

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

# Deep Learning X-Ray Analysis for Pulmonary Diseases

Ana Marques



Mestrado Integrado em Engenharia Informática e Computação

Supervisor: Rui Camacho

October 14, 2022



# **Deep Learning X-Ray Analysis for Pulmonary Diseases**

Ana Marques

Mestrado Integrado em Engenharia Informática e Computação

October 14, 2022



# Abstract

Lung disease identification through X-ray imaging is still an exclusively manual procedure. This means it is prone to human error, such as distraction, inexperience, or fatigue from the radiologist. A tool that could work as a support to the medical professionals, giving them a "second opinion," pointing to details that could have been missed, could lead to a significant increase in diagnosis success. Because it would help reduce uncertainty, it could also accelerate the diagnosis process and help initiate the treatment of the patient sooner. To evaluate the possibility of implementing such a tool, a study was made on the performance of deep learning algorithms in identifying several pathologies on chest X-ray images.

Deep learning models have been proven a powerful tool for medical data analytics. They can identify some implicit features that cannot be easily described. This capacity will be used in this thesis, in which the model must learn to distinguish several lung pathologies based on the features found in the dataset images.

We have based our work on an existing initial study and have used its dataset. It contained around 220 000 use cases and was put together by the CheXpert paper authors. It is available in two versions, one with the full resolution of the images, of around 400GB, and one with lower resolution images, of about 11GB. For this work, the latter version was used. This dataset consisted of actual case studies, so some patients had more than one underlying disease. We filtered the dataset to remove use cases with more than one pathology, reducing its original size by 30%. The original paper was done with a multi-label approach, differentiating the pathologies simultaneously. In contrast, this work was done with a binary label approach, "disease" or "no disease," repeated for each pathology.

The first part of the experimental phase consisted of running the algorithm for each of the twelve pathologies. The worst result of this phase was in the "Pleural Other" category, with an AUC of 0.9937. The best ones had an AUC of 0.9971 for Enlarged Cardiomedastinum and Consolidation classes. Then, in the second phase, each pathology sub-set was filtered by one of six attributes, such as projection or view of the radiography or sex of the patient, to see how performance was affected and draw conclusions from that. The worst results of the second phase, for the isolated attributes, were for the Posterior-Anterior projection of the Pleural Other category, with an AUC of 0.6645. The best was for the Anterior-Posterior projection of the Edema category, with an AUC of 0.9073.

The final results are very close to the ones obtained in the original CheXpert study and the leader board of the ongoing competition on the Stanford website, related to the same paper. This indicates that a binary-label approach is similar to a multi-label approach. These studies lead to belief in a promising future for medical diagnosis assisted by Artificial Intelligence.



# Resumo

O diagnóstico de doenças pulmonares através de raios-X de tórax é ainda um procedimento exclusivamente manual. Isto significa que é suscetível a erros humanos, tais como distração, inexperiência, ou fadiga por parte do radiologista. Uma ferramenta que funcionasse como apoio aos profissionais de saúde, oferecendo uma "segunda opinião", apontando detalhes que podem não ter sido detetados, pode levar a um aumento significativo no sucesso do diagnóstico. Pelo facto de ajudar na redução da incerteza, poderá também acelerar o processo de diagnóstico e fazer chegar o tratamento ao paciente mais cedo. Para avaliar a possibilidade de implementação de uma ferramenta com estas características, foi feito um estudo sobre a performance de algoritmos de *deep learning* na identificação de diversas patologias em imagens de raios-X do tórax.

Modelos de *deep learning* têm-se mostrado ferramentas poderosas para análise de dados médicos. Estes algoritmos são capazes de identificar características implícitas que não são facilmente descritas. Esta capacidade vai ser utilizada nesta tese para os algoritmos a serem construídos, dos quais os modelos terão de aprender a distinguir patologias de acordo com as características que conseguem detetar nas imagens.

Este trabalho é baseado num estudo já existente, "CheXpert", e utilizamos o seu conjunto de dados, ou *dataset*. Este contém mais de 220 mil casos de estudo e foi reunido pelos autores. Está disponível em duas versões, uma com imagens de alta resolução (a resolução original) e outra com a resolução reduzida. O tamanho destas duas versões é de aproximadamente 400 GB e 11 GB, respetivamente. Neste trabalho foi utilizada a segunda versão. Este *dataset* consiste em casos de estudo reais, portanto alguns pacientes acusaram mais do que uma patologia em simultâneo. Filtramos o conjunto total para remover estes casos de patologias sobrepostas, reduzindo em 30% o seu tamanho original. O estudo "CheXpert" foi feito com uma abordagem de múltiplas categorias identificadas e diferenciadas em simultâneo, enquanto que neste trabalho temos uma abordagem binária de "com anomalia" ou "saudável", repetida para cada patologia individualmente.

A primeira parte do processo experimental consistiu na execução de um algoritmo para cada uma das doze patologias. O pior resultado da experiência foi na categoria "*Pleural Other*", ou "Outras anomalias na pleura", com um AUC (*Area Under ROC Curve*, medida de 0 a 1) de 0.9937. Os melhores resultados foram nas categorias "*Enlarged Cardiomediastinum*" (Silhueta Cardiomiastinal Aumentada) e "*Consolidation*" (Consolidação). Na segunda parte do processo experimental, cada conjunto de imagens correspondente a cada patologia foi dividido dependendo de certos atributos isolados, tais como projeção e ângulo da radiografia ou sexo do paciente. O pior resultado desta fase correspondeu à projeção Posterior-Anterior da anomalia "Pleural Other", com um AUC de 0.6645 e o melhor resultado correspondeu à projeção Anterior-Posterior dos casos de Edema, com um AUC de 0.9073.

Os valores obtidos foram, no geral, muito próximos aos obtidos no estudo em que este trabalho foi baseado, o que indica que o método de classificação binária com dados filtrados tem uma validade semelhante à classificação múltipla simultânea. Ambos os estudos levam a acreditar num futuro promissor para o diagnóstico médico apoiado de inteligência artificial.





# Acknowledgements

I would firstly like to show my appreciation to my supervisor, Professor Rui Camacho, for his advice and guidance along this project.

I am also very grateful for my parents and family, who shaped me into the person I am today, and for my friends, who walked this path alongside me.

Finally, a special thanks to my partner for their consistent support throughout the last few years, with whom I have built beautiful memories that I will always cherish.

Ana Marques



*"Woof!"*

Nina





# Contents

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>   | <b>1</b>  |
| 1.1      | Context . . . . .   | 1         |
| 1.2      | Goals . . . . .   | 2         |
| 1.3      | Structure of the Document . . . . .                             | 2         |
| <b>2</b> | <b>Data Mining and Computer Vision</b>                          | <b>3</b>  |
| 2.1      | Data Mining . . . . .   | 3         |
| 2.1.1    | KDD . . . . .   | 4         |
| 2.1.2    | SEMMA . . . . .   | 5         |
| 2.1.3    | CRISP-DM . . . . .  | 6         |
| 2.2      | Computer Vision . . . . .                                       | 7         |
| 2.2.1    | Digital Image Processing . . . . .                              | 7         |
| 2.2.2    | Image Segmentation . . . . .                                    | 9         |
| 2.3      | Machine Learning . . . . .                                      | 10        |
| 2.3.1    | Decision Trees . . . . .  | 10        |
| 2.3.2    | Random Forest Classifier . . . . .                              | 10        |
| 2.3.3    | Support-Vector Machines . . . . .                               | 11        |
| 2.4      | Deep learning . . . . .   | 11        |
| 2.4.1    | Artificial Neural Networks . . . . .                            | 12        |
| 2.4.2    | Machine Learning, Deep Learning and Programming Tools . . . . . | 13        |
| 2.4.3    | DL Model Hyper-Parameters . . . . .                             | 14        |
| 2.4.4    | DL Model Evaluation . . . . .                                   | 14        |
| 2.5      | Summary . . . . .   | 15        |
| <b>3</b> | <b>Methodology</b>  | <b>17</b> |
| 3.1      | Problem Description . . . . .                                   | 17        |
| 3.2      | Setup description . . . . .                                     | 17        |
| 3.3      | Dataset Technical Parameters . . . . .                          | 17        |
| 3.4      | Deep Learning Algorithm . . . . .                               | 27        |
| 3.5      | Experimental Planning . . . . .                                 | 27        |
| 3.5.1    | Ideal Results . . . . .   | 27        |
| 3.5.2    | First Experiment and Base Values . . . . .                      | 27        |
| 3.5.3    | Second Experiment and Variations . . . . .                      | 28        |
| 3.6      | Summary . . . . .   | 28        |
| <b>4</b> | <b>Experiments</b>  | <b>29</b> |
| 4.1      | Conclusions . . . . .   | 32        |

|                                      |           |
|--------------------------------------|-----------|
| <b>5 Conclusions and Future Work</b> | <b>33</b> |
| <b>References</b>                    | <b>35</b> |

# List of Figures

|      |  |    |
|------|--|----|
| 2.1  | An Overview of the Steps That Compose the KDD Process . . . . .    | 5  |
| 2.2  | SEMMA steps . . . . .  | 5  |
| 2.3  | Partitioning polygons[23] . . . . .                                | 8  |
| 2.4  | DIP levels and how they relate to computer vision[23] . . . . .    | 9  |
| 3.1  | Directory tree of the dataset . . . . .                            | 18 |
| 3.2  | Example of the two views in images of a healthy patient . . . . .  | 18 |
| 3.3  | Atelectasis pathology example . . . . .                            | 19 |
| 3.4  | Cardiomegaly pathology example . . . . .                           | 20 |
| 3.5  | Consolidation example . . . . .                                    | 20 |
| 3.6  | Edema pathology example . . . . .                                  | 21 |
| 3.7  | Enlarged cardiomediastinum example . . . . .                       | 21 |
| 3.8  | Fracture example . . . . .   | 22 |
| 3.9  | Lung lesion example . . . . .                                      | 22 |
| 3.10 | Lung opacity example . . . . .                                     | 22 |
| 3.11 | Pleural effusion example . . . . .                                 | 23 |
| 3.12 | Pleural other example . . . . .                                    | 23 |
| 3.13 | Pneumonia example . . . . .  | 24 |
| 3.14 | Pneumothorax example . . . . .                                     | 24 |
| 3.15 | Distribution of the number of samples for each pathology . . . . . | 26 |





# List of Tables

|      |   |    |
|------|---|----|
| 3.1  | Class labels (pathologies) and attribute names . . . . .  | 19 |
| 3.2  | Number of cases with Pathology overlap . . . . .  | 24 |
| 3.3  | Dataset general dimensions . . . . .  | 25 |
| 3.4  | Number of samples of each pathology . . . . .   | 25 |
| 3.5  | Number of samples of individual attributes for each pathology . . . . .   | 26 |
| 3.6  | Values used for the Hyper Parameters . . . . .  | 27 |
| 4.1  | Average and standard deviation results of the runs of each pathology were evaluated in their accuracy, area under the ROC curve, and F1 score. The last column is the time it took the model to complete running . . . . .              | 30 |
| 4.2  | Average and standard deviation results of the runs of each attribute in each pathology, evaluated in their accuracy, area under the ROC curve, and F1 score. The last column is the time it took the model to complete running. . . . . | 31 |
| 5.1  | Atelectasis . . . . .   | 39 |
| 5.2  | Cardiomegaly . . . . .  | 39 |
| 5.3  | Consolidation . . . . .   | 39 |
| 5.4  | Edema . . . . .   | 40 |
| 5.5  | Enlarged Cardiomediatinum . . . . .   | 40 |
| 5.6  | Fracture . . . . .  | 40 |
| 5.7  | Lung Lesion . . . . .   | 41 |
| 5.8  | Lung Opacity . . . . .  | 41 |
| 5.9  | Pleural Effusion . . . . .  | 41 |
| 5.10 | Pleural Other . . . . .   | 42 |
| 5.11 | Pneumonia . . . . .   | 42 |
| 5.12 | Pneumothorax . . . . .  | 42 |
| 5.13 | Atelectasis . . . . .   | 44 |
| 5.14 | Cardiomegaly . . . . .  | 45 |
| 5.15 | Edema . . . . .   | 45 |
| 5.16 | Lung Lesion . . . . .   | 46 |
| 5.17 | Pleural Other . . . . .   | 47 |
| 5.18 | Pneumonia . . . . .   | 48 |



# Abbreviations

|          |   |
|----------|---|
| ADT      | Abstract Data Type                                      |
| AI       | Artificial Intelligence                                 |
| ANDF     | Architecture-Neutral Distribution Format                |
| ANN      | Artificial Neural Networks                              |
| AP       | Anterior-Posterior                                      |
| API      | Application Programming Interface                       |
| AUC      | Area Under ROC Curve                                    |
| API      | Application Programming Interface                       |
| BMU      | Best Matching Unit                                      |
| CAD      | Computer-Aided Design                                   |
| CASE     | Computer-Aided Software Engineering                     |
| CNN      | Convolutional Neural Networks                           |
| CORBA    | Common Object Request Broker Architecture               |
| CPU      | Central Processing Unit                                 |
| CRISP-DM | Cross-Industry Standard Process for Data Mining         |
| CSV      | Comma-Separated Values                                  |
| DIP      | Digital Image Processing                                |
| DL       | Deep Learning   |
| DM       | Data Mining   |
| GAN      | Generative Adversarial Networks                         |
| GPU      | Graphics Processing Unit                                |
| HSB      | Hue, Saturation and Brightness                          |
| HSI      | Hue, Saturation and Intensity                           |
| HSV      | Hue, Saturation, and Value                              |
| JSON     | JavaScript Object Notation                              |
| KDD      | Knowledge Discovery in Databases                        |
| K-NN     | K-Nearest Neighbour                                     |
| LSTM     | Long Short Term Memory Networks                         |
| ML       | Machine Learning  |
| MLP      | Multilayer Perceptrons                                  |
| MM       | Mathematical Morphology                                 |
| PA       | Posterior-Anterior                                      |
| RBFN     | Radial Basis Function Networks                          |
| ReLU     | Rectified Linear Units                                  |
| RGB      | Red, Green, and Blue                                    |
| ROC      | Receiver Operating Characteristic                       |
| RNN      | Recurring Neural Networks                               |
| SAS      | Statistical Analysis System                             |
| SEMMA    | Sampling, Exploring, Modifying, Modeling, and Assessing |
| SOM      | Self Organizing Maps                                    |
| SVM      | Support-Vector Machines                                 |
| UNCOL    | UNiversal Compiler-Oriented Language                    |



# Chapter 1

## Introduction

Lung disease identification through X-ray imaging is still an exclusively manual procedure. This means it is prone to human error, such as distraction, inexperience, or fatigue from the radiologist. A tool that could work as a support to the medical professionals, giving them a "second opinion," pointing to details that could have been missed, could lead to a significant increase in diagnosis success. Because it would help reduce uncertainty, it could also accelerate the diagnosis process and help initiate the treatment of the patient sooner.

Deep learning models have been proven a powerful tool for medical data analytics. They can identify some implicit features that cannot be easily described. This capacity will be used in this thesis, in which the model must learn to distinguish several lung pathologies based on the features found in the dataset images. The results will be analyzed in the final stage of the project.

### 1.1 Context

Chest X-ray medical radiographies are a globally adopted tool for medical imaging examination [20]. They are cost-effective compared to other diagnostic tools and hold critical information regarding various diseases. However, in these images, there is often an overlap between the pathology and other lung irregularities. Data mining techniques have been proven helpful in similar medical image classification studies, such as association rule mining for image categorization and neural networks. This work will use convolutional neural networks (CNN), a deep learning algorithm that is very effective for image classification problems [9].

A tool that could automatically identify diseases in X-ray images would be handy for a medical professional, so a positive result could indicate that it is a genuine possibility in the future and a relevant advance in diagnosis technology.

The automation of X-ray image analysis is a complex process. This process involves processing the images, which can be unclear and contain different medical conditions overlapping each other; building deep learning algorithms to correlate the samples with the previously assigned pathologies; and ending up with a model capable of detecting the pathologies autonomously.

## 1.2 Goals

This dissertation aims to provide some insight into the performance of deep learning models when seeking to identify diseases in chest X-ray images and to hypothesize the motives behind possible difficulties in this task.

The dataset used on this project is *CheXpert*. It has appropriate dimensions and consists of almost 230 000 samples belonging to approximately 65 000 patients, with relevant labels and 14 observations corresponding to pathologies or the absence of disease [20].

## 1.3 Structure of the Document

This report contains four additional chapters. In Chapter 2, we present an overview of the relevant context and state-of-the-art of the necessary tools and concepts for this thesis. Next, in Chapter 3, we describe the methodology adopted for the set of experiments. In Chapter 4, we present and interpret the experimental results. Finally, the conclusions and future work are described in Chapter 5.

## Chapter 2

# Data Mining and Computer Vision

The technological development over the last few years allowed an increased computational/processing power. As the cost of manufacturing powerful computers decreased, it became possible to execute processes so far done manually, which was time-consuming and costly. This resulted in the emergence of new tools and possibilities, like the ability to process large amounts of data, store large amounts of high-resolution images, such as medical images, useful for science studies, or automatically extract data from different kinds of written resources. The areas of computer science in which these possibilities are advantageous are data mining, computer vision, and text mining.

In this chapter, there will be an overview of the basic concepts of data mining, computer vision, and text mining. There will be a description of the state of the art and the main objective of applying these tools to the matter.

### 2.1 Data Mining

*"Data mining is the process of discovering meaningful correlations, patterns and trends by sifting through large amounts of data stored in repositories. Data mining employs pattern recognition technologies, as well as statistical and mathematical techniques." (Gartner Group[17])*

Nowadays, there is an incredible amount of data in various platforms and fields. Data collection has suffered massive growth, as have computing power and storage capacity. The urgency to develop fresh tools and methods to back analysts in the process of obtaining valuable knowledge from this dramatic growth of digital data led to the development of the field of knowledge discovery in databases (KDD) and, consequentially, of data mining [15].

However, there is not enough human capacity to analyze all of this data, and there is no software capable of doing so without human input. In order to do this data mining correctly, there has to be an insight of the structures in which the software is supported, which are based on mathematics and statistics.

Data mining, when done correctly, can accomplish various tasks, such as the description of patterns or trends within data, estimation models, prediction of future events, classification of target variables based on different predictors, clustering of the data set, and association of attributes. The resulting DM models should aim for transparency in the sense that one can intuitively interpret.

### 2.1.1 KDD

*"Knowledge Discovery in Databases (KDD) is an automatic, exploratory analysis and modeling of large data repositories. KDD is the organized process of identifying valid, novel, useful, and understandable patterns from large and complex data sets."*

(A. Adhikari et al[6])

KDD and DM are frequently mentioned interchangeably. A. Adhikari et al. state

The KDD process involves various approaches to extracting knowledge from data, "[...] including how the data are stored and accessed, how algorithms can be scaled to massive data sets and still run efficiently, how results can be interpreted and visualized, and how the overall man-machine interaction can usefully be modeled and supported". (Fayyad et al [15]).

There are nine steps within this process, which can be skipped or repeated, iteratively:

1. Understand the application domain, the knowledge already present, and determine the objective of the KDD process from the client's point of view;
2. Create a target data set
3. Clean and preprocess data
4. Reduce and project data
5. Apply a particular DM method
6. Select the method(s) to be applied to find data patterns and choose the DM algorithm
7. Data mining
8. Interpret the mined patterns
9. Act on the discovered knowledge

The Figure 2.1 was extracted from Fayyad et al.'s "From Data Mining to Knowledge Discovery in Databases"[15] and illustrates the steps of the KDD process.



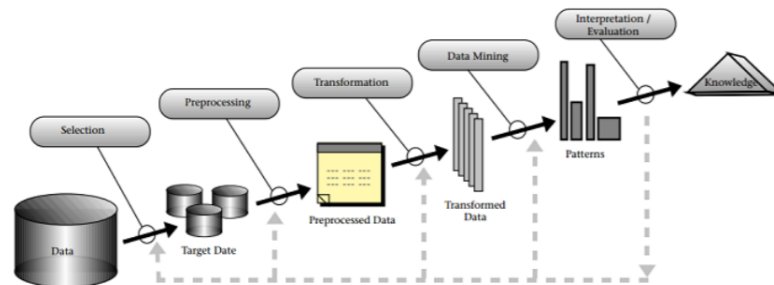


Figure 2.1: An Overview of the Steps That Compose the KDD Process

### 2.1.2 SEMMA

SEMMA (Sampling, Exploring, Modifying, Modeling, and Assessing) is a widely applicable data mining method. It consists of five steps, which can be adapted to the analysis by being removed or repeated in an iterative process. The following process flow diagram of Figure 2.2, illustrating this process, can be found on the SAS Help Center website:

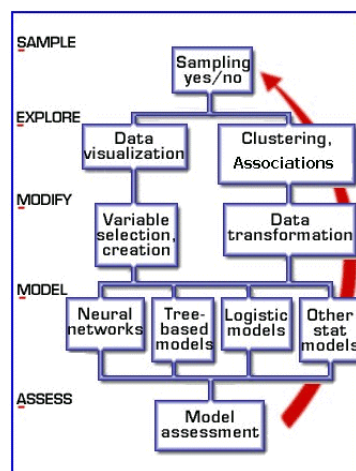


Figure 2.2: SEMMA steps

On the same page, there is also a short description of each step:

1. **Sample** the data by creating one or more data tables. The samples should be large enough to contain the significant information, yet small enough to process
2. **Explore** the data by searching for anticipated relationships, unanticipated trends, and anomalies in order to gain understanding and ideas
3. **Modify** the data by creating, selecting, and transforming the variables to focus the model selection process

4. **Model** the data by using the analytical tools to search for a combination of the data that reliably predicts a desired outcome
5. **Assess** the data by evaluating the usefulness and reliability of the findings from the data mining process [5].

### 2.1.3 CRISP-DM

CRISP-DM is a standard process to include data mining in problem-solving strategies. This standard is unrestricted to applications, industries, or tools. The concept is that there is a life cycle in each data mining project, which consists of six phases. The sequence between stages is not rigid, which means that the results obtained in a step will determine what the next step will be, which originates an iterative process [21]. The six phases are the following:

1. Business/Research understanding phase
  - (a) Enunciate the project objectives and requirements clearly
  - (b) Translate the objectives and restrictions into the formulation of a data mining problem definition
  - (c) Prepare a preliminary strategy for achieving these objectives.
2. Data understanding phase
  - (a) Collect the data
  - (b) Use exploratory data analysis to familiarize yourself with the data and discover initial insights
  - (c) Evaluate the quality of the data
  - (d) If desired, select interesting subsets that may contain actionable patterns.
3. Data preparation phase
  - (a) Prepare from the initial raw data the final data set that is to be used for all subsequent phases
  - (b) Select the cases and variables you want to analyze and that are appropriate for your analysis
  - (c) Perform transformations on certain variables, if needed
  - (d) Clean the raw data so that it is ready for the modeling tools.
4. Modeling phase
  - (a) Select and apply appropriate modeling techniques
  - (b) Calibrate model settings to optimize results
  - (c) Remember that often, several different techniques may be used for the same data mining problem
  - (d) If necessary, loop back to the data preparation phase to bring the form of the data into line with the specific requirements of a particular data mining technique.

### 5. Evaluation phase

- (a) Evaluate the one or more models delivered in the modeling phase for quality and effectiveness before deploying them for use in the field
- (b) Determine whether the model in fact achieves the objectives set for it in the first phase
- (c) Establish whether some important facet of the business or research problem has not been accounted for sufficiently
- (d) Come to a decision regarding use of the data mining results.

### 6. Deployment phase

- (a) Make use of the models created: Model creation does not signify the completion of a project
- (b) For businesses, the customer often carries out the deployment based on your model

## 2.2 Computer Vision

Computer Vision is a field that aspires to mimic biological vision and visually perceive the world [10]. It consists of various techniques to process, interpret and perceive images and real-life high-dimensional data to create digital information [13].

### 2.2.1 Digital Image Processing

A digital image contains a limited number of elements (*pixels*). It can be defined as a 2-dimensional function,  $f(x,y)$  where  $f$  is the intensity at the point of coordinates  $(x, y)$ . Image processing consists of manipulating these digital images to extract information from them. The fundamental steps in the processing of images involve image acquisition, enhancement and restoration, color processing, compression, morphological processing, segmentation, feature extraction, representation and description, and recognition [18].

Regarding the conversion from a continuous image to a digital one, the continuous image is sampled using a discrete partition of the continuous plane. These partitions can be constituted only by one of three regular polygon types: triangular, square, and hexagon. These are represented in Figure 2.3.

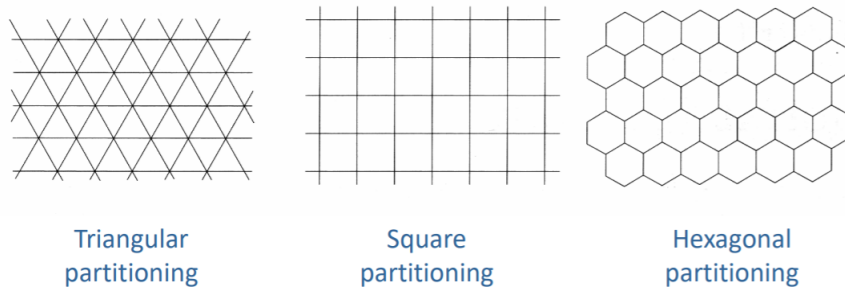


Figure 2.3: Partitioning polygons[23]

In this conversion process, there is also sampling and quantization. The first process will give us the number of samples evaluated as spatial resolution. The second one provides us with the number of intensity levels, then classified as intensity resolution.

Variations in spatial resolution will result in more or less pixelated images, depending on whether the number of samples is lower or higher. Variations in intensity resolution will affect the accuracy of the tones, so increasing the number of intensity levels will result in a more accurate image shading. These images are monochromatic. Monochromatic images, or 2D images, can be defined with the following equation:

$$F_M = \{f(i, j), i = 1, 2, \dots, N; j = 1, 2, \dots, M\} \quad (2.1)$$

where  $f(i, j)$  is the gray-level of a pixel with spatial coordinates  $(i, j)$ . Polychromatic images, or 3D images, on another hand, can be developed from that and defined with the following equation:

$$F_P = \{F_M(\text{Red}); F_M(\text{Green}); F_M(\text{Blue})\} \quad (2.2)$$

Regarding color images, these can have the following representations: RGB and HSV / HSI / HSB [23].

H – Hue

S – Saturation

V – Value

I – Intensity

B – Brightness

To go from digital image processing (DIP) to computer vision, a spectrum can be partitioned into low, mid, and high-level processes. Low-level processes include basic operations such as noise reduction, enhancement of features, an increase of contrast, edge enhancement, binarization or compensation of geometrical deformations, and both the input and output of the operation are images. These processes are often described as pre-processing, and other techniques include intensity transformations, linear filtering, and non-linear filtering. In mid-level processes, the input is also an image, but the output consists of attributes resulting from operations like object

recognition and segmentation. Finally, high-level processes take these attributes as input and return understanding as output and ideally perform cognitive functions associated with human vision, such as scene understanding and autonomous navigation [12]. Other techniques for processing images include intensity transformations, linear filtering, and non-linear filtering [7]. The way these concepts are related is represented in Figure 2.4.

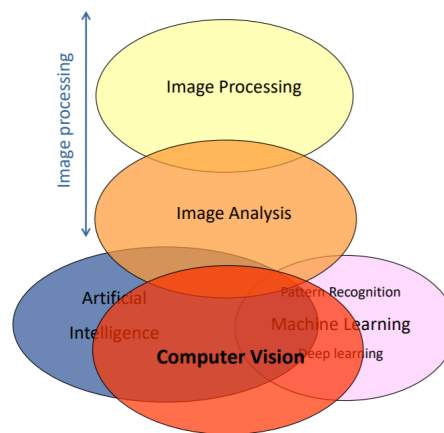


Figure 2.4: DIP levels and how they relate to computer vision[23]

## 2.2.2 Image Segmentation

Segmentation methods can be classified into two categories, according to pixel properties: feature-based and image-based segmentation. The first one includes thresholding or histogram-based methods and clustering methods. The latter includes region-based, edge-based, and point, line, edge, and corner detection methods.

There are issues within the images that complicate the process of analyzing them. These issues include non-uniform illumination and out-of-focus areas.

### 2.2.2.1 Thresholding

This technique turns a greyscale image into a binary image. The threshold value can be calculated manually or by applying specific methods.

### 2.2.2.2 Region Growing

This is an iterative, flexible approach for the sequential extraction of objects starting from an initial set of points (seeds). Seeds are obtained using a highly selective detection criterion. In each iteration, a pixel is aggregated if it belongs to the neighborhood of pixels that were previously detected or aggregated and if it verifies a less demanding condition, the aggregation criterion [25].

### 2.2.2.3 Feature Detection and Matching

These are essential components of many computer vision applications. The first kind of feature is noticing specific locations in the images, often called *keypoint features* (invariant features) or *interest points*. Other features are *corners* and *edges*. Edges can be grouped into longer curves and contours or straight lines segments. Feature matching consists of the comparison of features between two images. There is also feature tracking, which is finding the most similar feature from an image in another image [24].

### 2.2.2.4 Morphological Filtering

Morphological filters are local, non-linear operators resulting from the implementation of Mathematical Morphology concepts. Mathematical Morphology (MM) is a theory for analyzing spatial structures; it aims to explore objects' shape and form based on set theory, integral geometry, and lattice algebra. Morphological operations simplify images while quantifying and preserving the main shape characteristics of image objects.

## 2.3 Machine Learning

Machine Learning is an area of computational algorithms built to emulate human reasoning and apply it in future choice situations by gathering knowledge from the available data. This area has plenty of possible applications, such as computer vision and computational biology, both relevant to this thesis. Some machine learning tools and classification algorithms are decision trees, rule induction, support vector machine, and K-Nearest neighbor (K-NN) [34].

### 2.3.1 Decision Trees

Decision trees are very intuitive. They start with a decision regarding an attribute, an *internal node*, and its possible results are called *branches*. When a branch no longer splits, it becomes a decision, or *leaf*. They are computationally fast, easily interpretable, and can be converted directly to rules. There are **Classification Trees** and **Regression Trees**, but the former classifies individual classes, and the latter evaluates continuous values [35].

### 2.3.2 Random Forest Classifier

This is an algorithm used for classification and regression that obtains the outputs of a set of decision trees generated from random samples within the dataset.

The algorithm starts by picking random samples from the training set and making a decision tree, which can be displayed, for each sample, using an attribute selection indicator such as gain ratio, Gini index, or information gain. The output of each tree is a prediction result. This part is done for the number of trees chosen. Then, when the goal is classification, the number of predictions for each class is gathered, and the final prediction will be the one with the most "votes,"

which means that it is an essential feature for the classifier; it is also possible to get the probability of that feature besides the quality itself. In the case of regression, this final result is the average of all predictions.

This means that the algorithm is more robust and more accurate the more trees it comprises, but the slower it is in generating a result, considering that for each final prediction, all trees must generate their predictions. However, this does not mean it is not a good algorithm for large datasets. Because of the "divide and conquer" nature of the voting process, which cancels out the bias, Random Forests do not suffer from overfitting. They can handle datasets with a large number of features and entries. Also, when missing values exist in the dataset, their proximity-weighted average can be calculated for classification, or a median value can be used in the case of regression. A difficulty when using the Random Forest Classifier is regarding the interpretation, which is much more complex than a single decision tree. However, this can be facilitated by visualizing individual trees within it[28].

### 2.3.3 Support-Vector Machines

An algorithm appropriate for analyzing limited amounts of samples, such as in text classification, in two-group classification problems, although it can also be used for regressions. It consists of finding a hyperplane in a multi-dimensional space with as many dimensions as the number of features that divides the data points between the two specified groups with the maximum margin possible. This maximization is supported by a support vector, which comprises the data points closest to the hyperplane and influences its position and orientation. The calculation of this margin includes computing the hinge loss function with an added regularization parameter which balances maximization and loss. When the predicted value is the same as the actual value, the cost is 0; when the opposite happens, the cost is the calculated loss. After the definition of the loss function, partial derivatives are taken regarding the chosen weights to find the gradients. If the model makes an incorrect prediction, these gradients are updated in the regularization parameter along with the loss; if it makes a correct prediction, the loss is not included[16].

## 2.4 Deep learning

Deep learning is a type of machine learning. It is an approach to AI that tackles problems that humans solve intuitively. Because overlapping other learned notions support the human intuitive knowledge, a visual representation of it would look like a deep, layered image of concepts. With deep learning, a computer can use simple concepts to formulate more complex ones[19]. This means that, based on the characteristics of the human brain, artificial neural networks (ANN) will be capable of processing extensive amounts of information.

### 2.4.1 Artificial Neural Networks

These neural networks consist of stacks of nodes, which are artificial neurons. These nodes are stacked into three layers: an input layer, an output layer, and one or more hidden layers in between. First, the information is carried to a node as input. Then, there are calculations made inside this node, which include operations with random weights and added bias. The result of these calculations is applied activation functions, such as Rectified Linear Units (ReLUs), sigmoid functions, or tanh, which determines the nodes that will be fired and those that won't.

Deep Learning algorithms learn from examples to train a machine. During training, these algorithms find features, group objects, and valuable data patterns in the input data. To achieve this multi-level process, they need the support of artificial neural networks. There are a few types of ANNs, and they can be distinguished on a fundamental level as recurring (forming a cycle) or feedforward (no cycle formed)[11].

#### 2.4.1.1 Recurring Neural Networks

The structure of these networks is recurring because the outputs can be fed immediately back as inputs, forming directed cycles. Some examples of RNNs' applications are natural-language processing, handwriting recognition, and image captioning.

#### 2.4.1.2 Multilayer Perceptrons

This is a feedforward neural network type, meaning that the connections between the neurons do not form a cycle. The multiple layers of these networks are constituted by perceptrons (binary classifiers) and are fully connected in a graph, forcing a direction in the signal. MLPs can be used as a base for image or speech recognition software.

#### 2.4.1.3 Long Short Term Memory Networks

This is a type of Recurring Neural Network, and it has a chain-like structure in which four layers will interact and communicate. Due to their ability to retain past information over time, LSTMs can be helpful for pharmaceutical development and speech recognition.

#### 2.4.1.4 Generative Adversial Networks

These networks contain two key components: the generator and the discriminator. The generator receives input from the network and produces fake data based on it. Then, the discriminator learns to distinguish between the artificial and the actual data, and the cycle restarts. GANs are especially useful in generating realistic images and remastering low-resolution textures into high-resolution ones.



#### 2.4.1.5 Self Organizing Maps

SOMs aims to solve the human limitation regarding visualizing high dimensional data by reducing its dimensions using self-organizing Artificial Neural Networks. First, the weights are loaded, and a vector from the training data is chosen randomly. Secondly, there's an analysis of each node to find the Best Matching Unit (BMU). This step is repeated in any number of iterations. The SOM sees the BMU's neighborhood as a sample vector and assigns a winning weight to that sample vector. The weight of each node and how much it learns will vary inversely to the distance from the BMU. With each iteration, the BMU's neighborhood decreases.

#### 2.4.1.6 Radial Basis Function Networks

This type of feedforward neural network uses Radial Basis Functions in the hidden layer as activation functions. When the network receives inputs, it will compare them to examples in the training set and find their weighted sum. Then, the input vector goes to the hidden layers containing the transfer functions. Their output is inversely proportional to the distance from the node's center. Finally, the output layer is built of several nodes, one for each data class. RBFNs are helpful for classification, time-series predictions, and regressions.

#### 2.4.1.7 Convolutional Neural Networks

In this thesis, the Deep Learning algorithm known as *Convolutional Neural Network* (also known as *CNN* or *ConvNet*) will be of great importance. Inspired by the visual cortex structure, in which visual fields overlap and comprise the entire visible area, it takes an input image and initial weights and learns to differentiate them. This network takes less pre-processing than other classification algorithms and can retain filters that, in different methods, must be input by hand. The convolution operation consists of the application of several filters, and it happens in the Convolution Layer. Then, the elements pass through a Rectified Linear Unit (ReLU), resulting in a Rectified Feature Map. This feature map is pooled in the Pooling Layer, which means that it is downsampled to two-dimensional arrays, and then flattened into a continuous linear vector (flattened matrix), which allows the algorithm to classify the images or objects, resulting in a final Fully Connected Layer. This algorithm is capable of reducing the images while keeping critical features to guarantee results, which allows for great scalability[11].

### 2.4.2 Machine Learning, Deep Learning and Programming Tools

Some tools to be explored in the following steps of this thesis will be TensorFlow, Keras, and Jupyter Notebook.

TensorFlow is a machine learning platform that includes tools to process and load data, utilize pre-trained models, create new models, deploy them on different interfaces and automate and track model training[1].

Keras is a Python Deep Learning application programming interface (API) built on top of TensorFlow 2 . It is known for its flexibility across platforms and its beginner-friendly nature. It is also reliable enough to be used by many scientific organizations worldwide[2].

Jupyter Notebook is an open-source, interactive development environment. It is a web application, has a modular design, supports more than 40 programming languages, allows the production of interactive outputs and the usage of big data tools, and the notebooks can be shared across the web. These notebooks are in JSON-based document format and can store the code outputs and user sessions along with the code. It can be deployed with Anaconda Navigator, a graphical user interface with which a user can create separate virtual environments. These environments allow the same computer to run code on different versions of different python packages, providing great flexibility in program development[3].

### 2.4.3 DL Model Hyper-Parameters

The most important values to be defined when building a model are the data size, the number of epochs, batch size, and the learning rate. The data size is the number of samples in the dataset. The number of epochs is the number of times the entire dataset is run through the model[38]. Generally, the larger the number of epochs, the better. The batch size is how many samples are processed at the same time[27]. It is equivalent to a step, which means that a larger batch size equals making more significant jumps between groups of samples, increasing the training process's speed. The learning rate is a parameter that defines the step size in minimizing a loss function to avoid both slow learning and system divergence[39].

### 2.4.4 DL Model Evaluation

The models trained in this work will be evaluated regarding their Logarithmic Loss, Accuracy, Area Under ROC Curve, and F1 Score. They are binary classifications and have values between 0 and 1[33]. Then, the results' standard deviation and average will also be calculated.

#### 2.4.4.1 Logarithmic Loss

Log Loss is a classification metric based on probabilities that increases when the model's prediction diverges from the correct value.

#### 2.4.4.2 Accuracy

Accuracy is the proportion between the number of correct predictions and the total number of input samples.

$$Accuracy = \frac{CorrectPredictions}{TotalPredictions} \quad (2.3)$$

This metric should only be used in balanced data because if the number of samples belonging to each class differs, this ratio will be misleading.

### 2.4.4.3 Area Under ROC Curve

The Area Under Curve is a metric used in binary classification, which is the case in the models of this thesis. This curve is the result of the plot where  $x$  is the False Positive Rate, and  $y$  is the True Positive Rate. The False Positive Rate is the ratio between negative data points wrongly classified as positive and all the data points as negative.

$$FalsePositiveRate = \frac{FalsePositive}{TrueNegative + FalsePositive} \quad (2.4)$$

The True Positive Rate is the ratio between correctly classified positive data points and all positive data points. It is also known as Sensitivity or Recall.

$$TruePositiveRate = \frac{TruePositive}{FalseNegative + TruePositive} \quad (2.5)$$

### 2.4.4.4 F1 Score

The F1 Score evaluates a classifier by combining Precision and Recall metrics and computing their harmonic mean.

$$F1 = 2 \times \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (2.6)$$

Precision is the ratio between the number of True Positives and the total positive predictions of the classifier. Recall or Sensitivity has already been described.

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \quad (2.7)$$

## 2.5 Summary

This chapter contextualized the concepts needed for this work, the tools deemed appropriate to complete it, and the methods for evaluating the algorithm's performance. After this research, it was possible to move on to the next step, planning the experiment.



## Chapter 3

# Methodology

### 3.1 Problem Description

This work tackles the problem of classifying chest X-ray images according to the pathologies they include. From a software point of view, the challenge is for the program to analyze the radiography, identifying a set of features that allow for the intended differentiation between diseases. Contemporary machine learning techniques successfully have been proven to complete similar problems as the one in this thesis, so they were deemed more appropriate than the traditional image analysis and classification methods.

### 3.2 Setup description

This experiment was conducted with the support of Jupyter Notebook embedded in the Anaconda software, in a laptop with only a CPU in the early stages, and then on machines with GPU. The samples are from the CheXpert dataset and were stored on an external hard drive. This dataset was obtained through an email link after registration on the Stanford ML Group website. This email offers two download options according to the images' resolutions. The low-resolution zip file has 11GB, and the high-resolution zip file has 439GB. For this thesis, the low-resolution set was deemed the most appropriate, regarding the time and available setup limitations.

### 3.3 Dataset Technical Parameters

The dataset consists of JPG (JPEG) images and two CSV (Comma-Separated Values) format files. The images are chest X-Rays, and the CSV file contains information about every picture. The directories are organized as seen in Figure 3.1.

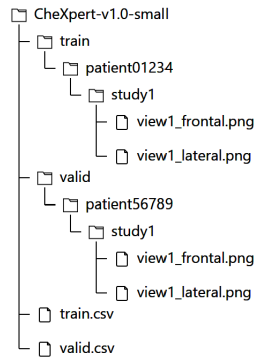
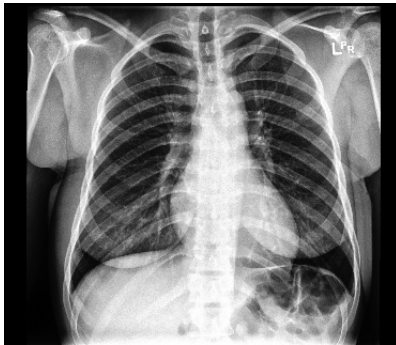
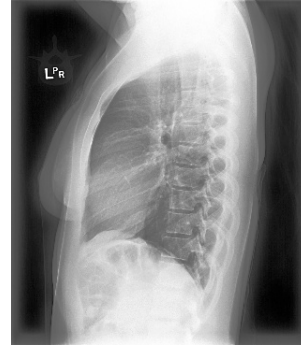


Figure 3.1: Directory tree of the dataset

This means that the dataset is divided into two sub-datasets with different dimensions but the same structure. There is a folder for each patient, which contains several studies, each with its directory. The images are inside these directories. The images may have two different angles, frontal and lateral, as seen in Figure 3.2.



(a) Frontal



(b) Lateral

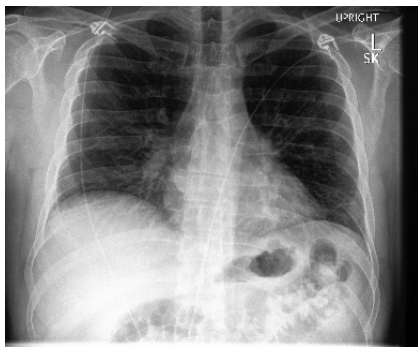
Figure 3.2: Example of the two views in images of a healthy patient

As mentioned before, this dataset is available in two CSV format files. Each file has the following information on the dataset's attributes: the first column contains the path to each image. The following columns contain data about the patient's sex and age, if it is a frontal or lateral view, if it's an AP (Anterior-Posterior, taken facing the back of the patient) or PA (Posterior-Anterior, taken facing the front of the patient, usually when they need to be lying down) projection, if nothing was detected in the X-Ray, and if the patient had any support devices. The remaining twelve columns have the value 1 or 0, signaling the existence or absence of the related pathology, respectively. Table 3.1 shows the existing pathologies. A single example from a patient can have more than one pathology identified.

Table 3.1: Class labels (pathologies) and attribute names

| Pathology Classes        | Attribute Names |
|--------------------------|-----------------|
| Atelectasis              | Path            |
| Cardiomegaly             | Sex             |
| Consolidation            | Age             |
| Edema                    | Frontal         |
| Enlarged Cardiomeastinum | Lateral         |
| Fracture                 | AP              |
| Lung Lesion              | PA              |
| Lung Opacity             | Support Devices |
| Pleural Effusion         |                 |
| Pleural Other            |                 |
| Pneumonia                |                 |
| Pneumothorax             |                 |
| No Finding               |                 |

Atelectasis is a condition in which a part of the lung has collapsed and does not hold air, while the rest of the lung looks normal; it is commonly caused by mucus blocking the airways [31]. An example of the aspect of this condition can be seen in Figure 3.3.



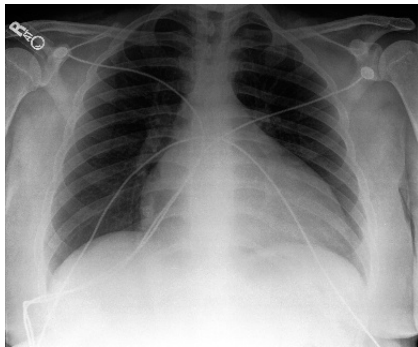
(a) Frontal



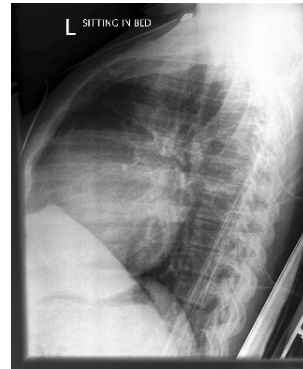
(b) Lateral

Figure 3.3: Atelectasis pathology example

Cardiomegaly is the increase in the size of the diameter of the heart; it is a symptom of another condition, which commonly is myocardial infarction or ischemia, but it can be caused by autoimmune diseases, toxins, and pregnancy, among others [8]. An example of this condition is in Figure 3.4.



(a) Frontal



(b) Lateral

Figure 3.4: Cardiomegaly pathology example

Consolidation is also a symptom of other diseases, such as pneumonia; in an X-ray, this condition makes the lung considerably blurry and conceals the details such as blood vessels, which indicates that a major portion of the air in the lungs was replaced by another fluid, such as blood [29]. Figure 3.5 shows an example of this condition.



(a) Frontal



(b) Lateral

Figure 3.5: Consolidation example

Edema is a condition in which there is a build-up of extravascular fluid in the lungs, and it can be caused by heart failure or injury [26]. An example of the condition is in Figure 3.6.



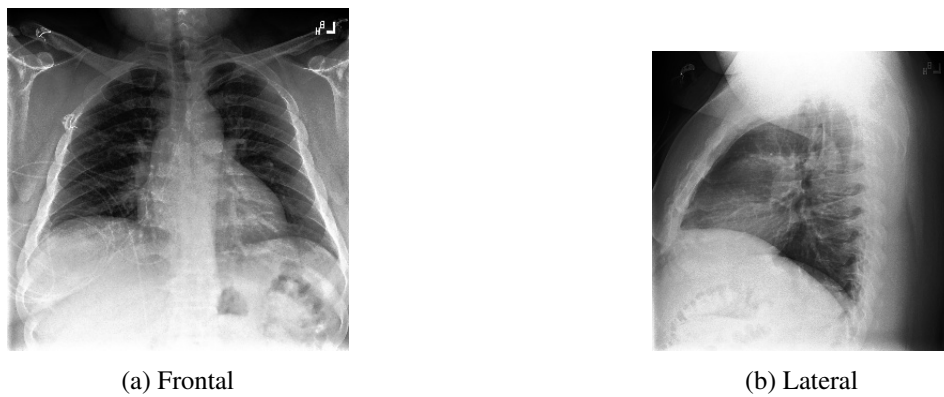


Figure 3.6: Edema pathology example

An enlarged cardiomeastinum means that the silhouette of the heart is larger than it should be; this condition can be seen both from a frontal and lateral X-ray, and it usually means that the patient suffers from cardiomegaly, even though this is not the only cause [4]. Example of this condition in Figure 3.7.

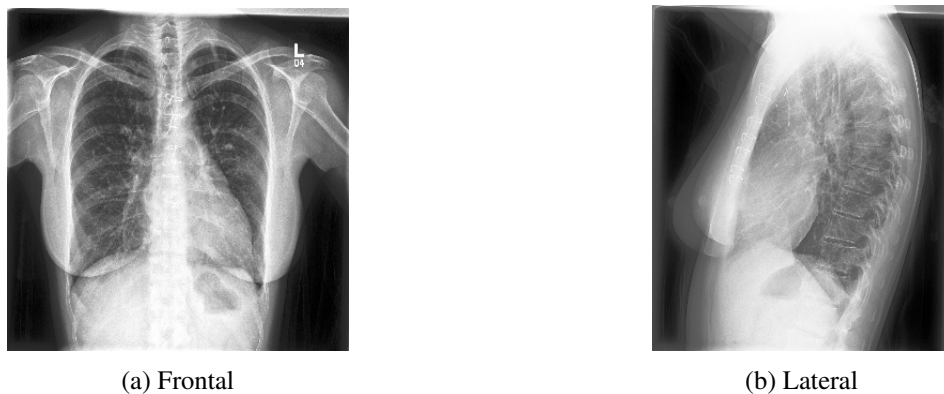


Figure 3.7: Enlarged cardiomeastinum example

A fracture happens in the ribs and leads to a series of complications. First, the pain the patient suffers when the rib is moving leads to shallow breathing, which predisposes to atelectasis. Secondly, an impact with enough force to cause a rib fracture will also cause a contusion in the lung, which involves bleeding and death of the tissues. Finally, if the force is also enough to cause a part of the chest wall to be destabilized, known as a *flail segment*. That destabilization provokes pressure in the opposite way of the normal intrapleural pressure, which means that this section will move inversely to the rest of the lung, shifting inward when the rib cage expands. This impedes the lung underneath the flail from expanding completely and thus leads to shallow breathing and deficient oxygenation of the tissues [22]. Figure 3.8 shows an example of a fracture.

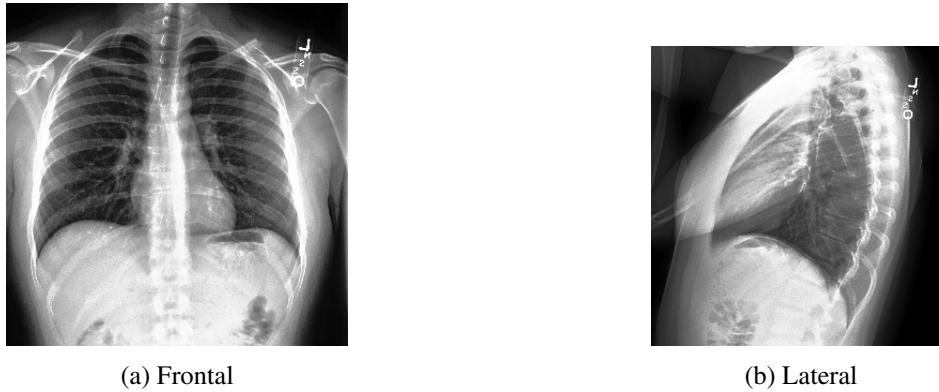


Figure 3.8: Fracture example

Lung lesions are usually tumors that can be benign or malignant, and they are often referred to as nodules [36]. These nodules can be seen in Figure 3.9.

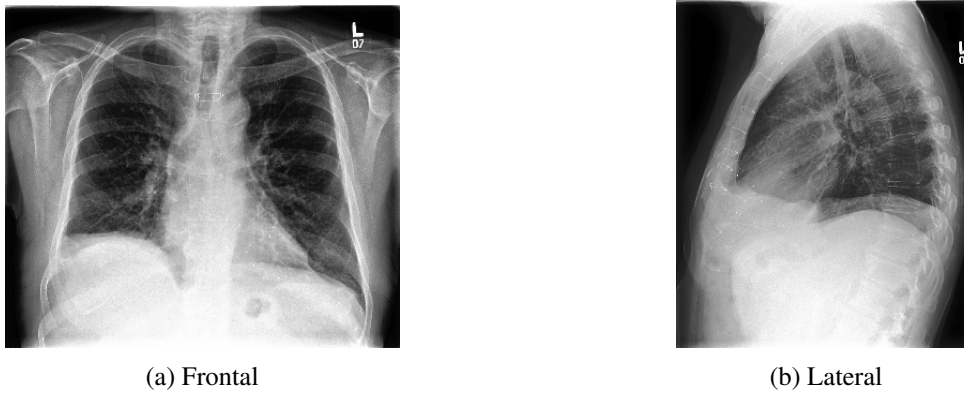


Figure 3.9: Lung lesion example

Lung opacity, often referred to as ground-glass opacity, happens when the portion of the lungs that contains air becomes partially filled with a denser substance, creating opacity in the image. This opacity, however, is not severe enough to conceal pulmonary vessels, at which point it would become consolidation [14]. This opacity is represented in Figure 3.10.

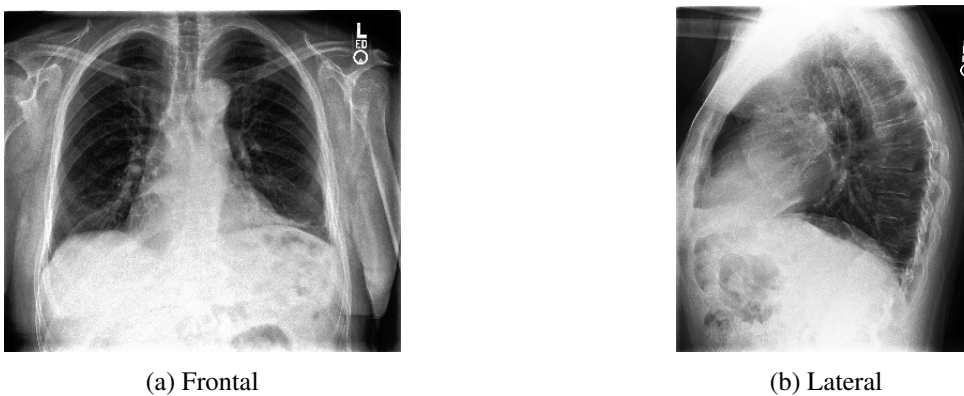


Figure 3.10: Lung opacity example

Pleural effusion is a condition in which there is an abnormally high amount of fluid in the pleura, which is the membrane that connects the lungs to the chest walls [32]. Figure 3.11 is an example of this.

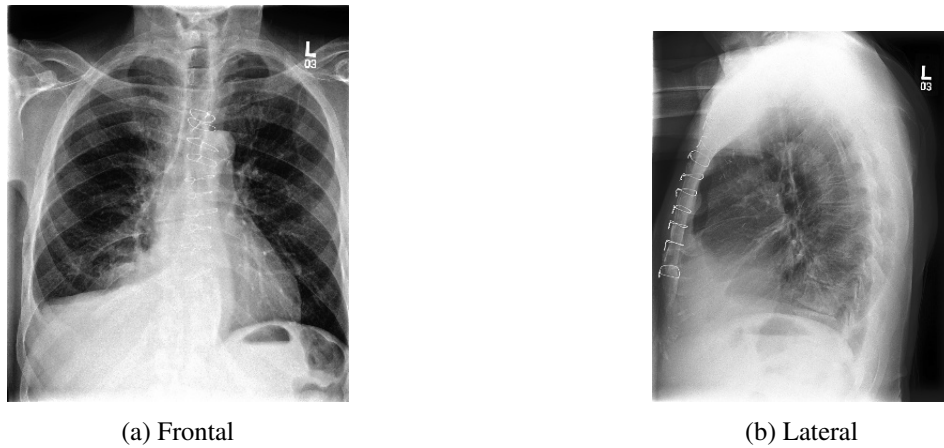


Figure 3.11: Pleural effusion example

Pleural other refers to other conditions regarding the pleural space. This is the most heterogeneous label. An example of these other conditions is represented in Figure 3.12.

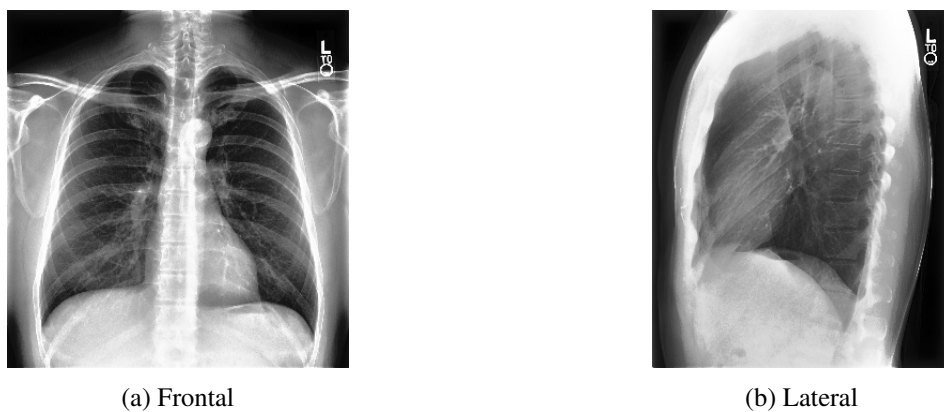
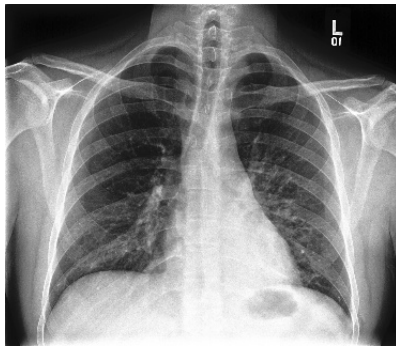
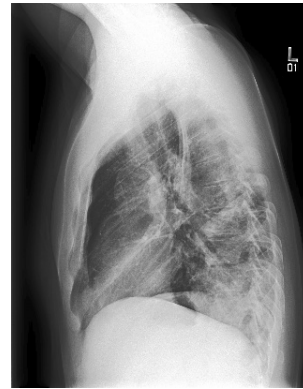


Figure 3.12: Pleural other example

Pneumonia is an infection that fills the lung's alveoli with fluid. In a chest X-ray, it is usually associated with the presence of pleural effusion [37]. Figure 3.13 is an example of pneumonia.



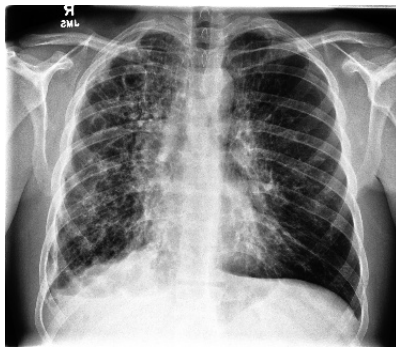
(a) Frontal



(b) Lateral

Figure 3.13: Pneumonia example

Pneumothorax consists of an accumulation of air in the pleural space, where the air is not supposed to enter. Pneumothorax can happen spontaneously, with or without another underlying lung disease, or during menstruation. Or due to trauma caused by other medical procedures or an impact that can be blunt or piercing [30]. This condition is represented in the Figure 3.14.



(a) Frontal



(b) Lateral

Figure 3.14: Pneumothorax example

These labels are not exclusive, meaning a patient can be diagnosed with more than one pathology. That is the case for more than a third of the samples as can be seen in Table 3.2.

Table 3.2: Number of cases with Pathology overlap

|                        |        |       |
|------------------------|--------|-------|
| <b>&gt;1 pathology</b> | 88572  | 39.6% |
| <b>≤ 1 pathologies</b> | 135077 | 60.4% |
| <b>Total</b>           | 223649 | 100%  |

The total dimensions of this dataset are described in Table 3.3.

Table 3.3: Dataset general dimensions

|              | Rows    | Columns | Patients |
|--------------|---------|---------|----------|
| <b>Train</b> | 223 414 | 19      | 64 540   |
| <b>Valid</b> | 235     | 19      | 182      |
| <b>Total</b> | 223 649 | 19      | 64 722   |

The dimensions of each pathology are not uniform. The number of samples per pathology can be seen in Table 3.4. These variations can be visually compared with the help of the histogram in Figure 3.15.

Table 3.4: Number of samples of each pathology

| Pathology                 | Train Samples | Valid Samples | Total Samples |
|---------------------------|---------------|---------------|---------------|
| Atelectasis               | 2651          | 467           | 3118          |
| Cardiomegaly              | 2269          | 400           | 2669          |
| Consolidation             | 1065          | 187           | 1252          |
| Edema                     | 4422          | 780           | 5202          |
| Enlarged Cardiomedastinum | 1298          | 229           | 1527          |
| Fracture                  | 1690          | 298           | 1988          |
| Lung Lesion               | 877           | 154           | 1031          |
| Lung Opacity              | 4246          | 747           | 4993          |
| Pleural Effusion          | 5181          | 913           | 6094          |
| Pleural Other             | 336           | 59            | 395           |
| Pneumonia                 | 459           | 80            | 539           |
| Pneumothorax              | 3798          | 670           | 4468          |
| No Finding                | 18963         | 3343          | 22306         |

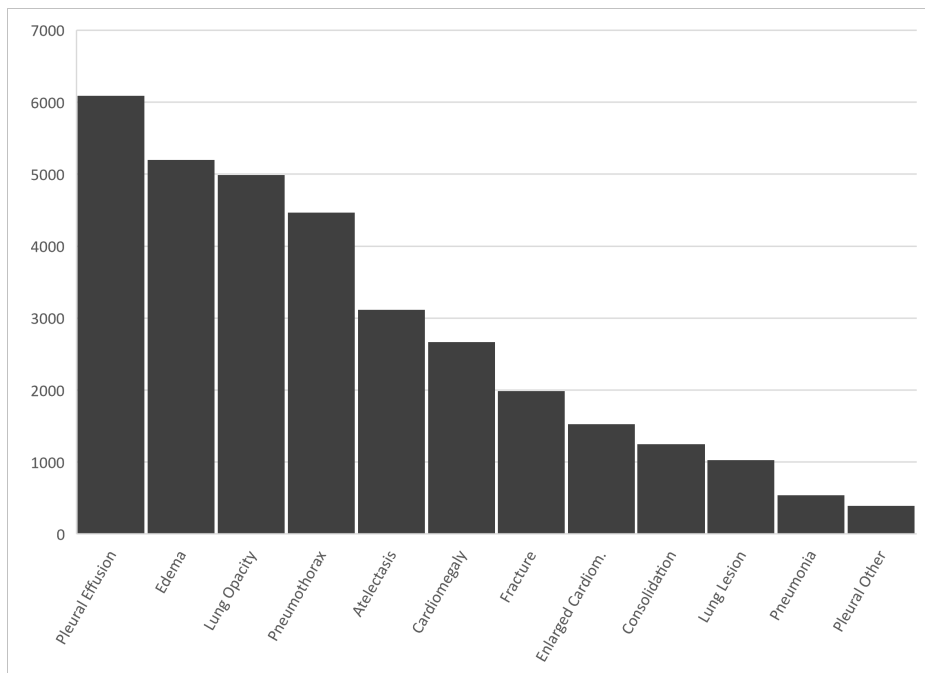


Figure 3.15: Distribution of the number of samples for each pathology

There is also, for each pathology, a varying number of samples for each non-pathology attribute. These attributes are *view*, *projection* and *sex*. These values are displayed in the Table 3.5.

Table 3.5: Number of samples of individual attributes for each pathology

| Pathology         | Frontal View | Lateral View | Anterior-Posterior Proj. | Posterior-Anterior Proj. | Male | Female |
|-------------------|--------------|--------------|--------------------------|--------------------------|------|--------|
| Atelectasis       | 2301         | 387          | 2688                     | 432                      | 1899 | 1221   |
| Cardiomegaly      | 1443         | 573          | 2016                     | 657                      | 1551 | 1119   |
| Consolidation     | 978          | 129          | 1107                     | 147                      | 717  | 537    |
| Edema             | 4761         | 189          | 4950                     | 255                      | 3051 | 2151   |
| Enlarged Cardiom. | 918          | 294          | 1212                     | 318                      | 969  | 558    |
| Fracture          | 1029         | 486          | 1515                     | 477                      | 1335 | 654    |
| Lung Lesion       | 492          | 288          | 777                      | 255                      | 618  | 417    |
| Lung Opacity      | 3446         | 750          | 4197                     | 798                      | 2931 | 2064   |
| Pleural Effusion  | 3495         | 1278         | 4773                     | 1323                     | 3666 | 2430   |
| pleural Other     | 108          | 150          | 258                      | 141                      | 252  | 144    |
| Pneumonia         | 192          | 171          | 363                      | 180                      | 327  | 216    |
| Pneumothorax      | 3531         | 504          | 4035                     | 435                      | 2703 | 1767   |

## 3.4 Deep Learning Algorithm

The Deep learning algorithm chosen for the models was the Convolutional Neural Networks. It is the most adequate for this work, considering its success in image analysis without needing an external attribute classifier such as decision trees, random forests, or support-vector machines.

## 3.5 Experimental Planning

The experimental part of this work aims to train one DenseNet model for each label, consisting of a binary "Pathology vs. No Finding" that will serve as a baseline. The following step is to experiment with parameter variations and analyze the results. For every case study, different hyperparameter values will be applied to the models to try and find the best results. The sample size will differ depending on the number of examples per sub-dataset. These hyper-parameters and their values are contained in the Table 3.6.

Table 3.6: Values used for the Hyper Parameters

| N <sup>er</sup> of Epochs | Batch Size | Learning Rate |
|---------------------------|------------|---------------|
| 20                        | 20         | 1E-4          |
| 40                        | 24         | 1E-5          |
| 50                        | 40         | 1E-6          |
| 60                        |            |               |

### 3.5.1 Ideal Results

According to the CheXpert website, the best results in this multi-labeling classification have an AUC of 0.930. Although this is not a fair comparison, considering those results were obtained with a multi-label classification and these models will be binary, this value will be taken into account as a reference, and the results obtained will be compared to it.

### 3.5.2 First Experiment and Base Values

The first experiment will be an algorithm that trains the models to differentiate each pathology in a binary fashion, "Pathology vs. No Finding." The original dataset will be filtered to include only the pathology in question. This means that samples with pathology overlap will not be considered, which is around 39.6% of the entire CheXpert dataset, leaving 135 077 relevant examples, according to Table 3.2. Then, because the "No Finding" sub-dataset is larger in samples than any other pathology sub-dataset, it will be downsampled to each of the dimensions of the respective disease sub-dataset to have balanced data, which means an equal number of samples. The number of samples per disease sub-dataset can be seen in Table 3.4. Each disease sub-dataset was divided into "train" and "test" datasets. These were scrambled into three groups for variability to have

three different main runs. The results of these models will work as a baseline for the variations implemented in the following experiment.

### 3.5.3 Second Experiment and Variations

In the following experiments, single attributes within each disease will be isolated. The attributes chosen are the following:

- Projection: AP (Anterior-Posterior) vs PA (Posterior-Anterior)
- View: Frontal vs Lateral
- Sex: Female vs Male

The goal is to analyze if the model has varying results depending on an attribute (Projection, View, Sex) and if it is possible to obtain satisfactory results with less data. Since these attributes are inherently binary, one of the two possibilities has better results.

This means there are six subsets for every pathology, divided into three groups of two. These single-attribute subsets will then be randomly distributed into three subsets each to increase the variability of the models. The attributes sub-sets are presented in the Table 3.5. In total, there will be 5832 models. The total number of models is calculated below.

There are 12 possible pathologies, six individual attributes per pathology, and three sub-datasets per attribute:

$$12_{pathologies} \times 6_{attributes} \times 3_{sub-datasets/attribute} = 216 \quad (3.1)$$

For each one of these models, the three hyper-parameters "number of epochs," "learning rate," and "batch size" are cycled and combined among three values each:

$$216 \times 3_{epoch-values} \times 3_{learning-rate-values} \times 3_{batch-size-values} = 5832_{models} \quad (3.2)$$

## 3.6 Summary

The practical aspect of this project was defined in this chapter. There was an understanding of the size of the entire dataset and each sub-dataset, the algorithm is chosen, and the desired results. The next stage will be a report on the outcome of the experiments and their possible interpretations.



## Chapter 4

# Experiments

Two python scripts were made to perform the experiments defined in the previous chapter, one for each experiment. The first one had a cycle that selected each pathology label, and for each label, it iterated the three hyper-parameters among three values each. All of this happened three times, one for each sub-dataset. This means that there were 81 models for each label. Each dataset was divided into a train (70%) and a test (30%) sub-dataset.

The first python script had the structure represented in the Algorithm 1. Writing hierarchical cycles allowed the desired input variability and covered all possible cases for the experiment in only one run.

---

**Algorithm 1** Pseudo-code for the first experiment

---

```
for PATHOLOGY LABELS = [Atelectasis,...,Pnemothorax] do
  for EPOCH VALUES = [20, 40, 50] do
    for BATCH SIZE VALUES = [20, 40, 60] do
      for LEARNING RATE VALUES = [1E-4, 1E-5, 1E-6] do
        1 - run model with "epoch value", "batch size value" and
          "learning rate value" currently in the cycle

        2 - Store performance values

        3- Update best and worst performance values
      end for
    end for
  end for
end for
RETURN Best and Worst models performances
```

---

The model had the same structure for the second experiment, but it went one level deeper because it also iterated the attribute. Hence, there were six more models for each pathology, 486 in total per pathology. The python script for this second part of the experiment had a similar structure to the first but with one more level for the individual attributes. This is represented in the Algorithm 2.

**Algorithm 2** Pseudo-code for the second experiment

---

```

for PATHOLOGY LABELS = [Atelectasis,...,Pneumothorax] do
  for ATTRIBUTE LABELS = [AP, PA][FRONTAL, LATERAL][FEMALE MALE] do
    for EPOCH VALUES = [20, 40, 50] do
      for BATCH SIZE VALUES = [20, 40, 60] do
        for LEARNING RATE VALUES = [1E-4, 1E-5, 1E-6] do
          1 - run model with "attribute label", "epoch value", "batch size value" and
            "learning rate value" currently in the cycle

          2 - Store performance values

          3- Update best and worst performance values
        end for
      end for
    end for
  end for
end for
RETURN Best and Worst models performances

```

---

The best results for the first experiment are displayed in Table 4.1. The best results for the second experiment are represented in Table 4.2, where each attribute is presented in a table for each pathology. Unfortunately, only a portion of the results was obtained due to practical limitations.

Table 4.1: Average and standard deviation results of the runs of each pathology were evaluated in their accuracy, area under the ROC curve, and F1 score. The last column is the time it took the model to complete running

| Pathology                | Accuracy        | AUC             | F1              | Time (s)   |
|--------------------------|-----------------|-----------------|-----------------|------------|
|                          | avg (stdev)     | avg (stdev)     | avg (stdev)     | avg(stdev) |
| <b>Atelectasis</b>       | 0.9893 (0.0080) | 0.9965 (0.0041) | 0.9844 (0.0073) | 239 (10)   |
| <b>Cardiomegaly</b>      | 0.9891 (0.0091) | 0.9966 (0.0040) | 0.9844 (0.0073) | 205 (13)   |
| <b>Consolidation</b>     | 0.9910 (0.0081) | 0.9971 (0.0035) | 0.9844 (0.0073) | 85 (5)     |
| <b>Edema</b>             | 0.9878 (0.0081) | 0.9960 (0.0042) | 0.9844 (0.0073) | 417 (18)   |
| <b>Enlarged Cardiom.</b> | 0.9906 (0.0087) | 0.9971 (0.0030) | 0.9844 (0.0073) | 102 (8)    |
| <b>Fracture</b>          | 0.9901 (0.0083) | 0.9968 (0.0041) | 0.9844 (0.0073) | 117 (12)   |
| <b>Lung Lesion</b>       | 0.9917 (0.0085) | 0.9970 (0.0032) | 0.9844 (0.0073) | 84 (5)     |
| <b>Lung Opacity</b>      | 0.9878 (0.0080) | 0.9961 (0.0047) | 0.9844 (0.0073) | 388 (27)   |
| <b>Pleural Effusion</b>  | 0.9878 (0.0081) | 0.9958 (0.0047) | 0.9844 (0.0073) | 306 (63)   |
| <b>Pleural Other</b>     | 0.9868 (0.0299) | 0.9937 (0.0175) | 0.9777 (0.0284) | 27 (4)     |
| <b>Pneumonia</b>         | 0.9863 (0.0276) | 0.9945 (0.0138) | 0.9783 (0.0272) | 40 (3)     |
| <b>Pneumothorax</b>      | 0.9883 (0.0081) | 0.9964 (0.0040) | 0.9844 (0.0073) | 357 (15)   |

Table 4.2: Average and standard deviation results of the runs of each attribute in each pathology, evaluated in their accuracy, area under the ROC curve, and F1 score. The last column is the time it took the model to complete running.

| Pathology            | Accuracy        | AUC             | F1               | Time (s)             |
|----------------------|-----------------|-----------------|------------------|----------------------|
|                      | avg (stdev)     | avg (stdev)     | avg (stdev)      | avg                  |
| <b>Atelectasis</b>   | 0.9893 (0.0080) | 0.9965 (0.0041) | 0.9844 (0.0073)  | 239 (10)             |
| AP                   | 0.8003 (0.2324) | 0.8328 (0.2194) | 0.7950 (0.2364)  | 240 (12)             |
| PA                   | 0.7644 (0.2004) | 0.7980 (0.2003) | 0.7596 (0.2059)  | 35 (3)               |
| Frontal              | 0.8136 (0.1880) | 0.8479 (0.1655) | 0.8096 (0.1922)  | 273 (10)             |
| Lateral              | 0.7807 (0.1882) | 0.8143 (0.1891) | 0.7757 (0.1942)  | 40 (2)               |
| Male                 | 0.8254 (0.1820) | 0.8601 (0.1625) | 0.8216 (0.1862)  | 185 (4)              |
| Female               | 0.7956 (0.2022) | 0.8261 (0.1920) | 0.7884 (0.2078)  | 118 (4) <sup>1</sup> |
| <b>Cardiomegaly</b>  | 0.9891 (0.0091) | 0.9966 (0.0040) | 0.9844 (0.0073)  | 205 (13)             |
| AP                   | 0.8278 (0.1910) | 0.8631 (0.1701) | 0.8233 (0.1935)  | 137 (3) <sup>2</sup> |
| <b>Edema</b>         | 0.9878 (0.0081) | 0.9960 (0.0042) | 0.9844 (0.0073)  | 417 (18)             |
| AP                   | 0.8738 (0.1365) | 0.9073 (0.1090) | 0.8701 (0.1390)  | 475 (7)              |
| PA                   | 0.7579 (0.1974) | 0.7933 (0.1990) | 0.7547 (0.2028)  | 17 (2)               |
| Frontal              | 0.8933 (0.1101) | 0.9305 (0.0783) | 0.8910 (0.11129) | 471 (8) <sup>3</sup> |
| <b>Lung Lesion</b>   | 0.9917 (0.0085) | 0.9970 (0.0032) | 0.9844 (0.0073)  | 84 (5)               |
| AP                   | 0.7564 (0.2117) | 0.7846 (0.2176) | 0.7491 (0.2187)  | 8 (1)                |
| PA                   | 0.7122 (0.1967) | 0.7516 (0.2000) | 0.7061 (0.2034)  | 11 (2)               |
| Frontal              | 0.7712 (0.2098) | 0.7981 (0.2099) | 0.7653 (0.2175)  | 21 (2)               |
| Lateral              | 0.7252 (0.1969) | 0.7582 (0.2030) | 0.7161 (0.2053)  | 10 (2)               |
| Male                 | 0.7461 (0.2130) | 0.7712 (0.2210) | 0.7385 (0.2211)  | 54 (1)               |
| Female               | 0.7496 (0.2028) | 0.7778 (0.2028) | 0.7420 (0.2106)  | 10 (2)               |
| <b>Pleural Other</b> | 0.9868 (0.0299) | 0.9937 (0.0175) | 0.9777 (0.0284)  | 27 (4)               |
| AP                   | 0.6941 (0.1916) | 0.7218 (0.1999) | 0.6853 (0.1961)  | 8 (1)                |
| PA                   | 0.6740 (0.1747) | 0.7119 (0.1862) | 0.6645 (0.1789)  | 10 (1)               |
| Frontal              | 0.7543 (0.1932) | 0.7838 (0.2024) | 0.7443 (0.2013)  | 21 (1)               |
| Lateral              | 0.6924 (0.1852) | 0.7202 (0.1957) | 0.6827 (0.1895)  | 10 (1)               |
| Male                 | 0.7459 (0.1871) | 0.7796 (0.1935) | 0.7379 (0.1968)  | 22 (2)               |
| Female               | 0.7248 (0.1946) | 0.7598 (0.2036) | 0.7179 (0.2013)  | 10 (1)               |
| <b>Pneumonia</b>     | 0.9863 (0.0276) | 0.9945 (0.0138) | 0.9783 (0.0272)  | 40 (3)               |
| AP                   | 0.7370 (0.1914) | 0.7774 (0.1923) | 0.7302 (0.1994)  | 17 (1)               |
| PA                   | 0.7175 (0.2033) | 0.7484 (0.2097) | 0.7092 (0.2090)  | 16 (1)               |
| Frontal              | 0.7445 (0.1972) | 0.7739 (0.2037) | 0.7364 (0.2050)  | 32 (2)               |
| Lateral              | 0.7042 (0.1983) | 0.7357 (0.2088) | 0.6963 (0.2039)  | 15 (1)               |
| Male                 | 0.7505 (0.1980) | 0.7818 (0.2052) | 0.7432 (0.2074)  | 30 (1)               |
| Female               | 0.7260 (0.1956) | 0.7581 (0.2056) | 0.7191 (0.2047)  | 19 (3)               |

## 4.1 Conclusions

The results were satisfactory, with an AUC above 0.7 in almost all models.

Regarding the first experiment, Enlarged Cardiomediatinum and Consolidation were the easiest conditions for the model to identify, with an AUC of 0.9971. This might indicate that these pathologies present clearer features, with more contrast than a healthy chest X-ray. The "Pleural Other" category had the worst performance, with an AUC of 0.9937, which is understandable considering that it contains heterogeneous, unspecified diseases, meaning there were not many evident characteristics in the images for successful classification.

The second experiment presented a general decrease in performance of approximately 30%, except for Edema and Cardiomegaly, in which this decrease is reduced to approximately 15%. In all cases, the Anterior-Posterior projection got better results than the Posterior-Anterior, and the frontal view got better results than the lateral view. The male patients were also usually easier to identify, except for Lung Lesions. The worst performance was the PA projection for the "Pleural Other" category, with an AUC of 0.6645, and the best value was the AP projection for Edema, with an AUC of 0.9073.

The most probable reason for such favorable results is that the dataset was filtered so that the input cases did not have super-positioning of labels, so there was no more than one pathology per image. This narrowed the learned characteristics of each label and led to a more precise outcome.

## Chapter 5

# Conclusions and Future Work

This study provided an overview of the possibility of using deep learning models to aid medical professionals in identifying diseases visible in lung X-rays. The experimental phase was divided into two phases, one for pathologies with all associated attributes and another for isolated attributes within each pathology. a second set of experiments where each value of each binary attribute was used to filter out the dataset. This filtering enabled us to evaluate if using just frontal X-rays (against using frontal and lateral) would degrade the model's performance. With this second set of experiments, we assessed if a more straightforward dataset could achieve the same performance as the whole set of attributes and values. The worst result of the first phase was in the Pleural Other category, with an AUC of 0.9937. The best ones had an AUC of 0.9971 for Enlarged Cardiomedastinum and Consolidation classes. The worst results of the second phase, for the isolated attributes, were for the Posterior-Anterior projection of the Pleural Other category, with an AUC of 0.6645. The best was for the Anterior-Posterior projection of the Edema category, with an AUC of 0.9073. Overall the results were very positive, which indicates a promising possibility for medical diagnosis assisted by artificial intelligence, leading to faster, more precise results and more successful treatments. Although we got very promising results, let us recall that we have used simplified datasets where instances with multiple classes were discarded.



# References

- [1] <https://www.tensorflow.org/>.
- [2] <https://keras.io/>.
- [3] <https://jupyter.org/>.
- [4] <https://radiopaedia.org/articles/enlargement-of-the-cardiac-silhouette>.
- [5] Introduction to semma, Aug 2017.
- [6] Animesh Adhikari and Jhimli Adhikari. *Advances in knowledge discovery in databases*. Springer, 2015.
- [7] Hussain Khidair Ibrahim Al-Ogalie. Image classification using non negative matrix factorization and ensemble methods for classification information technology, 2016.
- [8] Hina Amin and Waqas J Siddiqui. Cardiomegaly. In *StatPearls [internet]*. StatPearls Publishing, 2021.
- [9] Maria-Luiza Antonie, Osmar R Zaiane, and Alexandru Coman. Application of data mining techniques for medical image classification.
- [10] Nick Barnes. The role of computer vision in prosthetic vision. *Image and Vision Computing*, 30(8):478–479, 2012.
- [11] Avijeet Biswal. Top 10 deep learning algorithms you should know in 2022, 2022.
- [12] Fábio André da Silva Amarante. Extracting medical information from personal child health records. 3 2018.
- [13] E Roy Davies. *Computer vision: principles, algorithms, applications, learning*. Academic Press, 2017.
- [14] Christopher E Engeler, Joseph H Tashjian, Stephen W Trenkner, and James W Walsh. Ground-glass opacity of the lung parenchyma: a guide to analysis with high-resolution ct. *AJR. American journal of roentgenology*, 160(2):249–251, 1993.
- [15] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37–37, 1996.
- [16] Rohith Gandhi. Support vector machine — introduction to machine learning algorithms, 2018.
- [17] Gartner<sub>inc</sub>. *Definition of data mining – gartner information technology glossary*.

- [18] Rafael C Gonzalez, Richard E Woods, and Barry R Masters. Digital image processing, 2009.
- [19] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [20] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silvana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 590–597, 2019.
- [21] Daniel T Larose and Chantal D Larose. *Discovering knowledge in data: an introduction to data mining*, volume 4. John Wiley & Sons, 2014.
- [22] L May, C Hillermann, and S Patil. Rib fracture management. *Bja Education*, 16(1):26–32, 2016.
- [23] Ana Maria Mendonça and Tânia Melo. Digital image fundamentals.
- [24] Ana Maria Mendonça and Tânia Melo. Feature detection.
- [25] Ana Maria Mendonça and Tânia Melo. Image segmentation.
- [26] JF Murray. Pulmonary edema: pathophysiology and diagnosis. *The International journal of tuberculosis and lung disease*, 15(2):155–160, 2011.
- [27] Ali Narin and Ziyne Pamuk. Effect of different batch size parameters on predicting of covid19 cases. *arXiv preprint arXiv:2012.05534*, 2020.
- [28] Avinash Navlani. Understanding random forests classifiers in python tutorial. <https://www.datacamp.com/tutorial/random-forests-classifier-python>, 2018.
- [29] Peiman Nazerian, Giovanni Volpicelli, Simone Vanni, Chiara Gigli, Laura Betti, Maurizio Bartolucci, Maurizio Zanobetti, Francesca Romana Ermini, Cristina Iannello, and Stefano Grifoni. Accuracy of lung ultrasound for the diagnosis of consolidations when compared to chest computed tomography. *The American journal of emergency medicine*, 33(5):620–625, 2015.
- [30] Marc Noppen and Tom De Keukeleire. Pneumothorax. *Respiration*, 76(2):121–127, 2008.
- [31] DG Peroni and AL Boner. Atelectasis: mechanisms, diagnosis and management. *Paediatric respiratory reviews*, 1(3):274–278, 2000.
- [32] Jose M Porcel and Richard W Light. Diagnostic approach to pleural effusion in adults. *American family physician*, 73(7):1211–1220, 2006.
- [33] Hang Qiu, Lin Luo, Ziqi Su, Li Zhou, Liya Wang, and Yucheng Chen. Machine learning approaches to predict peak demand days of cardiovascular admissions considering environmental exposure. *BMC Medical Informatics and Decision Making*, 20(1):1–11, 2020.
- [34] Hugo José Freixo Rodrigues. Ferramenta para text mining em textos completos, 2016.
- [35] Lior Rokach and Oded Maimon. Decision trees. In *Data mining and knowledge discovery handbook*, pages 165–192. Springer, 2005.
- [36] Jiangdian Song, Caiyun Yang, Li Fan, Kun Wang, Feng Yang, Shiyuan Liu, and Jie Tian. Lung lesion extraction using a toboggan based growing automatic segmentation approach. *IEEE transactions on medical imaging*, 35(1):337–353, 2015.



- [37] Antoni Torres, Catia Cilloniz, Michael S Niederman, Rosario Menendez, James D Chalmers, Richard G Wunderink, and Tom van der Poll. Pneumonia. *Nature Reviews Disease Primers*, 7(1):1–28, 2021.
- [38] Hao Wang, Di Niu, and Baochun Li. Distributed machine learning with a serverless architecture. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 1288–1296. IEEE, 2019.
- [39] Matthew D Zeiler. Adadelata: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

---

## Appendix A - Experience 1 Results

---

This appendix displays the best and worst result of each one of the three runs that were done for each pathology.

Table 5.1: Atelectasis

|              |              | Loss    | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|--------------|--------------|---------|----------|--------|--------|------------------------|------------|---------------|
| <b>Run 1</b> | <b>Best</b>  | 4.74E-4 | 0.9956   | 1      | 0.9916 | 20                     | 60         | 1E-5          |
|              | <b>Worst</b> | 9.35E-7 | 0.9862   | 0.9832 | 0.9744 | 20                     | 20         | 1E-4          |
| <b>Run 2</b> | <b>Best</b>  | 3.37E-3 | 0.9940   | 1      | 0.9873 | 40                     | 40         | 1E-6          |
|              | <b>Worst</b> | 2.45E-6 | 0.9684   | 0.9851 | 0.9743 | 40                     | 20         | 1E-4          |
| <b>Run 3</b> | <b>Best</b>  | 6.78E-3 | 0.9982   | 1      | 0.9916 | 50                     | 60         | 1E-6          |
|              | <b>Worst</b> | 1.51E-6 | 0.9856   | 0.9851 | 0.9744 | 20                     | 20         | 1E-4          |

Table 5.2: Cardiomegaly

|              |              | Loss     | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|--------------|--------------|----------|----------|--------|--------|------------------------|------------|---------------|
| <b>Run 1</b> | <b>Best</b>  | 2.98E-4  | 0.9961   | 1      | 0.9916 | 20                     | 60         | 1E-5          |
|              | <b>Worst</b> | 1.64E-9  | 0.9725   | 0.9846 | 0.9744 | 50                     | 20         | 1E-4          |
| <b>Run 2</b> | <b>Best</b>  | 9.03E-3  | 0.9990   | 1      | 0.9916 | 50                     | 60         | 1E-6          |
|              | <b>Worst</b> | 4.68E-11 | 0.967    | 0.9867 | 0.9744 | 50                     | 20         | 1E-4          |
| <b>Run 3</b> | <b>Best</b>  | 4.39E-4  | 0.9922   | 1      | 0.9916 | 20                     | 60         | 1E-5          |
|              | <b>Worst</b> | 2.74E-5  | 0.9725   | 0.9804 | 0.9744 | 50                     | 20         | 1E-4          |

Table 5.3: Consolidation

|              |              | Loss    | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|--------------|--------------|---------|----------|--------|--------|------------------------|------------|---------------|
| <b>Run 1</b> | <b>Best</b>  | 2.21E-4 | 0.9958   | 1      | 0.9916 | 40                     | 60         | 1E-5          |
|              | <b>Worst</b> | 5.23E-9 | 0.9792   | 0.9861 | 0.9744 | 50                     | 20         | 1E-4          |
| <b>Run 2</b> | <b>Best</b>  | 4.47E-4 | 0.9976   | 1      | 0.9916 | 50                     | 60         | 1E-5          |
|              | <b>Worst</b> | 5.02E-8 | 0.9652   | 0.9876 | 0.9744 | 50                     | 20         | 1E-4          |
| <b>Run 3</b> | <b>Best</b>  | 3.94E-3 | 1        | 1      | 0.9916 | 20                     | 60         | 1E-5          |
|              | <b>Worst</b> | 3.92E-5 | 0.9708   | 0.9841 | 0.9744 | 40                     | 20         | 1E-4          |

Table 5.4: Edema

|              |              | Loss    | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|--------------|--------------|---------|----------|--------|--------|------------------------|------------|---------------|
| <b>Run 1</b> | <b>Best</b>  | 3.87E-3 | 0.9970   | 1      | 0.9916 | 40                     | 60         | 1E-6          |
|              | <b>Worst</b> | 2.1E-24 | 0.9667   | 0.9785 | 0.9744 | 50                     | 20         | 1E-4          |
| <b>Run 2</b> | <b>Best</b>  | 3.48E-3 | 0.9970   | 0.9998 | 0.9916 | 40                     | 60         | 1E-6          |
|              | <b>Worst</b> | 5.18E-8 | 0.9723   | 0.9818 | 0.9744 | 50                     | 20         | 1E-4          |
| <b>Run 3</b> | <b>Best</b>  | 1.27E-3 | 0.9951   | 1      | 0.9916 | 50                     | 60         | 1E-6          |
|              | <b>Worst</b> | 2.12E-8 | 0.9820   | 0.9875 | 0.9744 | 40                     | 20         | 1E-4          |

Table 5.5: Enlarged Cardiomediatinum

|              |              | Loss    | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|--------------|--------------|---------|----------|--------|--------|------------------------|------------|---------------|
| <b>Run 1</b> | <b>Best</b>  | 1.25E-4 | 0.9944   | 1      | 0.9916 | 50                     | 60         | 1E-5          |
|              | <b>Worst</b> | 2.93E-6 | 0.9931   | 0.9860 | 0.9744 | 20                     | 20         | 1E-4          |
| <b>Run 2</b> | <b>Best</b>  | 9.38E-5 | 0.9963   | 1      | 0.9916 | 50                     | 60         | 1E-5          |
|              | <b>Worst</b> | 2.87E-7 | 0.9724   | 0.9877 | 0.9744 | 50                     | 20         | 1E-4          |
| <b>Run 3</b> | <b>Best</b>  | 8.04E-4 | 0.9981   | 1      | 0.9916 | 40                     | 60         | 1E-5          |
|              | <b>Worst</b> | 5.54E-7 | 0.9750   | 0.9898 | 0.9744 | 40                     | 20         | 1E-4          |

Table 5.6: Fracture

|              |              | Loss     | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|--------------|--------------|----------|----------|--------|--------|------------------------|------------|---------------|
| <b>Run 1</b> | <b>Best</b>  | 1.31E-2  | 1        | 1      | 0.9916 | 50                     | 60         | 1E-6          |
|              | <b>Worst</b> | 6.99E-6  | 0.9763   | 0.9831 | 0.9744 | 20                     | 20         | 1E-4          |
| <b>Run 2</b> | <b>Best</b>  | 1.49E-3  | 0.9958   | 1      | 0.9916 | 20                     | 60         | 1E-5          |
|              | <b>Worst</b> | 3.35E-11 | 0.9645   | 0.9831 | 0.9744 | 40                     | 20         | 1E-4          |
| <b>Run 3</b> | <b>Best</b>  | 9.63E-3  | 1        | 1      | 0.9916 | 50                     | 60         | 1E-6          |
|              | <b>Worst</b> | 3.8E-5   | 0.9863   | 0.9804 | 0.9744 | 40                     | 20         | 1E-4          |

Table 5.7: Lung Lesion

|              |              | Loss    | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|--------------|--------------|---------|----------|--------|--------|------------------------|------------|---------------|
| <b>Run 1</b> | <b>Best</b>  | 5.20E-4 | 0.9976   | 1      | 0.9916 | 40                     | 60         | 1E-5          |
|              | <b>Worst</b> | 6.26E-7 | 0.9667   | 0.9864 | 0.9744 | 50                     | 20         | 1E-4          |
| <b>Run 2</b> | <b>Best</b>  | 7.57E-4 | 1        | 1      | 0.9916 | 40                     | 60         | 1E-5          |
|              | <b>Worst</b> | 2.09E-8 | 0.9725   | 0.9869 | 0.9744 | 50                     | 20         | 1E-4          |
| <b>Run 3</b> | <b>Best</b>  | 3.14E-6 | 0.9976   | 1      | 0.9916 | 20                     | 60         | 1E-4          |
|              | <b>Worst</b> | 7.39E-7 | 0.9833   | 0.9909 | 0.9744 | 50                     | 20         | 1E-4          |

Table 5.8: Lung Opacity

|              |              | Loss     | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|--------------|--------------|----------|----------|--------|--------|------------------------|------------|---------------|
| <b>Run 1</b> | <b>Best</b>  | 3.84E-3  | 0.9978   | 1      | 0.9916 | 40                     | 60         | 1E-6          |
|              | <b>Worst</b> | 1.96E-13 | 0.9647   | 0.9815 | 0.9744 | 50                     | 20         | 1E-4          |
| <b>Run 2</b> | <b>Best</b>  | 1.28E-4  | 0.9938   | 1      | 0.9916 | 20                     | 60         | 1E-5          |
|              | <b>Worst</b> | 1.7E-11  | 0.9750   | 0.9816 | 0.9744 | 40                     | 20         | 1E-4          |
| <b>Run 3</b> | <b>Best</b>  | 1.87E-3  | 0.9917   | 1      | 0.9873 | 40                     | 40         | 1E-6          |
|              | <b>Worst</b> | 1.75E-11 | 0.9696   | 0.9828 | 0.9744 | 40                     | 20         | 1E-4          |

Table 5.9: Pleural Effusion

|              |              | Loss     | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|--------------|--------------|----------|----------|--------|--------|------------------------|------------|---------------|
| <b>Run 1</b> | <b>Best</b>  | 8.99E-4  | 0.9930   | 0.9998 | 0.9873 | 50                     | 40         | 1E-6          |
|              | <b>Worst</b> | 2.72E-11 | 0.9724   | 0.9814 | 0.9744 | 40                     | 20         | 1E-4          |
| <b>Run 2</b> | <b>Best</b>  | 2.10E-3  | 0.9991   | 1      | 0.9916 | 40                     | 60         | 1E-6          |
|              | <b>Worst</b> | 2.18E-9  | 0.9746   | 0.9832 | 0.9744 | 40                     | 20         | 1E-4          |
| <b>Run 3</b> | <b>Best</b>  | 5.90E-3  | 0.9965   | 0.9998 | 0.9873 | 20                     | 40         | 1E-6          |
|              | <b>Worst</b> | 5.67E-9  | 0.9645   | 0.9841 | 0.9744 | 50                     | 20         | 1E-4          |

Table 5.10: Pleural Other

|              |              | Loss    | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|--------------|--------------|---------|----------|--------|--------|------------------------|------------|---------------|
| <b>Run 1</b> | <b>Best</b>  | 6.70E-3 | 1        | 1      | 0.9916 | 50                     | 60         | 1E-5          |
|              | <b>Worst</b> | 0.2416  | 0.9750   | 0.9868 | 0.9664 | 40                     | 60         | 1E-6          |
| <b>Run 2</b> | <b>Best</b>  | 1.39E-3 | 1        | 1      | 0.9916 | 20                     | 60         | 1E-4          |
|              | <b>Worst</b> | 0.5920  | 0.7583   | 0.8455 | 0.7647 | 20                     | 60         | 1E-6          |
| <b>Run 3</b> | <b>Best</b>  | 6.15E-4 | 1        | 1      | 0.9916 | 20                     | 60         | 1E-4          |
|              | <b>Worst</b> | 0.3145  | 0.9500   | 0.9849 | 0.9412 | 20                     | 60         | 1E-6          |

Table 5.11: Pneumonia

|              |              | Loss    | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|--------------|--------------|---------|----------|--------|--------|------------------------|------------|---------------|
| <b>Run 1</b> | <b>Best</b>  | 1.08E-2 | 1        | 1      | 0.9873 | 20                     | 40         | 1E-5          |
|              | <b>Worst</b> | 0.5739  | 0.8111   | 0.8930 | 0.8011 | 20                     | 60         | 1E-6          |
| <b>Run 2</b> | <b>Best</b>  | 7.39    | 1        | 1      | 0.9916 | 50                     | 60         | 1E-5          |
|              | <b>Worst</b> | 0.4731  | 0.8444   | 0.9317 | 0.8347 | 20                     | 60         | 1E-6          |
| <b>Run 3</b> | <b>Best</b>  | 6.05E-6 | 1        | 1      | 0.9916 | 40                     | 60         | 1E-4          |
|              | <b>Worst</b> | 1.86E-6 | 0.9850   | 0.9857 | 0.9744 | 40                     | 20         | 1E-4          |

Table 5.12: Pneumothorax

|              |              | Loss     | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|--------------|--------------|----------|----------|--------|--------|------------------------|------------|---------------|
| <b>Run 1</b> | <b>Best</b>  | 3.41E-3  | 0.9994   | 1      | 0.9916 | 40                     | 60         | 1E-6          |
|              | <b>Worst</b> | 2.21E-12 | 0.9693   | 0.9840 | 0.9744 | 50                     | 20         | 1E-4          |
| <b>Run 2</b> | <b>Best</b>  | 4.93E-3  | 0.9982   | 1      | 0.9916 | 40                     | 60         | 1E-6          |
|              | <b>Worst</b> | 2.09E-6  | 0.9779   | 0.9864 | 0.9744 | 50                     | 20         | 1E-4          |
| <b>Run 3</b> | <b>Best</b>  | 1.13E-3  | 0.9948   | 0.9991 | 0.9873 | 50                     | 40         | 1E-6          |
|              | <b>Worst</b> | 3.53E-12 | 0.9639   | 0.9853 | 0.9743 | 40                     | 20         | 1E-4          |

---

## Appendix B - Experience 2 Results

---

This appendix displays the best and worst result of each one of the three runs that were done for each attribute of each pathology

Table 5.13: Atelectasis

|                |              |              | Loss    | Accuracy | AUC    | F1      | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|----------------|--------------|--------------|---------|----------|--------|---------|------------------------|------------|---------------|
| <b>AP</b>      | <b>Run 1</b> | <b>Best</b>  | 1.84E-2 | 0.9879   | 0.9978 | 0.9881  | 50                     | 60         | 1E-5          |
|                |              | <b>Worst</b> | 0.4472  | 0.5259   | 0.5410 | 0.5181  | 20                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 1.93E-2 | 0.9902   | 0.9980 | 0.9887  | 50                     | 60         | 1E-5          |
|                |              | <b>Worst</b> | 0.4536  | 0.5161   | 0.5355 | 0.5065  | 20                     | 60         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 1.35E-2 | 0.9919   | 0.9977 | 0.9899  | 50                     | 60         | 1E-6          |
|                |              | <b>Worst</b> | 0.4484  | 0.5080   | 0.5332 | 0.4984  | 20                     | 60         | 1E-6          |
| <b>PA</b>      | <b>Run 1</b> | <b>Best</b>  | 2.70E-3 | 0.9917   | 0.9959 | 0.9916  | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5034  | 0.4931   | 0.4971 | 0.4752  | 20                     | 24         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 6.22E-4 | 0.9833   | 0.9979 | 0.9916  | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5246  | 0.500    | 0.4648 | 0.4725  | 20                     | 20         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 1.22E-3 | 0.9917   | 0.9979 | 0.9916  | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5532  | 0.4792   | 0.4508 | 0.46664 | 20                     | 60         | 1E-6          |
| <b>Frontal</b> | <b>Run 1</b> | <b>Best</b>  | 1.81E-2 | 0.9863   | 0.9988 | 0.9886  | 50                     | 60         | 1E-5          |
|                |              | <b>Worst</b> | 0.4309  | 0.5274   | 0.5589 | 0.5175  | 20                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 1.14E-2 | 0.9902   | 0.9983 | 0.9896  | 50                     | 60         | 1E-5          |
|                |              | <b>Worst</b> | 0.4382  | 0.5289   | 0.5592 | 0.5245  | 20                     | 60         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 1.09E-2 | 0.9905   | 0.9983 | 0.9911  | 50                     | 60         | 1E-5          |
|                |              | <b>Worst</b> | 0.4306  | 0.5357   | 0.5745 | 0.5301  | 20                     | 60         | 1E-6          |
| <b>Lateral</b> | <b>Run 1</b> | <b>Best</b>  | 1.98E-3 | 0.9867   | 0.9967 | 0.9916  | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.4961  | 0.5133   | 0.5009 | 0.5042  | 40                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 5.01E-4 | 0.9900   | 0.9984 | 0.9916  | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5046  | 0.5500   | 0.5291 | 0.5256  | 20                     | 20         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 1.57E-3 | 0.9900   | 0.9983 | 0.9916  | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.4486  | 0.5481   | 0.5402 | 0.5270  | 40                     | 24         | 1E-6          |
| <b>Female</b>  | <b>Run 1</b> | <b>Best</b>  | 1.39E-4 | 0.9967   | 0.9973 | 0.9916  | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.4841  | 0.5133   | 0.5261 | 0.5020  | 20                     | 60         | 1E-6          |
| <b>Male</b>    | <b>Run 1</b> | <b>Best</b>  | 2.81E-2 | 0.9847   | 0.9986 | 0.9839  | 50                     | 60         | 1E-5          |
|                |              | <b>Worst</b> | 0.4791  | 0.5063   | 0.5247 | 0.5000  | 20                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 2.05E-2 | 0.9903   | 0.9995 | 0.9881  | 50                     | 60         | 1E-5          |
|                |              | <b>Worst</b> | 0.4465  | 0.5549   | 0.5816 | 0.5469  | 20                     | 60         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 2.68E-2 | 0.9868   | 0.9980 | 0.9853  | 50                     | 60         | 1E-5          |
|                |              | <b>Worst</b> | 0.4531  | 0.5368   | 0.5620 | 0.5280  | 20                     | 60         | 1E-6          |



Table 5.14: Cardiomegaly

|           |              |              | Loss   | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|-----------|--------------|--------------|--------|----------|--------|--------|------------------------|------------|---------------|
| <b>AP</b> | <b>Run 1</b> | <b>Best</b>  | 0.0191 | 0.9889   | 0.9997 | 0.9879 | 50                     | 60         | 1E-5          |
|           |              | <b>Worst</b> | 0.4500 | 0.5444   | 0.5715 | 0.5350 | 20                     | 60         | 1E-6          |
|           | <b>Run 2</b> | <b>Best</b>  | 0.0241 | 0.9873   | 0.9984 | 0.9857 | 50                     | 60         | 1E-5          |
|           |              | <b>Worst</b> | 0.4862 | 0.5059   | 5060   | 0.4963 | 20                     | 60         | 1E-6          |

Table 5.15: Edema

|                |              |              | Loss    | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|----------------|--------------|--------------|---------|----------|--------|--------|------------------------|------------|---------------|
| <b>AP</b>      | <b>Run 1</b> | <b>Best</b>  | 5.93E-3 | 0.9886   | 0.9985 | 0.9896 | 50                     | 60         | 1E-5          |
|                |              | <b>Worst</b> | 0.3733  | 0.5889   | 0.6573 | 0.5880 | 20                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 9.98E-3 | 0.9911   | 0.9979 | 0.9902 | 40                     | 60         | 1E-5          |
|                |              | <b>Worst</b> | 0.3720  | 0.5933   | 0.6607 | 0.5910 | 20                     | 60         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 1.5E-2  | 0.9842   | 0.9982 | 0.9872 | 40                     | 60         | 1E-5          |
|                |              | <b>Worst</b> | 0.3665  | 0.5918   | 0.6635 | 0.5885 | 20                     | 60         | 1E-6          |
| <b>PA</b>      | <b>Run 1</b> | <b>Best</b>  | 3.95E-2 | 0.9722   | 0.9972 | 0.9645 | 20                     | 24         | 1E-4          |
|                |              | <b>Worst</b> | 0.5559  | 0.4653   | 0.4970 | 0.4539 | 20                     | 24         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 1.02E-4 | 1        | 1      | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5225  | 0.4214   | 0.4484 | 0.4103 | 20                     | 20         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 2.08E-3 | 0.9917   | 1      | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.6407  | 0.5000   | 0.4851 | 0.4958 | 20                     | 60         | 1E-6          |
| <b>Frontal</b> | <b>Run 1</b> | <b>Best</b>  | 4.91E-3 | 0.9926   | 0.9980 | 0.9911 | 50                     | 60         | 1E-5          |
|                |              | <b>Worst</b> | 0.3255  | 0.6775   | 0.7508 | 0.6747 | 20                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 2.01E-3 | 0.9922   | 0.9977 | 0.9916 | 50                     | 60         | 1E-5          |
|                |              | <b>Worst</b> | 0.3225  | 0.6723   | 0.7608 | 0.6704 | 20                     | 60         | 1E-6          |

Table 5.16: Lung Lesion

|                |              |              | Loss    | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|----------------|--------------|--------------|---------|----------|--------|--------|------------------------|------------|---------------|
| <b>AP</b>      | <b>Run 1</b> | <b>Best</b>  | 0.0297  | 0.9611   | 0.9978 | 0.9630 | 50                     | 20         | 1E-5          |
|                |              | <b>Worst</b> | 0.4987  | 0.4556   | 0.4574 | 0.4454 | 40                     | 24         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 2.15E-4 | 0.9944   | 0.9973 | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5199  | 0.4583   | 0.4450 | 0.4312 | 20                     | 24         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 0.0268  | 0.9667   | 0.9981 | 0.9715 | 50                     | 20         | 1E-5          |
|                |              | <b>Worst</b> | 0.5192  | 0.4944   | 0.4857 | 0.4738 | 20                     | 24         | 1E-6          |
| <b>PA</b>      | <b>Run 1</b> | <b>Best</b>  | 0.0243  | 0.9889   | 0.9938 | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5060  | 0.5056   | 0.5192 | 0.4930 | 50                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 4.38E-3 | 0.9944   | 0.9972 | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5520  | 0.4611   | 0.4274 | 0.4482 | 50                     | 60         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 4.52E-3 | 1        | 0.9972 | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.4749  | 0.4630   | 0.4753 | 0.4397 | 50                     | 24         | 1E-6          |
| <b>Frontal</b> | <b>Run 1</b> | <b>Best</b>  | 0.0136  | 0.9732   | 0.9959 | 0.9725 | 50                     | 20         | 1E-5          |
|                |              | <b>Worst</b> | 0.5047  | 0.4833   | 0.4697 | 0.4762 | 40                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 0.0189  | 0.9767   | 0.9957 | 0.9770 | 50                     | 24         | 1E-5          |
|                |              | <b>Worst</b> | 0.5100  | 0.5133   | 0.5092 | 0.5025 | 20                     | 60         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 8.47E-5 | 0.9900   | 0.9959 | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.4687  | 0.5100   | 0.5086 | 0.4906 | 20                     | 20         | 1E-6          |
| <b>Lateral</b> | <b>Run 1</b> | <b>Best</b>  | 1.64E-3 | 0.9889   | 1      | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.6960  | 0.3722   | 0.3670 | 0.3641 | 20                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 0.0125  | 0.9778   | 0.9991 | 0.9860 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5417  | 0.4583   | 0.4350 | 0.4309 | 40                     | 24         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 3.23E-3 | 0.9944   | 1      | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5241  | 0.5111   | 0.4835 | 0.4986 | 50                     | 60         | 1E-6          |
| <b>Female</b>  | <b>Run 1</b> | <b>Best</b>  | 5.29E-4 | 0.9967   | 0.9967 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.4929  | 0.4900   | 0.4989 | 0.4718 | 20                     | 20         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 1.10E-3 | 0.9967   | 0.9951 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5265  | 0.4967   | 0.4530 | 0.4874 | 50                     | 60         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 0.0503  | 0.9600   | 0.9953 | 0.9641 | 50                     | 20         | 1E-5          |
|                |              | <b>Worst</b> | 0.5331  | 0.4872   | 0.4794 | 0.4550 | 20                     | 24         | 1E-6          |
| <b>Male</b>    | <b>Run 1</b> | <b>Best</b>  | 2.29E-3 | 0.9905   | 0.9976 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5367  | 0.4762   | 0.4762 | 0.4682 | 20                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 8.89E-4 | 0.9976   | 0.9976 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5169  | 0.4714   | 0.4691 | 0.4610 | 20                     | 60         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 1.10E-3 | 0.9881   | 0.9977 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.4883  | 0.4772   | 0.4638 | 0.4476 | 20                     | 20         | 1E-6          |

Table 5.17: Pleural Other

|                |              |              | Loss    | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|----------------|--------------|--------------|---------|----------|--------|--------|------------------------|------------|---------------|
| <b>AP</b>      | <b>Run 1</b> | <b>Best</b>  | 5.57E-3 | 1        | 1      | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5973  | 0.3333   | 0.2963 | 0.3077 | 20                     | 20         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 6.09E-2 | 0.9833   | 1      | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5489  | 0.4583   | 0.4703 | 0.4397 | 20                     | 24         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 0.0105  | 0.9833   | 1      | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5615  | 0.500    | 0.4189 | 0.4872 | 40                     | 20         | 1E-6          |
| <b>PA</b>      | <b>Run 1</b> | <b>Best</b>  | 1.61E-4 | 0.9900   | 0.9857 | 0.9744 | 40                     | 20         | 1E-4          |
|                |              | <b>Worst</b> | 0.6981  | 0.4833   | 0.4650 | 0.4874 | 20                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 4.08E-3 | 0.9800   | 0.9990 | 0.9744 | 40                     | 20         | 1E-4          |
|                |              | <b>Worst</b> | 0.6861  | 0.5000   | 0.4161 | 0.4874 | 20                     | 60         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 4.32E-4 | 0.9800   | 0.9952 | 0.9744 | 50                     | 20         | 1E-4          |
|                |              | <b>Worst</b> | 0.5755  | 0.4600   | 0.4450 | 0.4308 | 20                     | 20         | 1E-6          |
| <b>Frontal</b> | <b>Run 1</b> | <b>Best</b>  | 1.09E-3 | 1        | 0.9973 | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5513  | 0.4667   | 0.4750 | 0.4594 | 40                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 3.48E-3 | 0.9833   | 1      | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5226  | 0.5179   | 0.5094 | 0.4985 | 20                     | 24         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 5.42E-4 | 0.9889   | 1      | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5723  | 0.4611   | 0.4055 | 0.4387 | 20                     | 20         | 1E-6          |
| <b>Lateral</b> | <b>Run 1</b> | <b>Best</b>  | 5.79E-3 | 0.9792   | 0.9946 | 0.9787 | 40                     | 24         | 1E-4          |
|                |              | <b>Worst</b> | 0.5923  | 0.4500   | 0.4276 | 0.4370 | 20                     | 60         | 1E-5          |
|                | <b>Run 2</b> | <b>Best</b>  | 1.28E-3 | 1        | 0.9881 | 0.9744 | 50                     | 20         | 1E-4          |
|                |              | <b>Worst</b> | 0.5903  | 0.4500   | 0.4475 | 0.4103 | 50                     | 20         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 1.67E-4 | 0.9700   | 1      | 0.9744 | 50                     | 20         | 1E-4          |
|                |              | <b>Worst</b> | 0.5750  | 0.4479   | 0.3812 | 0.4149 | 40                     | 24         | 1E-6          |
| <b>Female</b>  | <b>Run 1</b> | <b>Best</b>  | 0.0202  | 0.9833   | 0.9992 | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5932  | 0.5000   | 0.4875 | 0.5042 | 20                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 0.0118  | 1        | 1      | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.8334  | 0.4167   | 0.4310 | 0.4202 | 40                     | 60         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 2.07E-4 | 0.9688   | 0.9950 | 0.9787 | 50                     | 24         | 1E-4          |
|                |              | <b>Worst</b> | 0.5455  | 0.4792   | 0.4393 | 0.4574 | 20                     | 24         | 1E-6          |
| <b>Male</b>    | <b>Run 1</b> | <b>Best</b>  | 1.67E-3 | 0.9944   | 0.9973 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.4768  | 0.4944   | 0.5035 | 0.4786 | 50                     | 20         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 8.95E-4 | 0.9833   | 0.9973 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5441  | 0.5056   | 0.4828 | 0.4843 | 20                     | 20         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 2.63E-3 | 0.9833   | 0.9973 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5014  | 0.4889   | 0.5090 | 0.4672 | 40                     | 20         | 1E-6          |

Table 5.18: Pneumonia

|                |              |              | Loss    | Accuracy | AUC    | F1     | N <sup>er</sup> Epochs | Batch Size | Learning Rate |
|----------------|--------------|--------------|---------|----------|--------|--------|------------------------|------------|---------------|
| <b>AP</b>      | <b>Run 1</b> | <b>Best</b>  | 2.99E-4 | 0.9833   | 1      | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5036  | 0.4931   | 0.5165 | 0.4752 | 50                     | 24         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 5.76E-3 | 1        | 0.9959 | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5145  | 0.4750   | 0.4517 | 0.4706 | 40                     | 60         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 3.92E-3 | 0.9833   | 0.9999 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5206  | 0.5556   | 0.5271 | 0.5319 | 20                     | 24         | 1E-6          |
| <b>PA</b>      | <b>Run 1</b> | <b>Best</b>  | 8.27E-4 | 1        | 1      | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5704  | 0.4500   | 0.4635 | 0.4396 | 20                     | 20         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 0.0260  | 0.9917   | 0.9947 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5246  | 0.5000   | 0.4658 | 0.4689 | 20                     | 20         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 0.0135  | 1        | 0.9957 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5141  | 0.4583   | 0.4227 | 0.4359 | 40                     | 20         | 1E-6          |
| <b>Frontal</b> | <b>Run 1</b> | <b>Best</b>  | 3.08E-4 | 0.9875   | 0.9980 | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5357  | 0.5042   | 0.4960 | 0.4958 | 40                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 4.05E-4 | 0.9917   | 0.9980 | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5580  | 0.4625   | 0.4410 | 0.4496 | 20                     | 60         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 3.88E-4 | 0.9833   | 1      | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.4883  | 0.4848   | 0.4678 | 0.4603 | 40                     | 24         | 1E-6          |
| <b>Lateral</b> | <b>Run 1</b> | <b>Best</b>  | 1.33E-3 | 0.9917   | 1      | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.6160  | 0.4250   | 0.3986 | 0.4202 | 20                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 9.51E-4 | 0.9917   | 1      | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5583  | 0.4857   | 0.4556 | 0.4615 | 20                     | 20         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 2.02E-3 | 0.9833   | 1      | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5569  | 0.500    | 0.4989 | 0.4874 | 40                     | 60         | 1E-6          |
| <b>Female</b>  | <b>Run 1</b> | <b>Best</b>  | 0.0148  | 0.9917   | 0.9952 | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5792  | 0.4583   | 0.4452 | 0.4454 | 40                     | 60         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 5.24E-4 | 0.9833   | 0.9973 | 0.9916 | 50                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5724  | 0.4583   | 0.4716 | 0.4498 | 20                     | 24         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 0.0167  | 0.9917   | 0.9997 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.6188  | 0.4236   | 0.4602 | 0.4113 | 20                     | 24         | 1E-6          |
| <b>Male</b>    | <b>Run 1</b> | <b>Best</b>  | 1.27E-3 | 0.9917   | 0.9979 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5000  | 0.4875   | 0.4799 | 0.4638 | 20                     | 24         | 1E-6          |
|                | <b>Run 2</b> | <b>Best</b>  | 4.05E-3 | 0.9917   | 0.9978 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.4815  | 0.5000   | 0.4988 | 0.4766 | 50                     | 24         | 1E-6          |
|                | <b>Run 3</b> | <b>Best</b>  | 1.54E-3 | 0.9875   | 0.9979 | 0.9916 | 40                     | 60         | 1E-4          |
|                |              | <b>Worst</b> | 0.5401  | 0.5250   | 0.4910 | 0.4979 | 20                     | 24         | 1E-6          |