UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

NOEMI MARITZA LAPA ROMERO

COVID-VR: A Deep Learning COVID-19 Classification Model Using Volume-Rendered Computer Tomography

Thesis presented in partial fulfillment
of the requirements for the degree of
Master of Computer Science

Advisor: Prof. Dr. João Luiz Dihl Comba

Porto Alegre
April 2022

CIP — CATALOGING-IN-PUBLICATION

Lapa Romero, Noemi Maritza

COVID-VR: A Deep Learning COVID-19 Classification Model Using Volume-Rendered Computer Tomography / Noemi Maritza Lapa Romero. – Porto Alegre: PPGC da UFRGS, 2022.

85 f.: il.

Thesis (Master) – Universidade Federal do Rio Grande do Sul. Programa de Pós-Graduação em Computação, Porto Alegre, BR–RS, 2022. Advisor: João Luiz Dihl Comba.

1. Deep Learning. 2. Classification Models. 3. Computer Tomography. 4. Volume Rendering. 5. COVID-19. I. Dihl Comba, João Luiz. II. Título.

“True education is a kind of never ending story –
a matter of continual beginnings,
of habitual fresh starts, of persistent newness.”
— J. R. R. TOLKIEN

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer profundamente ao Professor João Comba pela oportunidade de trabalhar ao seu lado desde que me aceitou como aluna de mestrado. Por todo o conhecimento compartilhado, pelas tarefas dadas durante o desenvolvimento das aulas e projetos que me ajudaram a aprender muitas mais coisas do que eu imaginava. E acima de tudo, pelos conselhos, motivação e dedicação que pude apreciar ele mostra com seus alunos e estudos.

Agradeço também a todos os professores com quem consegui trabalhar ou ter aulas durante esses anos, por estarem sempre dispostos a compartilhar seus conhecimentos. Assim como aos meus colegas de laboratório e de projetos pela ajuda e sugestões constantes. A todos os amigos que conheci nesta cidade e aos amigos que estão longe também, vossa companhia tem sido muito importante nos últimos anos. Em especial a Vanessa por sua inestimável amizade, apoio e carinho de tantos anos, obrigada demais.

Por fim, gostaria de agradecer eternamente a minha família, meus pais Pablo e Maria que deram tudo pela nossa família e cuidaram de nós com muito amor e dedicação, os admiro muito por tudo que conquistaram e continuam conquistando. Às minhas irmãs Jenny e Julisa que sempre estiveram comigo e são meu maior orgulho, alegria e força; desejo ser o mesmo para elas. A todas minhas primas e familiares em geral por sempre me aconselharem e me apoiarem nas decisões que tomei, por me receberem com tanto carinho sempre que nos vemos; e me dar forças nestas últimas semanas para superar a perda da nossa maravilhosa avó Vicenta. A todos, muito obrigado.

Dissertação de Mestrado - INF/UFRGS

RESUMO

A pandemia da COVID-19 trouxe diversos desafios aos sistemas de saúde em todo o mundo. Como a maioria dos pacientes com COVID-19 tem infecções pulmonares, uma tomografia computadorizada (TC) do tórax se mostra eficiente na identificação de infecções por COVID-19, bem como outras classes de doenças pulmonares. Arquiteturas de redes profundas surgiram para identificar automaticamente classes de doenças pulmonares, usando as fatias de TCs como entrada para modelos de classificação. Este trabalho propõe COVID-VR, uma nova abordagem para classificar COVID-19 baseado na classificação de imagens tiradas de diferentes ângulos da renderização do volume dos pulmões, fornecendo assim uma visão global de todo o pulmão em cada imagem. Comparamos nossa proposta com as principais estratégias concorrentes com soluções de código aberto disponíveis, usando dados privados de hospitais parceiros e dados disponíveis publicamente. Os resultados mostram que nossa abordagem identifica lessões pulmonares de COVID-19 com sucesso e é competitiva em relação aos métodos baseados em fatias de TC. Embora nossos experimentos tenham sido focados em dados do COVID-19, nossa solução é extensível a outras doenças pulmonares.

Palavras-chave: Deep-learning, Modelos de Classificação, Tomografia Compuratizada, Renderização de Volume, COVID-19.

ABSTRACT

The COVID-19 pandemic brought several challenges to health systems worldwide. Since most patients with COVID-19 have lung infections, a Computer Tomography (CT) of the chest is often used to identify COVID-19 infections, as well as other classes of pulmonary diseases. Deep-learning architectures surfaced to automatically identify classes of pulmonary diseases, using the slices of CTs as inputs to classification models. This work proposes COVID-VR, a novel approach for classifying COVID-19 based on volume rendering images of the lungs taken from different angles, thus providing a global view of the entire lung in each image. We compared our proposal against leading competing strategies with available solutions, using private data from partner hospitals and publicly available data. Results show that our approach successfully identifies COVID-19 pulmonary lesions and is competitive against slice-based methods. Although our experiments were focused on COVID-19 data, our solution is extensible to other pulmonary diseases.

Keywords: Deep Learning. Classification Models. Computer Tomography. Volume Rendering. COVID-19.

LIST OF ABBREVIATIONS AND ACRONYMS

COVID-19	Coronavirus Disease 2019
CAP	Community-acquired pneumonia
RSNA	Radiology Society of North America
BSTI	British Society of Thoracic Imaging
CO-RADS	COVID-19 Reporting and Data System
COVID-RADS	COVID-19 imaging reporting and data system
DICOM	Digital Imaging and Communications in Medicine
NIfTI	Neuroimaging Informatics Technology Initiative
HU	Hounsfield Unit
COVID-VR	COVID-19 Volume Rendering Framework
COVNet	Coronavirus disease 2019 (COVID-19) detection neural network
DeCoVNet	3D deep Convolutional Neural Network to Detect COVID-19
P-HNN	Progressive Holistically Nested Networks
CAAE	Certificado de Apresentação de Apreciação Ética

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

The COVID-19 pandemics, caused by SARS-CoV-2, have been widely spread, reporting more than 540 million cases and making more than 6,32 million fatal victims around the world between December 2019 and June 2022[1]. This scenario leads to health system overload in multiple countries, as the disease diagnosis is often expensive, time-consuming, or requires specialized laboratories and doctors. The Real-Time Polymerase Chain Reaction (RT-PCR) test, a widely used method to diagnose the infection, has considerable drawbacks, such as the well-known possibility of false negatives (AREVALO-RODRIGUEZ et al., 2020).

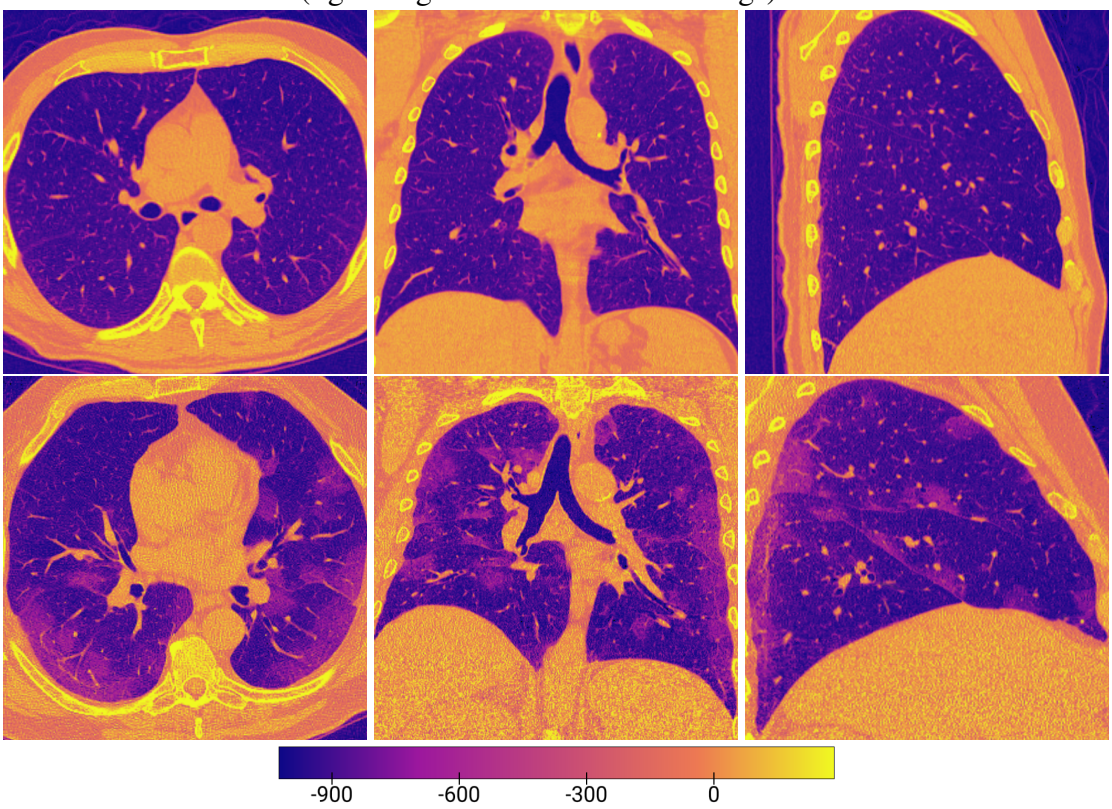
Chest computed tomography (CT) is a fast, non-invasive, painless, and accurate exam that can be helpful in the diagnosis of diseases that include chest symptoms. Although not all COVID-19 cases include lung infection, chest CT scanning constitutes an accurate method of diagnosing the target disease. Still, that method requires around 20 minutes of specialized radiologist time. Aiming to increase the efficiency of COVID-19 diagnosis, a variety of works presenting Machine Learning (ML) solutions emerged. The potential for using ML techniques comes from observations by radiologists about common patterns found in CTs, which are frequently associated with finding Ground-Glass Opacities (GGOs).

A GGO is a CT finding defined as an area of increased attenuation in the lung. It can be seen as a gray or hazy area, and it is non-specific, with wide etiology. In the COVID-19 context, there are well categorized pulmonary lesion patterns that are identified as typical of that disease. The GGO often caused by COVID-19 infection is characterized by being peripheral, bilateral, multifocal, rounded, possibly appearing with consolidation, etc. As an example, Figure 1.1 shows the comparison of two chest CTs (axial, coronal, and sagittal views) varying the colormap range in the HU values between -1000 and 300; one normal patient (with no signs of pneumonia) and another presenting those typical patterns of COVID-19 lung infection previously cited, the lesions are the light magenta areas inside the lungs. Finally, we observe that the disease cases may or may not cause a lung infection. Additional difficulties encountered in using ML for COVID-19 classification are the low availability of public datasets and the public models proposed.

In the present work, we propose a 3D Volume Rendering framework, COVID-VR, which includes the step of Lung Segmentation and utilizes a ResNet architecture (HE et

[1] https://coronavirus.jhu.edu/map.html

Figure 1.1: Axial, Coronal, and Sagittal CT slices of a normal (top) and a COVID-19 (bottom) patient. A color map was applied to highlight ground-glass opacities associated with COVID-19 lesions (light magenta areas inside the lungs).



al., 2016) as the backbone. The proposed model applied in two main cases (ternary and binary classification task) achieved an accuracy of 90.8% and an F1-score of 90.8% for ternary classification in the COVID-CT-MD (AFSHAR et al., 2021) public dataset (using classes COVID-19, Normal, and CAP). An accuracy of 92.2% with an F1-score of 87.2% in the binary classification COVID-19 vs. Others (Negative, Indeterminate and Atypical CT images) task; and accuracy of 96.1% with an F1-score of 96.3% in the binary classification task COVID-19 vs. Normal (Negative) task using COVID-19 class as the positive class in our private dataset given by Hospital das Clínicas de Porto Alegre (HCPA) and Hospital Moinhos de Vento (HMV) hospitals.

2 BACKGROUND

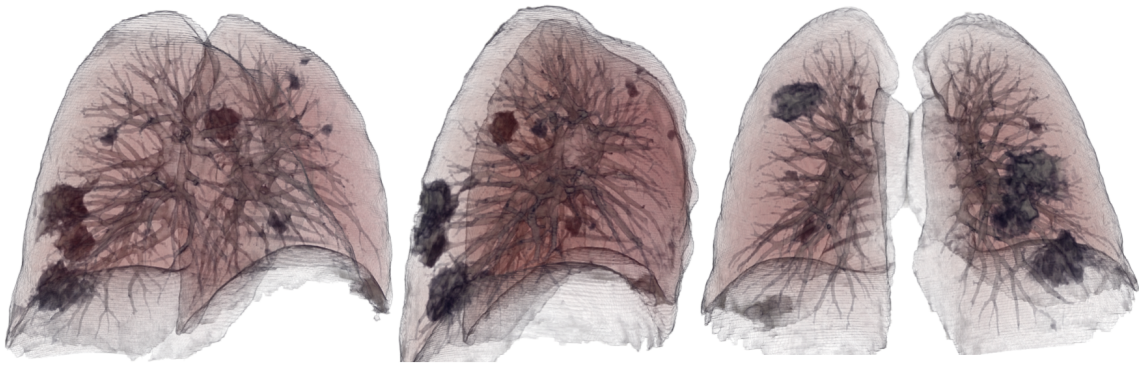
This chapter describes in more detail the concepts and definitions needed for a better understanding of this work.

2.1 CT Features

CT is a diagnostic image test that provides detailed images of organs, bones, soft tissues, and blood vessels using special X-Ray equipment. Chest CTs are very useful in the diagnosis of the cause of unexplained chest symptoms. Some of the worth-mentioning applications of that scanning test are to diagnose lung cancer and COVID-19. In our research, we used only the pulmonary window, not the mediastinal window. In the mediastinal window, the lungs are overexposed and appear black. On the other hand, in the pulmonary window, the lung tissues can be seen in detail, this includes vascular structures and anomalies (SIMPSON, 2009).

The Hounsfield unit (HU) is a dimensionless scale used to express CT numbers in a standardized form. It is obtained from a linear transformation of the measured absorption/attenuation of the X-Ray beam. As the physical density of each tissue is proportional to its coefficient, the HU unit has many well-known uses. The HU ranges that contain the lungs and ground-glass opacities will be discussed further.

Figure 2.1: Volume rendering of the segmented lung of a COVID-19 patient from three different views. Ground-glass opacities associated with COVID-19 lesions correspond to darker regions inside the lungs. The volume rendering image allows identifying the extension of lesions in the entire lung, while 2D slices indicate lesions in planar sections of the lung.



2.2 Classification Radiology Standards

Identifying typical patterns in CT images allowed the formalization of standards for the classification. Taking into consideration those well-known patterns, different radiology standards were developed in the pandemic context, such as the one proposed by the Radiology Society of North America (RSNA) (SIMPSON et al., 2020) and the British Society of Thoracic Imaging (BSTI) and other standards like CO-RADS (COVID-19 Reporting and Data System), or the COVID-RADS (COVID-19 imaging reporting and data system) as guidelines for the COVID-19 classifications from CT images.

The RSNA standard proposes four categories based on CT image features: Typical appearance, Indeterminate appearance, Atypical appearance, and Negative for COVID-19 pneumonia. The first classification corresponds to lesions associated with ground-glass opacities, usually at the walls of the lungs, such as the ones illustrated in Figure 2.1 and 2.2a. Indeterminate classification suggests the absence of typical features, plus the presence of multifocal, diffuse, perihilar, nonperipheral, nonrounded, or unilateral ground-glass opacity with or without consolidation, as seen in Figure 2.2b. Atypical classification denotes the absence of the previous classification features and presence of isolated consolidation like in Figure 2.2c without ground-glass opacities, lung cavitation, or discrete small nodules, and the latter classification appoints that the patient presents no signs of pneumonia as shown in Figure 2.2d.

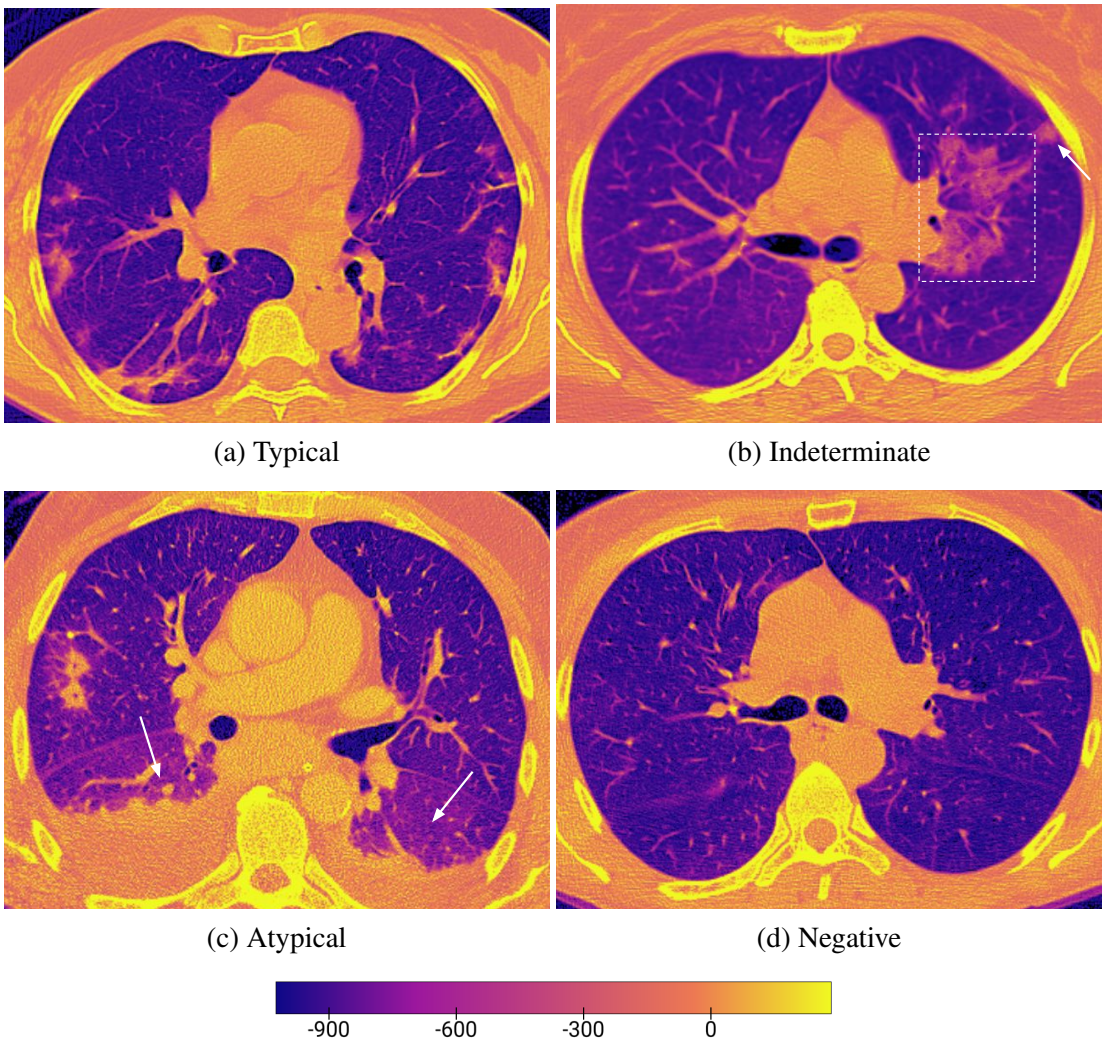
Since we only use the RSNA standard in our private dataset, we briefly describe the BSTI. We refer the reader to (INUI et al., 2020) for more details with alternative radiology standards like CO-RADS or COVID-RADS, and comparisons of Chest CT Grading Systems. The BSTI classification is made up of four classes:

- *CLASSIC COVID-19*, 100% of confidence for COVID-19.
- *PROBABLE COVID-19*, between 71% and 99% of confidence for COVID-19.
- *INDETERMINATE COVID-19*, less or equal to 70% of confidence for COVID-19.
- *NON-COVID-19*, 70% of confidence for an alternative disease.

2.3 Lung Segmentation

Using Image Segmentation, it is possible to separate images into segments of interest. Image segmentation has a wide field of applications such as object identification,

Figure 2.2: RSNA classification for COVID-19 using plasma colormap for [-1000, 300] range in HU scale.



(a) Typical (b) Indeterminate

(c) Atypical (d) Negative

content-based image retrieval, organ, and injury segmentation, etc. Numerous works reviewed in this dissertation (WANG et al., 2020; LI et al., 2020b; JADHAV et al., 2022; AMARA et al., 2022; YANG et al., 2021; HEIDARIAN et al., 2021) carry out, in their pre-processing stage, the segmentation of the lungs before the training of their models to manage to separate the background from the lung (everything that does not correspond to the region of interest). To achieve lung segmentation tasks, the literature explored in this work uses neural networks; however, it should be noted that other efforts like the one introduced by ALTIS (SOUSA et al., 2019) to address this task using Image Foresting Transform (FALCAO; STOLFI; LOTUFO, 2004).

The segmentation models used in this work receive as input the image or volume (set of images in DICOM or NIfTI format) and return a mask with each region identified with a different value. For example, P-HNN (HARRISON et al., 2017) introduces a neural network based on the VGG16 network that receives a volumetric CT image (NIfTI

format in our experiments) and returns a volumetric probability mask also in NIfTI format with the probabilities of each voxel being pulmonary region. Subsequently, they made a round of the values considering a threshold to specify what is a region of interest or not, returning a mask based on that threshold; if we choose a threshold of 0.75 as the probability for a voxel to be considered a lung region, all values greater than 0.75 will be 1. The rest will be 0 (where 0, black, is considered background and 1, white, is considered lung region). Finally, to get the segmented lung, it is enough to make a pixel-wise multiplication between the original image and the binary mask. In Figure 2.3, we depict this process.

We explore other techniques such as Lungmask (HOFMANNINGER et al., 2020a) that, through the training of neural networks, returns masks for each lung, as can be seen in Figure 2.4; and also lobe masks, unlike binary masks which contain 0s and 1s, these masks add more labels (2, 3, etc.) to denote different regions of interest depending on the number of classes needed. UBC (LENSINK et al., 2020), proposes L3-Net, a network based on 2D U-Net (RONNEBERGER; FISCHER; BROX, 2015), and PSPNet (ZHAO et al., 2017) as the backbone of its architecture, which offers masks with lung segmentation (differentiating right from left), pleura effusion, and Pure GGO, GGO w/ Smooth

Figure 2.3: Simplified pipeline for P-HNN to obtain segmented lung. In (A) receive the formatted volumetric image to proceed to the segmentation step. (B) shows the binary mask generated by the P-HNN approach. Finally, (C) presents the original image (A) cropped by the binary mask (B) to get a volumetric image with only the region of interest.

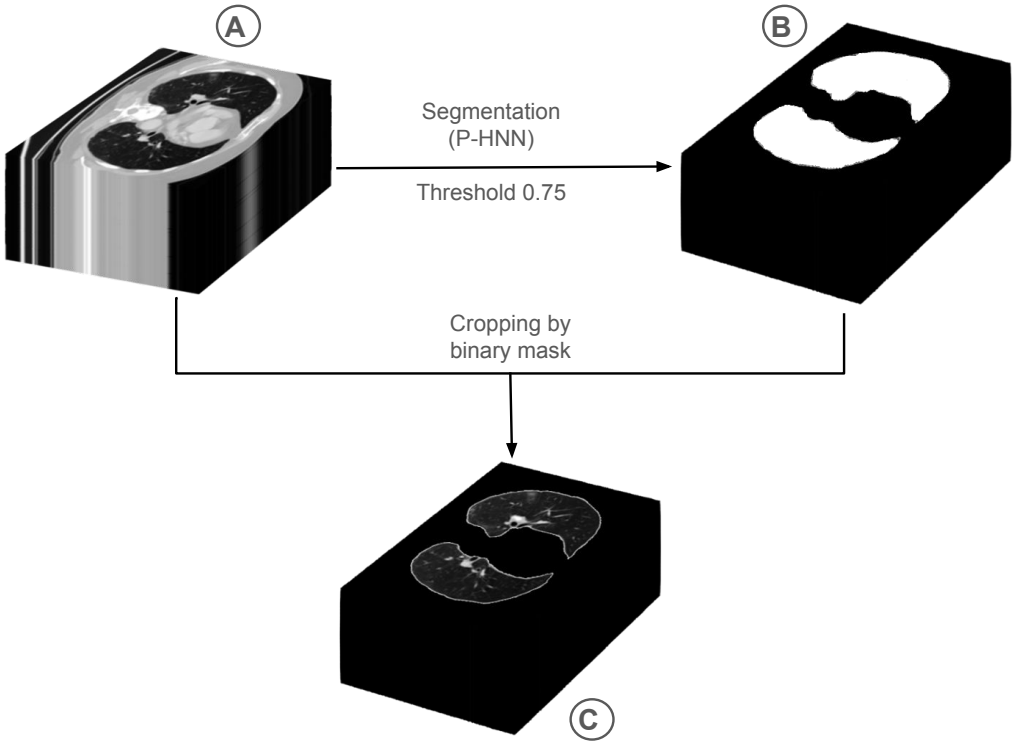
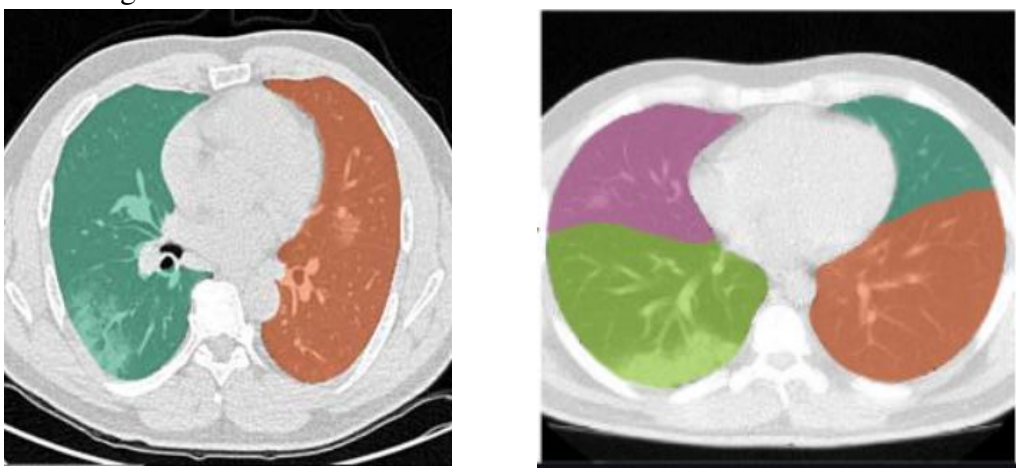


Figure 2.4: Lung and lobes masks overlapping original CT images to highlight regions of interest in images



Lungs segmentation Lobes segmentation

Interlobular Septal Thickening, Consolidation, etc.

3 RELATED WORK

In this section, we provide a review of related works. We restrict our analysis to solutions that use CT images as inputs to their classification model and have a patient-level result or explored volume rendering in COVID-19 context.

3.1 Deep Learning-based models for COVID-19 diagnosis from CT images

Since the start of the COVID-19 pandemic, several works described ML techniques to automatically classify chest radiography or CT scan images to make COVID-19 diagnostic less time-consuming.

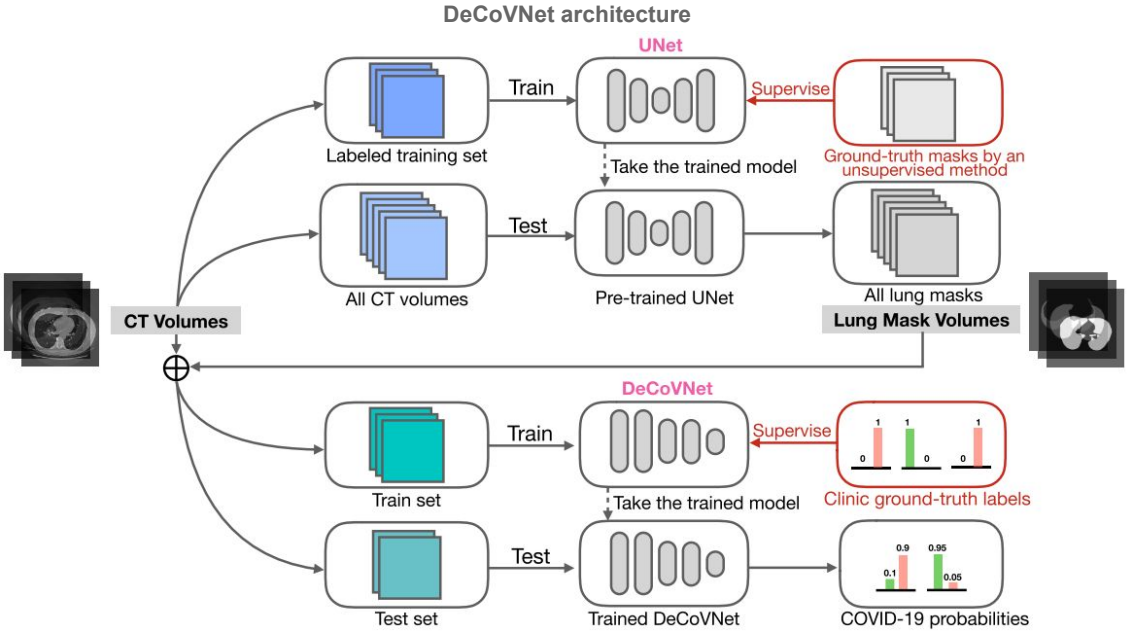
Even though both standards detailed in Section 2.2 propose a separation into four distinct classes, the first models to appear in the literature were either binary classifying in COVID-19 or non COVID-19 or ternary classifying in COVID-19, Normal (*i.e.,* with no perceptible lesions) or Community-Acquired Pneumonia (CAP). One good place to start reviewing many related works published so far is the survey by ROBERTS et al.. Their systematic review narrowed down 2212 studies initially found to 62 papers. The survey points out common pitfalls that work in this area are subjected to, arguing that none of them had successfully solved the proposed task due mainly to the lack of possible comparisons to other works and reproducibility of the experiments since most proposed works do not provide access to the training data utilized. Another recurring problem is the lack of annotated data because many of the reviewed models depend on slice-level annotations, which are more challenging to obtain in real situations.

Among the revised papers, we have a special interest in those that use CT images as inputs to their classification models and a 3D approach to classifying at the patient-level (JIN et al., 2020; WANG et al., 2020; LI et al., 2020b; HE et al., 2020; WANG; LIN; WONG, 2020; GOZES et al., 2020; JIN et al., 2020; HAN et al., 2020).

We highlight below the main competitors to our approach that has code available, thus allowing us to perform comparisons. The CT classification models reported early in the pandemic were binary and aimed to estimate the probability of a CT scan to be classified as COVID-19. DeCoVNet (WANG et al., 2020), as shown in Figure 3.1, uses a weakly-supervised 3D deep convolutional network. In a pre-processing step, DeCoV-Net receives CT images in DICOM format and performs the re-scale operation to obtain isometric volumes of $1\times1\times1$mm and normalizes the HU scale in the range [-1200, 600]

returning a 3D array for the following steps. Next, they use a U-NET to segment the lung and create a 3D binary mask. The original pre-processed volume and its 3D lung mask are sent to DeCoVNet, consisting of three stages: 3D convolution, 3D residual blocks, and a progressive classifier. The code for DeCoVNet is publicly available[1], thus allowing us to include it in our tests.

Figure 3.1: Architecture proposed by WANG et al.. A U-NET network performs lung segmentation generating a binary mask for the CT image. Afterward, the binary mask and CT volume are the inputs for the DeCoVNet model. Figure taken from Wang et al. (2020)

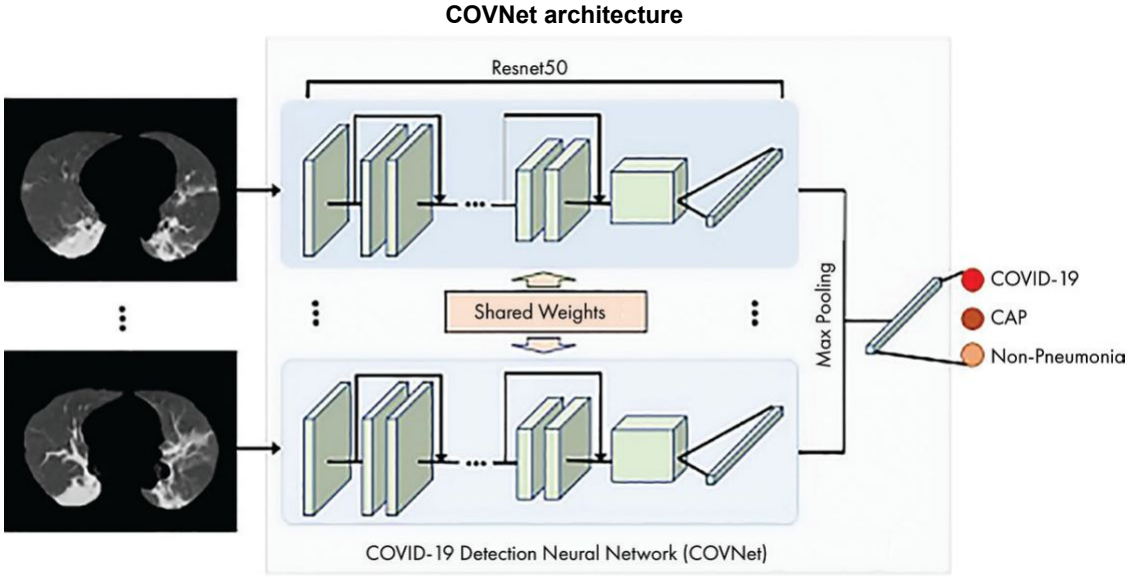


Ternary models soon followed binary models to allow the separation of COVID-19 from normal cases and other types of CAP. COVNet (LI et al., 2020b) uses a 3D deep learning framework that takes 3D slices as input and has ResNet50 as backbone architecture, generating features and combining them using a max-pooling operation, and finally generating a probability score for each class. Like DeCoVNet, COVNet also relies on a pre-processing step that uses a U-NET to create a segmented lung and mask for the lung, later resampled to the same spacing between slices of 1mm in the z-axis, downsampled five times, and scaled to S$\times$224$\times$224, where S is the number of slices after the downsampling. Since code for COVNet is also available[2], we include comparisons in our results. Figure 3.2 graphs their proposed architecture.

We include comparisons against the winner of the IEEE ICASSP 2021 Signal Processing Grand Challenge(SPGC) in the COVID-19 competition (top ten results in Ta-

[1]DeCoVNet source code: <https://github.com/sydney0zq/covid-19-detection>
[2]COVNet source code: <https://github.com/bkong999/COVNet>

Figure 3.2: Architecture proposed by LI et al.. COVNet uses as backbone ResNet50 and receives as input the CT volume and binary mask, then processes each CT slice, finally aggregating them by a max-pooling operation to produce a feature map fed to a fully connected layer returning the probability of each class. Figure taken from Li et al. (2020b)



ble 3.1). We will refer to this competition as SPGC 2021. SPGC 2021 brought an opportunity to perform comparisons among methods by providing a public dataset of CT images (COVID-CT-MD described by Afshar et al. (2021)) and a contest of performing classification in 3 possible classes: Normal, COVID-19, or (CAP). The six best-performing solutions for this competition were later presented at the ICASSP conference (CHAUDHARY et al., 2021; GARG et al., 2021; YANG et al., 2021; XUE; ABHAYARATNE, 2021; BOUGOURZI et al., 2021; LI et al., 2021), and we summarize their results and approaches below.

Table 3.1: Results of top ten methods for SPGC-ICASSP Competition 2021.

Team	Accuracy	COVID-19 Sens.	CAP Sens.	Normal Sens.
TheSaviours	90.00%	85.71%	90.00%	94.29%
IITDelhi	88.89%	88.57%	90.00%	88.57%
LLSCP	87.78%	80.00%	90.00%	100.00%
UniSheff_EEE	85.56%	82.86%	80.00%	91.43%
CNRIEMEN	81.11%	91.43%	45.00%	91.43%
Deepcam	81.11%	82.86%	100.00%	68.57%
FushionSTAR	80.00%	91.43%	30.00%	97.14%
Bingo	80.00%	88.57%	35.00%	97.14%
SAIVT	80.00%	68.57%	95.00%	82.86%
Winter	76.67%	85.71%	70.00%	71.43%

The first place (team TheSaviours (CHAUDHARY et al., 2021), we will use 'TheSaviours' as the name to refer to this approach and authors) achieved an accuracy of 90%.

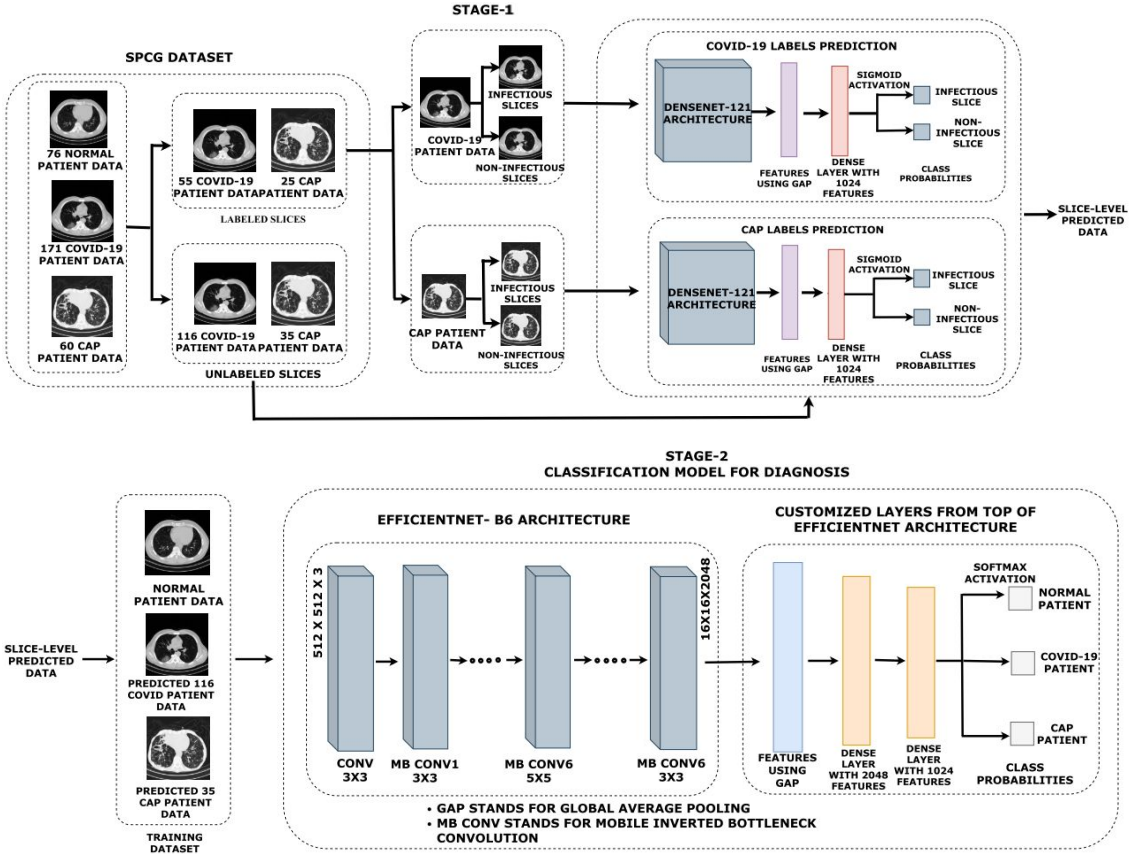
The received training CT images were not fully classified at the slice level as detailed by Afshar et al. (2021). Thus, TheSaviours describes an approach of two stages. In the first stage, as shown in stage-1 in Figure 3.3, they trained two networks that use DenseNet121 architecture as a backbone with a dense layer additional and sigmoid activation for binary classification with different tasks, the first one to classify images with the presence of COVID-19 infection against no infection; and the second to classify images with the presence of CAP infection against no infection. With these trained networks, they finish the annotation of the remaining slices. In the next stage, they perform a 90/10 split for training and validation for task classification with the set of annotated images obtained from Stage 1. They train a ternary model, as shown in stage-2 in Figure 3.3, that uses the EfficientNet-B6 architecture as the backbone, adding two dense layers of 2048 and 1024 features, ending with the softmax activation to classify images in COVID-19, CAP, and Normal classes. Finally, as this is an image-level approach to obtain a classification at the patient level, this approach classifies the 80 central images of the CT image. It proceeds to make a weighted sum giving greater weight to the 40 middle images, thus obtaining the final classification per patient. Since source code is public available[3] we included this approach in our comparisons.

The second place (team IITDelhi (GARG et al., 2021)) achieved an accuracy of 88.89%. Their approach follows a three-level approach. In the first level, they use a slice-level classifier that performs feature extraction from all the slices of the CT to learn different sizes of infection. The next level performs a patient-level classifier, using four classifiers designed to distinguish between infected and normal slices. The last level uses an ensemble-learning that combines the scores of the previous level classifiers.

The third place (team LLSCP (YANG et al., 2021)) achieved an accuracy of 87.78%. They use a multi-stage progressive learning approach composed of a 3D ResNet module, an ensemble binary classifier, and a final combining stage. The fourth place (team UniShef_EEE (XUE; ABHAYARATNE, 2021), achieved an accuracy of 85.56%, using a 3D Resnet50 classifier. The fifth place (team CNRIEMEN (BOUGOURZI et al., 2021)) achieved an accuracy of 81.11%). Their approach uses a 2-stage classifier composed of a slice-level classifier followed by a patient-level classifier. The sixth place (team FushionSTAR (LI et al., 2021)) achieved an accuracy of 80%. Their approach consists of an ensemble classifier composed of a slice classifier followed by a supervised-learning sequence classifier.

[3]TheSaviours source code: <https://github.com/shubhamchaudhary2015/ct_covid19_cap_cnn/>

Figure 3.3: Architecture proposed by team TheSaviours with two-stage. In stage 1, two proposed networks based on DenseNet121 fully annotated the slices; in stage 2, an EfficientNet B6 network to classify CT slices into COVID-19, CAP, and Normal class. Figure is taken from Chaudhary et al. (2021)



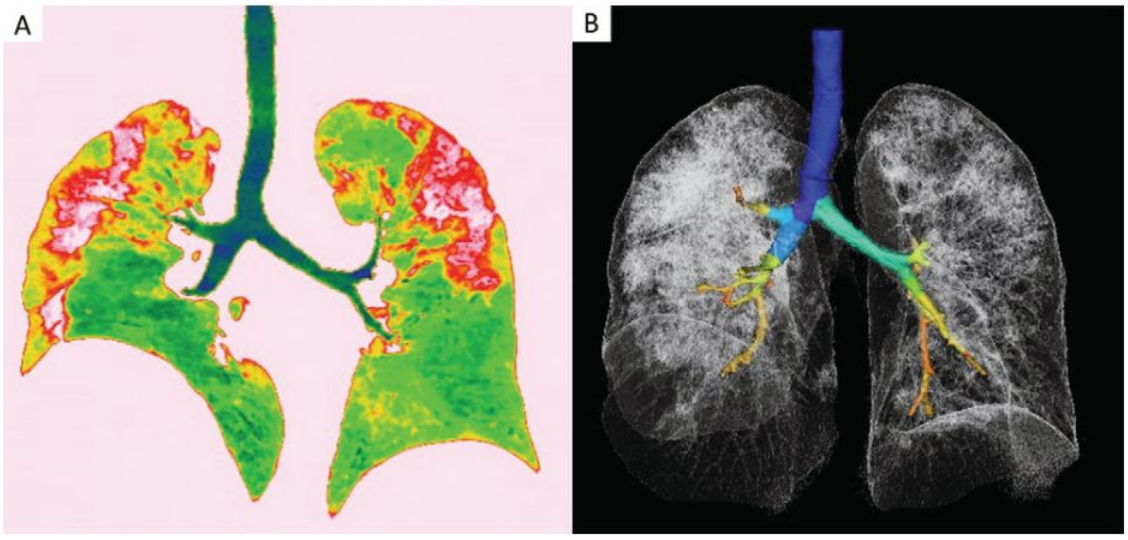
3.2 Volume Rendering Applications in COVID-19

Among other methods, including Multiplanar Reformation and Surface Rendering, Volume Rendering techniques are widely used in medical image visualization (ZHANG; EAGLESON; PETERS, 2011). Surface Rendering provides a 3D view from the surface of an object. Since this technique focuses on the surface of volumes, almost 90% of the data is lost. On the other hand, internal structures can be better observed through techniques that preserve the internal object information, such as Volume Rendering (DALRYMPLE et al., 2005). Volume Rendering combines opacity mapping and lighting effects through multiple rendering techniques (ARENS; DOMIK, 2010; J$^{\ddot{\text{i}}}$ONSSON et al., 2014; KNISS; KINDLMANN; HANSEN, 2002) to allow the appreciation of spatial relationships between structures. Cinematic rendering of the lungs is discussed in (SILVA; GELLADA, 2020) to produce more realistic images from CT and magnetic resonance data.

Despite the widespread application of Volume Rendering in medical fields, few studies addressed this technique in the COVID-19 context. Tang et al. (2020a) introduces

one of the first efforts visualizing COVID-19 pneumonia, exhibiting the lung lesions in a color coronal image and a tridimensional volume rendering of the lungs, bronchus, and trachea from a 54 years old patient shown in Figure 3.4 showing that volume rendering succeeds in highlighting lung infections. Li et al. (2020a) shows the advantage of 3D volume rendering to detect the extent of small pulmonary vessel microangiopathy and alveolar damage from an autopsy of a COVID-19 patient.

Figure 3.4: Images of a 54-year-old woman with severe COVID-19 pneumonia. (A) The red regions are the distribution of lesions in the patient. (B) Three-dimensional volume rendering image of lung and lesions. Figure taken from Tang et al. (2020a)



COVID-view (JADHAV et al., 2022) describes a system for COVID-19 diagnosis that combines a classification system with explainable visualizations of activations over 2D slices and volume rendering of the lungs. The user interface as shown in Figure 3.5 allows the user to modify the transfer functions used for volume rendering, either by choosing from a preset transfer function or by creating their own. They also use a coronal clipping tool that allows better inspection of the inner structures inside the lungs. However, their proposed classification model does not rely on volume rendering images but uses 2D slices instead. Its pipeline as shown in Figure 3.6 receive as input the 2D slices from CT images. Then, proceeding to the stage of Lung Segmentation, which uses the method proposed by Hofmanninger et al. (2020b) which also contains a model for lob lung segmentation. In the next stage, COVID-19 Lesion Localization, they included and adapted the model from Shamim et al. (2022) for COVID-19 lesion segmentation task. They proposed a model based on ResNet18 for their slice-based deep Multi-Instance Learning (MIL) for the classification step. Finally, in the Visualization Step, they integrate all the results from previous steps, offering 2D and 3D views of the segmented lungs, and

Grad-CAM for the CT slices of the activation maps.

Figure 3.5: Overview of COVID-View system. (a) The user interface for 2D and 3D views. (b) Classification result. (c) Axial view and Grad-CAM applied. (d) Lung volume-rendered showing the thick interlobular septation (pink surface-like structure in the right lung). The figure is taken from Jadhav et al. (2022)

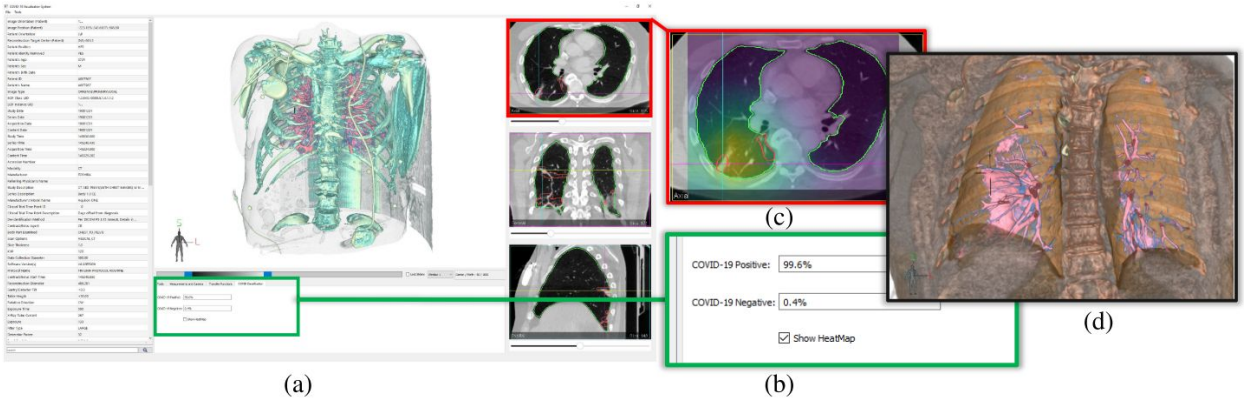
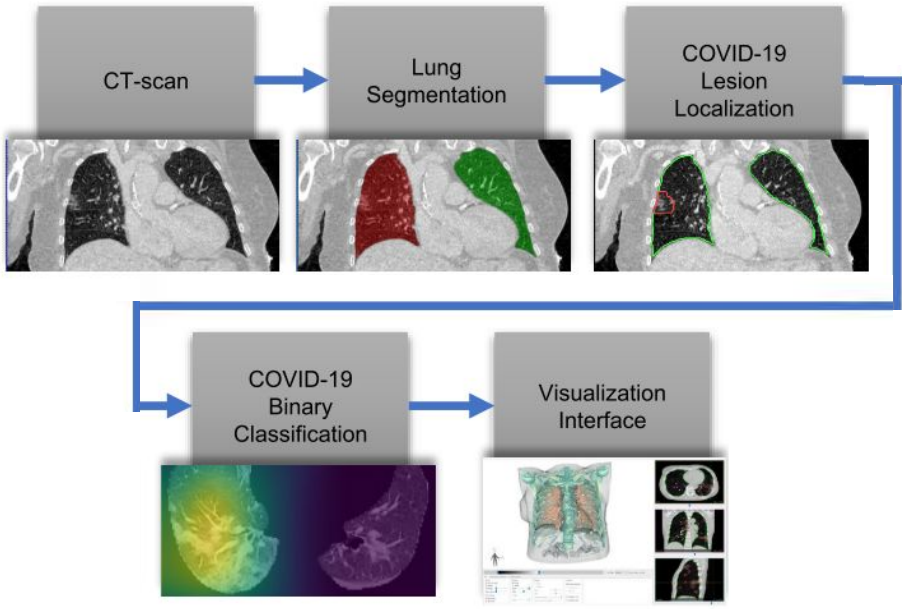


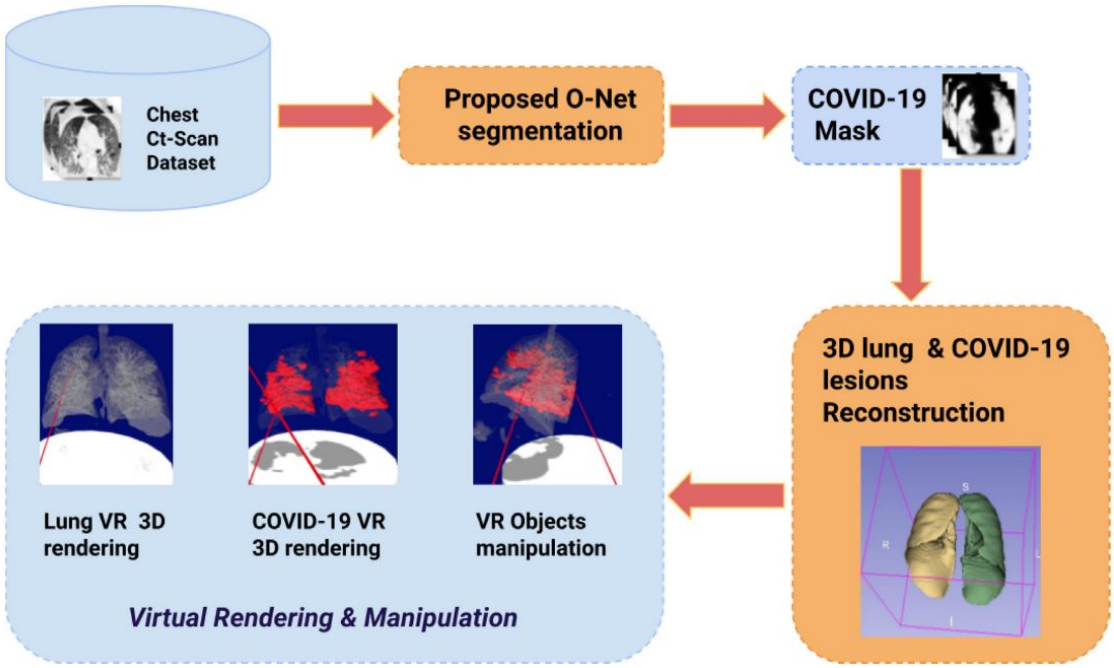
Figure 3.6: Proposed pipeline in COVID-View. The figure is taken and adapted from Jadhav et al. (2022)



Similarly, COVIR (AMARA et al., 2022) proposes a classification model based on 2D slices and a virtual reality platform to explore 3D reconstructed lungs and segmented infected lesions caused by COVID-19. Figure 3.7 presents the pipeline proposed, starting from the CT images in DICOM format as input and pre-processing these images, such as resize and border cropping. Then, they proposed a new model architecture named O-Net that consists of convolutional auto-encoders based on U-NET architecture generating lung and COVID-19 lesion segmentation. The 3D visualization relies on software such as Blender to fix problems with the models and Unity to produce the final renderings and

support the interaction through VR devices.

Figure 3.7: Proposed pipeline in COVIR. The figure is taken and adapted from Jadhav et al. (2022)



Finally, from the observed related works and as Roberts et al. (2021) details, many works fail in the reproducibility of their results due to the lack of public source code or the required annotations at the slice level and are found in private datasets. On the other hand, some works exploit volume rendering that was only used in the final stage to show/justify results. COVID-VR proposes to use these volumetric visualization techniques to train a classifier on images generated using both a private and public dataset requiring only patient-level annotations.

4 DATASETS

This study used COVID-19-related CT datasets from three different sources. The first source is the *public dataset* of CT scans prepared for the SPGC-ICASSP competition for positive and negative COVID-19 patients from Tehran, Iran. The other two sources are partner hospitals from the Rio Grande do Sul (Brazil) that provided CT scans for COVID-19 and non-COVID-19 cases. These datasets are referred to as our *private datasets*. All images were provided in DICOM format composed of studies with different spacing between slices, pixel intensity, origin source, etc. In what follows, we explain in more detail these datasets.

4.1 COVID-CT-MD Public Dataset

The COVID-CT-MD public dataset (AFSHAR et al., 2021) was released for the SPGC-ICASSP competition and was adopted for model development in several related works (*e.g.,* (CHAUDHARY et al., 2021; GARG et al., 2021)), thus allowing the comparison of our approach about state-of-the-art methods. The COVID-CT-MD dataset is composed of CT scans with patient-level annotations divided into three classes: confirmed positive COVID-19, Normal, and CAP cases. The COVID-19 cases were collected from February to April 2020, while CAP and normal cases were collected from April 2018 to December 2019 and January 2019 to May 2020, respectively. Although slice-level annotations are available for some patients, they were not explored in our approach. The COVID-CT-MD dataset is composed of 307 labeled CT scans that are used for model training and validation based on a stratified random split: 30% of these CT scans are randomly selected as the validation set, and the remaining are used as the train set. All patients were adults recruited from the Babak Imaging Center in Tehran, Iran, and exam labeling was conducted by three experienced radiologists as explained in (AFSHAR et al., 2021). Table 4.1 shows the number of CT scans per class.

In addition to the train/validation set, the SPGC 2021 competition released the SPGC-COVID Test Set (HEIDARIAN et al., 2021), with four independent test sets for models' evaluation. Three were used to calculate the competition's results and are applied for performance assessment in our work to enhance comparability with previous approaches. We do not use the fourth Test Set because it was not used to obtain the competition's winner, and we would not have fair comparisons concerning the results reported

Table 4.1: Public dataset: COVID-CT-MD, number of CT images obtained from each source, and distribution by class.

Class	**Train/Validation (307)**			**Test (90)**		
	F	M	Total	F	M	Total
COVID-19	63	108	171	9	26	35
Normal	36	40	76	15	20	35
CAP	26	34	60	7	13	20

in Table3.1.

Each of the three test datasets contains 30 CT scans, as described by HEIDARIAN et al.. The first set (Test Set 1) comprises COVID-19 and Normal cases (15 and 15, respectively), obtained from the same image center as the train/validation set. The second set (Test Set 2) contains the three classes, COVID-19, Normal, and CAP, with 10 cases for each one, obtained from another imaging center (Tehran Heart Center, Iran) using different scanner and scanner parameters. The third set (Test Set 3) contains the three classes, COVID-19, Normal (without lesions caused by lung infections), and CAP, with ten samples for each one, collected in the same imaging center as Test Set 1. We note that Test Set 2 differs from the others because it includes patients with a history of cardiovascular diseases and surgeries. Some of these differences in test sets are shown in Figure 4.1, where images corresponding to Test 1 have more noise, a cardiovascular-related complication in Test 2, and high image quality and contrast in Test 3. The distribution of samples per class for the SPGC-COVID Test Sets is also given in Table 4.1 and Table 4.2 present some important parameters for image quality and characteristics.

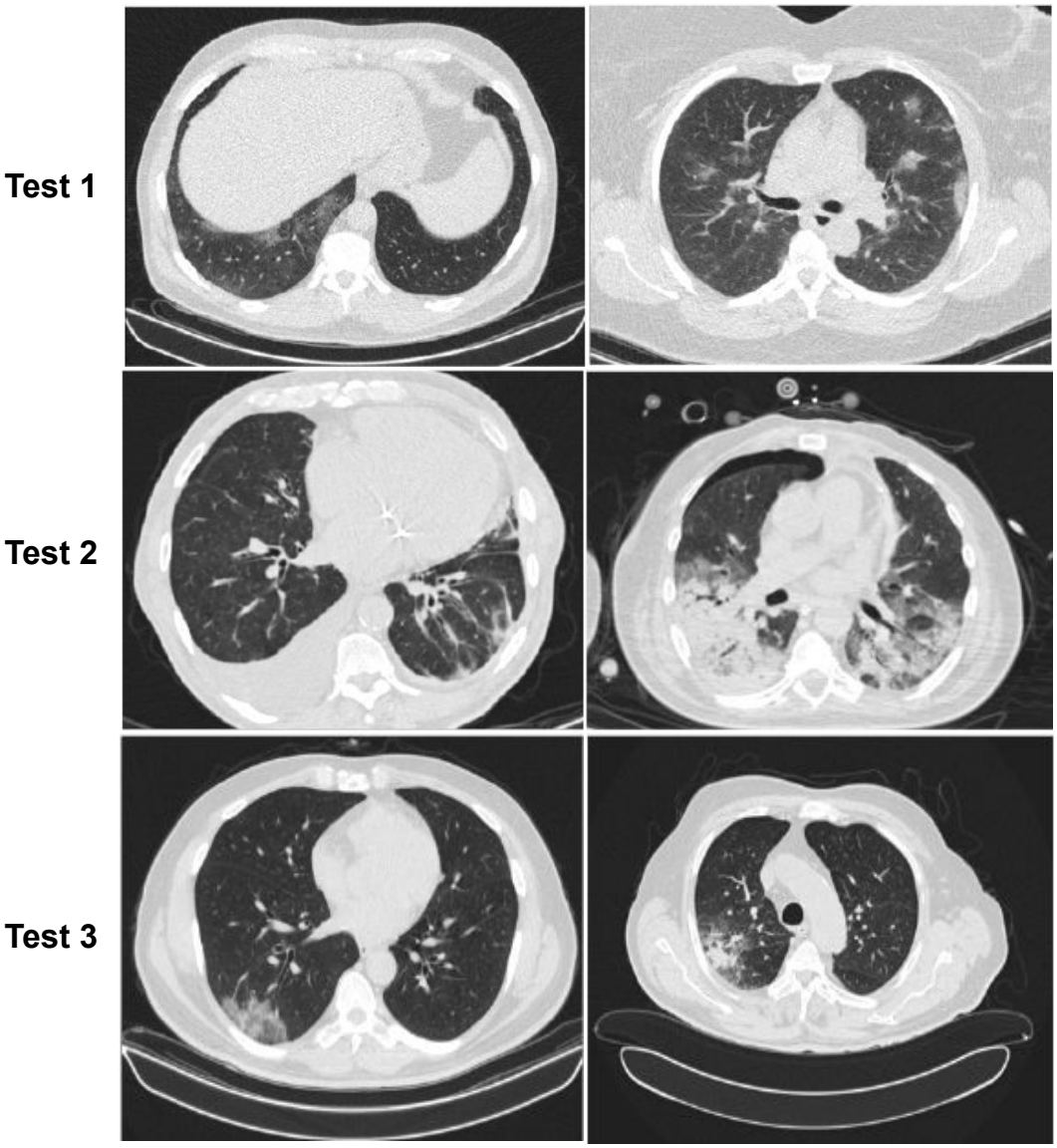
Table 4.2: Adquisition parameters used to obtain dataset COVID-CT-MD. Table taken from HEIDARIAN et al.

Dataset	**Slice Thickness (mm)**	**Reference Exposure (mAs)**	**kVp (kV)**	**Radiation (mSv)**	**Number of slices (per patient)**
Train/Val.	2	50	110−130	∼7	68−195
Test 1	2	15−20	110	∼0.3−1.5	126−169
Test 2	1.5−5	25	110−130	∼2	53−221
Test 3	2	50	100−110	∼7	115−183

4.2 Private Datasets

The private datasets were retrospectively obtained from patients admitted to two Brazilian hospitals in Porto Alegre: a private institution, Hospital Moinhos de Vento

Figure 4.1: Sample CT slices from the first three test sets. In Test set 1, the noise level is high. In Test 2, some cases reveal cardiovascular-related complications. In Test 3, the image quality and contrast are higher compared to other test sets. Taken from HEIDARIAN et al.



(HMV), and a public institution, Hospital de Clínicas de Porto Alegre (HCPA). The Research Ethics Committee approved the study of the participating hospitals (HCPA-CAAE: 32314720.8.0000.5327, HMV-CAAE: 32314720.8.3001.5330), and informed consent was waived due to the study's retrospective nature. All data were de-identified by providers to ensure patient privacy.

The HMV dataset contains 284 CT scans from patients admitted between March and May 2020, whereas the HCPA dataset comprises 105 CT scans collected from March to June 2020. Both datasets have patient-level annotations provided by expert radiologists, following the Radiology Society of North America (RSNA) standard for reporting

CT imaging findings potentially attributable to COVID-19 (SIMPSON et al., 2020). As previously explained (Chapter 2), the RSNA standard classifies images in one of four classes: Typical appearance, Indeterminate appearance, Atypical appearance, and Negative for pneumonia. The first category corresponds to exams showing CT features frequently and seen explicitly in patients with COVID-19 pneumonia (*e.g.,* bilateral, peripheral, and multifocal ground-glass opacities), thus representing our class of interest (*i.e.,* positive). The distribution of CT scans among classes is shown in Table 4.3.

Table 4.3: Private Dataset, number of CT images obtained from each source, and distribution by class.

Class	**HMV (289)**			**HCPA (105)**		
	Female	Male	Total	Female	Male	Total
Typical	34	58	92	13	17	30
Negative	54	38	92	11	4	15
Indeterminate	32	35	67	16	14	30
Atypical	16	17	33	13	17	30

4.2.1 Cross-Dataset Validation

As the CT images from our private dataset came from different sources and distinct types of equipment, we performed some experiments as a proof of concept to prove we could use the data as one unique dataset. We train a model with five folds cross-validation and use our pipeline (discussed in Chapter 5) for each case:

1. A binary model (with the same architecture described in Chapter 5) with the dataset from HMV and HCPA together using the origin hospital for each exam as labels for classification.
2. A model following the architecture described in Chapter 5 for binary classification (distinguishing typical COVID-19 from other classifications) trained and validated with the HMV dataset and tested with the HCPA dataset.
3. A model following the architecture described in Chapter 5 for binary classification (distinguishing typical COVID-19 from other classifications) trained and validated with the dataset from HCPA and tested with HMV exams.

The trained model for the Case 1 achieved an average accuracy of 93.9%. Therefore, we concluded that there are visible differences in data from different sources, as the

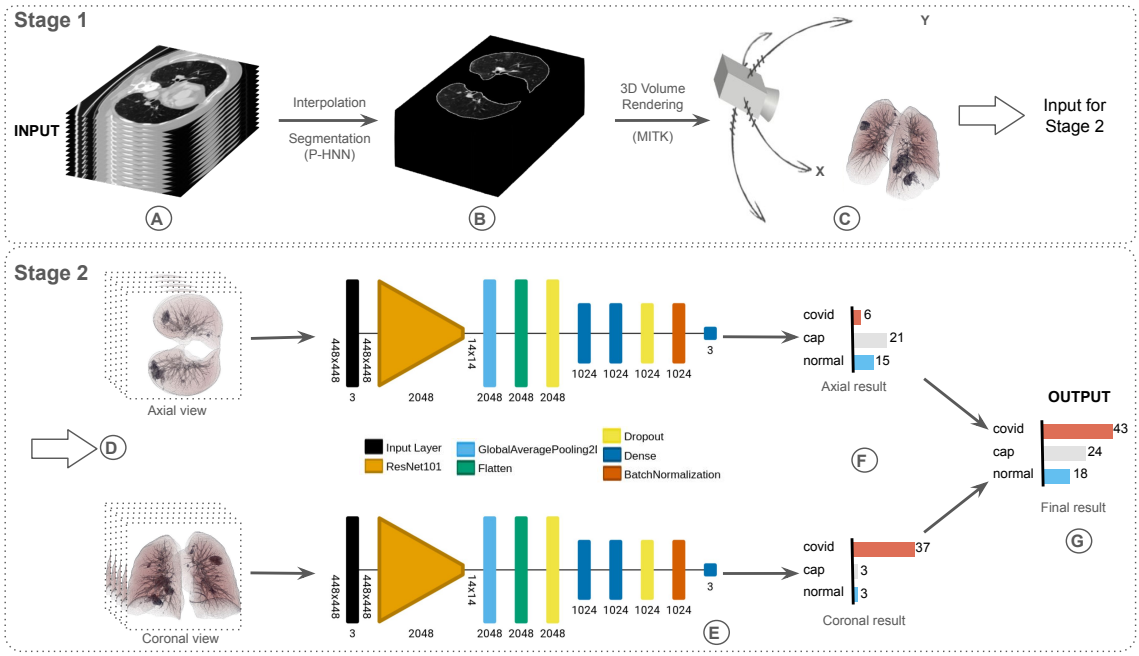
model can successfully discriminate the source of the exams.

Despite that, the trained models for Cases 2 and 3 obtained an average precision score of 85.7% and 81.8%, which suggests that those spottable differences in dataset sources do not interfere with the ability of our binary models to learn and generalize. Hence, we decided to treat these sets as a unique dataset.

5 THE CLASSIFICATION ARCHITECTURE

This section describes the proposed approach, named **COVID-VR**. Our pipeline involves two stages, as depicted in Figure 5.1, providing an end-to-end pipeline for analysis of CT images, from lung segmentation to patient-level classification regarding the presence of COVID-19-related infections based on volumetric information. The following sections will explain our methodology, covering initial data pre-processing and input preparation (Stage 1) to model development (Stage 2).

Figure 5.1: The pipeline of our proposed COVID-VR approach is divided into two stages. Stage 1 carries out data preparation to obtain the input images for model development, which is conducted in Stage 2. Data preparation involves the steps of image resizing by interpolation, lung segmentation, 3D volume rendering using pre-defined transfer functions, and snapshots generation for axial and coronal planes. In Stage 2, DL models are trained for each plane to distinguish among the classes of interest, and their outputs are combined to obtain a final patient-level classification.



5.1 Preprocessing Steps

5.1.1 Conversion to NIfTI

Initially in DICOM format, the received CT scans (Figure 5.1A) are first converted to the NIfTI format. We use dicom2nifti python package and additional configuration to make the conversion to NIfTI format and cropping the values of HU to [-1500, 2000]; *i.e.*,

all values lower than -1500 are set as -1500, all values above 2000 are set as 2000. This script is available in our repository [1].

5.2 Lung Segmentation

After converting the images to NIfTI format, we use the iftNormalize script (SOUSA et al., 2019) for interpolating and resizing the images to get isometric volumes with a slice spacing of 1mm in all dimensions, thus obtaining a more homogeneous representation across patients. Lung segmentation is carried out with the Progressive holistically nested networks (P-HNN) model (HARRISON et al., 2017). We use the pre-trained model of P-HNN to get the segmented lungs as shown in Figure 5.1B, choosing a threshold of 75% or above from the probability mask to consider a voxel as part of the lung. This output is saved in NIfTI format.

We note that other available methods for lung segmentation were considered, such as UBC (LENSINK et al., 2020) and Lungmask (HOFMANNINGER et al., 2020a), that perform additional tasks than lung segmentation as a whole. For example, Lungmask presents segmentation at the level of lobes and lungs, generating masks that manage to differentiate the right lung from the left and similarly achieve to differentiate the lobes. On the other hand, UBC accomplishes to identify COVID-19 lesions. While P-HNN only performs the segmentation of the lungs without separating the lungs, lobes, or lesions.

Nonetheless, P-HNN was selected after recommendation by expert radiologists against UBC, who compared the results of these methods and observed that P-HNN loses fewer lung portions; and doing some experiments (full train and validation of our pipeline) for P-HNN against Lungmask, P-HNN arranges better results than Lungmask, thus being more appropriate to our approach.

5.3 Volume Rendering from CT scans

With the lung segmented in NIfTI format, we perform a 3D volumetric rendering with the Medical Imaging Interaction Toolkit (MITK) framework (WOLF et al., 2004) by applying a customized transfer function (TF), as shown in Figure 5.1C. In-house scripts implemented in C++ based on MITK Framework were used to render the 3D volume

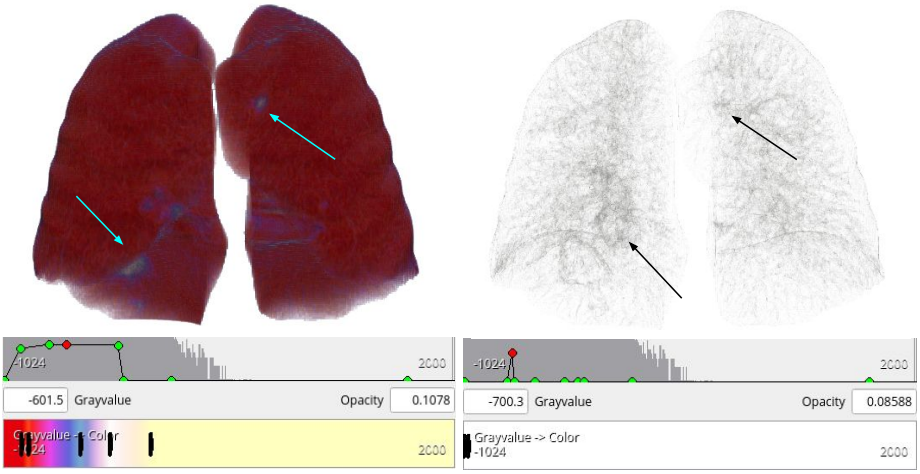
[1] <https://github.com/covid-vr/p-hnn-lung-segmentation>

of the segmented lungs, trying multiple TFs to map each voxel to color, transparency, and opacity. Our motivation for exploring different TFs is that each one of them can potentially highlight other regions of interest or reveal distinct relevant features present in CT exams. Thus, we define a set of TFs interactively, intuitively assigning color and opacity to original voxel intensities and observing the TF capacity to express relevant characteristics to the intended utility (*i.e.,* lung changes observed in COVID-19 patients).

The HU is a dimensionless scale used to express CT numbers in a standardized form. It is obtained from a linear transformation of the measured absorption/attenuation of the X-ray beam. As the physical density of each tissue is proportional to its coefficient, the HU unit has many well-known uses. Our first task was identifying the best range to map our TFs in the images' HU scale values to restrict our options. Related works (TANG et al., 2020b; LU et al., 2021) show that the normal HU scale for lung is [-1000, -700] and for ground-glass opacity is [-700, -300]. Therefore, we chose similar intervals to test while defining our TFs.

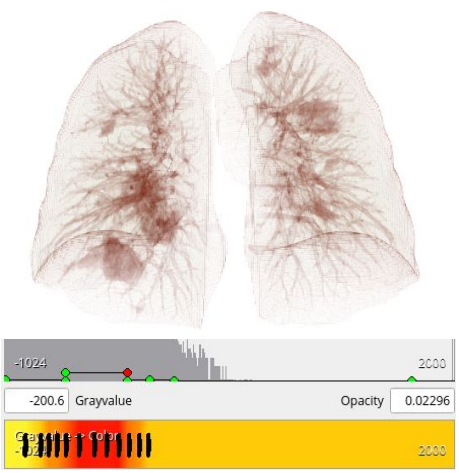
Figure 5.2 shows the six TFs (TF1 to TF6) that were defined after preliminary empirical analysis. Although they were all applied to the same CT exam, the obtained rendered volume differs among distinct TFs. TF1 (Figure 5.2a) highlights the outer layer of the lungs and provides information about the surface texture of the thoracic cavity, but a significant disadvantage is that it can miss the lesions that occur within the inner layer, which covers the lungs, blood vessels, nerves, and bronchi. On the other hand, transfer functions 2 to 6 (TF2 to TF6, in Figure 5.2) allow better visualization of internal lesions, with varying capabilities in terms of wall delineation and highlighting imaging features that were typically described for COVID-19 pneumonia. TF2 (Figure 5.2b) is defined within the lower range of values in the limit of lung and ground-glass opacity (*i.e.,* [-700, -300]), mapping only HU values in this interval. This function conserves a certain spatial distribution of the ground-glass opacity present in the lung but misses details on the lungs' texture and regions without any lesions. TF3 (Figure 5.2c) aims to replicate the behavior of 2D CTs that vary in a single color scale (usually gray-scale), mapping original values in the range of [-750, -200] with variations applied to the brightness. TF3 manages to preserve the lung as a three-dimensional image without losing critical information on the internal lesions caused by COVID-19, such as ground-glass opacity. TF4 to TF6 are obtained by including more colors to help differentiate lungs' textures and highlight different imaging features that may be present in the distinct classes comprised by our datasets (see Section 4). With TF4 (Figure 5.2d), we observe a thin delineation

Figure 5.2: Volume rendering of the lung using different transfer functions for COVID-19 diagnosis. The defined transfer functions explore a variety of mapping combinations for color and opacity, thus highlighting different regions or features of interest when applied to CT scans.

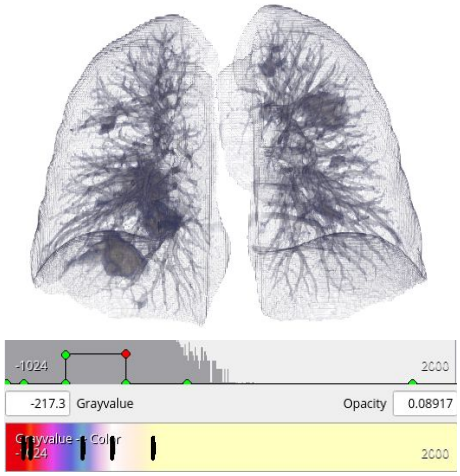


(a) Transfer Function 1 - TF1

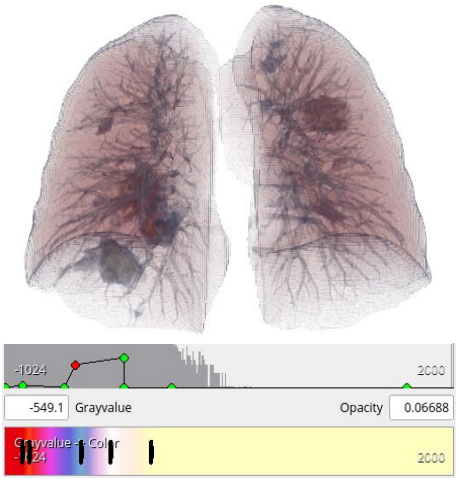
(b) Transfer Function 2 - TF2



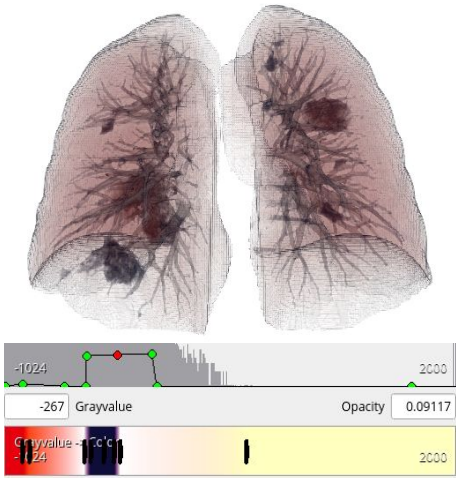
(c) Transfer Function 3 - TF3



(d) Transfer Function 4 - TF4



(e) Transfer Function 5 - TF5



(f) Transfer Function 6 - TF6

of the lungs' inner layer, in addition to a prominent highlighting of the bronchi, which in the presence of COVID-19 may show changes such as bronchial wall thickening (YE et al., 2020). With TF5 and TF6 (Figure 5.2e and 5.2f, respectively), we observe a clearer differentiation of the ground-glass opacity from other characteristics, which may be interesting for the classification model as this feature is the most common imaging finding in COVID-19 patients (YE et al., 2020). Thus, by exploring a wide range of mapping combinations for color and opacity, each TF can highlight different regions or features of interest when applied to CT scans, such as external surfaces of the lungs, as shown in Figure 5.2a, arteries as shown in Figure 5.2c, and ground-glass opacity as in Figures 5.2e and 5.2f.

5.4 Extraction of Coronal and Axial Views for Model Input

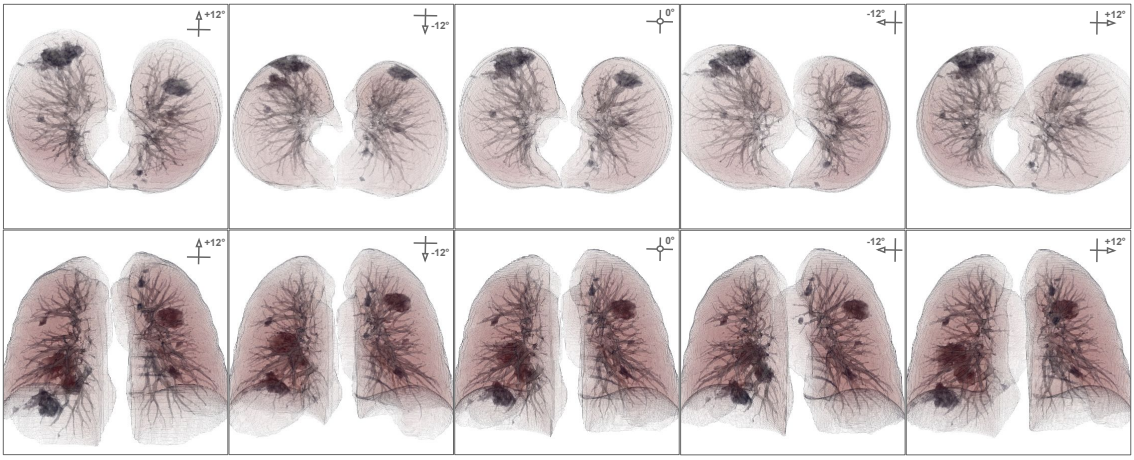
Our classification model receives as input images derived from CT scans, with a dimension of 448×448px (Figure 5.1D). To generate these images from the reconstructed 3D volumes, we obtain a series of snapshots for coronal and axial planes. We first place the virtual camera in the axial plane and displace it for $\pm 12.0°$ degrees in the Y and X-axis, taking pictures of the lungs every 1.2° degrees. We repeat the same process after positioning the camera in the coronal plane. With this procedure, we extract 42 different images per view (*i.e.,* coronal or axial), thus generating a total of 84 images per CT exam. Examples of the resulting images are shown in Figure 5.3.

5.5 Classification with a DNN Model

5.5.1 Model Architecture and Training

Our models are developed using deep neural networks (DNN) and the Tensorflow framework. The network architecture is summarized in Figure 5.1E. To improve the learning process for COVID-19 classification, we explore the concept of transfer learning, reusing a pre-trained model as the backbone for our architecture. We try multiple Convolution Neural Network (CNN) families such as VGG (SIMONYAN; ZISSERMAN, 2015), ResNet (HE et al., 2016), DenseNet (HUANG et al., 2017), and EfficientNet (TAN; LE, 2020). Figure 5.1E shows the Resnet101 as the backbone network, represented in or-

Figure 5.3: Some view images obtained from the reconstructed 3D volumes are used as inputs for our DL-based classification model. The first row represents images extracted for the axial plane. The third image in this row shows our initial position (0,0), to which the camera points for the axial view. From this position, we capture snapshots every 1.2° towards the horizontal and vertical axis directions, going from -12° to +12° on each axis. The same process is repeated for the coronal plane, as seen in the second row.



ange. To this backbone, we add a sequence of modern deep learning modules commonly explored in the literature to help model training. Specifically, following the backbone, we append a Global Average Pooling layer, a 20% Dropout layer to avoid overfitting, two fully-connected Dense layers, a new 20% Dropout layer, and a Batch Normalization layer. The final module is a Dense layer with a Sigmoid activation function for the binary classification task and a Softmax activation function for the ternary classification task. The definition of network architecture was performed empirically through several preliminary experiments. For model compilation, we use the Adam optimizer with a learning rate of 2×10^{-5}, adopting the Binary Cross-entropy loss function for binary models and Categorical Cross-entropy loss function for ternary models.

Model training and validation are carried out using stratified 5-fold cross-validation (CV) for the private datasets. For the public dataset, we follow the competition organizers' orientation and use a random 70%/30% split from the COVID-CT-MD dataset (AFSHAR et al., 2021) to generate the train and validation sets, also respecting original class distributions among splits. The model developed is further evaluated with three independent test sets provided by the SPGC-ICASSP competition (HEIDARIAN et al., 2021), as previously described in Section 4.1. Additionally, to increase the amount and variety of the dataset at the training step and reduce the chances of overfitting, we use data augmentation methods in the training partition, like rotation up to 15°, zoom ($\pm$5%), rescaling of 1/255 (RGB pixel value to float number between [0-1]), and shift for width and height image up to 10%.

We note that our approach trains an individual DNN for each defined plane, either axial or coronal (we did not use sagittal because it has overlapped images which do not represent better performance for our proposed approach as mentioned in Section 6.2), based on the set of image views extracted from it. Moreover, despite the use of patient-level annotations for the supervised learning task, the output of our DNNs is a classification per image view. Thus, the batch of image views obtained for a given CT scan (*e.g.,* 42 axial views or 42 coronal views) is classified by the corresponding network, specialized in either the axial or coronal view, generating a class label for each image analyzed. From this output, we obtain a distribution of class votes (Figure 5.1F) per image batch, which is further passed through a consensus-extraction module by majority voting to generate a patient-level prediction.

5.5.2 Patient-level Classification

As shown in Figure 5.1F, our classification approach is defined based on two submodels, each of which predicts the COVID-19 diagnosis using images obtained from a specific view of the reconstructed 3D volume, axial or coronal. We generate a patient-level consensus distribution based on the distribution of class votes received from each submodel by summing up the number of votes per class. For example, in the ternary classification task shown in Figure 5.1F, the 42 images generated from the axial view are classified by the corresponding submodel with the following distribution: 6 votes for class COVID-19, 21 votes for class CAP, and 15 votes for class Normal. On the other hand, the submodel trained for coronal view images assigns 37 votes for class COVID-19, 3 for class CAP, and 3 for class Normal. The ensemble-based solution generated from the sum of submodels' votes results in 43 images classified as COVID-19, 24 images classified as CAP, and 18 images classified as Normal. The final result, shown in Figure 5.1G, is used to predict a single label for the input CT scan, which is the class with the maximum number of votes according to both views. In this example, the input CT scan would be classified as COVID-19 by our approach.

5.5.3 Model Evaluation

Based on CT scans, we adopt traditional evaluation metrics to assess our model's performance for COVID-19 diagnosis. Using diverse metrics to reflect different types of correct and incorrect predictions helps investigate the model's clinical utility and eventual limitations. Thus, performance assessment was based on the accuracy (Acc), sensitivity (Sens, also called recall), specificity (Spec), precision (Prec, *i.e.,* positive predictive value), F1-score (*i.e.,* the harmonic mean between recall and precision), and the area under the Receiver Operating Characteristic (ROC) Curve (AUC score). For the ternary model, given its multiclass nature, we adopted the micro-, and macro-average (OPITZ; BURST, 2019) for all metrics except accuracy.

Additionally, to calculate the true positive rate (TPR) and the false positive rate (FPR) of positive classes for each configuration (we consider COVID-19 as a positive class), we need the probabilities of each class. Since our approach is not directly based on the probabilities per class at the image level (as seen in Figure 5.1E the network returns three probabilities for each of the input 42 images. We only consider the final labels from these values to proceed with the operations detailed in chapter 5.5.2). Taking the same example of the Figure 5.1 if we have 43 images classified as COVID-19, 24 as CAP, and 18 as Normal for a given patient, it will be defined that patient has a probability of 0.5243 of being COVID-19, 0.2927 of being CAP, and 0.2195 of being classified as Normal and we use these probabilities as values to calculate ROC curve in the results section (Chapter 7). These results per patient are also available in [2].

[2] <https://covid-vr.github.io/>

6 DESIGN CHOICES

This chapter aims to describe some of the choices we made during this study's development, presenting the respective justifications.

6.1 Environment for experiments

For the development of this work, we have three primary experimentation environments detailed below.

1. **Local Environment**: this consists of a computer with Intel(R) Core(TM) i7-4790 as CPU and GeForce GTX Titan graphics card with 6GB of RAM; additional information is reported in Table 6.1. We managed to start our first scripts for both steps, volume rendering, and first network models with this environment. Subsequently, we focused mainly on performing only the part of the volumetric rendering of the lungs that our *Server Environment* could not manage to run due to administrator permissions. Finally, using this environment, we build and execute the COVID-VR end-to-end deployment, deploying a Docker Image built with all the necessary packages for our presented approach detailed in Appendix B.
2. **Server Environment**: consists of a computer that has an Intel(R) Xeon(R) Silver 4208 CPU and GeForce RTX 2080 Ti graphics card with 22GB of RAM (separated into 2 GPUs with 11 GB of RAM each). Since this environment was available months after starting this work, the first steps were carried out in the *Local Environment*. Afterward, we pass to this instance the process of segmentation, training, validation, testing of the models, the data storage, and the implementation of the models to be compared (DeCoVNet, COVNet, TheSaviours). Some other important details of this environment are detailed in Table 6.1.
3. **COVNet Environment**: consists of a computer that has an Intel(R) Core(TM) i5-9400F CPU and GeForce RTX 3090 graphics card with 24GB of RAM. During the development process of our comparison experiments with other models, we faced a problem with the approach presented in COVNet due to memory errors during GPU allocation. Despite our GPU in the Server Environment having 22GB of memory, we only managed to use one of the GPU, which limited us to use 11 GB of RAM. We tried to go after the error. However, we did not get any workable solution.

Finally, we got access to this COVNet environment where we performed only the experiments in COVNet approach in this environment (COVID-VR, DeCoVNet, and TheSavioursexperiments were performed in the *Server Environment*). Some important details of this instance are detailed in Table 6.1.

Table 6.1: Environments of experiments are presented in this work.

	Local enviroment	**Server enviroment**	**COVNet enviroment**
CPU	Intel(R) Core(TM) i7-4790	Intel(R) Xeon(R) Silver 4208	Intel(R) Core(TM) i5-9400F
N° cores	8	32	6
RAM	24 GB	314 GB	16 GB
Operating System	Ubuntu 19.04	Debian 4.19	Ubuntu 20.04
Graphic card	GeForce GTX Titan	GeForce RTX 2080 Ti	GeForce RTX 3090
GPU RAM	6 GB	22 GB (2 GPU of 11 GB each one)	24 GB
Cuda version	11.0	11.2	11.1

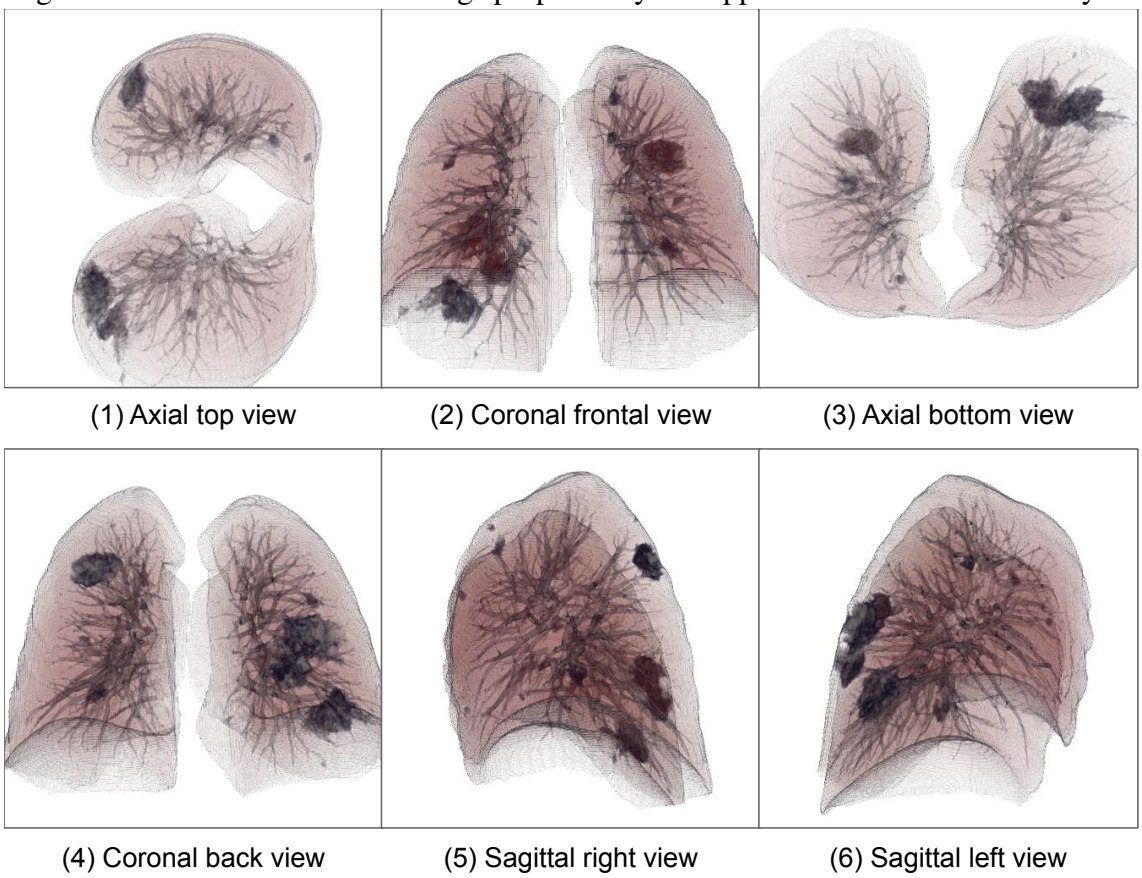
6.2 Volume Rendering cases

For the development of our work, we used the volumetric visualization part only as a visual aid. We noticed that according to the established transfer functions, we were able to highlight various parts of the desired lung and regions of interest that were increasingly interesting based on the change in transfer function (TF). Using the scripts developed and based on the MITK project, we can place a camera anywhere in the space surrounding the lung. Hence, we can take screenshots and videos from any point of view of the lung.

For simplicity, we decided to focus only on the six views that we found to be representative and similar to the views currently used in medicine, the Axial, Coronal, and Sagittal views. This corresponds taking camera shots from the top for the Axial view (Figure 6.1.1), the bottom (Figure 6.1.3), from the Coronal view front and back shots (Figure 6.1.2 and Figure 6.1.4, respectively), and the Sagittal view taking pictures from the right and left views of the lung (Figure 6.1.5 and 6.1.6, respectively).

Since our network works by training one model per view and we had six views available, making a majority vote on the results of all the networks, training six networks for each test was much more time-consuming. Additionally, the results obtained by the Sagittal views had worse performance than the others. One hypothesis could be due to

Figure 6.1: All six views of the lungs proposed by our approach COVID-VR to analyze.

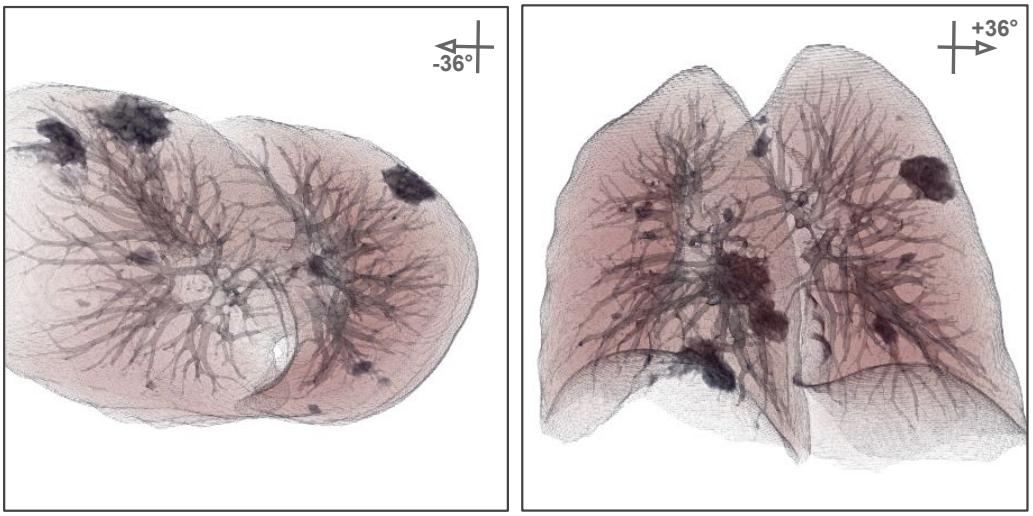


the overlapped lungs as seen in Figures 6.1.5 and 6.1.6. Subsequently, we continued the experiments; however, the training of four networks was still very time-consuming. From the observed accuracy metrics, using all four views had almost the same performance as using only two (an axial and a coronal view). We raised the idea of using three views, two axial and one coronal and vice versa, but the results were not better since one view ended up biasing the final result. Thus, we finally consider using only the Top Axial and Front Coronal views to continue the work.

The next point to discuss was that since we could generate numerous images (camera shots) as long as we managed to take captures from any point of view towards the lung. The challenge was how many picture shots were ideal. Without a metric of how many images are recommended to take, we used the idea presented in Section 5.4 where we consider taking the following amounts of pictures: 162, 82, 42, and 22 sweeping angles of a maximum 15° and a minimum 9° (since greater angles than these causes the image of the lungs begins to deform and lose the idea of still identifying the Axial or Coronal view like can be seen in Figure 6.2). Finally, we took the decision for image quantity based on the performance of the networks trained for each of the models, obtaining the best results

with 82 and 42 images per view. To reduce the training time amount, we decided to use 42 images per view for each network.

Figure 6.2: Pictures of lungs were taken from angles which causes overlapped images.



(a) Picture was taken from -36° (horizontal axis) in axial top view.

(b) Picture was taken from +36° (horizontal axis) in coronal front view.

In addition to the transfer functions detailed in Figure 5.2, we tested some others (we did not train a model for everyone) as shown in Figure 6.3. Another decision to make was the background of the images. We tested only two background colors, white and black. As reported in Section 5.3, when we change the background by applying the same transfer function we manage to highlight different lung regions. From the experiments that we carried out when we changed the background color from black to white with the selected transfer function TF6, we obtained an improvement of 3% of accuracy concerning the model used by training the images with a black background color in the ternary classification task using the COVID-CT-MD dataset.

Finally, similar to the step-by-step reported by TheSaviorsto use only the center 80 slices for classification, we perform the analog for our approach. Hence, obtaining an image without the top and bottom of the lungs. This changes the axial and top views as shown in Figure 6.5. With this attempt, we want to check if removing these edges from the bottom and top could give additional information. One example is removing these borders in the coronal model so that the network does not pay attention to these features. Another is removing in the axial view the top of the lungs. This case shows how the image achieves a difference because we remove part of the outer region to concentrate on the inner region of the lungs. This additional step did not improve the results obtained by the trained networks, which is why it is disregarded from the final model.

Figure 6.3: Some of the other transfer functions studied for COVID-VR

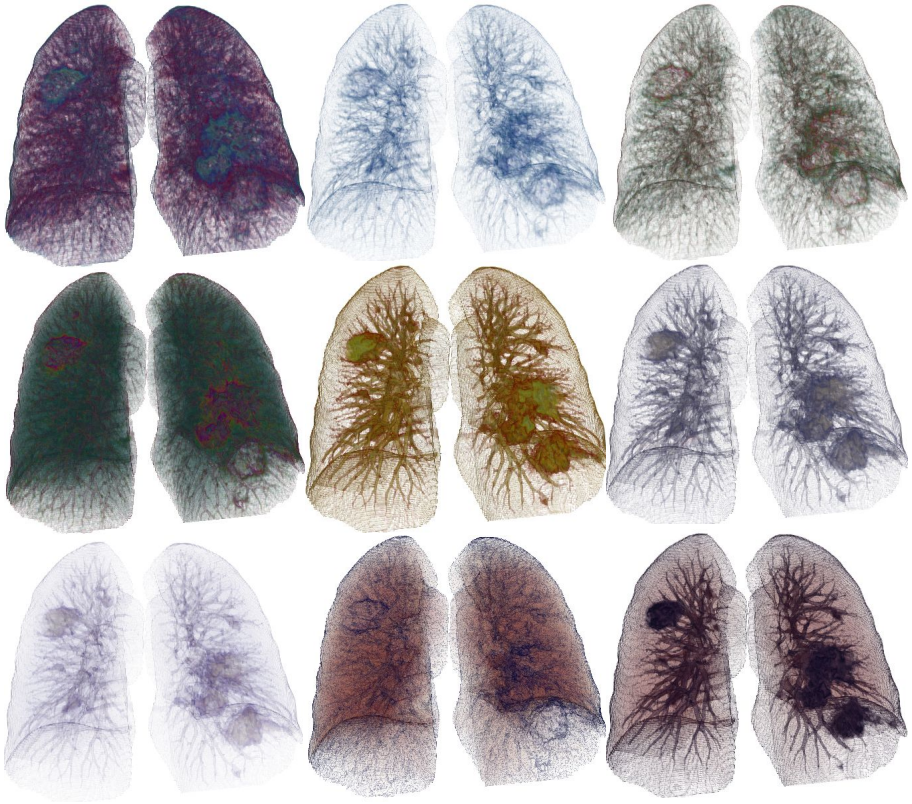


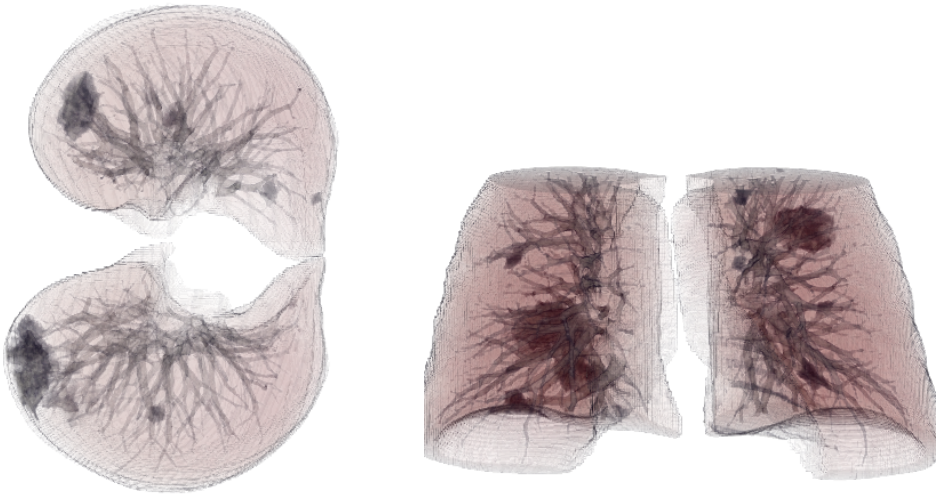
Figure 6.4: Same transfer function with different background colors. It should be noted how the lesions in the white background are easily seen in contrast with a black background.

(a) (b)

6.3 Network architectures

At this stage, we will provide details of some of the networks trained to carry out this work, likewise the change in the inputs of our neural networks.

Figure 6.5: Volume rendering from 80 center slices for COVID-CT-MD.



(a) Axial view, the image is a bit clear in contrast to Figure 6.11 because of the absence of the top lung region

(b) Coronal view

Among the first networks we tested are VGG16 and VGG19 using an architecture similar to the one presented in figure 5.1, where VGG was used as a backbone, continuing the following layers. From the additional layers of the backbone, we added a dropout of 20% (we also experimented with other dropout values of 10%, 15%, and 25%; however, we obtained the best results with 20%). Likewise, we discussed whether or not Batch Normalization should be used together with Dropout. We performed the experiments removing the Batch Normalization layer, concluding that the metrics did not reach the baseline, including the Dropout and BatchNormalization layers. For the Dense layers found after the first Dropout of our network, we also tried 2048×2048 and 2048×1024, but the results did not improve significantly. So, we kept the 1024×1024 dense layers. Finally, we used Adam as the Optimizer with a Learning Rate in the set $\{1 \times 10^{-3}$, 1×10^{-4}, 1×10^{-5}, 2×10^{-5}, 5×10^{-4}, $5 \times 10^{-5}\}$, obtaining better results with the value of 2×10^{-5} and using it for our experiments.

Some other architectures experimented with had a network with multiple inputs (one input per chosen view), and the output was unique for all inputs. That is, the network receives as input three images of each chosen view, and each of these images passed through a subnetwork similar (for each view) to the one shown in Figure 5.1. Still, the output passes through an assembly of a smaller dense network. We performed this experiment with the VGG16 network as the backbone; however, the returned results did not improve significantly and as the model was bigger, the training time did not validate the use of this architecture.

Another point to be discussed was the size of the inputs of our proposed network. To do this, we tried different image sizes in the following set $\{512 \times 512$px, 448×448px, 384×384px, 320×320px, 256×256px, and 224×224px$\}$, finding best between the first two values of 512px and 448px, using the latter since it manages to make the network smaller. Therefore the training time and the results obtained are highly similar to using a 512px input.

6.4 Views Aggregation

The justification for using two views and not focusing on just one is that in almost all the experiments carried out, we observed that the accumulated view obtains better results than the application independently. This can be seen in the confusion matrices in Figure 6.6 comparing them against their cumulative shown in Figure 7.1b.

The Table 6.2 also shows that almost all the metrics perform better in the cumulative model (except for specificity and precision in the Normal class for the axial view model). From the weighted metrics, we appreciate that the accumulated method has an improvement of 4.1% compared to the model of the axial view and 2% of the coronal view.

Figure 6.6: Confusion matrices for ternary classification using the public dataset COVID-CT-MD. (a) shows the confusion matrix for AXIAL view model; and (b) shows the confusion matrix for CORONAL view model, both using TF6 and Resnet101 as the backbone.

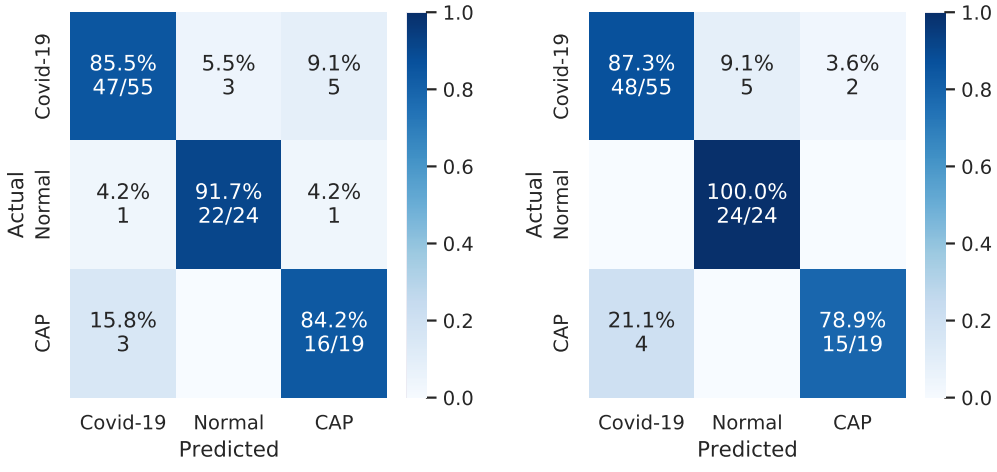


Table 6.2: Comparison among COVID-VR individual views against accumulated views (axial + coronal) using the public dataset provided by the SPGC 2021 competition.

	Metrics	COVID-VR (Axial)	COVID-VR (Coronal)	COVID-VR
Overall	Accuracy	86.7%	88.8%	**90.8%**
	F1-Score	86.9%	88.7%	**90.8%**
	Kappa	78.0%	81.2%	**84.6%**
	AUC score	94.3	94.0	**95.4**
COVID-19	Sensitivity	85.5%	87.3%	**89.1%**
	Specificity	90.7%	90.7%	**93.0%**
	Precision	92.2%	92.3%	**94.2%**
	F1-Score	88.7%	89.7%	**91.6%**
Normal	Sensitivity	91.7%	**100.0%**	**100.0%**
	Specificity	**95.9%**	93.2%	94.6%
	Precision	**88.0%**	82.8%	85.7%
	F1-Score	89.8%	90.6%	**92.3%**
CAP	Sensitivity	**84.2%**	78.9%	**84.2%**
	Specificity	92.4%	**97.5%**	**97.5%**
	Precision	72.7%	88.2%	**88.9%**
	F1-Score	78.0%	83.3%	**86.5%**

6.5 Backbone Network Choice

To perform CT scan classification with the model described in Section 5.5 and summarized in Figure 5.1, several popular CNN architectures were tested as the backbone network, namely ResNet (HE et al., 2016), DenseNet (HUANG et al., 2017), VGG (SIMONYAN; ZISSERMAN, 2015), and EfficientNet (TAN; LE, 2020) families. For each of these, we tried different depths (*i.e.,* ResNet50, ResNet101, DenseNet121, DenseNet201, EfficientNet-B0, EfficientNet-B1, EfficientNet-B6, VGG16, and VGG19) with fixed training and validation sets contained in the COVID-CT-MD dataset and compared their classification performance. Table 6.3 presents the most promising models obtained for each network family, using the TF6 (Figure 5.2f) as a fixed transfer function for 3D volume rendering. These results are obtained for the ternary classification models using the train and validation sets from the SPGC-ICASSP competition (*i.e.,* COVID-CT-MD dataset). We provide the performance for the COVID-19 class and overall performance for the three classes. Overall F1-score and AUC score are obtained with the micro-average; the macro-averages and the complete performance analysis per class can be found in the Appendix A.

Our model achieved the best results using the ResNet101 network as its backbone. The validation accuracy and F1-score were the highest among the best models obtained for each network family. Both accuracy and F1-score for the overall classifi-

cation in a ternary approach were 90.8%, approximately two percentage points above VGG16, which presented the second-best performance (Table 6.3). In terms of AUC score, DenseNet121 achieved the highest mark of 96.5 compared to 95.4 for ResNet101. Nonetheless, observing the classification results for the COVID-19 class (Table 6.3), we note that ResNet101 showed the best performance for all metrics. Moreover, it achieved outstanding performance for CAP cases and competitive performance for Normal cases (Appendix A). Thus, we chose ResNet101 as the backbone network for the proposed COVID-VR approach, using it for all the experiments further reported in our work.

Table 6.3: Comparison of distinct backbone network architectures in our COVID-VR approach. Model training and validation were carried out with the train and validation sets from the COVID-CT-MD public dataset for the ternary classification task (COVID-19 vs. CAP vs. Normal). F1-score and AUC scores are based on the micro-average.

	Metrics	VGG16	DenseNet121	EfficientNet-B2	ResNet101
Overall	Acc	88.8%	87.8%	86.7%	**90.8%**
	F1	88.6%	87.7%	86.7%	**90.8%**
	AUC	95.6	**96.5**	95.1	95.4
COVID-19	Sens	**89.1%**	85.5%	83.6%	**89.1%**
	Spec	88.4%	93.0%	90.7%	**93.0%**
	Prec	90.7%	94.0%	92.0%	**94.2%**
	F1	89.9%	89.5%	87.6%	**91.6%**

6.6 Transfer Function Choice

As explained in Chapter 5.3, different transfer functions were tested in the pre-processing step to generate the 3D rendered images. Results for TF4 and TF5 were omitted due to similar performance obtained with TF3 and TF2, respectively. We observed that TF6 presented the best accuracy and overall F1-score performance. Although it did not show the highest sensitivity among the TFs compared, it had a balanced performance in detecting COVID-19 cases. In addition, TF6 achieved the best sensitivity for CAP and normal cases (Appendix A).

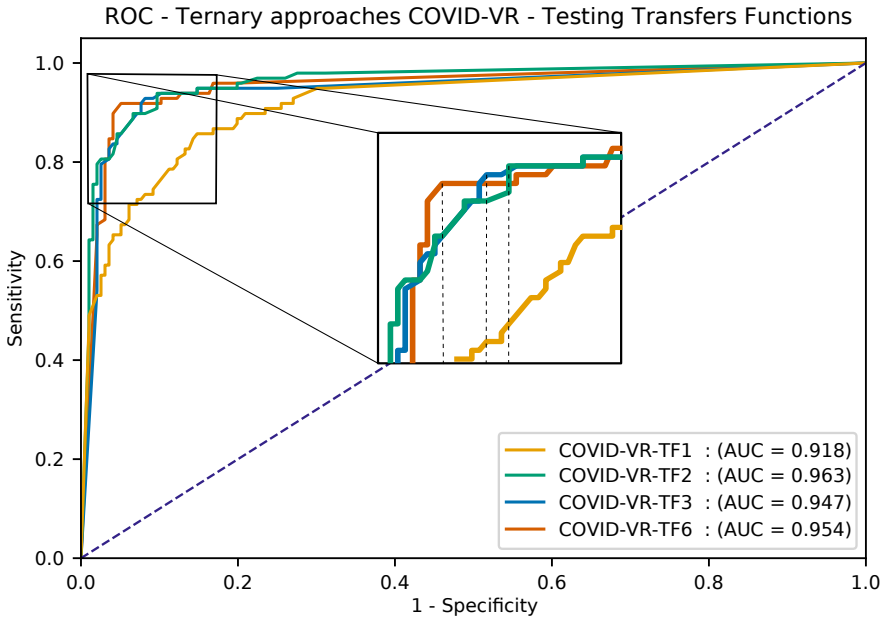
Figure 6.7 compares the ROC curves obtained for distinct TFs. Although a little difference is observed between the AUC score for TF6 and TF2, we note that the ROC curve for TF6 achieves the highest sensitivity for a 10% false positive rate. Therefore, TF6 was chosen as the standard transfer function for our approach. Besides the visual detection of different features highlighted by each TF, this test proved relevant, as switching the TF while keeping the same model architecture results in considerable accuracy variation, as

shown in Table 6.4.

Table 6.4: Comparison among Transfers Functions in the COVID-VR approach using train and validation sets from the COVID-CT-MD public dataset for the ternary classification task (COVID-19 vs. CAP vs. Normal). F1-score and AUC scores are based on the micro-average.

	Metrics	TF1	TF2	TF3	TF6
Overall	Acc	79.6%	85.7%	87.8%	**90.8%**
	F1	78.7%	84.5%	87.6%	**90.8%**
	AUC	91.8	**96.3**	94.7	95.4
COVID-19	Sens	89.1%	**96.4%**	89.1%	89.1%
	Spec	72.1%	72.1%	88.4%	**93.0%**
	Prec	80.3%	81.5%	90.7%	**94.2%**
	F1	84.5%	88.3%	89.9%	**91.6%**

Figure 6.7: Micro-average ROC curves for four distinct Transfer Functions (*i.e.,* TF1, TF2, TF3, and TF6) in our COVID-VR approach.



In the results discussed in the next chapter, we use ResNet101 as the backbone and TF6 as the transfer function.

7 RESULTS

This section reports the results obtained with the proposed COVID-VR approach.

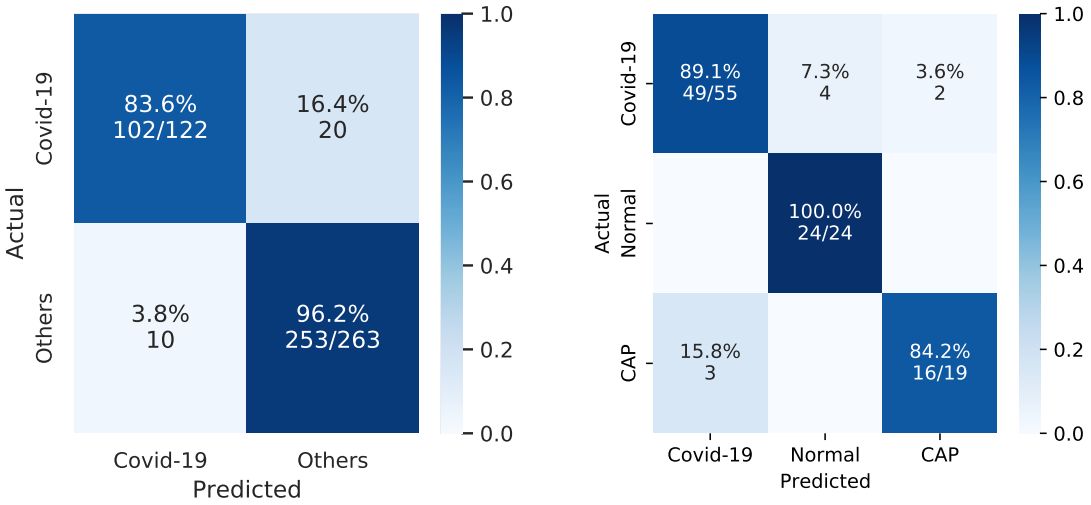
The first step was to perform binary classification with our private dataset from HCPA and HMV, distinguishing patients classified as having typical COVID-19 lung infection by a radiologist from one of the two hospitals. This binary classification was conceived by combining three radiological RSNA diagnoses: negative, atypical, and indeterminate. Tests with quaternary classification, preserving all four private dataset's original radiological classifications, were realized, but none of our designed models or related works tested achieved satisfactory results. Possible reasons include class imbalance and an insufficient number of instances in the dataset. While working with the private dataset, we performed a simple 5-fold Cross-Validation technique.

The second line of experiments consisted in using the public dataset from SPGC to perform ternary classification using the dataset original classes (COVID-19, Normal, and CAP), which provided a more direct comparison with other published works that reported their results training and validating with data from the competition. While working with the SPGC dataset, we defined training and validation sets according to the data split published by the SPGC 2021 competition and the dataset itself, in a nearly 0.68/0.32 (207/98 patients) proportion.

Our best results in each line of the experiment are displayed in Figure 7.1. The configuration utilized, the model backbone architecture, and chosen transfer function, will be detailed further. The following subsections contain the tests that justify our choices regarding backbone architecture and transfer function. Later, we present a comparison between our model using the chosen configuration and related works to each of the two lines of experiments described.

Later, we present the results for comparing different backbone networks and transfer functions using the public dataset, COVID-CT-MD. We extract the insights supporting model architecture definitions from these findings applied to the other experiments. Then, we provide the classification performance of our model in comparison with state-of-the-art approaches for detecting COVID-19-related pneumonia in the COVID-CT-MD dataset. Finally, we discuss the performance of the models developed from the private datasets. We note that for some experiments, we compare both ternary and binary classification models. Moreover, given the class imbalance in our datasets, we focus our discussion on the metrics' micro-average for the ternary models. Full details about the

Figure 7.1: Confusion matrices of our main results: (a) binary classification using hospitals (private) dataset and (b) ternary classification using SPGC (public) dataset.



(a) (b)

results obtained with our approach are provided in Appendix A.

7.1 Classification Performance for the Public Dataset

We conducted experiments to analyze COVID-VR's performance for a ternary classification model (COVID-19 vs. CAP vs. Normal) in the validation set and the three test sets (here unified in a single test set) released by the SPGC 2021 competition. All models presented in this section use the ResNet101 backbone network and TF6 for volume rendering to achieve better performance in previous results. We compared the proposed approach against the winner of the competition (TheSaviours) (CHAUDHARY et al., 2021), and two other state-of-the-art methods, DeCoVNet (WANG et al., 2020) and COVNet (LI et al., 2020b). We emphasize that the model for the TheSaviours approach in the validation test (Subsection 7.1.2) results was trained for us following the instructions in the provided repository[1], while in the model for test set we use the pre-trained weights also available in the public repository, the reason for these choices are described in Chapter 6.

The results for the ternary models are presented in Table 7.1 for validation set and Table 7.2 for test set.

[1]TheSaviours repository: <https://github.com/shubhamchaudhary2015/ct_covid19_cap_cnn/>

7.1.1 Results in Validation Set of Public Dataset

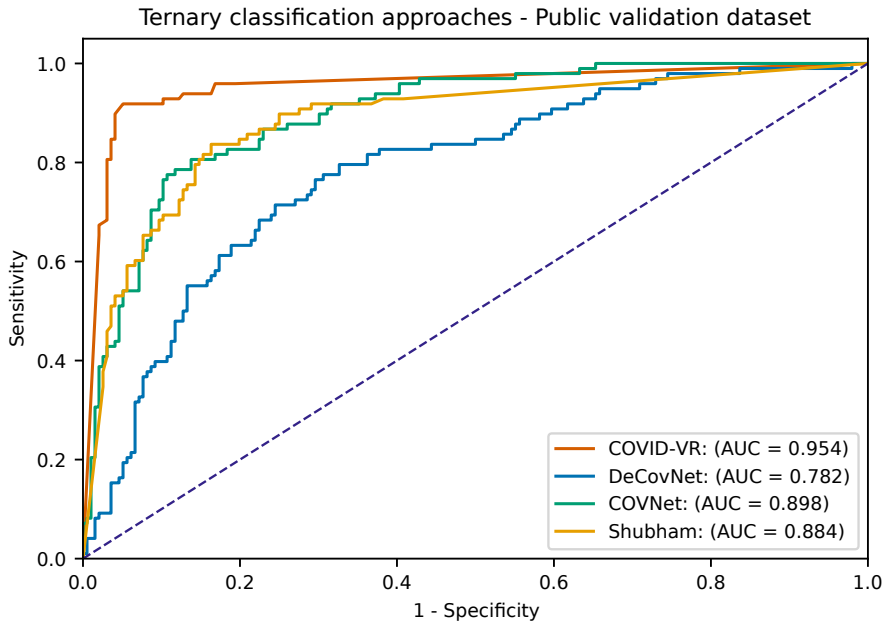
Our approach achieved the highest scores for overall performance in the validation set (Table 7.1): COVID-VR's accuracy was 90.8%, in contrast to 77.6% obtained with COVNet, ranked in second place. Regarding AUC score, our approach achieved 95.4 while COVNet achieved 89.8. Our method also had the best predictive power for COVID-19 cases, achieving high and balanced sensitivity and specificity values. We also observed the outstanding performance of our method when comparing the metrics computed for CAP and Normal cases: COVID-VR had a 100% sensitivity and 94% specificity for the Normal class, and 84.2% sensitivity and 97.5% specificity for the CAP class – in both cases surpassing the counterpart methods.

Table 7.1: Validation Set – Comparison among COVID-VR and state-of-the-art approaches for ternary classification using the VALIDATION sets from the public dataset provided by the SPGC 2021 competition.

	Metrics	COVID-VR	DeCovNet	COVNet	TheSaviours
Overall	Accuracy	**90.8%**	67.3%	77.6%	74.5%
	F1-Score	**90.8%**	66.8%	77.7%	75.5%
	Kappa	**84.6%**	44.9%	63.2%	58.8%
	AUC score	**95.4**	78.2	89.8	88.4
COVID-19	Sensitivity	**89.1%**	74.5%	74.5%	70.9%
	Specificity	**93.0%**	67.4%	83.7%	86.0%
	Precision	**94.2%**	74.5%	85.4%	86.7%
	F1-Score	**91.6%**	74.5%	79.6%	78.0%
Normal	Sensitivity	**100.0%**	41.7%	83.3%	75.0%
	Specificity	**94.6%**	**94.6%**	86.5%	77.0%
	Precision	**85.7%**	71.4%	66.7%	51.4%
	F1-Score	**92.3%**	52.6%	74.1%	61.0%
CAP	Sensitivity	**84.2%**	78.9%	78.9%	**84.2%**
	Specificity	**97.5%**	82.3%	93.7%	**97.5%**
	Precision	**88.9%**	51.7%	75.0%	**88.9%**
	F1-Score	**86.5%**	62.5%	76.9%	**86.5%**

Figure 7.3 compares the micro-average ROC curves for the ternary classification models using the validation test. The superior performance of COVID-VR in the validation set is clear, notably improving the true positive rate (*i.e.,* sensitivity) for false positive rates ranging from 0 to 0.3.

Figure 7.2: Micro-average ROC curves for the ternary classification task using the public dataset, considering the Validation set released by the SPGC 2021 competition.



7.1.2 Results in Test Set of Public Dataset

Regarding the test set (Table 7.2), COVID-VR achieved an accuracy of 86.7%, while TheSaviours correctly classified 90.0% of the test instances. Nonetheless, when analyzing the performance per class, our approach was the best among the four methods in detecting COVID-19 cases (94.3% sensitivity), keeping high specificity (92.7%), and F1-score (91.7%). In contrast, TheSaviours achieved the highest F1 score for Normal and CAP cases. It should be noted that our model is trained with labels at the patient level, while the approach presented by TheSaviours (CHAUDHARY et al., 2021) trains a model exploring labels at the slice level. Thus, we reach very competitive results despite employing a more coarse-grained annotation in CT scans.

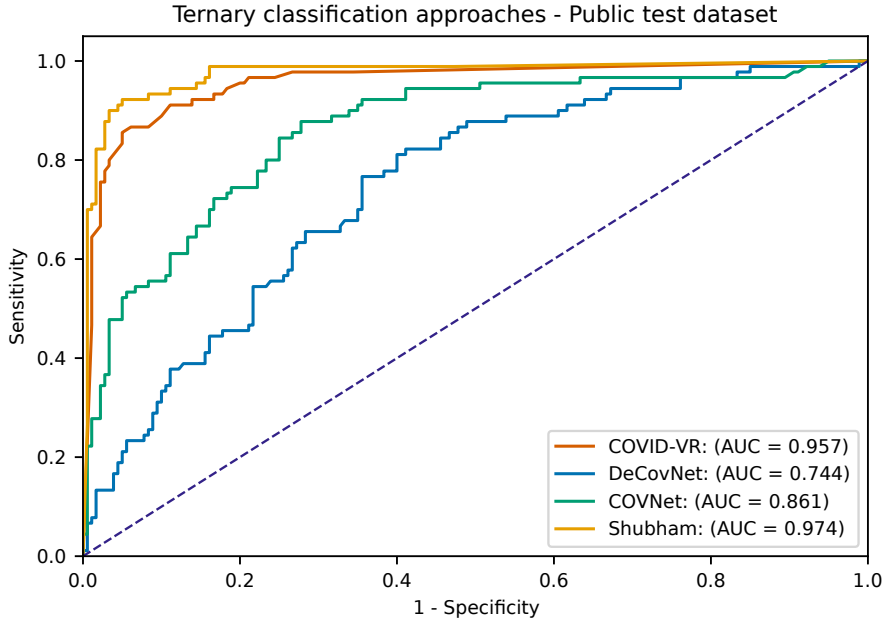
In Figure 7.3 we compare the micro-average ROC curves for the ternary classification models using the test set. Here, the performance of the TheSaviours model improves in relation to experiments with the validation set and surpasses our approach by about two points (*i.e.,* 97.4 vs. 95.7) in the AUC score. Nonetheless, we highlight that our approach had the most stable performance between validation and test sets, despite the clinical and technical differences introduced in the CT images from the COVID-CT-MD Test Set (HEIDARIAN et al., 2021).

Finally, we note that our approach had an accuracy close to that reported by the first places of the competition like IITDelhi (GARG et al., 2021) with 88.9%, LLSCP (YANG

Table 7.2: Test Set – Comparison among COVID-VR and state-of-the-art approaches for ternary classification using the TEST sets from the public dataset provided by the SPGC 2021 competition.

	Metrics	COVID-VR	DeCovNet	COVNet	TheSaviours
Overall	Accuracy	86.7%	52.2%	67.8%	**90.0%**
	F1-Score	86.8%	47.6%	67.7%	**90.0%**
	Kappa	79.7%	30.3%	50.0%	**84.6%**
	AUC score	95.7	74.4	86.1	**97.4**
COVID-19	Sensitivity	**94.3%**	57.1%	77.1%	85.7%
	Specificity	92.7%	61.8%	67.3%	**94.5%**
	Precision	89.2%	48.8%	60.0%	**90.9%**
	F1-Score	**91.7%**	52.6%	67.5%	88.2%
Normal	Sensitivity	77.1%	20.0%	57.1%	**94.3%**
	Specificity	**98.2%**	96.4%	87.3%	97.1%
	Precision	**96.4%**	77.8%	74.1%	89.2%
	F1-Score	85.7%	31.8%	64.5%	90.0%
CAP	Sensitivity	90.0%	**100.0%**	70.0%	90.0%
	Specificity	90.0%	71.4%	94.3%	**97.1%**
	Precision	72.0%	50.0%	77.8%	**90.0%**
	F1-Score	80.0%	66.7%	73.7%	**90.0%**

Figure 7.3: Micro-average ROC curves for the ternary classification task using the public dataset, considering the Test set released by the SPGC 2021 competition .



et al., 2021) with 87.8%, and UniSheff_EEE (XUE; ABHAYARATNE, 2021) with 85.56% reported previously in Table 3.1. The results for these approaches were not included in the table due to the lack of public code to reproduce the experiments when we developed this work.

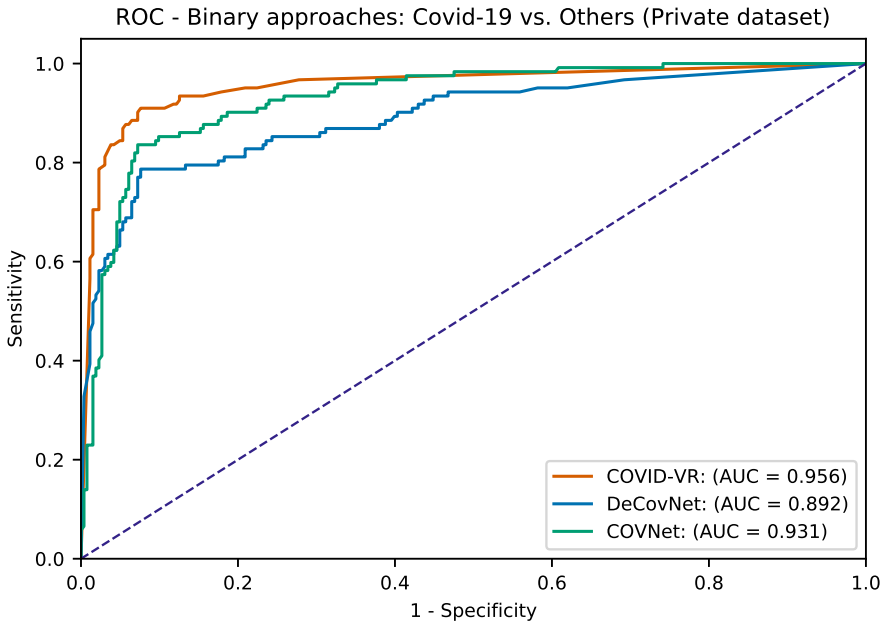
7.2 Classification Performance for the Private Dataset

We analyzed the methods' performance for a binary classification task, COVID-19 vs. non-COVID-19, using the private datasets. The experiments were run considering two different definitions for the negative class: in the first, we merged the Negative, Indeterminate, and Atypical classes into a unique non-COVID-19 class, and in the second, we considered only the original negative class (*i.e.,* Negative for pneumonia) as the classifiers' non-COVID-19 class. In both cases, the Typical classification was considered the positive class (*i.e.,* COVID-19). Performance assessment was based on a 5-fold CV, using the configuration of the same folds for COVID-VR, DeCoVNet, and COVNet. Results are presented in Table 7.3 demonstrating that our approach obtained the best results in all the metrics, obtaining 92.2% of accuracy and 95.6% of AUC for this binary classification task.

Table 7.3: Comparison of approaches in COVID-19 vs. Others task. Training and validation technique in the private (HMV+HCPA) dataset

Metrics	COVID-VR	DeCovNet	COVNet
Accuracy	**92.2%**	87.8%	89.4%
Sensitivity	**83.6%**	78.7%	83.6%
Specificity	**96.2%**	92.0%	92.0%
F1-score	**87.2%**	80.3%	83.3%
Cohen Kappa	**81.6%**	71.5%	75.5%
AUC	**95.6**	89.2	93.1

Figure 7.4: ROC curves comparison for binary classification task in private dataset

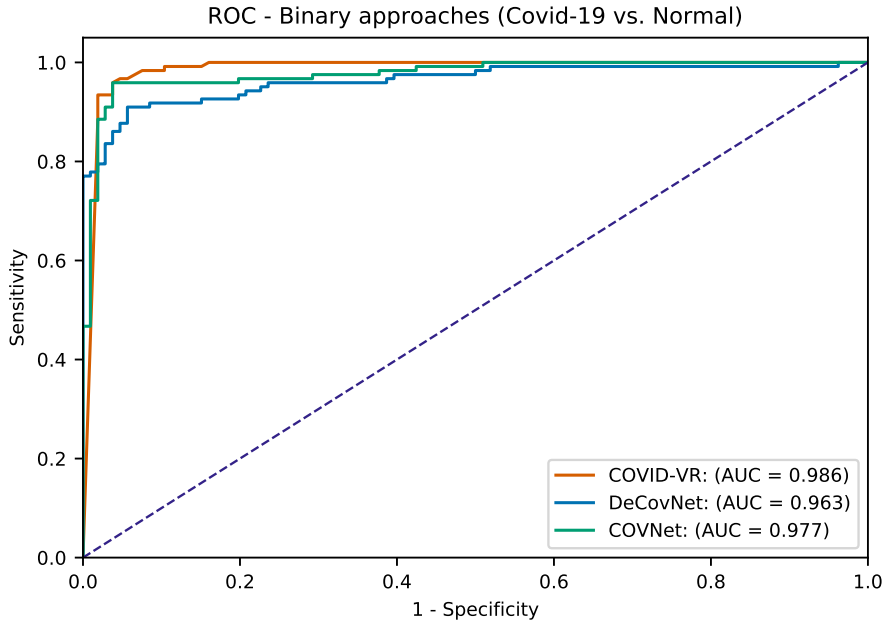


We also perform the comparison for a binary classification task of COVID-19 vs. Normal (*i.e.,* Negative for RSNA standard) in our private dataset. We compare our approach against DeCoVNet and COVNet using the 5-fold Cross-Validation technique (Table 7.4), achieving an accuracy of 96.1% in this task, similarly to COVNet and surpassing DeCoVNet all metrics, our method also get a greater value of AUC (98.6) against both other approaches as could be seen in micro-average ROC curves in Figure 7.5. It is worth noting that we execute the COVNet approach in another environment due to an error in memory in GPU allocation as mentioned in Section 6.1.

Table 7.4: Comparison of approaches in COVID-19 vs. Normal task. Training and validation technique in the private (HMV+HCPA) dataset

Metrics	**COVID-VR**	**DeCovNet**	**COVNet**
Accuracy	**96.1%**	92.5%	**96.1%**
Sensitivity	**96.7%**	91.0%	95.9%
Specificity	95.3%	94.3%	**96.2%**
F1-score	**96.3%**	92.9%	**96.3%**
Cohen Kappa	**92.1%**	85.1%	**92.1%**
AUC	**98.6**	96.3	97.7

Figure 7.5: ROC curves comparison for binary classification task (COVID-19 vs. Normal) in private dataset



7.2.1 Generalization Between Datasets

Finally, one of the questions we seek to answer in this work is the possible generalization and application of our trained models to other datasets. Thus, we analyze the generalization of the classification task of COVID-19 against Normal (Negative in RSNA standard), training a model with the preceding detailed parameters (ResNet101 as the backbone and using TF6 to render the segmented lungs).

We carry out the training and validation process with data distribution of 90% and 10%, respectively, using only our private dataset for this step. Afterward, we tested over the public dataset (COVID-CT-MD, training and validation set), obtaining an accuracy of 91.5% and 93.5% as F1-measure. Table 7.5 presents the rest of the metrics assuming COVID-19 as the positive class. We must highlight that we used only these two classes since they were the intersection labels in our datasets and argue that our approach achieves high values in generalization tasks despite the different sources of the datasets.

Table 7.5: Inference test of COVID-19 vs. Normal task classification using *Private Dataset* for training/validation and the Training/Validation Set of the COVID-CT-MD dataset (*Public Dataset*) to test.

Metrics	**COVID-VR**
Accuracy	91.5%
Precision	100.0%
Sensitivity	87.7%
Specificity	100.0%
F1-score	93.5%
Cohen Kappa	81.5%
AUC	96.7

It should be noted that we also did the reverse process of training our approach using the public dataset for the binary model with the split given by the competition. With this, we did not obtain metrics as high as those presented in the inverse process, for example, we reached an accuracy of 63.6% and an F1-score of 74.6%, and the other metrics are detailed in Table 7.6. One hypothesis is that the difference in values is the input data quality; private data has better image quality and higher resolution and number of slices.

Table 7.6: Inference test of COVID-19 vs. Normal task classification using *Public Dataset* for training/validation and the *Private Dataset* to test.

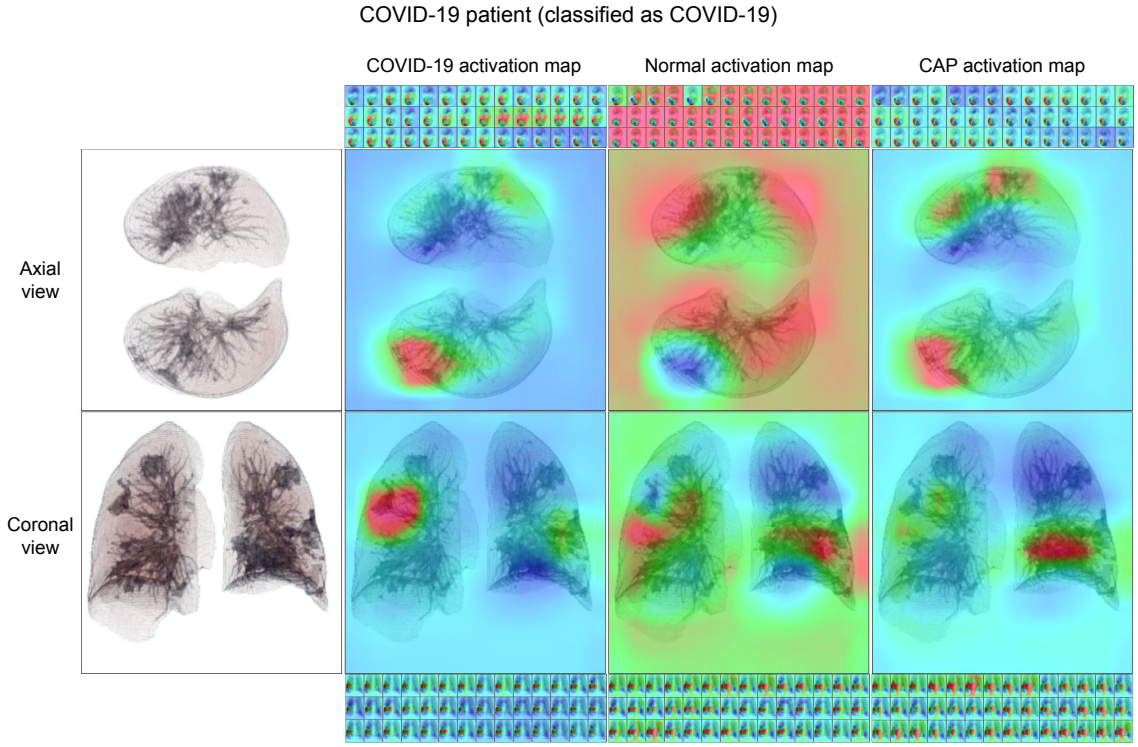
Metrics	**COVID-VR**
Accuracy	63.6%
Precision	59.5%
Sensitivity	100.0%
Specificity	21.7%
F1-score	74.6%
Cohen Kappa	22.9%
AUC	93.4

7.3 Visualization

To explain and intuitively verify the features our trained Convolutional Neural Network learned, we generated Grad-CAM visualizations (SELVARAJU et al., 2017). Figure 7.6 shows the last convolutional layer activation heatmaps for the three classes (COVID-19, Normal, and CAP) of our model of the axial and coronal view to a COVID-19 patient. Since our model works by classifying 42 images (rotated by small angles) per patient and view, each image generates an activation map for each class. Hence, Figure 7.6 shows these activation maps as thumbnails in the first row for the axial view and the fourth for the coronal view. In the second row, we display the simple mean of all activation maps (thumbnails shown in the first row) for each class overlapping the central image, point (0, 0) as shown in Figure 5.3, generated by our approach. The third row follows the same logic as the second row but for the coronal view.

In Figure 7.6 focusing on the second and third row, the first column shows the heatmaps' brightest areas (that most influenced the final classification) for COVID-19 class correspond to areas that are visibly injured in the central rendered image in axial and coronal views. In the second column, the Normal class activates almost the entire image avoiding one of the principal lesions from COVID-19 activation map in axial and coronal views and some other remarkable lesions in the coronal view. In the third column, we observe the activation of the CAP class for our case study, presenting that almost the entire image is activated but with less intensity than in the Normal class (aqua color to green color) and not centralizing the activation map only in the lesions but in their entire contour, unlike the activation map of the COVID-19 class.

Figure 7.6: Grad-CAM visualization results were obtained with our model, using a COVID-19 patient from the COVID-CT-MD dataset. The first column contains generated heatmaps indicating which areas from the input image most activated the model for the class COVID-19, leading it to classify that patient correctly. In contrast, the second and third column shows the heatmaps of activation maps for Normal and CAP class respectively. The first and last rows show thumbnails of all the activation maps for the 42 images generated by the patient; the second and third rows show the mean of the 42 activation maps for each view overlapping the central image generated by our model.



8 CONCLUSION AND FUTURE WORKS

The results of our experiments suggest that our proposed framework reasonably achieves the goal of learning to recognize typical COVID-19 patterns in chest CT images compared to other available methods that were tested. Our model can help specialists on COVID-19 diagnosis performing binary classification, returning a quick opinion that may constitute valuable information, as expressed by the feedback from radiologists of our partner hospitals, helping to identify typical cases or excluding the possibility of that diagnosis. We recognize that our ternary classification does not achieve results that may indicate the option of applying our classifier to segment the incoming patients into three classes in a real situation; however, we emphasize that our proposed approach gets a competitive accuracy using only patient-level annotations.

Among the possibilities for future work, we would like to explore additional transfer function generation methods. As discussed in this work, our method was a process of trial and error until reaching a transfer function that satisfied our requirements. We want to optimize the generation and selection of a transfer function using Deep Learning methods to choose the appropriate transfer function.

Another point left open for investigation is whether we could use the generated lung videos for patient classification models. In this way, having one video per patient could eliminate the final count part of our pipeline.

The classification of the four classes presented by RSNA is also pending for future tasks. Finally, the possibility of using volume rendering for classification tasks in other contexts like classification in other organs with other diseases (for example, tumors).

REFERENCES

AFSHAR, P. et al. COVID-CT-MD: COVID-19 Computed Tomography (CT) Scan Dataset Applicable in Machine Learning and Deep Learning. 9 2020. Available from Internet: <http://arxiv.org/abs/2009.14623>.

AFSHAR, P. et al. COVID-CT-MD, COVID-19 computed tomography scan dataset applicable in machine learning and deep learning. **Scientific Data**, Nature Research, v. 8, 12 2021. ISSN 20524463.

AMARA, K. et al. COVIR: A virtual rendering of a novel NN architecture O-Net for COVID-19 Ct-scan automatic lung lesions segmentation. **Computers & Graphics**, v. 104, p. 11–23, 2022. ISSN 0097-8493.

ARENS, S.; DOMIK, G. A survey of transfer functions suitable for volume rendering. In: **VG@ Eurographics**. [S.l.: s.n.], 2010. p. 77–83.

AREVALO-RODRIGUEZ, I. et al. False-negative results of initial rt-pcr assays for covid-19: a systematic review. **PloS one**, Public Library of Science San Francisco, CA USA, v. 15, n. 12, p. e0242958, 2020.

BOUGOURZI, F. et al. CNR-IEMN: A Deep Learning Based Approach to Recognise Covid-19 from CT-Scan. In: **ICASSP 2021 - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.: s.n.], 2021. p. 8568–8572.

CHAUDHARY, S. et al. Detecting covid-19 and community acquired pneumonia using chest ct scan images with deep learning. In: **ICASSP 2021 - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.: s.n.], 2021. p. 8583–8587.

DALRYMPLE, N. C. et al. Introduction to the language of three-dimensional imaging with multidetector ct. **Radiographics**, Radiological Society of North America, v. 25, n. 5, p. 1409–1428, 2005.

FALCAO, A. X.; STOLFI, J.; LOTUFO, R. de A. The image foresting transform: Theory, algorithms, and applications. **IEEE transactions on pattern analysis and machine intelligence**, IEEE, v. 26, n. 1, p. 19–29, 2004.

GARG, P. et al. Multi-scale residual network for covid-19 diagnosis using ct-scans. In: **ICASSP 2021 - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.: s.n.], 2021. p. 8558–8562.

GOZES, O. et al. Rapid AI development cycle for the coronavirus (COVID-19) pandemic: Initial results for automated detection & patient monitoring using deep learning CT image analysis. **arXiv**, 2020. ISSN 23318422. Available from Internet: <http://arxiv.org/abs/2003.05037>.

HAN, Z. et al. Accurate screening of covid-19 using attention-based deep 3d multiple instance learning. **IEEE transactions on medical imaging**, IEEE, v. 39, n. 8, p. 2584–2594, 2020.

HARRISON, A. P. et al. Progressive and multi-path holistically nested neural networks for pathological lung segmentation from ct images. In: **Medical Image Computing and Computer Assisted Intervention (MICCAI)**. Cham: Springer International Publishing, 2017. p. 621–629. ISBN 978-3-319-66179-7.

HE, K. et al. Deep Residual Learning for Image Recognition. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. [S.l.: s.n.], 2016. p. 770–778.

HE, X. et al. Sample-Efficient Deep Learning for COVID-19 Diagnosis Based on CT Scans. **medRxiv**, Cold Spring Harbor Laboratory Press, 2020.

HEIDARIAN, S. et al. Robust automated framework for covid-19 disease identification from a multicenter dataset of chest ct scans. **arXiv preprint arXiv:2109.09241**, 2021.

HOFMANNINGER, J. et al. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. **European Radiology Experimental**, SpringerOpen, v. 4, n. 1, p. 1–13, 2020.

HOFMANNINGER, J. et al. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. **European Radiology Experimental**, SpringerOpen, v. 4, n. 1, p. 1–13, 2020.

HUANG, G. et al. Densely Connected Convolutional Networks. In: **2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Los Alamitos, CA, USA: IEEE Computer Society, 2017. p. 2261–2269. ISSN 1063-6919.

INUI, S. et al. Comparison of chest ct grading systems in covid-19 pneumonia. **Radiology: Cardiothoracic Imaging**, Radiological Society of North America, v. 2, n. 6, p. e200492, 2020.

JADHAV, S. et al. <italic>covid</italic>-view: Diagnosis of covid-19 using chest ct. **IEEE Transactions on Visualization and Computer Graphics**, v. 28, n. 1, p. 227–237, 2022.

JIN, C. et al. Development and evaluation of an artificial intelligence system for COVID-19 diagnosis. **Nature Communications**, Springer US, v. 11, n. 1, 2020. ISSN 20411723.

J¨ONSSON, D. et al. A survey of volumetric illumination techniques for interactive volume rendering. In: WILEY ONLINE LIBRARY. **Computer Graphics Forum**. [S.l.], 2014. v. 33, n. 1, p. 27–51.

KNISS, J.; KINDLMANN, G.; HANSEN, C. Multidimensional transfer functions for interactive volume rendering. **IEEE Transactions on visualization and computer graphics**, IEEE, v. 8, n. 3, p. 270–285, 2002.

LENSINK, K. et al. Segmentation of pulmonary opacification in chest ct scans of covid-19 patients. **arXiv preprint arXiv:2007.03643**, 2020.

LI, B. et al. Diagnosing COVID-19 from CT Images Based on an Ensemble Learning Framework. In: **ICASSP 2021 - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.: s.n.], 2021. p. 8563–8567.

LI, G. et al. Multiscale 3-dimensional pathology findings of COVID-19 diseased lung using high-resolution cleared tissue microscopy. **Biorxiv**, Cold Spring Harbor Laboratory, 2020.

LI, L. et al. Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. **Radiology**, v. 296, p. E65–E71, 2020. ISSN 15271315.

LU, W. et al. Quantitative ct for detecting covid-19 pneumonia in suspected cases. **BMC infectious diseases**, Springer, v. 21, n. 1, p. 1–8, 2021.

OPITZ, J.; BURST, S. Macro f1 and macro f1. **arXiv preprint arXiv:1911.03347**, 2019.

ROBERTS, M. et al. Common Pitfalls And Recommendations For Using Machine Learning To Detect And Prognosticate For COVID-19 Using Chest Radiographs and CT Scans. **Nature Machine Intelligence**, v. 3, n. 3, p. 199–217, 3 2021. ISSN 2522-5839.

RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. **International Conference on Medical image computing and computer-assisted intervention**. [S.l.], 2015. p. 234–241.

SELVARAJU, R. R. et al. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: **Proceedings of the IEEE International Conference on Computer Vision (ICCV)**. [S.l.: s.n.], 2017.

SHAMIM, S. et al. Automatic covid-19 lung infection segmentation through modified unet model. **Journal of Healthcare Engineering**, Hindawi, v. 2022, 2022.

SILVA, C. J. A. da; GELLADA, N. Cinematic rendering for three-dimensional reconstructions of the chest wall: a new reality. **Einstein Journal**, v. 18, n. eMD5223, 2020.

SIMONYAN, K.; ZISSERMAN, A. **Very Deep Convolutional Networks for Large-Scale Image Recognition**. 2015.

SIMPSON, G. Thoracic computed tomography: principles and practice. 2009.

SIMPSON, S. et al. Radiological society of north america expert consensus document on reporting chest ct findings related to covid-19: endorsed by the society of thoracic radiology, the american college of radiology, and rsna. **Radiology: Cardiothoracic Imaging**, Radiological Society of North America, v. 2, n. 2, p. e200152, 2020.

SOUSA, A. M. et al. ALTIS: A fast and automatic lung and trachea CT-image segmentation method. **Medical Physics**, v. 46, p. 4970–4982, 2019. ISSN 00942405.

TAN, M.; LE, Q. V. **EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks**. 2020.

TANG, L. et al. Severe COVID-19 Pneumonia: Assessing Inflammation Burden with Volume-rendered Chest CT. **Radiology: Cardiothoracic Imaging**, v. 2, n. 2, p. e200044, 2020.

TANG, Z. et al. Severity assessment of coronavirus disease 2019 (covid-19) using quantitative features from chest ct images. **arXiv preprint arXiv:2003.11988**, 2020.

WANG, L.; LIN, Z. Q.; WONG, A. COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. **Scientific Reports**, Nature Research, v. 10, 12 2020. ISSN 20452322.

WANG, X. et al. A Weakly-Supervised Framework for COVID-19 Classification and Lesion Localization From Chest CT. **IEEE Transactions on Medical Imaging**, v. 39, n. 8, p. 2615–2625, 2020.

WOLF, I. et al. The medical imaging interaction toolkit (mitk): a toolkit facilitating the creation of interactive software by extending vtk and itk. In: SPIE. **Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display**. [S.l.], 2004. v. 5367, p. 16–27.

XUE, S.; ABHAYARATNE, C. Covid-19 diagnostic using 3d deep transfer learning for classification of volumetric computerised tomography chest scans. In: **ICASSP 2021 - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.: s.n.], 2021. p. 8573–8577.

YANG, Z. et al. A Multi-Stage Progressive Learning Strategy for Covid-19 Diagnosis Using Chest Computed Tomography with Imbalanced Data. In: **ICASSP 2021 - IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)**. [S.l.: s.n.], 2021. p. 8578–8582.

YE, Z. et al. Chest CT manifestations of new coronavirus disease 2019 (COVID-19): a pictorial review. **European Radiology**, Springer, v. 30, n. 8, p. 4381–4389, 2020.

ZHANG, Q.; EAGLESON, R.; PETERS, T. M. Volume visualization: a technical overview with a focus on medical applications. **Journal of digital imaging**, Springer, v. 24, n. 4, p. 640–664, 2011.

ZHAO, H. et al. Pyramid scene parsing network. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2017. p. 2881–2890.

APPENDIX A — DETAILED RESULTS FOR COVID-VR

Table A.1: Comparison of principal results obtained in this work using the Public Dataset COVID-CT-MD in ternary classification. All values in table are in percentage.

		By Classes														Mean				
		General		Covid-19				Normal				CAP				Weighted (AUC Micro)				
		Accu.	Kappa	Preci.	Sensi.	Speci	F-score	Preci.	Sensi.	Speci.	F-score	Preci.	Sensi.	Speci.	F-score	Preci.	Sensi.	Speci.	F-score	AUC
Transfer Function Comparison	TF 1	79.6	63.7	80.3	89.1	72.1	84.5	71.4	83.3	**89.2**	76.9	**100.0**	47.4	**100.0**	64.3	81.9	79.6	81.7	78.7	91.8
	TF 2	85.7	**74.0**	81.5	**96.4**	**72.1**	88.3	**95.7**	**91.7**	**98.6**	**93.6**	90.0	47.4	98.7	62.1	86.6	85.7	**83.7**	84.5	**96.3**
	TF 3	87.8	79.2	90.7	89.1	88.4	89.9	85.2	95.8	**94.6**	90.2	**82.4**	73.7	**96.2**	77.8	87.7	87.8	91.4	87.6	94.7
	TF 4	88.8	80.8	90.9	90.9	88.4	90.9	85.2	95.8	94.6	90.2	87.5	73.7	97.5	80.0	88.8	88.8	91.7	88.6	95.2
	TF 5	85.7	75.4	87.5	89.1	83.7	88.3	82.1	95.8	93.2	88.5	85.7	63.2	97.5	72.7	85.8	85.7	88.7	85.3	95.0
	TF 6	**90.8**	**84.6**	**94.2**	**89.1**	**93.0**	**91.6**	85.7	**100.0**	94.6	**92.3**	88.9	**84.2**	**97.5**	**86.5**	**91.1**	**90.8**	**94.3**	**90.8**	**95.4**
Architecture Comparison	VGG16	88.8	**81.0**	90.7	**89.1**	**88.4**	**89.9**	**88.9**	**100.0**	**95.9**	**94.1**	82.4	73.7	96.2	77.8	88.6	88.8	**91.7**	88.6	**95.6**
	DenseNet121	87.8	79.7	**94.0**	85.5	**93.0**	**89.5**	82.8	**100.0**	93.2	90.6	78.9	78.9	94.9	78.9	88.3	87.8	93.4	87.7	96.5
	EfficientNetB2	86.7	78.0	92.0	83.6	90.7	**87.6**	85.7	**100.0**	94.6	92.3	75.0	78.9	93.7	76.9	87.2	86.7	92.2	86.7	95.1
	ResNet101	**90.8**	**84.6**	**94.2**	**89.1**	**93.0**	**91.6**	85.7	**100.0**	**94.6**	**92.3**	**88.9**	**84.2**	**97.5**	**86.5**	**91.1**	**90.8**	**94.3**	**90.8**	**95.4**
Method Comparison (Train / Validation)	DeCovNet	67.3	44.9	74.5	74.5	67.4	74.5	**71.4**	41.7	**94.6**	52.6	51.7	78.9	82.3	62.5	69.3	67.3	76.9	66.8	78.2
	COVNet	77.6	63.2	85.4	74.5	83.7	79.6	66.7	83.3	86.5	74.1	75.0	78.9	93.7	76.9	78.8	77.5	86.3	77.7	89.8
	TheSaviours	74.5	58.8	86.7	70.9	86.0	78.0	51.4	75.0	**77.0**	**61.0**	**88.9**	**84.2**	**97.5**	**86.5**	78.5	74.5	86.0	75.5	88.4
	COVID-VR	**90.8**	**84.6**	**94.2**	**89.1**	**93.0**	**91.6**	**85.7**	**100.0**	**94.6**	**92.3**	**88.9**	**84.2**	**97.5**	**86.5**	**91.1**	**90.8**	**94.3**	**90.8**	**95.4**
Method Comparison (Test)	DeCovNet	52.2	30.3	48.8	57.1	61.8	52.6	77.8	20.0	96.4	**31.8**	50.0	**100.0**	71.4	66.7	60.3	52.2	77.4	47.6	74.4
	COVNet	67.8	50.0	60.0	77.1	67.3	67.5	74.1	57.1	87.3	64.5	77.8	70.0	94.3	73.7	69.4	67.7	81.1	67.7	86.1
	TheSaviours	**90.0**	**84.6**	**90.9**	85.7	**94.5**	88.2	89.2	**94.3**	**97.1**	**90.0**	**90.0**	**90.0**	**97.1**	**90.0**	**90.0**	**90.0**	**94.4**	**90.0**	**97.4**
	COVID-VR	86.7	**79.7**	89.2	**94.3**	**92.7**	**91.7**	**96.4**	77.1	**98.2**	85.7	72.0	90.0	90.0	80.0	88.2	86.7	94.2	86.8	95.7

Table A.2: Comparison of principal results obtained in this work using the Public Dataset COVID-CT-MD. All values in table are in percentage, continue from Table A.1 with Macro values

		Mean-Macro				
		Preci.	Sensi.	Speci.	F-score	AUC
Transfer Function Comparison	TF 1	86.1	86.2	93.1	86.0	94.1
	TF 2	89.1	78.5	89.8	81.3	**95.2**
	TF 3	83.9	73.3	87.1	75.2	90.2
	TF 4	87.9	86.8	93.5	87.0	94.7
	TF 5	85.1	82.7	95.1	83.2	95.2
	TF 6	**89.6**	**91.1**	**95.0**	**90.1**	**95.2**
Architecture Comparison	VGG16	87.3	87.6	93.5	87.3	94.9
	DenseNet121	85.2	88.1	93.7	86.3	**95.7**
	EfficientNetB2	84.2	87.5	93.0	85.6	94.0
	ResNet101	**89.6**	**91.1**	**95.0**	**90.1**	95.2
Method Comparison (Train / Validation)	DeCoVNet	65.9	65.0	81.4	63.2	76.8
	COVNet	75.7	78.9	88.0	76.9	89.3
	TheSaviours	75.7	76.7	86.8	75.2	87.7
	COVID-VR	**89.6**	**91.1**	**95.0**	**90.1**	**95.2**
Method Comparison (Test)	DeCoVNet	58.9	59.0	76.5	50.4	78.7
	COVNet	70.6	68.1	83.0	68.6	86.6
	TheSaviours	**90.0**	**90.0**	**94.8**	**90.0**	**98.0**
	COVID-VR	85.9	87.1	93.6	85.8	96.1

APPENDIX B — PROPOSED DOCKER IMAGE AND SUPPLEMENTARY REPOSITORIES

This appendix summarizes all the links to our repositories and results. Each repository contains an example of how to execute the included source code.

B.1 Full Pipeline

Our full pipeline presented and detailed in Chapter 5 with its requirements and examples of how to run are available in:

- **<https://github.com/covid-vr/covid-vr-docker>**.

B.2 Independent Repositories

In this Section, we detail the main components of our Full Pipeline, which can be used independently. Each repository contains an example of how to execute the code.

B.2.1 Lung segmentation

Required packages and minimum versions. For complete detail, see at requirements.txt in our adapted P-HNN version:

- Python 3.6
- Pytorch 1.3
- Cudatoolkit 10.1
- Anaconda

Link to repositories:

- Original P-HNN repository: <*https://adampharrison.gitlab.io/p-hnn/*>
- Our adapted P-HNN version: <*https://github.com/covid-vr/p-hnn-lung-segmentation*>
- Model Weights: <*https://drive.google.com/file/d/1l6yLFScULNw-oVoark0KZ-wnDFX8zwrN/view?usp=sharing*>

B.2.2 Visualization Repositories

Minimum required libraries:

- CMake 3.17
- QT 5.12
- MITK: <https://github.com/MITK/MITK>
- ffmpeg (For video-generator repository)

Link to repositories:

- Repository to capture view images, for step in Chapter 5.4: *<https://github.com/covid-vr/camera-shots-generator>*
- Repository for video generation from CT Image: *<https://github.com/covid-vr/video-generator>*

Both repositories need as input an image in NIfTI format and a XML for the transfer functions, samples of both were added to each repository. We highly recommend using a Desktop environment (with UI); however, it could be used in a server environment like our Full Pipeline, using VGLRUN command along with the steps described in B.1.

B.2.3 COVID-VR Proposed Network

Required minimum versions (the complete requirements are in the repository link):

- Python 3.6
- TensorFlow 2.0

Link to repositories:

- Repository for train and validation: *<https://github.com/covid-vr/covid-vr-network>*
- Model weights (For the two views): *<https://drive.google.com/drive/folders/1OXTliIhm7yGuBDIL7qZhQrjCoaxmGx0l>*

B.3 Supplementary Material

Link to repositories:

- Repository for get metrics (accuracy, precision, f1-measure, etc) and generate graphics used in this work: <*https://github.com/covid-vr/model-evaluation-metrics*>
- Repository to generate Grad-CAM visualizations: <*https://github.com/covid-vr/covid-vr-grad-cam*>
- Complete result by patients and images used in this work: <*https://covid-vr.github.io/*>

APPENDIX C — RESUMO EXPANDIDO

Título da Dissertação de Mestrado: COVID-VR um modelo de classificação de aprendizado profundo usando Renderização de Volumes de Tomografias Computarizadas.

Resumo expandido:

O contexto de COVID-19 apresentou uma série de desafios aos sistemas de saúde globais. Como a maioria dos pacientes com COVID-19 tem infecções pulmonares, as tomografias computadorizadas (TC) do tórax foram frequentemente utilizadas para diagnosticar infecções por COVID-19, bem como outras doenças pulmonares. Arquiteturas de aprendizado profundo foram desenvolvidas para identificar automaticamente classes de doenças pulmonares usando fatias de TC como entradas para algoritmos de classificação.

Neste trabalho apresentamos o COVID-VR, um modelo de classificação de tomografias computarizadas (TC) baseado em visualizações volumétricas para a identificação de pacientes com a presença de COVID-19, outras pneumonias ou ausência de infecções pulmonares. Comparamos nosso modelo com outros trabalhos, DeCoVNet (WANG et al., 2020), COVNet (LI et al., 2020b) e TheSaviours (CHAUDHARY et al., 2021) que também oferecem classificação em nível de paciente. Publicamos nossos resultados avaliando as métricas em um conjunto de dados público para comparações justas com as outras abordagens.

COVID-VR possui duas etapas principais, a parte de pré-processamento e a parte de treinamento de nossas redes e obtenção de resultados.

1. **PRÉ-PROCESSAMENTO.** Para esta etapa recebemos como entrada o TC no formato DICOM que transformamos para o formato NIfTI. Em seguida, segmentamos o TC para obter apenas a região de interesse (pulmão) em nossa imagem NIfTI. Posteriormente, usando scripts baseados em MITK, realizamos várias capturas do eixo axial superior e frontal coronal do volume renderizado.
2. **TREINAMENTO E OBTENÇÃO DE RESULTADOS.** Nesta etapa, treinamos duas redes convolucionais baseadas no ResNet101, uma para as imagens axiais e outra para as imagens coronais obtidas na etapa anterior. Na saída dessas redes nós obtemos uma classificação em nível de imagem e performamos uma voto majoritário para obter a classificação final no nível de TC.

No decorrer do desenvolvimento desta dissertação utilizamos dois tipos de dados e três modelos principais de classificação. O primeiro conjunto de dados COVID-CT-

MD (AFSHAR et al., 2020) são dados públicos compostos por dois subconjuntos, treinamento/validação e teste, divulgados pela competição ICASSP 2021. O segundo conjunto de dados é um conjunto de dados privado obtido de dois hospitais locais, Hospital Moinhos de Vento e Hospital de Clínicas de Porto Alegre. Pela natureza dos dados, o conjunto COVID-CT-MD consiste em três classes, COVID-19, CAP (pneumonias adquiridas na comunidade) e Normal (indica a ausência de infecções pulmonares); enquanto em nosso *Dataset Privado* possui as quatro classificações dadas pela RSNA no contexto da pandemia; Positivo (COVID-19), Indeterminado, Atípico e Negativo (ausência de doença).

Dada a diversidade de nossos conjuntos de dados, os estudos de caso a seguir foram orientados para classificação binária e ternária.

- **CLASSIFICAÇÃO BINÁRIA.** Para esta classificação, usamos nosso *Dataset Privado* e os agrupamos em dois experimentos principais usando a abordagem de validação cruzada de 5 folds. O primeiro caso consiste na classificação de COVID-19 vs. Negativo (Normal) onde nosso método consegue superar os resultados obtidos usando DeCoVNet e obter métricas semelhantes a COVNet. O segundo caso consiste na classificação de COVID-19 vs. Outros (agrupamento das classes Negativo, Indeterminado e Atípico), onde nosso método COVID-VR foi capaz de superar os resultados obtidos por DeCoVNet e COVNet em todas as métricas apresentadas.
- **CLASSIFICAÇÃO TERNÁRIA.** Nesta classificação usamos o conjunto COVID-CT-MD. Treinamos as abordagens usando o subconjunto de treinamento/validação com as divisões oferecidas para COVID-VR, DeCoVNet e COVNet; e seguindo os passos descritos na abordagem de treinamento porposto por TheSaviours. Deste primeiro subconjunto, COVID-VR consegue superar as outras abordagens em métricas (acurácia, micro F1-score) de classe e geral em vários pontos percentuais. Posteriormente, esses modelos treinados por COVID-VR, DeCoVNet e COVNet, foram testados com o conjunto de testes da competição, para abordagem TheSaviours usamos o modelo que eles divulgaram com o qual venceram a competição. A partir desses resultados, podemos mostrar que nossa abordagem superou DeCoVNet e COVNet, mas na métrica geral perdeu para TheSaviours em alguns pontos percentuais. No entanto, deve-se notar que nosso método obteve uma sensibilidade maior para a classe COVID-19 do que as outras abordagens e também que nossa abordagem é um modelo que trabalha apenas com anotações em nível de paciente enquanto TheSaviours precisa de anotações em nível de fatias. Demonstrando que o COVID-VR possui resultados de nível competitivo para esta tarefa.

Além dos experimentos citados, realizamos um teste de generalização do nosso modelo que consiste em treinar nossas redes com os dados do nosso *Dataset Privado* e testá-los no conjunto público COVID-CT-MD. Para este experimento utilizamos apenas as classes Positiva e Normal, pois são os dados que temos em comum entre os dois conjuntos. Obtemos 91,5% de acurácia, destacando assim a alta generalização do nosso método para esses conjuntos.

Por fim, apresentamos visualizações usando o Grad-CAM para mostrar as regiões ativadas em nossas imagens de entrada por classe e validar assim quais caracteristicas são as principais para cada classificação.