



# **UNIVERSIDAD DE INVESTIGACIÓN DE TECNOLOGÍA EXPERIMENTAL YACHAY**

**Escuela de Ciencias Biológicas e Ingeniería**

## **Early identification of different pulmonary pathologies using chest x-ray and machine learning algorithms.**

Trabajo de integración curricular presentado como requisito para la  
obtención del título de Ingeniero/a Biomédico/a.

**Autor:**

Kevin Andres Mena Urresta

**Tutor:**

Ph.D. Diego Almeida

Urcuqui, Mayo 2024

# Autoría

Yo, Kevin Andres Mena Urresta, con cédula de identidad 0402035216, declaro que las ideas, juicios, valoraciones, interpretaciones, consultas bibliográficas, definiciones y conceptualizaciones expuestas en el presente trabajo; así como, los procedimientos y herramientas utilizadas en la investigación, son de absoluta responsabilidad de el/la autor/a del trabajo de integración curricular. Así mismo, me acojo a los reglamentos internos de la Universidad de Investigación de Tecnología Experimental Yachay.

Urcuquí, Mayo de 2024.

---

Kevin Andres Mena Urresta

CI: 0402035216

# Autorización de publicación

Yo, Kevin Andres Mena Urresta, con cédula de identidad 0402035216, cedo a la Universidad de Investigación de Tecnología Experimental Yachay, los derechos de publicación de la presente obra, sin que deba haber un reconocimiento económico por este concepto. Declaro además que el texto del presente trabajo de titulación no podrá ser cedido a ninguna empresa editorial para su publicación u otros fines, sin contar previamente con la autorización escrita de la Universidad.

Asimismo, autorizo a la Universidad que realice la digitalización y publicación de este trabajo de integración curricular en el repositorio virtual, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior

Urcuquí, Mayo de 2024.

---

Kevin Andres Mena Urresta

CI: 0402035216

# Dedication

*To my mother, the strongest, most loving, and courageous woman that life has given me the chance to know.*

*And to all the incredible people who were part of this wonderful journey. Thank you for your love, your light and your wisdom. Let this be a sign of my gratitude and affection.*

*Kevin Andres Mena Urresta*

# Agradecimientos

En este apartado me gustaría agradecer enormemente a mi madre por su amor y cariño incondicional, por luchar contra viento y marea para permitirme soñar y nunca dejarme solo en este camino. En verdad mamá solo espero que la vida me alcance para poder devolver siquiera un poco de todo el amor que me das. Que dios siempre le tenga en su santa gloria, que se merece este y mil universos más por su amor y dedicación.

A Dios porque ahora más que nunca sé que su amor infinito está conmigo. Y su herencia de luz late en mí. Om Namaha Shivaya.

A los increíbles profesores y amigos que solo Yachay me habría dado la oportunidad de conocer. Frank, Graciela, Si Amar, Bryan, Ariel y Samanta, en verdad que bonita a sido la vida universitaria pudiendo aprender cada día de gente tan brillante como ustedes. Gracias Frank y Si Amar por abrirme los ojos al inmenso mundo de posibilidades que existen por explorar. Gracias Graciela por ser mi guía, mi amiga y una increíble maestra a la que siempre le agradeceré por aconsejarme y ayudarme a reunir el coraje para hoy estar aquí. De igual manera profes Diego, Fernando, Nelson, Hortensia, Vivian, Paola, y en fin todos en general. Que maravilloso a sido aprender de gente que en verdad ama su trabajo.

Gracias Bryan, Ariel, y Samanta por ser tan increíbles como amigos, como compañeros y como personas en general, de verdad me siento ganador solo de haberles conocido y aprendido de ustedes. Gracias por las risas, las experiencias nuevas y todos esos momentos que atesoro en el corazón.

Gracias a mi familia por siempre estar pendientes y darme sus palabras de aliento, gracias abuelitos Victoria y Felix que aunque no pudieron estar aquí, este triunfo también les pertenece.

Mi gratitud y amor infinito les acompañe siempre en luz.

Kevin Andres Mena Urresta

# Abstract

The present research work focuses on the study of several neural networks trained for the detection of lung diseases. The networks used were ResNet-18, ResNet-50, ResNet 101, VGG19 and Inception V3 and the data set was obtained from NIHCC being 112120 the total number of images with which different experiments and optimization were performed to obtain better results from the training of the networks.

**Keywords:** Pulmonary diseases, Neural networks, InceptionV3, Deep Learning, Image Classification.

# Resumen

El presente trabajo de investigación se centra en el estudio de varias redes neuronales entrenadas para la detección de enfermedades pulmonares. Las redes utilizadas fueron ResNet-18, ResNet-50, ResNet 101, VGG19 e Inception V3 y la base de datos fue obtenida de NIHCC siendo 112120 el número total de imágenes con las que se realizaron distintos experimentos y optimización para obtener mejores resultados de los entrenamientos de las redes.

**Palabras Clave:** Enfermedades pulmonares, Redes neuronales, InceptionV3, Deep Learning, Clasificación de imágenes.

# Contents

Dedication	IV
Agradecimientos	V
Abstract	VI
Resumen	VII
Contents	VIII
List of Tables	XI
List of Figures	XII
1. Introduction	1
1.1. Problem Statement	1
1.2. Justification	2
2. Objectives and Hypothesis Statement	4
2.1. Objectives	4
2.1.1. General Objective	4
2.1.2. Specific Objectives	4
2.2. Hypothesis Statement	5
3. Theoretical Framework	6
3.1. Lungs and Pulmonary conditions of interest	6
3.1.1. Other functions of the lungs	9
3.1.2. Respiratory diseases	10
• Inflammatory lung diseases	10



• Restrictive lung diseases	11
• Respiratory tract infections	12
• Lung cancer	13
• Pleural cavity diseases	14
• Pulmonary vascular diseases	15
3.2. Artificial Intelligence and Neural Networks	16
3.2.1. AI	16
3.2.2. Supervised Learning	19
3.2.3. Convolutional Neural Networks	20
• Convolutional Layer	21
• Activation Layers	21
• Pooling Layer	22
• Fully Connected Layers	23
• Batch Normalization	24
• Dropout	24
3.3. Image Classification	26
3.3.1. Deep Learning and Biomedical Imaging	27
4. Methodology	39
4.1. Data acquisition	40
4.2. Method Threats	41
4.3. Data Preprocessing	41
4.4. Models Development	42
4.4.1. Models Description	43
• VGG nets	43
• Inception	44
• ResNet	45

• DenseNet	46
4.4.2. Framework and hardware acceleration	47
4.5. Metrics	48
5. Results and Discussion	51
5.1. Data acquisition	51
5.2. Data Preprocessing	51
5.3. Testing different deep learning classification models.	55
5.4. Testing different Optimizers and Number of Epochs for the better classification models	57
5.5. Testing a different dataset for a better classification model.	58
5.6. Final Considerations	59
6. Conclusions	62
7. Bibliography	64

# List of Tables

<i>Table 1: Cases of each disease diagnosed per year from 2015 to 2019</i> .....	2
<i>Table 2: Bibliographic summary on neural networks.</i> .....	30
<i>Table 3: Training Initial Results</i> .....	56
<i>Table 4: InceptionV3 results for the initial dataset with different optimizers and epochs.</i> .....	57
<i>Table 5: InceptionV3 results with Adagrad, SGDM, and Adam,for the new dataset were Eddusion replace atelectasis(dataset5).Ccompared with InceptionV3+Adagrad with the initial dataset.</i> .....	59
<i>Table 6: Different models trained in the current experimental process with the precision results obtained per each of the pathologies analyzed</i> .....	60
<i>Table 7: Comparison of results among references.</i> .....	61

# List of Figures

Figure 1: Lungs Internal structure. Own elaboration [8].....	6
Figure 2: Internal and external respiration. Own elaboration. [51].....	7
Figure 3: Diffusion path in the lungs. Own elaboration. [52].....	8
Figure 4: Gas exchange in the lungs. Own elaboration [50], [51]. ....	8
Figure 5: Inflammatory disease with bronchus obstruction.[50].....	10
Figure 6: Restrictive lung disease diagram. Own elaboration. [50].....	11
Figure 7: Different stages of pneumonia and pulmonary infections. Own Elaboration. [50] .....	12
Figure 8: Different types of pulmonary neoplasms and tumors. Own elaboration. [2]. ....	13
Figure 9: Pneumothorax symptoms and stages. Own elaboration [50].....	14
Figure 10: Common Pulmonary vascular diseases. From left to right: Pleural Effusion due to heart disease, Embolism, and Emphysema. Own elaboration. [50].....	15
Figure 11: Subfields of Artificial Intelligence. Own elaboration.....	16
Figure 12: Venn diagram describing machine and deep learning as a part of AI [71].....	17
Figure 13: a) Timeline of some of the highest discoveries in deep learning history since perceptron development until GoogleNet presentation. b) Schematic view of artificial neural networks. Own elaboration [72], [73]. ....	18
Figure 14: Supervised leaning process.....	19
Figure 15: Convolution layer. ....	21
Figure 16: Average Pooling and Max Pooling example.....	23
Figure 17: Fully connected layer.....	24
Figure 18: a) fully connected neural network b) the same network after dropout. ....	25
Figure 19: Flow diagram of the methodology .....	39
Figure 20: Images obtained from the NIHCC data set. a. No finding, b. Nodule, c. Emphysema, d. Infiltration	40
Figure 21: VGG architecture [102]. ....	43
Figure 22: Variation on the location and size of the image .....	44
Figure 23: Inception modules [76] .....	45
Figure 24: Training and test error for a CIFAR-10 net [105]. ....	45
Figure 25: Residual block [105].....	46
Figure 26: A 5-layer dense block in a DenseNet [106]. ....	47
Figure 27: Possible predictions from the network. ....	48
Figure 28: Data set. ....	51

<i>Figure 29: Images labeled with more than one disease.....</i>	<i>52</i>
<i>Figure 30: Structured data from the dataset once the overlapping was eliminated. ....</i>	<i>53</i>
<i>Figure 31: Number of Images per class.....</i>	<i>54</i>
<i>Figure 32: Accuracy results for the models a) resnet18, b) resnet50, c) resnet101, d) densenet121, e) vgg19, y f) InceptionV3. ....</i>	<i>55</i>
<i>Figure 33: Confusion matrix of Resnet 18, InceptionV3, and VGG19 models for the initial dataset. ....</i>	<i>56</i>
<i>Figure 34: Confusion Matrix and Metrics of the model InceptionV3 with SGDM with.....</i>	<i>58</i>
<i>Figure 35: Confusion Matrix of InceptionV3 model for the initial dataset, and the second dataset (Effusion instead of atelectasis).....</i>	<i>59</i>



# 1. Introduction

## 1.1. Problem Statement

After the heart and brain, the lungs should be considered one of the most important organs for the maintenance of human lives[1]. Lungs are responsible for the gas exchange process called respiration, which allows every cell of the human body to take oxygen and liberate carbon dioxide as part of the cell living process[1]–[3]. Unfortunately, the current way humans live and the chemicals they are exposed to daily have become a source of illness that could seriously affect people’s lives. Some of these illnesses cause death if they are not detected at the right time to handle them correctly[4]–[7].

From pneumonia, pulmonary emphysema, pleural effusion, pulmonary hernia, and pulmonary fibrosis to pleural thickening, tuberculosis, and lung cancer, a wide variety of diseases could affect human lungs and produce irreparable harm to people’s quality of life. Some of those illnesses are among the leading causes of death worldwide[5]–[8]. As mentioned before, time is the most relevant aspect that could enormously decrease the probability of an untreatable diagnosis and even avoid the patient’s death. Any harmful or deadly disease could be treated and mostly stopped when it is diagnosed early[4], [8], [9].

Diagnostic imaging plays an essential role in the diagnosis of lung diseases[3], [6]. Symptoms of lung disease are often similar and discriminatory only in advanced stages when the prognosis for the eventually named pathology is already less favorable[10]–[12]. In addition to clinical testing, imaging methods provide valuable information that allows for a complete differential diagnosis. Currently, chest X-rays, computed tomography (CT), and magnetic resonance imaging (MRI) are common imaging modalities that are first prescribed after examining a patient[4], [13]–[15]. These modalities provide high spatial resolution anatomical and structural information about the lung. However, more accurate methods are needed to distinguish lung diseases with overlapping pathophysiology or diagnose them very early[6], [9].

Chest X-ray interpretation is critical for detecting many thoracic diseases, including pneumonia, tuberculosis, and lung cancer. Nowadays, radiologists trained in interpreting chest

X-rays look at the chest X-ray, analyzing the lungs, the heart, and other regions of the radiography[5], [8], looking for clues that might suggest the presence of pneumonia, tuberculosis, lung cancer, or another clinical condition. However, radiograph interpretation is a time-consuming task[16]. Usually, it requires expert radiologists to read the images, increasing the risk of a fatigue-based diagnostic error and a lack of the necessary diagnostic expertise in rural areas of the world, where expert radiologists are not available[15]–[19]. Even in many cities of the world, such as Latinamerica, a shortage of qualified, trained radiologists exists in the healthcare systems, being this gap higher in the public healthcare systems. That is why an artificial intelligence system is needed to interpret chest X-rays as experienced chest radiologists effectively, thus providing a significant benefit in numerous clinical settings, detecting important lung diseases easier[20]–[24]. Furthermore, avoiding fatigue-based diagnostic errors, improve the workflow and the prioritization of public health services[16], [17], and even makes an accessible option for people that do not have ways to contact expert radiologist to diagnose them.

## 1.2. Justification

In Ecuador, pneumonia, chronic obstructive pulmonary disease (COPD), lung cancer, Pulmonary Fibrosis, and bronchitis represent the fifth main pulmonary diseases that kill people in the country. According to the National Institute of Statistics of Ecuador, even in a context out of COVID-19, between 2015 and 2019. Pneumonia, COPD, Lung Cancer, and Pulmonary Fibrosis were related to a mean rate of 6700 deaths per year in Ecuador **Table 1**.

Pneumonia is the first of the pulmonary diseases related to people’s death in Ecuador. Between 2015 and 2018, pneumonia was related to more than three thousand deaths yearly. In 2019 this sign increased to more than four thousand deaths, a tendency of most of the first-fifth diseases mentioned. These numbers show that even in a context out of a chaotic situation such as the COVID-19 pandemic, the health systems of Ecuador were having important problems diagnosing and treating pulmonary diseases effectively.

*Table 1: Cases of each disease diagnosed per year from 2015 to 2019*

<b>DISEASE</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>
----------------	-------------	-------------	-------------	-------------



<b>PNEUMONIA</b>	3310	3427	3879	4132
<b>COPD</b>	1344	1422	1500	1513
<b>LUNG CANCER</b>	750	758	876	776
<b>P FIBROSIS</b>	595	597	734	727
<b>BRONCHITIS</b>	232	228	222	184

Nowadays, advancements in deep learning, image classification, and big datasets have allowed algorithms to equal the interpretation of medical professionals in various medical imaging tasks involving the detection and classification of pathologies such as diabetic retinopathy [25], [26], skin cancer [23], [27]–[30], and lymph node metastases [31]. Additionally, automated determination from chest imaging has received augmenting attention [14], [24], [32], [33], with specialized algorithms evolved for the detection and classification of pulmonary tuberculosis [34], COVID-19 [35]–[39], pneumonia [34], [35], [38], [40]–[42], COPD [20], [21], [34], [43], pneumothorax [44], lung nodules [10], [22], [32], and lung cancer [14], [32], [45]. Using chest radiographs to diagnose several pulmonary pathologies motivates the development of studies intended to detect several pathologies simultaneously [33], [43], [46], [47]. Recently, computing power and large data sets have enabled the development of different significant approaches. The implicit worth of these healthcare approaches is emphasized by the World Health Organization, which estimates that further than 4 billion people lack access to medical imaging expertise [48]. Indeed, even, in developed nations with evolved healthcare systems, an automated system with the ability to effectively interpret chest radiographs could give invaluable utility.

This tool could identify and prioritize the sickest patients, allowing those to receive early diagnosis and adequate treatments even in rural sectors where hospitals do not immediately have radiologists. Furthermore, deep learning algorithms avoid some common and, at some point, understandable limitations that could appear even with experienced radiologists, including fatigue, perceptual biases, and cognitive biases; human limitations could increase diagnostic errors [16], [17], [19], [31]. The algorithm helps the radiologist avoid those perceptual errors and biases by providing information on the presence and the location of different chest X-ray abnormalities [13], [30], [37], [40]. In this way, deep learning algorithms that effectively detect pulmonary pathologies could be used as a novel system to perform faster and more accurate computer-aided diagnoses.

## 2. Objectives and Hypothesis Statement

### 2.1. Objectives

#### 2.1.1. General Objective

Develop a convolutional neural network (CNN) algorithm capable of effectively detecting several pulmonary pathologies from chest X ray images with efficient machine cost and able to be runned in rural parts of Ecuador giving them access to more people to an accurate and on time diagnosis.

#### 2.1.2. Specific Objectives

- Generate and depurate a congruent dataset for the development of the CNN algorithm
- Design several CCN algorithms that effectively and efficiently identify pulmonary pathologies avoiding extensive processing requirements.
- Validate and compare the results of the CNN's designed in terms of accuracy and efficiency with each other and o algorithms reported in the literature to contrast the results obtained and select the better arrangement to be implemented as aid for diagnostics.

## 2.2. Hypothesis Statement

The current thesis project is developed with two important hypothesis to prove:

- About the efficacy of detection by the CCN algorithms designed

**Ho:** It is not possible to develop a CNN algorithm able to effectively identify pulmonary pathologies such as pneumonia, pulmonary nodules, atelectasis, pneumothorax, and pleural thickening, avoiding extensive processing requirements.

**Ha:** It is possible to develop a CNN algorithm to effectively identify pulmonary pathologies such as pneumonia, pulmonary nodules, atelectasis, pneumothorax, and pleural thickening, avoiding extensive processing requirements.

- About the efficiency of the algorithm developed

**Ho:** It is not possible to develop a CNN algorithm able to efficiently identify pulmonary pathologies such as pneumonia, pulmonary nodules, atelectasis, pneumothorax, and pleural thickening, avoiding extensive processing requirements.

**Ha:** It is not possible to develop a CNN algorithm able to efficiently identify pulmonary pathologies such as pneumonia, pulmonary nodules, atelectasis, pneumothorax, and pleural thickening, avoiding extensive processing requirements.

### 3. Theoretical Framework

In this chapter, essential concepts of the study are presented. First, the lungs, their structure, functions, and pulmonary diseases are presented. Then the definitions of artificial intelligence, artificial neural networks, computer vision, supervised learning, deep learning, and image classification are described. After that, the concepts of convolution and how a convolutional neural network works, especially those commonly used for image classification tasks and biomedical image classification tasks.

For the development of this thesis, different specialized search engines, such as PubMed, Google Scholar, Scopus, and Web of Science, were consulted in order to obtain data with the highest quality standards of reliability and reproducibility as the foundation of state of the art and this project in general. This work intends to be descriptive and systematically analytic. Experimental studies are described as a comparative analysis between authors. The documents below cited were selected according to their relevance and contributions to the present work.

#### 3.1. Lungs and Pulmonary conditions of interest

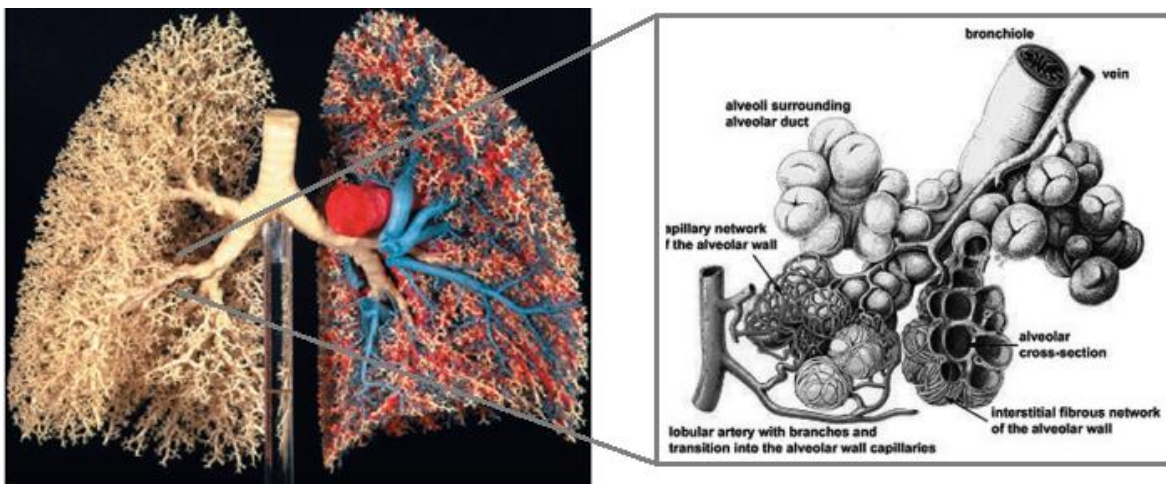


Figure 1: Lungs Internal structure. Own elaboration [8].

The lungs Figure 1 are the two major organs and the centerpiece of the entire respiratory system [2], [49]; Lungs provide the body with a stable and continuous supply of oxygen [1], [6], [50]. As part of their normal work, lungs can take even more than 6 million breaths per year while supporting the correct functioning of the entire body [8].

In an easy way, the lung's main function is to get air from the atmosphere and then introduce the oxygen from the atmospheric air into the bloodstream. From the lungs, oxygen circulates to the rest of the body. Lungs require help from other surrounding structures in the body in order to breathe properly. To breathe human body uses the diaphragm, the intercostal muscles, the muscles of the abdomen, and even some neck muscles [6].

The lungs are located inside the chest cavity on each side of the human heart. There is a slight difference in size between the lungs; the left lung is slightly smaller than the right lung since the heart occupies some space on the left side. Consequently, the right and the left lungs are divided into different number of lobes (sections). The right lung is divided into three lobes while the left is divided into two [4]. When someone breathes in, air enters the airways and travels down into the alveoli. Alveoli are where gas exchange takes place [6], [50].

The circulatory system supports the respiratory system by carrying blood to and from the lungs. The circulatory system supplies nutrients and oxygen to tissues and organs throughout the body. The circulatory system also helps remove carbon dioxide and waste products from cellular respiration Figure 2. Moreover, other body systems, such as the nervous, lymphatic, and immune systems, also work with the respiratory system [3].

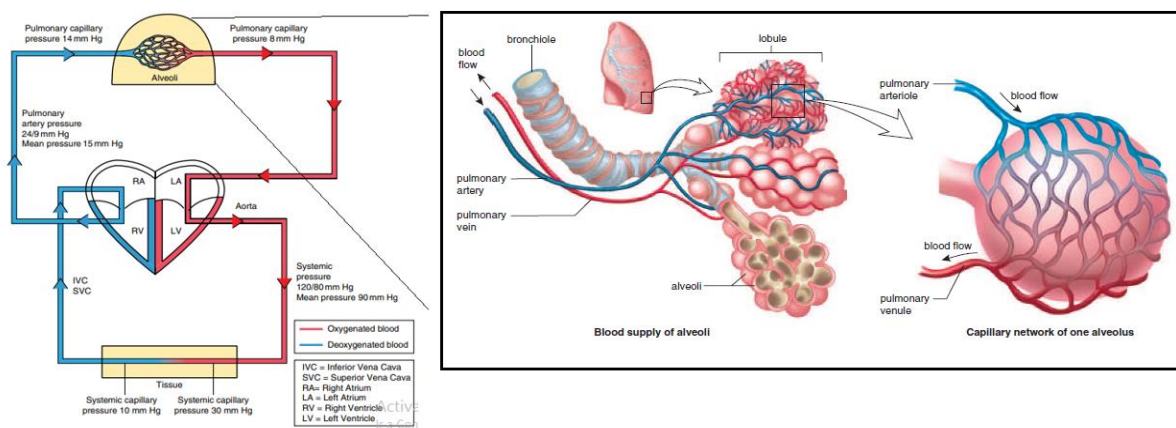


Figure 2: Internal and external respiration. Own elaboration. [51]

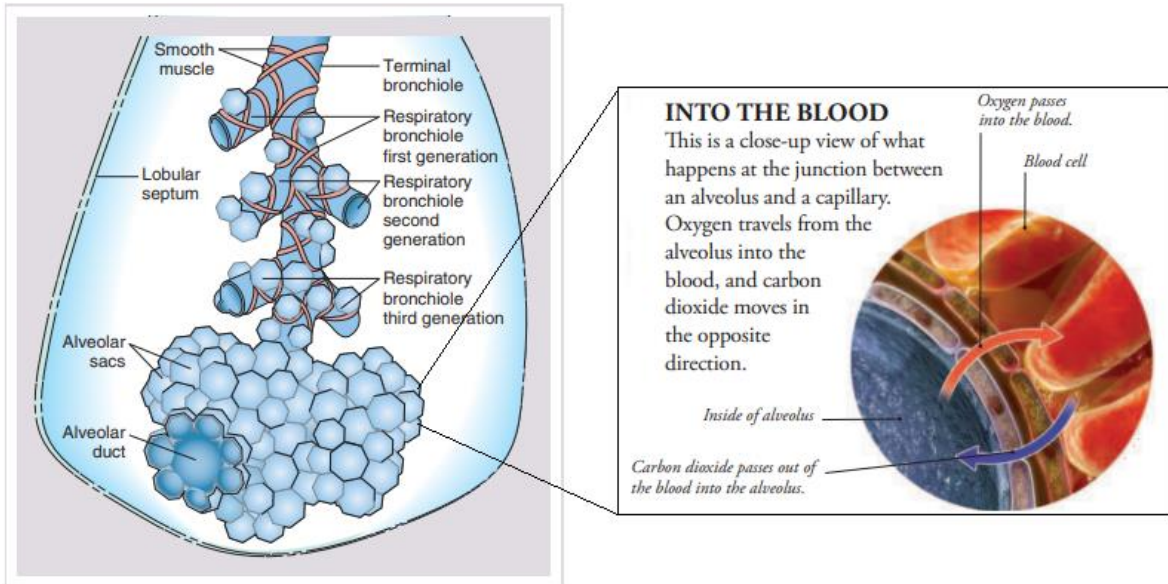


Figure 3: Diffusion path in the lungs. Own elaboration. [52]

Human lungs have about 150 million alveoli. Alveoli are naturally elastic, meaning they can expand and contract at will. Furthermore, they are coated with a substance that confers them the ability to inflate easily [2], [51]. Lungs in their insides pose a substance called surfactant. The surfactant reduces the work it takes the lungs to expand and contracts during breathing. Thus, surfactant helps the lungs inflate more easily when a person breathes in while preventing the lungs from collapsing when you breathe out [4].

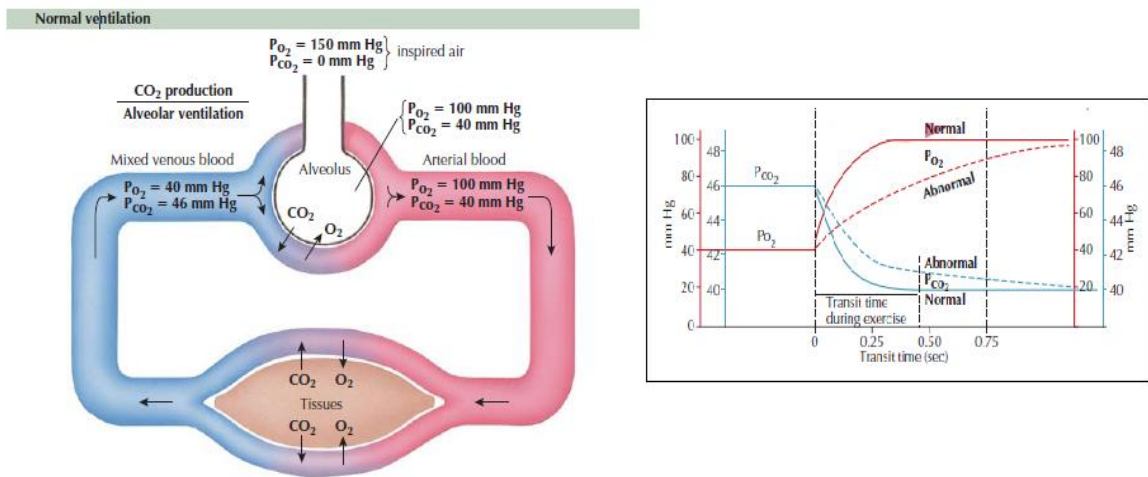


Figure 4: Gas exchange in the lungs. Own elaboration [50], [51].

Blood deficient in oxygen flows through the lungs to the alveoli, where the carbon dioxide in the blood is exchanged for pure oxygen from breathing via hemoglobin Figure 3. Then the oxygen-enriched blood travels to the left side of the heart through the pulmonary veins. From the heart, blood is delivered to the organs and tissues of the body via the circulatory system to ensure proper organ function and to initiate another cycle where carbon dioxide is exchanged [3], [4]. Each of these alveoli is formed by an arrangement of small blood vessels called capillaries that connect to the body's circulatory system, thus allowing the exchange of gases necessary for life. Once in the circulation, carbon dioxide returns to the heart and enters the right side. It then goes to the lungs through the pulmonary artery, which flows from the capillaries back into the alveoli in exchange for the incoming oxygen Figure 4 [52], [53].

### 3.1.1. Other functions of the lungs

Although respiration is the best-known function of the lungs, they carry out other important functions, including:

**pH Balancing:** The presence of high amounts of carbon dioxide can cause the body to become acidic. When the body detects a rise in acidity, the lungs increase the ventilation rate to expel more of this unwanted gas, which helps to balance pH [2].

**Body protection:** Lungs can secrete immunoglobulin A and use mucociliary clearance. Those substances help protect the body from harmful pathogens and infections [51].

**Speech:** Without airflow, a person would be unable to speak [49], [51]. The vocal cords make the sounds that allow communication; these cords are separated when a person is silent. However, when speaking, the cords become tight so that they vibrate and create sounds when the air pushes between them.

### 3.1.2. Respiratory diseases

Respiratory diseases is the name to designate those diseases that can affect any part of the respiratory system, from the upper respiratory tract to the bronchi and down into the alveoli. Some conditions that affect the lungs include inflammatory and restrictive lung diseases, respiratory tract infections, lung cancer, pleural cavity diseases, and pulmonary vascular diseases [4]. Symptoms of lung disease can be subtle. Commonly an early sign of lung disease can be fatigue. Other signs could be trouble breathing, shortness of breath, inability or decreased ability to exercise, coughing with or without blood or mucus, and pain when breathing in or out [8].

- Inflammatory lung diseases

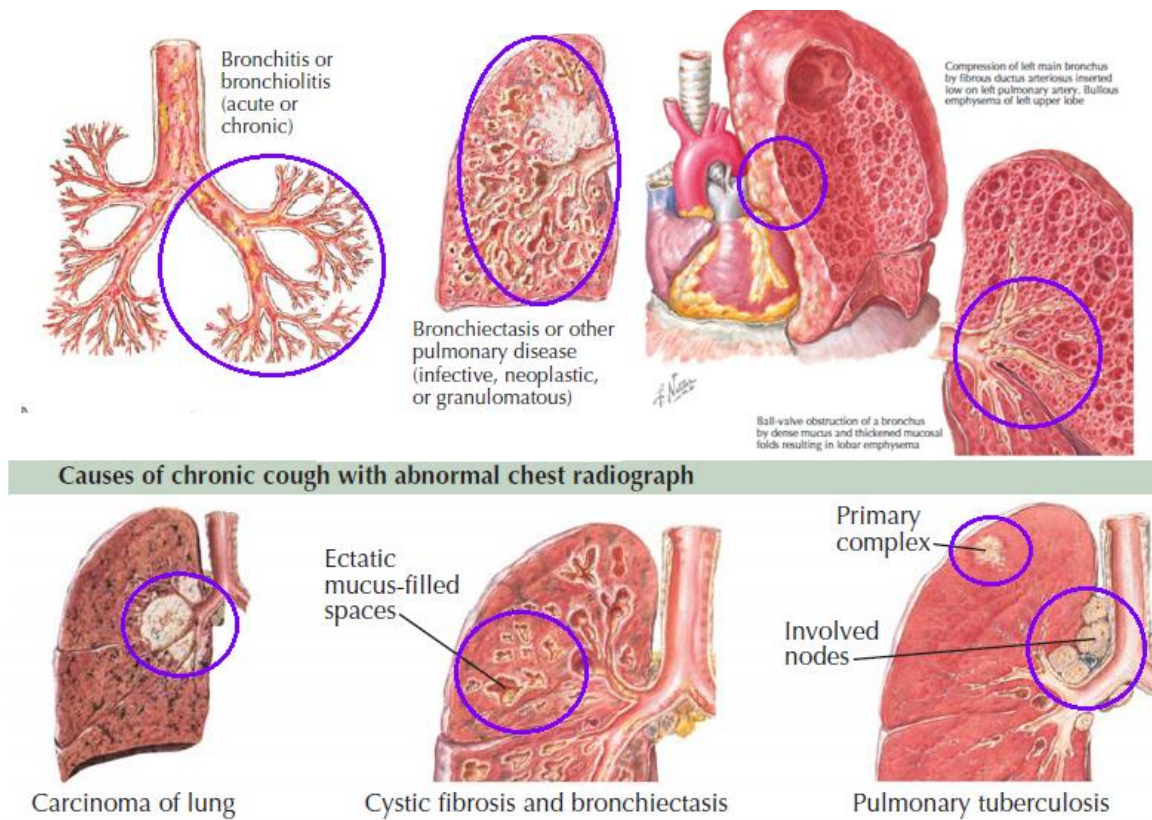


Figure 5: Inflammatory disease with bronchus obstruction.[50].

As their name suggests, inflammatory diseases are diseases where the airways of the respiratory system could be sporadically or constantly inflamed, usually due to congenital sensitivity to some environmental or chemical elements [7]. Asthma, cystic fibrosis, acute respiratory



distress syndrome, and chronic obstructive pulmonary disease (COPD), which encompasses emphysema and chronic bronchitis, are all included in this category Figure 5. Frequently, asthma symptoms are wheezing, chest tightness, coughing, and shortness of breath [5], [6]. COPD patients typically appear with a persistent cough with excessive mucus production and symptoms comparable to asthma [43], [54]. A dry cough, weariness, unexplained weight loss, and musculoskeletal discomfort are among the symptoms of pulmonary fibrosis. COPD is typically caused by lung damage caused by cigarette smoking [54], [55].

- Restrictive lung diseases

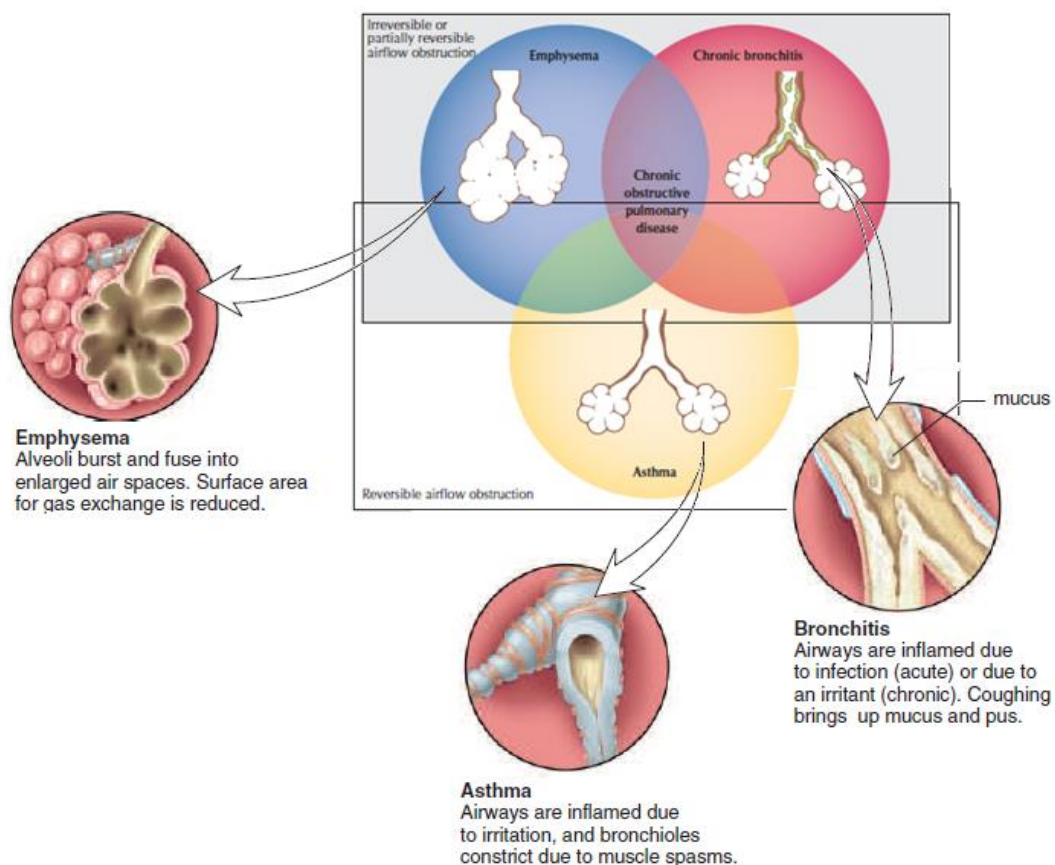


Figure 6: Restrictive lung disease diagram. Own elaboration. [50].

Restrictive lung diseases Figure 6 indicate that the airways are constricted, reducing the quantity of air a person can take in and making breathing more difficult. It can arise due to a clinical condition that causes the lungs to stiffen, the airways to fill with phlegm, and problems with the chest wall or breathing muscles [8], such as pulmonary fibrosis [56], spinal curvature, or obesity [57], among other causes [58].

- Respiratory tract infections

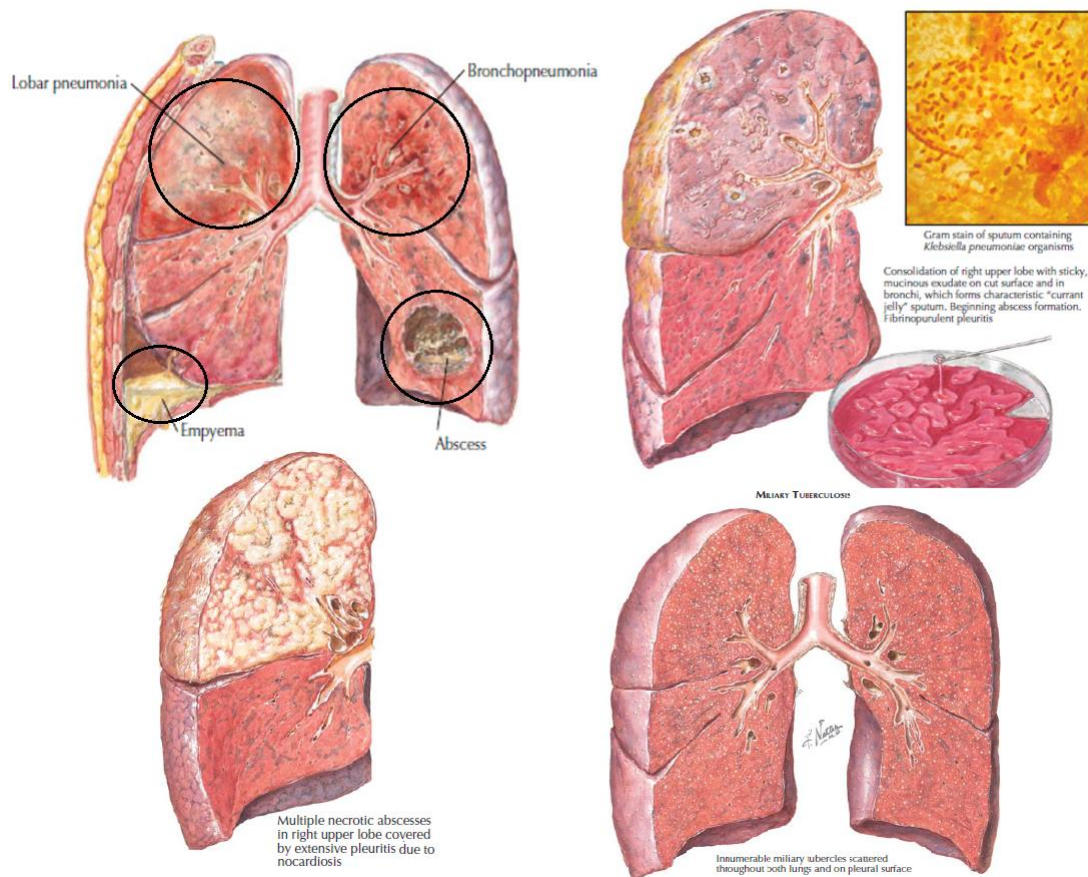


Figure 7: Different stages of pneumonia and pulmonary infections. Own Elaboration. [50]

Infections Figure 7 caused by bacteria and viruses can arise at any point in the respiratory system. The diseases are then classified as upper respiratory tract infection (URTI) and lower respiratory tract infection (LRTI) [4], [8]. The common cold is the most prevalent URTI. Others include tonsillitis, laryngitis, and pharyngitis. In comparison, bacterial pneumonia is the most prevalent kind of LRTI [9]. Viruses and fungi are two more causes of LRTI [4]. These infections can cause complications such as lung abscesses and spreading the infection to the pleural cavity. Pneumonia is an acute pulmonary infection that bacteria, viruses, or fungi can cause [59]. It infects the lungs, causing inflammation of the alveoli and pleural effusion, a condition in which the lung is filled with fluid. Pneumonia is more common in non-developed and developing countries, where overpopulation, pollution, and precarious environmental conditions exacerbate the situation, and medical resources are scanty [60], [61].

- Lung cancer

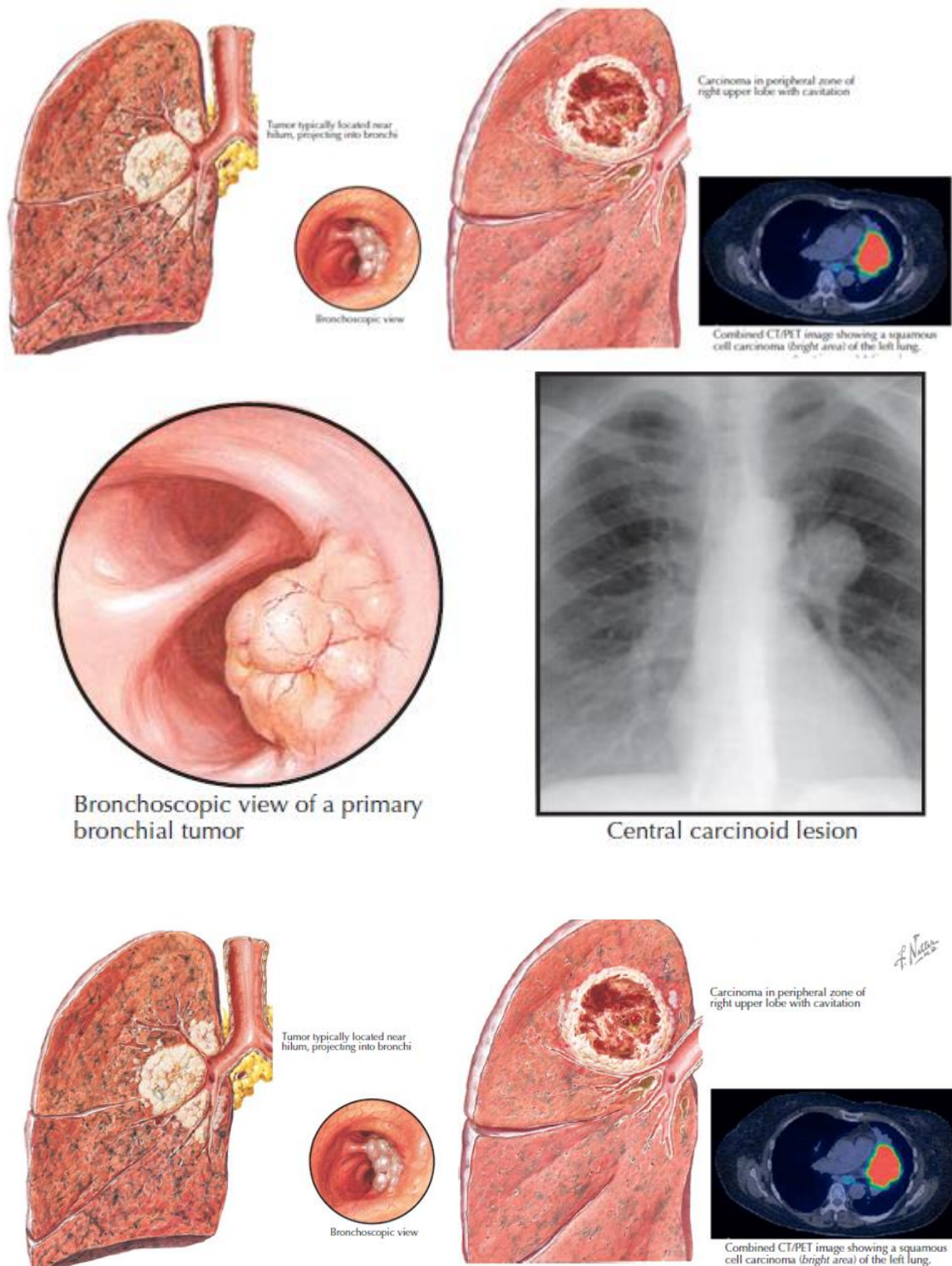


Figure 8: Different types of pulmonary neoplasms and tumors. Own elaboration. [2].

Lung cancer (LC) Figure 8 is a type of cancer where cells in the lungs divide uncontrollably. LC can cause tumors to appear and grow, reducing a person's ability to

breathe [62]. In advanced cases, tumors could spread to other vital body parts, producing a systemic and untreatable disease [63]. According to evidence, lung cancer is the third most frequent cancer and the leading cause of cancer-related mortality in the United States [62]. Furthermore, while smoking is the most prevalent cause of lung cancer, other risk factors might include a family history of the illness and exposure to radiation or certain chemicals [9].

- Pleural cavity diseases

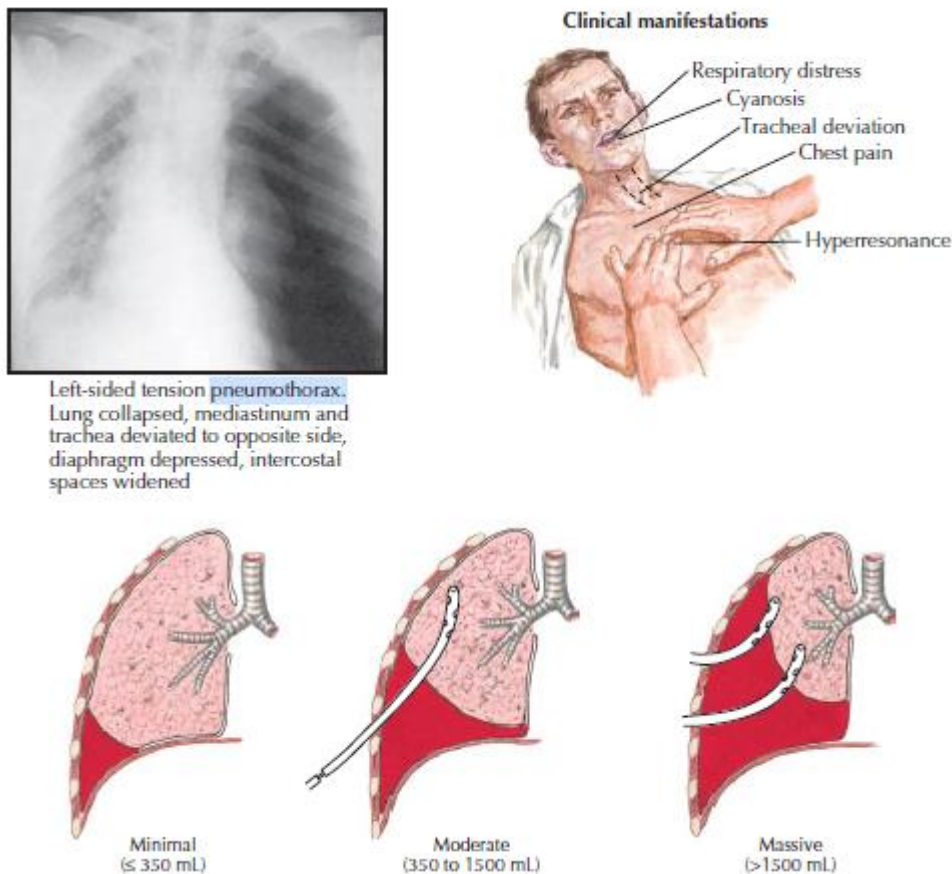


Figure 9: Pneumothorax symptoms and stages. Own elaboration [50].

As mentioned before, the pleural cavity is a space between the inner and outer pleural membranes surrounding the lungs' exterior. Pleural effusion is a condition that causes an accumulation of fluid in the pleural cavity [4]. It is always the outcome of another ailment, such as cancer, heart failure, or liver cirrhosis [8]. Another pleural cavity condition that happens when air enters the area between the chest wall and the lung,

known as the pleural space, is a collapsed lung, also known as pneumothorax Figure 9 [9]. A pneumothorax can compress the lungs highly; in extreme cases, the lungs might collapse like a balloon.

- Pulmonary vascular diseases

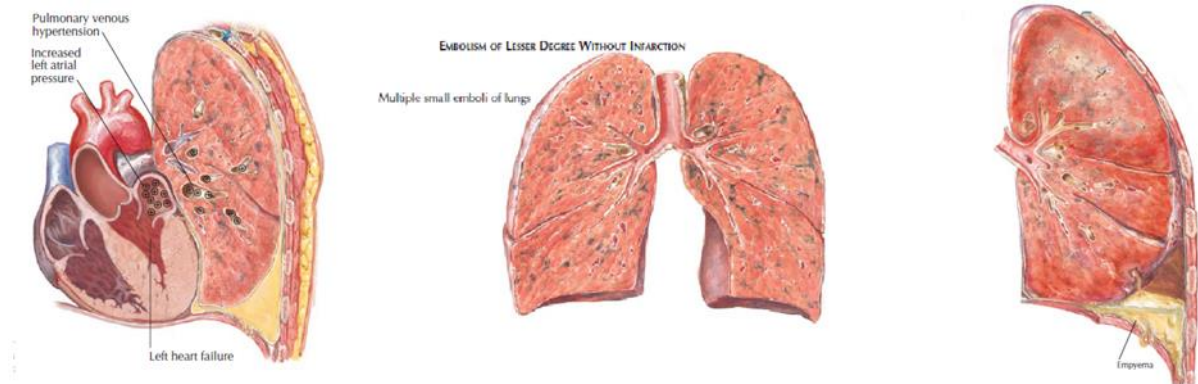


Figure 10: Common Pulmonary vascular diseases. From left to right: Pleural Effusion due to heart disease, Embolism, and Emphysema. Own elaboration. [50]

Pulmonary vascular diseases Figure 10 are clinical conditions that affect the vessels that carry blood through the lungs. This group includes pulmonary artery embolism, where a blood clot could travel from elsewhere in the body to the lungs, where it becomes lodged [4]. Pulmonary arterial hypertension is a sudden increase in the pressure of the pulmonary arteries, sometimes for unclear reasons [64], [65]. Pulmonary edema occurs when fluid leaks from the capillaries into the air spaces within the alveoli. It most often occurs due to congestive heart failure [66].

## 3.2. Artificial Intelligence and Neural Networks

### 3.2.1. AI

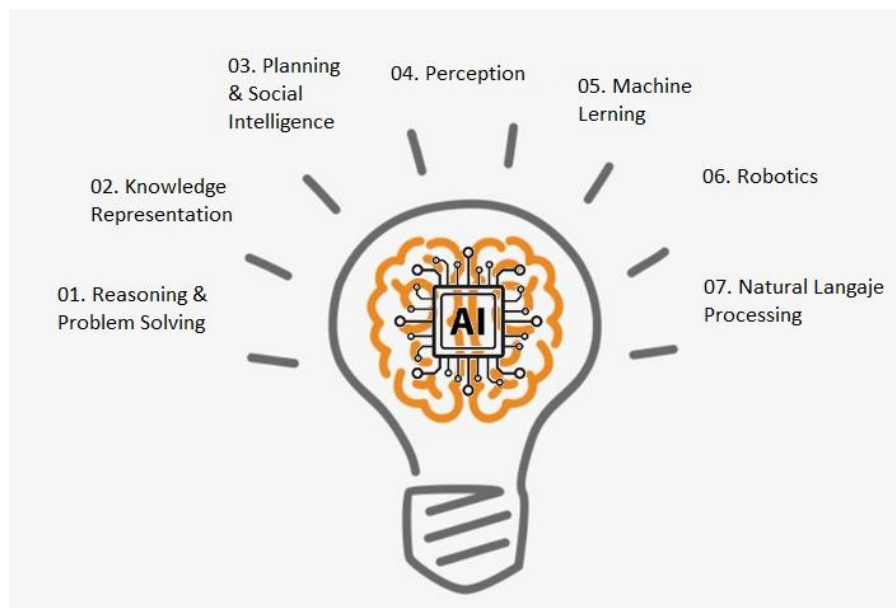


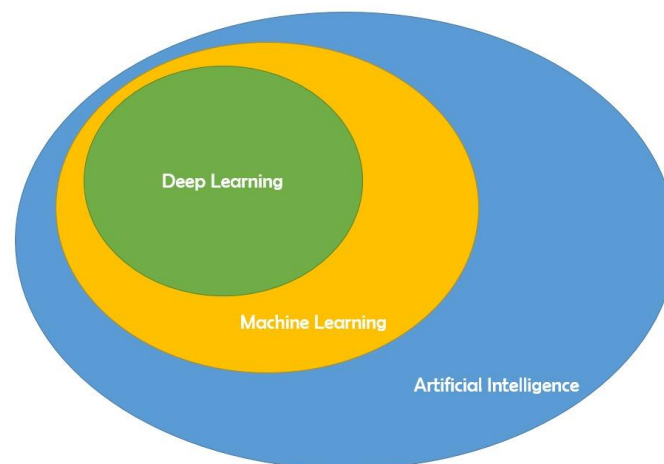
Figure 11: Subfields of Artificial Intelligence. Own elaboration.

Artificial Intelligence is a branch of computational sciences dedicated to developing different programs and machines that can solve problems and rationally make decisions Figure 11. The overall goal of AI is to develop algorithms and techniques to solve problems that humans might perform intuitively and almost automatically but pose a great challenge to computers, to then adapt those algorithms to solve problems that humans cannot because of their complexity or extensiveness [67]. An example of this class of AI problems is the ability to interpret and understand the content of an image: this assignment is something a human can do with relatively minimal effort. However, it has been proven to be challenging for machines. Therefore, when the task is about interpreting and understanding the content of millions of images, it is better to have an algorithm that helps performing the task efficiently [30], [47], [68].

Machine learning is one of the multiple branches of artificial intelligence that endows systems with the ability to self-learn and self-improve from experience without being explicitly programmed [67]. While AI combines extensive work related to automatic machine reasoning, machine learning tends to focus on recognizing patterns and learning from data. Machine learning allows developing algorithms and programs with the capacity to learn to extract

patterns and data relationships by themselves [68], [69]. Then, these patterns can be used to predict future behaviors and make evidence-based decisions.

Deep learning is a special sub-field of machine learning Figure 12 that employs artificial neural networks (ANNs) to perform multifactorial tasks using a structure inspired by the human brain processing method [67]. A deep learning architecture consists of multiple layers or stacks of simple nodules designed by engineers to learn from data using a structured learning procedure. As its name indicates, a neural network is an arrangement of simple processing units (neurons) that get specific inputs and, based on them, obtain the desired output [67]. ANNs are not intended to be realistic models of the human brain. ANNs are a mathematical approach that allow establishing parallelisms within a basic model of the brain and deduce how to mimic some of its behaviors using computational methods in which the inputs are equivalent to the dendrites of neurons. The outputs are like the axons of neurons [70].



*Figure 12: Venn diagram describing machine and deep learning as a part of AI [71].*

Deep learning mimics the process of human learning by an abstraction of the process. As mentioned before, ANNs are systematic arrangements of multiple layers of simple processing units connected to each other. Furthermore, a common neural network is comprised of four main components: inputs, weights, a bias or threshold, and an output Figure 13 [67].

Each layer will have a different “weight”. This weighting reflects what was known about the images and their components and how important is the information provided.

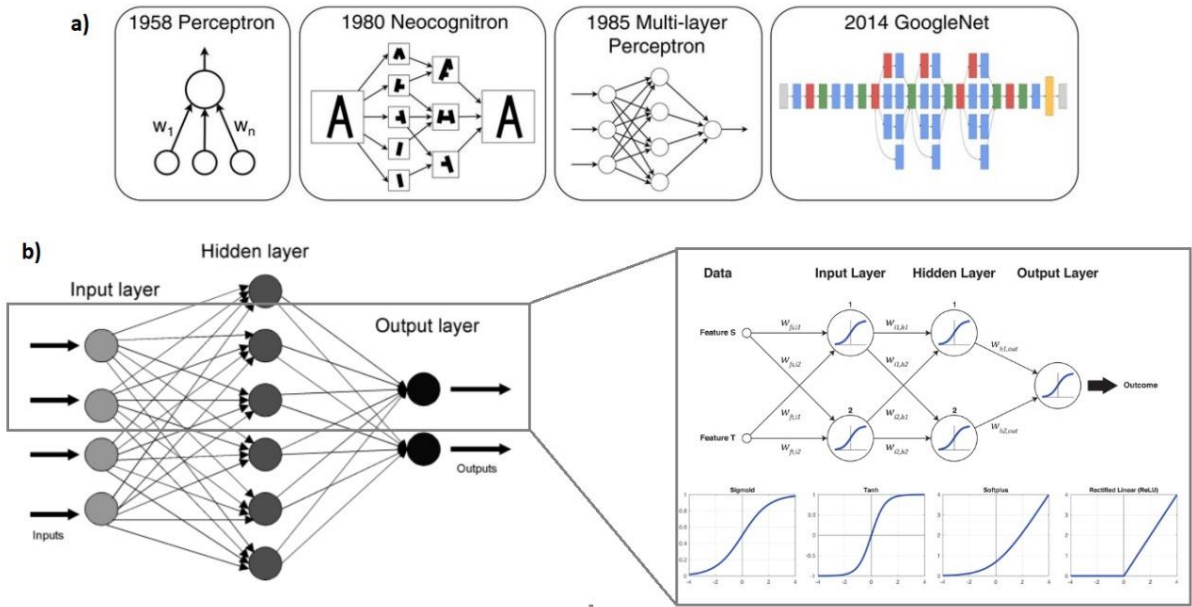


Figure 13: a) Timeline of some of the highest discoveries in deep learning history since perceptron development until GoogleNet presentation. b) Schematic view of artificial neural networks. Own elaboration [72], [73].



### 3.2.2. Supervised Learning

*Supervised learning* is a term used to describe a model or algorithm that makes predictions using class or labels distribution in terms of predictor features. According to the training data, a model is created through a training process in which the model makes predictions about the input data and then is corrected when the predictions are erroneous. The training process continues until the model reaches a certain desired shutdown criterion, like a maximum number of training iterations or a low error rate. The resultant classifier is used then to assign labels of its respective class to test images where the values of the predictor features are already known, but the values of the class label are not known Figure 14 [73].

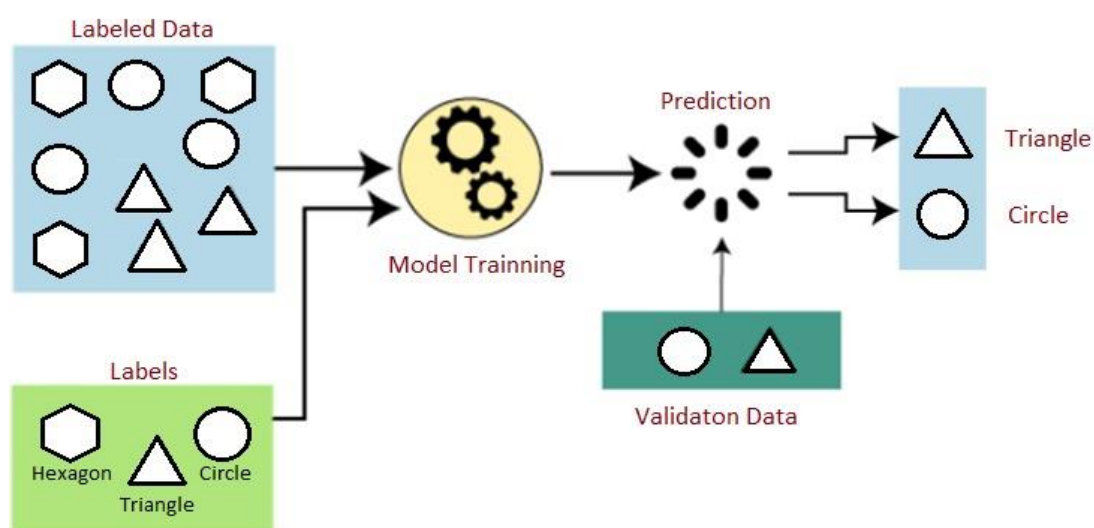


Figure 14: Supervised learning process

Supervised learning aims to develop an artificial intelligence system capable of learning the mapping between the input and output signals and predicting the outcome of the system given the new inputs. As its names show, this field of machine learning is supervised because, in the training process, the data labels indicate to the program which is the class to which this image belongs[67]. Based on that, later, the model will be able to determine whether future images belong to it or not. Supervised learning is the most known and studied machine learning algorithm to develop classifier and detector software for natural and medical images.

### 3.2.3. Convolutional Neural Networks

In deep learning, a convolutional neural network (CNN) is a particular type of artificial neural network (ANN), often designed to process structured arrays of data such as images and other visual resources. CNN's, also known as Space Invariant Artificial Neural Networks, is based on a shared-weight architecture. The kernels or convolution filters of the network glide along the input features and provide equal translational responses known as feature maps.

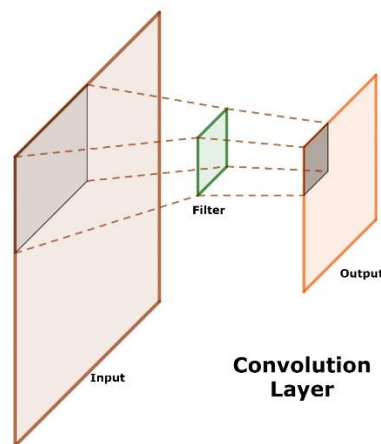
A convolution is a mathematical operation, more precisely, an integral that conveys the degree of superposition of one function over another, yielding a third function that expresses how the shape of one is altered by the other [74]. In deep learning terms, a convolution (of images) is a multiplication by elements of two matrices followed by an addition. The convolution operations consider the distribution of the information in a matrix space, making them suitable for the detection or classification even in complex images [74]. The convolutional kernel is a small matrix with an anchor point used to determine the position of the kernel concerning the image and numbers in each cell. The anchor point starts at the image's top-left corner to be analyzed and moves over each pixel sequentially. At each position, the kernel overlaps a few pixels on the image. Each overlapping pair of numbers is multiplied and added. Finally, the value at the current position is set to this sum [67]. The CNN obtains high-level features in the first filters, like edges, contours, and colors, among others, whereas the inner layers extract abstract characteristics.

Convolutions are one of the fundamental building blocks in computer vision and image processing algorithms. Convolutions have applications in image and video recognition, natural language processing, image classification and segmentation, medical image analysis, and brain-computer interfaces, among many others [75]. Concerning the overall architecture, there are many types of layers used to build a Convolutional Neural Network, but the most likely used are:

- Convolutional (CONV).
- Activation (activation function).
- Pooling (POOL).
- Fully connected (FC).

- Batch normalization (BN).
- Dropout (DO).
- Convolutional Layer

As mentioned before, convolution is a mathematical operation that extracts features from an image taking an image matrix and a filter or kernel. The convolutional layers of a CNN apply learned filters to the incoming images to construct feature maps that summarize the existence of those features in the input. Convolutional layer feature maps predict the class probabilities for each feature by applying this filter to the whole image, a few pixels at a time [76]. The output of the filters goes through a linear or nonlinear activation function to form the output feature maps [67]. See Figure 15 .



*Figure 15: Convolution layer.*

- Activation Layers

Activation layers are not technically recognized as “layers” (because no parameters or weights are learned within an activation layer). They are sometimes omitted in network architecture diagrams since it is assumed that an activation immediately follows a convolution. However, it is important to note that the activation function is being implemented within the network architecture. After each CONV layer in a CNN, applying a nonlinear activation function, such as Sigmoid, ReLU, ELU, or any other activation function variants, is necessary.

- Pooling Layer

Spatial Pooling layers aim to gradually decrease the number of data points in images that are too large [74]. A pooling layer is commonly added after a nonlinear activation function (i.e. Sigmoid) is applied to feature maps produced by the convolutional layer.

Spatial pooling or down sampling reduces the dimensionality of each map but retains essential information [74]. Additionally, as the pooling layer operates upon each feature map separately, spatial pooling creates a new dataset of the same number of the feature map pooled.

The pooling operation or filter is smaller than the feature map; in particular, it is nearly always 2x2 pixels applied with a stride of 2 pixels [77]. This means that a pooling operation will lower the size of each feature map by a factor of two, i.e., each dimension is halved, decreasing the number of pixels or values in each feature map to one-quarter of its original size. A pooling layer applied to an 8x8 (64 pixels) feature map produces an output feature map of 4x4 pixels (16 pixels).

There are two forms of spatial pooling: max pooling and average pooling [69]. Max Pooling returns the maximum value from the Kernel-covered region of the picture. On the other hand, average pooling returns the average of all values from the region of the picture covered by the kernel [74]. An example is shown in Figure 16.

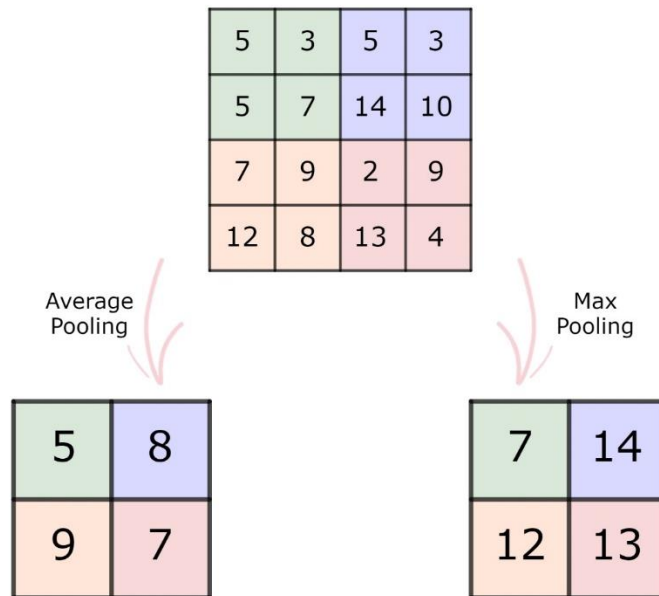
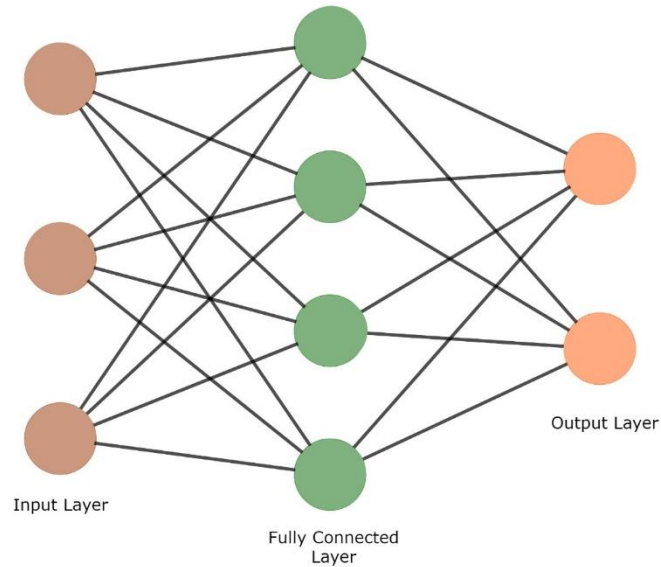


Figure 16: Average Pooling and Max Pooling example.

- Fully Connected Layers

As is customary for feed-forward neural networks, neurons in FC layers are completely linked to all activations in the preceding layer, see Figure 17. In the previous phases, a fully connected layer or classification layer computes the score of each class using the retrieved features from a convolutional layer [78]. Now, once the input picture is turned into a format suited for the Multi-Level Perceptron, it will flatten into a column vector. The flattened output is input into a feed-forward neural network, and backpropagation is used for each training iteration. The model can discriminate between dominant and certain low-level characteristics in pictures and categorize them using the Softmax Classification algorithm across a series of epochs [68]. FC layers are always applied at the network's end (i.e., the model does not apply a CONV layer, then an FC layer, then another CONV layer).



*Figure 17: Fully connected layer.*

- Batch Normalization

Batch normalization (or BN) layers are used to normalize the activations of a particular input dataset before passing it to the next layer of the network, as the name implies. Ioffe and Szgedy introduced BN in 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [79]. Using this equation, activations coming out of a batch normalization layer will have a mean of zero and a variance of one (i.e., centered at zero).

- Dropout

Dropout is a type of regularization that seeks to minimize overfitting by boosting testing accuracy, sometimes at the price of training accuracy. Dropout layers randomly disconnect inputs from the upstream layer to the downstream layer in the network architecture for each mini-batch in the training set with probability  $p$  [80]. Figure 18 shows an example of a neural network after dropout.

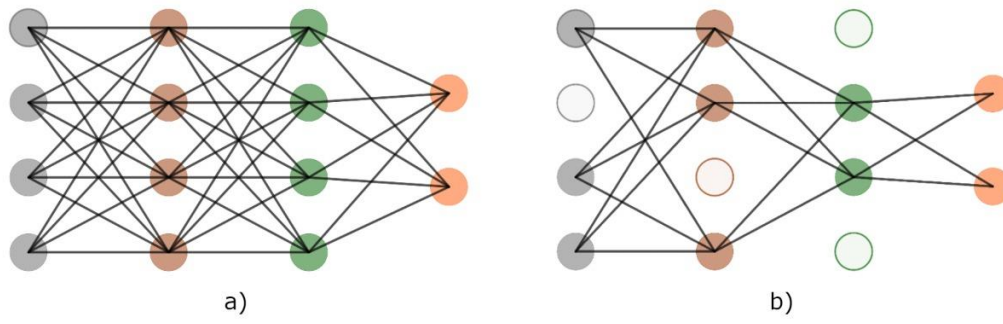


Figure 18: a) fully connected neural network b) the same network after dropout.

### 3.3. Image Classification

For computers, trying to identify the objects inside an image is much harder than for humans. An image is just interpreted by machines like a large arrangement of binary information. In this way, it is also important to understand that a computer will not understand anything about the thoughts, knowledge, or meaning the image is trying to transmit.

In order to help computers understand the contents of an image, it is necessary to apply an image classification task. Image classification consists of implementing computer vision and machine learning algorithms that allow the computer to extract the meaning from a set of images. This action could be simple as assigning a label to what a single object image contains, to as advanced as interpreting the content of an image with many objects and returning a human-readable sentence with the number of objects per class from a list of classes designed. Image classification is a vast and very promising field of study, covering a wide variety of techniques and a wide range of fields of application both inside and outside the medical field.

As mentioned before, “Image classification, at its very core, is the task of assigning a label to an image from a predefined set of categories.” In practice, the goal in image classification is to parse an input image and output a label that categorizes the image into one of the categories set. For example, assume that one set of possible categories includes: categories = {car, house, tree}. Then a car photograph is presented to the classification system. The aim is to take this entry image and assign it a label from the categories set. In this case, the correct label is a car. Besides that, the classification systems could also assign multiple labels to a single image indicating the probability of fitting for each category (i.e., car: 95%; house: 4%; tree: 1%).



### 3.3.1. Deep Learning and Biomedical Imaging

Deep learning has been present since the 1940s, going by several names and incarnations depending on different schools of thought and major research trends. Since then, a very long way of discoveries and improvements have been assembled by scientists around the world. Those discoveries/improvements help this field grow and develop to the point that nowadays, DL algorithms surpass the performance of many medical professionals in a broad spectrum of medical imaging tasks.

While difficult to access, it is still important to access a medically validated data set to run tests with CNNs. Emerging progress in deep learning and large data sets has permitted algorithms efficiently outperform a broad range of medical imaging challenges, from diabetic retinopathy detection [25], [26] and skin cancer classification [29], [30], to COVID-19 detection [35], [37], [39], and pulmonary nodules identification [22], [32]. Furthermore, the automated diagnosis of chest radiographs gained increasing attention with algorithms intended to identify and classify a broad spectrum of pulmonary diseases.

Using the publicly available OpenI dataset, researchers discovered that the same deep convolutional network architecture does not perform well across all abnormalities [81]. Ensemble models significantly improved classification accuracy compared to single models, and deep learning methods improved accuracy compared to rule-based methods.

Furthermore, it is essential to remember that medical and daily life images differ. The regions of interest or ROI, the specific points in the image aimed to be identified because they indicate relevant information to the classification process, are significantly smaller in medical images than those in lifelike pictures. As a result, recognizing them may be significantly more difficult for the neural network.

On the other hand, some medical photos quality is enormous compared to conventional images, and more pixels equals more information. This large disparity in the amount of information per image makes treating the program more difficult. Medical imaging data is scarce, costly, and loaded with legal patient privacy issues. As a result, the vast majority of such data cannot be used for broad public study.

However, despite all the limitations, important advances have been made in medical image classification. Some of the most relevant ones are summarized in Table 2. As [82] whom using the freely accessible OpenI dataset, investigated the performance of several convolutional architectures on various anomalies. [83] Whom published ChestX-ray-14, a dataset that is an order of magnitude larger than prior datasets of its sort. Also, they benchmarked several convolutional neural network architectures pre-trained on ImageNet. [84] used statistical dependencies between labels to improve prediction accuracy, outperforming [83] in 13 of 14 classes.

Then, between 2018 and 2019, several scientists developed proposals for the detection and classification of chronic obstructive pulmonary disease, tuberculosis[34], pneumonia, asthma, malignant pulmonary neoplasm, and lung cancer [33], [34], [85], [86]. The comparative work from [34] is of remarkable interest. They considered Convolutional neural networks (CNNs), backpropagation neural networks (BPNNs) with supervised learning, and competitive neural networks (CpNNs) with unsupervised learning for the detection of five of the diseases previously mentioned. At the end of their experimentations, scientists conclude that compared to other networks, CNNs have attained the greatest recognition rate for training and testing data. In contrast, the superiority of CNN over other networks requires more time and a greater number of learning iterations. Besides that, [86] developed a deep learning-based algorithm using single-center data collected where the algorithm demonstrated significantly higher performance than all three physician groups in both image-wise classifications (0.983 vs 0.814-0.932; all  $P < .005$ ) and lesion-wise localization (0.985 vs 0.781-0.907; all  $P < .001$ ) increasing the precedents of the potency and scope of deep learning in the detection of pulmonary diseases. In 2020 with the COVID-19 pandemic, a new epoch for the detection of pulmonary pathologies has emerged. With dozens of papers published and algorithms developed for the early detection and classification of COVID-19 patients in Japan, China, India, and Italy, among others. 2020 was the year data scientists from every point in the world focused on deep learning trying to stop the huge record of COVID deaths [35], [37], [38], [62], [87], [88] In this part, it is remarkable the job performed by [35]. They developed eleven convolutional neuronal networks based on different CNN models for detecting infected coronavirus patients. From their several experimentations, scientists discover that ResNet201 plus J48 CNN method results in better

classification for detection of COVID-19 with Accuracy, Recall, Specificity, Precision, and F1-Score of 98.50%, 100%, 97.20%, 100%, and 98.40%, respectively.

Table 2: Bibliographic summary on neural networks.

Name of the study	Authors	Year	Pathology of interest	Model	Main Methods	Performance	Ref.
CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning.	Rajpurkar, Pranav; Irvin, Jeremy; Zhu, Kaylie; Yang, Brandon; Mehta, Hershel; Duan, Tony; Ding, Daisy; Bagul, Aarti; Langlotz, Curtis; Shpanskaya, Katie; Lungren, Matthew P.; Ng, Andrew Y.	2017	Pneumonia.	CheXNet, is a 121-layer CNN trained on ChestX-ray14, currently the largest publicly available chest X-ray dataset, containing over 100,000 frontal view X-ray images with 14 diseases.	Four practicing academic radiologists annotate a test set, on which the performance of CheXNet is compared to that of radiologists.	They find that CheXNet exceeds average radiologist performance on the F1 metric. CheXNet was extended to detect all 14 diseases in ChestX-ray14 and achieve state of the art results on all 14 diseases.	[89]
Deep Convolutional Neural Networks for Chest Diseases Detection.	Abiyev, Rahib H.; Sallam Ma'aitah, Mohammad Khaleel.	2018	Chronic obstructive pulmonary disease, pneumonia, asthma, tuberculosis, and lung diseases.	Convolutional neural networks (CNNs, backpropagation neural networks (BPNNs) with supervised learning, and competitive neural networks (CpNNs) with unsupervised learning.	All the considered networks CNN, BPNN, and CpNN were trained and tested on the same chest X-ray data set, and the performance of each network is discussed.	CNN has achieved the highest recognition rate for training and testing data, compared to other employed networks. In contrast, this outperformance of CNN over other networks requires longer time and a larger number of learning iterations than that of BPNN2 and CpNN2.	[34]

<p>Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists.</p>	<p>Rajpurkar, Pranav; Irvin, Jeremy; Ball, Robyn L.; Zhu, Kaylie; Yang, Brandon.</p>	<p>2018</p>	<p>Tuberculosis and lung cancer.</p>	<p>CNN to concurrently detect the presence of 14 different pathologies.</p>	<p>CheXNeXt was trained and internally validated on the ChestX-ray8 dataset, with a held-out validation set consisting of 420 images, sampled to contain at least 50 cases of each of the original pathology labels. On this validation set, the majority vote of a panel of 3 board-certified cardiothoracic specialist radiologists served as reference standard.</p>	<p>CheXNeXt achieved radiologist-level performance on 11 pathologies and did not achieve radiologist-level performance on 3 pathologies. The radiologists achieved statistically significantly higher AUC performance on cardiomegaly, emphysema, and hiatal hernia. CheXNeXt performed better than radiologists in detecting atelectasis, statistically significantly higher than radiologists; there were no statistically significant differences in the detection of other 10 pathologies.</p>	<p>[33]</p>
<p>Development and Validation of a Deep Learning-Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs.</p>	<p>Hwang, Eui Jin; Park, Sunggyun; Jin, Kwang-Nam; Kim, Jung Im; Choi, So Young; Lee, Jong Hyuk; Goo, Jin Mo; Aum, Jaehong; Yim, Jae-Joon.</p>	<p>2019</p>	<p>Pulmonary malignant neoplasm, active tuberculosis, pneumonia, and pneumothorax.</p>	<p>Deep learning-based algorithm using single-center data collected.</p>	<p>The algorithm was externally validated with multicenter data collected between May 1 and July 31, 2018. A total of 486 chest radiographs with normal results and 529 with abnormal results from 5 institutions were used for external validation. Fifteen physicians, including non-radiology physicians, board-certified radiologists, and thoracic radiologists, participated in observer performance testing.</p>	<p>The algorithm demonstrated a median (range) area under the curve of 0.979 (0.973-1.000) for image-wise classification and 0.972 (0.923-0.985) for lesion-wise localization; the algorithm demonstrated significantly higher performance than all 3 physician groups in both image-wise classification (0.983 vs. 0.814-0.932; all P &lt; .005) and lesion-wise localization (0.985 vs. 0.781-0.907; all P &lt; .001).</p>	<p>[86]</p>

Development and Validation of Deep Learning-based Automatic Detection Algorithm for Malignant Pulmonary Nodules on Chest Radiographs.	Nam, Ju Gang; Park, Sunggyun; Hwang, Eui Jin; Lee, Jong Hyuk; Jin, Kwang-Nam; Lim, Kun Young; Vu, Thienkai Huy; Sohn, Jae Ho; Hwang, Sangheum; Goo, Jin Mo; Park, Chang Min.	2019	Malignant pulmonary nodules.	Deep learning-based automatic detection algorithm (DLAD).	DLAD was developed by using 43 292 chest radiographs in 34 676 patients (healthy-to-nodule ratio, 30 784:3892; 19230 men [mean age, 52.8 years; age range, 18–99 years]; 15 446 women [mean age, 52.3 years; age range, 18–98 years]) obtained between 2010 and 2015.	Radiograph classification and nodule detection performances of DLAD were a range of 0.92–0.99 (AUROC) and 0.831–0.924 (JAFROC FOM), respectively.	[90]
An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare.	Stephen, Okeke; Sain, Mangal; Maduh, Uchenna Joseph; Jeong, Do-Un.	2019	Pneumonia.	CNN model from scratch to features extraction and image classification.	CNN model trained from scratch to classify and detect the presence of pneumonia from a collection of chest X-ray image samples. Several data augmentation algorithms were used to improve the validation and classification accuracy of the CNN.	Average training and validation accuracies of 0.94814 and 0.93012 respectively were obtained along with different changes in the experiment set up related to the image sizes.	[91]

Performance of a Deep Learning Algorithm Compared with Radiologic Interpretation for Lung Cancer Detection on Chest Radiographs in a Health Screening Population.	Lee, Jong Hyuk; Sun, Hye Young; Park, Sunggyun; Kim, Hyungjin; Hwang, Eui Jin; Goo, Jin Mo; Park, Chang Min.	2020	Lung cancer.	A commercially available deep learning algorithm (Lunit Insight for Chest Radiography, version 4.7.2; Lunit, Seoul, Korea).	The algorithm provides both an image-wise probability value of a chest radiograph being abnormal and a per-pixel localization map overlaid on the input chest radiograph identifying the location of abnormalities.	In a validation test comprising 10 285 radiographs from 10202 individuals with 10 radiographs of visible lung cancers, the algorithm's AUC was 0.99 (95% confidence interval: 0.97, 1), and it showed comparable sensitivity (90% [nine of 10 radiographs]) to that of the radiologists (60% [six of 10 radiographs]; P = .25) with a higher FPR (3.1% [319 of 10 275 radiographs] vs. 0.3% [26 of 10275 radiographs]; P , .001).	[92]
Explainable Deep Learning for Pulmonary Disease and Coronavirus COVID-19 Detection from X-rays.	Brunese, Luca; Mercaldo, Francesco; Reginelli, Alfonso; Santone, Antonella.	2020	COVID-19.	CNN for image classification.	Fine-tuning transferred learning of a VGG-16 network trained on the ImageNet data set trained for weeks using NVIDIA Titan Black GPU's and 6,523 chest X-rays: 250 related to patients afflicted by COVID-19, 2,753 related to patients with other pulmonary diseases and 3,520 related to healthy patients. All the diagnosis were confirmed by expert radiologists.	The model for healthy and generic pulmonary diseases discrimination obtains a sensitivity equal to 0.96 and a specificity of 0.98. The model for generic pulmonary diseases and COVID-19 discrimination exhibits a sensitivity of 0.87 and a specificity equal to 0.94. Finally, in regard to the accuracy, the first model reaches an accuracy of 0.96, while the second model obtains a value of 0.98 with an average time for COVID-19 detection of approximately 2.5 seconds.	[39]

Multi-View Ensemble Convolutional Neural Network to Improve Classification of Pneumonia in Low Contrast Chest X-Ray Images.	Ferreira, Jose Raniery; Cardona Cardenas, Diego Armando; Moreno, Ramon Alfredo; De Sa Rebelo, Marina de Fatima; Krieger, Jose Eduardo; Gutierrez, Marco Antonio.	2020	Pneumonia.	VGG-16 CNN and replaced its fully-connected layers with a customized multilayer perceptron.	Four different training strategies were proposed and evaluated: original chest X-ray image (baseline), chest-cavity-cropped image (A), and histogram-equalized segmented image (B). The last strategy method (C) implemented is based on ensemble between strategies A and B.	The ensemble model C yielded the highest performances: AUC of 0.97 (CI: 0.96–0.99) to classify pneumonia vs. normal, and AUC of 0.91 (CI: 0.88–0.94) to classify bacterial vs. viral cases.	[93]
Modality-specific deep learning model ensembles toward improving TB detection in chest radiographs.	Rajaraman, Sivaramakrishnan; Antani, Sameer K.	2020	Tuberculosis.	CNN and selected popular pre-trained CNNs.	CNNs are trained to learn modality-specific features from large-scale publicly available chest X-ray collections including (i) RSNA dataset, (ii) Pediatric pneumonia dataset and (iii) Indiana dataset. The models are evaluated through cross-validation (n = 5) at the patient-level with an aim to prevent over fitting, improve robustness and generalization.	It is observed that a stacked ensemble of the top-3 retrained models demonstrates promising performance (accuracy: 0.941; 95% confidence interval (CI): [0.899, 0.985], area under the curve (AUC): 0.995; 95% CI: [0.945, 1.00]). The ensemble model resulted in reduced prediction variance and sensitivity to training data fluctuations. Results from their combined use are superior to the state-of-the-art.	[94]



Automated segmentation and diagnosis of pneumothorax on chest X-rays with fully convolutional multi-scale ScSE-DenseNet: a retrospective study.	Wang, Hongyu; Gu, Hong; Qin, Pan; Wang, Jia.	2020	Pneumothorax.	Fully convolutional DenseNet (FC-DenseNet).	Fully convolutional DenseNet (FC-DenseNet) with multi-scale module and spatial and channel squeezes and excitation modules in the detection and segmentation of pneumothoraces.	Mean pixel-wise accuracy of 93% and Dice score of 0.92, with diagnostic accuracy 93.45% and F1-score 92.97%.	[95]
Validation of a Deep Learning Algorithm for the Detection of Malignant Pulmonary Nodules in Chest Radiographs.	Yoo, Hyunsuk; Kim, Ki Hwan; Singh, Ramandeep; Digumarthy, Subba R.; Kalra, Mannudeep K.	2020	Malignant pulmonary nodules - lung cancer.	Deep learning-based AI algorithm using separate training (in-house) and validation (NLST) data sets.	A deep CNN that uses residual information of 34 layers (ResNet-34) was selected. The raw pixel map of the Digital Imaging and Communications in Medicine file was normalized with windowing information, and the normalized pixel map was used as input for the AI model.	The area under the ROC curve (AUROC) of the AI algorithm was 0.93 (95% CI, 0.90-0.96) for all chest radiographs, 0.99 (95% CI, 0.97-1.00) for digital radiographs, and 0.86 (95% CI, 0.79-0.93) for computed radiographs.	[32]

Deep learning-based automated detection algorithm for active pulmonary tuberculosis on chest radiographs: diagnostic performance in systematic screening of asymptomatic individuals.	Lee, Jong Hyuk; Park, Sunggyun; Hwang, Eui Jin; Goo, Jin Mo, Lee, Woo Young; Lee, Sangho; Kim, Hyungjin; Andrews, Jason; Park, Chang Min.	2021	Pulmonary tuberculosis.	DLAD algorithms in systematic screening for active pulmonary tuberculosis.	Out-of-sample testing of a pre-trained DLAD algorithm, using CRs from 19,686 asymptomatic individuals (ages, $21.3 \pm 1.9$ years) as part of systematic screening for tuberculosis between January 2013 and July 2018.	With high specificity thresholds, DLAD showed comparable diagnostic measures with the pooled radiologists (p values > 0.05). For the radiologically identifiable relevant abnormality (n = 28), DLAD showed an AUC value of 0.967 (95% confidence interval, 0.938-0.996) with sensitivities of 0.821 and 0.679, specificities of 0.960 and 0.997, PPVs of 0.028 and 0.257, and NPVs of both 0.999 at high sensitivity and high specificity thresholds, respectively.	[24]
Prediction of Obstructive Lung Disease from Chest Radiographs via Deep Learning Trained on Pulmonary Function Data.	Schroeder, Joyce D.; Lanfredi, Ricardo Bigolin; Li, Tao; Chan, Jessica; Vachet, Clement; Paine, Robert; Srikumar, Vivek; Tasdizen, Tolga.	2021	Obstructive lung disease.	CNN trained with near concurrent pulmonary function test (PFT) data.	The Image Model (Resnet18 pre-trained with ImageNet CNN) is trained using frontal and lateral radiographs and PFTs with 10% of the subjects for validation and 19% for testing. The NLP Model is trained using radiologist text reports and PFTs.	The Image Model achieves an AUC of 0.814 for prediction of obstructive lung disease (FEV1/FVC <0.7) from chest radiographs. The Image Model predict severe or very severe COPD (FEV1 <0.5) with an AUC of 0.837.	[96]

A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and COVID-19 pneumonia from chest X-ray images.	Wang, Guangyu; Liu, Xiaohong; Shen, Jun; Wang, Chengdi; Li, Zihuan.	2021	COVID-19.	Fully automated deep-learning pipeline for the standardization of chest X-ray images, for the visualization of lesions and for disease diagnosis can identify viral pneumonia caused by coronavirus disease 2019 and other types of pneumonia.	Deep neural network based on the DenseNet-121 architecture and 120,702 chest X-ray images from 92,327 patients with labels of 14 common thoracic pathologies, was used for model training.	AUC of 0.914 for differentiating pneumonia from all other groups and an AUC of 0.935 for the overall classification of lung opacity. Furthermore, the AI system achieved comparable performance to the senior radiologists' level with an AUC of 0.981 (95% CI: 0.970-0.990) for the viral pneumonia diagnosis.	[88]
Deep Learning Algorithm for COVID-19 Classification Using Chest X-Ray Images.	V. J., Sharmila; D., Jemi Florinabel.	2021	COVID-19.	Deep convolutional generative adversarial networks (DCGANs) that classify chest X-ray images into normal, pneumonia, and COVID-19.	The proposed CNN method was trained with four distinct publicly accessible datasets of chest X-ray images (COVID-19 X-ray, COVID Chest X-ray, COVID-19 Radiography, and CoronaHack-chest X-ray) datasets based on the DCGAN synthetic images.	The proposed CNN model was more efficient in detecting COVID-19 from four different datasets than the pre-trained models. However, although the proposed DCGAN-CNN provided a significant advantage over COVID-19 detection, it contained some shortcomings with respect to multilabel classification, time consumption, and efficiency.	[97]
DON: Deep Learning and Optimization-Based Framework for Detection of Novel Coronavirus	Dhiman, Gaurav; Vinoth Kumar, V; Kaur, Amandeep; Sharma, Ashutosh.	2021	COVID-19.	Eleven different CNN-based models (AlexNet, VGG-16, VGG-19, GoogleNet, ResNet18, ResNet500, ResNet101, InceptionV3, InceptionResNetV2,	Multiobjective optimization and a deep-learning methodology according to the J48 algorithm for the detection of infected coronavirus patients with X-rays.	ResNet101 plus J48 CNN method result better classification for detection of COVID-19 with Accuracy, Recall, Specificity, Precision, and F1-Score are 98.50%, 100%, 97.20%, 100%, and 98.40%, respectively.	[35]

Disease Using X-ray Images.				DenseNet201 and XceptionNet).			
UBNet: Deep learning-based approach for automatic X-ray image detection of pneumonia and COVID-19 patients.	Widodo, Chomsin S; Naba, Agus; Mahasin, Muhammad M; Yueniwati, Yuyun; Putranto, Terawan A; Patra, Pangeran I.	2022	COVID-19.	Three CNN architectural hierarchies to classify between normal images and pneumonia images, bacterial and viral pneumonia images, and pneumonia virus images and COVID-19 virus infected images.	An open-source data set with 9,250 chest X-ray images including 3,592 COVID-19 images were used in this study to train and test the developed deep learning models.	CNN architecture with a hierarchical scheme developed in UBNet v3 using a simple architecture yielded following performance indices to detect chest X-ray images of COVID-19 patients namely, 99.6%accuracy, 99.7%precision, 99.7%sensitivity, 99.1%specificity, and F1 score of 99.74%.	[98]

## 4. Methodology

Based on the information presented in the previous chapters, this chapter will describe the methodology used to develop this study. In addition, important methodological notions related to data acquisition and preprocessing will be addressed, as well as the implementation of different convolutional models and the metrics used to measure the performance of the models. In the same way, a section in this chapter is dedicated to analyze the possible sources of error or uncertainty in the models and how to avoid them.

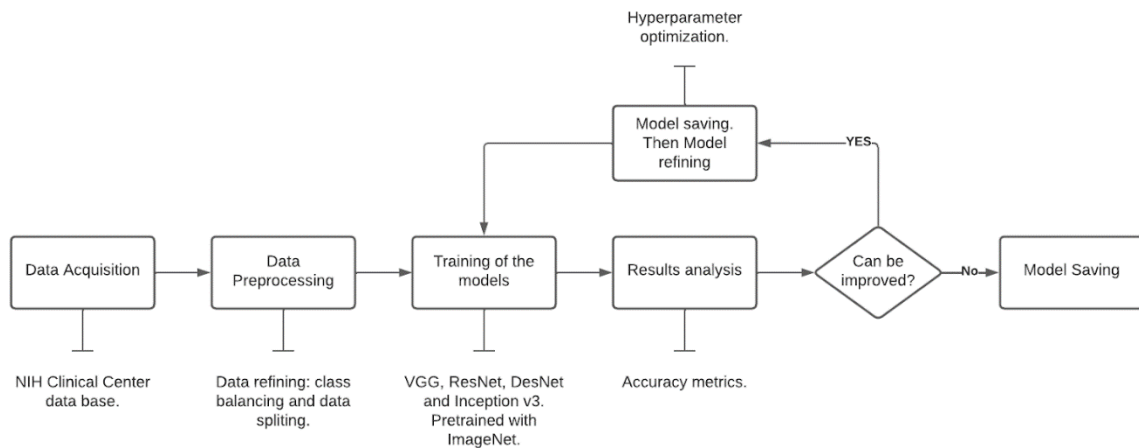


Figure 19: Flow diagram of the methodology

The methodology of this study is summarized in Figure 19. The images and dataset preprocessing, as well as the development of the Convolutional models proposed, were performed in Google Colaboratory (colab); it is a hosted version of Jupyter Notebooks with access to Google hardware, where Notebooks are run in Linux-based virtual machines (VMs) provided and maintained by Google and the computation can be performed with central processing units (CPUs) or accelerated through specialized graphical processing units (GPUs) and tensor processing units (TPUs) that allows to run and share modern AI and ML techniques [99]

## 4.1. Data acquisition

The data used for this research are Chest X-ray images from ChestX-Ray14, a Hospital-scale Chest X-ray data set with up to 112,120 frontal-view X-ray images of 30,805 unique patients (collected between 1992 and 2015) with the text-mined fourteen common disease labels, mined from the text radiological reports via NLP techniques.

The ChestX-Ray14 benchmark was recently introduced in [83]. The publicly available dataset was obtained from the NIH Clinical Center Library, see Figure 20, a repository of freely-available medical research data managed by the NIH Clinical Center. The dataset contains 112,120 high-resolution frontal view chest X-ray images from 30,805 patients, and each image is labeled with one or multiple common thorax diseases or “Normal” otherwise. Images are available in PNG format, numbered from “00000001\_000.png” to “00030805\_000.png”. The data entry protocol and data collection can be found in the original study.

As mentioned before, the original data set generated in this study [83] is a large dataset of radiological evidence from Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia, Pneumothorax, Consolidation, Edema, Emphysema, Fibrosis, Pleural Thickening, and Hernia. Unfortunately, the number of diseases is not always limited to one per image, and these images would make the training of the models harder. Thus the .csv file with the labels’ information associated with each image must be filtered and reorganized.



Figure 20: Images obtained from the NIHCC data set. a. No finding, b. Nodule, c. Emphysema, d. Infiltration

Chest X-ray images are the most common diagnostic analysis available and easy to obtain for the detection and treatment of most pulmonary diseases. They have been proven to be crucial in providing different features for detecting and diagnosing several pulmonary and thoracic diseases and, in this case, for training the networks proposed. Chest X-ray images are available

in every experiment in the selected data set, so these images have been chosen for their convolutional feature extraction and multi-label classification study. However, the results obtained by different scientists in studies related to this dataset are not as accurate as needed for developing an accepted medical-grade computer-aided diagnosis software. Therefore, the data should be treated based on the number of diseases the models are trying to learn; this in order to understand if the number of diseases categorized has any relevance in the learning capacity each model can develop.

## 4.2. Method Threats

Before training the neural networks, the data set must be filtered first to avoid images with misleading information, thus allowing the networks to learn to recognize the proposed diseases correctly. The data should be analyzed to identify the number of images per class/disease in the data set. It is essential to avoid frequent problems such as the well-known class imbalance problem, which consists of the imbalance of information generated when one of the classes present in the dataset has a much higher number of images than the others.

If the images present class imbalance, the number of images will be manually balanced, and discriminating factors will be determined to select the diseases that are more apt for the training of the networks. i.e., the diseases with the highest number of images, the diseases with the highest medical relevance, and diseases that do not generate noise among themselves. It is also crucial to avoid the information leakage that occurs when files or data used in the training of the networks are also used in the testing or validation processes.

## 4.3. Data Preprocessing

ChestX-Ray14 benchmark was recently introduced in [56]. The dataset contains 112,120 high-resolution frontal view chest X-ray images from 30,805 patients. Each image is labeled with one or multiple common pulmonary and Thorax diseases or “Normal” otherwise. The first step is to refine data to obtain a new dataset only with images with a single label instead of two or more. For reading the images from the main .csv file, the python Pandas and NumPy libraries have been used to read, annotate and perform other data refinement and splitting techniques. For preprocessing and data refinement, some practical methods, such as the .loc and .asnumpy methods, will help refine the data for training the implemented models.

First, the initial .csv file containing the data of all the images hosted in the dataset must be filtered by eliminating the data of the images with more than one label. Then the data of the remaining images will be read and stored in a data frame of the Pandas library, where the data will be registered by counting the number of images per class. Then the images will be separated into individual folders for each disease.

In case of class imbalance problems, the data for model training will be decided based on three discriminating factors:

- the diseases with the highest number of images,
- the diseases with the highest medical relevance, and
- diseases that do not generate noise among themselves

Once the images have been divided, and the presence or absence of class imbalance has been determined and overcome based on the discriminant factors presented, the number of images per class will be established.

Then the previously determined number of images per class will be randomly taken, and the information concerning these and their respective classes will be loaded into a new .csv file. Subsequently, this new .csv file containing all the information concerning the images for training and their classes will be divided into three sets: training, validation, and testing, with 80%, 10%, and 10% of the images, respectively. Since the data is separated after each image has been individually identified so that they do not have the same name or information, there is no risk of data leakage. Then the files are ready to feed the models prepared for the study.

#### 4.4. Models Development

For the development of the present study, different convolutional models were developed under the methodology known as transferred learning. In particular, four types of networks, the VGG, Resnet, Densenet, and Inception models were chosen for classifying the medical images on the first mapping of results. These networks were chosen after the bibliographic review. It was determined that they were the most appropriate for classifying and recognizing the diseases in the data set.



The mentioned networks were chosen in their stock version to map the results obtained under the fine tuning and feature extraction modalities. After obtaining the initial results generated by the proposed networks, the best performing networks were selected in order to study their performance after hyperparameter optimization.

#### 4.4.1. Models Description

- VGG nets

VGG stands for Visual Geometry Group; first introduced in [100], it is a standard deep Convolutional Neural Network (CNN) architecture with multiple layers. The “deep” refers to the number of layers with VGG-16 or VGG-19 consisting of 16 and 19 convolutional layers. The VGG architecture Figure 21 is the basis of ground-breaking object recognition models. It is now still one of the most popular image recognition architectures. For example in [101] they use a VGG-19 network for melanoma thickness prediction.

As for the architecture, VGG-16 has 16 layers (VGG-19 has 19 layers) consisting of 13 convolution and 3 fully connected layers as shown in Figure 21; the first and second layers have 64 kernels; the third and fourth layers are of 128 kernels. The fifth, sixth, and seventh layers use 256 kernel filters. Eighth to thirteenth have 512 kernel filters. All these layers are followed by a max pooling layer each, and finally, layers fourteen and fifteen layers are fully connected hidden layers of 4096 units, followed by a softmax output layer in layer sixteenth [102].

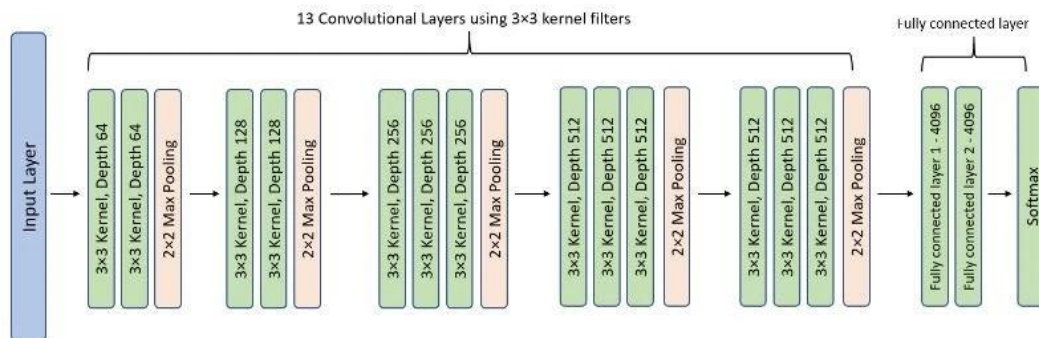
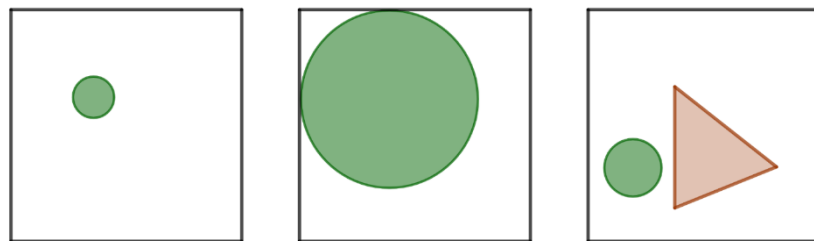


Figure 21: VGG architecture [102].

- Inception

The Inception network was an essential milestone in the development of CNN classifiers. It was first introduced in [76] and was designed to solve a problem of a variation in the location of information in an image, see Figure 22. Before its introduction, it was usual to stack convolutional layers on CNNs to obtain a better performance over time, which did not always work. The idea behind Inception networks was to have multiple-sized filters on the same levels instead of making multiple levels, going wider instead of deeper.



*Figure 22: Variation on the location and size of the image*

The Inception network was complex, and its constant evolution led to several versions of the network. The most popular versions are Inception v1, which will be briefly explained; Inception v2 and Inception v3, the last being the one implemented in this study, were both introduced in [103]; Inception v4 and Inception-ResNet, introduced in [104].

The Inception v1 architecture consists of nine inception modules stacked linearly. It is known as the GoogLeNet. In Figure 23, can be seen the modules for the Inception v1. The naive version (a) performs convolution with three different sizes of filters (1x1, 3x3, 5x5) and max pooling. The outputs are concatenated and sent to the next module. Moreover, to make the computational cost cheaper and reduce the number of input channels, an extra 1x1 convolution before the 3x3 and 5x5 convolutions is added, and a 1x1 convolution after the max pooling layer [76] as shown in (b).

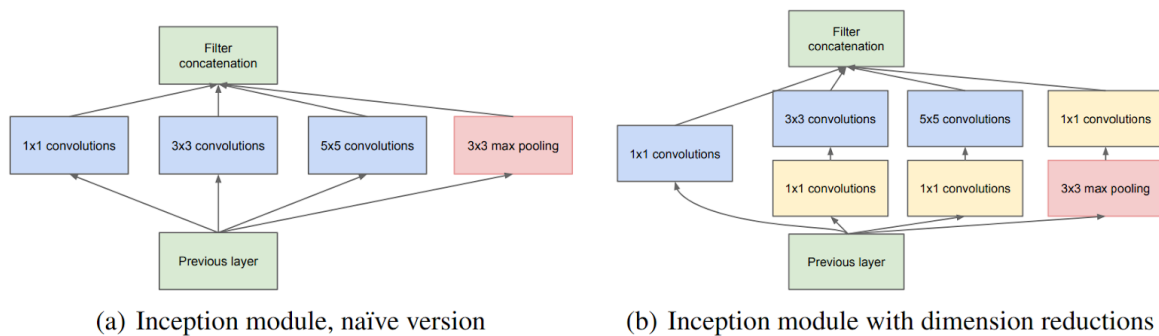


Figure 23: Inception modules [76]

Inception v2 and Inception v3 were proposed to improve the first version since there were a few problems. There was a representational bottleneck due to the dimension alteration. It was solved in the Inception v2 model by changing the 5x5 convolution by two 3x3 convolutions or factoring an  $n \times n$  convolution into a combination of  $1 \times n$  and  $n \times 1$  convolutions [103]. While this improved the computational cost, the auxiliary classifiers contributed little near the end of the training process. The inception v3 model implements the ideas of the Inception v2 model without drastically changing the modules; they implemented an RMSProp Optimizer, factorized  $7 \times 7$  convolutions, BatchNorm in the auxiliary classifiers, and Label Smoothing to prevent overfitting [103].

- ResNet

While working with neural networks, there was the idea that increasing the number of layers increases precision, which is true to a certain level because, after some depth, the performance degrades. This degradation of the performances is a problem in deep learning known as the *Vanishing/Exploding gradient*, which causes the gradient to become 0 or too large. In Figure 24 one can see the degradation of training accuracy; on the left, the training error, and on the right, the test error for a CIFAR-10 net with 20 and 56 layers; the deeper network (56 layers) has higher training and test error [105].

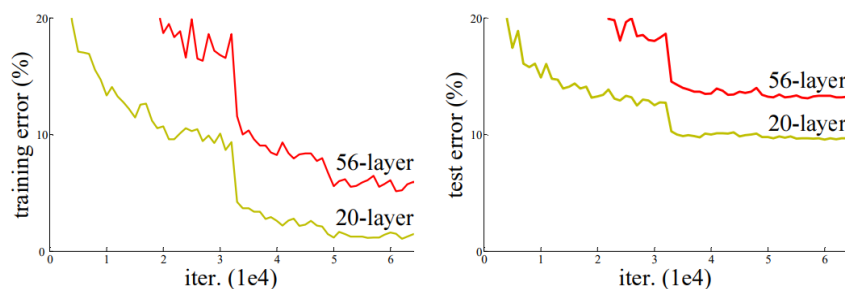


Figure 24: Training and test error for a CIFAR-10 net [105].

ResNet architecture was first introduced in [105] for solving vanishing/exploiding gradient problems. The architecture introduces the concept of *Residual Blocks* by using the *Skip Connections* technique; it consists of connecting activations of a layer to further layers, skipping some layers in between, and forming a residual block. By stacking these residual blocks, a ResNet CCN can be created. In general, the network uses a 34-layer network architecture similar to a VGG-19 net, in which the shortcut connection is added.

Neural networks are good function approximators; they should be able to solve the identity function  $f(x) = x$  easily [105]; hence the network is allowed to fit the residual mapping, where the output of a function becomes the input itself, i.e., let  $h(x)$  be the initial mapping. The function needed to be filled by the network is:

$$F(x) := h(x) - x \Leftrightarrow h(x) := F(x) + x.$$

Graphically this idea can be interpreted as in Figure 25.

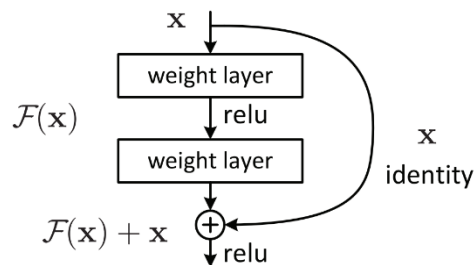


Figure 25: Residual block [105].

- DenseNet

DenseNets were introduced in [106]. This network connects every layer to every other layer. DenseNets alleviate the vanishing-gradient problem, strengthen feature propagation, encourage feature reuse, and substantially reduce the number of parameters [106]. DenseNet consists of several dense blocks, see Figure 26; between two adjacent blocks, there are two layers performing convolution and pooling, known as transition layers.

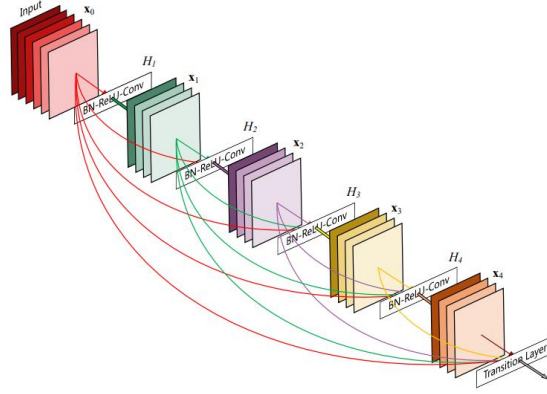


Figure 26: A 5-layer dense block in a DenseNet [106].

Regarding connectivity, in each layer, the feature maps of all the previous layers are not summed; instead, they are concatenated and used as inputs; consequently, DenseNets require fewer parameters. Mathematically [106], the  $l$ -th layer receives the feature maps of previous layers as input

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}])$$

The multiple inputs of  $H_l$  are concatenated into a single tensor to ease implementation.

There are several models of DenseNet. Due to the implementation of the Dense blocks, these networks can become deeper without worrying about the vanishing gradient problem. In this investigation work, the implemented networks are: a DenseNet 121 that consists of 1  $7 \times 7$  Convolution, 58  $3 \times 3$  Convolution, 61  $1 \times 1$  Convolution, 4 AvgPool, and 1 Fully Connected Layer. In summary, it has 120 Convolutions and 4 AvgPool.

#### 4.4.2. Framework and hardware acceleration

##### Machine learning framework

PyTorch was used to develop and implement the trained models in the present study since all the pre-trained models under study are implemented in the framework and can be easily imported using torchvision. As previously mentioned, the implemented models are pre-trained and pre-tested on the ImageNet data set, a data set generated for object detection with more than 14 million images and 20 000 categories or classes. Thus, the proposed models will be developed under the transferred learning methodology applying fine tuning and feature extraction methods.

##### Hardware acceleration

Several graphics processing units (GPUs) were employed (Nvidia A100, Nvidia Tesla P100, and Nvidia TeslaT4 of 40GB, 16GB, and 16GB, respectively) from Google Colaboratory with a 2.00GHz Intel® Xeon(R) CPU for hardware acceleration. GPUs are vital in the training of a DL model since training might take more or less time depending on their capacity. It is critical to understand that the Nvidia A100 GPU has a performance of 19.5 TFLOPs, the Tesla P100 GPU has a performance of 10.6 TFLOPs, and the Tesla T4 GPU has a performance of 8.07TFLOPs. Unfortunately, the Nvidia A100 and Tesla P100 GPUs are only available in the pro edition of Google Colab. Hence the Nvidia Tesla T4 GPU was the most utilized for network training and testing.

#### 4.5. Metrics

In order to objectively determine the results of the trained models, the following metrics are used. See Figure 27 for reference on the formulas and definitions.

		Predicted Value	
		Positive	Negative
Ground-Truth	Positive	<b>TP</b> True Positive	<b>FN</b> False Negative
	Negative	<b>FP</b> False Positive	<b>TN</b> True Negative

Figure 27: Possible predictions from the network.

#### Accuracy

As illustrated in Formula (1), accuracy is defined mathematically as the fraction or percentage of correct predictions (TP+TN) towards the total number of predictions (TP + FP + TN + FN).

$$Accuracy = \frac{TP + TN}{Total\ of\ predictions} = \frac{TP + TN}{TP + FP + TN + FN} \quad (1)$$

Where:

When the model accurately predicts positive outcomes, it is said to have a true positive value (TP). A true negative (TN) value is obtained when the model accurately predicts negative outcomes. A false positive value (FP) occurs when the model predicts positive outcomes inaccurately. Furthermore, a false negative (FN) value occurs when the model predicts a negative outcome inaccurately.

Accuracy is used to evaluate and fine-tune the model's objectives function. However, accuracy is only sometimes sufficient to indicate a model's total performance because accuracy might be high while other measures diverge.

### **Precision**

Precision see Formula (2), also called the Positive Predictive Value (PPV). It is a metric that quantifies the number of "True Predictions" (True Positives values) out of all the positive predicted values (all the diagnoses classified as "positive" by the model), whether or not they have been correctly classified. High Precision is the aim when the objective is to minimize false positives.

$$Precision = \frac{TP}{Positive\ values} = \frac{TP}{TP + FP} \quad (2)$$

### **Recall**

Recall, Formula (3), is another performance indicator that counts the number of positive predictions (TP) that the model successfully identifies out of all possible positive predictions. It is determined by dividing the number of true positive values by the total number of true positives + false negatives values It is also known as a model's True Positive Rate or Sensitivity. Recall shows missed positive predictions, as opposed to Precision, which only comments on the right positive predictions out of all positive predictions. In this way, Recall offers some idea of the positive class's coverage. When the goal is to limit false negatives, a high recall is desired.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

## **F1 Score**

Even though it is intuitive to use accuracy to evaluate a model, it is also recommendable to use Precision and Recall measures as there can be situations where accuracy is very high, but Precision or Recall is low. The aim is to avoid situations leading to mistreating subjects; for example, when the subject presents a disease, but the model classifies it as healthy, i.e., a FN, or when the subject does not present any disease, but the model classifies the opposite, i.e., a FP.

Maximizing Precision will minimize the number of FP, whereas maximizing the Recall will minimize the number of FN. Although the aim is to obtain high precision and recall values, it is impossible to achieve both simultaneously. Both metrics are important in medical diagnosis, so the aim is not only for a high recall but also for high Precision. Therefore, these two metrics are balanced using an F1 Score see formula (4). It is the harmonic mean of the Precision and Recall

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

F1 Score can be improved to indicate good Precision and recall value, so it is easier to work with than improving Precision or Recall.

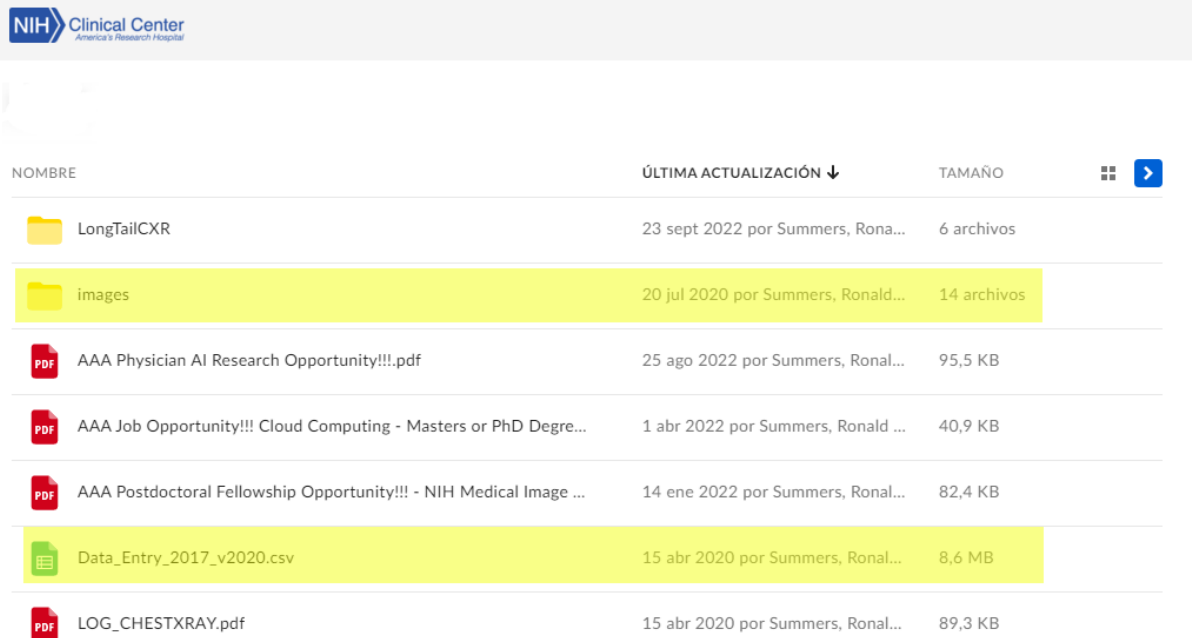


## 5. Results and Discussion

### 5.1. Data acquisition

The label information of each image in the dataset as well as the entire number of images were accessed via cloud technology from the NIH Clinical Center. Images in form 14 zip archives that contain a total of 45 GB of information as well as a main csv file called “Data\_Entry\_2017\_v2020.csv”, see Figure 28, and were downloaded from the NIHCC cloud data set.

Subsequently the data stored in the .zip files were unzipped, stored in a single folder, compressed into a mega file named "images.zip", and finally uploaded together with the csv file to the google drive space designated for the development of the studies which will then be accessed through google colab.



NOMBRE	ÚLTIMA ACTUALIZACIÓN ↓	TAMAÑO	
LongTailCXR	23 sept 2022 por Summers, Rona...	6 archivos	
images	20 jul 2020 por Summers, Ronald...	14 archivos	
AAA Physician AI Research Opportunity!!!.pdf	25 ago 2022 por Summers, Rona...	95,5 KB	
AAA Job Opportunity!!! Cloud Computing - Masters or PhD Degre...	1 abr 2022 por Summers, Ronald ...	40,9 KB	
AAA Postdoctoral Fellowship Opportunity!!! - NIH Medical Image ...	14 ene 2022 por Summers, Rona...	82,4 KB	
Data_Entry_2017_v2020.csv	15 abr 2020 por Summers, Rona...	8,6 MB	
LOG_CHESTXRAY.pdf	15 abr 2020 por Summers, Rona...	89,3 KB	

Figure 28: Data set.

### 5.2. Data Preprocessing

After uploading the .zip file containing the images and the .csv file with the information of the classes per image to the google colab environment, the information in the dataset was read and loaded into a pandas data frame structure. The information present in the dataset was read and loaded into a pandas data frame structure. Once the data was loaded into the data frame, it was

possible to visualize the presence of images reported with more than one disease see Figure 29, which, if not eliminated, would generate problems and errors in the learning process of the implemented models.

	Image Index	Finding Labels
0	0000001_000.png	Cardiomegaly
1	0000001_001.png	Cardiomegaly Emphysema
2	0000001_002.png	Cardiomegaly Effusion
3	0000002_000.png	No Finding
4	0000003_001.png	Hernia
5	0000003_002.png	Hernia
6	0000003_003.png	Hernia Infiltration
7	0000003_004.png	Hernia
8	0000003_005.png	Hernia
9	0000003_006.png	Hernia
10	0000003_007.png	Hernia
11	0000003_008.png	Hernia
12	0000004_000.png	Mass Nodule
13	0000005_000.png	No Finding
14	0000005_001.png	No Finding
15	0000005_002.png	No Finding
16	0000005_003.png	No Finding
17	0000005_004.png	No Finding
18	0000005_005.png	No Finding
19	0000005_006.png	Infiltration

[20 rows x 22 columns]  
(112120, 11)

Figure 29: Images labeled with more than one disease.

In this way, the information in the initial dataset was filtered, eliminating the information of the images whose report indicated more than one detected disease, thus reducing the number of images from 112 120 images to 89 237. After eliminating images with more than one class, the filtered dataset classes were separated by columns, where a positive diagnosis for a given class was marked with the number one. In contrast, a negative diagnosis for the rest of the diseases was marked with zero's see Figure 30.

```

      id Atelectasis Cardiomegaly Consolidation Edema \
0 00000583_003.png      1          0          0      0
1 00000583_008.png      1          0          0      0
2 00000583_009.png      1          0          0      0
3 00000583_010.png      1          0          0      0
4 00000583_031.png      1          0          0      0

Effusion  Emphysema  Fibrosis  Hernia  Infiltration  ... \
0      0      0      0      0      0      ...
1      0      0      0      0      0      ...
2      0      0      0      0      0      ...
3      0      0      0      0      0      ...
4      0      0      0      0      0      ...

Pleural_Thickening  Pneumonia  Pneumothorax  Pneumoperitoneum \
0      0      0      0      0
1      0      0      0      0
2      0      0      0      0
3      0      0      0      0
4      0      0      0      0

Pneumomediastinum  Subcutaneous Emphysema  Tortuous Aorta \
0      0      0      0
1      0      0      0
2      0      0      0
3      0      0      0
4      0      0      0

Calcification of the Aorta  No Finding  subject_id
0      0      0      583
1      0      0      583
2      0      0      583
3      0      0      583
4      0      0      583

[5 rows x 22 columns]
(89237, 22)

```

Figure 30: Structured data from the dataset once the overlapping was eliminated.

Thus, based on the results obtained and represented in Figure 31, it was possible to verify that the dataset does not maintain the balance between the different classes that compose it. The number of images belonging to the "no finding" class or "healthy patients" equals 58,509 of the 89,237 usable images, while classes such as pneumonia, emphysema, hernia, and pneumoperitoneum have less than five hundred images each. This type of class imbalance is very problematic because it dramatically affects the ability of the networks to learn and recognize each of the classes of interest effectively.

```
df_csv_labels = pd.read_csv('/content/drive/MyDrive/tesis/labels_originales/labels.csv', sep=';')
del df_csv_labels["id"]
del df_csv_labels["subject_id"]

for column in df_csv_labels:
    print(f'{column}: {df_csv_labels[column].sum()}')
```

Atelectasis: 4048  
Cardiomegaly: 1079  
Consolidation: 1279  
Edema: 644  
Effusion: 3849  
Emphysema: 360  
Fibrosis: 734  
Hernia: 134  
Infiltration: 9237  
Mass: 2113  
Nodule: 2647  
Pleural\_Thickening: 1113  
Pneumonia: 343  
Pneumothorax: 2049  
Pneumoperitoneum: 148  
Pneumomediastinum: 85  
Subcutaneous Emphysema: 115  
Tortuous Aorta: 491  
Calcification of the Aorta: 230  
No Finding: 58539

Figure 31: Number of Images per class

Thus, taking class imbalance into account, two different actions were required to reduce the class imbalance observed in the dataset. First, the discriminant method mentioned previously in the methodology section is employed. By counting the classes with the highest number of images, and the classes with the highest medical relevance, three classes, in addition to the "no finding" class, were selected for the training of the proposed neural networks. The selected classes were: "pneumonia", "atelectasis", and "mass".

Second, since "pneumonia" is one of the classes with the fewest images in the dataset, but it is also the disease with the highest medical relevance for the study, according to INEC. An specialized search for data sets that contained X-ray images for the detection of pneumonia was conducted. Chest X-Ray Images (Pneumonia), an open-access repository of Kaggle, a subsidiary of Google users may use to search and post data sets, was selected to increase the number of images of this important class.

Further data set filtering was performed to obtain the names of the images belonging to each selected class. Once the names of the images of interest were collected, the data of two thousand images per class were randomly taken to generate the definitive repository of images that would be used for the training of the models. In the case of the "pneumonia" class, all the existing

images in the primary dataset, ChestXnet, were taken, and the remaining images were randomly taken from the second repository found to optimize the learning of this class. Once the two thousand images were collected for each class, the repository was divided as described in the methodology. 80% of the images were for model training, and the remaining 20% was divided for validation and testing, 10% each. Finally, the filtered and separated images were stored in 2 folders, the first one called "training" (which in turn is divided into two subfolders, "training" and "validation"), and a second one called "testing" to facilitate the access to the information of the models during the learning process.

### 5.3. Testing different deep learning classification models.

As part of this experimental procedure, the first approach to image classification with balanced classes was performed using the convolutional neural models Resnet18, Resnet 50, and Resnet 101 with 25 epochs each, as well as the models Densenet 121, VGG19, and Inception V3 with 15 epochs respectively. The optimizer used to carry out the six experiments was the Stochastic Gradient Descent with Momentum (SGDM), and the results obtained are summarized in Table 3 and Figure 32.

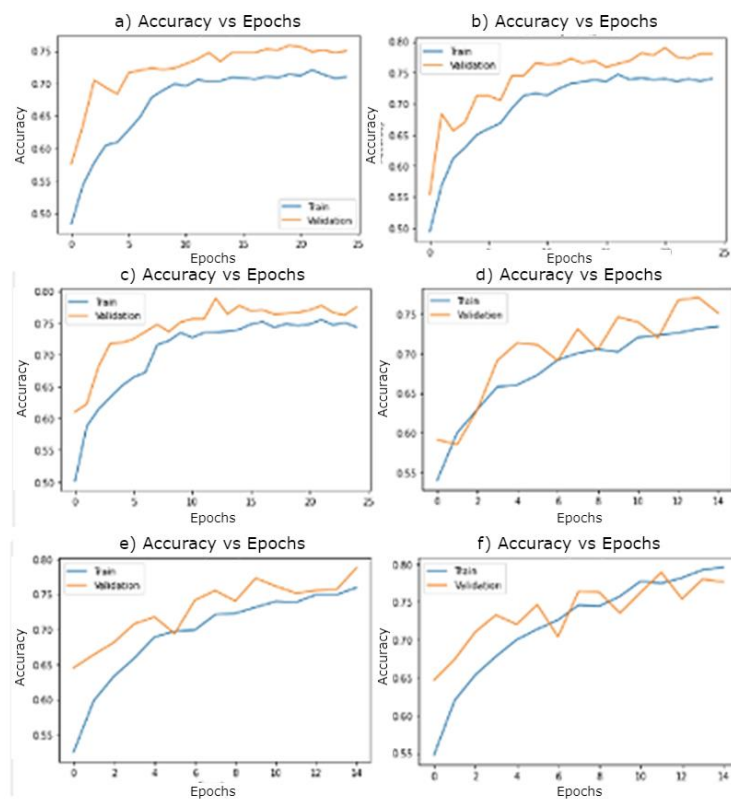


Figure 32: Accuracy results for the models a) resnet18, b) resnet50, c) resnet101, d) densenet121, e) vgg19, y f) InceptionV3.

Table 3: Training Initial Results

Model Name	Val acc	Test acc	Optimizer
resnet18_1_25epochs_SGDM	75,87	72,98	SGDM
resnet50_1_25epochs_SGDM	79,00	75,50	SGDM
resnet101_1_25epochs_SGDM	78,87	72,99	SGDM
densenet121_1_15epochs_SGDM	77,12	75.00	SGDM
vgg19_60_15epochs_SGDM	78,87	75.62	SGDM
inceptionv3_1_15epochs_SGDM	78,75	75,87	SGDM

The model name includes the number of training epochs; for each of the models, it can be seen that the Test Acc values vary from 72.98 to 75.87, showing that the network with the best performance in this first experiment was the Inception V3 network, followed by the VGG19 network, and the Densenet121 network in third place. On the other hand, the Resnet type networks did not present better performance within the learning process, and it is essential to note that for this task, it is observed that a greater number of layers does not ensure a higher accuracy value.

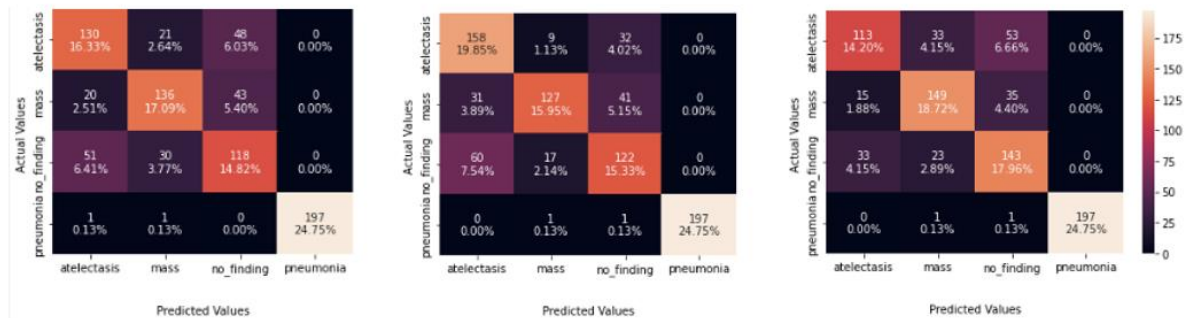


Figure 33: Confusion matrix of Resnet 18, InceptionV3, and VGG19 models for the initial dataset.

Also, in Figure 33 it is possible to observe the confusion matrices corresponding to the Resnet 18, Inception V3, and VGG19 networks. Here one can observe these networks performance in classifying each of the selected classes.

#### 5.4. Testing different Optimizers and Number of Epochs for the better classification models

For the subsequent experiments, the network with the highest initial performance (Inception V3) was taken for further training to optimize its metrics and achieve the best possible classification capacity. Thus, the previously trained InceptionV3 network was taken, and the value of the training epochs and the optimizer used were modified, as shown in Table 4.

*Table 4: InceptionV3 results for the initial dataset with different optimizers and epochs.*

<b>Model Name</b>	<b>Val acc</b>	<b>Test acc</b>	<b>Optimizer</b>
inceptionv3_1_14epochs_SGDM	78,75	75,87	SGDM
inceptionv3_2_25epochs_SGDM.pt	79,25	78,26	SGDM
inceptionv3_4_50epochs_SGDM	79,50	75,12	SGDM
inceptionv3_5_25epochs_ADAGRAD	79,62	79,27	ADAGRAD
inceptionv3_10_15epochs_ADAGRAD	79,36	77,28	ADAGRAD

From the results obtained, two essential things were inferred:

1. The model that best fits the classification of the analyzed medical images is the InceptionV3 model with 25 training epochs and the Adagrad optimizer achieving a final result of 79.27% Acc in the testing process.
2. A higher number of epochs does not always ensure a better accuracy result because the model may start to memorize the training images and not develop a good generalization capacity in the testing. This can be evidenced in Figure 34 where the InceptionV3 network with 50 epochs and SGDM optimizer obtained a test acc of 75.12 compared to the model with 25 epochs and the same optimizer, which obtained an acc test of 78.26.

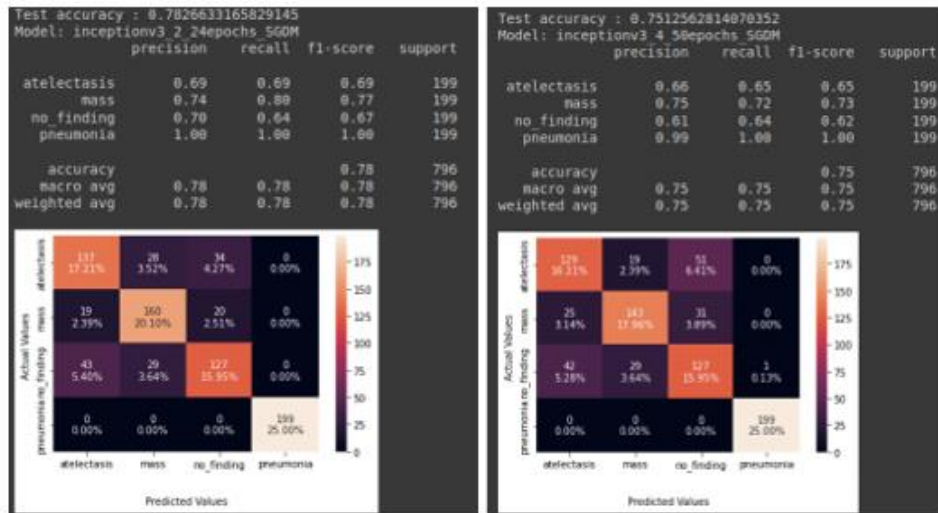


Figure 34: Confusion Matrix and Metrics of the model InceptionV3 with SGDM with

In addition, after performing an analysis of the confusion matrices generated by the two best models previously mentioned, Figure 34 and the models trained in the initial stage, Figure 33. There is a pattern in the results that reduce the overall accuracy of the implemented networks. In confusion matrices of the models, it is possible to observe that the highest number of errors in the classification is between the classes "atelectasis" and "no finding".

### 5.5. Testing a different dataset for a better classification model.

With this discovery, the next test proceeded based on the third discriminant condition for the treatment of the data mentioned in the methodology "classes will be chosen that do not generate errors between them." Thus, the data was checked once again to obtain a new class that would allow testing a new hypothesis (atelectasis and non-finding are classes that generate noise among themselves and do not allow a better performance of the networks).

"Effusion" was the class selected to replace "atelectasis" in the subsequent experiments. So again, three convolutional networks of the InceptionV3 type were retrained with a new dataset (Dataset5) and three different optimizers (Adagrad, SGDM, and Adam).



Table 5: InceptionV3 results with Adagrad, SGDM, and Adam, for the new dataset were Eddusion replace atelectasis(dataset5). Compared with InceptionV3+Adagrad with the initial dataset.

Model Name	Val acc	Test acc	Optimizer
inceptionv3_5_25epochs_ADAGRAD	79,625	79,27	ADAGRAD
inceptionv3_52_15epochs_ADAGRAD_d atase5	83,625	84,5	ADAGRAD
inceptionv3_55_15epochs_SGDM_datas et5	81,25	82,25	SGDM
inceptionv3_54_15epochs_ADAM_datas et5	80,125	78,625	ADAM

The results obtained in these three experiments are summarized in Table 5. Here it is possible to observe that, indeed, after changing the class "atelectasis" to the class "effusion," the test acc increases its value from 79.27% to 84.5% in the case of the best-performing network (inceptionv3\_52\_15epochs\_ADAGRAD\_dataset5). Moreover, based on the confusion matrices of the last trained models, Figure 35, it can be observed that this increase in the accuracy value is due to better discrimination between the "effusion" and "no finding" classes.

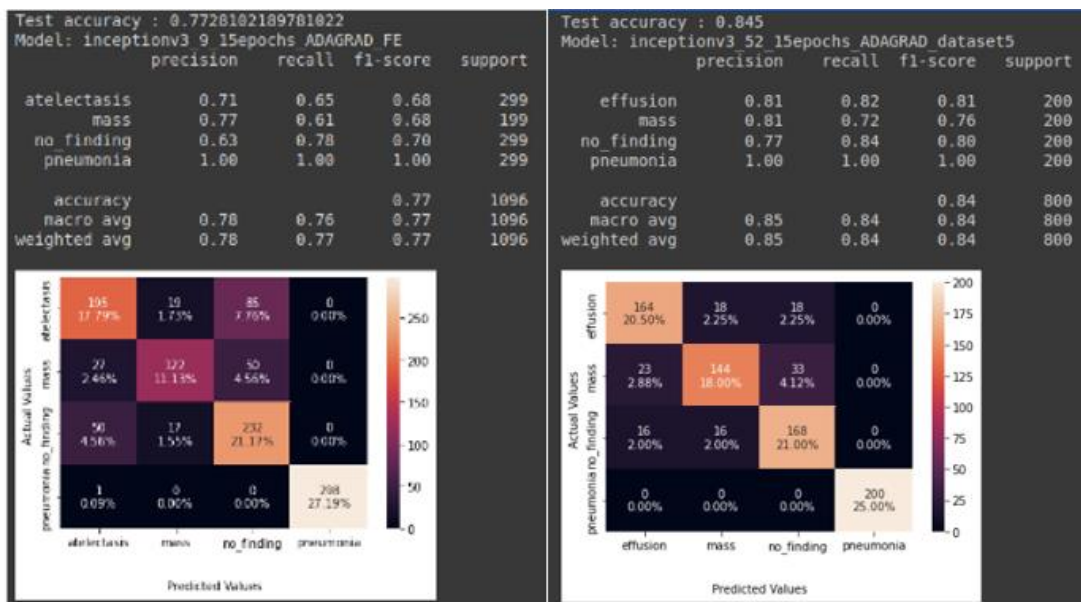


Figure 35: Confusion Matrix of InceptionV3 model for the initial dataset, and the second dataset (Effusion instead of atelectasis)

## 5.6. Final Considerations

Table 6: Different models trained in the current experimental process with the precision results obtained per each of the pathologies analyzed

Model Name / Precision per disease.	Atelectasis	Mass	Pneumonia	No Finding	Effusion
resnet18_1_25epochs_SGDM	0.64	0.72	1.0	0.56	NA
resnet50_1_25epochs_SGDM	0.68	0.77	1.0	0.59	NA
resnet101_1_25epochs_SGDM	0.67	0.73	1.0	0.56	NA
densenet121_1_15epochs_SGDM	0.61	0.84	1.0	0.61	NA
vgg19_60_15epochs_SGDM	0.70	0.72	1.0	0.62	NA
inceptionv3_1_15epochs_SGDM	0.80	0.74	1.0	0.75	NA
inceptionv3_52_15epochs_ADAGR AD_dataset5	NA	0.81	1.0	0.77	0.81
inceptionv3_55_15epochs_SGDM_d ataset5	NA	0.74	1.0	0.75	0.80
inceptionv3_54_15epochs_ADAM_d ataset5	NA	0.78	1.0	0.71	0.68

Negative interaction during the training process

Positive interaction during the training process

Different approaches to developing an efficient classifier for detecting pulmonary pathologies were developed during the current experimental process. Between these, it is possible to observe again that the best configuration among the number of epochs, optimizers, and diseases of interest was the model “inceptionv3\_52\_15epochs\_ADAGRAD\_dataset5” with an overall accuracy of 84.5% compared with a similar model with the same optimizer and number of epochs one different disease that obtained just 79.27%. In the same way, in Table 6 it can be observed that this change in the accuracy of between models responds to the negative/positive interaction the diseases “atelectasis” and “Effusion” have with the class no finding, and it must be considered in future experiments.

Finally, as can be seen in Table 7, among the results obtained in the literature search, it is possible to note that in [83], scientist report accuracies of 70.69%, 66.42%, 56.44%, and 63.33% for the classes atelectasis, effusion, mass, and pneumonia, respectively, diagnosed through a Resnte50 type network pre-trained with the imagenet data set. While in [84], results of 77.2%, 85.9%, 79.2%, and 71.3% accuracy are reported through the training of networks elaborated from scratch. Thus, one can say that the experiments performed through the InceptionV3 network have managed to surpass the results obtained by [83]. In contrast, these results are slightly below those of [84]. This improvement in the results of [84] is because the implemented models are pre-trained in the Imagenet data set, which can present particular types of optimization biases that the networks trained from scratch do not generate and can improve

their accuracy. It is also important to mention that thanks to the addition of the specialized image data set for pneumonia detection, InceptionV3 model far outperforms both previously performed studies and outperforms the results reported by [91]. They obtained an accuracy of 94% in their intentional study for pneumonia detection.

*Table 7: Comparison of results among references.*

<b>Pathology</b>	<b>Wang [83]</b>	<b>Yao[84]</b>	<b>Okeke [91]</b>	<b>Our Results</b>
<b>Atelectasis</b>	0.716	0.772	-	0.71
<b>Effusion</b>	0.784	0.859	-	0.81
<b>Mass</b>	0.706	0.792	-	0.81
<b>Pneumonia</b>	0.633	0.713	0.9373	1

## 6. Conclusions

After knowing the results of the work performed it can be concluded that although it was possible to create and debug a congruent dataset for the convolutional neural networks training, it was also possible to train and improve the proposed models efficiently with a mapping of their enhancement parameters. It was also possible to train and improve the proposed models efficiently by mapping their improvement parameters. The trained models still need to improve their accuracy to be accepted and used within a hospital environment. Therefore, it was not possible to build a model that effectively and efficiently identifies the proposed lung pathologies.

However, this model could be upgraded for use in rural areas. If this is the case, a minimum of technical specifications will be required such as a computer with average capabilities with respect to processor and ram memory specifications and without the need for a GPU, but with good internet connection to run the program on the Google Collab platform or in case of not having internet access a computer with an intel core i5 8th gen processor or higher, a graphics card with at least 8 Gb of dedicated memory and 12 Gb of RAM. Once the network is trained, it will not require the same amount of memory used during the training phase.

The inceptionV3 convolutional network with a fine-tuning transfer learning methodology is the network best suited to the proposed problem. Similarly, Adagrad is the optimizer that generates the best results among the proposed optimizers. However, despite the different optimizers, the number of epochs, and the learning coefficients proposed, a single variable seems to have the power to drastically change results previously obtained in the literature, the quantity, and the quality of images used to feed the network. This was evidenced by the significant change in accuracy observed for the pneumonia category, whose images were 100% identified by the InceptionV3 model.

Therefore, it is of utmost importance to encourage the creation and development of public repositories of images and medical data, which in an organized way, allow scientists to train

and improve their computational models and generate solutions that help detect, discover, and overcome important pathologies for nowadays humanity.

Finally, it is essential to recognize that more technology and data are needed to develop a system to help detect lung pathologies efficiently. Nevertheless, the results obtained in this investigation work are promising, especially for areas with no radiology specialists. Furthermore, with the accession of new data sets with quality images, achieving results that will help thousands of people in Ecuador and the world, as they helped with the detection of COVID-19 during the 2020 pandemic, is possible.

## 7. Bibliography

- [1] J. T. M. Kay, “Color Atlas and Textbook of Human Anatomy,” *J. Clin. Pathol.*, vol. 32, no. 8, 1979.
- [2] F. Netter, *Netter’s Atlas of Human Anatomy*, vol. 44, no. 8. 2019.
- [3] M. G. Levitzky, *Lange Pulmonary Physiology*. 2007.
- [4] S. E. Weinberger, B. A. Cockrill, and J. Mandel, *Principles of pulmonary medicine*. 2018.
- [5] J. B. West, “Environmental, neoplastic, and infectious diseases,” in *Pulmonary Pathophysiology: The Essentials*, 2013.
- [6] N. C. Gonzalez, “Pulmonary Pathophysiology: The Essentials, 5th Edition,” *Med. Sci. Sport. Exerc.*, vol. 31, no. 1, 1999.
- [7] John B West, *West’s Pulmonary Pathophysiology - The Essentials 2017*, vol. 53, no. 9. 2019.
- [8] D. H. Ingbar, “Fishman’s pulmonary diseases and disorders, 5th edition,” *Annals of the American Thoracic Society*, vol. 12, no. 8. 2015.
- [9] M. Sapti, *Harrison’s Pulmonary And Critical Care Medicine 3rd Edition.*, vol. 53, no. 9. 2019.
- [10] Q. Zhao *et al.*, “[A review of deep learning methods for the detection and classification of pulmonary nodules].,” *Sheng Wu Yi Xue Gong Cheng Xue Za Zhi*, vol. 36, no. 6, 2019.

- [11] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine learning for medical imaging," *Radiographics*, vol. 37, no. 2, 2017.
- [12] D. D. Miller and E. W. Brown, "Artificial Intelligence in Medical Practice: The Question to the Answer?," *American Journal of Medicine*, vol. 131, no. 2. 2018.
- [13] E. Baratella, C. Marrocchio, A. M. Bozzato, E. Roman-Pognuz, and M. A. Cova, "Chest X-ray in intensive care unit patients: what there is to know about thoracic devices," *Diagnostic Interv. Radiol.*, vol. 27, no. 5, 2021.
- [14] G. Chassagnon, M. Vakalopoulou, N. Paragios, and M. P. Revel, "Artificial intelligence applications for thoracic imaging," *Eur. J. Radiol.*, vol. 123, 2020.
- [15] R. Fitzgerald, "Error in radiology," *Clinical Radiology*, vol. 56, no. 12. 2001.
- [16] A. P. Brady, "Error and discrepancy in radiology: inevitable or avoidable?," *Insights into Imaging*, vol. 8, no. 1. 2017.
- [17] S. Waite, J. M. Scott, A. Legasto, S. Kolla, B. Gale, and E. A. Krupinski, "Systemic error in radiology," *Am. J. Roentgenol.*, vol. 209, no. 3, 2017.
- [18] D. Moreno-Zambrano, D. Santamaria, A. Barco, and R. Santibanez, "Assessment of cognitive deficits in middle-aged type II diabetics View project," 2016.
- [19] D. J. Manning, S. C. Ethell, and T. Donovan, "Detection or decision errors? Missed lung cancer from the posteroanterior chest

- radiograph,” *Br. J. Radiol.*, vol. 77, no. 915, 2004.
- [20] M. Remy-Jardin *et al.*, “Machine Learning and Deep Neural Network Applications in the Thorax: Pulmonary Embolism, Chronic Thromboembolic Pulmonary Hypertension, Aorta, and Chronic Obstructive Pulmonary Disease,” *Journal of Thoracic Imaging*. 2020.
- [21] C. T. Wu *et al.*, “Acute exacerbation of a chronic obstructive pulmonary disease prediction system using wearable device data, machine learning, and deep learning: Development and cohort study,” *JMIR mHealth uHealth*, vol. 9, no. 5, 2021.
- [22] X. Zhao *et al.*, “Deep CNN models for pulmonary nodule classification: Model modification, model integration, and transfer learning,” *J. Xray. Sci. Technol.*, vol. 27, no. 4, 2019.
- [23] A. Takiddin, J. Schneider, Y. Yang, A. Abd-Alrazaq, and M. Househ, “Artificial intelligence for skin cancer detection: Scoping review,” *Journal of Medical Internet Research*, vol. 23, no. 11. 2021.
- [24] J. H. Lee *et al.*, “Deep learning–based automated detection algorithm for active pulmonary tuberculosis on chest radiographs: diagnostic performance in systematic screening of asymptomatic individuals,” *Eur. Radiol.*, vol. 31, no. 2, 2021.
- [25] V. Gulshan *et al.*, “Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs,” *JAMA - J. Am. Med. Assoc.*, vol. 316, no. 22, 2016.



- [26] R. Raman, S. Srinivasan, S. Virmani, S. Sivaprasad, C. Rao, and R. Rajalakshmi, "Fundus photograph-based deep learning algorithms in detecting diabetic retinopathy," *Eye (Basingstoke)*, vol. 33, no. 1. 2019.
- [27] J. Premaladha and K. S. Ravichandran, "Novel Approaches for Diagnosing Melanoma Skin Lesions Through Supervised and Deep Learning Algorithms," *J. Med. Syst.*, vol. 40, no. 4, 2016.
- [28] R. D. Seeja and A. Suresh, "Deep learning based skin lesion segmentation and classification of melanoma using support vector machine (SVM)," *Asian Pacific J. Cancer Prev.*, vol. 20, no. 5, 2019.
- [29] M. Dildar *et al.*, "Skin cancer detection: A review using deep learning techniques," *International Journal of Environmental Research and Public Health*, vol. 18, no. 10. 2021.
- [30] E. Shoen, "DermIA: Machine Learning to Improve Skin Cancer Screening," *Journal of Digital Imaging*, vol. 34, no. 6. 2021.
- [31] D. F. Steiner *et al.*, "Impact of Deep Learning Assistance on the Histopathologic Review of Lymph Nodes for Metastatic Breast Cancer," *Am. J. Surg. Pathol.*, vol. 42, no. 12, 2018.
- [32] H. Yoo, K. H. Kim, R. Singh, S. R. Digumarthy, and M. K. Kalra, "Validation of a Deep Learning Algorithm for the Detection of Malignant Pulmonary Nodules in Chest Radiographs," *JAMA Netw. Open*, vol. 3, no. 9, 2020.
- [33] P. Rajpurkar *et al.*, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to

- practicing radiologists,” *PLoS Med.*, vol. 15, no. 11, 2018.
- [34] R. H. Abiyev and M. K. S. Ma’aitah, “Deep Convolutional Neural Networks for Chest Diseases Detection,” *J. Healthc. Eng.*, vol. 2018, 2018.
- [35] G. Dhiman, V. Vinoth Kumar, A. Kaur, and A. Sharma, “DON: Deep Learning and Optimization-Based Framework for Detection of Novel Coronavirus Disease Using X-ray Images,” *Interdiscip. Sci. – Comput. Life Sci.*, vol. 13, no. 2, 2021.
- [36] T. S. Qaid, H. Mazaar, M. Y. H. Al-Shamri, M. S. Alqahtani, A. A. Raweh, and W. Alakwaa, “Hybrid Deep-Learning and Machine-Learning Models for Predicting COVID-19,” *Comput. Intell. Neurosci.*, vol. 2021, 2021.
- [37] S. Albahli and G. N. A. Hassan Yar, “Fast and accurate detection of COVID-19 along with 14 other chest pathologies using a multi-level classification: Algorithm development and validation study,” *J. Med. Internet Res.*, vol. 23, no. 2, 2021.
- [38] J. D. Arias-Londono, J. A. Gomez-Garcia, L. Moro-Velazquez, and J. I. Godino-Llorente, “Artificial Intelligence applied to chest X-Ray images for the automatic detection of COVID-19. A thoughtful evaluation approach,” *IEEE Access*, 2020.
- [39] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, “Explainable Deep Learning for Pulmonary Disease and Coronavirus COVID-19 Detection from X-rays,” *Comput. Methods Programs Biomed.*, vol. 196, 2020.

- [40] D. A. Moses, “Deep learning applied to automatic disease detection using chest X-rays,” *J. Med. Imaging Radiat. Oncol.*, vol. 65, no. 5, 2021.
- [41] R. Kundu, R. Das, Z. W. Geem, G. T. Han, and R. Sarkar, “Pneumonia detection in chest X-ray images using an ensemble of deep learning models,” *PLoS One*, vol. 16, no. 9 September, 2021.
- [42] V. Chouhan *et al.*, “A novel transfer learning based approach for pneumonia detection in chest X-ray images,” *Appl. Sci.*, vol. 10, no. 2, 2020.
- [43] Y. Feng, Y. Wang, C. Zeng, and H. Mao, “Artificial intelligence and machine learning in chronic airway diseases: Focus on asthma and chronic obstructive pulmonary disease,” *International Journal of Medical Sciences*, vol. 18, no. 13. 2021.
- [44] Y. Cho, J. S. Kim, T. H. Lim, I. Lee, and J. Choi, “Detection of the location of pneumothorax in chest X-rays using small artificial neural networks and a simple training process,” *Sci. Rep.*, vol. 11, no. 1, 2021.
- [45] S. Barbiero *et al.*, “Single-fraction flattening filter-free volumetric modulated arc therapy for lung cancer: Dosimetric results and comparison with flattened beams technique,” *Med. Dosim.*, vol. 41, no. 4, pp. 334–338, 2016.
- [46] J. Ma, Y. Song, X. Tian, Y. Hua, R. Zhang, and J. Wu, “Survey on deep learning for pulmonary medical imaging,” *Frontiers of Medicine*, vol. 14, no. 4. 2020.
- [47] A. R. Clark, E. J. Her, R. Metcalfe, and C. A. Byrnes, “Could

- automated analysis of chest X-rays detect early bronchiectasis in children?,” *Eur. J. Pediatr.*, vol. 180, no. 10, 2021.
- [48] M. E. Kruk *et al.*, “High-quality health systems in the Sustainable Development Goals era: time for a revolution,” *The Lancet Global Health*, vol. 6, no. 11. 2018.
- [49] E. E. Kim, H. Im, D. S. Lee, and K. W. Kang, *Atlas and Anatomy of PET/MRI, PET/CT and SPECT/CT*. 2016.
- [50] F. H. NETTER and D. ROSE, “Respiratory System (The Ciba Collection of Medical Illustrations, Vol. 7),” *Anesthesiology*, vol. 52, no. 5, 1980.
- [51] B. A. Al Enizi, N. Saquib, M. S. A. Zaghloul, M. S. A. Alaboud, M. S. Shahid, and J. Saquib, “Ross and Willson Anatomy and Physiology,” *Int. J. Health Sci. (Qassim)*., vol. 10, no. 3, 2016.
- [52] R. A. Rhoades and D. Bell, *Medical Physiology Fifth Edition*, vol. 53, no. 9. 2017.
- [53] J. E. Hall and A. C. Guyton, *GUYTON, A.C ; HALL, J. E. Guyton & Hall: Tratado de Fisiologia Médica 11<sup>a</sup> ed., Elsevier, Rio de Janeiro, 2006.*, vol. 53. 2011.
- [54] K. Szalontai *et al.*, “Chronic obstructive pulmonary disease: Epidemiology, biomarkers, and paving the way to lung cancer,” *Journal of Clinical Medicine*, vol. 10, no. 13. 2021.
- [55] D. Elia *et al.*, “Pulmonary hypertension and chronic lung disease: Where are we headed?,” *European Respiratory Review*, vol. 28, no. 153. 2019.

- [56] Q. Y. Yu and X. X. Tang, “Irreversibility of Pulmonary Fibrosis,” *Aging and Disease*, vol. 13, no. 1. 2022.
- [57] S. Miethe, A. Karsonova, A. Karaulov, and H. Renz, “Obesity and asthma,” *Journal of Allergy and Clinical Immunology*, vol. 146, no. 4. 2020.
- [58] W. Seeger *et al.*, “Pulmonary hypertension in chronic lung diseases,” in *Journal of the American College of Cardiology*, 2013, vol. 62, no. 25 SUPPL.
- [59] K. Brown and T. J. Lee-Chiong, *Oxford American Handbook of Pulmonary Medicine*. 2009.
- [60] G. Waterer, “What is pneumonia?,” *Breathe*, vol. 17, no. 3. 2021.
- [61] C. W. Lanks, A. I. Musani, and D. W. Hsia, “Community-acquired Pneumonia and Hospital-acquired Pneumonia,” *Medical Clinics of North America*, vol. 103, no. 3. 2019.
- [62] K. C. Thandra, A. Barsouk, K. Saginala, J. S. Aluru, and A. Barsouk, “Epidemiology of lung cancer,” *Wspolczesna Onkologia*, vol. 25, no. 1. 2021.
- [63] López *et al.*, “Non-small cell lung cancer,” *Med.*, vol. 13, no. 25, 2021.
- [64] S. D. Nathan *et al.*, “Pulmonary hypertension in chronic lung disease and hypoxia,” *Eur. Respir. J.*, vol. 53, no. 1, 2020.
- [65] A. Sydykov *et al.*, “Pulmonary hypertension in acute and chronic high altitude maladaptation disorders,” *International Journal of Environmental Research and Public Health*, vol. 18, no. 4. 2021.
- [66] S. Weinberger, B. Cockrill, and J. Mandel, *Principles of*

*Pulmonary Medicine: Sixth Edition.* 2013.

- [67] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, “Deep Learning for Computer Vision: A Brief Review,” *Computational Intelligence and Neuroscience*, vol. 2018. 2018.
- [68] J. Brownlee, “Deep Learning for Computer Vision Image Classification , Object Detection , and Face Recognition in Python UNLOCK Computer Vision With Deep Learning,” *Deep Learn. Comput. Vis.*, 2019.
- [69] A. F. Gad, *Practical Computer Vision Applications Using Deep Learning with CNNs.* 2018.
- [70] A. Rosebrock, *Deep Learning for Computer Vision with Python - Practitioner Bundle*, vol. 53, no. 9. 1981.
- [71] Dr. Adrian Rosebrock, *1. Deep Learning for Computer Vision with Python - Starter Bundle.* 2017.
- [72] F. Nensa, A. Demircioglu, and C. Rischpler, “Artificial intelligence in nuclear medicine,” *J. Nucl. Med.*, vol. 60, no. 9, 2019.
- [73] C. W. Connor, “Artificial Intelligence and Machine Learning in Anesthesiology,” *Anesthesiology*, vol. 131, no. 6, 2019.
- [74] N. Milosevic, *Introduction to Convolutional Neural Networks.* 2020.
- [75] J. Gu *et al.*, “Recent advances in convolutional neural networks,” *Pattern Recognit.*, vol. 77, 2018.

- [76] C. Szegedy *et al.*, “Going deeper with convolutions,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2015, vol. 07-12-June-2015.
- [77] D. A. Rosebrock, “Deep Learning for Computer Vision -- Starter Edition,” *Learn Comput. Vis. Using OpenCV*, 2019.
- [78] R. Venkatesan and B. Li, *Convolutional Neural Networks in Visual Computing*. 2017.
- [79] U. Michelucci, *Advanced applied deep learning: Convolutional neural networks and object detection*. 2019.
- [80] A. Rosebrock, “Deep Learning for Computer Vision with Python - Starter Bundle,” *Pyimagesearch*, 2017.
- [81] S. Rajaraman, I. Kim, and S. K. Antani, “Detection and visualization of abnormality in chest radiographs using modality-specific convolutional neural network ensembles,” *PeerJ*, vol. 2020, no. 3, 2020.
- [82] M. Tariqul Islam, M. A. Aowal, A. Tahseen Minhaz, and K. Ashraf, “Abnormality Detection and Localization in Chest X-Rays using Deep Convolutional Neural Networks,” *arXiv e-prints*, p. arXiv:1705.09850, May 2017.
- [83] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*

- 2017, 2017, vol. 2017-January.
- [84] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels," *arXiv e-prints*, p. arXiv:1710.10501, 2017.
- [85] E. Escalera *et al.*, "Small cell lung cancer," *Med.*, vol. 13, no. 25, 2021.
- [86] E. J. Hwang *et al.*, "Development and Validation of a Deep Learning-Based Automated Detection Algorithm for Major Thoracic Diseases on Chest Radiographs," *JAMA Netw. open*, vol. 2, no. 3, 2019.
- [87] D. Zhao *et al.*, "A comparative study on the clinical features of coronavirus 2019 (COVID-19) pneumonia with other pneumonias," *Clin. Infect. Dis.*, vol. 71, no. 15, 2020.
- [88] G. Wang *et al.*, "A deep-learning pipeline for the diagnosis and discrimination of viral, non-viral and COVID-19 pneumonia from chest X-ray images," *Nat. Biomed. Eng.*, vol. 5, no. 6, 2021.
- [89] S. H. Ahammad *et al.*, "Chexnet reimplementaion for pneumonia detection using pytorch," *Int. J. Pharm. Res.*, vol. 12, no. 2, 2020.
- [90] J. G. Nam *et al.*, "Development and validation of deep learning-based automatic detection algorithm for malignant pulmonary nodules on chest radiographs," *Radiology*, vol. 290, no. 1, 2019.
- [91] O. Stephen, M. Sain, U. J. Maduh, and D. U. Jeong, "An Efficient Deep Learning Approach to Pneumonia Classification in



- Healthcare,” *J. Healthc. Eng.*, vol. 2019, 2019.
- [92] J. H. Lee *et al.*, “Performance of a deep learning algorithm compared with radiologic interpretation for lung cancer detection on chest radiographs in a health screening population,” *Radiology*, vol. 297, no. 3, 2020.
- [93] J. R. Ferreira, D. Armando Cardona Cardenas, R. A. Moreno, M. De Fatima De Sa Rebelo, J. E. Krieger, and M. Antonio Gutierrez, “Multi-View Ensemble Convolutional Neural Network to Improve Classification of Pneumonia in Low Contrast Chest X-Ray Images,” in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2020, vol. 2020-July.
- [94] S. Rajaraman and S. K. Antani, “Modality-Specific Deep Learning Model Ensembles Toward Improving TB Detection in Chest Radiographs,” *IEEE Access*, vol. 8, 2020.
- [95] Q. Wang *et al.*, “Automated segmentation and diagnosis of pneumothorax on chest X-rays with fully convolutional multi-scale ScSE-DenseNet: a retrospective study,” *BMC Med. Inform. Decis. Mak.*, vol. 20, 2020.
- [96] J. D. Schroeder *et al.*, “Prediction of Obstructive Lung Disease from Chest Radiographs via Deep Learning Trained on Pulmonary Function Data,” *Int. J. COPD*, vol. 15, 2020.
- [97] S. V J and J. F. D, “Deep Learning Algorithm for COVID-19 Classification Using Chest X-Ray Images,” *Comput. Math. Methods Med.*, vol. 2021, 2021.

- [98] C. S. Widodo, A. Naba, M. M. Mahasin, Y. Yueniwati, T. A. Putranto, and P. I. Patra, "UBNet: Deep learning-based approach for automatic X-ray image detection of pneumonia and COVID-19 patients," *J. Xray. Sci. Technol.*, vol. 30, no. 1, 2022.
- [99] M. J. Nelson and A. K. Hoover, "Notes on Using Google Colaboratory in AI Education," in *Annual Conference on Innovation and Technology in Computer Science Education, ITiCSE*, 2020.
- [100] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.
- [101] J. Jaworek-Korjakowska, P. Kleczek, and M. Gorgon, "Melanoma thickness prediction based on convolutional neural network with VGG-19 model transfer learning," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2019, vol. 2019-June.
- [102] S. Tammina, "Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images," *Int. J. Sci. Res. Publ.*, vol. 9, no. 10, 2019.
- [103] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-

December.

- [104]C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-ResNet and the impact of residual connections on learning,” in *31st AAAI Conference on Artificial Intelligence, AAAI 2017*, 2017.
- [105]K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016, vol. 2016-December.
- [106]G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-January.