People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

University M'Hamed BOUGARA – Boumerdès

Institute of Electrical and Electronic Engineering

Department of Electronics

Project Report Presented in Partial Fulfilment of

the Requirements of the Degree of

'MASTER'

In Computer Engineering

Title:

Chest Medical Image Classification using Deep Neural Network

Presented By:

- **Mohammed Amine TIS**

Supervisor:

Pr. Abdelhamid DAAMOUCHE

Registration number: ……./2023

Abstract

Human lung which is among the most important parts in human body is facing mortal diseases especially after the COVID-19 pandemic. The scientific world is rapidly developing the health-care field to face these disorders and save millions of lives all around the world. The primary objective was to find a precise and efficient strategy for the accurate and early detection and classification of lung diseases. To achieve this goal, we used the power of two essential medical imaging techniques: computerized tomography (CT-scan) and X-ray imaging. Additionally, we employed three deep learning models: Inception-v3, ResNet, and DenseNet, coupled with two distinct classification ; binary classification and multi-class classification. Our research journey started with binary classification, focusing on distinguishing between COVID-19 and non COVID-19, using both CT-scan and X-ray datasets in total of 17,599, all three models delivered outstanding results, with the highest accuracy reaching an impressive accuracy of 96%, achieved by DenseNet using CT-scan images. These results underscore the potential of deep learning in helping healthcare professionals with highly accurate disease classification. Shifting to the multi-class classification dictated by the need for a more comprehensive and realistic approach to diagnosing and identifying a wide range of medical conditions in clinical practice and research. The new class added to COVID-19, non COVID-19 is: Community-acquired pneumonia (CAP), in total of 17,104 CT-scan images,and using the same models we challenged the system using different splitting data ratios. Through a series of experiments and evaluations, our system achieves an overall accuracy of 98% in classifying chest images across multiple categories, using DenseNet model and the 80:10:10 splitting ratio. The results showcase the significant potential of deep learning in assisting healthcare.

Acknowledgment

I would like to express my sincere gratitude to my teacher and supervisor **Pr. Abdelhamid DAAMOUCHE** for his support, guidance, and for allowing me to duly embrace this project under his auspices despite his busy agenda. He has been an ideal supervisor and mentor. I am proud of and grateful for, my time working with him. I am thankful for his confidence and the freedom he gave me throughout the course of this work.

I am deeply grateful to the members of the jury for taking the time to read and analyze this project report. My warmest thanks go to the teachers of IGEE for the devotion they have shown and the knowledge they have passed to me.

Dedication

To my beloved **Parents**, who have always been the guiding stars of my life. Your unwavering love, sacrifices, and belief in my dreams have shaped me into the person I am today. This thesis is a tribute to your boundless support.

And to my dear **Grandmother**, whose wisdom and love have been a guiding light throughout my life. Your stories and experiences have enriched my perspective and added depth to my journey.

To my three remarkable brothers, **Abderrezak**, **Riyadh**, and **Zakariya**, whose constant encouragement and sibling camaraderie have filled my life with joy and purpose.

To my amazing roommates, **Tarek**, **Yaakoub**, and **Taha**, who have turned our shared space into a home and provided the laughter and support that made late-night study sessions bearable.

To my friends and classmates, who have embarked on this academic journey alongside me, your collective wisdom, camaraderie, and shared experiences have made the pursuit of knowledge all the more fulfilling.

This thesis stands as a tribute to the incredible network of love and support that surrounds me. Each of you has played a unique and invaluable role in my life, and I am profoundly grateful for your presence.

With heartfelt thanks,

Amine.

Contents

List of Figures

List of Tables

List of Abbreviations

AI Artificial Intelligence

ANNs Artificial Neural Networks

APIs Application Programming Interface

BUN Blood Urea Nitrogen

CAD Computer-Aided Design

CAP Community Acquired Pneumonia

CBC Complete Blood Cell

CNN Convolutional Neural Networks

COPD Chronic Obstructive Pulmonary Disease

CSV Comma-separated values

CT scan Computerized Tomography scan

DBN Deep Belief Network

DL Deep Learning

DNN deep Neural Network

FDA Food and Drug Administration

GPUs Graphic processing units

JSON JavaScript Object Notation

K-NN K-Nearest Neighbors

ML Machine Learning

MLP Multilayer Perception

MRI Magnetic Resonance Imaging

NAAT Nucleic Acid Amplification Test

NSCLC Non-Small Cell Lung Cancer

PCR Polymerase Chain Reaction

PET Positron Emission Tomography

PFA Power Flow Analysis

SCLC Small Cell Lung Cancer

SVM Support Vector Machines

TPUs Tensor processing unit

V/Q Ventilation-Perfusion

WHO World Health Organization

XML Extensible Markup Language

General introduction

The human respiratory system is a marvel of biological engineering, intricately created to perform the vital function of oxygen exchange, which is fundamental for sustaining life. This complex system includes a network of airways, including the bronchi and alveoli, as well as a complex set of physiological mechanisms that allow the exchange of oxygen and carbon dioxide to occur efficiently. However, despite its remarkable design and functionality, the respiratory system is remarkably delicate and vulnerable to a wide array of diseases and disorders that can disrupt its intricate workings.

Within this context, this work embarks on a profound exploration of the complexities that surround lung diseases, peeling back the layers to unveil the intricacies of this vital organ. It goes deeply into the anatomy of the lungs, revealing the construction and function of these amazing organs. By going into the anatomical intricacies, this work aims to give a comprehensive foundation for understanding the wide range of disorders that can affect these critical structures.

Furthermore, this exploration extends to the diverse origins of lung ailments. Lung diseases can arise from a multitude of factors, including infectious agents such as bacteria, viruses, and fungi, as well as environmental factors like exposure to pollutants or carcinogens. Notably, the global COVID-19 pandemic has highlighted the profound impact that infectious agents can have on the respiratory system. Additionally, lung cancer, one of the most prevalent forms of cancer worldwide, underscores the significance of understanding how environmental factors can lead to devastating lung conditions.

The emergence of medical imaging has revolutionized the field of medical diagnosis and treatment, allowing healthcare professionals to visualize and assess the condition of the lungs in unprecedented detail. Moreover, the revolutionary impact of deep learning within the field of medical imaging is highlighted, showcasing how advanced computer algorithms, particularly deep neural networks, have transformed our ability to interpret complex medical images.

As a result of this extensive exploration, the thesis culminates in an in-depth analysis of chest CT scan image classification. This represents a state-of-the-art application of deep learning, where deep neural networks are employed to decipher and categorize pulmonary conditions, by using the power of artificial intelligence and machine learning, this work tries to show how artificial intelligence and machine learning have the potential to increase the precision and effectiveness of lung disease diagnosis, eventually leading to better patient treatment

and better health outcomes.

This report is divided into the following sections:

Chapter 01: It sets the stage for this journey into the complexities of lung diseases. It begins by delving into the intricate anatomy of human lungs, highlighting their essential role in oxygen exchange. Moving forward, it explores the diverse causes behind lung diseases, encompassing everything from infectious agents to environmental factors, emphasizing the global health impact. Moreover, this chapter categorizes these conditions, shedding light on their significance in the realm of pulmonary health. Ultimately, it underscores the critical importance of precise diagnostics as a cornerstone for effective management, serving as a crucial foundation for our exploration into the world of lung diseases.

Chapter 2: It delves into the transformative landscape of medical imaging, elucidating its pivotal role in precision medicine and disease detection. As science and technology advance, the manual interpretation of complex medical imaging data becomes increasingly challenging. Recognizing this, radiologists turn to Computer-Aided Design (CAD) systems as indispensable tools for image analysis. Within this context, Deep Learning takes center stage, progressively supplanting traditional machine learning techniques. Deep Learning methods, particularly deep neural networks, emerge as indispensable tools for tasks like image classification, often achieving remarkable accuracy comparable to human performance. Machine learning techniques empower computers to autonomously learn from data, enabling them to generate algorithms and make predictions based on information gleaned. The chapter explores the rapid evolution of deep learning techniques over the past decade, driven by advancements in computational capabilities. It highlights the role of Artificial Neural Networks (ANNs) as the foundation of supervised learning, laying the groundwork for deep architectures that enhance information processing. The chapter culminates in an exploration of Deep Neural Network (DNN) techniques, aimed at simplifying mapping functions for human utilization.

Chapter 3: Presents the project's approach and implementation, highlighting the tools, techniques,performance metrics, and transfer learning using specific architectures.

Chapter 4: It represents the highest point of our chest classification research. Using deep neural networks such as ResNet, Inception-V3, and DenseNet on a Kaggle datasets, including binary and multi-class classification. The main objective is to evaluate these models, classification methods, and medical imaging techniques methodically in order to determine the most accurate and efficient strategies. The results have profound implications, particularly in healthcare, where accurate and timely diagnosis is important to patient care and outcomes. This chapter discusses the study's results and shows the significance of correct early diagnosis, which may have a substantial influence on patient lives and healthcare delivery.

Finally, a general conclusion is given with the contributions and the possible further work.

Chapter 1

Lung diseases overview

1.1 Introduction

Lung disease refers to several types of diseases that damage the function of lungs. Lung disease can affect respiratory function, or the ability to breathe, and pulmonary function, which is how well lungs function.

There are many different lung diseases, some of which are caused by bacterial, viral, or fungal infections. Other lung diseases are associated with environmental factors, including Lung cancer, COVID-19, Community Acquired Pneumonia (CAP) [1].

Both COVID-19 and CAP primarily affect the lungs, causing inflammation, damage to lung tissue, and impaired gas exchange. Early detection and appropriate management are very important for both conditions to prevent complications and improve patient prognosis.

This chapter lays the groundwork by exploring the fundamental anatomy of human lungs, delving into the diverse causes of lung diseases, and providing an overview of the distinct categories of lung disorders. it navigates through common lung diseases, shedding light on their significance in the context of global health. it also underscores the critical importance of accurate and efficient diagnostic methods in tackling the challenges posed by these intricate conditions.

1.2 Literature review

In the ever-evolving landscape of medical science, the quest for effective and accurate methods of diagnosing lung diseases stands as a testament to our commitment to improving patient care and saving lives. This comprehensive literature review delves into the historical evolution, existing research, and recent advancements in the diagnosis of lung diseases, with a particular focus on the pivotal role played by radiological imaging techniques.

- **Historical Perspective:** The journey toward understanding and diagnosing lung diseases

traces its roots back to ancient medical practices, where clinical observations and rudimentary physical examinations formed the bedrock of diagnostic approaches. Over the centuries, this practice evolved, marked by seminal contributions such as René Laennec's invention of the stethoscope in 1816 , which revolutionized the physical examination of the chest [2].

- **Evolution of Radiological Imaging:** The seismic shift in lung disease diagnosis occurred with the discovery of X-rays by Wilhelm Conrad Roentgen in 1895 [3]. This groundbreaking discovery provided the medical community with a non-invasive means to visualize the internal structures of the human body. Subsequent developments, including the advent of computed tomography (CT) scans, transformed the diagnostic landscape further, allowing for the creation of detailed cross-sectional views of the lungs and surrounding tissues. In this section, we focus on the profound impact of these innovations on the field of medical diagnostics.

- **Previous Studies and Research:** In the quest for more accurate and effective diagnostic tools for lung diseases, numerous studies and research papers have contributed significantly. These studies range from the identification of distinct patterns in chest X-rays indicative of specific lung conditions to the quantification of nodule characteristics using advanced CT scan technologies. The collective knowledge derived from these studies has not only deepened our understanding of lung diseases but has also laid the foundation for contemporary diagnostic practices.

- **Advancements in Radiology:** Recent years have witnessed remarkable advancements in radiological imaging, driven by technological innovations. Enhanced image resolution, reduced radiation exposure, and the introduction of cutting-edge imaging modalities have revolutionized lung disease diagnosis. These advancements have not only improved diagnostic accuracy but have also made diagnostic methods more patient-friendly and accessible. In this section, we explore the latest developments in radiology and their implications for lung disease diagnosis.

- **Role of Radiology in Lung Disease Diagnosis:** Radiology, particularly X-rays and CT scans, occupies a central and indispensable role in the diagnosis of lung diseases. These techniques provide precise, detailed images of the lungs and surrounding structures, aiding healthcare professionals in accurate diagnosis and treatment planning. Throughout this section, we highlight specific cases and instances where radiological imaging has played a transformative role in diagnosing lung diseases, underscoring its clinical significance.

- **Challenges and Future Directions:** Despite the manifold advantages of radiological imaging in lung disease diagnosis, challenges persist. False positives, cost considerations, and disparities in accessibility remain significant concerns. Moreover, the emergence of artificial intelligence (AI) and machine learning (ML) technologies offers promise in addressing some of these challenges. We explore the potential of AI and ML in improving

diagnostic accuracy and efficiency. Additionally, we outline areas for future research and development, offering insights into the lung disease diagnostics.

1.3 Human Lungs anatomy

The lungs are the major organs of the respiratory system, and are divided into sections, or lobes. The right lung has three lobes and is slightly larger than the left lung, which has two lobes.

The lungs are separated by the mediastinum. This area contains the heart, trachea, esophagus, and many lymph nodes. The lungs are covered by a protective membrane known as the pleura and are separated from the abdominal cavity by the muscular diaphragm [4]. Figure1.1 represents the anatomy of human lungs.

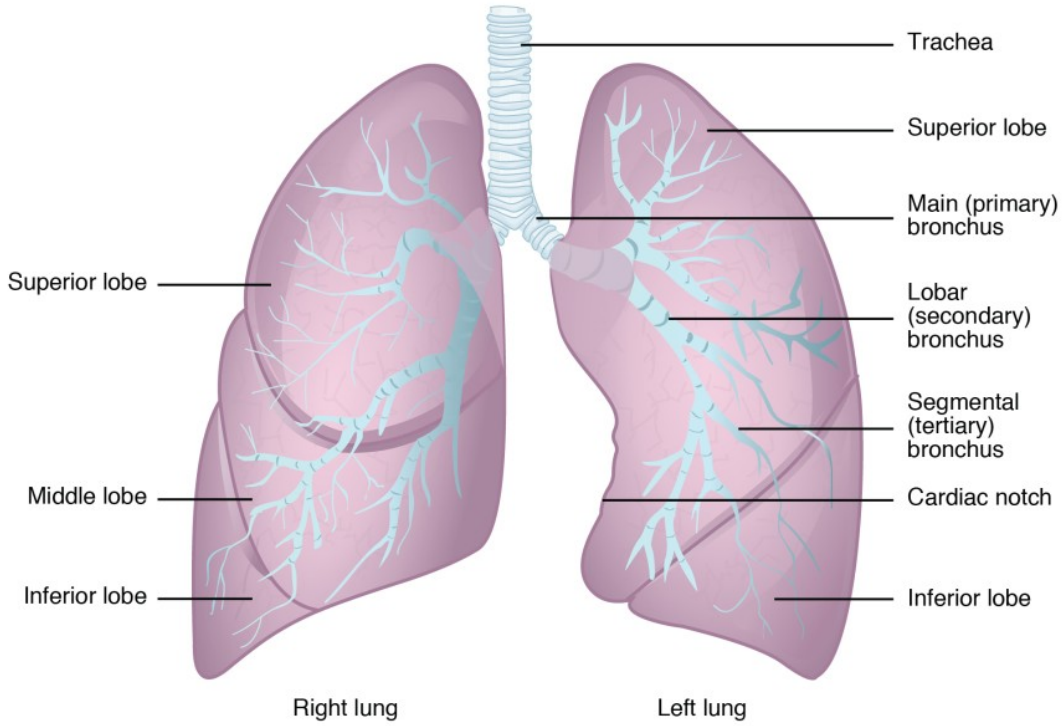


Figure 1.1: Human Lungs anatomy [5].

The right lung is divided into three LOBES or sections. Each lobe is like a balloon filled with tissue. Air moves in and out through one opening—a branch of the bronchial tube. The left lung, on the other hand, is divided into two LOBES. The PLEURA are the two membranes, actually, one continuous one folded on itself, that surround each lobe of the lungs and separate the lungs from your chest wall. The bronchial tubes are lined with CILIA (like very small hairs) that move like waves. This motion carries MUCUS (sticky phlegm or liquid) upward and out into the throat, where it is either coughed up or swallowed. Mucus catches and holds much of the dust, germs, and other unwanted matter that has invaded the lungs. This matter is ejected outside the body through coughing, sneezing, clearing the throat, or swallowing.

The smallest branches of the bronchial tubes are called BRONCHIOLES, at the end of which are the air sacs or alveoli, these lattes are the very small air sacs where the exchange of oxygen and carbon dioxide takes place. CAPILLARIES are blood vessels in the walls of the alveoli. Blood passes through the capillaries, entering through the PULMONARY ARTERY and leaving via the PULMONARY VEIN. While in the capillaries, blood gives off carbon dioxide through the capillary wall into the alveoli and takes up oxygen from air in the alveoli [5].

1.4 Lung disease causes

The affect of disease on health is fast expanding as a result of environmental changes, climate change, lifestyle, and other reasons. In 2016, over 3.4 million people died from chronic obstructive pulmonary disease (COPD), which is typically caused by pollution and smoking, while 400,000 people passed away from asthma.

The danger of lung diseases is substantial, especially in developing and low middle income countries, where millions of people are facing daily the poverty and air pollution. According to the estimation of World Health Organization (WHO), over 4 million premature deaths occur annually from household air pollution-related diseases, including asthma, and pneumonia. Hence, it is necessary to take necessary steps to reduce air pollution and carbon emission. It is also essential to implement efficient diagnostic systems which can assist in detecting lung diseases. Since late December 2019, a novel coronavirus disease 2019 (COVID-19) has been causing serious lung damage and breathing problems. In addition, pneumonia, a form of lung disease can be due to the causative virus of COVID-19 or may be caused by other viral or bacterial infection. Hence, early detection of lung diseases has become more important than ever [6].

The main causes of lung disorders are the following:

- Diseases.
- Radiation.
- Aging.
- Pollution.
- Environment.
- Genetics.
- Asbestos (silicates).
- Tobacco (smoking/second hand).

1.5 Lung diseases

Lung disease, as previously stated in the introduction, is any disorder associated with the lungs that stops them from functioning normally. Airway diseases, lung tissue diseases, and lung circulation diseases are the three primary categories of lung disease. Each of these diseases has a unique set of symptoms, testing procedures, and therapies.

1.5.1 Airway diseases

These diseases affect the tubes (airways) that carry oxygen and other gases into and out of the lungs. They usually cause a narrowing or blockage of the airways. Airway diseases include asthma, COPD, bronchiolitis, and bronchiectasis (which also is the main disorder for persons with cystic fibrosis). People with airway diseases often say they feel as if they are "trying to breathe out through a straw" [7].

1.5.2 Lung tissue diseases

These diseases affect the structure of the lung tissue. Scarring or inflammation of the tissue makes the lungs unable to expand fully (restrictive lung disease). This makes it hard for the lungs to take in oxygen and release carbon dioxide. People with this type of lung disorder often say they feel as if they are "wearing a too-tight sweater or vest." As a result, they can't breathe deeply. Pulmonary fibrosis and sarcoidosis are examples of lung tissue disease [7].

1.5.3 Lung circulation diseases

These diseases affect the blood vessels in the lungs. They are caused by clotting, scarring, or inflammation of the blood vessels. They affect the ability of the lungs to take up oxygen and release carbon dioxide. These diseases may also affect heart function. An example of a lung circulation disease is pulmonary hypertension. People with these conditions often feel very short of breath when they exert themselves [7].

1.6 Common Lung Diseases

As previously explained, lung diseases encompass a diverse range of conditions that can vary significantly depending on their underlying cause and the affected area within the respiratory system. For the purpose of this study, the focus is specifically on three prominent respiratory illnesses: Lung Cancer, COVID-19, and CAP. These diseases have garnered considerable attention due to their widespread impact on public health, making them highly relevant subjects for investigation and analysis in the context of lung disease detection using deep learning methods. By narrowing the scope, the study aims to contribute valuable insights into the accurate and efficient detection of these diseases, which can have significant implications for patient management, public health preparedness, and medical research.

1.6.1 Lung Cancer

Lung cancer is a disease caused by uncontrolled cell division in the lungs. The cells divide and replicate themselves as a part of their normal function. But sometimes, they get changes (mutations) that cause them to keep making more of themselves when they should not. Damaged cells dividing uncontrollably create masses, or tumors, of tissue that eventually keep the organs from functioning properly.

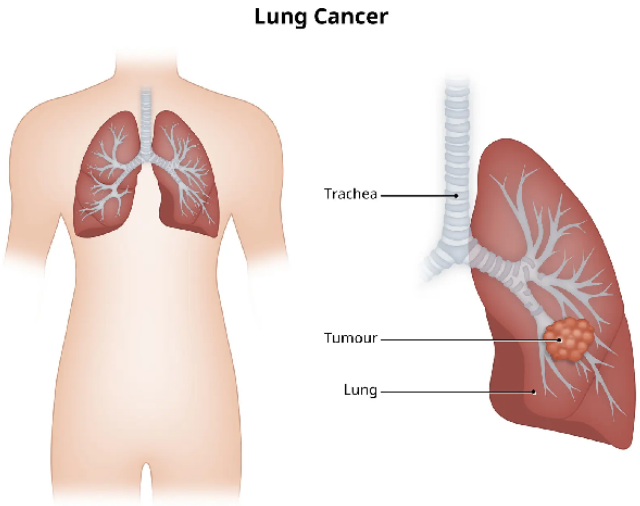


Figure 1.2: Lung Tumour [8].

1.6.1.1 Lung Cancer symptoms

Lung cancer is generally characterized by the absence of noticeable signs and symptoms during its initial stages. The manifestation of indicators and clinical manifestations of lung cancer usually becomes apparent as the disease progresses to an advanced stage.

The clinical presentation of lung cancer includes:

- An enduring cough that fails to resolve over time.
- Coughing up blood, even a small amount.
- Shortness of breath.
- Chest pain.
- Hoarseness.
- Losing weight without trying.
- Bone pain.
- Headache.

1.6.1.2 Types of Lung Cancer

There are many cancers that affect the lungs. However, the term “lung cancer” is used for two main kinds: non-small cell lung cancer and small cell lung cancer.

- **Non-small cell lung cancer (NSCLC).**
- **Small cell lung cancer (SCLC).**

It is worth mentioning that other types of cancer can start in or around the lungs, including lymphomas (cancer in the lymph nodes), sarcomas (cancer in the bones or soft tissue) and

pleural mesothelioma (cancer in the lining of the lungs). These are treated differently and usually are not referred to as lung cancer.

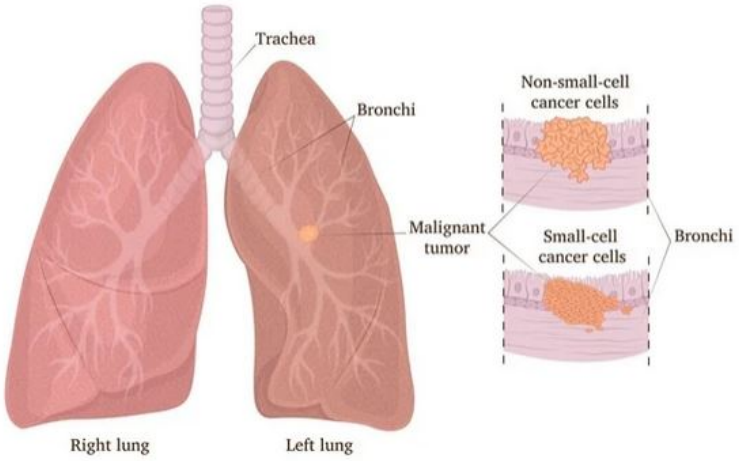


Figure 1.3: Histological types of Lung Cancer [8].

1.6.1.3 Diagnosis tests for Lung Cancer

- **Blood tests:** Blood tests can not diagnose cancer on their own, but, can help the provider check how the organs and other parts of the body are working.
- **Imaging:** Chest X-rays and CT scans give the provider images that can show changes in the lungs. PET/CT scans are usually done to evaluate a concerning finding on a CT scan or after a cancer diagnosis to determine whether cancer has spread.
- **Biopsy:** There are a number of procedures the provider can use to look more closely at what is going on inside the chest. During the same procedures, the provider can take samples of tissue or fluid (biopsy), which can be studied under a microscope to look for cancer cells and determine what kind of cancer it is. Samples can also be tested for genetic changes (mutations) that might affect the treatment.
- **Molecular tests:** As part of a biopsy, the provider might have the tissue sample tested for gene changes (mutations) that special drugs can target as part of the treatment plan.

1.6.2 COVID-19

COVID-19 is a disease caused by a virus named SARS-CoV-2. It is very infectious and spreads rapidly. COVID-19 has killed over one million individuals in the United States.

COVID-19 most often causes respiratory symptoms that can feel much like a cold, the flu, or pneumonia. COVID-19 may cause damage to other sections of your body after attacking your lungs and respiratory system. The majority of COVID-19 patients have mild symptoms, however some develop serious illness.

Some persons, even those with no or minor symptoms will develop Post-COVID Conditions, often known as "Long COVID." [9]

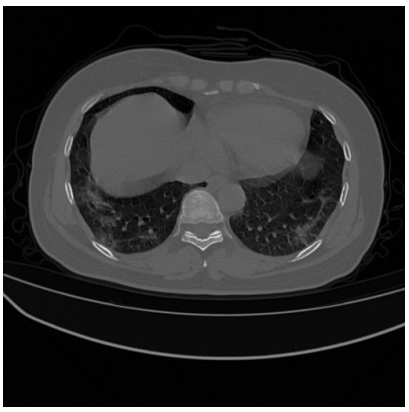


Figure 1.4: CT scan for a COVID sample [10].

1.6.2.1 COVID-19 symptoms

People with COVID-19 have reported experiencing a wide range of symptoms, from little discomfort to serious sickness. 2 to 14 days after viral contact, symptoms may start to show [11]. The symptoms include:

- Fever or chills.
- Cough.
- Shortness of breath or difficulty breathing.
- Muscle or body aches.
- Headache.
- New loss of taste or smell.
- Sore throat.
- Congestion or runny nose.
- Nausea or vomiting.
- Diarrhea.

1.6.2.2 Covid-19 testing

- **PCR test:** PCR tests are the “gold standard” for COVID-19 tests. They are a type of Nucleic Acid Amplification Test (NAAT), which are more likely to detect the virus than antigen tests, As shown in Figure 1.5a. The sample will usually be taken by a healthcare provider and transported to a laboratory for testing. It may take up to 3 days to receive results [12].
- **Antigen Tests:** antigen tests are rapid tests that usually produce results in 15-30 minutes. Positive results are very accurate and reliable. However, in general, antigen tests are less likely to detect the virus than PCR tests, especially when symptoms are not present. Therefore, a single negative antigen test cannot rule out infection. Food and Drug Administration (FDA) recommends 2 negative antigen tests for individuals with symptoms or 3 antigen

tests for those without symptoms, performed 48 hours apart. A single PCR test can be used to confirm an antigen test result [12]. As shown in Figure 1.5b

Figure 1.5 illustrates the two Covid-19 tests.

a PCR test. **b** Antigen test.

Figure 1.5: Covid-19 tests [12].

1.6.3 Community-acquired pneumonia (CAP)

CAP is one of the most frequent infectious diseases and a major source of death and morbidity around the world. Streptococcus pneumonia, Haemophilus influenza, and Moraxella catarrhalis are common bacterial infections that cause CAP. However, with the development of new diagnostic technology, viral respiratory infections are becoming more common causes of CAP. Human rhinovirus and influenza are the most prevalent viral infections isolated from hospitalized CAP patients [13].

1.6.3.1 CAP symptoms

Historical clues and physical examination findings may suggest a causative pathogen, but the clinical signs and symptoms of CAP are not sufficiently specific to reliably differentiate the exact etiologic agent. Therefore, additional testing remains necessary to identify the pathogen and to optimize therapy in CAP. However, extra-pulmonary signs and symptoms seen in some forms of atypical CAP may include the following:

- Mental confusion.
- Myalgias.
- Ear pain.
- Abdominal pain.
- Diarrhea.
- Rash (Horder spots in psittacosis; erythema multiforme in Mycoplasma pneumonia).
- Nonexudative pharyngitis.
- Hemoptysis.
- Splenomegaly.
- Relative bradycardia.

1.6.3.2 CAP diagnosis

In the field of diagnosing CAP, a range of standard diagnostic studies play a critical role in accurately identifying and differentiating this common respiratory infection. These essential tests assist healthcare workers in determining the existence of infiltrates, excluding out other similar diseases and determining the severity of the infection. The primary diagnostic modalities include the following:

- Chest radiography (X-rays and CT scans).
- Complete Blood Cell (CBC) count with differential.
- Serum Blood Urea Nitrogen (BUN) and creatinine levels.

Since the main topic of this study revolves around chest scans, the focus will be solely on chest radiography in the context of diagnosing CAP. Chest radiographs serve as a fundamental diagnostic tool, allowing for the assessment of infiltrates and the exclusion of conditions that may imitate CAP symptoms. It is important to note that in early CAP cases, radiography results may appear negative, necessitating a repeat examination within 24 hours to enhance diagnostic accuracy. Additionally, in situations involving immunocompromised patients with suggestive symptoms and negative chest radiography, consideration of X-rays and CT scans may be prudent. Furthermore, serial chest radiography can be employed to monitor the progression of CAP; however, it is essential to be aware that radiographic improvement might not immediately align with clinical improvement.

Figure 1.6: Typical bacterial CAP [10].

1.7 The use of Radiology in the diagnosis

Radiology plays a crucial role in diagnosing lung diseases. Various imaging techniques are used to visualize the internal structures of the lungs and detect abnormalities. Some of the most common radiological techniques used for diagnosing lung diseases include:

- X-rays (Chest Radiography).
- Computed Tomography (CT) Scan.
- Magnetic Resonance Imaging (MRI).
- Positron Emission Tomography (PET) Scan.
- Ultrasound.
- Fluoroscopy.
- Ventilation-Perfusion (V/Q) Scan.

This study is fundamentally oriented towards an in-depth examination of X-rays and CT scans. Its principal aim is to enhance the comprehension of these methodologies. In pursuit of this goal, the subsequent comparative analysis is formulated:

1.7.1 CT scans & X-rays

As explained in the previous section, during the diagnostic stage, a variety of imaging examinations are conducted to assist the medical team in achieving a precise diagnosis. These examinations also play a crucial role in determining the most suitable course of treatment. Each imaging procedure employs distinct technologies to generate visuals that aid physicians in recognizing specific medical issues. This study primarily concentrates on CT scans, yet for the purpose of comparison, it is beneficial to elucidate both CT scans and X-rays and draw a comparison between them.

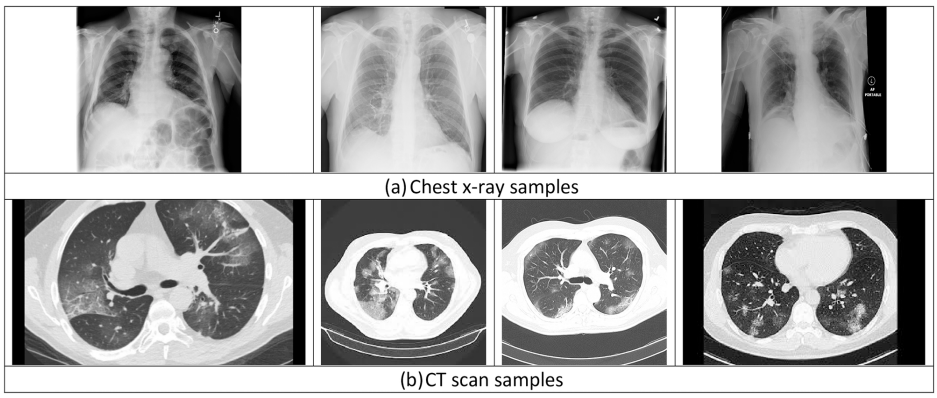


Figure 1.7: Sample of the two datasets: (a) chest X-ray images; (b) CT scans. [14]

Table 1.1 compares the main aspects between X-ray and CT-scan.

Table 1.1: Comparison: X-rays vs. CT Scans

Aspect	X-rays	CT Scans
Technology	Uses X-rays for imaging.	Utilizes rotating X-ray tube and detectors for cross-sectional imaging.
Image Detail	Provides 2D images of tissue density differences.	Produces cross-sectional images for better tissue differentiation and 3D reconstructions.
Applications	Diagnoses fractures, joint issues, and some lung conditions.	Detects tumors, soft tissue problems, internal injuries, and guides medical procedures.
Radiation Exposure	Lower radiation dose.	Higher radiation dose due to detailed imaging.
Exposure Time	Shorter exposure time.	Longer due to machine rotation.
Speed and Comfort	Quick and requires brief patient positioning.	Longer procedure with the patient lying still.
Image Formation	Denser tissues appear whiter, less dense appear darker.	Constructs detailed cross-sectional images.
Clinical Purpose	Good for bones, joints, and some lung issues.	Preferred for detailed investigations, tumors, blood vessels, and complex procedures.
Frequency of Use	Can be used more often.	Reserved for specific cases due to higher radiation exposure.
Safety Considerations	Generally safe with precautions.	Requires careful consideration due to higher radiation exposure.

1.8 The need for Accurate and Efficient Diagnosis Methods

The need for accurate and efficient diagnosis methods in the field of medicine is of paramount importance due to its direct impact on patient care, treatment outcomes, and overall healthcare systems. Accurate diagnosis refers to the precise identification of a medical condition or disease that a patient is experiencing. Efficient diagnosis, on the other hand, involves achieving this identification in a timely and resource-effective manner. Let us delve into the details of why both accuracy and efficiency are crucial in medical diagnosis:

- **Optimal Patient Care and Treatment:** Accurate diagnosis is the foundation of effective medical care. It provides essential information about the nature, severity, and specific characteristics of a patient's condition. With accurate diagnosis, healthcare professionals can tailor treatment plans to address the underlying causes and characteristics of the

disease, leading to better patient outcomes. Incorrect or delayed diagnosis can result in inappropriate treatments, worsening of the condition, and even patient harm.

- **Reduced Healthcare Costs:** Efficient diagnosis methods can significantly reduce healthcare costs. Timely and accurate diagnosis helps avoid unnecessary medical tests, treatments, and hospitalizations. Unneeded procedures can burden patients financially and strain healthcare systems. By streamlining the diagnostic process, healthcare resources can be allocated more effectively, reducing both financial and time-related burdens.
- **Prevention and Early Intervention:** Accurate diagnosis enables the early identification of diseases, which is critical for effective interventions and preventive measures. Detecting diseases at an early stage often provides more treatment options, higher success rates, and potentially lower healthcare costs. For conditions like cancer, early detection can mean the difference between successful treatment and advanced, less treatable stages.
- **Public Health and Contagious Diseases:** Rapid and accurate diagnosis is essential in identifying contagious diseases and preventing their spread. During outbreaks or pandemics, such as the case with COVID-19, swift and precise diagnosis helps isolate and treat affected individuals, limiting the disease's transmission and protecting public health.
- **Research and Innovation:** Accurate and efficient diagnostic methods contribute to medical research and innovation. Researchers can use reliable diagnostic tools to study disease patterns, develop new treatments, and identify emerging health threats. The ability to accurately identify diseases also aids in monitoring the effectiveness of new treatments and interventions.
- **Personalized Medicine:** Precision medicine relies on accurate diagnosis to tailor treatments to individual patients' genetic, molecular, and clinical characteristics. Accurate diagnostics enable the identification of specific biomarkers that can guide treatment decisions, maximizing therapeutic benefits while minimizing adverse effects.
- **Building Patient Trust:** Accurate diagnosis fosters trust between patients and healthcare providers. When patients receive timely and precise diagnoses, they have confidence in their healthcare professionals and the recommended treatment plans. Trust in the healthcare system encourages patient compliance with treatments and enhances overall patient satisfaction.
- **Global Health Challenges:** In regions with limited access to healthcare resources, accurate and efficient diagnosis becomes even more critical. Diagnostic tools that are portable, cost-effective, and easy to use can help address global health challenges, enabling the rapid identification and management of diseases in resource-constrained settings.

Incorporating accurate and efficient diagnostic methods is essential for improving patient outcomes, optimizing healthcare resources, and advancing medical knowledge. Ongoing advancements in technologies like medical imaging, molecular diagnostics, and artificial intelligence are contributing to the development of innovative diagnostic tools that hold the potential to revolutionize medical practice and patient care.

1.9 Conclusion

This chapter introduced the anatomy of human lungs, causes of lung diseases, and various types of lung diseases including airway, lung tissue, and lung circulation disorders. Notably, Lung Cancer, Covid-19, and Community-acquired pneumonia (CAP) were highlighted. The imperative for accurate and efficient diagnostic methods in lung disease detection was emphasized, as they are pivotal for tailored treatments and optimal patient outcomes. The chapter sets the stage for in-depth exploration of lung disease mechanisms, diagnostic advancements, and their transformative impact on healthcare.

Chapter 2

Deep learning background and Implementation

2.1 Introduction

Deep learning is a type of machine learning that is based on artificial neural networks, which are designed to learn from large amounts of data. Deep neural networks have become increasingly popular in recent years due to the advances in processing power and the availability of large datasets. In this chapter we will provide an overview of the theoretical foundations of deep neural networks, including the backpropagation algorithm, activation functions, and different types of neural networks.

2.2 Machine Learning & Deep learning

2.2.1 Machine Learning (ML)

Machine learning is an evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment. They are considered the working horse in the new era of the so-called big data. Techniques based on machine learning have been applied successfully in diverse fields ranging from pattern recognition, computer vision, spacecraft engineering, finance, entertainment, and computational biology to biomedical and medical applications [15].

2.2.1.1 Machine Learning process stages

As shown in the flow diagram below in Figure 2.1, the Machine Learning process stages are as follows:

Figure 2.1: Machine Learning process Steps [8].

- **Collection of Data:** The initial step in the machine learning process is to gather relevant data that will be used to train and test the model, This data may be obtained from a variety of sources, including databases, APIs, and sensors. this data might be in any format(CSV, XML, JSON, etc.)
- **Data Cleaning and Feature Engineering:** After collecting the date it is frequently necessary to clean and preprocess data. This include dealing with outliers, handling missing values, normalizing or scaling characteristics, and converting the data into a format that will be used for future research.
- **Model Training:** Once the data is collected, cleaned and processed, a machine learning model is chosen and trained using this data.
- **Evaluate Model:** After the model has been trained, it needs to be evaluated to determine how well it performs and how well it generalizes. Evaluation is typically done using a standard data set called the test set, which is unseen data and it was not used during the training. Depending on the particular learning task, common performance measures include accuracy, precision, recall, sensitivity, or mean squared error, . . . ect.
- **Model Deployment:** Based on the evaluation outcomes, the model can be fine-tuned or optimized to enhance its performance. This could entail modifying hyperparameters, experimenting with various algorithms or architectures, or gathering more data to address any discovered constraints or shortfalls [8].

2.2.1.2 Classifiers in Machine Learning

A classifier in machine learning is an algorithm that automatically organises or categorizes data into one or more of a set of "classes". One of the most common examples is an email classifier, which evaluates emails and filters them based on whether they are spam or not. Machine learning algorithms can be useful for automating processes that previously required

manual labor. They can significantly speed the process, cut costs and increase productivity for companies. The difference between a Classifier and a Model is that a classifier is an algorithm, or the set of rules that computers use to categorize data. However, a classification model is the outcome of the classifier's machine learning. The classifier is used to train the model, which eventually uses the classifier to classify the data. Classifiers can be supervised or unsupervised. Only unlabeled datasets are supplied to unsupervised machine learning classifiers, and they categorize the data based on patterns, structures, and anomalies found in the data. Training datasets are provided to supervised and semi-supervised classifiers so they can learn how to classify data into specific groups or classes [16].

2.2.1.3 Five Types of Classification Algorithms

These top 5 categorization algorithms must be able to deal with both needs and data.

- **Decision Tree:** A Decision Tree is a supervised machine learning algorithm used for classification and regression. It is a flexible and interpretable model that makes decisions based on a set of rules that can be represented as a tree-like structure. Each leaf node in the tree represents a class label (in classification) or a numerical value (in regression) [16].

- **Naive Bayes Classifier:** The Naive Bayes classifier is a probabilistic machine learning algorithm commonly used for classification tasks. The probability that a given data point will fall into one or more of a group of categories (or not) is calculated using the Naive Bayes. In text analysis, Naive Bayes is used to group texts into subjects, topics, or tags in order to arrange them in accordance with specified criteria. Naive Bayes algorithms determine the probability of each tag for a given text, then they produce the tag with the highest probability: the probability of A if B is true, is equal to the probability of B, if A is true, times the probability of A being true, divided by the probability of B being true [16].

$$P(A\backslash B) = \frac{P(B\backslash A) \times P(A)}{P(B)} \tag{2.1}$$

- **K-Nearest Neighbors:** K-Nearest Neighbors (K-NN) is a simple and intuitive machine learning algorithm used for both classification and regression tasks. It's a non-parametric and instance-based learning algorithm, meaning it doesn't make any assumptions about the underlying data distribution and instead relies on the data itself to make predictions. In text analysis, the k-NN algorithm would classify a given word or phrase by determining its closest neighbor: k is decided by a plurality vote of its neighbors. It would be placed in the class closest to 1 if $k = 1$ [16].

- **Support Vector Machines (SVM):** SVM is a flexible and powerful supervised machine learning algorithm that could be utilized for classification and regression tasks. SVM

is very useful when there is a clear margin of separation between classes or when the data cannot be separated linearly. SVM, which was developed in the 1990s by Vapnik and his colleagues, it has acquired significant acceptance and popularity in a variety of applications, including image classification, text classification, and bioinformatics [17].

- **Artificial Neural Networks:** The artificial neural network (ANN), often known as a neural network, is a machine learning technology that was developed to simulate the human brain. The data explosion in modern drug discovery research requires sophisticated analysis methods to uncover the hidden causal relationships between single or multiple responses and a large set of properties.The ANN is one of several flexible tools available to meet the growing need for drug discovery modeling. When compared to standard regression methods, the ANN can describe complicated nonlinear interactions. The ANN also has excellent fault tolerant, quick, and highly scalable with parallel processing [18].

2.2.2 Deep Learning (DL)

Deep learning is a subfield of machine learning that focuses on the development and training of artificial neural networks with many layers, also known as deep neural networks. Deep learning has gained significant attention and popularity due to its remarkable performance in a wide range of applications, including image and speech recognition, natural language processing, and especially healthcare like medical imaging interpretation, disease diagnosis, drug discovery, personalized treatment planning, and remote patient monitoring. It also powers healthcare chatbots, assists with genomics analysis, and automates tasks in radiology and pathology.However, it also poses challenges related to data requirements, model interpretability, and computational resources, which researchers and practitioners continue to address.

Deep learning algorithms can be broadly categorized into three types:

2.2.2.1 Supervised Learning

In supervised deep learning, the model is trained using labeled data, where each input is paired with a matching target or output. By reducing the difference between its predictions and the real labels, the model learns to map inputs to outputs. This method is often used for image classification and natural language processing tasks.

2.2.2.2 Semi- Supervised

Semi-supervised learning combines supervised and unsupervised learning elements. To train the model, it uses a combination of labeled and unlabeled data. Deep learning models in this category can benefit from the extra knowledge offered by labeled data while continuing to learn from unlabeled samples.

2.2.2.3 Unsupervised Learning

Unsupervised deep learning is used to process unlabeled data. The model aims to find hidden patterns or representations within the data without explicit supervision. Common techniques include clustering, dimensionality reduction, and generative modeling. Some examples of unsupervised deep learning are models autoencoders and deep belief networks.

2.2.3 Historical Development and Relevance of Deep Learning

In the late 20th century, deep hierarchical patterns of human voice perception and production systems served as inspiration for deep learning algorithms. Breakthroughs on deep learning have been achieved since Hinton presented a unique deep structured learning architecture dubbed the Deep Belief Network (DBN) in 2006 [19].

Artificial neural networks (ANNs) were the subject of research that gave rise to the idea of deep learning. Since 1980, backpropagation, an effective gradient descent technique, has been a key component of ANNs. Due to its local optima and overfitting issue, its performance may not be sufficient when applied to testing data. To achieve global optimums with reduced power consumption, several efficient machine learning methods (seen in classifiers title) such support vector machine (SVM), boosting, and K-nearest neighbor (K-NN) have been implemented.

Deep learning methods were initially introduced with Geoffrey Hinton's [19] proposal of layer-wise greedy learning in 2006. In this approach, unsupervised learning is utilized to pre-train the network before proceeding with layer-by-layer training. This technique significantly decreases the number of dimensions in data and produces a short representation. The growth in popularity of deep learning can be related to advancements in big-data analysis techniques, which, in part, address the issue of overfitting during training by providing the network with non-random initial values, this leads to faster convergence rates and better performance by locating lower values in the loss function.

Deep learning methods have piqued significant attention and achieved outstanding results in various fields. For example, in 2012, using deep learning techniques, a research team led by Hinton won the ImageNet Image Classification competition by surpassing the runner-up by 10% [20]. Google and Baidu have improved their image search engines by using Hinton's deep learning architecture. Furthermore, in 2016, Google's DeepMind project demonstrated outstanding performance in predicting the activity of pharmacological compounds and their effects on gene expression [21].

2.3 Artificial Neural Networks:

ANNs are computational models inspired by the form and function of biological neural networks observed in the human brain as shown in Figure 2.2 . ANNs are made up of linked artificial neurons or nodes that are grouped into layers. They are used in machine

learning and computational neuroscience to handle complicated problems including as pattern recognition, classification, regression, and optimization. ANNs are distinguished by their capacity to learn from data, adapt to new information, and generalize patterns, making them an important tool for supervised and unsupervised learning tasks. They've assisted in a variety of fields, including image and speech recognition, natural language processing, and autonomous systems.

In ANNs, two different kinds of neurons are employed: the biological neurons present in the human brain that serve as inspiration, and the artificial neurons, also known as perceptrons, that mimic their activity in the ANN design.

- **The biological neuron:** The living neuron is: The biological neuron, which has a cell body, dendrites, an axon, and synapses, is the basic building block of the neural network in the human brain. Axons transfer information while dendrites receive it as neurons join to transmit chemical and electrical impulses. Synapses are crucial in determining how strongly neurons are connected to one another [22].

  The Figure 2.2 represent the biological neuron.

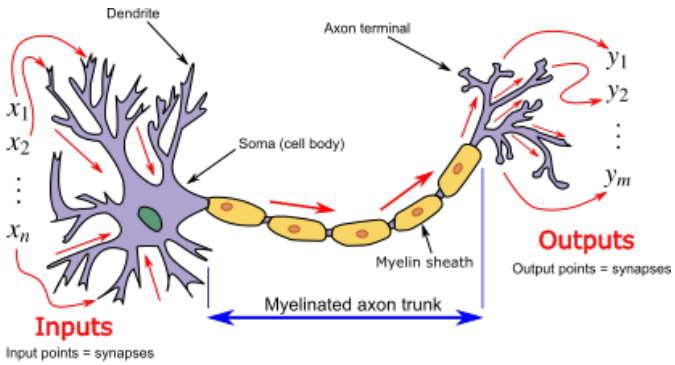


Figure 2.2: Neuron and myelinated axon, with signal flow from inputs at dendrites to outputs at axon terminals [23].

- **Artificial neuron:** Emulates the capabilities of a biological neuron in ANNs. It determines the weighted sum after receiving various inputs that have all been multiplied by a synaptic weight. Non-linearity is introduced by including a bias term and using an activation function, which enables ANN_s to recognize complicated relationships. The network's succeeding neurons receive an artificial neuron's output as input, promoting communication and learning [24] [25].

  The Figure 2.3 represent the artificial neuron.

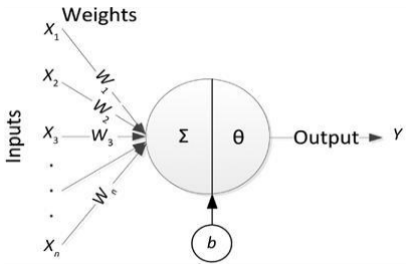


Figure 2.3: An artificial neuron [26].

– **Multilayer Merceptron (MLP):** The MLP architecture, which has several hidden layers between the input and output layers, is a more robust and adaptable ANN design as the diagram in Figure 2.4 shows. The weighted total of the inputs is applied by each neuron in the hidden layers as an activation function. MLPs are well suited for a variety of applications because they can learn intricate nonlinear correlations between inputs and outputs [25]

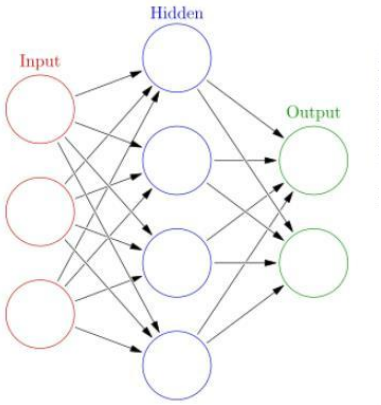


Figure 2.4: Connection between input, hidden and output layers [23].

2.4 Neural Network Training:

Neural networks training is a critical step in machine learning research. It requires incrementally improving network parameters using error minimization techniques such as gradient descent. Neural networks Training allows them to catch detailed patterns from large datasets, improving their capacity to generalize and generate accurate predictions across a wide range of applications such as computer vision, natural language processing, and reinforcement learning.

2.4.1 Loss function

A loss function, which is also known as a cost function, considers the probability or uncertainty of a forecast depending on how much it varies from the real value. This provides us with a deeper understanding of how the model is doing.

2.4.2 Gradient Descent

Gradient descent is an optimization approach that iteratively modifies the weights and biases of the network. The amplitude and direction of updates are chosen using the gradient of the loss function (L) with respect to the parameters. The parameters are changed as follows:

$$\theta = \theta - learningrate \times \Delta L \tag{2.2}$$

where θ represents the network's parameters (weight and bias), and the learningrate controls the step size in each iteration.

2.4.3 Backward Propagation

Backward propagation is the tool that gradient descent uses to calculate the gradient of the loss function (cost function). It is an algorithm used to determine the gradients of a neural network by propagating errors backward through the network, it requires applying the chain rule of calculus to compute the partial derivatives of the loss function with respect to the network's parameters. The gradients are then used to update the parameters during training via gradient descent. The gradients (Δ) of the loss function (L) with respect to the parameters (θ) are computed mathematically during back propagation as follows:

$$\Delta = \frac{dL}{d\theta} \tag{2.3}$$

These gradients are then used to update the parameters in the opposite direction of the gradient as shown in Figure 2.5:

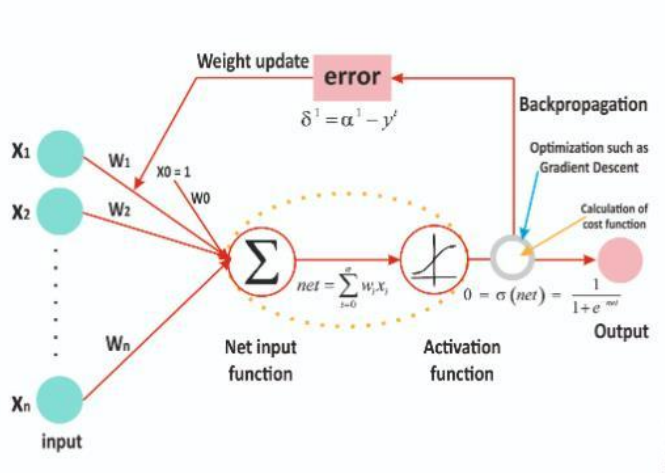


Figure 2.5: Gradient descent and weights update in Backpropagation algorithm [27].

2.4.4 Dropout

Dropout is a regularization method commonly used in deep learn used to stop overfitting in deep learning models, during training, it randomly cuts out (sets to zero) certain number of neurons from the neural network. This prevents the network from being overly dependent on any specific features or neurons, and forces it to learn more reliable ones by promoting better generalization to unseen data. Dropout is a widely adopted technique that enhances the robustness and performance of deep neural networks in various applications.

As shown in Figure 2.6

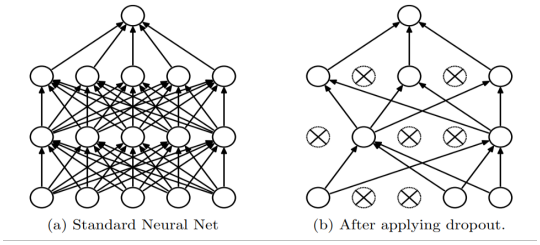


Figure 2.6: Dropout Regularization [28]

2.4.5 Batch Normalization

Batch normalization is a technique for normalizing the inputs to each layer in a neural network. It helps improve training stability and speed, by minimizing the internal covariate shift (changes in the distribution of a network's internal activations through training). To understand more the Figure 2.7 shows the neuron before and after applying batch normalization.

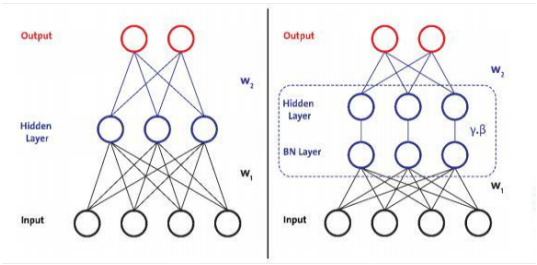


Figure 2.7: A Neural Network before and after Batch Normalization [28].

2.5 Hyperparameters

Hyperparameters are parameters whose values control the learning process and determine the values of model parameters that a learning algorithm ends up learning. The prefix 'hyper-' indicates that they are 'top-level' parameters that control the learning process and the model parameters that come from it [29]. These parameters can have a direct impact on the behavior of the training algorithm and have a significant impact on the

performance of the model being trained and they are set by the data scientist or machine learning practitioner before the training process starts. The learning rate, the number of epochs, the batch size, the dropout rate, and other factors that control how the model learns and generalizes from data are examples of hyperparameters.

2.5.1 Learning rate

The learning rate is a hyperparameter that controls the magnitude of parameter changes during training. It controls how fast or slowly a machine learning model converges on the ideal set of parameters for minimizing the loss function. To calculate the amount of each parameter update in gradient-based optimization, the learning rate is multiplied by the gradient of the loss function with respect to the model's parameters.

2.5.2 Number of epochs

The number of epochs is another critical hyperparameter that determines how many times the entire training dataset is processed by the model during training. each epoch is made up of numerous iterations (mini-batches). The number of epochs determines how long the training process will last.

2.5.3 Batch size

The batch size is a hyperparameter that defines how many data samples are utilized to compute the gradient and update the model's parameters during one training iteration. A larger batch size may lead to faster training but can require more memory, while a smaller batch size may result in slower training but potentially better convergence.

2.6 Activation Functions

An important part of artificial neural networks is the activation function, which decides whether or not to activate a neuron based on input. By controlling which neurons are engaged and have an impact on the subsequent layer, this function plays a critical role in aiding neural networks in learning complicated patterns in data. The activation function transforms the output from one neuron into an input for the following neuron, much as how neurons in the human brain do so. Since a neural network would only be able to perform as a linear regression model in the absence of an activation function, its main function is to add non-linearity into the network. Applying an activation function to the weighted sum of a neuron's inputs, which includes the bias term, yields the neuron's activity as shown below:

$$A = G(W_1X_1 + W_2X_2 + \cdots + W_nX_n + B) \tag{2.4}$$

Such that:
- A: output of the neuron.
- G: activation function.
- W: weight.
- X_i: features.
- B: bias.

2.6.1 Sigmoid

The sigmoid activation function, also known as the logistic function, generates outputs between 0 and 1 by compressing the input into an S-shaped curve. As a result, it is applied to models that need to do binary classification. The mathematical formulation of sigmoid function is:

$$sigmoid = \frac{1}{1+e^{-x}} \tag{2.5}$$

While sigmoid activation functions have their applications, particularly in specific output layers and historical contexts, other activation functions, such as ReLU and its variants, have gained popularity in modern deep learning architectures due to their ability to address some of the limitations associated with sigmoid functions.

2.6.2 Softmax

Softmax is a powerful generalization of sigmoid , it is an essential component in neural network architectures for multi-class classification tasks, such as image classification, natural language processing, and many other applications where the model needs to assign inputs to one of several possible categories. It uses a softmax operation to normalize the input vector into a probability vector, where each value denotes the probability or confidence score for the relevant class. It is frequently used at the topmost layer of the neural network as the final activation function. Softmax is mathematically represented by:

$$softmax(x) = \frac{e^{xi}}{\sum_{j=1}^{k} e^{xj}} \tag{2.6}$$

Such that:
- x: input vector.
- K: number of classes in the multi-class classifier.
- e^{xi} : $exp(xi)$.
- e^{xj} : $exp(xj)$.

2.6.3 ReLU

Rectified Linear Unit ReLU is a very common activation function in neural networks. it generates non-linearity by outputting the input directly if it's positive, and zero otherwise. Its formula is

$$f(x) = max(0,x) \tag{2.7}$$

ReLU has a number of benefits, including the ability to learn more quickly than sigmoid activations and avoid saturation because its gradient is either 0 or 1. Additionally, it reduces the likelihood of gradient-related issues such as exploding gradients and vanishing gradients. Furthermor, it surpass sigmoid in striking a compromise between computational effectiveness and biological realism. However, it can suffer from the "dying ReLU" problem, where some neurons can become inactivee during training.

As a result, ReLU has become a preferred choice for many deep learning architectures, contributing to their stability and faster training convergence.

2.7 Deep Neural Network:

DNNs have been suggested as an extension of ANN shallow architectures due to improvements in hardware and processing capacity. DNNs do not seek to represent the biological brain in the same way that cortical algorithms (CAs) or other machine learning methods with biological inspiration do, even if they use the idea of neurons from the biological brain. The Fukushima (1980) neocognitron model is where DNN ideas come from. DNN architectures are designed to build powerful AI models and are broadly characterized as a collection of machine learning algorithms that tries to learn in a hierarchical fashion and that involves various levels of abstraction for knowledge representation. The learning at the higher level is determined by and builds on the statistical learning that occurs at the lower-level layers of these systems as information spreads through higher levels. With such a broad definition of deep learning, we might consider backpropagation algorithms, which have been there since 1974, to be the forerunners of deep architectures, along with recurrent neural networks and convolution neural networks, which were developed in the 1980s. Research on deep architectures has been going, though, only after Hinton, Osindero, and Teh's (2006) contribution to deep learning training [26]. In actuality, linear models or shallow neural networks might not have enough expressiveness to accurately predict the task. It has been suggested that deep neural networks (DNNs) could be used to create more accurate predictive models. They can be represented abstractly as a series of layers,

$$f(x) = f_L \circ \cdots \circ f_1(x) \tag{2.8}$$

Each of these applies a linear transformation and then introduces an elementwise nonlinearity. By combining many of these layers, the model has a tremendous capacity for prediction. Computer vision tasks have shown particular success for DNNs. However, DNN models are also much more complex and nonlinear, and quantities entering into the simple explanation model become considerably harder to compute and estimate reliably [30]. Four main deep learning architectures are restricted: Boltzmann machines (RBMs), deep belief networks (DBNs), autoencoder (AE), and convolutional neural networks (CNNs). Convolutional neural networks will be the main focus of this study.

2.8 Convolutional Neural Networks CNNs:

CNNs, a subclass of the discriminative deep architecture, have demonstrated acceptable processing performance for two-dimensional data with grid-like topology, such as images and videos. CNNs use convolution processes to identify visual patterns by imitating the visual cortex [31]. Convolution has taken the place of the universal matrix multiplication in standard NNs in CNNs. By doing this, the number of weights is reduced and the network's complexity is lowered. Additionally, the photos can be instantly integrated into the network as raw inputs without having to go through the feature extraction process that is required by conventional learning techniques. Due to the effective training of the hierarchical layers, CNNs are the first fully successful deep learning architecture. By utilizing spatial relationships, the CNN architecture minimizes the amount of parameters in the network, which improves performance when using conventional backpropagation techniques. The GPU-accelerated computing techniques have been used to train CNNs more effectively due to the quick development of calculation techniques. CNNs are now successfully used in recommender systems, handwriting recognition, face detection, behavior recognition, speech recognition, and particularly image classification [32]. CNNs are made up of a stack of convolution and pooling layers that are alternated, along with several fully connected layers as shown in Figure 4.14 and Figure 4.14.

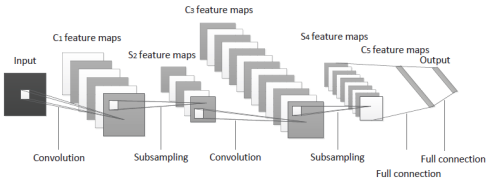


Figure 2.8: Schematic of CNNs [32].

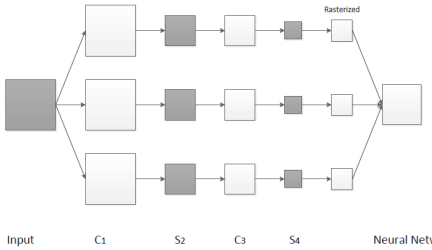


Figure 2.9: Conceptual structure of CNNs [32].

2.8.1 Convolutional Layer

When extracting features from an input image, a Convolutional Layer is essential. The image is convolved using a smaller matrix known as a kernel or filter. The dot product between the kernel and the image is calculated as the kernel slides across the image. The sliding distance is determined by stride length. When working with colored images, the kernel's channel count must coincide with the image's channel count. With their sizes maintained constant, many kernels can be used to extract numerous features. Convoluted features are layered together to create an output with the number of channels equal to the number of filters applied. An activation function, such as ReLu or Tanh, is applied to the convolved features to induce non-linearity. The output, sometimes referred to as a feature map, is what this method produces.

2.8.2 Pooling Layer

The pooling layer is an essential component of convolutional neural networks (CNNs) that plays a number of important roles in the processing of visual and spatial input. It usually comes after convolutional layers to reduce dimensionality, provide translation invariance, and select features. By reducing the size of the input image and increasing the reliability of recognized features, it enhances computational speed and memory management. Pooling increases the network's ability to recognize features independent of their actual geographical locations, promoting strength. It systematically collects essential information from tiny locations, either by max-pooling (keeping the maximum value) or by average-pooling (calculating the mean). This reduces the network's parameter count dramatically, decreasing overfitting while increasing generalization.

2.8.3 Fully Connected Layer

The Fully Connected Layer (FCL), also known as the Dense Layer, is an important component in artificial neural networks, especially in feedforward neural networks and deep learning designs such as multi-layer perceptrons (MLPs). It connects the feature extraction layers (such as the convolutional and pooling layers in CNNs) to the output layer, which is responsible for generating final predictions or classifications [33]. It is

also useful for tasks like image classification and natural language processing, where complex relationships between features need to be captured to make high-level predictions.

Chapter 3

Approach and Implementation

3.1 Introduction

This chapter presents the approach and implementation details for our project. We discuss the tools and technologies used, we delve into the concept of transfer learning, specifically focusing on the ResNet, Inception-V3, and DenseNet architectures, and introduce performance metrics.

3.2 Tools

In our way to develop and evaluate deep learning models for chest image classification, powerful tools and technologies were relied . These tools played a critical role in the successful execution of our study, enabling us to manipulate, process, and analyze our dataset effectively. Our toolkit included: Kaggle, Python, TensorFlow, Keras,and OpenCV.

3.2.1 Kaggle

Kaggle [34] is an online platform and community dedicated to data science, machine learning, and artificial intelligence. It provides access to a diverse array of datasets and hosts competitions that challenge data scientists and machine learning enthusiasts to solve real-world problems. It serves as a hub for data-driven collaboration, enabling knowledge sharing, model development, and insights discovery within the global data science community.

3.2.2 Python

Python [35] is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together.

Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance.

3.2.3 TensorFlow

TensorFlow [36] is an open-source numerical computing library developed by Google that focuses on machine learning and deep learning applications. It provides an adaptive and fast framework for executing multi-dimensional array-based mathematical computations known as tensors. TensorFlow is intended to make it easier to construct, train, and deploy machine learning models, particularly neural networks, across a variety of platforms and computing environments.

3.2.4 Keras

Keras [37] is a high-level, deep learning API developed by Google for implementing neural networks. It is written in Python and is used to make the implementation of neural networks easy. It also supports multiple back-end neural network computation.

3.2.5 OpenCV

OpenCV [38], or Open Source Computer Vision Library, is a free and open-source computer vision and machine learning software library. It provides a wide number of tools and functionalities for tasks relating to computer vision, image processing, and machine learning, making it a significant resource for developers and academics working in disciplines such as robotics, augmented reality, facial recognition, image and video analysis, and more.

OpenCV is built in C++ and provides bindings for other computer languages, including Python, Java, and MATLAB, making it accessible to a large community of developers. It supports a variety of operating systems, including Windows, Linux, macOS, Android, and iOS.

3.3 Transfer Learning

Transfer learning is a powerful technique in which knowledge gained from one model is applied to another related task. In our project, we have employed transfer learning using the ResNet, Inception-V3, and DenseNet architectures, utilizing pretrained models trained on large datasets such as ImageNet. The architectures are renowned for their effectiveness in medical imaging classification. We have specifically chosen these models due to their capability to achieve remarkable performance while having a relatively smaller number of parameters. This choice is well-suited for our dataset, which comprises a large number of high-resolution images.

Figure 3.1: This is an example image.

3.3.1 ResNet

ResNet stands for Residual Neural Network is a Convolution Neural Network with 50 Deep Layers developed in 2015 by Kaimimg He. There are two blocks in this Model: Convolution Block and Identity Block. ResNet50 uses the principle of Skip Connection. There are 48 Convolution Layers, 1 Max Pool Layer, and 1 Average Pool Layer among the 50 Layers, it is used for computer vision applications designed to support hundreds or thousands of convolutional layers. Residual neural networks are a type of artificial neural network (ANN) that forms networks by stacking residual blocks [39].

3.3.2 Inception-V3

Inception-V3 is a convolutional neural network for assisting in image analysis and object detection in a total of 48 layers, including convolutional layers, pooling layers, fully connected layers, and auxiliary classifiers. It utilizes factorized convolutions and aggressive regularization to reduce the number of parameters and prevent overfitting.

3.3.3 DenseNet

DenseNet is a type of convolutional neural network that utilizes dense connections between layers, through Dense Blocks, where we connect all layers (with matching feature-map sizes) directly with each other [40]. It consists of 121 layers, including densely connected convolutional layers. The key innovation of DenseNet is its dense connectivity pattern, where each layer receives direct input from all preceding layers. This design promotes feature reuse, enhances information flow, and reduces the risk of vanishing gradients.

3.4 Performance Metrics

3.4.1 Confusion Matrix

It is N by N matrix, where N is the number of classes or outputs, and it is one of the most intuitive performance measurements for machine learning classification problems where output can be two or more classes. It displays a tabular representation of the model's correct and wrong predictions, classified according to the actual and projected classifications. This matrix is very useful when dealing with unbalanced datasets with varying class distributions. By studying error patterns, it is possible to discover areas where the model may be failing, allowing for targeted modifications to improve overall performance [41]. It gives a comparison between actual and predicted values and summarizes the count of the parameters as follows in Figure 3.2:

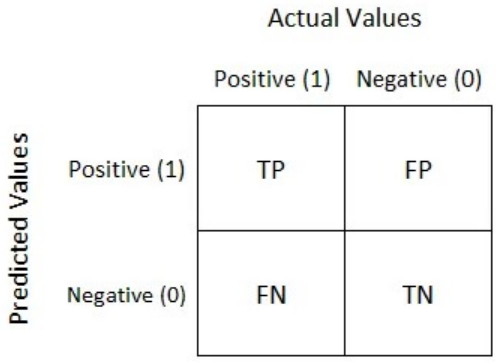


Figure 3.2: Confusion Matrix [42].

- **True Positive(TP):** The actual value and the predicted values are the same.
- **True Negative(TN):** The actual value and the predicted values are the same.
- **False Positive(FP):** The actual value is negative, but the model has predicted it as positive,
- **False Negative(FN):** The actual value is positive, but the model has predicted it as negative.

3.4.2 Accuracy

Accuracy is a metric for classification models that measures the number of predictions that are correct as a percentage of the total number of predictions that are made [43]. Formally, accuracy has the following definition:

$$\text{Accuracy} = \frac{\text{Sum of Correct Predictions for All Classes}}{\text{Total Number of Predictions for All Classes}} \tag{3.1}$$

For binary classification, accuracy can also be calculated in terms of positives and negatives as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3.2}$$

3.4.3 Recall

Recall (or sensitivity) measures the proportion of actual positive instances that were correctly identified by the model. It provides an indication of the model's ability to avoid missing positive cases, making it particularly relevant when the cost of false negatives is high. Recall is commonly used in medical diagnostics, information retrieval, and other applications where the focus is on finding relevant items [44].

The equation for recall is as follows:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.3}$$

3.4.4 Specificity

Specificity measures the proportion of actual negative instances that were correctly identified as negative by the model. It provides insight into the model's ability to avoid classifying negative cases as positive, which is essential when the cost of false positives is high.

The equation for specificity is as follows:

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{3.4}$$

3.4.5 Precision

Precision in classification is the measurement of the accuracy of positive predictions made by a model. It quantifies the proportion of instances that the model correctly classified as positive out of all instances it predicted as positive [44].

The equation for precision is as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.5}$$

3.4.6 F1-Score

The F1-Score is a machine learning evaluating metric in classification that combines precision and recall (sensitivity) scores of a model into a single value. The F1-Score ranges between 0 and 1, with higher values indicating better model performance [45]. It is especially useful when there is an imbalance between the classes.

The equation for F1-Score is as follows:

$$\text{F1-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3.6}$$

3.5 Conclusion

Chapter 3 provides an overview of the methodology and tools used in the project. It highlights the evaluation metrics, and the used transfer learning models. Overall the chapter establishes the foundation for the next chapter's results and discussion laying the groundwork for the development and evaluation of a deep learning model for chest image classification.

Chapter 4

Results and Discussion

4.1 Introduction

In this chapter, we unveil the outcomes of our comprehensive analysis in the field of chest image classification. Building upon the foundation laid in the previous chapters where we explored the background of chest diseases and deep learning methodologies, and the approach used, this chapter serves as the culmination of our research efforts.

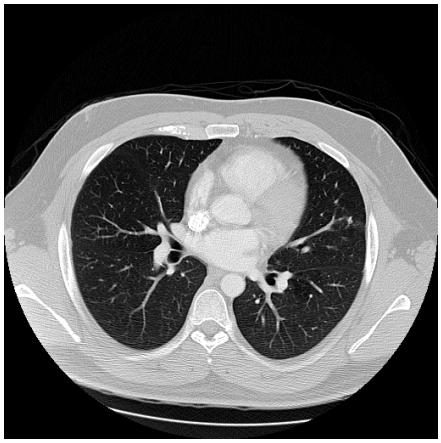
Our investigation centers on the utilization of deep neural networks, including ResNet, Inception-V3, and DenseNet, to classify chest images, utilizing datasets sourced from Kaggle. This exploration extends to both binary classification with two different types of datasets: CT-scan and X-ray, where we discern specific conditions, and multi-class classification, which broadens our scope to encompass a range of diagnostic categories.

Our primary objective is to conduct a meticulous comparison between these deep learning models and classification approaches, aiming to identify the most accurate and efficient strategies. These findings carry profound implications, particularly in the domain of healthcare, where accurate and efficient diagnoses play a critical role in saving lives and improving patient outcomes.

4.2 Experiments and Results

4.2.1 Binary classification

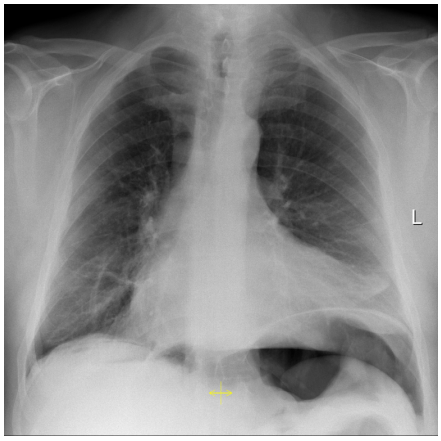
In our binary classification experiment,we used two datasets for distinguishing COVID-19 from non-COVID-19 cases, both datasets downloaded from kaggle platform under the name of "COVID 19 X-Ray and CT-Scan Image" [10] in total of 17099 X-ray and CT-scan images. The dataset contains two main folders, one for the X-ray images, which includes two separate sub-folders of 5500 Non-COVID images and 4044 COVID images. The other folder contains the CT images. It includes two separate sub-folders of 2628 Non-COVID images and 5427 COVID images.



a Example from COVID CT-scan dataset

b Example from Non COVID CT-scan dataset

Figure 4.1: Examples from CT-scan dataset [10].



a Example from COVID X-ray dataset

b Example from Non COVID X-ray dataset

Figure 4.2: Examples from X-ray dataset [10]

Now we will present the results of our classification experiment. We used three distinct models with a fixed splitting ratio at 60:20:20.

In addition, we carefully evaluated several combinations of hyperparameter values to determine the most successful ones. After extensive testing.

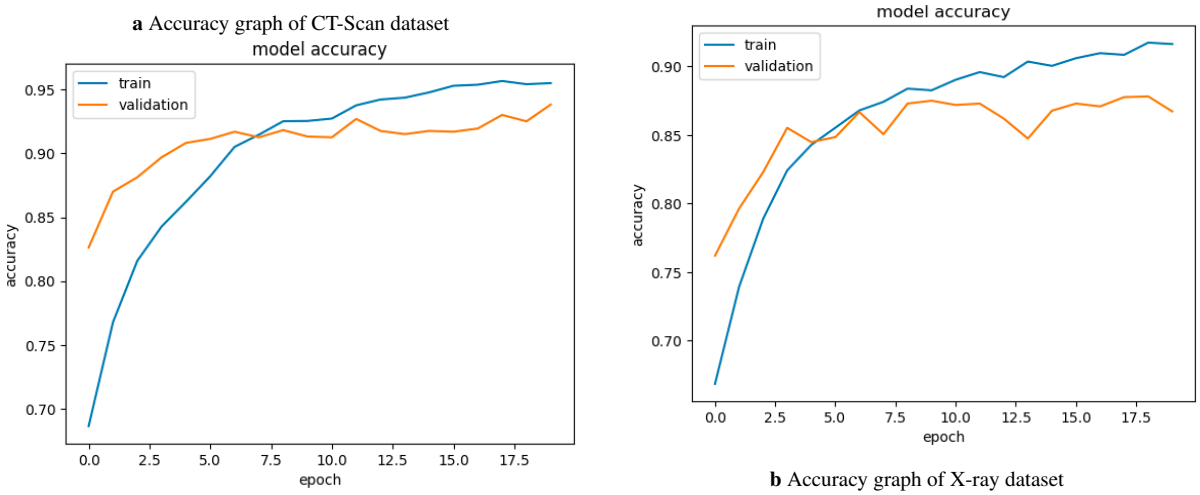
We chose the hyperparameters presented in Table 4.1. The subsections that follow give thorough results, analyses, and discussions for each model and imaging technique, giving light on their adaptability to different data situations and their implications for practical medical picture classification.

Table 4.1: Selected hyperparameters values for Binary classification

Hyperparameter	Fixed value
Number of Epochs	20
Learning rate	0.001
Batch size	32
Optimizer	Adam
Loss function	categorical crossentropy

4.2.1.1 Inception V3

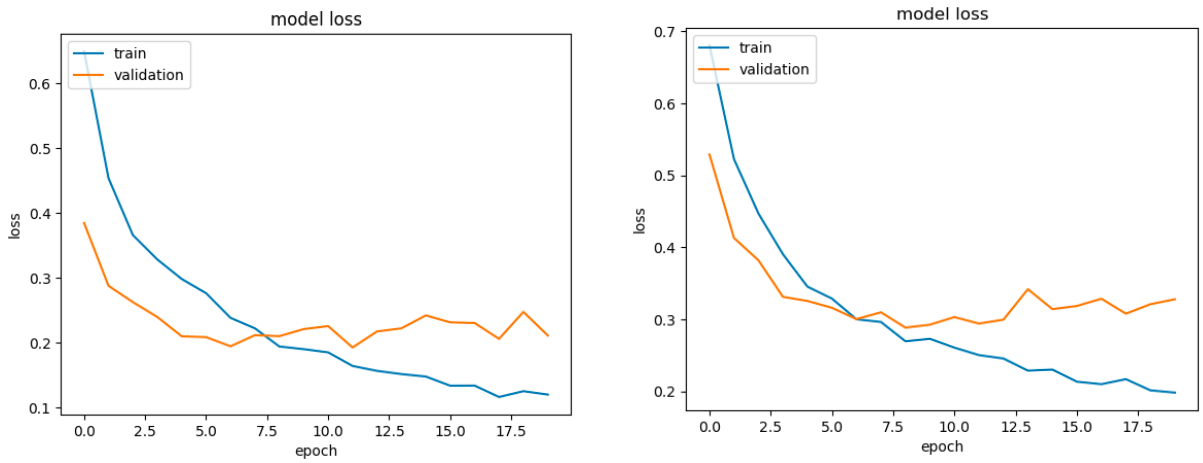
In the exploration of chest image classification, the first step was deploying the Inception-V3 model on both CT scan and X-ray dataset. With a splitting ratio of 60:20:20 for train, validation and test respectively, and a fixed number of epochs at 20 epochs. As we present the curves of accuracy and loss of this experiment in Figure 4.3 and Figure 4.4:



a Accuracy graph of CT-Scan dataset

b Accuracy graph of X-ray dataset

Figure 4.3: Accuracy graphs of Inception-V3 model

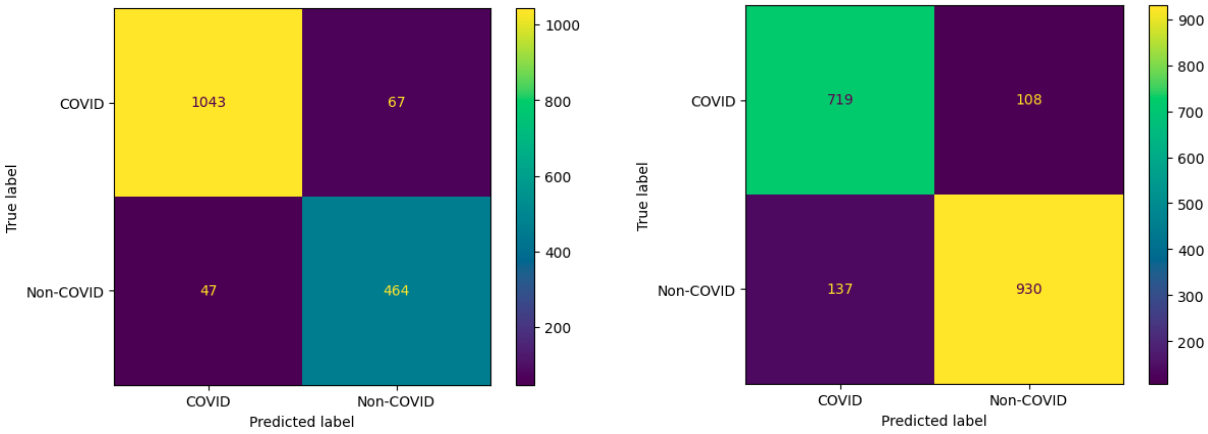


a Loss graph of CT-scan dataset

b Loss graph of X-ray dataset

Figure 4.4: Loss graphs of Inception-V3 model

As we can see from Figure 4.3, both graphs exhibit a noticeable upward trend, indicating positive learning progress, while The loss graphs in Figure 4.4 indicate a consistent downward trend, indicating effective convergence during training, with a remarkable advantage in CT-scan curves. Further information is needed from the confusion matrix in Figure 4.5 :



a Confusion matrix of CT-scan dataset

b Confusion matrix of X-ray dataset

Figure 4.5: Confusoin matrix of Inception-V3

As we can see from the confusion matrix CT-scan is giving a slightly better results. in addition, we have other metrics to cover in Table 4.2 :

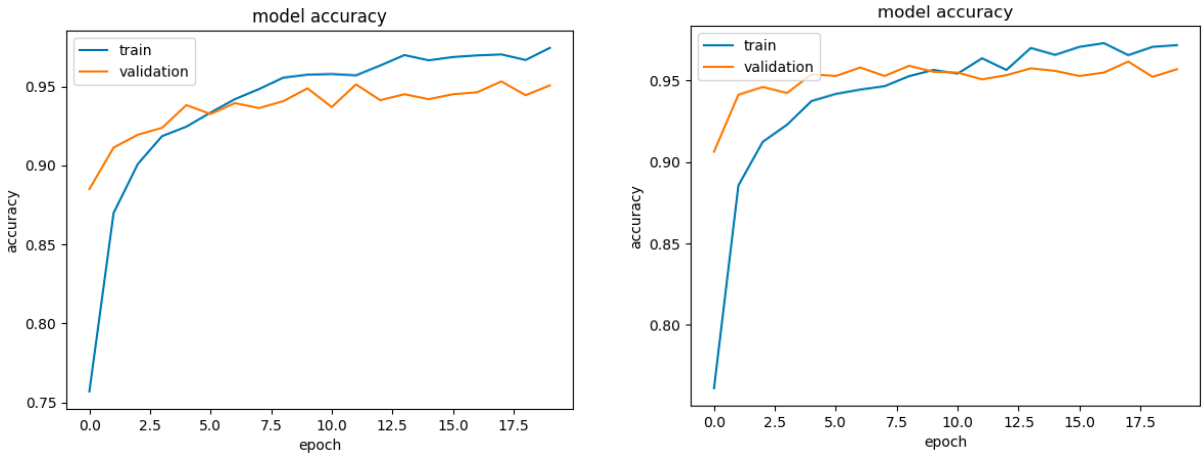
Table 4.2: Performance Metrics for Inception-V3 Model on COVID-19 Binary Classification

Dataset	Class	Accuracy	Precision	Recall	F1-Score
CT Scan	COVID	0.92	0.95	0.95	0.95
CT Scan	Non-COVID	0.92	0.89	0.88	0.89
X-ray	COVID	0.88	0.83	0.90	0.86
X-ray	Non-COVID	0.88	0.92	0.85	0.88

Table 4.2 compares the performance between CT-scan and X-ray datasets for COVID-19 binary classification using the Inception-V3 model. Notably, for the CT Scan dataset, both COVID and Non-COVID classes demonstrate high accuracy at 0.92, indicating strong classification. In particular, the COVID class achieves a precision, recall, and F1-Score of 0.95, showing outstanding performance in detecting COVID cases. On the other hand, the X-ray dataset also performs well with an accuracy of 0.88 for both classes. However, the Non-COVID class has a greater precision of 0.92, demonstrating its ability to properly identify Non-COVID instances. The COVID class has a high recall of 0.90, highlighting its efficacy in collecting COVID instances.

4.2.1.2 ResNet

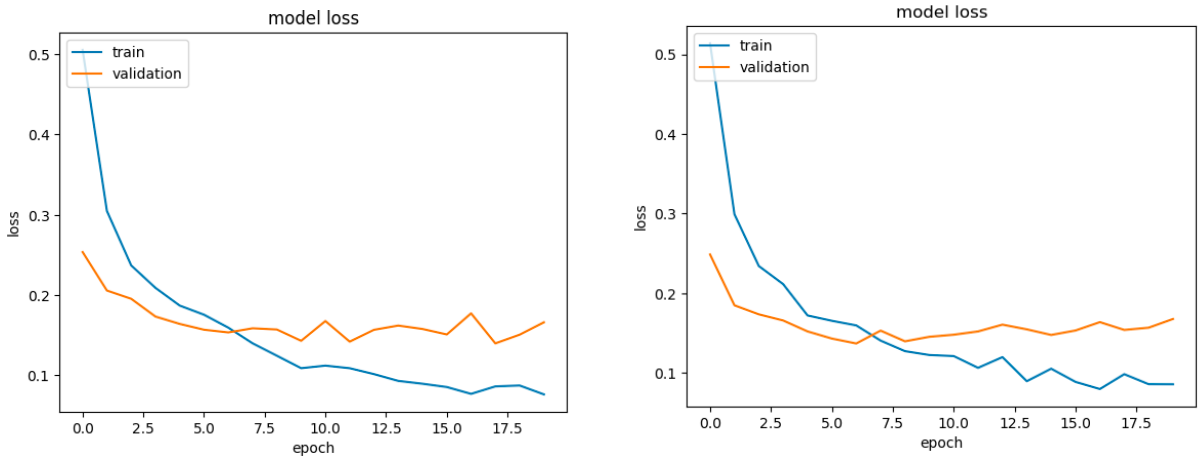
With the same hyperparameters and steps of inception-V3, we deployed the ResNet model on both CT scan and X-ray datasets and the curves shown in Figure 4.6 and Figure 4.7 as follows :



a Accuracy graph of CT-scan dataset.

b Accuracy graph of X-ray dataset.

Figure 4.6: Accuracy graphs of Resnet models

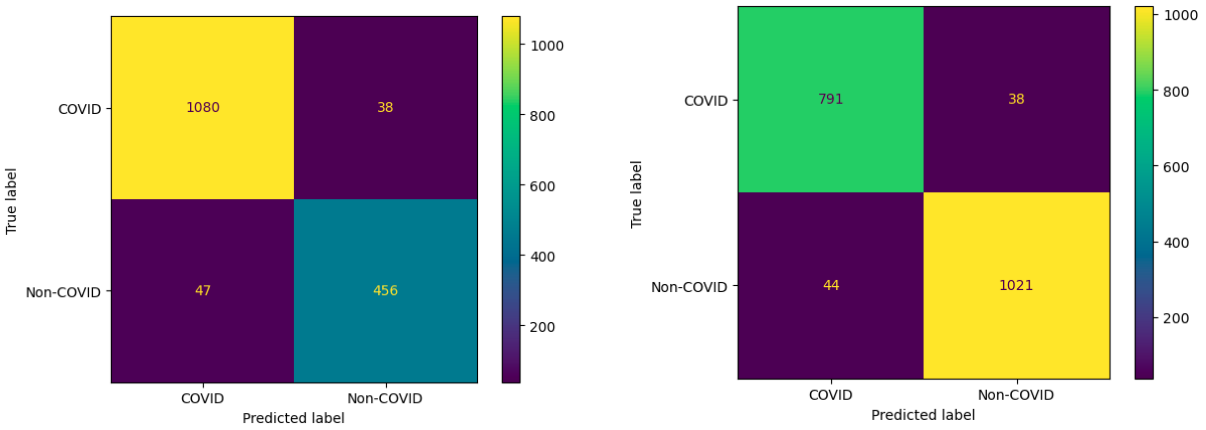


a Loss graph of CT-scan dataset.

b Loss graph of X-ray dataset.

Figure 4.7: Loss graphs of ResNet models

In Figure 4.6, we observe consistent upward trends in accuracy for both CT Scan and X-ray datasets, indicating that the model steadily improves its ability to classify COVID-19 cases over training epochs. This suggests that the model is effectively learning from the data. Moreover, the loss curves in Figure 4.7 show a continuous downward trajectory, affirming that the model is converging during training. The decreasing loss values reflect the diminishing error in the model's predictions. More details from the confusion matrix in Figure 4.8.



a Confusion matrix of CT-scan dataset

b Confusion matrix of X-ray dataset

Figure 4.8: Confusion matrix of ResNet

The confusion matrix results in Figure 4.8 are indicative of the model's performance in both the CT Scan and X-ray datasets. Notably, in the CT Scan dataset, the model misclassified 38 cases as non-COVID when they were not, and 47 cases as COVID when they were non-COVID. Similarly, in the X-ray dataset, there were 38 false positives (non-COVID misclassified as COVID) and 44 false negatives (COVID misclassified as non-COVID).

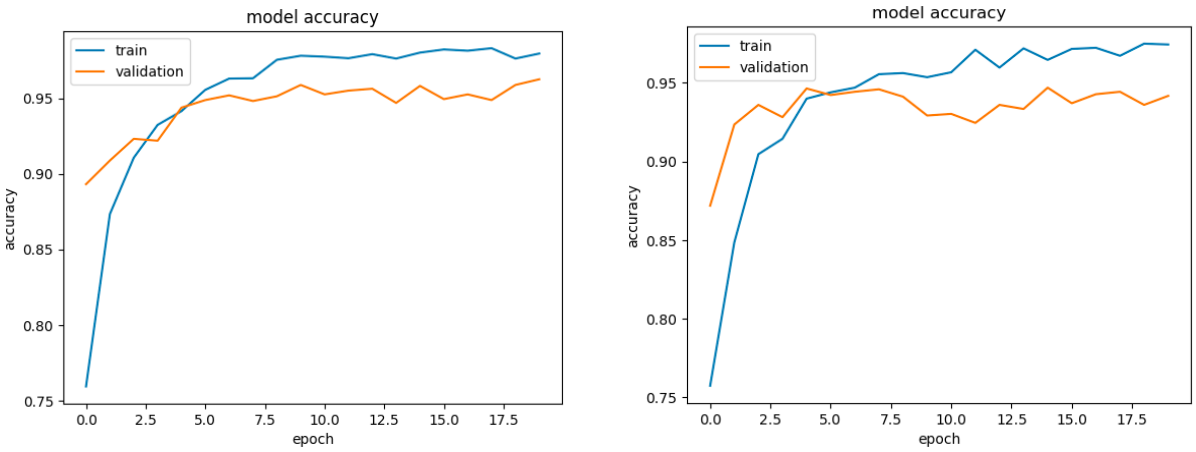
Table 4.3: Performance Metrics for ResNet Model on COVID-19 Binary Classification

Dataset	Class	Accuracy	Precision	Recall	F1-Score
CT Scan	COVID	0.95	0.96	0.97	0.96
CT Scan	Non-COVID	0.95	0.93	0.93	0.93
X-ray	COVID	0.92	0.93	0.94	0.93
X-ray	Non-COVID	0.92	0.92	0.93	0.92

In Table 4.1 the model demonstrates strong accuracy, with CT Scan achieving 0.95 for both COVID and Non-COVID classes and X-ray achieving 0.92 for both classes. Notably, CT Scan exhibits exceptional precision, recall, and F1-Score values of 0.96, 0.97, and 0.96 for COVID, demonstrating its proficiency in identifying COVID cases. Similarly, X-ray maintains a solid precision, recall, and F1-Score of 0.93 for COVID.

4.2.1.3 DenseNet

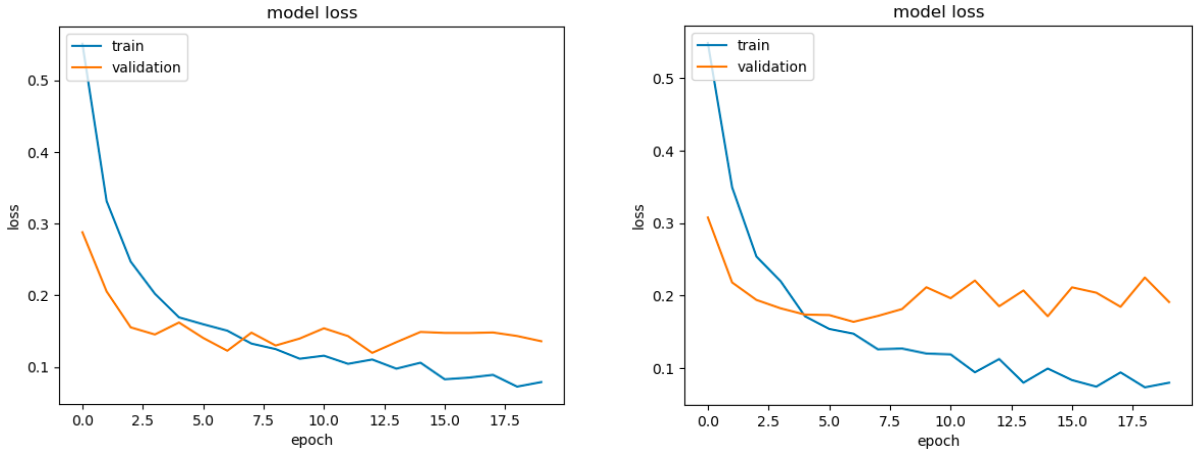
With the same parameters and steps of inception-V3 and ResNet, we deployed the DenseNet model on both CT scan and X-ray datasets and the graphs of accuracy and loss are shown in Figure 4.9 and Figure 4.10, respectively:



a Accuracy graph of CT-scan dataset

b Accuracay graph of X-ray dataset

Figure 4.9: Accuracy graphs of DenseNet models

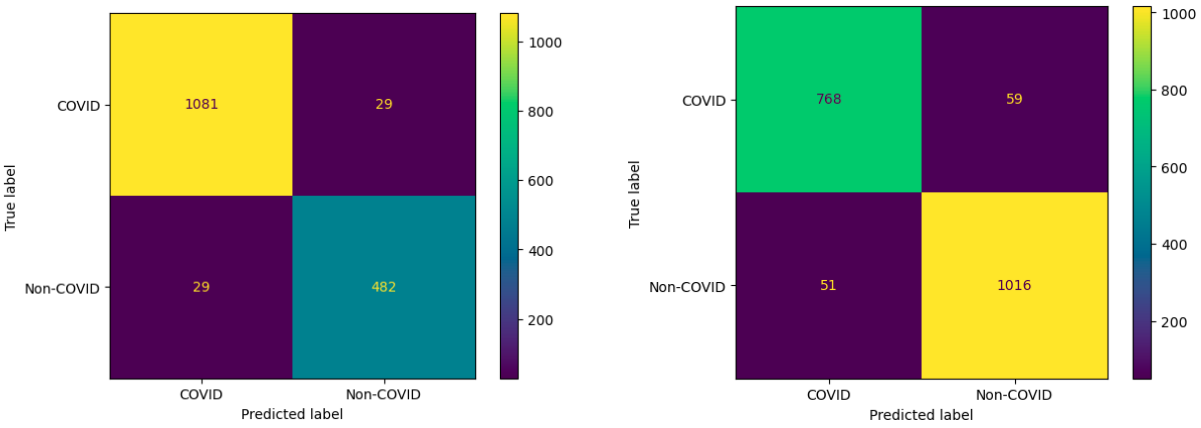


a Loss graph of CT-scan dataset

b Loss graph of X-ray dataset

Figure 4.10: Loss graphs of DenseNet models

in Figure 4.9 and Figure 4.10 the accuracy and loss curves for the DenseNet model for both datasets reveal promising trends. The accuracy consistently improves, demonstrating effective learning, while the loss steadily decreases, indicating convergence. These patterns suggest the model's potential for accurate COVID-19 classification, although further evaluation is necessary from the confusion matrices in Figure 4.11.



a Confusion matrix of CT-scan dataset

b Confusion matrix of X-ray dataset

Figure 4.11: Confusion matrix of DenseNet model

The confusion matrix results that the model's performance in both the CT Scan and X-ray datasets looks better than the matrices of ResNet and Inception-V3 with less misclassified samples. Notably, in the CT Scan dataset, the model misclassified 29 cases as COVID when they were not and 29 cases as non-COVID when they were COVID. Similarly, in the X-ray dataset, there were 29 false positives (COVID misclassified as non-COVID) and 51 false negatives (non-COVID misclassified as COVID).

Table 4.4: Performance Metrics for DenseNet Model on COVID-19 Binary Classification

Dataset	Class	Accuracy	Precision	Recall	F1-Score
CT Scan	COVID	0.96	0.97	0.97	0.97
CT Scan	Non-COVID	0.96	0.94	0.94	0.94
X-ray	COVID	0.94	0.94	0.93	0.93
X-ray	Non-COVID	0.94	0.94	0.93	0.92

The performance metrics table for the DenseNet model in binary classification, which includes both CT Scan and X-ray datasets, demonstrates outstanding results. In the CT Scan dataset, the model achieves exceptional accuracy scores of 0.96 for both COVID and Non-COVID classes. Moreover, precision, recall, and F1-Score values consistently reach 0.97 for COVID, underlining the model's precision and effectiveness in identifying COVID cases. Similarly, the DenseNet model's performance in the X-ray dataset is commendable, with an accuracy of 0.94 for both COVID and Non-COVID classes. The COVID class maintains solid precision, recall, and F1-Score values of 0.94, confirming its capability in COVID-19 detection.

These results underscore the superiority of the DenseNet model, which outperforms both the ResNet and Inception-V3 models. Its consistently high performance metrics highlight its potential as a valuable asset in medical diagnostics, with significance for enhancing health care and disease management. Notably, the CT-scan dataset performs more efficiently than the X-ray dataset, demonstrating its strength in COVID-19 classification.

4.3 Multi-Class Classification

In our earlier exploration, we started on a binary classification journey, employing three popular deep learning models(Inception-V3, ResNet, and DenseNet) to distinguish between COVID-19 cases and non-COVID-19 within CT-Scan and X-ray datasets. This initial study provided significant results, including the superiority of CT scans over X-rays and the prominence of the DenseNet architecture for binary classification tasks. we now turn our attention to a more intricate and comprehensive challenge: multi-class chest image classification. Beyond the dichotomy of COVID-19 and non COVID-19 diagnoses, we interact on a mission to distinguish between three distinct classes COVID-19, non COVID-19, and Community-Acquired Pneumonia (CAP).

The dataset used for multi-class classification task, downloaded from kaggle platform under the name of "Large COVID-19 CT scan slice dataset" [46]. To ensure the dataset quality, the closed lung normal slices that do not carry information about inside lung manifestations have been removed. Additionally, they did not include images lacking clear class labels or patient information. In total, they collected 7,593 COVID-19 images (44.39%) from 466 patients, 6,893 normal images (40.30%) from 604 patients, and 2,618 CAP images (15.31%) from 60 patients. Figure 4.12 represents samples from the dataset.

a Example of COVID-19 CT-scan. **b** Example of non COVID-19 CT-scan. **c** Example of CAP CT-scan.

Figure 4.12: Examples from CT-scan dataset of Multi-class classification [46].

Now we will present the outcomes of our classification experiment. We employed three different splitting ratios to properly evaluate model performance: 80:10:10, 70:15:15, and 60:20:20 for training, validation, and testing, respectively. Additionally, we carefully evaluated various combinations of hyperparameter values to identify the most effective ones. After thorough experimentation.

We selected the hyperparameters shown in Table 4.5. The following subsections present detailed results, analyses, and discussions for each model and splitting ratio, shedding light on their adaptability to varying data scenarios and their implications for practical medical image classification.

Hyperparameter	Fixed value
Number of Epochs	50
Learning rate	0.001
Batch size	32
Optimizer	Adam
Loss function	categorical crossentropy

Table 4.5: Selected hyperparameters values for Multi-class classification

4.3.1 Inception V3

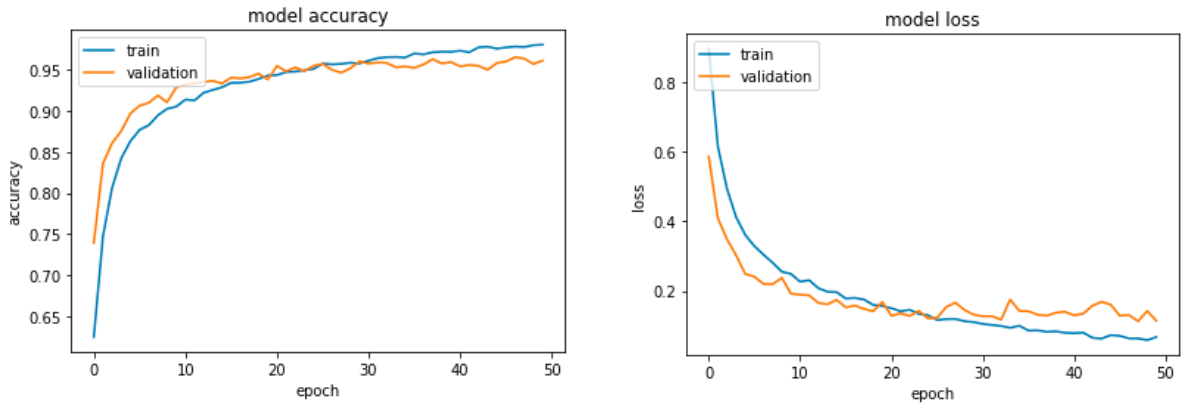
4.3.1.1 First splitting ratio

In this subsection, we discuss the results obtained when employing the Inception-V3 model with an 80:10:10 data splitting ratio. This scenario represents the conventional approach of an 80% training set, 10% validation set, and 10% testing set. The dataset splitting is shown in Table 4.6.

Table 4.6: Distribution of Dataset Samples

	Training	validation	test
Number of samples	13684	1696	1724

Figure 4.13 shows the accuracy and loss graphs obtained after 50 training epochs:



a Accuracy graph. **b** Loss graph.

Figure 4.13: Accuracy and Loss graphs of Inception-V3's first ratio.

From both graphs in Figure 4.13 we see the accuracy and loss curves displaying consistent patterns. The model learned and generalized well. The absence of overfitting or underfitting suggests strong generalization capabilities. These findings open the way to further analysis through the confusion matrix.

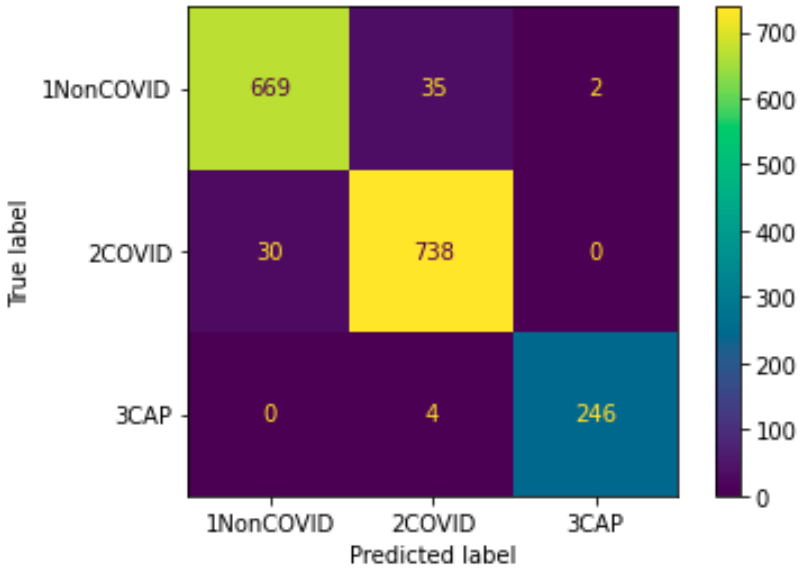


Figure 4.14: Confusion matrix of Inception-V3's first ratio.

The confusion matrix offers valuable insights into the classification performance of the Inception-V3 model, reflecting its proficiency in correctly identifying COVID-19 cases with a notable true positive rate. However, there were instances of misclassifications between COVID and Non-COVID cases and few misclassifications with CAP cases. As we explore the performance metrics, we will gain more comprehensive understanding of its strengths for further enhancement.

Table 4.7: Performance Metrics table of Inception-V3's first ratio.

Class	Accuracy	Precision	Recall	F1-Score
COVID	0.96	0.95	0.96	0.96
Non-COVID	0.96	0.96	0.95	0.95
CAP	0.96	0.96	0.97	0.96

From Table 4.7 the model achieves a commendable accuracy of 96% with splitting ratio of 80:10:10, demonstrates its ability to make accurate predictions. The precision values, averaging at 0.96, indicate a low rate of false positives, while the recall values, averaging at 0.96, signify the model's effectiveness in correctly identifying true positives. This balance between precision-recall is reflected in an F1-Score of 0.96, indicating strong performance.

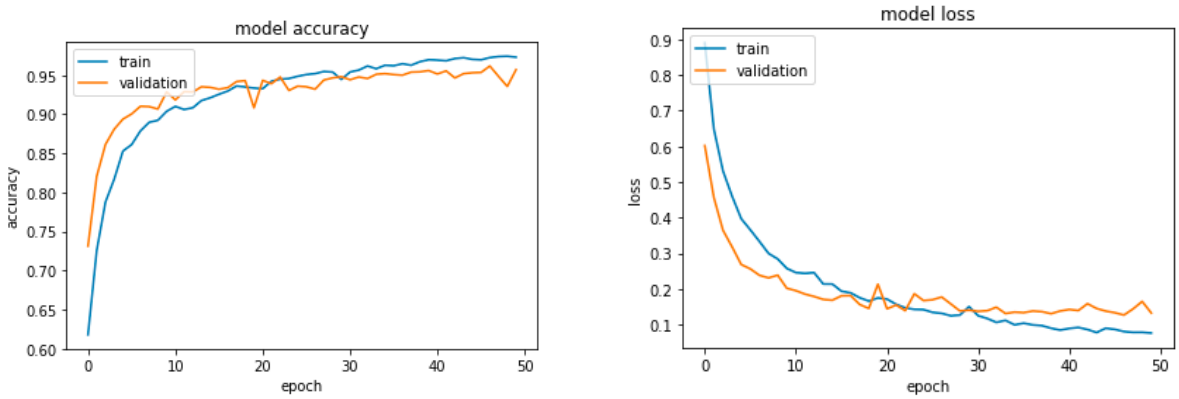
4.3.1.2 Second Splitting ratio

In this part we will put more challenges on the performance by reducing the training data to 70%, and increasing the validation and test data to 15% each as shown in Table 4.8:

Table 4.8: Distribution of Dataset Samples

	Train	validation	test
Number of samples	11973	2560	2571

Figure 4.15 and Figure 4.16 show the accuracy and loss graphs and the confusion matrix obtained after 50 training epochs:



a Accuracy graph. **b** Loss graph.

Figure 4.15: Accuracy and Loss graphs of Inception-V3' second ratio.

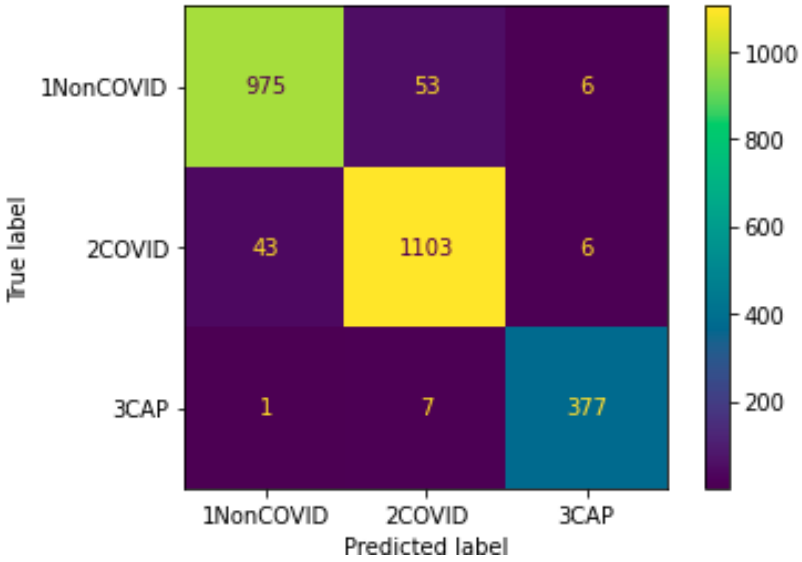


Figure 4.16: Confusion matrix of Inception-V3's second ratio.

The accuracy and loss graphs illustrate consistent patterns and provide confidence in the model's training process and its potential for accurate classification. The accompanying confusion matrix, although revealing a few misclassifications, underscores the model's overall proficiency in distinguishing between classes.

Table 4.9: Performance Metrics table of Inception-V3's second ratio

Class	Accuracy	Precision	Recall	F1-Score
COVID	0.95	0.96	0.94	0.95
Non-COVID	0.95	0.95	0.96	0.95
CAP	0.95	0.96	0.97	0.96

With accuracy of 0.95, precision, recall, and F1-Score values all hovering between 0.94 and 0.96, the model demonstrates a balanced classification performance.

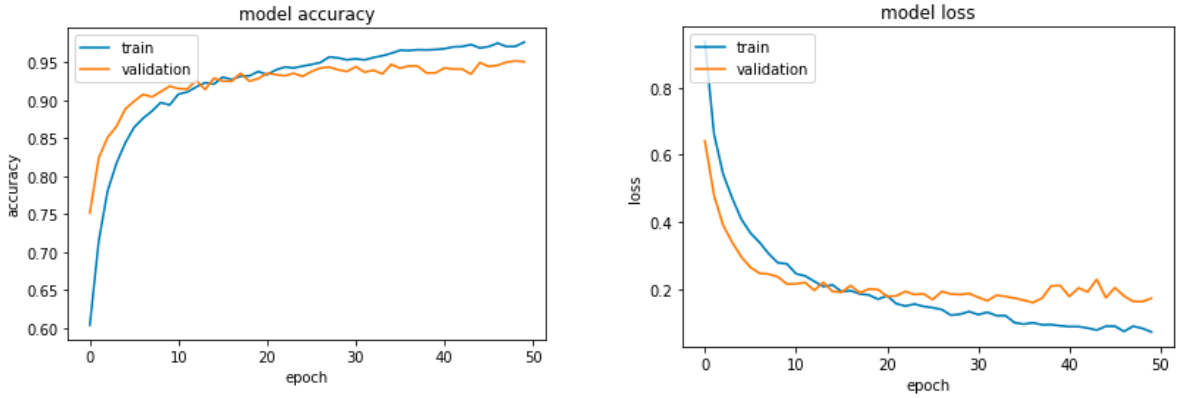
Changing the dataset splitting ratio from 80:10:10 to 70:15:15 can impact the size of the training, validation, and test sets, which, in turn, can affect model performance, generalization, and the ability to detect overfitting. However, despite this shift, the model maintains its proficiency with admirable accuracy, precision, recall, and F1-Score scores across all classes.

4.3.1.3 Third Splitting ratio

In this part we will put more challenges than the last part on the performance by reducing the the training data to 60%, and increasing the validation and test data to 20% each,the distribution shown in Table 4.10

Table 4.10: Distribution of Dataset Samples

	Train	validation	test
Number of samples	10263	3424	3417



a Accuracy graph.

b Loss graph.

Figure 4.17: Accuracy and Loss graphs of Inception-V3's third ratio.

Throughout training, and with reduced training data both training and validation accuracy exhibited consistent upward trends as shown in Figure 4.17, indicating effective learning

from the data, and loss curves showed steady convergence, reflecting efficient error minimization. Further information needed from the confusion matrix.

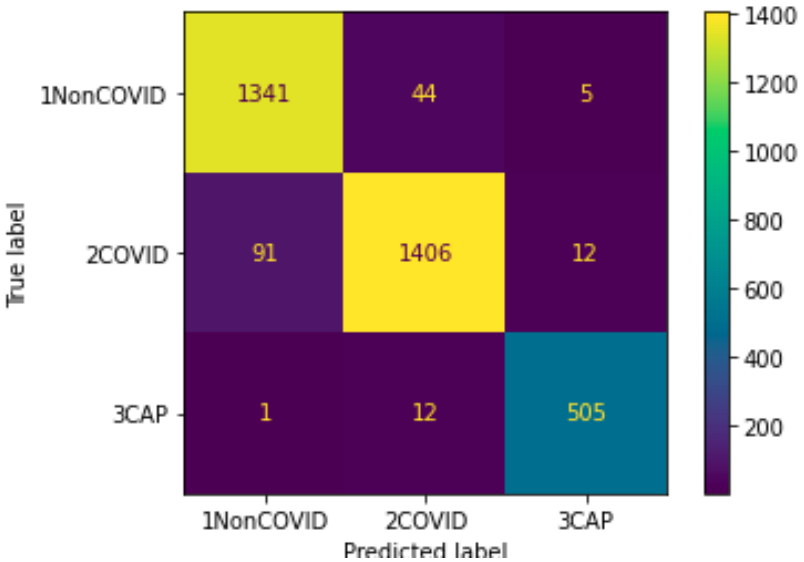


Figure 4.18: Confusion matrix of Inception-V3's third ratio

The confusion matrix shown in Figure 4.18 shows similar trend with the previous confusion matrix but with more missclassified samples.

Further analysis through performance metrics will help us discern any nuanced differences and offer insights into the model's adaptability across varying data distributions.

Table 4.11: Performance Metrics table of Inception-V3's third ratio

Class	Accuracy	Precision	Recall	F1-Score
COVID	0.95	0.93	0.95	0.95
Non-COVID	0.95	0.95	0.93	0.95
CAP	0.95	0.96	0.96	0.96

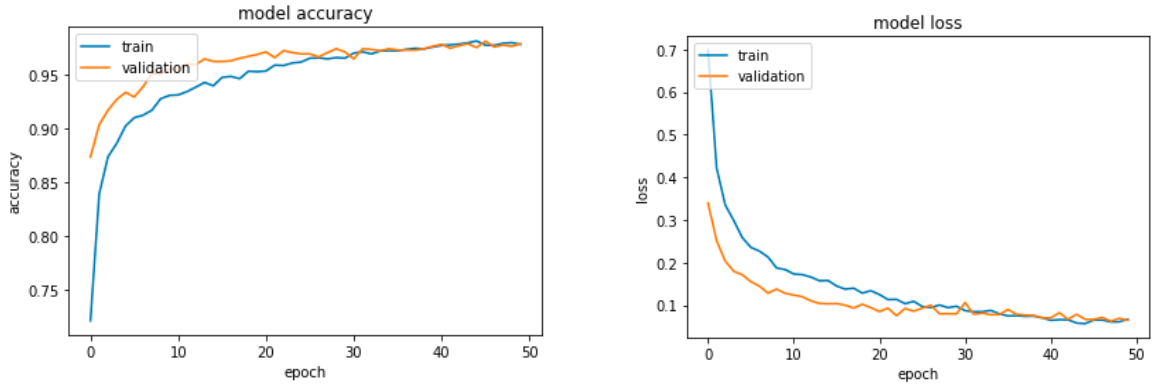
From Table 4.11 The model show overall good performance starting by high accuracy of 0.95 and balance between precision and recall leading to high F1-score, when considering different data split ratios, we find that the ratio of 70:15:15 produces even better results. Moreover, the most impressive performance is achieved with the 80:10:10 data split ratio.

4.3.2 ResNet

4.3.2.1 First Splitting ratio

In this section, we present the results obtained from training the ResNet model under the 80:10:10 dataset splitting ratio. We begin by analyzing the accuracy and loss graphs in Figure 4.19b to gain more details into the model's training progress. Subsequently,

we move to the confusion matrix to assess classification performance.Finally, we present the performance metrics, including accuracy, precision, recall, and F1-Score to have the overall analysis.



a Accuracy graph.

b Loss graph.

Figure 4.19: Accuracy and Loss graphs of ResNet's first ratio.

From Figure 4.19 the accuracy and loss graphs for the ResNet model built a highly successful training process. In addition, both training and validation accuracy consistently improved, indicating the model's ability to learn effectively from the data.

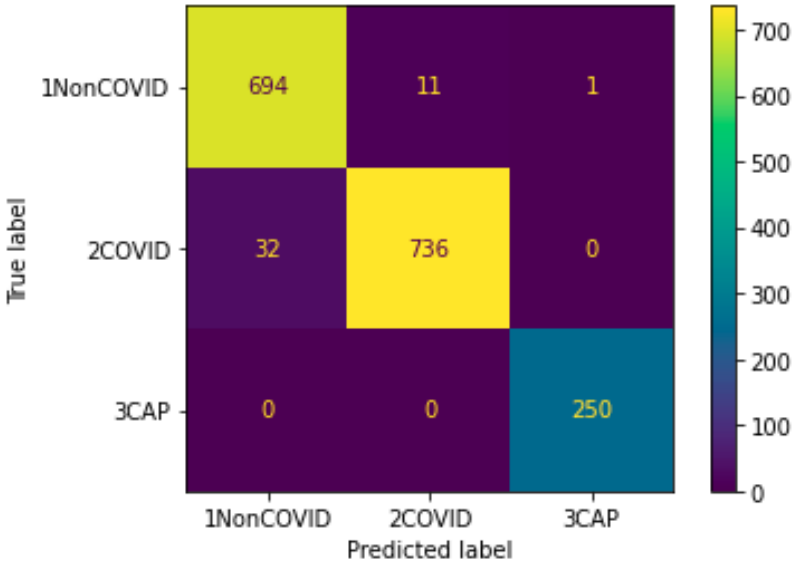


Figure 4.20: Confusion matrix of ResNet's first ratio.

In the confusion matrix shown in Figure 4.20 we see strong classification performance, with high true positive rate for the CAP class with no misclassified samples. While there are 32 and 12 misclassified samples for COVID and Non-COVID, respectively. In addition with the graphs and confusion matrix, further information will be discovered from the performance metrics in Table 4.12

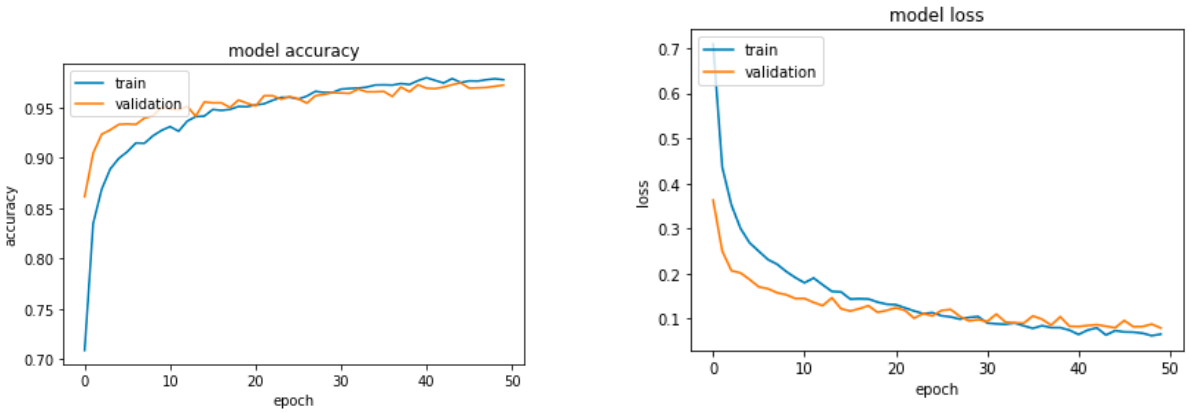
Table 4.12: Performance Metrics table of ResNet's first ratio.

Class	Accuracy	Precision	Recall	F1-Score
COVID	0.97	0.98	0.96	0.97
Non-COVID	0.97	0.96	0.98	0.97
CAP	0.97	1.00	1.00	1.00

The performance metrics in Table 4.12 showcases exceptional results for our ResNet model trained under the 80:10:10 dataset splitting ratio. Across all classes(COVID, Non-COVID, and CAP) the model consistently reaches outstanding accuracy of 0.97, precision, recall, and F1-Score values, all hovering around the remarkable 0.96 to 0.98 range for both COVID and Non COVID classes and strong 1.00 for CAP class. Of particular note is the model's flawless performance in distinguishing CAP cases, with a precision and recall of 1.00, underscoring its absolute accuracy in identifying this condition. These results collectively affirm the ResNet model's exceptional classification capabilities, highlighting its potential as a highly reliable diagnostic tool for CT scans across various medical conditions.

4.3.2.2 Second Splitting ratio

With the same parameters we shift from 80:10:10 to 70:15:15.The shift to 70:15:15 ratio was prompted by a need to allocate a larger proportion of the dataset to both the validation and test sets.



a Accuracy graph. **b** Loss graph.

Figure 4.21: Accuracy and Loss graphs of ResNet's second ratio.

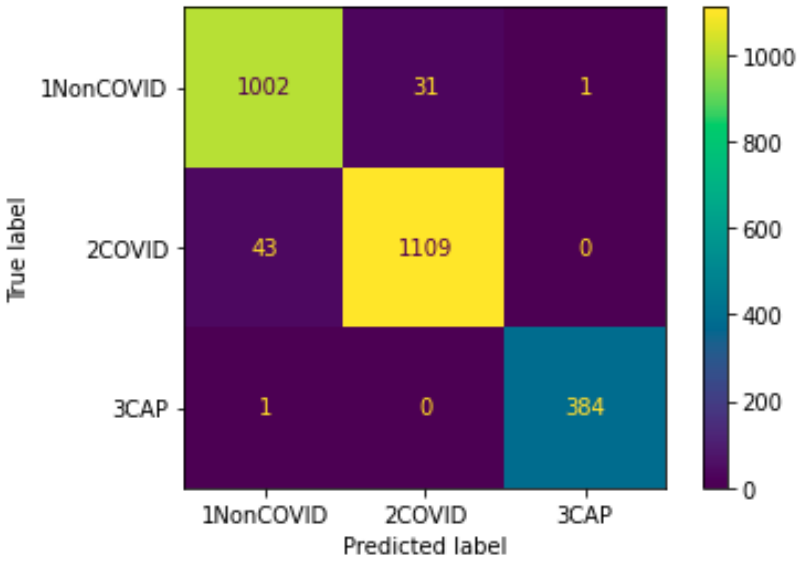


Figure 4.22: Confusion matrix of ResNet's second ratio.

The accuracy and loss graphs in Figure 4.21 , along with the confusion matrix in Figure 4.22, present a comprehensive view of our model's performance under the 70:15:15 dataset splitting ratio. The accuracy and loss graphs demonstrate consistent progress and convergence, suggesting effective learning and stability during training. The confusion matrix reveals a proficient classification performance. While the presence of misclassified samples highlights the inherent complexity of the task.

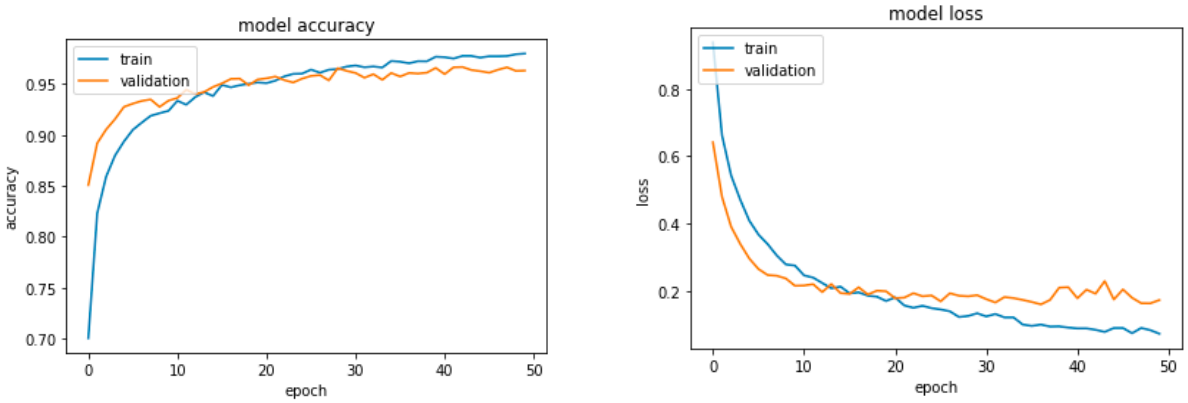
Table 4.13: Performance Metrics table of ResNet's second ratio.

Class	Accuracy	Precision	Recall	F1-Score
COVID	0.96	0.97	0.96	0.97
Non-COVID	0.96	0.96	0.97	0.96
CAP	0.96	1.00	1.00	1.00

From the Table 4.13 we can see strong performance and high metrics for all classes, but looking to the previous ratios metrics we can note that the ratio of 80:10:10 gives the highest performance followed by 70:15:15 then 60:20:20.

4.3.2.3 Third Splitting ratio

Moving to 60:20:20 ratio to challenge more the model performance. Accuracy and Loss graphs shown in Figure 4.23.



a Accuracy graphs.

b Loss graph.

Figure 4.23: Accuracy and Loss graphs of ResNet's third ratio.

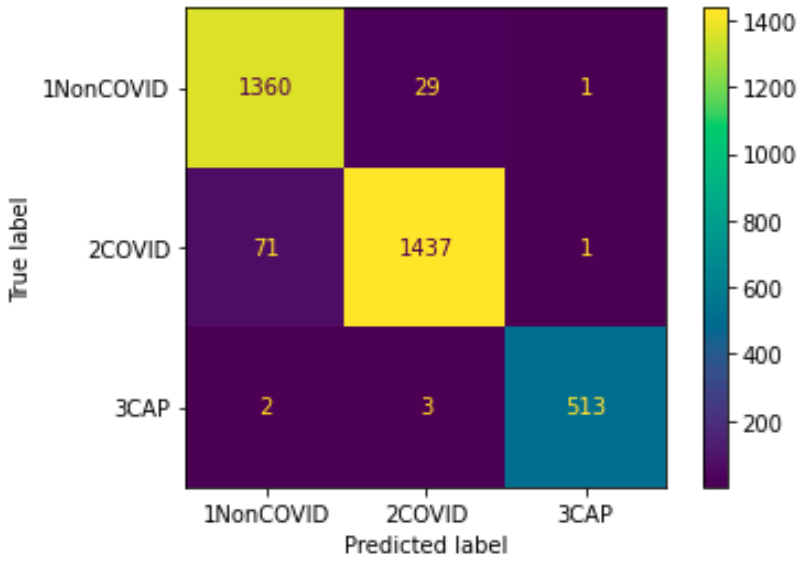


Figure 4.24: Confusion matrix of ResNet's third ratio.

Both training and validation accuracy displayed steady rising trends during training and after lowering training data to 60%, as seen in Figure 4.23, indicating effective learning from the data, and loss curves showed steady convergence, reflecting efficient error minimization. In addition the confusion matrix in Figure 4.24 demonstrates an efficient classification performance. While the existence of misclassified samples emphasizes the task's essential difficulty.

Table 4.14: Performance Metrics table of ResNet's third ratio.

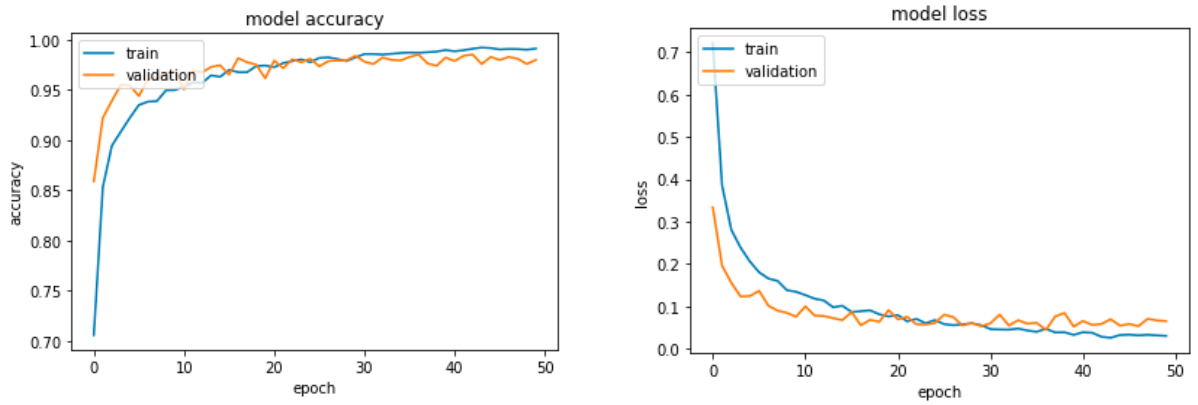
Class	Accuracy	Precision	Recall	F1-Score
COVID	0.96	0.97	0.95	0.97
Non-COVID	0.96	0.95	0.97	0.96
CAP	0.96	1.00	1.00	1.00

From Table 4.14 the model exhibits generally strong performance, start with a high accuracy of 0.96 and a balance between recall and precision, resulting in a high F1-score. When other data split ratios are considered, we discover that a ratio of 70:15:15 generates even better results. Furthermore, the best performance is obtained with an 80:10:10 data split ratio.

4.3.3 DenseNet

4.3.3.1 First Splitting ratio

In this part, we state the results of training the DenseNet model with different splitting ratios starting with ratio of 80:10:10. To get insight into the model's training process, we begin by studying the accuracy and loss graphs. The confusion matrix is then used to evaluate classification performance.Finally, we provide the performance measures, which include accuracy, precision, recall, and F1-Score.



a Accuracy graph.

b Loss graph.

Figure 4.25: Accuracy and Loss graphs of DenseNet's first ratio.

Both graphs in Figure 4.25 demonstrate consistent progress throughout the 50 epochs. Further information is described in the confusion matrix in Figure 4.26 :

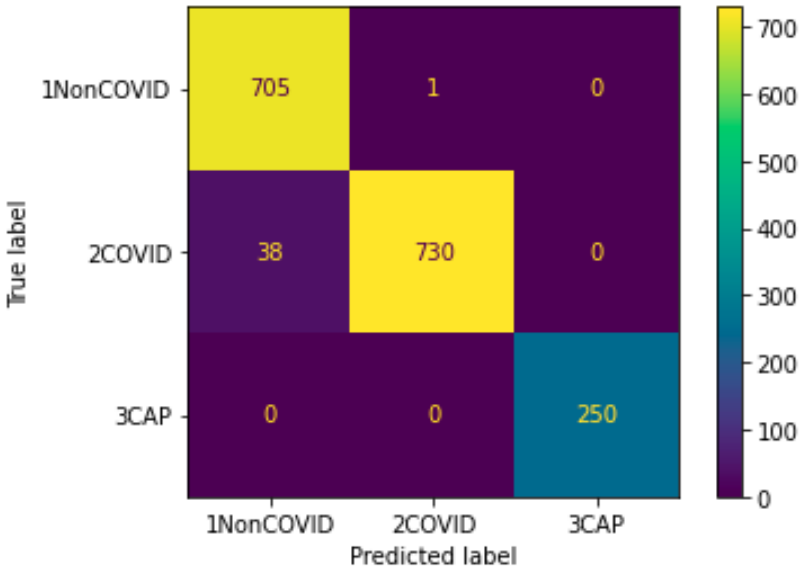


Figure 4.26: Confusion matrix of DenseNet's first ratio.

The confusion matrix for our DenseNet model trained with the 80:10:10 dataset splitting ratio and 50 epochs indicates an exceptional performance, especially in identifying CAP cases with no misclassifications and only one missclassified sample for Non COVID. This perfect detection of CAP and Non COVID displays the model's excellent accuracy in detecting this essential disease. While there have been occasional instances of misclassification, most notably 38 false negatives for COVID, overall classification performance has remained strong. These findings show the model's ability to identify between COVID, Non-COVID, and CAP instances, as well as its promise as a reliable diagnostic tool for CT scan categorization, with a particular strength in CAP and Non COVID identification. Table 4.15 below gives the performance metrics of our model.

Table 4.15: Performance Metrics table of DenseNet's first ratio.

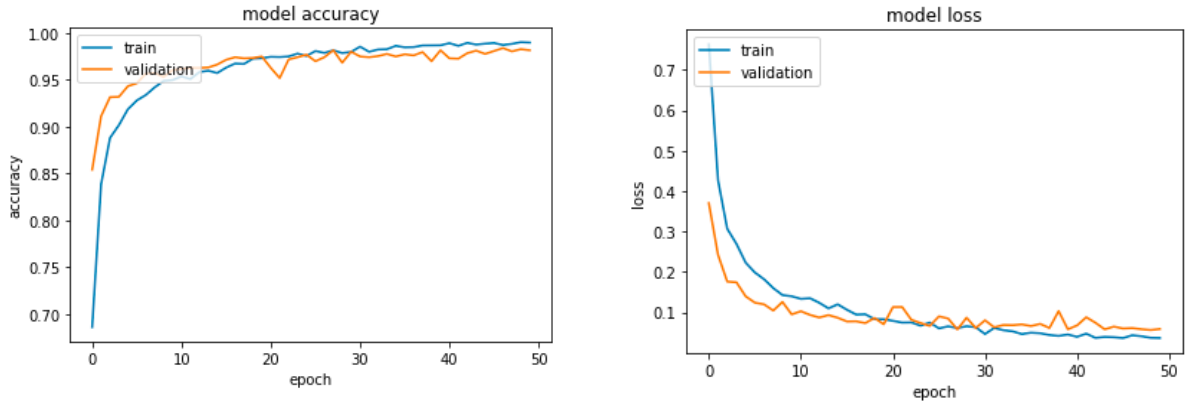
Class	Accuracy	Precision	Recall	F1-Score
COVID	0.98	1.00	0.95	0.97
Non-COVID	0.98	0.95	1.00	0.97
CAP	0.98	1.00	1.00	1.00

Performance metrics shown in Table 4.15 demonstrates an extraordinary degree of performance across all classes, starting with the high accuracy value of 0.98 for all classes and the notable superior level of CAP class at 1.00 with all metrics, and high scores for other classes with precision, recall, and F1-Score, all hanging around the 0.95 to 0.98 area.

4.3.3.2 Second Splitting ratio

With the same parameters we shift from 80:10:10 to 70:15:15. This shift will put additional challenges on the performance evaluation. As shown in the graphs of accuracy and loss

in Figure 4.27 below:



a Accuracy graph. **b** Loss graph.

Figure 4.27: Accuracy and Loss graphs of DenseNet's second ratio.

Both graphs in Figure 4.27 demonstrate consistent progress throughout the 50 epochs. Further information is described in the confusion matrix in Figure 4.28 :

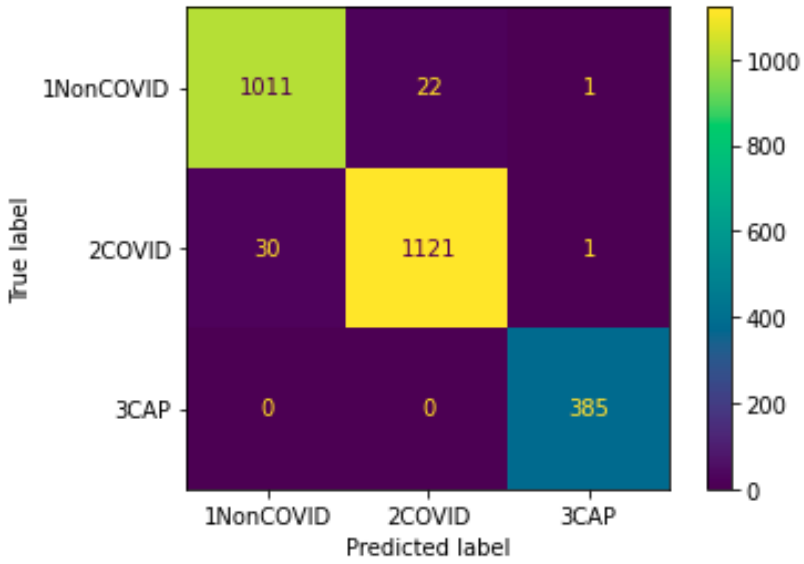


Figure 4.28: Confusion matrix of DenseNet's second ratio.

Our DenseNet's confusion matrix displays a mixed performance. While the model succeeds at correctly recognizing CAP situations with no misclassifications, it exhibits slightly less performance in discriminating between COVID and Non-COVID cases. Notably, 23 Non-COVID cases and 31 COVID cases have been misclassified. This suggests that the model is having difficulties distinguishing between these two classes, maybe due to similarities in imaging features.In addition, further information will be displayed in Table 4.16 below.

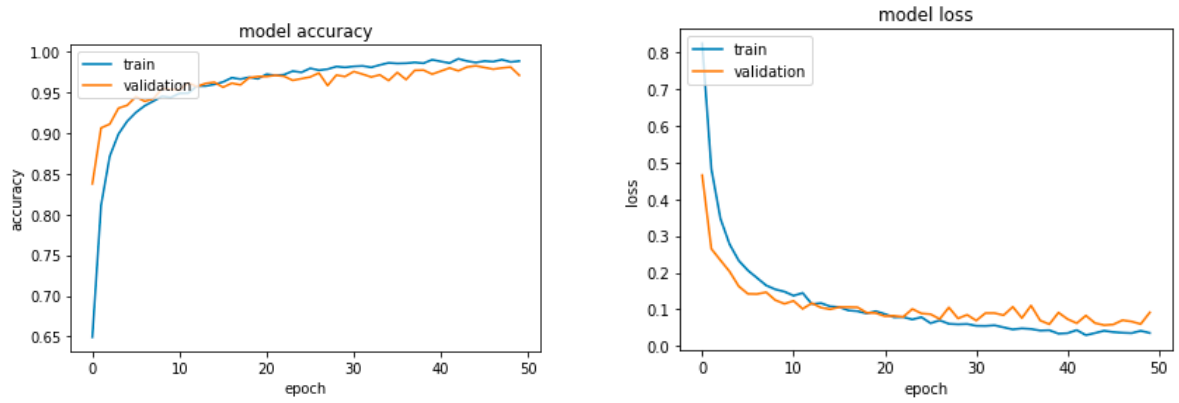
Table 4.16: Performance Metrics table of DenseNet's second ratio.

Class	Accuracy	Precision	Recall	F1-Score
COVID	0.97	0.98	0.97	0.98
Non-COVID	0.97	0.97	0.98	0.98
CAP	0.97	0.99	1.00	1.00

From Table 4.16 we see high performance with accuracy of 0.97 and high performance measures for all class ranges between 0.97 and 0.99 and strong value of recall and F1-score for CAP class at 1.00 .

4.3.3.3 Third Splitting ratio

Putting additional challenges on the model's performance leads to the results shown in the following, starting by the accuracy and loss graphs shown in the Figure 4.29.



a Accuracy graph.

b Loss graph.

Figure 4.29: Accuracy and Loss graphs of DenseNet's third ratio.

The graphs in Figure 4.29 shows positive trend through the 50 epochs with consistent pattern.

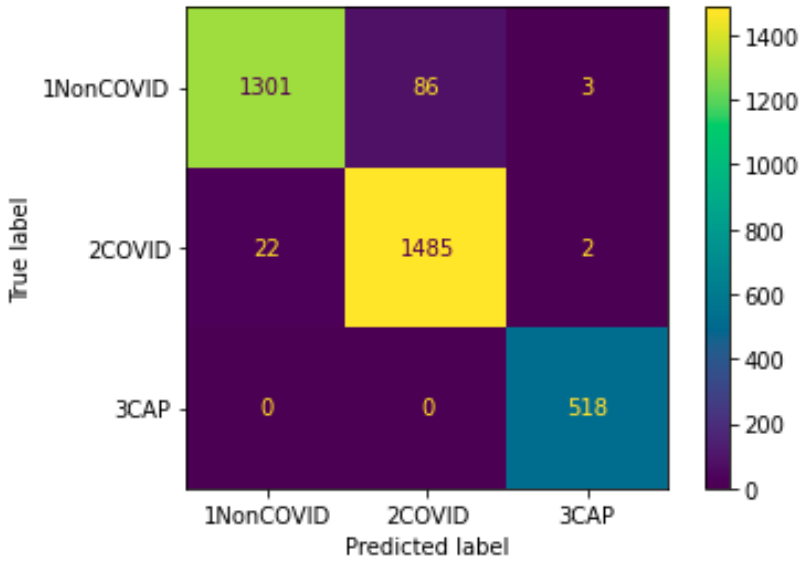


Figure 4.30: The classification results obtained by DenseNet's third ratio.

From confusion matrix in Figure 4.30 even after putting additional challenges the model's ability to correctly identify CAP cases without any misclassifications is remarkable. While there are 24 misclassified COVID cases and 89 misclassified Non-COVID cases. While the additional information will be discovered from the performance metrics in Table 4.17

Table 4.17: Performance Metrics table of DenseNet's third ratio.

Class	Accuracy	Precision	Recall	F1-Score
COVID	0.97	0.95	0.98	0.96
Non-COVID	0.97	0.98	0.94	0.96
CAP	0.97	0.99	1.00	1.00

from Table 4.17 we see DenseNet demonstrates exceptional results across all classes.the model achieves an accuracy of 0.97 for COVID, Non-COVID, and CAP classes, indicating its adaptability in accurately classifying diverse medical conditions. Precision values for all classes consistently exceed 0.95, underlining the model's precision in correctly identifying each class. Moreover, recall values are notably high, with the CAP class achieving perfect recall, signifying the model's ability to effectively capture all instances of this condition. The F1-Score values of approximately 0.96 demonstrate the balanced performance of the DenseNet model across all classes. These outstanding results reflect the model's strength in multi-class classification, making it an effective tool in medical diagnostics and disease classification.

4.4 General Discussion

The binary classification of chest images using CT scan and X-ray datasets was the core focus of the first part of this study, which used three deep learning models (Inception-V3, ResNet, and DenseNet) and fixed 20 training epochs, and a combination of multiple hyperparameters settings presented in Table 4.1. The distribution data of 60% for training, 20% for validation and test each, comes up from the unique characteristics of the research and data constraints. With limited data availability, allocating a larger share (60%) for training enhances the model's ability to learn complex patterns. The 20% reserved for validation facilitates rigorous model evaluation and hyperparameter optimization. Simultaneously, the remaining 20% for testing enables thorough performance evaluation on previously unseen data. This split balances the need for model learning and generalization while maintaining statistical significance in the results.

Inception-V3 reaches commendable accuracy across both datasets,with clear representation of accuracy and loss curves and a confusion matrix particularly excelling in CT scan data with an accuracy of 0.92 for both COVID and non-COVID classes. These findings are corroborated by precision and recall values that underline its proficiency in classification. However, as compared to X-rays, CT scan data has a more balance between precision and recall.

ResNet also exhibited remarkable confusion matrix performance with a bit higher accuracy of 0.95 for both CT scan and X-ray datasets, coupled with high precision and recall values, signifying its strong performance in accurately discerning COVID-19 cases. This consistency and precision-recall balance across both datasets make ResNet a compelling candidate for the best model.

Moreover, DenseNet showcased exceptional accuracy of 0.96 and an outstanding F1-Score 0.97 for CT-scan dataset, highlighting its strong performance and the reliability of its predictions. These results position DenseNet as the strongest model at this task, with clear superiority of CT-scan dataset.

Moving to the second part, our study has moved on from binary classification to multi-class classification. In addition to COVID-19 and non-COVID classes, we included Community-Acquired Pneumonia (CAP) . This approach shift introduces new difficulties and insights on model performance and dataset relevance, we expanded our investigation by introducing a single CT scan dataset and varying splitting ratios of 80:10:10, 70:15:15, and 60:20:20 for each model (Inception-V3, ResNet, and DenseNet). The aim was to impose additional challenges on the models and ascertain their adaptability to diverse scenarios.

Comparing the performance of the Inception-V3 model with different ratios, it is evident that 80:10:10 achieved the strongest performance, with only 37 misclassified samples for non-COVID and 30 for COVID cases, and none for CAP in the confusion matrix. The performance metrics table (Table 4.7) showed Inception-V3's capabilities, with accuracy,

precision, recall, and F1-Score values of 0.96 across all classes. These results exhibit Inception-V3's ability to effectively classify COVID-19, non-COVID-19, and CAP cases under the specified conditions.

Moving to ResNet model also presented an outstanding performance, especially with the ratio of 80:10:10 clearly with highest accuracy and best confusion matrix performance with only 12 misclassifications for non-COVID and 32 for COVID cases, and none for CAP. The performance metrics table (Table 4.12) reflected ResNet's exceptional capabilities, with accuracy, precision, recall, and F1-Score values of 0.97 or higher across all classes.

Moreover, DenseNet delivered the best performance in the confusion matrix, with just one misclassification for non-COVID cases, 38 for COVID, and none for CAP under the 80:10:10 splitting ratio with total of 39 misclassified samples.The performance metrics table (Table 4.15) highlighted the proficiency of the model, with accuracy, precision, recall, and F1-Score values consistently at or above 0.98 across all classes.

4.5 Conclusion

In conclusion, this chapter has presented an analysis of binary and multi-class chest image classification using three different deep learning models (Inception-V3, ResNet, and DenseNet) across varying datasets and different splitting ratios. Our findings confirmed the significance of model selection and dataset choice in achieving accurate and reliable respiratory condition diagnosis. Notably, DenseNet stood up as the best performer, excelling in both binary and multi-class classification with an 80:10:10 splitting ratio across various evaluation metrics. Furthermore, the consistent superiority of the CT-scan dataset over X-ray dataset across different models and scenarios highlights its critical role in improving the accuracy and clinical relevance of respiratory disease detection.

General Conclusion

After the COVID-19 pandemic, respiratory disorders are a developing global concern that requires early and precise identification for effective care and control. In this research project We proceed properly into the topic of chest image classification using deep learning techniques, with the primary objective of improving the identification of respiratory disorders. During this study, we used deep learning's strong capabilities, deploying three differentiated models Inception-V3, ResNet, and DenseNet, to examine the complex environment of chest image categorization. Our work extended beyond model selection to include the essential choice of dataset (CT-scans and X-rays) as a major component.

The findings from our deep research are stunning. Among the models under evaluation, DenseNet stood out among the models tested, constantly giving high accuracy, precision, recall, and F1-Score. Its outstanding performance, especially in the multi-class classification scenario using the 80:10:10 splitting ratio, highlights its potential for clinical applications, where the precise identification of respiratory conditions holds life-saving significance.

Furthermore, our findings shed light on the importance of dataset quality. It became evident that the CT scan dataset consistently outperformed X-rays across all models,due to their superior resolution, 3D imaging capabilities, tissue differentiation, quantitative analysis, and enhanced visualization of pathologies. confirming its key significance in producing solid chest classification outcomes. This discovery illustrates the critical relevance of data quality in medical imaging and highlights why CT scans are better than X-rays.

The significant limitations encountered in our study, particularly with CT scan and X-ray datasets, involved several critical factors. These included the challenge of limited dataset sizes; which impacted model generalization. While Class imbalance; affecting the model's performance. And the presence of noise, variability, and artifacts in medical images posed complexities demanding advanced pre-processing solutions. Furthermore, The limitations regarding hardware and resource requirements involve the need for significant computational power, including high-performance GPUs or TPUs for model training and inference, large memory for storing models and data, and enough storage space for datasets. These limitations collectively emphasize the need for ongoing research and methodological improvement in order to fully realize the potential of deep learning in applications.

The future work of this study holds a great promise for advancing the field. Firstly, We can use this research work presented as the groundwork for another research in the medical field and extend the current binary and multi-class classification models to include more types of respiratory conditions and diseases, this can help in the accurate and early diagnosis of a larger range of diseases. Moreover, developing more advanced deep learning models that can achieve higher accuracy rates and be developed for specific diseases. Furthermore, focus on making AI-based medical image classification accessible and affordable to healthcare providers and facilities worldwide, including those in underserved regions.

Finally, this study represents a step forward in the search of more precise and efficient medical imaging systems, with the potential to improve healthcare outcomes and the quality of life for countless people worldwide. The need for accuracy and efficiency in healthcare remains critical, and our study confirms the revolutionary impact of deep learning in medical image classification.

References

[1] "Lung disease - national institute of environmental health sciences (niehs)," National Institute of Environmental Health Sciences (NIEHS). (), [Online]. Available: https://www.niehs.nih.gov/health/topics/conditions/lung-disease/index.cfm.

[2] A. Roguin, "Rene theophile hyacinthe laënnec (1781–1826): The man behind the stethoscope," *Clinical medicine & research*, vol. 4, no. 3, pp. 230–235, 2006.

[3] M. Tubiana, "Wilhelm conrad röntgen and the discovery of x-rays," *Bulletin de l'Academie nationale de medecine*, vol. 180, no. 1, pp. 97–108, 1996.

[4] "Lung anatomy - seer training," SEER Cancer Statistics. (), [Online]. Available: https://training.seer.cancer.gov/lung/anatomy/.

[5] A. L. Association. "How lungs work." (), [Online]. Available: https://www.lung.org/lung-health-diseases/how-lungs-work.

[6] S. Bharati, P. Podder, and M. R. H. Mondal, "Hybrid deep learning for detecting lung diseases from x-ray images," *Informatics in Medicine Unlocked*, vol. 20, p. 100 391, 2020.

[7] MedlinePlus. "Atelectasis." (), [Online]. Available: https://medlineplus.gov/ency/article/000066.htm.

[8] C. Clinic. "Lung cancer." (), [Online]. Available: https://my.clevelandclinic.org/health/diseases/4375-lung-cancer.

[9] C. for Disease Control and Prevention. "About covid-19." (), [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/your-health/about-covid-19.html.

[10] *Covid-19 x-ray and ct scan image dataset*, https://www.kaggle.com/datasets/ssarkar445/covid-19-xray-and-ct-scan-image-dataset.

[11] C. for Disease Control and Prevention. "Covid-19 symptoms." (), [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html.

[12] C. for Disease Control and Prevention. "Covid-19 testing overview." (), [Online]. Available: https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/testing.html.

[13] B. Zhou, B. Lou, J. Liu, and J. She, "Serum metabolite profiles as potential biochemical markers in young adults with community-acquired pneumonia cured by moxifloxacin therapy," *Sci Rep*, vol. 10, no. 1, p. 4436, Mar. 2020. DOI: 10.1038/s41598-020-61290-x.

[14] J. Smith and A. Johnson, "The influence of polyphenol compounds on human gastrointestinal tract microbiota," *PLOS ONE*, vol. 16, no. 5, e0250688, 2021. DOI: 10.1371/journal.pone.0250688.

[15] I. El Naqa and M. J. Murphy, *What is machine learning?* Springer, 2015.

[16] MonkeyLearn. "What is a classifier?" (), [Online]. Available: https://monkeylearn.com/blog/what-is-a-classifier/.

[17] W. S. Noble, "What is a support vector machine?" *Nature biotechnology*, vol. 24, no. 12, pp. 1565–1567, 2006.

[18] J. Zou, Y. Han, and S.-S. So, "Overview of artificial neural networks," *Artificial neural networks: methods and applications*, pp. 14–22, 2009.

[19] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, 2012, pp. 1097–1105.

[21] D. Silver, A. Huang, C. J. Maddison, *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[22] M. Banoula. "What is perceptron: A beginners guide for perceptron." (2023).

[23] Wikipedia. "Artificial neural network." (), [Online]. Available: https://en.wikipedia.org/wiki/Artificial_neural_network.

[24] A. Geron, "Neural networks and deep learning," in *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O'Reilly, 2019.

[25] "Perceptrons and multi-layer perceptrons." (2022), [Online]. Available: https://indiantechwarrior.com/perceptrons-and-multi-layer-perceptrons/.

[26] M. Awad and R. Khanna, "Deep neural networks," in *Research Gate*, 2015, pp. 127–147.

[27] R. Sharma. "Deep learning activation functions: Their mathematical implementation." (2022), [Online]. Available: https://medium.com/nerd-for-tech/deep-learning-activation-functions-their-mathematical-implementation-b620d536d39b.

[28] M. Ashraf and T. H., "Brain tumor detection using convolutional neural," Dhaka, Bangladesh, Jun. 2019.

[29] K. Nyuytiymbiy. "Parameters and hyperparameters." (), [Online]. Available: https://towardsdatascience.com/parameters-and-hyperparameters-aa609601a9ac.

[30] W. Samek, G. Montavon, S. Lapuschkin, C. J. Anders, and K. Müller, "Explaining deep neural networks and beyond: A review of methods and applications," *Proceedings of the IEEE*, vol. 109, no. 3, Mar. 2021.

[31] S. Saha. "A comprehensive guide to convolutional neural networks: The eli5 way." (2023), [Online]. Available: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53.

[32] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu, and F. E. Alsaadi, "A survey of deep neural network architectures and their applications," *IEEE Access*, vol. 8, pp. 132 907–132 935, 2020.

[33] A. Esteva, A. Robicquet, B. Ramsundar, *et al.*, "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, 2019. DOI: 10.1038/s41591-018-0316-z.

[34] "Kaggle website." (), [Online]. Available: https://www.kaggle.com/.

[35] "Python official website." (), [Online]. Available: https://www.python.org/.

[36] "Tensorflow documentation." (), [Online]. Available: https://www.tensorflow.org/.

[37] "Keras documentation." (), [Online]. Available: https://keras.io/.

[38] "Opencv official website." (), [Online]. Available: https://opencv.org/.

[39] S. Kumari, E. Ranjith, A. Gujjar, S. Narasimman, and H. A. S. Zeelani, "Comparative analysis of deep learning models for covid-19 detection," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 559–565, 2021.

[40] P. with Code. "Densenet." (), [Online]. Available: https://paperswithcode.com/method/densenet.

[41] "Understanding confusion matrix in machine learning." (), [Online]. Available: https://www.analyticsvidhya.com/blog/2020/04/confusion-matrix-machine-learning/.

[42] "Understanding confusion matrix." (), [Online]. Available: https://towardsdatascience.com/understanding-confusion-matrix-a9ad42dcfd62.

[43] G. Developers. "Classification: Accuracy." (), [Online]. Available: https://developers.google.com/machine-learning/crash-course/classification/accuracy.

[44] G. Developers. "Classification: Precision and recall." (), [Online]. Available: https://developers.google.com/machine-learning/crash-course/classification/precision-and-recall.

[45] Educative. "What is the f1 score?" (), [Online]. Available: https://www.educative.io/answers/what-is-the-f1-score.

[46] *Large covid-19 ct slice dataset*, https://www.kaggle.com/datasets/maedemaftouni/large-covid19-ct-slice-dataset.