate the air volume portion coefficients for each voxel which, together, construct the lung air map (Figure 2-5)

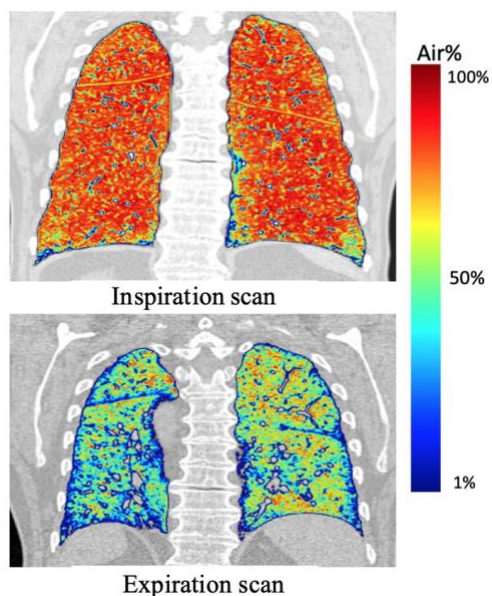


Figure 2-5: An example of a lung air map for inspiration and expiration scans.

Three groups of features were extracted from lung air maps of inspiration and expiration scans. These groups include features related to lung volume variation, inspiration and expiration air distribution, and overall air distribution in the lung. The first group, lung volume variation features, was calculated based on the difference in lung air volume between inspiration and expiration images in relation to the expiration volume (V_{exh}) or inspiration volume (V_{inh}).

The second group of features, inspiration and expiration air distribution features, was calculated as the air percentage per voxel. It includes seven percentages of air in voxels, which are 95%, 90%, 85%, 80%, 75%, 70%, and 65%. High concentrations of air in voxels from inspiration scans are related to emphysema, while high concentrations of air in voxels from expiration scans are related to air-trapping.

Lastly, the third group of features pertained to the overall air distribution in the lung. This group included the mean and standard deviation for each of the inspiration (MV_{inh} , SDV_{inh}) and expiration (MV_{exh} , SDV_{exh}) air distribution features.

2.2.3.2 COPD Phenotyping Features

COPD encompasses a range of lung pathologies, known as phenotypes, each with distinct characteristics. Emphysema is characterized by damage to the lung's air sac

walls, leading to reduced elasticity and impaired gas exchange. Air trapping occurs when air becomes trapped in the lungs during expiration, contributing to hyperinflation and airflow limitation. Chronic bronchitis involves inflammation and narrowing of the airways due to chronic irritation, resulting in excessive mucus production and persistent coughing. Identifying the specific phenotype of COPD present in a patient is crucial to tailor treatment strategies and optimize management approaches.

To assess emphysema and air trapping within the lung, we employed thresholding based on Hounsfield Units (HU). For the inspiration scan, a threshold value of -950 HU was applied to quantify the extent of emphysema present in the lung tissue [18]. Conversely, for the expiration scan, a threshold value of -856 HU was utilized to evaluate the degree of air trapping within the lung [18]. These threshold values were chosen to accurately delineate between healthy lung tissue and areas affected by emphysema or air trapping, providing valuable insights into lung pathology associated with COPD.

We have also utilized a technique called Parametric Response Mapping (PRM) [15] to extract additional phenotype features. PRM is used to evaluate the impact of airway destruction in the lungs by analyzing air trapping in expiration scans. First, lung volumes at both the inspiration and expiration phases were segmented. Next, the Free Form Deformation (FFD) registration was applied to align the expiration and inspiration scans. Then, thresholding was applied to the joint histogram formed using all voxel pairs within the registered inspiration-expiration lungs. This process resulted in three distinct categories: healthy lung tissue, designated by green; functional small airway disease (fSAD), represented by yellow; and emphysema, indicated by red. Figure 2-6 depicts two examples of voxel distribution, where the concentration of voxels forms an elongated elliptical pattern. The voxels are highly concentrated in the center and coded by red color, while decreasing towards the periphery, as shown in blue. The first example shown in Figure 2-6a was generated for a healthy subject, where the voxels are primarily located in green areas. The second example shown in Figure 2-6b was generated for a COPD patient with a severe lung pathology measured with PFT ($FEV_1 = 18$).

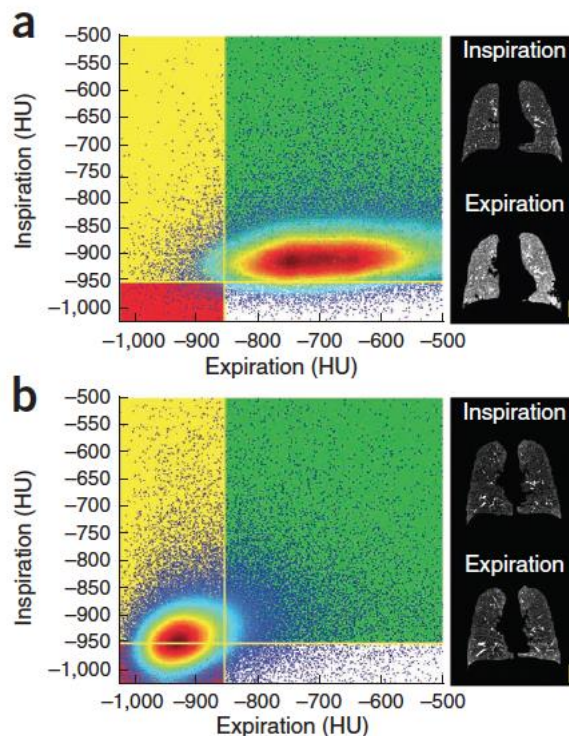


Figure 2-6: The distribution of voxels with varying values at inspiration and expiration for an individual with normal lung function (a), and another with severe lung function, FEV1 = 18 (b) [11].

2.2.4 NN Model Training

Two neural network models were considered in this project: the NN-CT and NN_Hybrid models. The NN-CT model was trained using only lung CT features, while the NN_Hybrid model incorporated the level of symptoms and the frequency of COPD exacerbation.

The NN-CT model architecture consisted of six layers: the input layer, four hidden layers, and the output layer. Two dropout layers were included to prevent overfitting. Hyperparameters such as activation function, batch size, optimizer, epochs, and patience for Early Stopping were optimized using the Bayesian optimization algorithm. The ReLU activation function and Categorical Cross Entropy loss function were employed. The model was trained with the Adam optimizer for 250 epochs and a batch size of 32.

On the other hand, the NN_Hybrid model comprised five layers: the input layer, three hidden layers, and the output layer. A dropout layer was added to mitigate overfitting. Similarly, hyperparameters were optimized using Bayesian optimization. The Sigmoid

activation function and Categorical Cross Entropy loss function were utilized. The model was trained with the Adam optimizer for 199 epochs and a batch size of 16.

Both models were trained to perform multiclassification of the eight-stage COPD using extracted features from 320 subjects, including lung air features and phenotype features. The NN_Hybrid model included two additional inputs: exacerbation frequency and symptom level. The dataset was balanced, with 40 subjects used for each class.

2.3 Results

2.3.1 Features Cross-Correlation Coefficients.

The effectiveness of the extracted features was evaluated by calculating their correlation with PFT measurements. Table 2-2 displays the correlation coefficients between PFT measurements and the extracted lung air features. Notably, features such as Vexh80, Vexh75, Vexh70, Vexh65, Vin, MVexh, and Vinh90 exhibit high correlation values with PFT features.

Furthermore, the correlation coefficients between phenotypic features and PFT measurements are summarized in Table 2-3. The results indicate that the air-trapping feature demonstrates a greater correlation with both FEV1 and FEV1/FVC (0.70 and 0.65, respectively). Additionally, the two emphysema features derived from inhalation and PRM mapping exhibit a high correlation with PFT measurements.

Table 2-2: Correlation coefficients between the lungs' air features and corresponding PFT measurements.

Exhalation features	FEV1/FVC	FEV1	Inhalation features	FEV1/FVC	FEV1
Vexh	0.63	0.72	Vinh	0.65	0.62
V95exh	0.60	0.72	V95inh	0.58	0.54
V90exh	0.68	0.71	V90inh	0.64	0.56
V85exh	0.74	0.71	V85inh	0.59	0.48

V80exh	0.78	0.71	V80inh	0.52	0.40
V75exh	0.81	0.72	V75inh	0.46	0.34
V70exh	0.83	0.72	V70inh	0.42	0.30
V65exh	0.83	0.71	V65inh	0.39	0.26
MVexh	0.81	0.72	MVinh	0.58	0.47
SDVexh	0.76	0.64	SDVinh	0.33	0.37

Table 2-3: Correlation coefficients between phenotype features and PFT measurements.

Features	Emphysema	Air Trapping	PRM Emphysema	PRM fSAD
FEV1/FVC	0.68	0.70	0.64	0.59
FEV1	0.62	0.65	0.60	0.52

2.3.2 Neural Network Models Evaluation

The neural network models were evaluated with Stratified fold cross-validation with $K = 7$. This method was used to preserve the class distribution in the training and test sets. The performance of the model was evaluated in terms of Accuracy, Precision, Recall, and F1 score for each fold. The mean of the 7 folds was calculated to present the overall performance of the model.

Figure 2-7 illustrates the results obtained for the NN-CT model. The model achieved an accuracy of 0.63, with an average calculated from 7 folds. The accuracy scores varied from 0.54 to 0.76 among the folds. The mean precision was determined to be 0.68, ranging from 0.57 to 0.81. The average recall and F1 score were 0.63 and 0.62, respectively. The recall and F1 score consolidate the accuracy results, with a recall range from 0.54 to 0.78 and an F1-score range from 0.49 to 0.77. The confusion matrix for the NN-CT is presented in Figure 2-9-a, indicating that most misclassifications occur among neighboring classes, i.e., classes that share one or two severity factors in common.

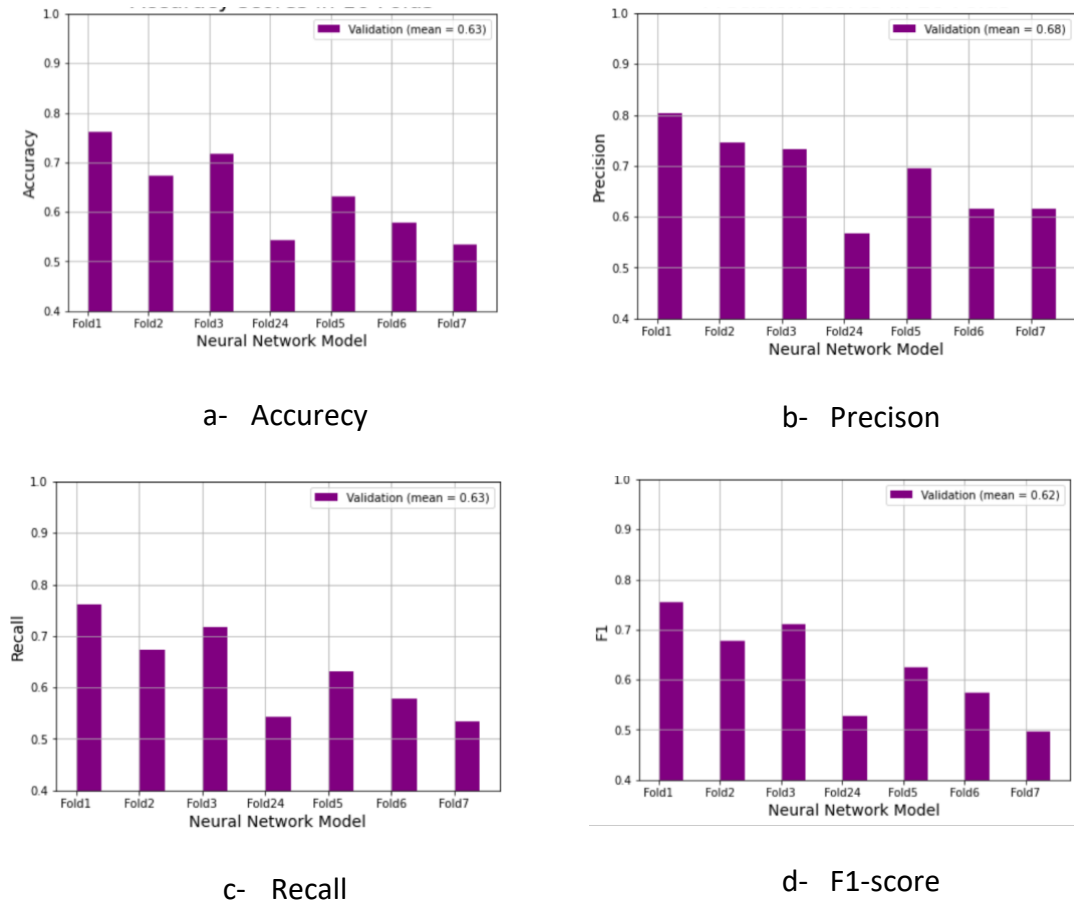
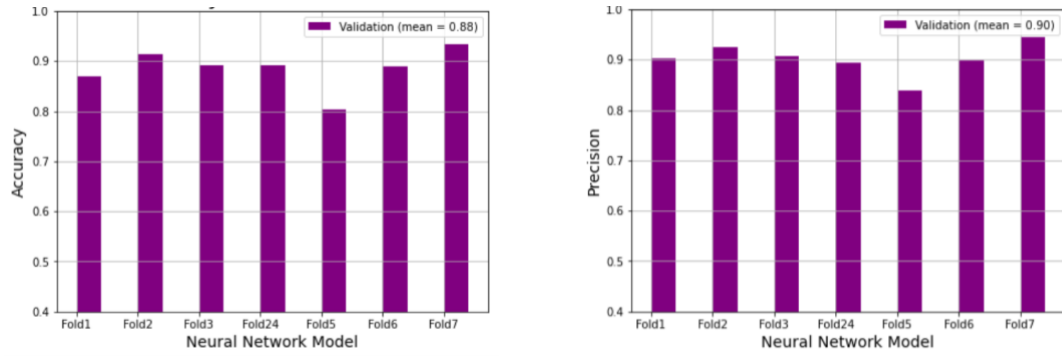


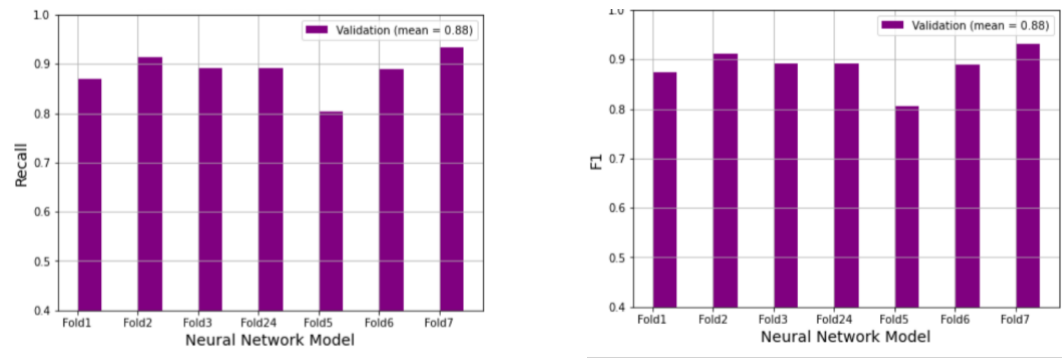
Figure 2-7: NN-CT Performance in terms of Accuracy, Precision, Recall, and F1 score.

Figure 2-8-a displays the accuracy results of NN-Hybrid in each of the 7 folds, ranging from 0.81 to 0.94, with an average of 0.88. The precision scores of the model in the 7 folds are illustrated in Figure 2-8-b, ranging from 0.84 to 0.96. The average precision score across the 7 folds is 0.90, indicating that 90% of the samples classified to a certain class are correctly assigned to the true classes. The recall of the model in the 7 folds is 0.88, as shown in Figure 2-8-c, indicating that 88% of the samples in a certain class are classified to the true class. Lastly, Figure 2-8-d illustrates the F1 scores in the 7 folds, ranging from 0.81 to 0.94 with an average of 0.88. The recall and F1 score corroborate the accuracy results in each fold in addition to the resulting mean value. The confusion matrix for the 8 classes is presented in Figure 2-9-b. Misclassifications were observed mainly between classes D2 and D3, C2 and C3, and D1 and B.



a- Accuracy

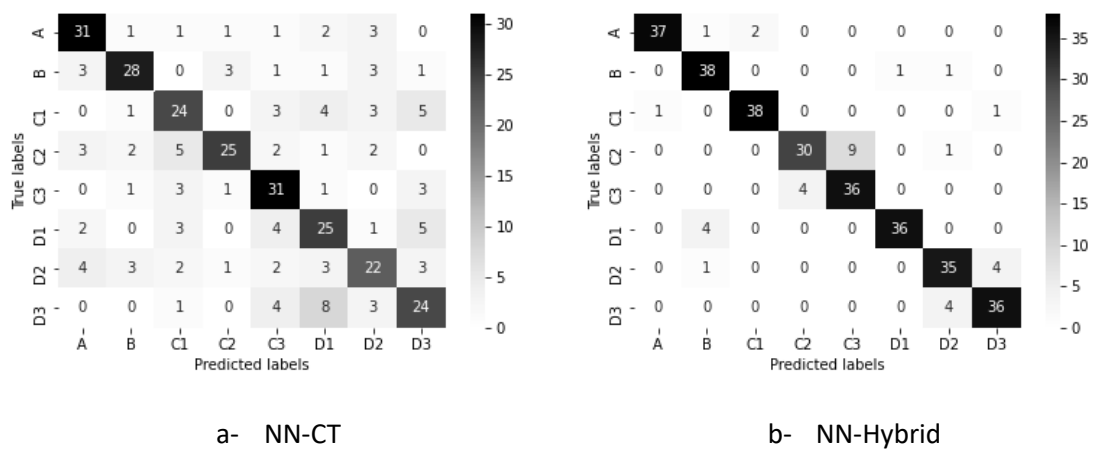
b- Precision



c- Recall

d- F1-score

Figure 2-8: NN-Hybrid Performance in terms of Accuracy, Precision, Recall, and F1 score.



a- NN-Ct

b- NN-Hybrid

Figure 2-9: Confusion Matrices for NN-Ct and NN-Hybrid

We conducted a comparison of our model's accuracy with three other four-stage COPD classification models, as presented in Table 2-4. **Error! Reference source not found.**

Table 2-4: Comparison between our proposed COPD classification model and the other COPD classification techniques.

Method	Nimura Four-staging	Lung Air Four-staging	2D-CNN Four- staging	NN-CT Eight- staging	NN- Hybrid Eight- staging
# of subjects	49	69	7,983	320	320
Accuracy	53%	84%	51%	63%	88%

2.4 Discussion

The GOLD guidelines recommend an approach to COPD treatment that considers lung function, symptom severity, and exacerbation risk. To put these guidelines into practice, an eight-stage COPD system has been developed.

Our novel classification models are designed to perform an eight-stage COPD assessment. They utilize lung air features and phenotype features from paired lung CT volumes to train NN models. These features detect lung pathology and assess disease severity accurately. This means that, using these models, patients would no longer have to undergo PFT to determine the severity of their disease. Instead, lung CT scans taken during inspiration and expiration would be employed to automatically assess disease severity. This method has the advantage of detecting the disease accurately in its early stages while not being dependent on patient compliance.

The effectiveness of the extracted features was validated by calculating the correlation coefficients between these features and clinical measurements. Both the lung air features and the phenotype features demonstrated reasonably high correlation values with PFT

measurements, indicating the efficacy of the extracted features. Lung pathologies are primarily characterized by a high concentration of air in certain parts of the lung. Therefore, pathology detection can be more effectively accomplished from expiration scans compared to inspiration scans. Consequently, lung air features extracted from expiration scans exhibited higher correlation values than those extracted from inspiration scans.

Out of the two neural network models, the NN-CT model presents a low classification accuracy. Upon observing the confusion matrix of NN-CT, it becomes evident that there is high misclassification among classes that share one or more severity factors, such as D1 and D3. Both of these classes demonstrate a high level of symptoms and severe lung function; the only discernible difference between them lies in the risk of exacerbation. Furthermore, Class C1 is misclassified with classes C3, D1, D2, and D3, all of which pose a high risk of either exacerbation or lung function impairment.

To enhance the model's performance, we proposed incorporating exacerbation frequency and symptom level as additional inputs. These factors are more accessible to measure in clinical settings than PFT. By doing so, the features extracted from CT scans can solely assess lung function, while exacerbation frequency and symptom level serve as complementary indicators. These experiments were conducted on the NN-Hybrid model, which achieved a much higher classification accuracy of 0.88.

The NN-Hybrid model exhibits high misclassification between classes that share two severity factors, namely C2 and C3, D2 and D3, or one severity factor, such as class D1 with class B. One common distinguishing factor among these classes is the severity of lung function. For instance, C2 and D2 are classes with a high risk of exacerbation but low lung function. Similarly, D1 and B are classes with a high level of symptoms, but class B has low lung function. This discrepancy could be attributed to limitations in PFT's ability to assess lung function accurately. Samples in classes C2, D2, and B may indeed have severe lung function impairment, which has not been adequately detected through PFT testing.

Among the various COPD classification models, the 2D-CNN [43] and Nimuramodel [44] achieved the lowest accuracy results. For Nimura model, the outcome could be attributed to the dominance of pulmonary function test (PFT) features over the image-based features in their model, indicating that PFT features are not highly discriminative. Furthermore,

beside the low discriminative features utilized, the Nimura model was trained with a small number of samples (49 subjects), which may have impacted its performance negatively. On the other hand, in the 2D-CNN model, a large number of subjects (more than 7000 subjects) were used to build their classification model. However, it yielded a low accuracy result of 51%. This can be attributed to the fact that only a limited portion of the lung was considered, whereas COPD severity should ideally be measured by assessing the amount of destruction in the entire lung. The Lung Air model developed by Moghadas et al. [45] demonstrates high accuracy, owing to its utilization of highly discriminative features. Trained with a small number of samples (69), it focuses on four COPD staging only, in contrast to our models, which combine lung air and PRM phenotype features. Our proposed models have been trained with 320 subjects to perform eight COPD staging. While the lung air four staging model may excel in its specific scope of disease severity assessment, our developed models represent a valuable tool for identifying COPD severity and incorporating the disease phenotype based on image data, facilitating the formulation of effective therapy plans.

2.5 Conclusions

To make this eight-stage system accessible and repeatable, two neural network (NN) models for the eight-stage COPD scheme have been developed based mainly on features extracted from thoracic CT image data. These features include lung air features and COPD phenotypes features.

The extracted features achieved reasonably high correlation values with PFT measurements and were used to train the two NN models. The first model was trained solely with features extracted from lung CT scans, while the subsequent model incorporated exacerbation frequency and symptom level alongside other extracted features. The latter model achieved high classification accuracy, primarily because the extracted features were specifically tailored to predict lung function severity.

The developed COPD classification and staging systems have achieved high accuracy in assessing COPD severity, making them valuable tools for identifying COPD severity

primarily based on image data. Such a staging system could potentially replace the need for Pulmonary Function Tests (PFT) and enable physicians to develop treatment plans based on lung CT data alone. Although CT imaging is not yet a standard practice for diagnosing mild to moderate COPD, our developed models can be a strong incentive to include CT scanning as a routine diagnostic tool for COPD patients, particularly for smokers.

2.6 References

- [1] World Health Organization WHO, “Chronic obstructive pulmonary disease (COPD).” Accessed: Mar. 07, 2024. [Online]. Available: [https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-\(copd\)](https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd))
- [2] “Global Initiative For Chronic Obstructive Lung Disease Global Strategy For The Diagnosis, Management, And Prevention Of Chronic Obstructive Pulmonary Disease (2023 REPORT),” 2022. [Online]. Available: www.goldcopd.org
- [3] “Global Initiative for Chronic Obstructive Lung Disease Global Initiative for Chronic Obstructive Lung Disease POCKET GUIDE TO COPD DIAGNOSIS, MANAGEMENT, AND PREVENTION A Guide for Health Care Professionals,” 2017. [Online]. Available: www.goldcopd.org
- [4] P. Lange *et al.*, “Prediction of the clinical course of chronic obstructive pulmonary disease, using the new GOLD classification: A study of the general population,” *Am J Respir Crit Care Med*, vol. 186, no. 10, pp. 975–981, Nov. 2012, doi: 10.1164/rccm.201207-1299OC.
- [5] M. L. K. Han *et al.*, “GOLD 2011 disease severity classification in COPDGene: A prospective cohort study,” *Lancet Respir Med*, vol. 1, no. 1, pp. 43–50, Mar. 2013, doi: 10.1016/S2213-2600(12)70044-9.
- [6] E. A. Regan *et al.*, “Genetic epidemiology of COPD (COPDGene) study design,” *COPD: Journal of Chronic Obstructive Pulmonary Disease*, vol. 7, no. 1, pp. 32–43, 2010, doi: 10.3109/15412550903499522.

- [7] S. Erol *et al.*, “Does the 2017 revision improve the ability of GOLD to predict risk of future moderate and severe exacerbation?,” *Clinical Respiratory Journal*, vol. 12, no. 8, pp. 2354–2360, Aug. 2018, doi: 10.1111/crj.12914.
- [8] D. Singh and A. Ravi, “Management of Chronic Obstructive Pulmonary Disease: A Personalized Interpretation of the Global Initiative for Chronic Obstructive Lung Disease (GOLD) - ABCD Recommendations,” *Barcelona Respiratory Network*, vol. 2, no. 1, Jan. 2016, doi: 10.23866/brnrev:2016-m0014.
- [9] A. Agustí *et al.*, “Clinical and prognostic heterogeneity of C and D GOLD groups,” *European Respiratory Journal*, vol. 46, no. 1. European Respiratory Society, pp. 250–254, Jul. 01, 2015. doi: 10.1183/09031936.00012215.
- [10] A. S. Oh *et al.*, “Visual emphysema at chest CT in GOLD stage 0 cigarette smokers predicts disease progression: Results from the COPDGene study,” *Radiology*, vol. 296, no. 3, pp. 641–649, Sep. 2020, doi: 10.1148/radiol.2020192429.
- [11] E. Pompe *et al.*, “Five-year progression of emphysema and air trapping at ct in smokers with and those without chronic obstructive pulmonary disease: Results from the COPDGene study,” *Radiology*, vol. 295, no. 1, pp. 218–226, 2020, doi: 10.1148/radiol.2020191429.
- [12] Y. Nimura *et al.*, “Assessment of COPD severity by combining pulmonary function tests and chest CT images,” *Int J Comput Assist Radiol Surg*, vol. 8, no. 3, pp. 353–363, 2013, doi: 10.1007/s11548-012-0798-y.
- [13] H. Moghadas-Dastjerdi, M. Ahmadzadeh, E. Karami, M. Karami, and A. Samani, “Lung CT image based automatic technique for COPD GOLD stage assessment,” *Expert Syst Appl*, vol. 85, pp. 194–203, Nov. 2017, doi: 10.1016/j.eswa.2017.05.036.
- [14] H. Moghadas-Dastjerdi, M. Ahmadzadeh, and A. Samani, “Towards computer based lung disease diagnosis using accurate lung air segmentation of CT images in exhalation and inhalation phases,” *Expert Syst Appl*, vol. 71, pp. 396–403, Apr. 2017, doi: 10.1016/j.eswa.2016.11.013.

- [15] C. J. Galbán *et al.*, “Computed tomography-based biomarker provides unique signature for diagnosis of COPD phenotypes and disease progression,” *Nat Med*, vol. 18, no. 11, pp. 1711–1715, Nov. 2012, doi: 10.1038/nm.2971.
- [16] S. M. Ryan, B. Vestal, L. A. Maier, N. E. Carlson, and J. Muschelli, “Template Creation for High-Resolution Computed Tomography Scans of the Lung in R Software,” *Acad Radiol*, vol. 27, no. 8, pp. e204–e215, Aug. 2020, doi: 10.1016/j.acra.2019.10.030.
- [17] M. Ibanez. Johnson, *ITK Software Guide: Design and Functionality*”, Fourth. Kitware Inc, 2015.
- [18] D. A. Lynch and M. A. Al-Qaisi, “Quantitative computed tomography in chronic obstructive pulmonary disease,” in *Journal of Thoracic Imaging*, Sep. 2013, pp. 284–290. doi: 10.1097/RTI.0b013e318298733c.
- [19] G. Gonzalez *et al.*, “Disease staging and prognosis in smokers using deep learning in chest computed tomography,” *American Journal of Respiratory and Critical Care Medicine*, vol. 197, no. 2. American Thoracic Society, pp. 193–203, Jan. 15, 2018. doi: 10.1164/rccm.201705-0860OC.

Chapter 3

3 « CNN with a Single Lung 3D CT Volume for COPD Eight Staging System»

A version of this chapter will be submitted to the Expert Systems with Applications journal for peer review and potential publication.

3.1 Introduction

Chronic obstructive pulmonary disease (COPD) represents a significant global health challenge, contributing to substantial morbidity and mortality rates. It is currently the third leading cause of death and was responsible for the loss of 3.23 million lives in 2019 [1]. Due to its significant impact, there is an increased interest in improving diagnostic methods for COPD. The main objectives of assessing COPD are to determine the extent of airflow restriction, how it affects the patient's health, and the likelihood of future exacerbations.

Airflow limitation severity is typically evaluated using the Pulmonary Function Test (PFT), which measures airflow during breathing. This test helps assessing disease severity following the Global Initiative for Chronic Lung Disease guidelines (GOLD) [2]. The GOLD staging system classifies COPD severity into four stages —mild, moderate, severe, and very severe— based on the forced expiratory volume in 1 second (FEV1) and the forced vital capacity (FVC). While this staging system provides a framework for assessing the severity of whole lung airflow limitation, it lacks critical information necessary for determining the treatment course required to manage the disease and prevent its progression. As such, a revised ABCD classification system for COPD was introduced in 2011 by the GOLD committee [2]. The degree of symptoms and history of exacerbations are considered in this system to provide more reliable criteria for determining the severity of the disease. The St. George's Respiratory Questionnaire (SGRQ) (Appendix A), the COPD Assessment Test (CAT) score, or the modified Medical Research Council (mMRC) dyspnea score can all be used to assess the severity of symptoms(Appendix B) [2]. An individual's history of exacerbations is determined by the frequency of emergency room

visits and/or hospital admissions during the preceding year. Clinicians use this information to develop treatment plans tailored to the symptoms and exacerbation history of their patients. As per Lange et al., the ABCD GOLD classification demonstrated better accuracy in predicting future exacerbations compared to the PFT GOLD staging, highlighting the significant role of symptoms and past exacerbations in the progression of COPD [3].

Another COPD staging system was developed through the COPDGene study which is an Eight COPD Staging System that incorporates three key factors for guiding COPD therapy [5]. This system classifies COPD patients into eight stages based on their symptoms and risk, where risk is determined by their exacerbation history and airflow limitation as measured by PFT ($FEV1\% < 50$). The eight stages are defined as A, low symptoms/low risk; B, high symptoms/low risk; C, low symptoms/high risk; and D, high symptoms/high risk. The C and D classes are subdivided according to the cause of the high risk as C1 and D1 are high risk due to lung function only ($FEV1\% < 50$); C2 and D2 meet exacerbation criteria only; C3 and D3 meet both exacerbation and FEV1 criteria. While more complex, the eight-staging system is advantageous over other simpler staging systems as, unlike other systems that provide information pertaining to disease severity, it can also be used effectively for treatment guidance. Based on this system, depending on the cause of the disease severity, different kinds of bronchodilators are used for treatment. A short-acting bronchodilator is recommended for class A with low symptoms, while a long-acting bronchodilator is required for high symptoms. Patients with a high risk of lung function and/or exacerbation require combining two or three kinds of long-acting bronchodilators depending on the risk cause. Different studies investigated the necessity of subgrouping C/D patients according to factors contributing to risk elevation [7], [8]. Ferreira et.al. reported that different therapeutic options must be considered for the sub-groups in C and D [7].

Generally, COPD staging systems require evaluation of 1) lung function in the clinic based on the Pulmonary Function Test (PFT) and 2) symptom severity and exacerbation. The former requires a higher level of patient cooperation, understanding of instructions and compliance compared to evaluating symptom severity and the risk of exacerbation. Lung function strongly depends on lung pathology, which can be captured effectively through assessment and can be effectively performed using lung Computed Tomography (CT). This

imaging modality has been increasingly used for COPD assessment [8], [9]. This imaging modality can detect various COPD phenotypes, such as emphysema, bronchial wall thickening, and gas trapping. Moreover, lung CT enables physicians to monitor changes in lung morphology and density as the disease advances, providing valuable insights into disease progression and treatment efficiency.

Deep learning techniques can be integrated to leverage CT imaging for assessing the severity of COPD. Several deep learning models have been developed to assess COPD severity. Gonzalez et al. developed a 2D-CNN network based on lung 3D CT data to classify COPD into four stages [10]. In their proposed model, from each scan, four slices are extracted: an axial slice, a coronal slice, and two sagittal slices at the level of the right and left lung. These slices are concatenated to form an image of size 512×512 , which serves as input for training the CNN model. Another deep learning model has been developed to assess the severity of parenchymal emphysema as trace, mild, moderate, confluent, and advanced destructive emphysema [11]. The model combines CNN and LSTM (Long Short-Term Memory) architectures and was trained using 2407 participants from the COPDGene study. More recently, Li et al. developed an unsupervised 3D convolutional autoencoder model, which was integrated with a feature constructor to build a classifier. Exploratory factor analysis was applied to explore latent traits (factors) among pattern clusters. Two of these factors were utilized to train a logistic regression model, enabling the prediction of COPD exacerbation severity [12].

For this project, we employed deep learning 3D CNN techniques in combination with thoracic CT images to evaluate the severity of COPD based on the high-resolution 8-staging system. We developed four deep learning models for the COPD 8-staging system. Two of these models utilize ResNet-18 and ResNet-34 architectures with pre-trained weights. The remaining two models were 3D-CNN models trained from scratch. One of the scratch-trained models takes only lung CT scans as input while the other model incorporates additional features along with lung CT scans.

3.2 Materials & Methods

3.2.1 Dataset

The dataset for this project is obtained from the NIH COPDGene study [6], one of the largest investigations involving 21 clinical centers dedicated to COPD research. It encompasses imaging and PFT data of 10,192 participants, including both COPD patients and controls. Participants were included from both genders with an age range between 45 and 80 years. The dataset comprises volumetric lung CT images acquired during both inhalation and exhalation phases. The voxel resolution varies from $(512 \times 512 \times 100)$ to $(512 \times 512 \times 736)$. COPDGene labels COPD subjects according to the Eight COPD Staging System. The distribution of subjects among these classes is provided in Table 3-1. The dataset exhibits class imbalance, particularly notable for class C2 and C3.

Table 3-1: Distribution of COPD subjects among the Eight classes

A	B	C1	C2	C3	D1	D2	D3
1425	1118	174	38	10	1192	254	410

3.2.1.1 Data Augmentation Method

For this project, we utilized 326 lung CT scans taken from the COPDGene. To ensure balanced class data distribution, each of the 8 classes must have ~40 subjects which are available except with classes C2 and C3. To address the data imbalance issue observed in these classes, a novel lung CT image augmentation method has been developed, comprising two primary steps: image deformable registration and image warping. In the first step, Free Form Deformation (FFD) was employed for image registration to align the lung CT scan at expiration with the one at inspiration, and vice versa. This process yielded transformation matrices that aligned the paired scans.

In the second step, three new scans were generated from each original scan through image warping using incrementally varying fractions of the transformation matrix. For expiration scans, the transformation matrix aligning the expiration scan to the inspiration one was

utilized. Subsequently, image warping at 10%, 15%, and 20% of the transformation matrix was applied to produce three new post expiration scans. This image augmentation step was applied to data samples belonging to classes C2 and C3, ensuring that the total number of samples in these classes becomes 40 samples. Detailed descriptions of the methodology can be found in Chapter 2.

3.2.2 Data Preprocessing

The 326 lung CT scans used in this project were separated into three portions: a training set, a validation set, and a testing set (divided 60:20:20). To maintain the distribution of classes within the training, validation, and test sets, we employed a Stratified split. The class representation for A, B, C1, C2, C3, D1, D2, and D3 was preserved in each set, with 40 subjects in each class except for D1 and D2, which had 43 subjects each.

To prepare the data for training, several preprocessing steps were applied. These steps included lung segmentation using a fully automated image segmentation method [14] to separate the lungs from surrounding tissues, volume cropping to eliminate surrounding areas, and Hounsfield unit normalization to adjust the voxel values to a range between 0 and 1. The volumes were resampled to a size of $128 \times 128 \times 128$ using cubic spline interpolation.

Other than the previously described data augmentation employed to address the data imbalance issue, another data augmentation was applied during training to enrich the data and further empower the classifier being developed. This augmentation included rotation, flipping, and blurring of the selected scans. ITK (Insight Toolkit), which is an open-source library renowned for its extensive tools for medical image analysis [15], was used for data preprocessing and augmentation.

3.2.3 Transfer Learning for COPD Eight Staging

State-of-the-art architectures such as Alex Net [16], VGG16 [17], and ResNet34 [18] often serve as robust starting points for transfer learning. Their capacity to learn intricate representations from large datasets renders them well-suited for fine-tuning on specific tasks, thereby contributing to the efficacy of transfer learning approaches. Given the

exceptional performance of Residual Neural Networks (ResNet) in the computer vision domain, we opted to utilize ResNet architectures.

Acquiring a pretrained 3D-CNN model specifically trained on medical images was not a straightforward task. However, the Med3D [19] framework addresses this challenge by providing pretrained models based on the ResNet architecture with varying depths (10, 18, 34, 50, 101). These models were trained on a diverse set of publicly available medical datasets, comprising 23 different medical datasets used in various competitions. The datasets encompass different medical imaging modalities, such as magnetic resonance imaging (MRI) and computed tomography (CT), covering a range of scan regions, target organs, and pathologies.

Two versions of ResNet were employed in this project: ResNet-18 and ResNet-34. Since the models were initially trained for segmentation tasks, the decoder components were replaced with fully connected layers to perform classification.

Both models underwent fine-tuning using our training data, consisting of 230 lung CT scans. Validation sets were utilized during training for fine-tuning the hyperparameters, including optimizer selection, learning rate adjustment, and the implementation of early stopping techniques. In both models, Adam optimizer was used along with an initial learning rate of 0.0001.

3.2.4 3D-CNN model for COPD Eight Staging

The architecture of the 3D-CNN model comprises eight 3D convolutional layers and four max-pooling layers, as shown in Figure 3-1. The initial two convolutional layers employ 16 filters each, while the remaining six layers utilize 32 filters each, enabling optimal feature extraction from the input data. To reduce spatial dimensions while retaining vital information, max-pooling layers are utilized to down-sample the feature maps. Finally, after the convolutional and pooling layers, five fully connected layers utilize the learned features to produce predictions.

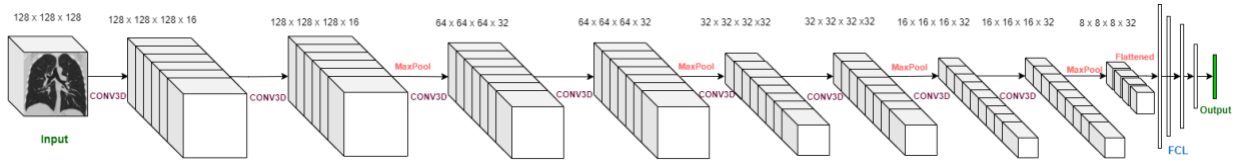


Figure 3-1: 3D-CNN architecture for the proposed eight staging schemes.

The 3D-CNN utilizes ReLU (Rectified Linear Unit) activation functions, except for the final layer, which uses SoftMax activation for multi-class classification. To prevent overfitting, early stopping is implemented by monitoring the validation set's performance, and training is optimized using the Adam optimizer. Hyperband optimization is used to tune hyperparameters, including the activation function, learning rate, number of filters per layer, and number of neurons in dense layers.

3.2.5 3D-CNN model with hybrid inputs for COPD Eight Staging

The eight-stage system for COPD considers three key factors: lung function, symptom severity, and the likelihood of future exacerbation. As such, for enhanced accuracy, we have integrated two inputs: lung CT scan and features. These features comprise two variables reflecting the frequency of exacerbations and the severity of symptoms.

The frequency of exacerbation is assessed based on the number of times a patient experienced exacerbation in the previous year, and whether these episodes necessitated a visit to the emergency room or resulted in hospitalization. Symptom severity is evaluated using the St. George's Respiratory Questionnaire (SGRQ) test (Appendix A), which yields a score ranging from 0 to 100. SGRQ employs a threshold of 25 to distinguish between high and low symptom levels. To align the model with other symptom questionnaires commonly utilized in clinical settings, we have categorized symptom scores as either high (>25) or low (<25).

The lung CT scan is used as an input to the convolutional layers, which maintain the same model structure described in Section 3.2.4. This consists of eight 3D convolutional layers with a kernel size of $3 \times 3 \times 3$, alongside four max-pooling layers. Following extraction, features derived from the convolutional layers are flattened and merged with the two input features representing symptom severity and exacerbation frequency. Subsequently, this

concatenated feature set is forwarded through five fully connected layers to process the classification (Figure 3-2).

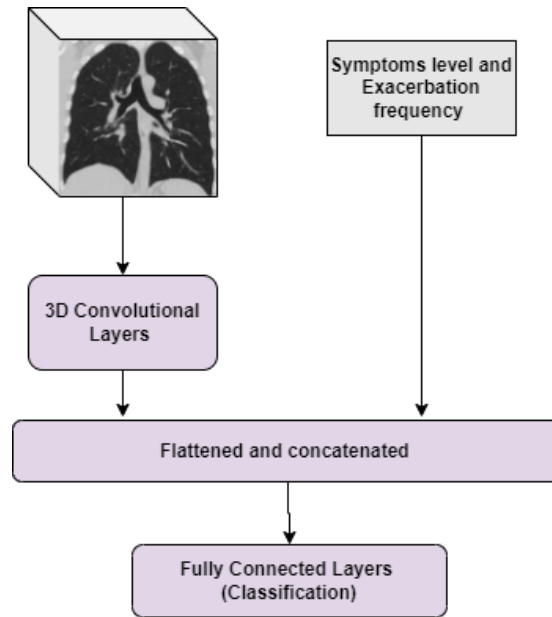
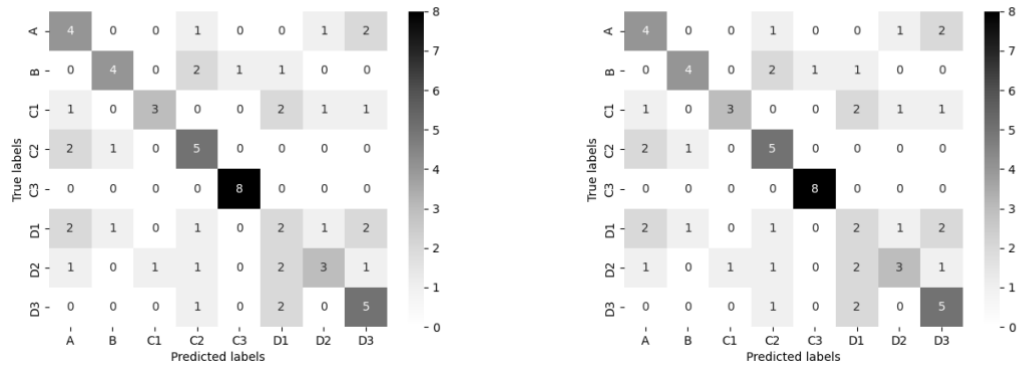


Figure 3-2: Diagram showing the illustration of the classification model

3.3 Results

3.3.1 Transfer Learning for COPD Eight Staging Evaluation

Our evaluation of model performance involved assessing classification accuracy and examining the confusion matrix. We found that the ResNet-18 model had classification accuracy rates of 0.70, 0.62, and 0.54 in the train, validation, and test sets, respectively. A visual representation of the confusion matrix for ResNet-18 can be found in Figure 3-3-a. The ResNet-34 model achieved a classification accuracy of 0.8, 0.65, and 0.51 on the training, validation, and test sets, respectively. The confusion matrix for the ResNet-34 model, depicting its performance across 8 classes, is also presented in Figure 3-3-b. The number of test samples in each class is 8, except for classes D1 and D2, which have 9 samples each.



(a) ResNet-18 Confusion Matrix.

(b) ResNet-34 Confusion Matrix.

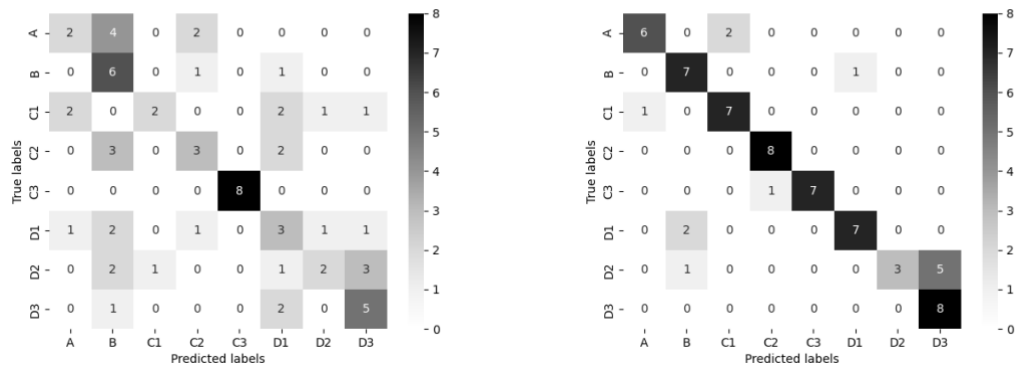
Figure 3-3: Confusion matrices for the eight staging models using transfer learning.

3.3.2 3D-CNN model for COPD Eight Staging Evaluation

Classification accuracy and confusion matrix have been used to assess the 3D-CNN model classification performance. We found that the 3D-CNN model has classification accuracy rates of 0.84, 0.54, and 0.46 in the training, validation, and test sets, respectively. Figure 3-4a presents the confusion matrix for the eight classes where the number of samples in each class is 8 except classes D1 and D2 which have 9 samples each.

3.3.3 3D-CNN model with hybrid inputs for COPD Eight Staging Evaluation

The 3D-CNN model that used hybrid inputs achieved the highest classification accuracy when compared to the other eight staging models. In the test set, the classification accuracy score was 0.80, while training and validation had an accuracy of 0.87 and 0.83, respectively. A confusion matrix was created for the eight classes, where each class had 8 samples except for classes D1 and D2, which had 9 samples each. Figure 3-4b displays the confusion matrix for these classes.



(a) 3D-CNN Confusion Matrix

(b) 3D-CNN with hybrid input Confusion Matrix

Figure 3-4: Confusion matrices for the eight staging models using the proposed 3D CNN model

3.4 Discussion

Deep learning techniques have been utilized to evaluate the severity of COPD according to the high-resolution eight-stage system. This staging system assists in treatment planning as different types of bronchodilators are used depending on the cause of the disease's severity. For example, patients with a high risk of lung function and/or exacerbation require the combination of two or three kinds of long-acting bronchodilators depending on the risk factor.

To facilitate using the eight-staging system, we developed four different deep-learning models. The first two models used transfer learning, leveraging the ResNet-18 and ResNet-34 structures. The other two models we developed from scratch, with optimized model structures and parameters. One of these models integrates two input modalities: 1) images, which capture lung function, and 2) exacerbation and symptoms features. It achieved a classification accuracy of 0.8, the highest among the four models. In this model, the convolutional layers were specifically trained to predict lung function, with exacerbation frequency and symptom level provided as additional inputs. Using CT image data in the model as an effective alternative to PFT was pursued for two reasons. First, it is becoming

increasingly common to use CT imaging for COPD assessment. Second, PFT requires patient's cooperation, understanding of instructions and their compliance. The level of patient compliance in PFT compared to evaluating symptom severity and the risk of exacerbation is known to be higher. Moreover, our results indicate that image information is less capable of capturing symptom severity and exacerbation.

Although the other three models did not perform well in terms of classification accuracy, they still provide valuable information. The models perform an eight-class classification, hence random guessing would yield an accuracy of only 0.12 whereas the classification accuracies attained in the test set are 0.54, 0.51, and 0.46 for ResNet-18, ResNet-34, and the 3D-CNN trained with a single input, respectively. This indicates that these models are achieving performance levels approximately 4 to 5 times better than random guessing.

The confusion matrix supports this observation as most of the misclassifications occur between classes that have at least one factor in common. For instance, the samples belonging to class D2 are frequently misclassified as class D3 in all four models. These two classes share two factors: severe symptoms and a high risk of exacerbation. The only difference between them is that class D2 indicates moderate lung pathology while class D3 indicates severe lung pathology. This highlights one of the limitations of our data. The samples that are classified as D3 instead of D2 might have severe lung pathology, but the lung function test fails to accurately capture it.

An important advantage of using CT over PFT for COPD assessment is the observation that quantitative CT measurements can detect disease progression before any measurable change is detected with PFT (Pulmonary Function Test). A study conducted by Pompe et al. [20] investigated longitudinal changes in emphysema and air-trapping, which were quantitatively measured from lung CT scans, alongside PFT measurements in cigarette smoker subjects with and without COPD. Their findings revealed a significant increase in emphysema and air-trapping over a 5-year follow-up period in both smokers with and without COPD. However, they found that a substantial proportion of the progression in emphysema and air-trapping was not correlated with PFT. Therefore, assessing lung function with lung CT is crucial for detecting the disease in its early stages.

Previous studies have been conducted to evaluate the severity of Chronic Obstructive Pulmonary Disease (COPD). The 2D-CNN model developed by Gonzalez et al. for 4-stage classification achieved a classification accuracy of 51.1% [10]. It was trained using data from 7,983 COPDGene participants. The CNN-LSTM model developed by Humphries et al. classified COPD in terms of emphysema level into 5 levels using data from 2407 participants [11]. However, only 45% of the classification results matched the correct emphysema level.

Our deep learning models are the first to perform eight-stage classifications for COPD severity. Moreover, our models' classification performance is comparable to 4 and 5-stage models. When hybrid inputs are used, our 3D-CNN model achieves high classification accuracy, potentially making it suitable for clinical settings.

3.5 Conclusions

The Global Initiative for Chronic Obstructive Lung Disease (GOLD) uses a combination of three factors to assess COPD, which are lung function, symptoms, and exacerbation history. This helps to determine the severity of the disease and create a treatment plan that is tailored to the individual's needs. The COPDGene study has designed an eight COPD staging system that classifies COPD patients into eight different categories to provide critical information necessary for guiding therapy. In this study, we used deep learning models we developed from scratch in addition to transfer learning techniques to assess COPD severity based on the high-resolution eight-stage system. This involves analyzing the thoracic CT images of patients to provide a more accurate assessment of their condition.

Three of the models were developed for COPD eight-stage system utilizing only lung CT scan data as input. The first two models employed transfer learning, utilizing the state-of-the-art architectures ResNet-18 and ResNet-34. The third model is the 3D-CNN model, which was trained from scratch using optimized model parameters.

The hybrid model integrated additional features related to symptom level and the risk of exacerbation along with lung CT scan data. It achieved high classification accuracy,

primarily because the convolutional layers were trained specifically to predict lung function.

Given that the eight-stage system requires physicians to conduct three tests for devising an appropriate treatment plan, our deep learning models streamline the physician's task, eliminating the need for multiple tests and enabling a more efficient and effective treatment plan.

3.6 References

- [1] World Health Organization WHO, “Chronic obstructive pulmonary disease (COPD).” Accessed: Mar. 07, 2024. [Online]. Available: [https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-\(copd\)](https://www.who.int/news-room/fact-sheets/detail/chronic-obstructive-pulmonary-disease-(copd))
- [2] “Global Initiative for Chronic Obstructive Lung Disease Global Initiative for Chronic Obstructive Lung Disease POCKET GUIDE TO COPD DIAGNOSIS, MANAGEMENT, AND PREVENTION A Guide for Health Care Professionals,” 2017. [Online]. Available: www.goldcopd.org
- [3] P. Lange *et al.*, “Prediction of the clinical course of chronic obstructive pulmonary disease, using the new GOLD classification: A study of the general population,” *Am J Respir Crit Care Med*, vol. 186, no. 10, pp. 975–981, Nov. 2012, doi: 10.1164/rccm.201207-1299OC.
- [4] M. L. K. Han *et al.*, “GOLD 2011 disease severity classification in COPD Gene: A prospective cohort study,” *Lancet Respir Med*, vol. 1, no. 1, pp. 43–50, Mar. 2013, doi: 10.1016/S2213-2600(12)70044-9.
- [5] S. Erol *et al.*, “Does the 2017 revision improve the ability of GOLD to predict risk of future moderate and severe exacerbation?,” *Clinical Respiratory Journal*, vol. 12, no. 8, pp. 2354–2360, Aug. 2018, doi: 10.1111/crj.12914.
- [6] D. Singh and A. Ravi, “Management of Chronic Obstructive Pulmonary Disease: A Personalized Interpretation of the Global Initiative for Chronic Obstructive Lung Disease

- (GOLD) - ABCD Recommendations,” *Barcelona Respiratory Network*, vol. 2, no. 1, Jan. 2016, doi: 10.23866/brnrev:2016-m0014.
- [7] A. Agustí *et al.*, “Clinical and prognostic heterogeneity of C and D GOLD groups,” *European Respiratory Journal*, vol. 46, no. 1. European Respiratory Society, pp. 250–254, Jul. 01, 2015. doi: 10.1183/09031936.00012215.
- [8] D. A. Lynch and J. D. Newell, “Quantitative Imaging of COPD,” *J Thorac Imaging*, vol. 24, no. 3, pp. 189–194, Aug. 2009, doi: 10.1097/RTI.0b013e3181b31cf0.
- [9] G. Washko, “Diagnostic Imaging in COPD,” *Semin Respir Crit Care Med*, vol. 31, no. 03, pp. 276–285, Jun. 2010, doi: 10.1055/s-0030-1254068.
- [10] G. Gonzalez *et al.*, “Disease staging and prognosis in smokers using deep learning in chest computed tomography,” *American Journal of Respiratory and Critical Care Medicine*, vol. 197, no. 2. American Thoracic Society, pp. 193–203, Jan. 15, 2018. doi: 10.1164/rccm.201705-0860OC.
- [11] S. M. Humphries *et al.*, “Deep learning enables automatic classification of emphysema pattern at CT,” *Radiology*, vol. 294, no. 2, pp. 434–444, 2020, doi: 10.1148/radiol.2019191022.
- [12] F. Li *et al.*, “Latent traits of lung tissue patterns in former smokers derived by dual channel deep learning in computed tomography images,” *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-84547-5.
- [13] E. A. Regan *et al.*, “Genetic epidemiology of COPD (COPDGene) study design,” *COPD: Journal of Chronic Obstructive Pulmonary Disease*, vol. 7, no. 1, pp. 32–43, 2010, doi: 10.3109/15412550903499522.
- [14] S. M. Ryan, B. Vestal, L. A. Maier, N. E. Carlson, and J. Muschelli, “Template Creation for High-Resolution Computed Tomography Scans of the Lung in R Software,” *Acad Radiol*, vol. 27, no. 8, pp. e204–e215, Aug. 2020, doi: 10.1016/j.acra.2019.10.030.
- [15] M. Ibanez. Johnson, *ITK Software Guide: Design and Functionality*”, Fourth. Kitware Inc, 2015.

- [16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks.” [Online]. Available: <http://code.google.com/p/cuda-convnet/>
- [17] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Sep. 2014, [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [18] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition.” [Online]. Available: <http://image-net.org/challenges/LSVRC/2015/>
- [19] S. Chen, K. Ma, and Y. Zheng, “Med3D: Transfer Learning for 3D Medical Image Analysis,” Apr. 2019, [Online]. Available: <http://arxiv.org/abs/1904.00625>
- [20] E. Pompe *et al.*, “Five-year progression of emphysema and air trapping at ct in smokers with and those without chronic obstructive pulmonary disease: Results from the COPDGene study,” *Radiology*, vol. 295, no. 1, pp. 218–226, 2020, doi: 10.1148/radiol.2020191429.

Chapter 4

4 « CNN with a Single Lung 3D CT Volume for COPD Classification Based on The GOLD 2023 Staging System»

A version of this chapter will be submitted to the Expert Systems with Applications journal for peer review and potential publication.

4.1 Introduction

Chronic Obstructive Pulmonary Disease (COPD) is a heterogeneous lung disease that is characterized by airflow limitation which causes difficulty breathing. Forced expiratory volume in one second (FEV1) and forced vital capacity (FVC) are often measured by pulmonary function tests (PFTs) to diagnose COPD. However, in the early stages of COPD, these measurements are unable to detect local structural and functional changes. Moreover, while requiring patient's understanding and compliance with instructions given in the tests, spirometry has a poor correlation with the degree of breathlessness or any other COPD symptoms [1]. Therefore, using spirometry measures to diagnose the disease and provide individualized treatment in clinical settings is lacks desirable reliability.

Different updates have been designed by the Global Initiative for Chronic Lung Disease (GOLD) committee to manage the disease and prevent its progression. Toward higher level of practicality in the clinic, in 2023, a new ABE staging system for COPD was designed [2]. Compared to the eight-staging system, this system is substantially lower resolution and relies on the level of symptoms and COPD Exacerbation history to assess the disease severity (Figure 4-1). The symptoms level can be measured using a dyspnea measure (the modified Medical Research Council [mMRC] dyspnea score (Appendix B) or a health status measure (the COPD Assessment Test [CAT] score), or the SGRQ (St. George's Respiratory Questionnaire)(Appendix A). COPD Exacerbation is defined as the acute worsening of symptoms resulting in the necessity of additional therapy; it is measured as the frequency of requiring assessment in the Emergency Department or hospital admission in the prior year. Patients' symptoms and exacerbation history allow clinicians to initiate a

treatment plan at the individual patient level. Lange et al. stated that symptoms and exacerbation history can be used as good predictors of future COPD exacerbation, and can be effective in devising therapy plans aiming at preventing or slowing down COPD progression as it has [3]. The frequency of severe exacerbations is associated with an increased risk of mortality, especially if it requires hospital admission [4]. Therefore, it is crucial to determine whether the patient is at high risk of experiencing COPD exacerbation.

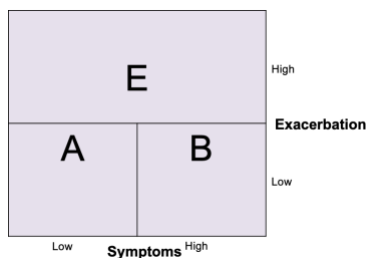


Figure 4-1: COPD staging scheme based on GOLD2023 staging system.

The criterion used to assess the severity of exacerbation is not reliable as not all patients who experience COPD exacerbation and require hospitalization visit the hospital. The rate of hospitalization for COPD exacerbations is known to decrease where patients perceive higher risk for hospitalization compared to its potential benefits. For example, COPD exacerbation-related hospitalization was reduced by 53% compared to the pre-COVID pandemic [5]. Avoiding visiting emergency rooms during COVID-19 possibly caused a change in the treatment of COPD exacerbations from inpatient to outpatient. To cope with such situations, other assessment tools are essential to detect the severity of COPD exacerbation in the GOLD2023 staging system.

Computed tomography (CT) has a rapidly developing role in COPD investigation as it can reveal valuable information pertaining to the lung pathology. It can detect the different COPD phenotypes of emphysema, bronchial wall thickening and gas trapping. As COPD symptoms and exacerbation are a manifestation of these phenotypical pathologies, it is conceivable that lung CT data can encode information pertaining to COPD symptoms and exacerbation. As a potentially viable tool for decoding this information, deep learning can be employed to take advantage of CT imaging in assessing COPD severity based on the framework of GOLD2023 staging system. Given the pitfalls of PFTs, this approach can be used as a more effective alternative in this staging system.

While to our knowledge, so far no studies have attempted to incorporate deep learning to develop classifiers for the GOLD2023 staging system. However, relevant investigations studies have been conducted to assess COPD severity using deep learning and thoracic CT. Gonzalez et al. developed a 2D-CNN network that utilizes 3D CT scans of lungs [6]. The network was trained using lung CT scans from the COPDGene study. Four slices were extracted from each scan, including an axial slice, a coronal slice, and two sagittal slices at the level of the left and right lung. These four slices were used as input to train the CNN model. The model was trained to perform multi-classification of 1) COPD vs. non-COPD (binary classification), and 2) to the disease staging. Another deep learning model was developed to assess the severity of parenchymal emphysema as trace, mild, moderate, confluent, and advanced destructive emphysema [7]. The model combines CNN and LSTM (Long Short-Term Memory) architectures and was trained using data pertaining to 2407 participants from the COPDGene study. Axial slices were extracted from each scan and used as input to train the CNN. Feature vectors obtained from all slices were concatenated and used as an input to the LSTM model. Then, A composite feature vector was produced by this LSTM model which was used as input to the fully connected layers to produce the predicted emphysema level. In the model developed by Ho et al. paired 3D volumes were extracted from 4D-CT data and used to create a 3D Parametric Response Map (PRM) before a 3D-CNN model was trained [8]. More recently, an unsupervised 3D convolutional autoencoder model was developed and integrated with a feature constructor as a classifier. An exploratory factor analysis was applied to explore the latent traits (factors) among pattern clusters. Two of the factors were used to train a logistic regression model to predict the severity of COPD exacerbation [9].

In this research project, we developed a classification technique based on the GOLD2023 staging system. To develop this model, we used lung CT images integrated with deep learning to measure the symptoms level and the severity of COPD exacerbation. Toward this goal, five models were trained and evaluated for this measurement. The first model was trained from scratch with optimized model parameters, while the other four were trained using transfer learning from the state-of-the-art architectures ResNet18 and ResNet34.

4.2 Materials & Methods

4.2.1 Proposed GOLD2023 based COPD classification

GOLD2023 The staging system is based on the patient's symptom severity and future exacerbation risk. To determine the stage of COPD according to this system, we developed deep learning models for symptom and exacerbation measurement using CT scans. We employ both of the developed models concurrently to determine COPD staging based on the scheme illustrated in Figure 4.1. The classification process begins with the lung CT scan being passed through the exacerbation deep learning model. If the model indicates a high risk of exacerbation, the patient is classified as GOLD stage E. If the model indicates a low risk, then the lung CT scan is passed to the symptom detection model. If the symptom detection model indicates a positive result, the patient is classified as GOLD stage B, and if it indicates a negative result, the patient is classified as GOLD stage The diagram of the classification scheme is illustrated in Figure 4-2.

4.2.2 Dataset:

The data required for this project is available through the NIH COPDGene (COPD Genetic Epidemiology) study [10]. It is one of the largest studies exploring COPD. COPDGene is a multicenter study that collected data from 21 clinical centers and using different multi-detector CT scanners. It includes volumetric lung CT data, subjects' PFT measurements, exacerbation frequency, and symptom levels of 10,192 participants aged between 45 to 80 years from both genders. The voxel volume in the lung CT varied from 0.244 mm^3 to 1.376 mm^3 , with image resolution ranging from $(512 \times 512 \times 100)$ to $(512 \times 512 \times 736)$.

In this project, we utilized thoracic CT scans at the expiration phase. The lung CT scan in the expiration phase is valuable for detecting air trapping. This approach is particularly relevant when assessing lung pathology in patients with COPD. Air trapping, a characteristic feature of COPD, provides a more comprehensive assessment of overall lung pathology, including emphysematous lung damage and airway obstruction. We selected 562 subjects to develop the symptoms models. Among them, 284 and 278 subjects belong to the high and low symptom classes, respectively. Exacerbation models were developed

using 468 subjects, where 234 subjects belonged to the high exacerbation class and another 234 subjects belonged to the low exacerbation class.

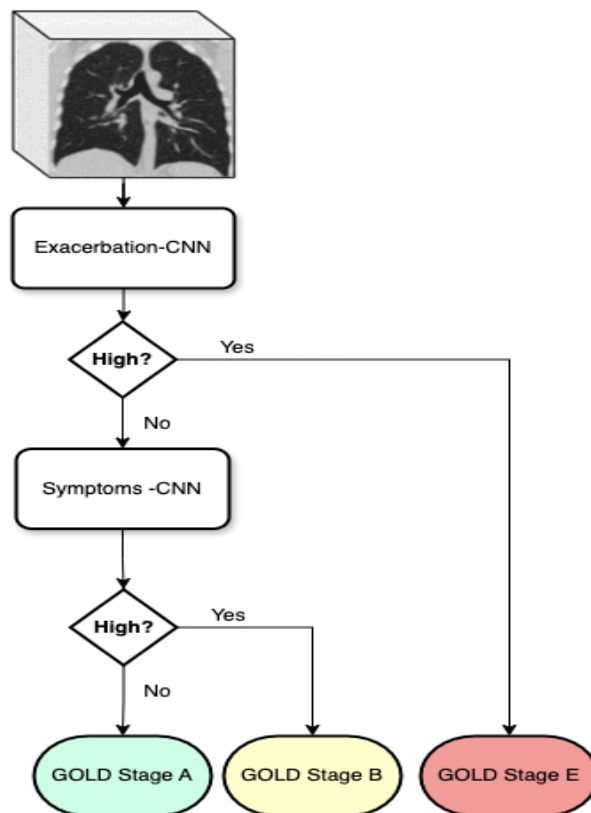


Figure 4-2: Proposed GOLD2023 classification scheme.

4.2.3 Transfer Learning

Deep Learning models developed in this investigation were trained wither from scratch or using Transfer Learning. Transfer Learning allows models to use information from one task to improve performance on another related task. Pretrained models contain knowledge learned from large datasets and complex tasks. By starting with a pre-trained model, the training process for a new task can be faster and more efficient as the model has already learned useful features and representations. Thus, it helps models generalize better to new and unseen data.

State-of-the-art architectures like Alex Net [11], VGG16 [12], ResNet34 [13] often serve as strong starting points for transfer learning. Their ability to learn rich representations on large datasets makes them excellent candidates for fine-tuning on specific tasks, contributing to the success of transfer learning approaches. Due to the outstanding performance of Residual Neural Networks in the computer vision domain, we used Residual Network (ResNet) architectures.

4.2.4 Thoracic CT Preprocessing

Several preprocessing steps were applied to the lung CT volume. First, each lung volume was segmented using a fully automated image segmentation method [14]. The segmented lung volume was used to apply volume cropping and remove background area. Volumes were resampled to size $128 \times 128 \times 128$ using cubic spline interpolation. Lastly, the Hounsfield unit (HU) of each voxel was normalized to values in the range of [0, 1]. All data preprocessing was applied using ITK, an open source toolkit that has extensive tools for processing medical images [15].

4.2.5 Pretrained Models for Transfer Learning

Several pre-trained 2D CNN models with downloadable parameters are readily available. These models have been pretrained using millions of natural images from the ImageNet dataset. However, there is a scarcity of pre-trained 3D CNN models available for download. Med3D [16] used varying depths to train multiple ResNet models (10, 18, 34, 50, 101). Med3D used 23 different medical datasets for training, which is a collection of several publicly available 3D segmentation datasets from different medical imaging modalities, e.g., magnetic resonance imaging (MRI) and computed tomography (CT), with various scan regions, target organs and pathologies.

In this project, we used two versions of ResNet: ResNet-18 and ResNet-34. The decoder was replaced with fully connected layers to perform the classification. ResNet-18 is a deep neural network architecture consisting of 18 layers, where 17 of them are convolutional layers followed by a fully connected layer and the output layer. The model begins with a convolutional layer followed by eight residual blocks. Each residual block is made up of

two convolutional layers. The convolutional layers in the model use $3\times 3\times 3$ kernels, and the fully connected layer has 512 neurons.

ResNet-34, on the other hand, is a deeper neural network architecture consisting of 34 layers, where 33 of them are convolutional layers with a kernel size of $3\times 3\times 3$. It is followed by a fully connected layer with 512 neurons. The last output layer uses the SoftMax activation function to perform classification.

4.2.6 Symptoms Detection

In the COPDGene study, the level of symptoms was evaluated using two questionnaires—the mMRC questionnaire (Appendix B) and the SGRQ questionnaire (Appendix A). The mMRC questionnaire is a simple questionnaire with only 5 items while the SGRQ questionnaire is more comprehensive with several questions. To minimize errors from both questionnaires, we redefined the level of symptoms as follows: If both questionnaires indicated a high level of symptoms, then the symptoms' level was considered high. Conversely, if both questionnaires indicated a low level of symptoms, then the symptoms' level was considered low. However, if one of the questionnaires indicated a high level of symptoms while the other indicated a low level of symptoms, then other data such as increased medication in the prior year and frequency of using antibiotics or steroids at home were considered to assess the level of symptoms. If the subject used antibiotics, steroids, or increased medication at home in the prior year, their sample is considered high-symptom.

We used five models to determine the level of symptoms from lung CT scans. The first model, called Symptoms-CNN, was trained from scratch using optimized model parameters. The other four models were trained using pre-trained, state-of-the-art architectures called ResNet-18 and ResNet-34. To train the models, the dataset was split into training, validation, and testing sets using a ratio of 75:25. Specifically, 75% of the dataset, comprising 422 CT scans, was allocated for training the models. The remaining 25% of the dataset was evenly split between validation and testing sets, resulting in 70 CT scans allocated to each set. The models were trained to perform binary classification to categorize COPD subjects based on the level of symptoms as either high or low.

4.2.6.1 *Symptoms-CNN Model* for Symptoms Detection

Symptoms-CNN was trained from scratch where the model parameters were optimized. The CNN starts with a 3D convolutional layer (CONV3D) with a kernel size of $7 \times 7 \times 7$ and 8 filters (Figure 4-3). Four Residual Blocks follow the first layer, where each Residual Block consists of two CONV3D with a kernel size of $3 \times 3 \times 3$, two ReLU activations, and one addition. The number of filters in these blocks are 16, 16, 32, and 32, respectively. The resulting features are flattened and passed to three fully connected layers. The total trainable parameters are $\sim 2M$ and the model was trained with Adam optimizer in 89 epochs (Loss learning curve is shown in Figure 4-5-a). Model hyper parameters were assigned using the Hyperband optimization method, which is a hyperparameter optimization method. It employs a concept called "successive halving" to simultaneously explore multiple configurations of hyperparameters and allocate resources to promising configurations across the entire network. Hyper parameters included in the optimization are activation function, learning rate, number of filters in each of the fully connected layers, and number of neurons in each of the fully connected layers.

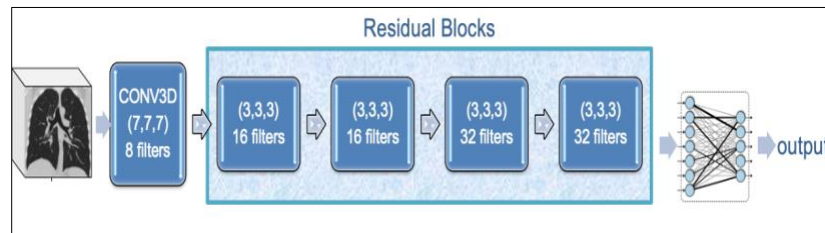


Figure 4-3: Proposed symptoms-CNN model structure.

4.2.6.2 Pretrained Models for Symptom Detection

State-of-the-art architectures, namely ResNet-18 and ResNet-34, are being used for symptom detection with pre-trained weights. Two transfer learning approaches have been implemented. The first approach involves using the convolutional layers as a feature extractor while keeping the pre-trained weights frozen and fine-tuning the fully connected layers only. The second approach involves initializing the network with pre-trained weights

and fine-tuning the entire network. In both approaches, Adam optimizer is used in conjunction with the Early Stopping technique used to determine the number of epochs.

4.2.7 Exacerbation Detection

The severity of COPD exacerbation has been measured by evaluating the frequency of experiencing exacerbation over a year. If the patient has undergone at least one exacerbation that led to hospital admission or two or more exacerbations that led to emergency room visits, then it is considered a severe exacerbation. Otherwise, it is considered a low exacerbation. The training dataset comprised 356 CT scans, while the remaining 112 samples were evenly split between validation and testing, with 56 scans allocated to each. The model was trained to perform binary classification, categorizing the risk of future exacerbation as high or low.

4.2.7.1 *Exacerbation-CNN Model* for Exacerbation Detection

The lung CT scan data is fed into a CONV3D layer that uses a $7 \times 7 \times 7$ kernel with 16 filters (as shown in Figure 4-4). After the first layer, four Residual Blocks are employed, each with a different number of filters (16, 16, 32, and 128). The resulting features are then flattened and passed to three fully connected layers for classification. The model, which has approximately 8 million trainable parameters, was trained using the Adam optimizer in 86 epochs (Loss learning curve is shown in Figure 4-5-b). The Hyperband optimization method was used to determine the hyperparameters for the model, which include the activation function, learning rate, number of filters in each fully connected layer, and number of neurons in each fully connected layer.

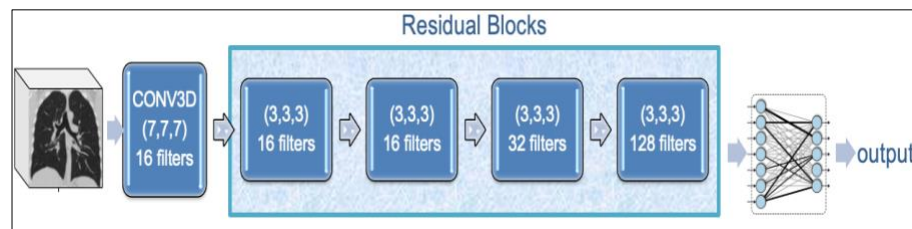


Figure 4-4: Exacerbation-CNN Model Structure

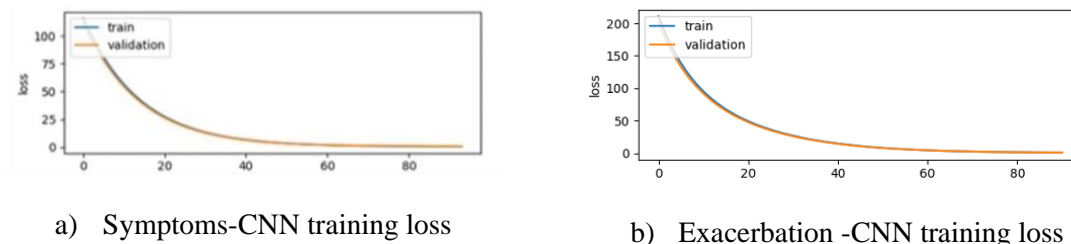


Figure 4-5: Loss learning curve for training and validation sets

4.2.7.2 Pretrained Models for Exacerbation Detection

State-of-the-art ResNet-18 and ResNet-34 architectures, which come equipped with pre-trained weights, were used to detect exacerbation. Two transfer learning approaches were employed. In the first approach, the convolutional layers function as a feature extractor, utilizing pre-trained weights. In this approach, the weights in the convolutional layers are frozen, and fine-tuning is only applied to the fully connected layers. The second approach involves initializing the entire network with pre-trained weights, and then fine-tuning is carried out across the whole network. For both methods, training utilizes the Adam optimizer. The Early Stopping technique is employed to determine the optimal number of epochs required for training.

4.3 Results

4.3.1 *Symptoms Detection Evaluation*

4.3.1.1 Symptoms-CNN model evaluation

The accuracy score is used to evaluate the Symptoms-CNN model classification performance. The model achieved a classification accuracy of 0.8, 0.71, and 0.7 in training, validation, and test sets, respectively. The precision, recall, and F1-score were employed to assess the model's classification performance. For precision, recall, and F1-score, values

of 0.74, 0.66, and 0.70 were attained, respectively (Table 4-1). The Confusion Matrix and the Receiver Operating Characteristic (ROC) were calculated for the test set. The test set has 70 samples where 35 samples belong to each of high and low symptom classes. The Confusion Matrix is shown in Figure 4-6-e; it shows that 0.74 of samples with high symptoms are predicted correctly as high symptoms samples. The ROC curve with Area Under the Curve (AUC) of 0.73 was achieved in the Symptoms-CNN model as illustrated in Figure 4-6-f.

4.3.1.2 ResNet-18 models evaluation

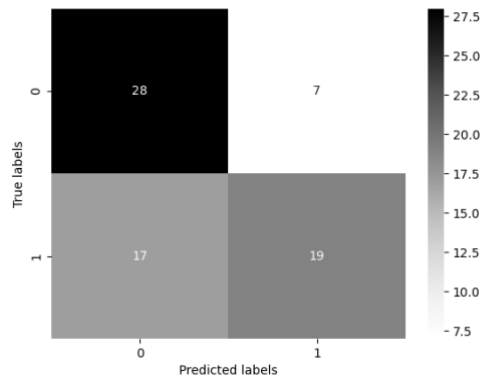
Two approaches were considered with ResNet-18 pretrained model. The first approach where the ResNet-18 model was used as a feature extractor, resulted in accuracy scores of 0.62, 0.55, and 0.5 for the train, validation, and test sets, respectively. For the test set, precision, recall, and F1-score values of 0.5, 0.54, and 0.52 were attained as given in Table 4-1. The fine-tuned ResNet-18 demonstrated improved classification scores, with values of 0.66, 0.68, and 0.66 achieved for classification accuracy in the train, validation, and test sets, respectively. Precision, recall, and F1-score achieved values of 0.74, 0.47, and 0.58 respectively, as shown in Table 4-1. The confusion matrix for the fine-tuned model is illustrated in Figure 4-6-a. The ROC curve, depicted in Figure 4-6-b, indicates an AUC of 0.68 for the fine-tuned ResNet-18 model.

4.3.1.3 ResNet-34 model evaluation

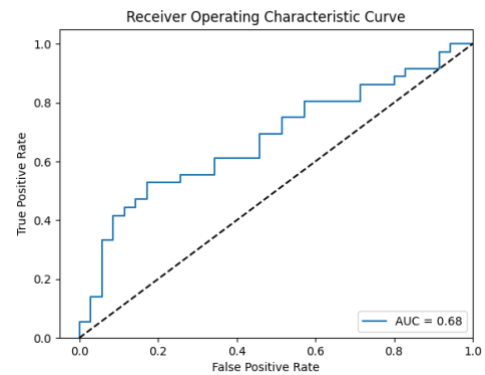
Pretrained ResNet-34 models were evaluated in terms of accuracy, precision, recall, and F1-score. The accuracy scores for the train, validation, and test sets were 0.65, 0.6, and 0.54 in the first approach, which used the ResNet-34 model as the feature extractor. A value of 0.56 was attained across all three metrics of precision, recall, and F1-score in the test set. The second approach exhibits higher classification performance, with accuracy scores of 0.9, 0.70, and 0.71 achieved for the training, validation, and test sets, respectively. In Table 4-1, precision, recall, and F1-score metrics are reported for the test set as 0.74, 0.69, and 0.71, respectively. Figure 4-6-c presents the confusion matrix for the two methods. The fine-tuned ResNet-34 has an AUC of 0.73 according to the ROC curve shown in in Figure 4-6-d. The classification evaluation metrics for all symptoms models are reported in Table 4-1.

Table 4-1: Symptoms' models classification metrics.

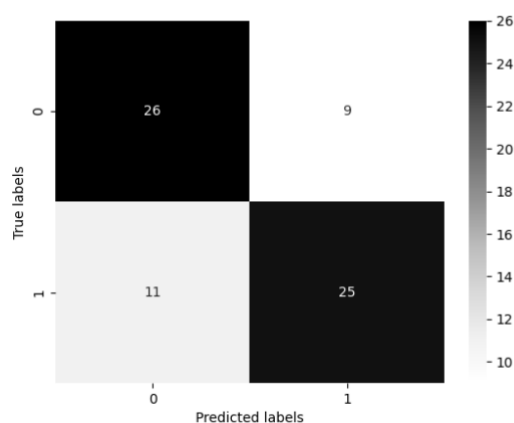
Evaluation Metrics	ResNet-18 “Feature extractor”	ResNet-18 “fine-tuned”	ResNet-34 “Feature extractor”	ResNet-34 “fine-tuned”	Symptoms-CNN
Accuracy	0.50	0.66	0.54	0.71	0.70
Precision	0.5	0.74	0.56	0.74	0.74
Recall	0.54	0.47	0.56	0.69	0.66
F1-score	0.52	0.58	0.56	0.71	0.70
AUC	0.51	0.68	0.55	0.73	0.73
Training epochs	20 epochs	16 epochs	38 epochs	13 epochs	139 epochs
Training time	1 min	4 min	5 min	5 min	13 min + 5 hours optimization



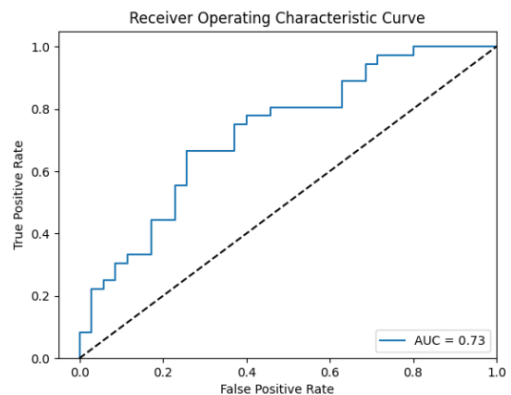
(a)



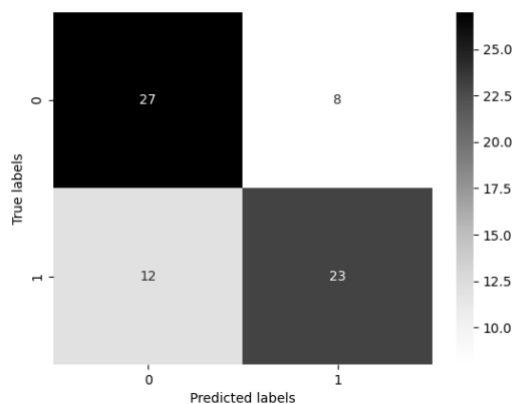
(b)



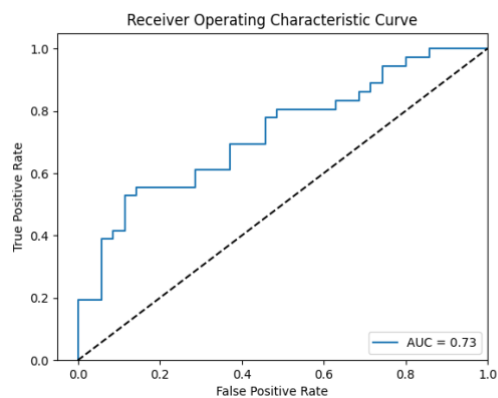
(c)



(d)



(e)



(f)

Figure 4-6: Confusion Matrix and ROC curve for symptoms models: (a) fine-tuned ResNet-18 Confusion Matrix, (b) fine-tuned ResNet-18 ROC curve, (c) fine-tuned ResNet-34 Confusion Matrix, (d) fine-tuned ResNet-18 ROC curve, (e) Symptoms-CNN Confusion Matrix, (f) Symptoms-CNN ROC curve.

4.3.2 Exacerbation Detection Evaluation

4.3.2.1 Exacerbation-CNN model evaluation

The Exacerbation-CNN model was evaluated in terms of accuracy score, Confusion Matrix and ROC curve. The model's classification accuracy scores in training, validation and testing sets are 0.84, 0.76, and 0.75, respectively. Precision, recall, and F1-score were used to evaluate model classification performance. As shown in Table 4-2, values of 0.72, 0.82,

and 0.77 were achieved for precision, recall, and F1-score, respectively. The confusion matrix for the test set is shown in Figure 4-7-e, where the number of samples in the test sets is 56 samples (28 samples for each of the low and high risk categories). The confusion matrix shows that 23 out of 28 samples were correctly identified as high exacerbation risk, and only 5 samples were misclassified as low exacerbation risk. Figure 4-7-f shows the ROC curve for the Exacerbation-CNN model. The model achieved an AUC of 0.78 in the test set.

4.3.2.2 ResNet-18 models evaluation

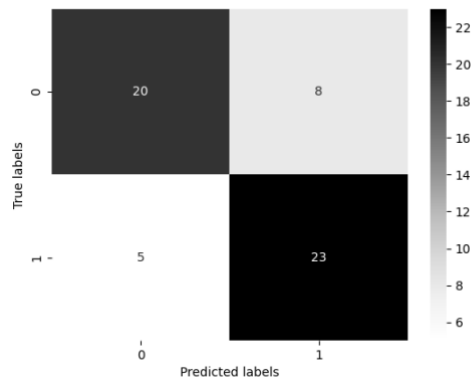
The accuracy scores for the train, validation, and test sets were 0.71, 0.62, and 0.62 in the first approach, which used the ResNet-18 model as the feature extractor. Precision, recall, and F1-score values for the test set were 0.67, 0.5, and 0.57, respectively. With classification accuracy of 0.74, 0.76, and 0.76 in the train, validation, and test sets, respectively, the fine-tuned ResNet-18 showed improved classification scores. Precision, recall, and F1-score in the fine-tuned ResNet-18 were reported as 0.86, 0.64, and 0.73, respectively (Table 4-2). The Confusion Matrix for the fine-tuned ResNet-18 is shown in Figure 4-7-a. The ROC curve in Figure 4-7-b indicates that the AUC of the fine-tuned ResNet-18 is 0.83.

4.3.2.3 ResNet-34 models evaluation

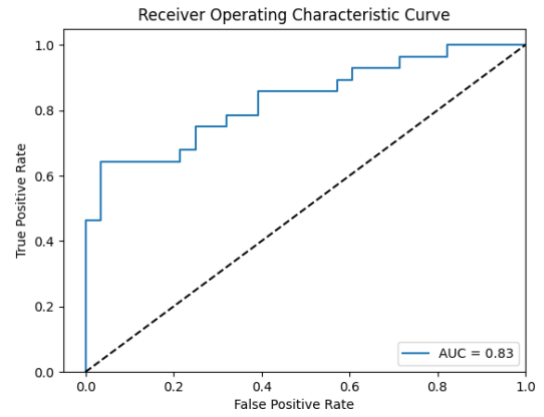
In the first approach, which utilized the ResNet-34 model as a feature extractor, accuracy scores of 0.72, 0.64, and 0.6 were achieved for the train, validation, and test sets, respectively. As shown in Table 4-2, precision, recall, and F1-score values pertaining to the test set were calculated at 0.65, 0.46, and 0.54, respectively. In the fine-tuned ResNet-34, values of 0.68, 0.71, 0.73 were achieved for model classification accuracy in train, validation, and test sets, respectively. Also, for the same test set, values of 0.76, 0.68, and 0.72 were calculated for precision, recall, and F1-score, respectively. Figure 4-7-c shows the Confusion Matrix for the fine-tuned ResNet34 while the AUC of this model was calculated at 0.77, according to the ROC curve shown in Figure 4-7-d. The classification evaluation metrics for all exacerbation models are illustrated in Table 4-2.

Table 4-2: Exacerbation detection models evaluation metrics

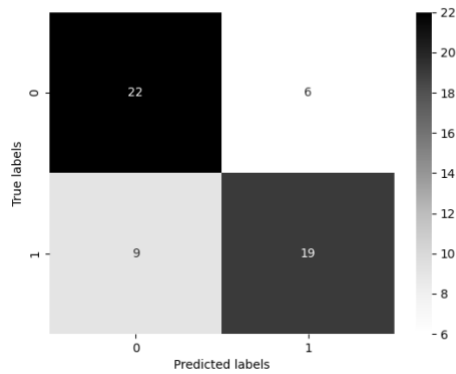
Evaluation Metrics	ResNet-18 “Feature extractor”	ResNet-18 “fine-tuned”	ResNet-34 “Feature extractor”	ResNet-34 “fine-tuned”	Exacerbation-CNN
Accuracy	0.62	0.76	0.60	0.73	0.75
Precision	0.67	0.86	0.65	0.76	0.72
Recall	0.50	0.64	0.46	0.68	0.82
F1-score	0.57	0.73	0.54	0.72	0.77
AUC	0.62	0.83	0.61	0.73	0.78
Training Epochs	19 epochs	18 epochs	21 epochs	16 epochs	189 epochs
Training Time	1 min	3 min	2 min	5 min	21 min + 4 hours optimization



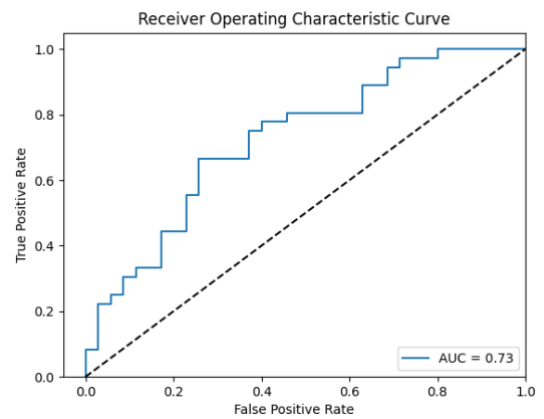
(a)



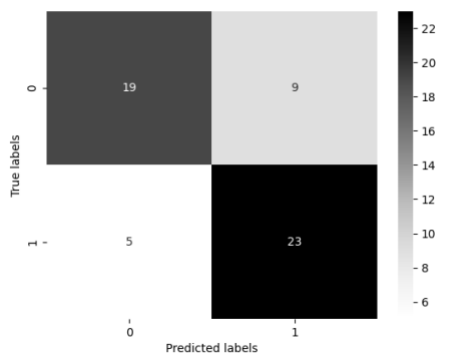
(b)



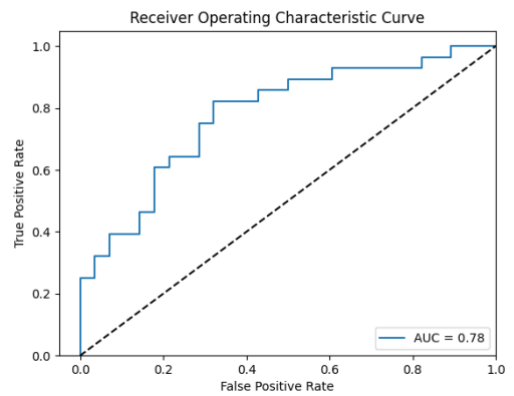
(c)



(d)



(e)



(f)

Figure 4-7: Confusion Matrix and ROC curve for the exacerbation models: (a) fine-tuned ResNet-18 Confusion Matrix, (b) fine-tuned ResNet-18 ROC curve, (c) fine-tuned ResNet-34 Confusion

Matrix, (d) fine-tuned ResNet-18 ROC curve, (e) Exacerbation-CNN Confusion Matrix, (f) Exacerbation-CNN ROC curve.

4.3.3 Evaluation of the GOLD2023 COPD Staging Model

To evaluate the classification performance of the classification algorithm developed for the GOLD2023 COPD staging, we utilized samples from the test sets used to evaluate symptoms and exacerbation models, totaling 98 samples. These samples were distributed as follows: 35 samples for each of classes A and B, and 28 samples for class E of the GOLD 2023 stages. The classification accuracy score was used as the evaluation metric. The classification performance is reported for models trained from scratch and the fine-tuned models ResNet-18 and ResNet34. Table 4-3 shows the accuracy score for all models' configurations. This table shows that, using the Symptoms-CNN model for symptoms assessment and the Exacerbation-CNN model for exacerbation assessment achieved a classification accuracy score of 0.72. The fine-tuned Symptom ResNet-18 and Exacerbation ResNet-18 achieved classification accuracy of 0.70. A value of 0.73 was achieved when using the fine-tuned Symptom ResNet-34 and Exacerbation ResNet-34. The highest classification accuracy is 0.74, which was obtained by using Exacerbation ResNet-18 for exacerbation assessment, and Symptom ResNet-34 for symptoms assessment, these are the two models with the highest classification performance for exacerbation and symptoms, respectively.

Table 4-3: GOLD2023 classification accuracy

Models Configuration	Accuracy score
Exacerbation-CNN + Symptoms-CNN	0.72
Fine-tuned Exacerbation ResNet-18 + Symptom ResNet-18	0.70
Fine-tuned Exacerbation ResNet-34 + Symptom ResNet-34	0.73
Fine-tuned Exacerbation ResNet-18 + Symptom ResNet-34	0.74

4.4 Discussion

Deep learning of lung CT images has been applied to enhance COPD assessment. In this investigation, we developed a deep learning based classification algorithm in conjunction with the recently developed GOLD2023 staging system where a single lung 3D CT scan is used as input to predict the patient's COPD stage. The algorithm utilizes two deep learning classifiers we developed to determine symptom's severity and exacerbation. Different deep learning models have been developed to determine the severity of COPD symptoms. Among these models, the fine-tuned ResNet-34 exhibited the highest classification performance, followed by the Symptoms-CNN model. In the latter, the optimization of model hyperparameters, which was specifically carried out for COPD symptoms determination using limited lung CT data, allowed the model to achieve classification accuracy comparable to that of a prominent deep learning model pretrained with extensive medical datasets.

For predicting the risk of future exacerbation, the fine-tuned ResNet-18 demonstrated the highest classification accuracy, followed by the fine-tuned ResNet-34 and Exacerbation-CNN. While ResNet-18 and ResNet-34 were pretrained with extensive medical datasets, Exacerbation-CNN achieved comparable classification performance due to the optimization of model hyperparameters specifically for COPD exacerbation determination using lung the limited CT data. It is worth mentioning that while the optimized model achieved comparable performance to the pretrained models, it required more computational time and cost as shown in Table 4-2. As such, it is conceivable that both the Symptoms-CNN and Exacerbation-CNN models can lead to substantially higher accuracy if they are trained using substantially larger subset of the COPDGene dataset.

The higher recall of Exacerbation-CNN makes it more effective in identifying true positives, although it may result in more false positives. Fine-tuned ResNet18, on the other hand, excels in distinguishing between classes, as indicated by its higher AUC. However, in a medical diagnosis scenario, it is important to reduce false negatives as overlooking a positive case can have serious consequences if the disease progresses.

Therefore, even though Exacerbation-CNN has a lower AUC, its higher recall makes it more suitable for this task.

It is worth noting that in the context of exacerbation determination, a deeper neural network does not always perform better. In fact, ResNet-18 was found to perform better than ResNet-34. This could be because the underlying pathology of exacerbation, which results from severe lung parenchyma destruction, is simple enough to be effectively determined by a relatively shallower neural network. The developed exacerbation models demonstrate overall higher classification performance compared to the symptom models. This, once again, could be attributed to the severity of lung parenchyma destruction being the primary cause of exacerbation. Lung CT scans, being highly detailed and capable of capturing structural abnormalities, may more effectively detect features indicative of exacerbation related pathologies, which often involves substantial lung tissue damage or inflammation. This emphasizes the importance of leveraging advanced imaging techniques like lung CT scans in developing accurate diagnostic models for COPD exacerbation. In contrary, the underlying pathologies of COPD symptoms are more complex, hence possibly requiring deeper neural networks in conjunction to larger training datasets.

Combing the Symptoms-CNN Exacerbation-CNN models trained with a limited subset of the COPDGene dataset led to an overall accuracy of COPD stage determination of 0.72 while the best accuracy score achieved through transfer learning was 0.74. While these accuracy scores are quite comparable, it is conceivable that training the Symptoms-CNN Exacerbation-CNN models using a substantially larger data subset may improve the accuracy substantially. Among previous different studies that used lung CT to assess the COPD severity, the binary 3D-CNN classifier trained with the PRM maps obtained from paired lung scans to classify subjects to COPD and non-COPD [8] achieved a classification accuracy 0.89. Moreover, the lung air model developed by Moghadas [17] classifies COPD subjects based on a previous four-class GOLD staging system designed using PFT data. That model achieved a classification accuracy of 0.84. While both previous methods achieved high accuracy, it should be noted that they use paired 3D CT images pertaining to inhalation and exhalation phases in contrast to the proposed technique that uses a single 3D CT scan which is more accessible in the clinic. Another classifier is the unsupervised

model of Li et al. [9] which achieved a classification accuracy of 0.66 for classifying COPD patients based on the risk of future exacerbation.

Identifying the severity of COPD exacerbation using image information is crucial as not all patients who experience exacerbation visit the hospital while the tool used to measure COPD exacerbation is based on the history of visiting the hospital. The recent GOLD2023 guidelines recommend classifying patients with COPD by the severity of their symptoms and exacerbations before recommending proper medications following these measurements. The recommendations aim at minimizing disease progression. To our knowledge, the proposed classification algorithm is the first classifier used to perform GOLD2023 COPD staging through inputting a single 3D CT scan. The developed algorithm can potentially serve as a valuable tool for identifying COPD stage/severity based on image data before devising a therapy plan.

4.5 Conclusions

A COPD GOLD2023 staging model was developed using 3D lung CT data in conjunction with deep learning. While sufficiently simple for clinical assessment of COPD, the GOLD2023 staging system can be used effectively for devising treatment. To develop the model, CNNs were developed and utilized along with lung CT scans at expiration phase to detect the two main predictors, symptom and exacerbation, that were then used for COPD staging. Five different CNNs were trained/fine-tuned to detect each of the predictors. For symptoms level determination, the Symptoms-CNN model was developed using residual blocks and model hyper parameters were optimized to obtain the model's optimal parameters. Pretrained ResNet-18 and ResNet-34 were used following two approaches of feature extraction with and without and fine-tuning. Models obtained from both approaches were employed for symptoms detection. Symptoms-CNN and the fine-tuned ResNet-18 were superior to the other approaches for detecting symptoms level. Similarly, the risk of future exacerbation was determined using five models, where four of them were the pretrained ResNet-18 and ResNet-34 employed in the same two approaches of feature extractors with and without fine-tuning. The Exacerbation-CNN model was developed

using residual blocks, with model hyperparameters optimized to obtain the optimal parameters for determining the risk of future exacerbation.

The GOLD2023 classification model was developed by utilizing the models developed for symptoms and exacerbation determination. This classifier showed ability of effective COPD staging with reasonably high accuracy score of 0.74, hence holding the potential to be deployed in clinical settings for COPD assessment using image data.

4.6 References

- [1] A. Agusti *et al.*, “Characterisation of COPD heterogeneity in the ECLIPSE cohort,” *Respir Res*, vol. 11, Sep. 2010, doi: 10.1186/1465-9921-11-122.
- [2] “Global Initiative For Chronic Obstructive Lung Disease Global Strategy For The Diagnosis, Management, And Prevention Of Chronic Obstructive Pulmonary Disease (2023 REPORT),” 2022. [Online]. Available: www.goldcopd.org
- [3] P. Lange *et al.*, “Prediction of the clinical course of chronic obstructive pulmonary disease, using the new GOLD classification: A study of the general population,” *Am J Respir Crit Care Med*, vol. 186, no. 10, pp. 975–981, Nov. 2012, doi: 10.1164/rccm.201207-1299OC.
- [4] J. J. Soler-Cataluña, M. Á. Martínez-García, P. Román Sánchez, E. Salcedo, M. Navarro, and R. Ochando, “Severe acute exacerbations and mortality in patients with chronic obstructive pulmonary disease,” *Thorax*, vol. 60, no. 11, pp. 925–931, Nov. 2005, doi: 10.1136/thx.2005.040527.
- [5] J. Y. So *et al.*, “Population Decline in COPD Admissions During the COVID-19 Pandemic Associated with Lower Burden of Community Respiratory Viral Infections,” *American Journal of Medicine*, vol. 134, no. 10, pp. 1252-1259.e3, Oct. 2021, doi: 10.1016/j.amjmed.2021.05.008.
- [6] G. Gonzalez *et al.*, “Disease staging and prognosis in smokers using deep learning in chest computed tomography,” *American Journal of Respiratory and Critical Care*

- Medicine*, vol. 197, no. 2. American Thoracic Society, pp. 193–203, Jan. 15, 2018. doi: 10.1164/rccm.201705-0860OC.
- [7] S. M. Humphries *et al.*, “Deep learning enables automatic classification of emphysema pattern at CT,” *Radiology*, vol. 294, no. 2, pp. 434–444, 2020, doi: 10.1148/radiol.2019191022.
- [8] T. T. Ho *et al.*, “A 3D-CNN model with CT-based parametric response mapping for classifying COPD subjects,” *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-020-79336-5.
- [9] F. Li *et al.*, “Latent traits of lung tissue patterns in former smokers derived by dual channel deep learning in computed tomography images,” *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-84547-5.
- [10] E. A. Regan *et al.*, “Genetic epidemiology of COPD (COPDGene) study design,” *COPD: Journal of Chronic Obstructive Pulmonary Disease*, vol. 7, no. 1, pp. 32–43, 2010, doi: 10.3109/15412550903499522.
- [11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks.” [Online]. Available: <http://code.google.com/p/cuda-convnet/>
- [12] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” Sep. 2014, [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [13] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition.” [Online]. Available: <http://image-net.org/challenges/LSVRC/2015/>
- [14] S. M. Ryan, B. Vestal, L. A. Maier, N. E. Carlson, and J. Muschelli, “Template Creation for High-Resolution Computed Tomography Scans of the Lung in R Software,” *Acad Radiol*, vol. 27, no. 8, pp. e204–e215, Aug. 2020, doi: 10.1016/j.acra.2019.10.030.
- [15] M. Ibanez. Johnson, *ITK Software Guide: Design and Functionality*”, Fourth. Kitware Inc, 2015.
- [16] S. Chen, K. Ma, and Y. Zheng, “Med3D: Transfer Learning for 3D Medical Image Analysis,” Apr. 2019, [Online]. Available: <http://arxiv.org/abs/1904.00625>

- [17] H. Moghadas-Dastjerdi, M. Ahmadzadeh, E. Karami, M. Karami, and A. Samani, "Lung CT image based automatic technique for COPD GOLD stage assessment," *Expert Syst Appl*, vol. 85, pp. 194–203, Nov. 2017, doi: 10.1016/j.eswa.2017.05.036.

Chapter 5

5 « Summary, Conclusion, and Future Work »

The overarching objective of this thesis is to leverage lung CT imaging in conjunction with deep learning techniques for assessing COPD. This framework utilizes two staging systems: the high-resolution eight-stage system and the recently introduced GOLD2023 three-stage system. The models developed for classifying COPD according to these staging systems are detailed in Chapters 2, 3, and 4 of this thesis. Below, a summary and concluding remarks for each chapter are provided.

5.1 Summary

In Chapters 2 and 3, the focus was on training models to classify COPD using the eight-stage COPD system while Chapter 4 delves explicitly into the GOLD 2023 three-stage COPD staging system.

In **Chapter 2**, we employed the Neural Network (NN) approach in conjunction with features extracted from lung CT scans to develop a classification system founded on the eight-stage COPD system. Two sets of imaging features were derived: lung air volume features and COPD phenotypes features. The lung air volume features included three groups: lung volume variation features, inspiration and expiration air distribution features, and lung overall air distribution features. The phenotype features measured emphysema, air trapping, and function small airway disease (fSAD) using Parametric Response Mapping (PRM). Two NN models were developed for the eight-stage COPD classification system which we referred to as NN-CT and NN-Hybrid. The NN-CT model was trained solely using lung imaging features, while the NN-Hybrid model incorporated two additional features related to exacerbation frequency and symptom level.

The eight-stage COPD system is based on the three severity factors of lung function, exacerbation frequency, and symptom level. As such, the NN-CT model essentially predicted the three severity factors before performing staging accordingly. In contrast, the NN-Hybrid model focused exclusively on predicting lung function, with exacerbation

frequency and symptom level provided as inputs. This decision was made because lung function assessment requires more time and effort to predict among the three severity factors, as it involves spirometry tests or radiologists to identify the severity from CT scans. The results revealed that the NN-CT model had a relatively low classification accuracy. It was noted that most misclassifications occurred among classes that shared severity factors. In order to enhance the model's performance, we proposed integrating exacerbation frequency and symptom level as supplementary inputs. This approach was tested on the NN-Hybrid model, which yielded a significantly higher classification accuracy of 0.88. The relatively low accuracy and high accuracy of the NN-CT and NN-hybrid models indicate that while CT image data was capable of capturing pulmonary function characterized by PFT accurately, it was incapable of capturing symptoms severity or exacerbation frequency accurately. This may have been achievable through the development of separate NNs to predict the latter two severity factors.

Our contribution in Chapter 2 focused on the development of classification systems for the eight-stage COPD system employing the Neural Network (NN) approach. We explored two approaches: one utilizing only lung imaging features and the other incorporating additional features related to exacerbation frequency and symptom level. Both approaches required paired lung CT scans for feature extraction and model development.

In **Chapter 3**, we utilized the Convolutional Neural Network (CNN) approach for COPD eight-stage classification, presenting two CNN models: 3D-CNN and 3D-CNN-Hybrid. The 3D-CNN model used only lung CT scan data as input, while the hybrid model integrated additional features related to symptom level, exacerbation frequency, and lung CT scan data. Consequently, the 3D-CNN model predicted the three severity factors and performed eight-stage classification accordingly, whereas the convolutional layers in 3D-CNN-Hybrid focused exclusively on predicting lung function. In addition to these models, a transfer learning approach was employed for the eight-stage COPD classification. State-of-the-art architectures ResNet-18 and ResNet-34 were utilized, and the models were fine-tuned with training data to perform the classification.

Out of the four CNN models developed for eight-stage COPD, the 3D-CNN_Hybrid model achieved the highest classification accuracy of 0.8. Although the other three models did

not achieve high classification results, they still provided valuable information. These models achieved accuracy levels approximately 4 to 5 times better than random guessing in an eight-class classification. Once again, the observations of the classification models developed in this chapter indicate that the developed single CNN was incapable of capturing symptoms severity or exacerbation frequency accurately as well as lung function through the lung CT data. This may have been achievable through the development of separate NNs to predict the latter two severity factors.

Our contribution in Chapter 3 focuses on utilizing the Convolutional Neural Network (CNN) approach for COPD eight-stage classification, employing either single lung CT scans at the expiration phase or integrating additional features related to symptom level and exacerbation frequency. Additionally, we incorporated transfer learning approach for eight-stage COPD classification, leveraging the 3D versions of state-of-the-art architectures ResNet-18 and ResNet-34, fine-tuned with our training data to perform the classification.

In **Chapter 4**, we applied the Convolutional Neural Network (CNN) approach to classify COPD based on the GOLD2023 staging system. CNN models were developed to predict the two severity factors essential for the GOLD2023 system: symptom severity and the risk of exacerbation. For symptom level prediction, Symptoms-CNN was developed using residual blocks with optimized model hyperparameters. Additionally, pretrained ResNet-18 and ResNet-34 architectures were utilized in two approaches: feature extraction and fine-tuning. Similarly, we utilized five models, including four pretrained ResNet-18 and ResNet-34 models, to detect the risk of future exacerbation. These models were employed in two different ways: as feature extractors and with fine-tuning. We also developed the Exacerbation-CNN, which utilized residual blocks and optimized hyperparameters to detect future exacerbation risks. The GOLD2023 COPD classification algorithm was generated using the developed models for symptom and exacerbation detection.

Among the different models developed for symptom detection and future exacerbation prediction, fine-tuned ResNet-34 exhibited the highest classification performance for detecting symptom severity, while fine-tuned ResNet-18 demonstrated the highest accuracy for predicting future exacerbation risk. The optimized symptom detection and

exacerbation prediction models achieved comparable performance to the pretrained models but required more computational time.

In Chapter 4, we have three main contributions: Firstly, we utilize the Convolutional Neural Network approach to detect the risk of COPD exacerbation. Secondly, we employ the Convolutional Neural Network approach to detect the level of symptoms. Lastly, we utilize the two developed models for exacerbation detection and symptom detection to classify COPD based on the GOLD2023 staging system.

5.2 Conclusions and Future Work

Given that COPD staging systems require physicians to conduct two or three tests to devise an appropriate treatment plan, with available CT scans, the proposed deep learning models can be more effective as it eliminates the need for multiple tests and enables a more efficient and effective treatment plan.

Our novel eight-stage classification models show promising potential in assessing lung function mainly based on image data, particularly the Hybrid models (NN-Hybrid and CNN-Hybrid), which achieved high classification performance by exclusively utilizing lung imaging information to assess lung function. This suggests that COPD patients may not have to undergo pulmonary function tests (PFTs) to determine the severity of their disease, as lung CT scans can automatically assess disease severity. Utilizing lung CT in evaluating lung function offers the advantage of accurately detecting the disease in its early stages and not being dependent on patient compliance.

Comparing the NN-Hybrid and CNN-Hybrid in terms of classification performance, the NN-Hybrid achieved higher classification performance as it was trained with features extracted using lung imaging methods. However, it requires paired lung CT scans at inspiration and expiration breathing phases to extract these features. In contrast, the CNN-Hybrid only requires a single lung CT scan at the expiration phase for classification.

COPD assessments and treatment planning are mainly based on three severity factors: lung function, symptom severity, and the risk of exacerbation. In this thesis, we successfully developed models capable of predicting all three severity factors with reasonably high

performance; the symptom's severity and exacerbation risk have been predicted with different CNN models, while the lung function is implicitly predicted with the NN and CNN parts of the hybrid models, NN-Hybrid 3D-CNN-Hybrid. Predicting the severity of COPD exacerbation from images is crucial. Some patients may not visit the hospital despite experiencing exacerbation, while traditional methods for measuring COPD exacerbation rely on hospital visits.

COPD staging systems have been created using these factors, with the eight-stage COPD system utilizing all three factors of COPD severity, while the GOLD 2023 staging system uses only symptom severity and exacerbation risk. As such, all eight staging models implicitly predict the three severity factors but to various accuracy levels. By utilizing specialized algorithms for lung imaging, the NN-CT model achieved superior classification performance compared to other CNN models trained exclusively on image data. Additionally, the NN-Hybrid model delivered the highest classification performance in comparison to all models, utilizing both solely image data and image data with additional features as input. The algorithms used to extract features in NN-CT and NN-Hybrid models integrate both image information and physics-based knowledge. Specifically, they incorporate the concept of tissue incompressibility and air exchange dynamics between the background air and lung air to extract lung air features. By combining these principles, the algorithms aim to capture essential aspects of lung physiology and pathology, enhancing the models' ability to extract relevant information from CT scans for accurate assessment of lung function and pathology.

Transfer learning has proven to be an effective technique for computer vision tasks, as models trained with transfer learning demonstrate superior classification performance compared to models developed from scratch. In addition, fine-tuning time for transfer learning models is significantly shorter than for models undergoing hyperparameter optimization. It is worth noting that the pretrained models were initially trained with medical images, which made the source dataset similar to a great extent to the target dataset. This similarity likely contributed to the effectiveness of transfer learning in this context, allowing for efficient knowledge transfer from the source domain to the target domain, ultimately improving model performance.

There were some limitations to our study:

- COPDGene, being conducted across multiple clinical centers, resulted in CT scanner heterogeneity, with varying models and manufacturers. Although protocols were designed to minimize this variability, it is possible that slight discrepancies in quantitative imaging measurements may have arisen as a result.
- Another limitation was related to the dataset is that PFT may not be sensitive enough to identify early stage lung pathology. To enhance the accuracy of lung function prediction models utilizing imaging data, a possible solution is to conduct a thorough review of lung CT scans for any signs of lung pathology prior to model training. This approach could enable the models to detect even the most subtle lung abnormalities and significantly improve their sensitivity in predicting lung function. It is an area that warrants further research and attention.

For future works:

- Re-developing hybrid models using the CNN models developed for detecting symptom levels and exacerbation severity, instead of relying on clinical data features. This approach allows us to create fully automated models that are solely based on image data, which eliminates the need for separate tests to evaluate symptoms or exacerbation severity. By integrating these CNN-based assessments directly into the hybrid models, we can simplify the COPD assessment process, potentially improving efficiency and accuracy in diagnosis and treatment planning.
- A large training dataset is always beneficial in deep learning, and our developed models could be further improved by utilizing more samples. Increasing the size of the training dataset can enhance the models' ability to learn complex patterns and generalize well to unseen data. With a larger and more diverse dataset, our models may capture a broader range of variations and nuances present in lung imaging data, leading to improved performance and robustness. Therefore, incorporating a larger number of samples in future iterations of our models can

contribute to their continued advancement and refinement for a more accurate and reliable assessment of lung function and pathology.

Appendices

Appendix A: St. George's Respiratory Questionnaire (SGRQ)

St. George's Respiratory Questionnaire

The St. George's Respiratory Questionnaire (SGRQ) is designed to measure health impairment in patients with asthma and COPD.

Four scores are calculated
Symptoms, Activity, Impacts, and Total.
Lower scores indicate better health.

Maximum raw scores: 662.5, 1209.1, 2117.8, 3989.4, respectively.
Normal (healthy) mean scores: 12, 9, 2, 6, respectively.

Symptoms score uses Part 1, questions 1-8
Activity score uses Part 2, Section 2 and Section 6
Impact score uses Part 2, Sections 1, 3, 4, 5, 7 and 8

Total uses all questions.

Note that each question's responses get separate weighting.

Source of weighted values
St. George's Resp. Questionnaire Manual, 2003
Jones, Paul, et al.
St. George's, University of London
London, SW17 0RE, UK

For more information, the manual is available on-line at
http://www.healthstatus.sgu.ac.uk/SGRQ_download/SGRQ%20Manual.pdf



7144



Interviewer (initials)

COPD Gene ID

Month

Day

Year

Center (eg, NJC)

St. George's Respiratory Questionnaire

This questionnaire is designed to help us learn more about how your breathing is troubling you and how it affects your life. We are using it to find out which aspects of your illness cause you the most problems, rather than what the doctors and nurses think your problems are.

Please read the instructions carefully.
Do not spend too long deciding about your answers.

Part 1

Please describe how often your respiratory problems have affected you over the past 4 weeks.

1. Over the past 4 weeks, I have coughed
 - Almost every day
 - Several days a week
 - A few days a month
 - Only with respiratory infections
 - Not at all

2. Over the past 4 weeks, I have brought up phlegm (sputum)
 - Almost every day
 - Several days a week
 - A few days a month
 - Only with respiratory infections
 - Not at all

3. Over the past 4 weeks, I have had shortness of breath
 - Almost every day
 - Several days a week
 - A few days a month
 - Only with respiratory infections
 - Not at all

7144





7144

COPDGene ID

--	--	--	--	--	--

4. Over the past 4 weeks, I have had wheezing attacks
- Almost every day
 - Several days a week
 - A few days a month
 - Only with respiratory infections
 - Not at all
5. How many times during the past 4 weeks have you suffered from severe or very unpleasant respiratory attacks?
- More than 3 times
 - 3 times
 - 2 times
 - 1 time
 - None of the time
6. How long did the worst respiratory attack last? (Go to question 7 if you did not have a severe attack)
- A week or more
 - 3 or more days
 - 1 or 2 days
 - Less than a day
7. Over the past 4 weeks, in a typical week, how many good days (with few respiratory problems) have you had?
- No good days
 - 1 or 2 good days
 - 3 or 4 good days
 - Nearly every day is good
 - Every day is good
8. If you wheeze, is it worse when you get up in the morning?
- No
 - Yes



7144



7144

COPDGene ID

--	--	--	--	--	--	--	--

Part 2**Section 1**

1. How would you describe your respiratory condition?
 - The most important problem I have
 - Causes me quite a lot of problems
 - Causes me a few problems
 - Causes me no problem

2. If you have ever held a job
 - My respiratory problems made me stop working altogether
 - My respiratory problems interfere with my job or made me change my job
 - My respiratory problems do not affect my job

Section 2

These are questions about what activities usually make you feel short of breath these days.

- | | <i>True</i> | <i>False</i> |
|--|-----------------------|-----------------------|
| 3. Sitting or lying still | <input type="radio"/> | <input type="radio"/> |
| 4. Washing yourself or dressing | <input type="radio"/> | <input type="radio"/> |
| 5. Walking in the house | <input type="radio"/> | <input type="radio"/> |
| 6. Walking outside on level ground | <input type="radio"/> | <input type="radio"/> |
| 7. Walking up a flight of stairs | <input type="radio"/> | <input type="radio"/> |
| 8. Walking up hills | <input type="radio"/> | <input type="radio"/> |
| 9. Playing sports or other physical activities | <input type="radio"/> | <input type="radio"/> |



7144

COPDGene ID

--	--	--	--	--	--	--	--

Section 3

These are more questions about your cough and shortness of breath these days.

- | | <i>True</i> | <i>False</i> |
|--|-----------------------|-----------------------|
| 10. Coughing hurts | <input type="radio"/> | <input type="radio"/> |
| 11. Coughing makes me tired | <input type="radio"/> | <input type="radio"/> |
| 12. I am short of breath when I talk | <input type="radio"/> | <input type="radio"/> |
| 13. I am short of breath when I bend over | <input type="radio"/> | <input type="radio"/> |
| 14. My coughing or breathing disturbs my sleep | <input type="radio"/> | <input type="radio"/> |
| 15. I become exhausted easily | <input type="radio"/> | <input type="radio"/> |

Section 4

These are questions about other effects that your respiratory problems may have on you these days.

- | | <i>True</i> | <i>False</i> |
|---|-----------------------|-----------------------|
| 16. My coughing or breathing is embarrassing in public | <input type="radio"/> | <input type="radio"/> |
| 17. My respiratory problems are a nuisance to my family, friends or neighbors | <input type="radio"/> | <input type="radio"/> |
| 18. I get afraid or panic when I cannot catch my breath | <input type="radio"/> | <input type="radio"/> |
| 19. I feel that I am not in control of my respiratory problems | <input type="radio"/> | <input type="radio"/> |
| 20. I do not expect my respiratory problems to get any better | <input type="radio"/> | <input type="radio"/> |
| 21. I have become frail or an invalid because of my respiratory problems | <input type="radio"/> | <input type="radio"/> |
| 22. Exercise is not safe for me | <input type="radio"/> | <input type="radio"/> |
| 23. Everything seems too much of an effort | <input type="radio"/> | <input type="radio"/> |

Section 5

These are questions about your treatment and medication (including oxygen, inhalers and pills).

a. Are you receiving any treatment for your respiratory problems?

- Yes No If *No* go to *Section 6* (top of next page)

- | | <i>True</i> | <i>False</i> |
|---|-----------------------|-----------------------|
| 24. My treatment does not help me very much | <input type="radio"/> | <input type="radio"/> |
| 25. I get embarrassed using my medication in public | <input type="radio"/> | <input type="radio"/> |
| 26. I have unpleasant side effects from my medication | <input type="radio"/> | <input type="radio"/> |
| 27. My treatment interferes with my life a lot | <input type="radio"/> | <input type="radio"/> |

7144





7144

COPDGene ID

--	--	--	--	--	--

Section 6

These are questions about how your activities might be affected by your respiratory problems.

- | | <i>True</i> | <i>False</i> |
|---|-----------------------|-----------------------|
| 28. I take a long time to get washed or dressed | <input type="radio"/> | <input type="radio"/> |
| 29. I cannot take a bath or shower, or I take a long time to do it | <input type="radio"/> | <input type="radio"/> |
| 30. I walk slower than other people my age, or I stop to rest | <input type="radio"/> | <input type="radio"/> |
| 31. Jobs such as household chores take a long time, or I have to stop to rest | <input type="radio"/> | <input type="radio"/> |
| 32. If I walk up one flight of stairs, I have to go slowly or stop | <input type="radio"/> | <input type="radio"/> |
| 33. If I hurry or walk fast, I have to stop or slow down | <input type="radio"/> | <input type="radio"/> |
| 34. My breathing makes it difficult to do things such as walk up hills, carry things up stairs, light gardening such as weeding, dance, bowl or play golf. | <input type="radio"/> | <input type="radio"/> |
| 35. My breathing makes it difficult to do things such as carry heavy loads, dig in the garden or shovel snow, jog or walk briskly (5 miles per hour), play tennis or swim | <input type="radio"/> | <input type="radio"/> |
| 36. My breathing makes it difficult to do things such as very heavy manual work, ride a bikes, run, swim fast or play competitive sports | <input type="radio"/> | <input type="radio"/> |

Section 7

We would like to know how your respiratory problems usually affect your daily life.

- | | <i>True</i> | <i>False</i> |
|--|-----------------------|-----------------------|
| 37. I cannot play sports or do other physical activities | <input type="radio"/> | <input type="radio"/> |
| 38. I cannot go out for entertainment or recreation | <input type="radio"/> | <input type="radio"/> |
| 39. I cannot go out of the house to do the shopping | <input type="radio"/> | <input type="radio"/> |
| 40. I cannot do household chores | <input type="radio"/> | <input type="radio"/> |
| 41. I cannot move far from my bed or chair | <input type="radio"/> | <input type="radio"/> |



7144



COPDGene ID

--	--	--	--	--	--	--	--

Section 8

Here is a list of other activities that your respiratory problems may prevent you from doing. (You do not have to check these, they are just to remind you of ways your shortness of breath may affect you.)

- Going for walks or walking the dog
- Doing activities or chores at home or in the garden
- Sexual intercourse
- Going to a place of worship, or a place of entertainment
- Going out in bad weather or into smoky rooms
- Visiting family or friends or playing with children

Please write in any other important activities that your respiratory problems may stop you from doing.

42. Now please check the box (one only) that you think best describes how your respiratory problems affect you.

- It does not stop me from doing anything I would like to do
- It stops me from doing one or two things I would like to do
- It stops me from doing most of the things I would like to do
- It stops me from doing everything I would like to do

Thank you for completing this questionnaire. Before you finish, would you please make sure that you have answered all the questions?

Copyright reserved
 P.W. Jones, PhD FRCP
 Professor of Respiratory Medicine,
 St. George's University of London,
 Jenner Wing,
 Cranmer Terrace,
 London SW17 0RE, UK.



Appendix B: Modified Medical Research Council) Dyspnea Scale**Modified MRC (mMRC) Questionnaire**

PLEASE TICK IN THE BOX THAT APPLIES TO YOU
(ONE BOX ONLY)

mMRC Grade 0. I only get breathless with strenuous exercise.

mMRC Grade 1. I get short of breath when hurrying on the level or walking up a slight hill.

mMRC Grade 2. I walk slower than people of the same age on the level because of breathlessness, or I have to stop for breath when walking on my own pace on the level.

mMRC Grade 3. I stop for breath after walking about 100 meters or after a few minutes on the level.

mMRC Grade 4. I am too breathless to leave the house or I am breathless when dressing or undressing.

Appendix C: Permissions for Reproduction of Scientific Articles

3/13/24, 12:13 PM

RightsLink Printable License

WOLTERS KLUWER HEALTH, INC. LICENSE TERMS AND CONDITIONS

Mar 13, 2024

This Agreement between Halimah Alsurayhi ("You") and Wolters Kluwer Health, Inc. ("Wolters Kluwer Health, Inc.") consists of your license details and the terms and conditions provided by Wolters Kluwer Health, Inc. and Copyright Clearance Center.

License Number	5747120423311
License date	Mar 13, 2024
Licensed Content Publisher	Wolters Kluwer Health, Inc.
Licensed Content Publication	WK Health Book
Licensed Content Title	West's Respiratory Physiology
Licensed Content Author	John B. West MD, PhD, DSc, Andrew M. Luks MD
Licensed Content Date	Sep 29, 2020
Type of Use	Dissertation/Thesis
Requestor type	University/College
Sponsorship	No Sponsorship
Format	Electronic
Will this be posted online?	Yes, on a secure website
Portion	Figures/tables/illustrations
Number of figures/tables/illustrations	1

3/13/24, 12:13 PM

RightsLink Printable License

Author of this Wolters Kluwer article	No
Will you be translating?	No
Intend to modify/change the content	No
Current or previous edition of book	Current edition
Title of new work	Machine Learning Classifiers for Chronic Obstructive Pulmonary Disease Assessment Using Lung CT Data
Institution name	WESTERN UNIVERSITY
Expected presentation date	Apr 2024
Portions	HUMAN AIRWAYS GENERATION
	Halimah Alsurayhi 66- 275 Callaway rd
Requestor Location	London, ON N6G 0N8 Canada Attn: Halimah Alsurayhi
Publisher Tax ID	895524239 RT0001
Total	0.00 CAD
Terms and Conditions	

Wolters Kluwer Health Inc. Terms and Conditions

1. **Duration of License:** Permission is granted for a one time use only. Rights herein do not apply to future reproductions, editions, revisions, or other derivative works. This permission shall be effective as of the date of execution by the parties for the maximum period of 12 months and should be renewed after the term expires.
 - i. When content is to be republished in a book or journal the validity of this agreement should be the life of the book edition or journal issue.

- ii. When content is licensed for use on a website, internet, intranet, or any publicly accessible site (not including a journal or book), you agree to remove the material from such site after 12 months, or request to renew your permission license
- 2. **Credit Line:** A credit line must be prominently placed and include: For book content: the author(s), title of book, edition, copyright holder, year of publication; For journal content: the author(s), titles of article, title of journal, volume number, issue number, inclusive pages and website URL to the journal page; If a journal is published by a learned society the credit line must include the details of that society.
- 3. **Warranties:** The requestor warrants that the material shall not be used in any manner which may be considered derogatory to the title, content, authors of the material, or to Wolters Kluwer Health, Inc.
- 4. **Indemnity:** You hereby indemnify and hold harmless Wolters Kluwer Health, Inc. and its respective officers, directors, employees and agents, from and against any and all claims, costs, proceeding or demands arising out of your unauthorized use of the Licensed Material
- 5. **Geographical Scope:** Permission granted is non-exclusive and is valid throughout the world in the English language and the languages specified in the license.
- 6. **Copy of Content:** Wolters Kluwer Health, Inc. cannot supply the requestor with the original artwork, high-resolution images, electronic files or a clean copy of content.
- 7. **Validity:** Permission is valid if the borrowed material is original to a Wolters Kluwer Health, Inc. imprint (J.B Lippincott, Lippincott-Raven Publishers, Williams & Wilkins, Lea & Febiger, Harwal, Rapid Science, Little Brown & Company, Harper & Row Medical, American Journal of Nursing Co, and Urban & Schwarzenberg - English Language, Raven Press, Paul Hoeber, Springhouse, Ovid), and the Anatomical Chart Company
- 8. **Third Party Material:** This permission does not apply to content that is credited to publications other than Wolters Kluwer Health, Inc. or its Societies. For images credited to non-Wolters Kluwer Health, Inc. books or journals, you must obtain permission from the source referenced in the figure or table legend or credit line before making any use of the image(s), table(s) or other content.
- 9. **Adaptations:** Adaptations are protected by copyright. For images that have been adapted, permission must be sought from the rightsholder of the original material and the rightsholder of the adapted material.
- 10. **Modifications:** Wolters Kluwer Health, Inc. material is not permitted to be modified or adapted without written approval from Wolters Kluwer Health, Inc. with the exception of text size or color. The adaptation should be credited as follows: Adapted with permission from Wolters Kluwer Health, Inc.: [the author(s), title of book, edition, copyright holder, year of publication] or [the author(s), titles of article, title of journal, volume number, issue number, inclusive pages and website URL to the journal page].
- 11. **Full Text Articles:** Republication of full articles in English is prohibited.
- 12. **Branding and Marketing:** No drug name, trade name, drug logo, or trade logo can be included on the same page as material borrowed from *Diseases of the Colon & Rectum, Plastic Reconstructive Surgery, Obstetrics & Gynecology (The Green Journal), Critical Care Medicine, Pediatric Critical Care Medicine, the American Heart Association publications and the American Academy of Neurology publications.*
- 13. **Open Access:** Unless you are publishing content under the same Creative Commons license, the following statement must be added when reprinting material in Open Access journals: "The Creative Commons license does not apply to this content. Use of the material in any format is prohibited without written permission from the publisher, Wolters Kluwer Health, Inc. Please contact permissions@lww.com for further information."
- 14. **Translations:** The following disclaimer must appear on all translated copies: Wolters Kluwer Health, Inc. and its Societies take no responsibility for the accuracy of the translation from the published English original and are not liable for any errors which may occur.

15. **Published Ahead of Print (PAP):** Articles in the PAP stage of publication can be cited using the online publication date and the unique DOI number.
 - i. **Disclaimer:** Articles appearing in the PAP section have been peer-reviewed and accepted for publication in the relevant journal and posted online before print publication. Articles appearing as PAP may contain statements, opinions, and information that have errors in facts, figures, or interpretation. Any final changes in manuscripts will be made at the time of print publication and will be reflected in the final electronic version of the issue. Accordingly, Wolters Kluwer Health, Inc., the editors, authors and their respective employees are not responsible or liable for the use of any such inaccurate or misleading data, opinion or information contained in the articles in this section.
16. **Termination of Contract:** Wolters Kluwer Health, Inc. must be notified within 90 days of the original license date if you opt not to use the requested material.
17. **Waived Permission Fee:** Permission fees that have been waived are not subject to future waivers, including similar requests or renewing a license.
18. **Contingent on payment:** You may exercise these rights licensed immediately upon issuance of the license, however until full payment is received either by the publisher or our authorized vendor, this license is not valid. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of Wolters Kluwer Health, Inc.'s other billing and payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.
19. **STM Signatories Only:** Any permission granted for a particular edition will apply to subsequent editions and for editions in other languages, provided such editions are for the work as a whole in situ and do not involve the separate exploitation of the permitted illustrations or excerpts. Please view: [STM Permissions Guidelines](#)
20. **Warranties and Obligations:** LICENSOR further represents and warrants that, to the best of its knowledge and belief, LICENSEE's contemplated use of the Content as represented to LICENSOR does not infringe any valid rights to any third party.
21. **Breach:** If LICENSEE fails to comply with any provisions of this agreement, LICENSOR may serve written notice of breach of LICENSEE and, unless such breach is fully cured within fifteen (15) days from the receipt of notice by LICENSEE, LICENSOR may thereupon, at its option, serve notice of cancellation on LICENSEE, whereupon this Agreement shall immediately terminate.
22. **Assignment:** License conveyed hereunder by the LICENSOR shall not be assigned or granted in any manner conveyed to any third party by the LICENSEE without the consent in writing to the LICENSOR.
23. **Governing Law:** The laws of The State of New York shall govern interpretation of this Agreement and all rights and liabilities arising hereunder.
24. **Unlawful:** If any provision of this Agreement shall be found unlawful or otherwise legally unenforceable, all other conditions and provisions of this Agreement shall remain in full force and effect.

For Copyright Clearance Center / RightsLink Only:

1. **Service Description for Content Services:** Subject to these terms of use, any terms set forth on the particular order, and payment of the applicable fee, you may make the following uses of the ordered materials:
 - i. **Content Rental:** You may access and view a single electronic copy of the materials ordered for the time period designated at the time the order is placed. Access to the materials will be provided through a dedicated content viewer or other portal, and access will be discontinued upon expiration of the

3/13/24, 12:13 PM

RightsLink Printable License

designated time period. An order for Content Rental does not include any rights to print, download, save, create additional copies, to distribute or to reuse in any way the full text or parts of the materials.

- ii. **Content Purchase:** You may access and download a single electronic copy of the materials ordered. Copies will be provided by email or by such other means as publisher may make available from time to time. An order for Content Purchase does not include any rights to create additional copies or to distribute copies of the materials

Other Terms and Conditions:

v1.18

Questions? customercare@copyright.com.

Curriculum Vitae

Name: Halimah Alsurayhi

Post-secondary Education and Degrees: University of Umm Al-Qura
Mecca, Saudi Arabia
2008-2013 B.Sc. (Computer Science)

The University of Western Ontario
London, Ontario, Canada
2016-2019 M.Sc. (Computer Science)

The University of Western Ontario
London, Ontario, Canada
2019-2024 Ph.D. (Computer Engineering)

Honours and Awards: Saudi Arabia Graduate Research Scholarship
2015- 2019, 2019-2024

Related Work Experience: Teaching Assistant
The University of Western Ontario
2017, 2023

Teaching Assistant
Umm AlQura University, Mecca
2013 - 2015

Publications:

H. Alsurayhi and A. Samani, "Deep Learning Model for COPD Classification/Staging Using Lung CT", CMBES Proc., vol. 45, May 2023.

H. Alsurayhi and A. Samani, “Novel COPD Classification/Staging Technique Using Lung CT Data and Neural Network”, 21st Annual Imaging Network Ontario Symposium Proceedings, March 2023

Alsurayhi, H., & El-Sakka, M. R. (2019). An Improved Weighted Average Reprojection Image Denoising Method. In *Image Analysis and Recognition: 16th International Conference, ICIAR 2019, Waterloo, ON, Canada, August 27–29, 2019, Proceedings, Part I 16* (pp. 207-215). Springer International Publishing.