INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Covid-19 Classification From Chest X-Ray Analysis

José Miguel Nunes Fernandes

Master's in Integrated Business Intelligence Systems

Supervisor:
PhD João Pedro Afonso Oliveira da Silva, Associate Professor,
Iscte-iul

Co-Supervisor:
PhD Luís Manuel Nobre de Brito Elvas, Invited Assistant Professor,
Iscte-Iul

September, 2024

iscte
TECNOLOGIAS
E ARQUITETURA

Department of Information Science and Technology

Covid-19 Classification From Chest X-Ray Analysis

José Miguel Nunes Fernandes

Master's in Integrated Business Intelligence Systems

Supervisor:
PhD João Pedro Afonso Oliveira da Silva, Associate Professor,
Iscte-iul

Co-Supervisor:
PhD Luís Manuel Nobre de Brito Elvas, Invited Assistant Professor,
Iscte-Iul

September, 2024

INSTITUTO UNIVERSITÁRIO DE LISBOA

Covid-19 Classification From Magnetic Resonance Imaging Analysis

José Miguel Nunes Fernandes

Acknowledgements

I would like to sincerely thank and convey my gratitude to all of the people who have continuously supported me during this difficult road.

First and foremost, I want to thank Professor Luís Elvas for being both my respected professor and an incredible friend. Throughout my two-year Master's program, as well as during my thesis, I am incredibly appreciative of his assistance and patience.

To Professor Joao Pedro Oliveira, thank you for guiding me in crafting this thesis. Your opinions have been good in helping me develop the best research work possible.

To my family, thank you for your support and timely encouragement. Your belief in me and occasional pressure helped me stay focused and inspired.

To my girlfriend, Diana Cruz, that has been with me since day one and has encouraged me to continue this work until the end, despite worst days with less motivation, her support was essential to keep me pushing until the end and finish this thesis.

I must also express my gratitude to my friends that I have met when I was a little kid and have been helping me grow, to my friends from Iscte Junior Consulting, and to my good friend Pedro Águas that was doing his work alongside me and helped me through the process. Your support and strength made this research journey manageable. From challenging me to persevere to providing help when needed, you have been there since the beginning.

I am sincerely grateful to everyone who supported me and had the determination to believe in me. Your assistance was essential to this work's successful conclusion.

I dedicate this work to all of you with great joy.

Resumo

Com a rápida evolução da pandemia de Covid-19, o papel da Inteligência Artificial tornou-se cada vez mais importante para obter um diagnóstico preciso da doença e do seu impacto no corpo humano, sem direta intervenção no mesmo. Este trabalho centrar-se-á principalmente na análise da radiografia do peito (CXR) e na posterior classificação da Covid-19 utilizando técnicas de *Deep Learning* (DL), destacando a importância destas tecnologias no combate à pandemia. Esta revisão sistemática sobre DL demonstrou que as radiografias estão entre as técnicas de imagem médica mais utilizadas para a classificação da Covid-19, assim como as Redes Neuronais Convolucionais (CNN) que assumem um papel importante nesta investigação, devido aos bons resultados que apresentam quando se trata de análise de imagens. Ao longo desta dissertação, iremos (1) concentrar-nos em encontrar o modelo que obtém os melhores resultados, (2) fazer o fine tune do mesmo e (3) aplicar um *heat map* GRAD-CAM para fornecer uma representação visual. O modelo que obteve os melhores resultados foi o VGG19, que alcançou 97% de *accuracy* e *recall*. Ao identificar o modelo ótimo e ao aplicar a análise visual das condições dos pacientes, permitimos diagnósticos mais rápidos e mais precisos para os médicos, o que ajuda a reduzir os tempos de espera, fornecendo também ao paciente uma imagem gráfica para uma melhor compreensão do problema.

Palavras-chave: *Machine Learning, Medical Image Analysis, AI, Deep Learning, Chest X-Ray, Covid-19.*

Abstract

With the quick evolution of the Covid-19 pandemic, the role of AI became more and more important in order to get accurate diagnosis of the disease and its impact on the human body, without human intervention. This work will focus mainly on the analysis of Chest X-Ray (CXR) and further classification of Covid-19 by using Deep Learning (DL) techniques, highlighting the importance of these technologies when it comes to fighting the pandemic. A systematic review on DL demonstrated that chest CXRs are among the most frequently used medical imaging techniques for COVID-19 classification, as well as Convolutional Neural Networks (CNN) that take a great part in this investigation, due to the good results they present when it comes to image analysis. Throughout this dissertation we will (1) focus on finding the model that gets the best results, (2) fine tune it and (3) apply a GRAD-CAM heat map to provide a visual representation. The model that produced the best results was VGG19 reaching 97% recall and precision. By identifying the optimal model and applying visual analysis of patients' conditions, we enable quicker and more accurate diagnoses for doctors, which helps reduce waiting times, also providing the patient with a graphical image for better understanding of the problem.

Keywords: Machine Learning, Medical Image Analysis, AI, Deep Learning, Chest X-Ray, Covid-19.

Index

List of Figures and Tables

Glossary

Artificial Intelligence – AI

CNN – Convolutional Neural Network

CRISP-DM – Cross Industry Standard Process for Data Mining

CV – Computer Vision

CXR – Chest X-Ray

CT – Computed Tomography

DL – Deep Learning

HSM – Hospital de Santa Maria

ML – Machine Learning

PRISMA – Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RQ – Research Question

WoS – Web of Science

CHAPTER 1

Introduction

We are currently living in an era characterized by rapid advancements in Artificial Intelligence (AI) , transforming various sectors, revolutionizing the way we approach complex problems and raising a lot of questions about the future [1]. In the healthcare industry, AI holds immense potential to enhance the capabilities of medical professionals, particularly in the realm of diagnostics.

The aim of this study is to explore and implement AI technologies to assist doctors in making faster and more accurate diagnoses. Using AI to analyse massive volumes of medical data and spot patterns that would be invisible to the human eye, this research seeks to improve diagnostic efficiency and accuracy, ultimately contributing to an easier process to all involved parties, as well as a better understanding on the patients side.

1.1 Context

By the end of 2019, the World Health Organization (WHO) noted a media statement from the Wuhan Municipal Health Commission about atypical cases of "viral pneumonia", marking the beginning of the Covid-19 outbreak. Starting in China, the virus rapidly spread worldwide, leading the WHO to declare Covid-19 a pandemic on March 11, 2020 [2] . Common symptoms include fever, cough, and loss of taste or smell, typically appearing 5 to 6 days post-infection. While healthy individuals can manage these symptoms at home, severe cases, particularly those involving difficulty breathing, chest pain, or shortness of breath, often require hospitalization, especially when compounded by conditions like diabetes, heart disease, obesity, and heart failure [3].

Combining a large amount of data with Machine Learning approaches, such as Deep Learning (DL), is proven to be one of the most efficient ways to mix technology and medicine [4], in a way that computers can provide humans with critical, quick and precise diagnosis of a disease.

These AI algorithms are in constant development, continuing to strengthen the medical image processing system from the infection analysis, to the categorization of the infection, all the way to further diagnosis. This way, AI empowered models, are reliable enough in order to assist radiologists and health experts in order to make better clinical decisions.

Chest X-Rays (CXR), Computed Tomography (CT) and Magnetic Resonance Imaging (MRI), are the most common medical imaging techniques to address the detection and evolution of Covid-19 on patients [5]. From these techniques, and regarding Covid-19 classification, CXR is the one that is most used worldwide due to its cost being way less when compared to the other two exams [6].

1.2 Objectives

The objective of this research is to build a Convolutional Neural Network (CNN) model for the classification of Covid-19 in chest X-rays by combining it with computer vision techniques. The central research question (RQ) is: "How can we effectively use AI to automatically classify CXR's of patients with Covid-19?"

To address this RQ, a CNN model integrated with computer vision techniques will be implemented and tested to accurately distinguish between CXR's showing Covid-19 and those without it. Making a binary conclusion about the disease's presence or absence is required for this categorization task.

To accomplish our objectives and address our research question, we have adopted the CRISP-DM (Cross-Industry Standard Process for Data Mining) [7] methodology for image data. This methodology guides us through various stages, including Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation and Deployment, to achieve our research goals.

Throughout this research, we aim to identify the most effective model for the binary classification of Covid-19 from chest CXR images. By testing and comparing various models, we seek to determine which one provides the best results in distinguishing Covid-19 cases from non-Covid-19 cases.

Building on this result, we will apply a heat map technique, Grad-CAM [8], to create visual representations of the model's predictions. These heat maps will highlight the areas of the X-ray images that contributed most to the model's decision, providing valuable insights that can be used in a medical context to enhance the diagnostic process and support healthcare professionals in their assessments.

1.3 Methodology

The structure of this dissertation adheres to the CRISP-DM methodology, a prevalent framework in data mining and analytics projects. This methodology comprises six stages, as depicted in Figure 1:

1. **Business Understanding**: Evaluate the business context to comprehend available and necessary resources, clearly outline the data mining objectives and success criteria, and develop an essential project plan.

2. **Data Understanding**: Gather and examine data from diverse sources, verify data quality, and utilize statistical analysis to describe the data, identifying key attributes and their relationships.

3. **Data Preparation**: Establish criteria for data inclusion and exclusion, rectify poor data quality through cleaning, and create derived attributes based on the model selected in the initial phase.

4. **Modelling**: Choose a modelling technique suited to the business problem and data, configure specific parameters for constructing the model, and assess the model against predefined criteria to select the most effective one.

5. **Evaluation**: Validate results against established business objectives, discuss findings, outline future work, and review the entire process.

6. **Deployment**: Organize the deployment phase, which could result in a final report or a software component; this phase includes planning, monitoring, and maintenance activities.

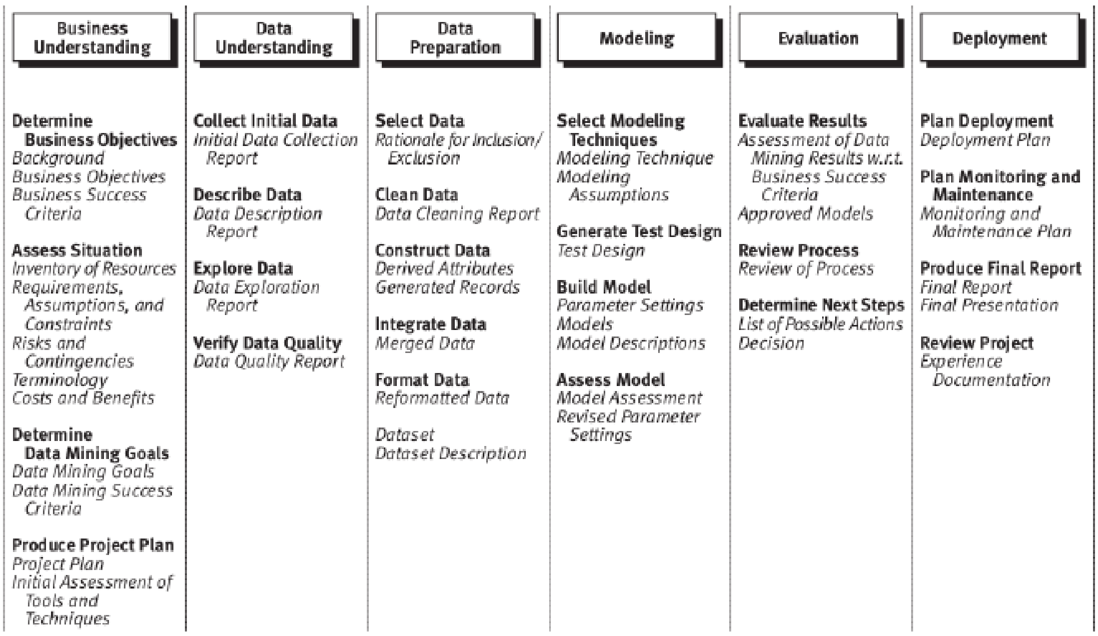

Figure 1 – CRISP-DM Methodology [9]

1.4 Dissertation’s Outline

The primary aim of this dissertation is to automate the detection of Covid-19 in patients by identifying the presence of the disease in chest CXR images. To achieve this, we will (1) implement and evaluate the performance of five different models to determine which yields the best results, (2) fine-tune the top-performing model to optimize its accuracy, and (3) apply a heat map to the refined model to enhance the interpretability and accuracy of the results.

This dissertation is structured into five chapters (including introduction):

- **Chapter 2**: A literature review on the state-of-the-art techniques for Covid-19 detection from CXR images using computer vision, including an explanation and application of the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [10] methodology.
- **Chapter 3**: Following the CRISP-DM [11] methodology, we create the dataset, apply data augmentation techniques, and provide detailed explanations of the models used in the research.
- **Chapter 4**: Finding out what is the model that gets the best results, fine tuning it and applying a heat map over the final images.
- **Chapter 5**: A conclusion of the conducted project and areas for improvement, along with delineation of potential future advancements.

CHAPTER 2

Literature Review

2.1. Search Strategy and Inclusion Criteria

The PRISMA methodology [10] was conducted throughout this systematic review, which is also aiming to tackle de following research question: “What is the state of the art on AI models for automatic Covid-19 classification on Ches R-Rays?”.

Scopus and Web of Science (WoS) databases were used in order to get articles capable of answering the research question. This study was conducted on the 20th of April 2024, filtering the results to only get papers, articles and reviews from the last five years (2019-2024), and written in English.

The same query was used in both databases, allowing to take into account the number of articles as well as the concept, context and population.

2.2. Study Selection

After removing the duplicates, the papers selection was made by analysing the title and the abstract in most cases. If this was not enough, the whole text was analysed in order to obtain the required piece of information, that was needed from the article.

2.3. Data Extraction and Synthesis

The title, author, year, journal, subject area, keywords and abstract were all managed and stored through both Zotero and Microsoft Excel. Taking into account the results obtained, a qualitative review of the articles was conducted in order to analyse and synthetize the data. Scopus and WoS were both used to systematically search the existent work regarding “medical image analysis” or “machine learning” as concept, the target population “Chest X-Ray” and the context of study being “Covid-19”.

2.4. Results

The search results were obtained by applying the concept, target population and context mentioned previously to WoS and Scopus. The same query was applied to both databases, resulting in some duplicate values, as expected. The first batch of articles is described on Table 1.

Table 1 - Keyword selection using PRISMA

<table>
<tr><th>Concept</th><th>Population</th><th>Context</th><th>Limitations</th></tr>
<tr><td>machine learning</td><td rowspan="2">Chest X-Ray</td><td rowspan="2">Covid-19</td><td rowspan="2">2020 - 2024</td></tr>
<tr><td>medical image analysis</td></tr>
<tr><td>AI</td><td rowspan="2">cxr</td><td rowspan="2">binary classification</td><td rowspan="4">Only journal papers, articles and reviews</td></tr>
<tr><td>deep learning</td></tr>
<tr><td>489 517 documents</td><td></td><td></td></tr>
<tr><td colspan="2">4 119 documents</td><td></td></tr>
<tr><td colspan="4">49 documents</td></tr>
</table>

By analysing Table 1, the search query was made by taking the keywords from each column (Concept AND Population AND Context AND Limitations) returning a total of 49 documents. After manually eliminating all the duplicates, as well as filtering the obtained results in order to exclude the less important, 17 documents were collected. It is possible to see the detailed process on Figure 2.

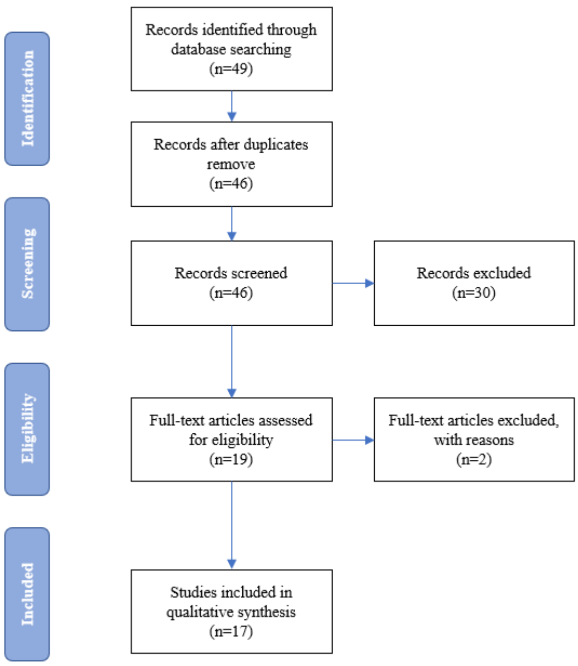

Figure 2 - PRISMA workflow diagram

From the documents obtained, it was only expected for them to be relatively recent due to the fact this study focuses on the detection of Covid-19 impact on the patient's lungs and, even though the pandemic only started in 2020, relevant articles for this review have only been published on the year after, despite having only 1 article from that same year. Therefore, 24% of the results are from 2021, 41% from 2022, 24% from 2023 and only 1 article, representing 6%, on both 2020 and 2024.

Having the goal of this literature review in mind, a list of main topics was put together in order to associate them with each document collected. A detailed analysis of the topics is present on Table 2, crossing the topics with the references and the number of documents found for each one. The highlight of this analysis is definitely the fact that there were a lot of articles regarding binary classification of Covid-19 on CXR's.

Table 2 - Studies by topics

Topic	Reference	# of Documents
Pre trained model	[12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]	17
Covid-19	[12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]	17
Classification	[12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]	17
ML	[12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]	17
Deep Learning	[12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]	17
CXR	[12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]	16
VGG	[4], [5], [6], [9], [10], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]	15
Resnet	[7], [8], [9], [11], [12], [13], [16] , [18], [20]	9
DenseNet	[5], [7], [9], [11], [12], [13], [15]	7
Heatmap	[4], [8], [16]	3

2.5. Goals and Outcomes

By analysing Table 2 it is possible to notice that almost all of the collected documents focus about more general terms such as Covid-19, Machine Learning or even pre trained models. During the research, a few documents were discarded due to the fact that the pandemic led to the writing of a lot of reviews, making it harder to filter it for the required topics, such as the utilization of heatmaps to identify Covid-19 on medical exams. Apart from this, it is possible to notice the relevance of technology in modern medicine, taking into account the amount of articles found in this area during the systematic review.

Authors on [13], [14], [20], [23], [24] reveal the effects of the pandemic on the respiratory system, highlighting Middle East respiratory syndrome (MERS) and severe acute respiratory syndrome (SARS). Radiological imaging methods like chest CXR and CT scans have been extensively utilized for diagnostic purposes. In comparison to CT scans, though, CXR offers advantages in terms of convenience, cost-effectiveness, and patient safety [29].

In what comes to models, the VGG family has been highly effective when it comes to classifying medical images. In this case we highlight the presence of VGG 16 and VGG19, where the first, in what comes to binary classification of Covid-19, was able to get accuracy results like 93.44% [14], 94% [22], 98.81% [23], 94.16% [24] and 93.6% [28]. VGG-16 comprises 13 convolutional layers and 3 fully-connected layers, while VGG-19 incorporates 16 convolutional layers and 3 fully-connected layers [30]. Consequently, VGG-19 is regarded as a more intricate CNN architecture when compared to VGG-16 [30]. Even though VGG16 ended up showing better results, this complexity also resulted in a higher performance, achieving values like 90% [13], 92.15% [21], 93.54% [22], 97.53% [23], 90.72% [24] and 90.8% [28].

Two other models that were found in the literature and were used for image classification, are Densenet201 and ResNet50. The first one was able to get accuracy results such as 98.7% [21], 92.5% [18], 93.02% [20] and 96,25%, proving that it comes in as one of the best models to use in picture classification. The second one, was a bit less common in the literature, due to the fact that it shows a worst performance when compared to any of the above models. As an example, in [21], it got 80% of accuracy compared to 98% with Densnet201. Despite poor results when comparing to other models, in [22], [24], [26], the authors managed to achieve over 92% of accuracy by using Resnet50.

Authors on [12], [16], [24] have tried to use a visual representation of the model in order to provide further explainability to the end user on how the classification is made, and on what aspects of an image does the model focus on. To achieve this, an heatmap was used to overlay the original image and highlight the areas that matter the most. On articles [8], [16] the authors used the Gradient-weighted Class Activation Mapping (Grad-CAM) [8] in order to get the desired results.

2.6. Conclusions

Based on the literature analysis, it has been concluded that the most effective models for classifying medical images, particularly for detecting Covid-19, are Densenet201, ResNet50, VGG19, and VGG16. These models have demonstrated high accuracy in various studies, with VGG16 and VGG19 achieving accuracy rates as high as 98.81% and 97.53% respectively. Densenet201 has also shown to be effective, with accuracies up to 98.7%, while ResNet50, though less common, still achieves over 92% accuracy in some instances. The literature highlights the extensive use of radiological imaging methods such as CXR and CT scans, with CXR being favoured for its convenience, cost-effectiveness, and safety.

Moreover, the integration of technology and AI have been proven critical in early disease detection [31]. Deep Learning, specifically, is identified as the most suitable machine learning mechanism for detecting Covid-19 from CXR, CT, and MRI images. For model explainability, studies [12], [16], [24] have employed heatmaps, such as Grad-CAM, to visually represent the areas of images that the models focus on during classification. This approach enhances the transparency and interpretability of the AI models, providing end users with a clearer understanding of the classification process. Consequently, our research will test these models - Densenet201, ResNet50, VGG19, and VGG16 - along with Grad-CAM heatmaps for visualization to validate their performance in medical image classification.

CHAPTER 3

CRISP-DM Methodology

The objective of the research is to classify CXR and to identify the impact of Covid-19 on the lungs. This study follows the CRISP-DM Methodology [7], as depicted in Figure 3, commencing with data preparation and concluding with the Deployment phase. Subsequent sections will delve into more comprehensive explanations.

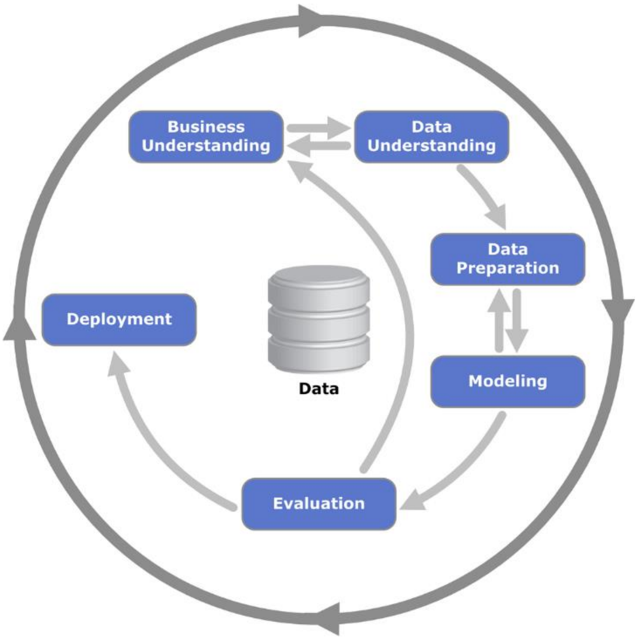

Figure 3 - CRISP-DM workflow [32]

3.1. Business and Data Understanding

In this chapter, our focus shifts to a thorough examination of the foundational elements of our dataset. We aim to gain a comprehensive understanding of its origins, composition, and intended application within our research framework.

The initial phase of this study involved securing access to imaging data under a confidentiality agreement established in collaboration with Santa Maria's Hospital. This agreement was designed to safeguard patient privacy and ensure ethical compliance in line with institutional and regulatory guidelines.

Upon initial exploration of the provided dataset, it became evident that it lacked key imaging data essential for this research. This shortfall represented a significant limitation, despite being a dataset with a lot of patient detail and information that would be important to analyse together with the images. The lack of images prompted the need to identify alternative data sources capable of addressing the research objectives.

A search for external datasets led to the identification of a publicly available repository developed by Qatar University [33]. This dataset contains a collection of 33 920 CXR images, categorized into three diagnostic classes: COVID-19 (with 11 956 images), pneumonia (with 11 263), and healthy cases (with 10 701 images). While this dataset does not include detailed patient-level metadata, offering only JPG image files and associated segmentation masks, it provides an extensive and diverse sample population critical for robust model development.

Despite its advantages, we consider the absence of patient metadata in the Qatar University dataset [33] a limitation. This lack limits analyses involving demographic or clinical variables, such as age, gender, or comorbidities. To address this, future work may integrate complementary datasets or focus on patient diagnostic patterns visible in CXR images.

Through the integration of the Qatar University dataset [33], this research gains access to a comprehensive resource for investigating pulmonary disease diagnostics. The utilization of this dataset enables the development of ML models while addressing the constraints posed by the initial dataset, thus ensuring alignment with the study's scientific and clinical objectives.

3.2. Data Preparation

The dataset utilized in this master's thesis proved to be very advantageous, as it was pre-prepared and readily accessible for analysis. Consisting of 256x256 PNG images, the dataset exhibited consistent standards throughout. This ensured an easier integration into our research framework, significantly reducing the need for extensive preprocessing tasks. With the dataset's standardized format, our focus shifted towards optimizing the data for analysis, thereby maximizing the potential for obtaining better results. Thus, our attention was devoted to fine-tuning the dataset preparation process, underscoring our commitment to extracting the most valuable insights from the images at hand.

In this work, we present a detailed analysis of CXR images, serving as emblematic representations of the diversity inherent in our dataset. With a repository comprising 11,956 COVID-19 images and 10,701 images from healthy patients, we confront a broad spectrum of anatomical variations. From differences in lung sizes to unique configurations, these images underscore the complexity of the dataset. Such diversity poses a significant challenge, demanding innovative approaches to extract precise and reliable diagnostic insights. As we work through these different details, our goal is to make the most of our dataset, trying to get the best possible results. Detailed in Figure 4, it is possible to find 4 different examples of the diversity present within our dataset.

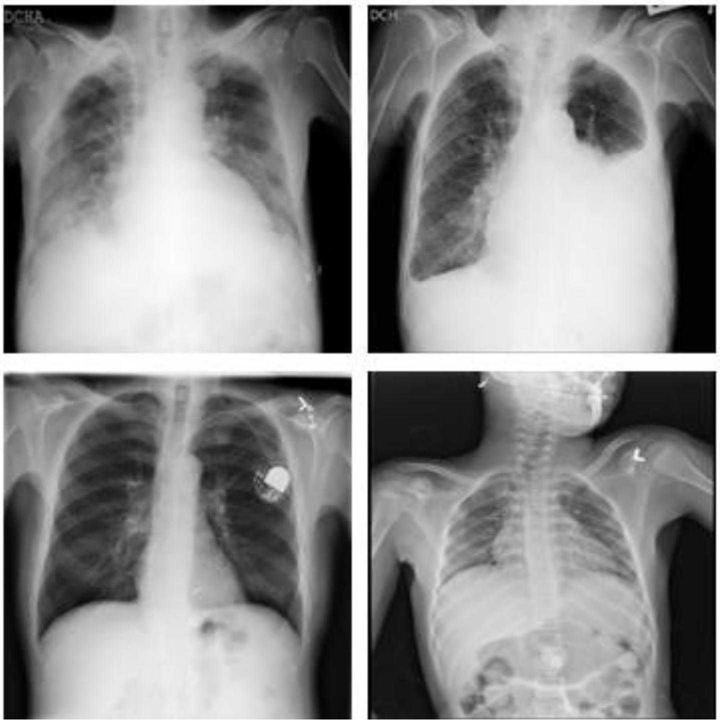

Figure 4 - Four raw images from the dataset

One aspect of this dataset is its organization, further reinforced by the inclusion of masks for every image. This feature enhances our process, eliminating the need for manual mask generation and ensuring consistency across the dataset. With masks readily available, we can bypass this time-consuming step and focus our efforts on refining our analysis techniques. By providing the model with merged images combining the original X-Rays and their corresponding masks, we aim to enhance the accuracy and reliability of our results. This approach not only expedites our workflow but also sets the stage for more precise and insightful analyses, ultimately advancing the field of medical imaging research. This step is described in Figure 5, as both image and mask are merged into one final image, that allows the model to focus only on the desired area, making the heatmap in the final stage easier to read by the final patient.

Figure 5 - Application of the mask over the original image

In our experimentation process, we employed three distinct values for the Gamma parameter of the images [34]. The primary objective was to ascertain the optimal setting that would maximize the performance of the models utilized in our research. By varying the Gamma parameter, we sought to discern how different levels of image enhancement influence the models' ability to exploit various characteristics within the images. Figure 6 illustrates the impact of Gamma parameter variation on an image, providing valuable insight into how these adjustments affect image quality and interpretability. Through this systematic exploration, we aim to gain a deeper understanding of how our models interact with enhanced image features, ultimately refining our approach to image analysis and model optimization.

In this investigation, data augmentation was consciously omitted from our methodology.

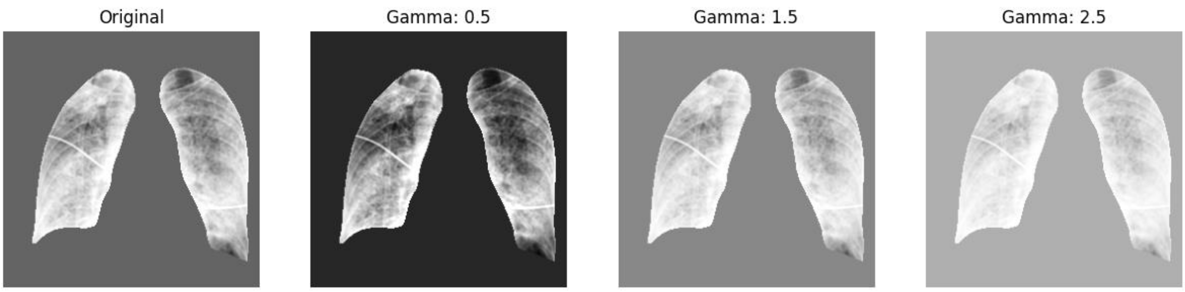

Figure 6 - Impact of the Gamma parameter on the images

Traditionally, data augmentation serves to expand datasets and enhance model training by introducing variations in images, such as rotation or amplification. However, given the ample availability of images in our dataset and the consistent orientation of future input images, we deemed data augmentation unnecessary. Our preliminary tests indicated that implementing data augmentation resulted in significantly poorer outcomes than anticipated. Therefore, we opted to forego this step, recognizing it as an unnecessary complication that could potentially detract from the model's performance. Instead, we directed our focus towards refining other aspects of our methodology to optimize model performance and achieve our research objectives efficiently.

From our analysis of a complex dataset, we have devised a two-stage approach to optimize our model's performance:

- First Stage:
  - Utilized masked images from a large dataset.
  - Experimented with different gamma parameters to evaluate their impact on model performance.
  - Tested various models using both merged and unmerged images to assess the influence of image enhancement on results.

- Second Stage:
  - Focused on the most promising model identified from the initial stage.
  - Conducted extensive parameter tuning to fine-tune the model.
  - Explored the optimal combination of parameters to achieve the best results in medical imaging analysis.

By dividing our approach into these two distinct stages, we aim to systematically enhance our model's performance and unlock its full potential in accurately analysing medical imaging data.

3.3. Modelling

In this section, we explore the core findings of our study by using CNNs to classify CXRs and detect the presence of Covid-19. Here, we offer an introduction to CNNs and highlight essential techniques to our research.

CNNs are essential to deep learning, and they perform highly on image-based tasks. Their unique architecture allows them to learn hierarchical features directly from input data [35], making them highly effective in capturing intricate patterns and feature hierarchies. CNNs have shown outstanding performance in classification problems by using convolutional layers to automatically extract features from input data, where subtle patterns are critical for accurate classification, which is the case we have in hands.

To enhance the capabilities of our dataset, we used Convolutional Neural Networks (CNNs) with transfer learning, utilizing pre-trained weights sourced from the publicly available "ImageNet" dataset [36], renowned for its extensive class coverage. Specifically, we adopted the VGG16 [35], VGG19 [38], DenseNet 201 [39], ResNet-50 [40], and Xception [41] models, selected based on established research findings. An initial base model was also used as a start point in order to compare the differences and improvements of the pre-built models. Each model was tailored to perform feature extraction, crucial for delineating distinct regions within CXR scans and accurately identifying the impact of the disease. In order to make training with our dataset more efficient, we added an additional three dense layers to each model (Figure 7). These adjustments provided intriguing outcomes that closely matched the unique feature needs of our investigation.

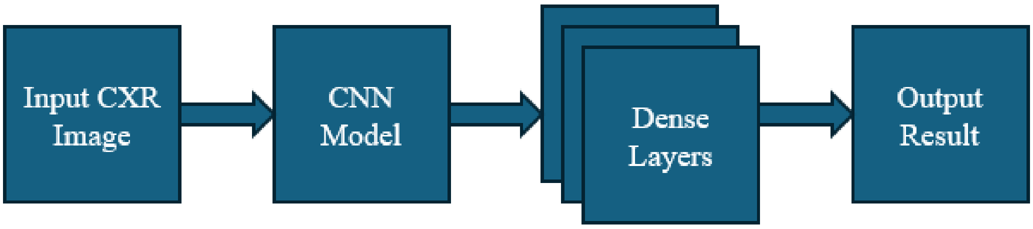

Figure 7 - Standard model architecture

To completely maximize these models' performance, we did, however, realize that some improvements were necessary. In light of our goal of binary classification, we set up the model's output layer to consist of a single neuron with sigmoid activation. To reinforce the models' performance and address concerns of overfitting, we employed a five-fold cross-validation strategy. To improve resilience and encourage generalization, this method involved rearranging the dataset and training the model five times.

Furthermore, as suggested by the literature analysis, we explored the idea of fine-tuning. Using the pre-trained model's knowledge and feature extraction skills, this strategy retrains the model's final layers while leaving the higher layers untrainable, allowing it to be more successfully tailored to our particular job of classifying Covid-19. By doing so, we not only improve the model's performance but also optimize time and computational resources.

In our study, given the amount of images available, we divided our dataset into three primary sets: validation, testing, and training, with proportions of 15%, 15%, and 70%, respectively. This stratification allowed us to effectively train and evaluate our models while ensuring robustness and generalization. The validation set was utilized during the training process to tune hyperparameters and assess model performance, while the testing set served as an independent evaluation to gauge the model's effectiveness. The largest portion, the training set, was used to train the models and optimize their parameters. Additionally, in Figure 8, we present a visualization illustrating the distribution and balance of the data between the training and validation tests, providing insights into the dataset's composition and aiding in the interpretation of our results.

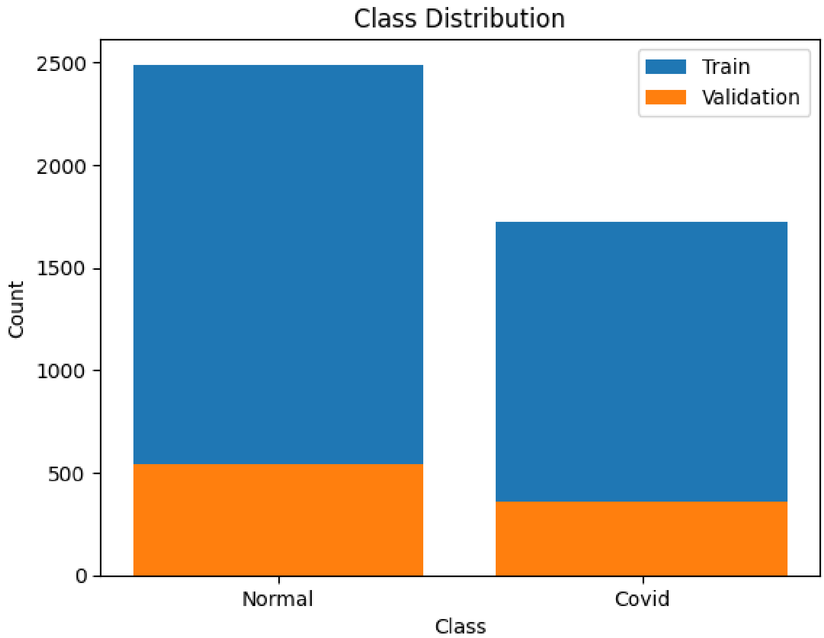

Figure 8 - Class Distribution

CHAPTER 4

Methods and Discussion

This chapter compares and evaluates the results and functionality of the six CNN models used in this study. Our evaluation is conducted using a dataset comprised of images sourced from our dataset.

To tackle this challenge and optimize our results, we adopted a two-stage approach. Initially, we looked to identify the most effective model out of the six available options. Following this selection process, we proceeded to fine-tune the chosen model, aiming to maximize its performance and achieve the best possible outcome. This systematic approach allowed us to iteratively refine our methodology and ultimately enhance the quality of our results. Despite having two steps, these parameters were stable throughout the whole process:

- The standardized shape consistently maintained was (224, 224, 3) across all stages of the process.
- The number of batches was determined to be 32 using the following formula below, where N represents the total number of samples, B denotes the batch size, and E signifies the number of epochs [42].

$$Number\ of\ Batches = N\ B * E$$

- Initially, we conducted tests using 25 epochs for the model. However, after implementing cross-validation techniques, we opted to reduce this number to 20. This adjustment was made as it was observed that the difference in performance between 20 and 25 epochs was negligible, while the use of fewer epochs helped conserve computational resources.

Finally, after identifying the best-performing model, we will employ a heatmap overlay technique [43] on the images to pinpoint the areas most affected by the disease, ensuring a visual representation of the classification.

4.1. Finding the Best Model

In this new chapter, we dig into the details of Stage 1 of our process, where we meticulously outline the steps taken to determine the most optimal model for our research. Here, we provide a comprehensive overview of the various models utilized and elucidate the criteria employed to select the best candidate. Through detailed analysis and evaluation, we aim to offer transparency into our decision-making process, ultimately laying the foundation for the subsequent stages of our investigation.

Initially, a base model was employed to establish a foundational understanding of the task at hand. However, recognizing the need to delve deeper to achieve optimal results, we proceeded to explore additional CNN models to enhance our outcomes. To this end, VGG16 [37], VGG19 [36] DenseNet201 [39], Xception [41], and ResNet50 [40] were utilized to assess their effectiveness. By systematically testing each model, we aimed to identify the one that would yield the most favourable results for our research objectives.

In this initial stage of our analysis, cross-validation was incorporated to ensure the robustness and reliability of our results. By employing cross-validation techniques, we aimed to validate the performance of our models across multiple iterations, thereby reducing the risk of overfitting and enhancing the generalizability of our findings.

Table 3 presents a comprehensive overview of the results obtained through this evaluation process, shedding light on the effectiveness of each model in addressing our research objectives.

Table 3 - Results of Cross Validation vs Normal Approach

Approach	Model	Recall	Precision	F1-Score
Normal Approach	Base Model	0,77	0,81	0,8
	ResNet 50	0,8	0,82	0,81
	DenseNet 201	0,89	0.9	0,89
	VGG 19	**0,91**	**0,92**	**0,92**
	VGG 16	0,88	0,9	0,88
	Xception	0,81	0,82	0,82
Cross Validation	Base Model	0,79	0,83	0,82
	ResNet 50	0,85	0,86	0,86
	DenseNet 201	0,88	0,89	0,88
	VGG 19	**0,93**	**0,93**	**0,93**
	VGG 16	0,88	0,9	0,89
	Xception	0,84	0.84	0,84

These results provide valuable insights into the performance of various CNN models under both the normal approach and cross-validation. In the normal approach, DenseNet201, VGG16 and VGG19 emerge as the top performers, exhibiting high recall, precision, and F1-score values, indicating their effectiveness in accurately classifying instances of the target class. It is also possible to see the gap when comparing to the other models, especially the Base Model, that confirms our need to use pre-trained models due to its poor results comparing to the remaining ones.

When employing cross-validation, VGG19 consistently outperforms the other models, achieving the highest recall, precision, and F1-score values. This suggests that VGG19 not only excels in the normal approach but also maintains its superiority when subjected to cross-validation, highlighting its reliability across different techniques. Additionally, it's observed that all models generally maintain similar performance trends between the normal approach and cross-validation, albeit with slight variations in specific metric values. Overall, these results underscore the importance of selecting the appropriate CNN model based on the specific requirements and objectives of the classification task.

In conclusion, based on the evaluation of various CNN models using both the normal approach and cross-validation, VGG19 emerges as the top-performing model. With consistently high recall, precision, and F1-score values, VGG19 demonstrates superior performance in accurately classifying instances of the target class while minimizing false positives and false negatives. This underscores its robustness and reliability across different evaluation techniques. Therefore, VGG19 stands out as the most effective model for the classification task at hand and we decided to proceed to the next stage of our investigation with it. We are confident that leveraging the capabilities of this model will enable us to obtain the most accurate and reliable results for our research objectives.

4.2. Optimizing Model Performance

Having selected VGG19 as the model for optimization, we will run a comprehensive exploration of its performance by varying the learning rate across five different parameters, ranging from 0.000001 to 0.01 (Table 4). In addition, we will implement cross-validation techniques using 5 k-folds, each comprising 20 epochs, to ensure robustness and reliability in our evaluation process. This meticulous approach aims to fine-tune the model's hyperparameters and maximize its effectiveness in accurately classifying instances of the target class.

Table 4 - Comparison of different learning rates

Approach	LR	Recall	Precision	F1-Score
Different Learning Rates	0,01	0,946	0,946	0,946
	0,001	0,971	0,971	0,971
	0,0001	**0,973**	**0,974**	**0,973**
	0,00001	0,961	0,961	0,961
	0,000001	0,898	0,904	0,897

From Table 4, the results reveal a notable impact of the learning rate (LR) on the performance of the VGG19 model. When LR is set to 0.01, the model demonstrates high recall, precision, and F1-score values of 0.946. Subsequently reducing LR to 0.001 leads to improved performance across all metrics, with recall, precision, and F1-score values of 0.971, suggesting enhanced classification accuracy. Further decreasing LR to 0.0001 results in marginal improvements in recall, precision, and F1-score values, reaching 0.973, indicating continued enhancement in classification performance. However, lowering LR to 0.00001 leads to a decrease in performance, particularly in recall and F1-score values, suggesting less effective classification performance compared to previous learning rates. Finally, setting LR to 0.000001 results in a significant decline, demonstrating a notable decrease in classification performance. Overall, the findings suggest that a learning rate between 0.001 and 0.0001 achieves optimal performance for the VGG19 model, obtaining the best classification results from the selected LR values. This way we can conclude that the best learning rate for the best model is 0.0001.

From the graph below (Figure 9), it is evident that increasing the learning rate (LR) up to 0.0001 results in improved performance in our findings. However, as the LR continues to increase beyond this point, the results begin to decline, leading to lower values across the metrics evaluated. This observation underscores the importance of selecting an appropriate LR within the optimal range to achieve the best performance in our classification task.

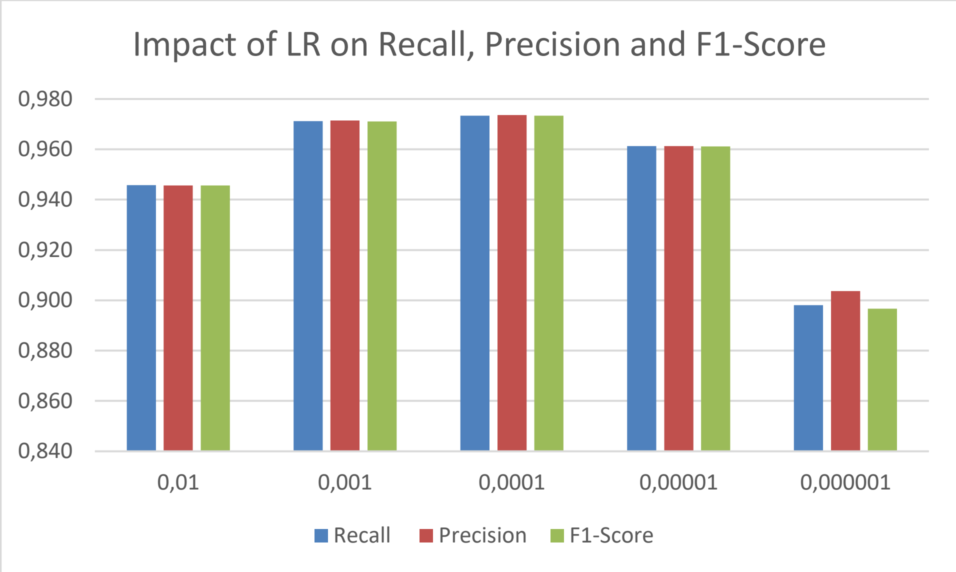

Figure 9 - The impact of Learning Rate variation

As a final validation test to ensure that our model is not overfitting, we evaluated the model using the best parameters identified on a previously untouched dataset comprising 15% of images that were neither used for training nor validation. The results obtained from this evaluation closely resemble those achieved previously, indicating consistency and robustness in our model's performance across different datasets. This consistency suggests that the model is effectively generalizing to new, unseen data, reinforcing the notion that overfitting is not a concern in our model, as can be seen from the Confusion Matrix on Figure 10 as well as from Table 5.

Table 5 - Best results

Model	LR	Recall	Precision	F1-Score
VGG19	0,0001	0,970	0,970	0,970

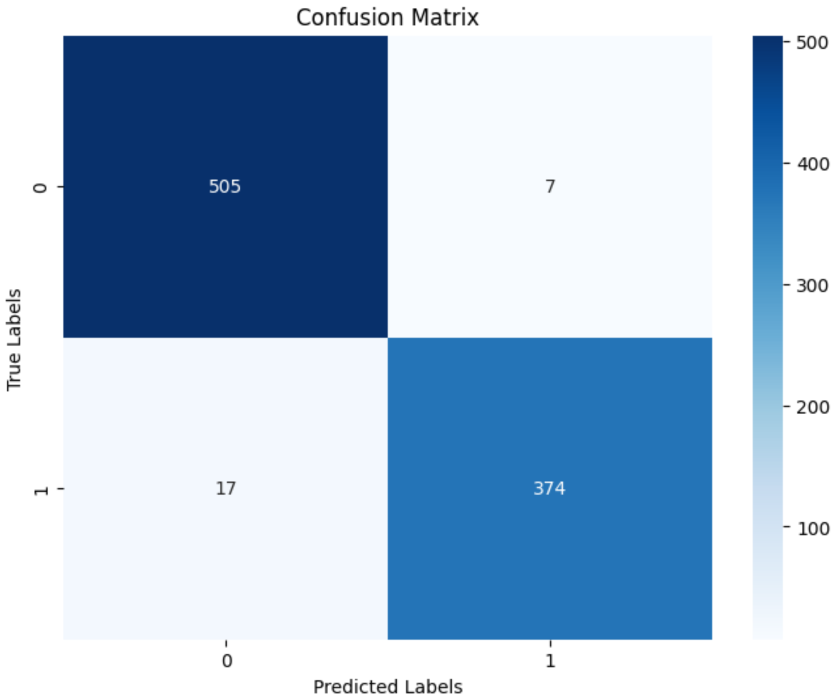

Figure 10 - Confusion Matrix

4.3. Heatmap

For the final stage of our investigation, we will proceed by applying a heatmap [44] over the images classified by our model. This heatmap will provide a visual representation of the main areas that are scrutinized by the computer during the classification process. By overlaying the heatmap onto the images, we aim to gain insights into the regions of interest identified by the model, further enhancing our understanding of its decision-making process. This visual analysis will complement our quantitative results and offer valuable interpretability into the model's predictions, ultimately contributing to the overall transparency and trustworthiness of our findings.

The heatmap process begins by generating a pixelized heatmap highlighting the most relevant points on the image, as perceived by the model. Subsequently, a superimposed picture is created by merging the original image with the heatmap. In our case, as the model is trained on masked images, we remove the image mask in the final stage to enhance the interpretability of the final image for the average viewer. This process is illustrated in Figure X and Y for both COVID and non-COVID cases, respectively. Through this visualization technique, we aim to provide a clearer understanding of the areas of interest identified by the model and facilitate the interpretation of its classifications.

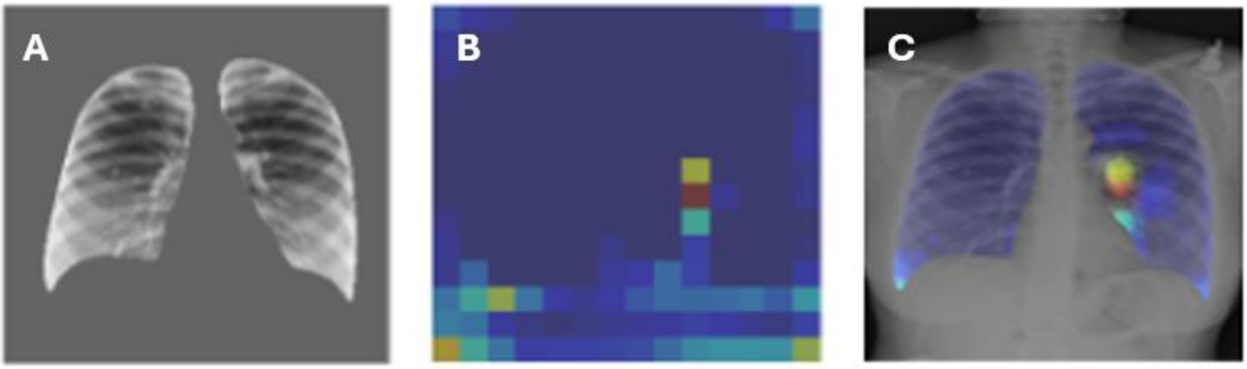

Figure 11 – (A) Healthy CXR; (B) Heatmap; (C) Superimposed image

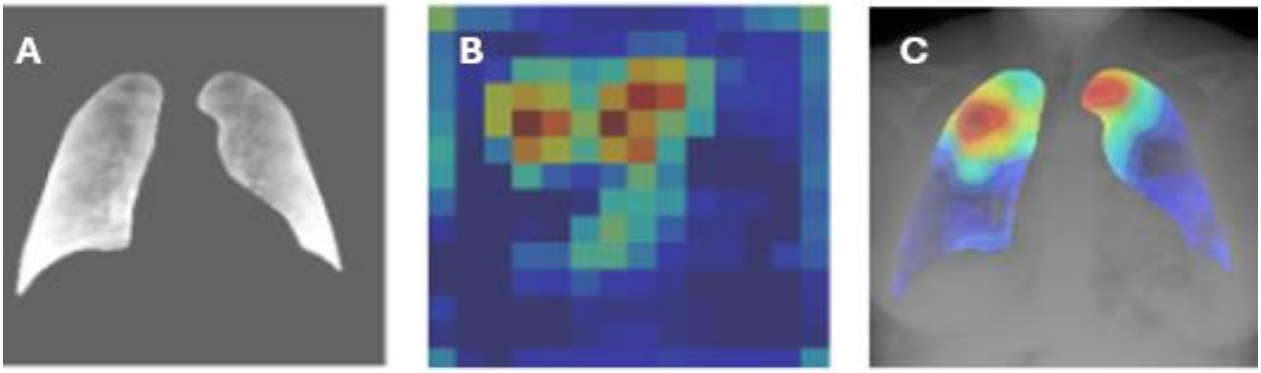

Figure 12 - (A) Covid-19 CXR; (B) Heatmap; (C) Superimposed image

Figure 13 presents a comparison of chest X-ray images analysed for Covid-19 detection using deep learning visualization techniques. The top row shows the original X-rays, with the first labelled as Covid-19 positive and the others as normal. The middle row displays feature maps from a neural network, highlighting regions within each X-ray that the model identifies as significant. Finally, the bottom row overlays these activation maps on the original images, with colours indicating the level of model attention—red areas suggest higher importance. In the Covid-19 positive case, there is a noticeable concentration of red and yellow in the lungs, potentially corresponding to Covid-19-specific abnormalities. This visualization helps demonstrate how the model distinguishes between Covid-19 and normal cases based on specific lung features.

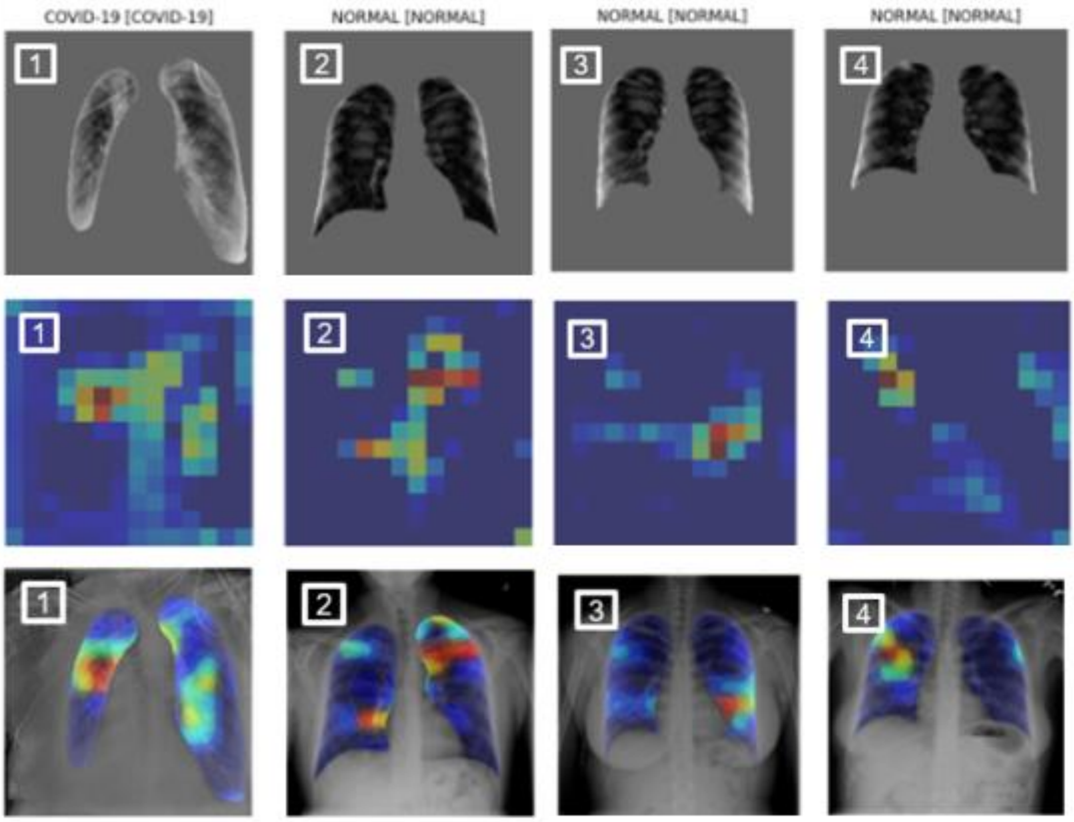

Figure 13 - Model-identified regions of interest in Covid-19 positive and normal cases.

CHAPTER 5

Conclusion and Future Work

5.1. Conclusion

The central aim of this dissertation was to effectively classify and visually identify Covid-19 in Chest X-Rays. To achieve this objective, machine learning techniques and pre-trained models were utilized, enabling us to attain the highest possible level of accuracy and efficacy in our analysis.

Despite initially intending to source data solely from Hospital de Santa Maria, the necessity of utilizing a new dataset unexpectedly resulted in significantly improved overall results. This unexpected shift provided substantial advantages, particularly concerning the application of the heatmap method.

Our analysis revealed that the VGG19 model emerged as the top performer, achieving 97% recall rate. Furthermore, the optimal learning rate was identified to be 0.0001, further enhancing the model's effectiveness. Other models also demonstrated notable performance, particularly considering the origin of the dataset. Models such as VGG16 and DenseNet201 achieved remarkably high values, further validating the efficacy of our approach. Cross-validation played a pivotal role in attaining the desired results by mitigating overfitting throughout the study. This technique effectively reinforced the reliability of our findings and improved all results during the initial stage of the investigation.

This study underscores the importance of integrating science and technology in the modern era. AI emerges as a pivotal force shaping various aspects of our daily lives. Particularly from a medical perspective, the utilization of AI to assist in disease identification and classification through medical exams represents a significant advancement. This integration holds the potential to mitigate human error and enhance diagnostic accuracy, paving the way for more efficient and effective healthcare practices in the years ahead.

5.2. Future Work

Future work in this investigation will involve deploying the identified best-performing model in various scenarios with diverse sets of images. This exploration will allow us to assess the model's robustness and generalizability across different contexts and datasets. By subjecting the model to a range of scenarios, including different image types and environments, we can evaluate its performance under varying conditions and uncover any potential limitations or areas for improvement. Additionally, conducting real-world validation tests with new data sources will provide valuable insights into the model's applicability in practical settings. Overall, this continued experimentation will contribute to enhancing the model's reliability and effectiveness in real-world applications.

Additionally, incorporating an extra model to predict masks will be crucial for addressing unknown datasets that may differ in accessibility from the one used in this study. This mask prediction model will enhance the adaptability of our approach to diverse datasets and ensure accurate classification even in unfamiliar environments.

Given our “narrow-down” approach to find the best model, we did not use techniques such as Ensemble Modelling [45]. This would be good in order to use multiple models (for example, the ones used in the first stage of our approach) in order to predict an outcome.

Finally, the ultimate step would involve deploying this model in a real-life scenario to fulfil its mission of aiding both doctors and patients in better understanding and classifying medical images, ideally in HSM, which was the original plan for this work. Our model's capabilities can be used to increase diagnostic accuracy, expedite workflows, and ultimately improve patient care by integrating it into clinical practice. This real-world application represents the culmination of our research efforts and underscores the potential impact of our findings on healthcare delivery.

References

[1] European Parliament, Directorate-General for Parliamentary Research Services, and P. Boucher, *Artificial intelligence – How does it work, why does it matter, and what we can do about it?* European Parliament, 2020. doi: 10.2861/44572.

[2] "Listings of WHO's response to COVID-19." Accessed: Nov. 06, 2022. [Online]. Available: https://www.who.int/news/item/29-06-2020-covidtimeline

[3] "Most COVID-19 hospitalizations due to four conditions," National Institutes of Health (NIH). Accessed: Dec. 06, 2022. [Online]. Available: https://www.nih.gov/news-events/nih-research-matters/most-covid-19-hospitalizations-due-four-conditions

[4] M. I. Razzak, S. Naz, and A. Zaib, "Deep Learning for Medical Image Processing: Overview, Challenges and Future." arXiv, Apr. 22, 2017. doi: 10.48550/arXiv.1704.06825.

[5] A. M. Rahmani *et al.*, "Automatic COVID-19 detection mechanisms and approaches from medical images: a systematic review," *Multimedia Tools and Applications*, vol. 81, no. 20, pp. 28779–28798, 2022, doi: 10.1007/s11042-022-12952-7.

[6] J. S. M. Kim, J. Z. Dong, S. Brener, P. C. Coyte, and Y. R. Rampersaud, "Cost-Effectiveness Analysis of a Reduction in Diagnostic Imaging in Degenerative Spinal Disorders," *Healthc Policy*, vol. 7, no. 2, pp. e105–e121, Nov. 2011.

[7] H. Palacios, R. Toledo, G. Hernandez, and A. Navarro, "A comparative between CRISP-DM and SEMMA through the construction of a MODIS repository for studies of land use and cover change," *Advances in Science, Technology and Engineering Systems Journal*, vol. 2, pp. 598–604, Jun. 2017, doi: 10.25046/aj020376.

[8] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice: IEEE, Oct. 2017, pp. 618–626. doi: 10.1109/ICCV.2017.74.

[9] J. Saltz, "CRISP-DM for Data Science Teams: 5 Actions to Consider," Data Science Process Alliance. Accessed: May 27, 2024. [Online]. Available: https://www.datascience-pm.com/crisp-dm-for-data-science-teams-5-actions-to-consider/

[10] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, Mar. 2021, doi: 10.1136/bmj.n71.

[11] N. Hotz, "What is CRISP DM?," Data Science Process Alliance. Accessed: Nov. 25, 2023. [Online]. Available: https://www.datascience-pm.com/crisp-dm-2/

[12] A. A. Barbhuiya, R. Kumar Karsh, R. Jain, and H. Phukan, "A deep neural network-based architecture for automated detection of COVID-19 from chest X-ray images," *International Journal of Medical Engineering and Informatics*, vol. 16, no. 2, pp. 173–184, 2024, doi: 10.1504/IJMEI.2024.136964.

[13] C. C. Ukwuoma *et al.*, "Deep learning framework for rapid and accurate respiratory COVID-19 prediction using chest X-ray images," *Journal of King Saud University - Computer and Information Sciences*, vol. 35, no. 7, 2023, doi: 10.1016/j.jksuci.2023.101596.

[14] M. N. Toroghi, U. U. Sheikh, and S. S. Irani, "Classification of COVID-19 and lung opacity using vision transformer on chest x-ray images," presented at the Journal of Physics: Conference Series, 2023. doi: 10.1088/1742-6596/2622/1/012016.

[15] M. V. L. Costa, E. J. De Aguiar, L. S. Rodrigues, J. S. Ramos, C. Traina, and A. J. M. Traina, "A Deep Learning-based Radiomics Approach for COVID-19 Detection from CXR Images using Ensemble Learning Model," presented at the Proceedings - IEEE Symposium on Computer-Based Medical Systems, 2023, pp. 517–522. doi: 10.1109/CBMS58004.2023.00272.

[16] K. Imagawa and K. Shiomoto, "Performance Change with the Ratio of Training Data A Case Study on the Binary Classification of COVID-19 Chest X-Ray by using Convolutional Neural Networks," presented at the 2023 IEEE International Symposium on Medical Measurements and Applications, MeMeA 2023 - Conference Proceedings, 2023. doi: 10.1109/MeMeA57477.2023.10171929.

[17] L. Kong and J. Cheng, "Classification and detection of COVID-19 X-Ray images based on DenseNet and VGG16 feature fusion," *Biomedical Signal Processing and Control*, vol. 77, 2022, doi: 10.1016/j.bspc.2022.103772.

[18] G. U. Nneji *et al.*, "Fine-Tuned Siamese Network with Modified Enhanced Super-Resolution GAN Plus Based on Low-Quality Chest X-ray Images for COVID-19 Identification," *Diagnostics*, vol. 12, no. 3, 2022, doi: 10.3390/diagnostics12030717.

[19] B. Dwarakanath, G. Shrivastava, R. Bansal, P. Nandankar, V. Talukdar, and M. A. Usmani, “Explainable Machine Learning Techniques in Medical Image Analysis Based on Classification with Feature Extraction,” *International Journal of Communication Networks and Information Security*, vol. 14, no. 3, pp. 342–357, 2022.
[20] R. Shakil, B. Akter, F. M. Javed Mehedi Shamrat, N. Jahan, S. Hasan, and A. Khater, “Systematic Analysis of Several Deep Learning Approaches for COVID-19 Detection Using X-ray Images,” presented at the 3rd International Conference on Smart Electronics and Communication, ICOSEC 2022 - Proceedings, 2022, pp. 1301–1307. doi: 10.1109/ICOSEC54921.2022.9952147.
[21] T. Makarovskikh, A. Salah, A. Badr, A. Kadi, H. Alkattan, and M. Abotaleb, “Automatic classification Infectious disease X-ray images based on Deep learning Algorithms,” presented at the 2022 8th International Conference on Information Technology and Nanotechnology, ITNT 2022, 2022. doi: 10.1109/ITNT55410.2022.9848538.
[22] R. Biju, W. Patel, K. Suresh Manic, and V. Rajinikanth, “Framework for Classification of Chest X-Rays into Normal/COVID-19 Using Brownian-Mayfly-Algorithm Selected Hybrid Features,” *Mathematical Problems in Engineering*, vol. 2022, 2022, doi: 10.1155/2022/6475808.
[23] W. Abdulsalam Hamwi and M. M. Almustafa, “Development and integration of VGG and dense transfer-learning systems supported with diverse lung images for discovery of the Coronavirus identity,” *Informatics in Medicine Unlocked*, vol. 32, 2022, doi: 10.1016/j.imu.2022.101004.
[24] G. S. Tang, L. S. Chow, M. I. Solihin, N. Ramli, N. F. Gowdh, and K. Rahmat, “Detection of COVID-19 Using Deep Convolutional Neural Network on Chest X-Ray (CXR) Images,” presented at the Canadian Conference on Electrical and Computer Engineering, 2021. doi: 10.1109/CCECE53047.2021.9569064.
[25] S. Liaqat and G. Raja, “Computer-Aided detection of COVID-19 using chest imaging,” *IET Conference Proceedings*, vol. 2021, no. 1, pp. 170–175, 2021, doi: 10.1049/icp.2021.1458.
[26] V. S. N. Mehar *et al.*, “COVID-19 PREDICTION THROUGH CHEST X-RAY IMAGE DATASETS USING DEEP LEARNING,” presented at the 2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021, 2021. doi: 10.1109/ICCCNT51525.2021.9579985.
[27] E. Peláez, R. Serrano, G. Murillo, and W. Cárdenas, “A comparison of deep learning models for detecting COVID-19 in chest X-ray images,” presented at the IFAC-PapersOnLine, 2021, pp. 358–363. doi: 10.1016/j.ifacol.2021.10.282.
[28] R. Mohammadi, M. Salehi, H. Ghaffari, A. A. Rohani, and R. Reiazi, “Transfer learning-based automatic detection of coronavirus disease 2019 (COVID-19) from chest x-ray images,” *Journal of Biomedical Physics and Engineering*, vol. 10, no. 5, pp. 559–568, 2020, doi: 10.31661/jbpe.v0i0.2008-1153.
[29] R. S. Saeed and B. K. O. Chabor Alwawi, “A binary classification model of COVID-19 based on convolution neural network,” *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 3, pp. 1413–1417, 2023, doi: 10.11591/eei.v12i3.4832.
[30] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition.” arXiv, Apr. 10, 2015. doi: 10.48550/arXiv.1409.1556.
[31] M. M. Ahsan and Z. Siddique, “Machine learning based disease diagnosis: A comprehensive review.” arXiv, Dec. 31, 2021. Accessed: Jun. 01, 2024. [Online]. Available: http://arxiv.org/abs/2112.15538
[32] “Cross Industry Standard Process for Data Mining,” *Wikipédia, a enciclopédia livre*. Feb. 27, 2024. Accessed: May 27, 2024. [Online]. Available: https://pt.wikipedia.org/w/index.php?title=Cross_Industry_Standard_Process_for_Data_Mining&oldid=67545979
[33] “COVID-QU-Ex Dataset.” Accessed: Nov. 26, 2023. [Online]. Available: https://www.kaggle.com/datasets/anasmohammedtahir/covidqu
[34] A. K. Jain, *Fundamentals of digital image processing*. USA: Prentice-Hall, Inc., 1989.
[35] M. Maashi *et al.*, “Modeling of Reptile Search Algorithm With Deep Learning Approach for Copy Move Image Forgery Detection,” *IEEE Access*, vol. 11, pp. 87297–87304, 2023, doi: 10.1109/ACCESS.2023.3304237.
[36] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “ImageNet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255. doi: 10.1109/CVPR.2009.5206848.
[37] G. Learning, “Everything you need to know about VGG16,” Medium. Accessed: Jan. 18, 2023. [Online]. Available: https://medium.com/@mygreatlearning/everything-you-need-to-know-about-vgg16-7315defb5918

[38] "VGG-19 Convolutional Neural Network - an overview | ScienceDirect Topics." Accessed: May 20, 2024. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/vgg-19-convolutional-neural-network

[39] "densenet201 — Torchvision 0.18 documentation." Accessed: May 20, 2024. [Online]. Available: https://pytorch.org/vision/stable/models/generated/torchvision.models.densenet201.html

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

[41] B. Gülmez, "A novel deep neural network model based Xception and genetic algorithm for detection of COVID-19 from X-ray images," *Ann Oper Res*, vol. 328, no. 1, pp. 617–641, Sep. 2023, doi: 10.1007/s10479-022-05151-y.

[42] R. Pramoditha, "All You Need to Know about Batch Size, Epochs and Training Steps in a Neural Network," Data Science 365. Accessed: May 20, 2024. [Online]. Available: https://medium.com/data-science-365/all-you-need-to-know-about-batch-size-epochs-and-training-steps-in-a-neural-network-f592e12cdb0a

[43] "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization | IEEE Conference Publication | IEEE Xplore." Accessed: May 15, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/8237336

[44] K. Team, "Keras documentation: Grad-CAM class activation visualization." Accessed: May 20, 2024. [Online]. Available: https://keras.io/examples/vision/grad_cam/

[45] "Ensemble Modeling - an overview | ScienceDirect Topics." Accessed: Jun. 29, 2024. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/ensemble-modeling