COVID-19 DIAGNOSIS AND SEGMENTATION USING MACHINE LEARNING ANALYSES OF LUNG COMPUTERIZED TOMOGRAPHY

Bhuvan Mittal

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

August 2021

APPROVED:

JungHwan Oh, Major Professor
Song Fu, Committee Member
Qing Yang, Committee Member
Bill Buckles, Committee Member
Yan Huang, Chair of Department of Computer Science and Engineering
Hanchen Huang, Dean, College of Engineering
Victor Prybutok, Dean of the Toulouse Graduate School

Mittal, Bhuvan. *COVID-19 Diagnosis and Segmentation Using Machine Learning Analyses of Lung Computerized Tomography*. Doctor of Philosophy (Computer Science and Engineering), August 2021, 127 pp., 41 tables, 46 figures, 1 appendix, 139 numbered references.

COVID-19 is a highly contagious and virulent disease caused by the severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2). COVID-19 disease induces lung changes observed in lung computerized tomography (CT) and the percentage of those diseased areas on the CT correlates with the severity of the disease. Therefore, segmentation of CT images to delineate the diseased or lesioned areas is a logical first step to quantify disease severity, which will help physicians predict disease prognosis and guide early treatments to deliver more positive patient outcomes. It is crucial to develop an automated analysis of CT images to save their time and efforts. This dissertation proposes CoviNet, a deep three-dimensional convolutional neural network (3D-CNN) to diagnose COVID-19 in CT images. It also proposes CoviNet Enhanced, a hybrid approach with 3D-CNN and support vector machines. It also proposes CoviSegNet and CoviSegNet Enhanced, which are enhanced U-Net models to segment ground-glass opacities and consolidations observed in computerized tomography (CT) images of COVID-19 patients. We trained and tested the proposed approaches using several public datasets of CT images. The experimental results show the proposed methods are highly effective for COVID-19 detection and segmentation and exhibit better accuracy, precision, sensitivity, specificity, F-1 score, Matthew's correlation coefficient (MCC), dice score, and Jaccard index in comparison with recently published studies.

ACKNOWLEDGMENTS

I owe my sincere gratitude to my Major Professor, Dr. JungHwan Oh. His constant guidance, prompt feedback and invaluable suggestions made this research feasible. I am highly thankful to Dr. JungHwan Oh for his constant support. I respect him a lot for having sparked my interest in research.

I am highly appreciative of my Ph.D. Committee Members: Dr. Bill Buckles, Dr. Song Fu and Dr. Qing Yang for their constructive feedback and reviews during the dissertation research work. The Department of Computer Science and Engineering at the University of North Texas has offered me excellent learning opportunities such as participating in national and international conferences to further my studies and research work.

I am also indebted to all my parents for serving as excellent role models and providing constant encouragement and inspiration. They have always served as a role model, as they are highly respected medical researchers. I am highly thankful to my sisters for supporting me in countless ways on a daily basis during my scholarship journey. My adorable daughter has also been highly supportive of my studies, and I cannot thank her enough for allowing me to work diligently on my research. Her unmeasurable sacrifices will hopefully motivate other women to pursue their dreams.

I am also very thankful to my friends, neighbors, colleagues, mentors, and several other folks. I am also thankful to all the researchers, physicians, and patients whose sacrifices and contributions made this research feasible, and hopefully of some miniscule significance.

And above all, I dedicate this work to Sri Ramakrishna with humility and thank him for guiding me through tumultuous and peaceful times.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION: COVID-19 CLASSIFICATION AND SEGMENTATION

1.1 What is COVID-19?

COVID-19 is a highly infectious and virulent disease with human-to-human transmission [78] and has flu-like symptoms caused by the severe acute respiratory syndrome - corona virus – 2 (SARS-CoV-2). It can cause a severe inflammatory response in the lungs with SARS-like symptoms. COVID-19 was declared a global pandemic in March 2020 and has caused over 175 million cases and 3.8 million deaths worldwide as of June 15, 2021 [1].

1.2 Why COVID-19 Detection and Segmentation via Lung Imaging?

A multinational consensus from the Fleischner Society reported that detecting patients at an early stage and isolating them from the people with a high risk of exposure is essential. X-ray, computerized tomography (CT) and ultrasound are three imaging modalities we are aware of with published research for the detection of COVID-19 from lung imaging. We chose the CT modality for this research per advice from an expert radiologist since more lung tissue details are seen in CT images.

Lung computerized tomography (CT) is a relevant tool for this purpose due to its high sensitivity in detecting early pneumonic changes. It can also contribute to the management and triage of the disease by detecting severe cases [2]. However, this classification involves radiologists' time and efforts significantly. It is crucial to develop an automated analysis of CT images to save their time and efforts.

However, COVID-19 and other types of pneumonia in CT have similar imaging

characteristics, which can hamper correct diagnosis by radiologists. A radiologists' performance study in differentiating COVID-19 from other viral pneumonia reported that the median values of sensitivity and specificity were 83% (ranging 67%-97%) and 96.5% (ranging 7%-100%), respectively [3]. Therefore, it is crucial to develop an automated analysis of CT images to save radiologists' time and efforts, and to increase the classification sensitivity and specificity.

COVID-19 disease induces lung changes observed in lung computerized tomography (CT) or radiographic images and those predominantly include ground-glass opacification (GGO) with occasional consolidation in the peripheral regions of the lungs [54]. The Fleischner Society glossary of terms [57] defines ground-glass opacities as an increase in opacification of the lung that does not obscure the blood vessels and airways. The Fleischner Society defines consolidation as a homogeneous opacification that obscures blood vessels and airway walls.

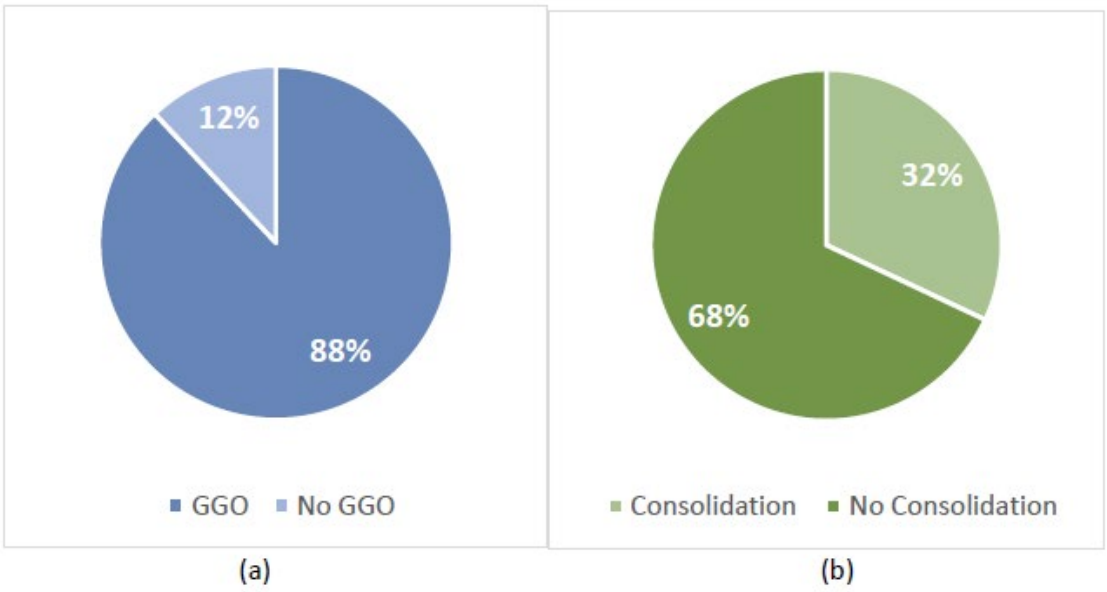

Figure 1.1: COVID-19 lung changes: (a) percentage of COVID-19 positive patients' CT showing GGO, (b) percentage of COVID-19 positive patients' CT showing consolidation.

Anatomically, the COVID-19 disease causes the little sacs of lung tissue called alveoli to collect fluid, which causes inflammation of the lungs. CT imaging shows this inflammation as

ground-glass opacities (GGOs) which further progress into consolidations. Salehi et al.'s [55] review of 30 medical studies consisting of 919 patients revealed that 88% of COVID-19 positive patients' CT scans showed ground-glass opacification (GGO) and 32% showed consolidation as shown in Figure 1.1.

Additionally, the percentage of the lung affected by ground-glass opacities and consolidation on the computerized tomography is a measure of the severity of the disease [58]. Therefore, segmentation of CT images to delineate the diseased or lesioned areas is a logical first step in helping physicians quantify disease severity and disease prognosis [33] as shown in Figure 1.2.

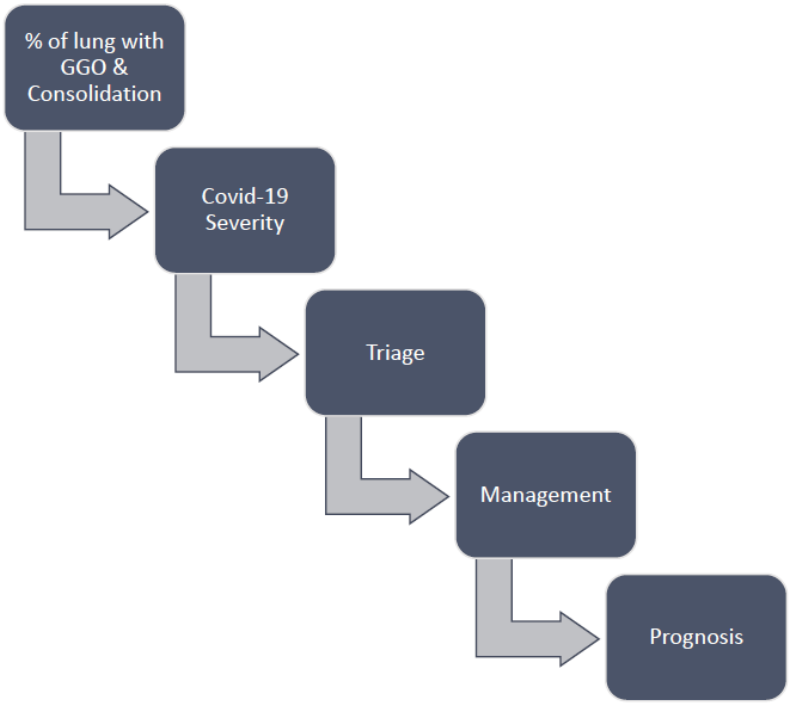

Figure 1.2: Why COVID-19 segmentation?

To summarize, the artificial intelligence-powered COVID-19 diagnosis techniques applied to lung CT can help save radiologist's time and can provide an efficient and early diagnosis and severity assessment for COVID-19.

1.3 The Research Problem Addressed in This Dissertation.

The major research problem that is addressed in this dissertation is automatic early COVID-19 detection and segmentation. This problem is subdivided into different phases which include the following:

- Can DNN based methods help in efficient and effective COVID-19 detection? What approaches could show a superior generalizability?
- How can COVID-19 detection be accomplished efficiently from lung imaging to assist radiologists in diagnosis?
- How can the diseased areas of the lungs having ground-glass opacities and consolidations be quantified?
- Can DNN based methods learn efficiently and effectively from the highly sparse representations of COVID-19 diseased regions?

Most of these problems are addressed in the subsequent chapters of this dissertation.

1.4 What is the Significance of the Problem?

The problem of COVID-19 detection and disease quantification is highly significant and worthy of attention. As depicted in Figure 1.3, early patient identification helps in early treatment which leads to more positive patient outcomes, and patient isolation prevents the spread.

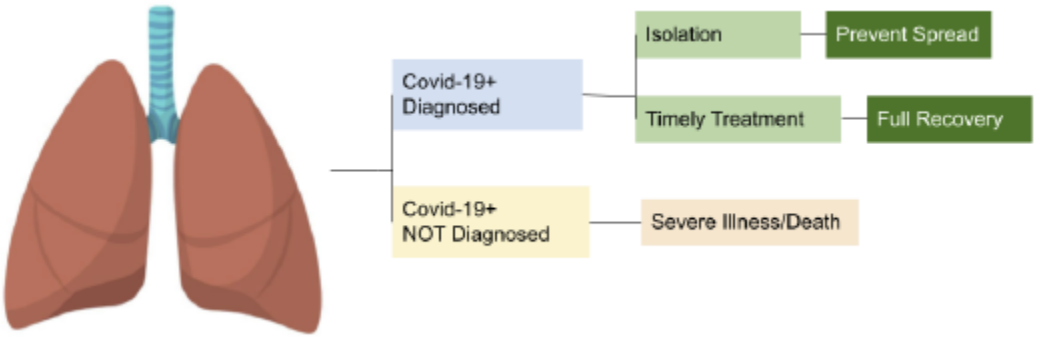

Figure 1.3: Impact of early COVID-19 diagnosis on patient outcomes and disease spread.

Segmentation of computerized tomography (CT) images to delineate the diseased or

lesioned areas helps quantify disease severity and assists physicians in predicting disease prognosis and to administer appropriate treatments to save lives [33].

1.5 The Solution Explored in this Dissertation

This dissertation proposes the use of supervised deep learning methods to solve the problem of covid-19 detection and disease quantification.

Convolutional neural networks are a well-established method used for disease diagnosis via medical imaging. This dissertation explores the various convolutional neural network models that can help diagnose COVID-19 from lung computerized tomography (CT). We also explore the enhancements to the U-Net-based deep learning methods to help assess the COVID-19 disease severity.

1.6 What are the Challenges?

For automated analysis of CT images, several methods have been published in [12], [16] which are based on convolutional neural networks (CNN) originally proposed by LeCun [21]. Roberts et al. [28] systematically screened all published papers and preprints from January 2020 to October 2020 on new machine learning models for the diagnosis or prognosis of COVID-19 from Chest X-Ray (CXR) or CT images. From 2,212 studies, 415 studies were included after initial screening, and a systematic review on 62 studies revealed major methodical deficiencies due to the high likelihood of duplicated images across different sources that result in so-called "Frankenstein datasets", underlying biases such as including samples from non-representative populations, and low-quality data [29].

Additionally, in most studies, the CT image data were split randomly into training and

testing such that different images from one patient's CT scan end up in both training and validation. This violates the independence and identically distributed (IID) assumption.

Furthermore, a high variation in infection characteristics, low-intensity contrast between infections and normal tissues, and an insufficient amount of data inhibiting the training of a deep model are other challenges in disease classification and segmentation from CT scans [4]. Accurate segmentation of computerized tomography (CT) volumes is a challenge due to complex structures, pathological changes, individual differences in infection characteristics, and low image quality [34]. Moreover, class imbalance and annotation errors also make the segmentation task more challenging [43].

Due to these challenges, it is unclear which study (if any) is of potential clinical utility. To solve these issues and make the research reproducible, higher-quality datasets, heavily documented research, and external validation are needed. Lie et al. [13] recommend that artificial intelligence-based approaches for COVID-19 could be expanded to embrace all sorts of respiratory illnesses, even new coronaviruses that may arise in the future. With Artificial Intelligence (AI), a standard CT scan or X-ray becomes a versatile tool to assist with a speedy diagnosis to contain disease spread. Even after the pandemic is over, such techniques can be expanded to diagnose and prognosticate all respiratory illnesses including new viruses when doing any chest radiograph or low-dose CT for lung screening.

1.7 The Novel Contributions of This Dissertation

- A novel approach of using a 3-dimensional convolutional neural network (CNN) on three-dimensional CT volumes for COVID-19 detection.

Accurate predictions of COVID-19 diagnosis in terms of various metrics including

Accuracy, Precision, Sensitivity, Specificity, F-1 score, and the Matthews correlation coefficient (MCC) in comparison to the recent studies.

- A novel approach of segmenting COVID-19 diseased regions using spatial and channel-attention on the pre-trained DenseNet169 backbone-based U-Net model on lung CT.

This model's dice score and Jaccard index of COVID-19 segmentation is statistically significantly higher than previously reported values in recent studies, as evaluated on three different public datasets via the held-out test set approach.

This novel approach of using the pre-trained DenseNet as the encoder backbone with the transfer learning strategy exhibited a superior performance on the segmentation task versus training a deep model from scratch.

This approach is also a faster and more efficient method than training a model from scratch.

1.8 Role of CT Chest in COVID-19 Infection

Chest CT has an important role in the diagnosis, detection of complications and prognostication of coronavirus disease 2019. Per Rubin et. al [81], CT Imaging is not routinely indicated as a screening test for COVID-19 in asymptomatic individuals, and it is not indicated for patients with mild features of COVID-19 unless they are at risk for disease progression. However, CT Imaging is indicated for patients with moderate to severe features of COVID-19 regardless of COVID-19 test results. It is also indicated for patients with COVID-19 and evidence of worsening respiratory status. Additionally, CT is indicated in recovered cases with functional impairment and/or hypoxaemia.

CT chest may be normal in up to 10.6% cases [60]. These were mostly patients in the first 4-5 days after symptom onset, however a small though non-negligible number of patients also showed normal CT in the later stages of infection.

Zhou et al. [93] reported the common CT findings of COVID-19 infection as listed below:

- Ground glass opacities
- Consolidation
- Reticular pattern
- Vacuolar sign
- Microvascular dilation sign
- Fibrotic streaks
- Subpleural line
- Pleural thickening

Adams et al. [60] reported the various CT findings as detailed below:

- The commonest findings (>70% incidence) on chest CT scan are ground glass opacities, vascular enlargement, bilateral involvement, preferential lower lobe involvement and peripheral, posterior, and basal predilection [60]. Kwee et. al [76] and Ng et al. [80] also confirmed that ground glass opacities and consolidation in the lung periphery is the imaging hallmark of COVID-19. In summary, a peripheral predominance of lung opacities is observed in lung CT of COVID-19 patients.

- Several chest CT findings have been reported with intermediate frequency, at 10-70% of Reverse Transcriptase Polymerase Chain Reaction (RT-PCR) proven cases. These findings include consolidation, linear opacities, septal thickening, reticulation (crazy paving pattern), air bronchogram, pleural thickening, halo sign, bronchial wall thickening and dilatation [60].

- Several chest CT abnormalities are seen less commonly (<10% of RT-PCR proven cases) and these include pleural effusion, lymphadenopathy, tree in bud sign, central lesion distribution, pericardial effusion, and cavitating lung lesions. These findings usually occur in combination with other more common findings and usually later in course of disease [60].

1.8.1 CT Findings for Severe COVID-19 Requiring Intensive Care Unit (ICU) Care

Saburi et al. [82] revealed that mild COVID-19 pneumonia mainly starts as small subpleural, unilateral or bilateral GGO in the lower lobes as shown in Figure 1.4. These lesions develop into subsequent consolidation and a crazy-paving pattern. The residual GGO and subpleural parenchymal bands appear gradually after two weeks, indicating a decrease in the severity of the disease. What are the findings in computerized tomography predictive of severe disease which will require intensive care transfer? Cases with the highest lung severity score at high resolution computerized tomography (HRCT) were admitted to the intensive care unit. Patients in the emergency group are more likely to have the following CT findings: architectural distortion, traction bronchiectasis, and higher CT involvement score. Additionally, bilateral patchy consolidation and interstitial abnormalities on CT and CXR are reported to occur twice as frequently in severe patients than in non-severe patients.

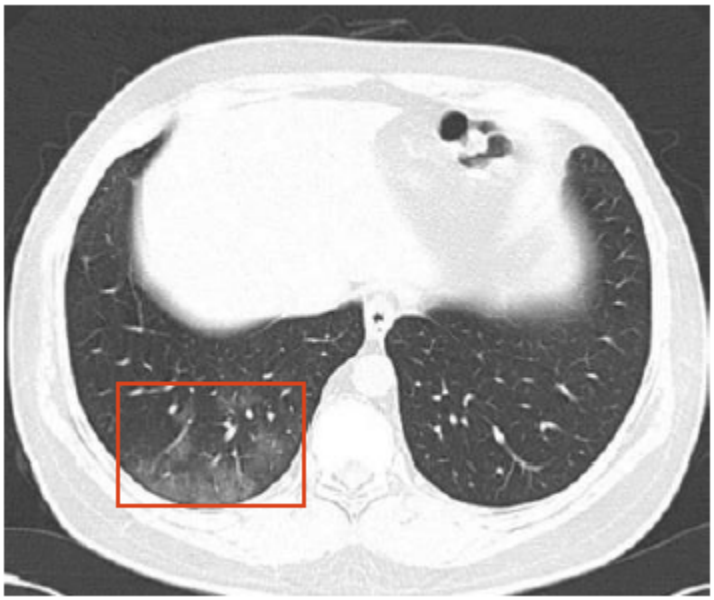

Figure 1.4: CT image of a COVID-19 positive patient with the COVID-19 disease-affected region shown as the red bounding box. Image sourced from Saburi et al. [82].

1.8.2 Temporal Evolution of Disease

Knowledge of the natural temporal evolution of lung abnormalities in COVID-19 may be helpful in determining the stage of disease and in distinguishing them from potential complications [47],[52].

1. Early Stage: (0-5 days) Normal or mainly ground glass opacities.
2. Progressive Stage (5-8 days) Increased ground glass opacity and crazy paving appearance.
3. Peak Stage (9-13 days) Characterized by progressive consolidation
4. Late Stage (>= 14 days) Gradual decrease of consolidation and ground glass opacities. Appearance of signs of fibrosis i.e., parenchymal bands, traction bronchiectasis.

1.8.3 CT Chest Severity Score

Chest CT severity score is designed to quantitate the extent of lung involvement and assess the severity of COVID-19. This could expedite the identification and management of patients with moderate and severe disease. A scoring method of 0-40 is described in which both lungs are divided into 20 regions [61]. Each region is given a score of 0-2 based on the percentage of lung involvement:

- Score 0: 0% involvement
- Score 1: < 50% involvement
- Score 2: > 50% involvement

It should be noted that the CT findings of COVID-19 infection may have overlap with other viral pneumonias. It is not always the case that first a patient develops GGOs and later develops consolidation or other imaging features. In fact, for elderly or pediatric cases, there are some atypical manifestations [59].

Note that the axial lung CT images could show either an open lung or the closed lung as shown in Figure 1.5, depending on which part of the lung is being scanned in that axial slice.

We can see that the closed lung image only shows a small portion of the lungs. The open lung images show a larger proportion of the lung region. While some published works discarded the closed lung images from their dataset to supposedly make the model train better, an expert radiologist advised us not to do that. This is because COVID-19 disease can occur in any specific part of the lung alone including the upper or lower regions of the lungs only.

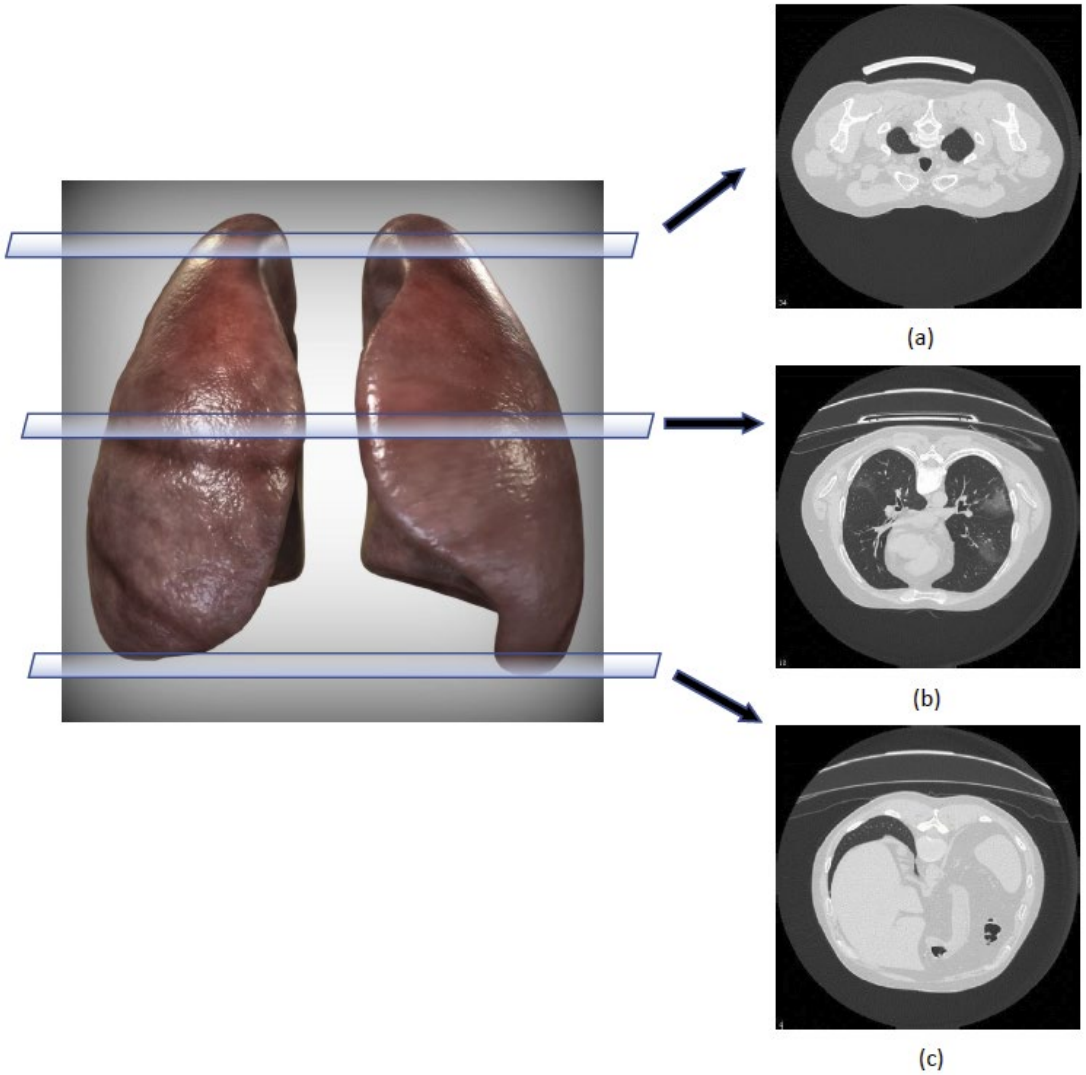

Figure 1.5: (a) Closed Lung for an axial slice from the top part of the lung, (b) Open Lung for an axial slice from the middle part of the lung, (c) Closed Lung for an axial slice from the bottom part of the lung.

The lung CT image samples are shown in Figure 1.6, with both COVID-19 positive and COVID-19 negative cases.

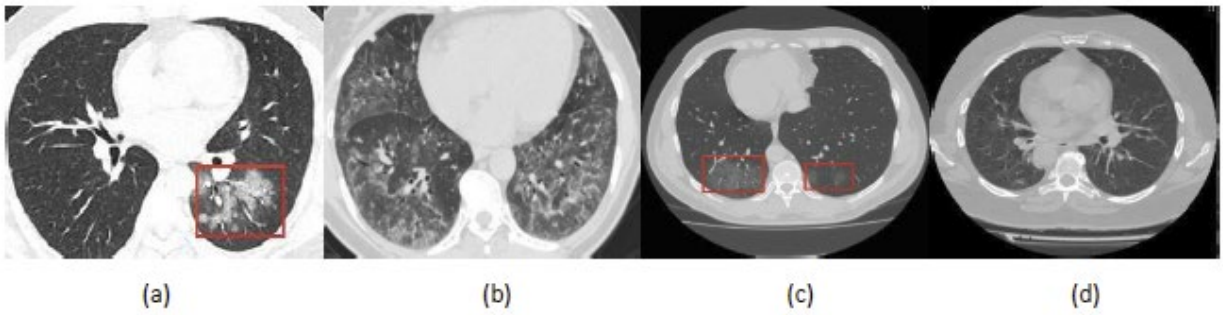

Figure 1.6: Examples of chest CT scans of patients from UCSD-AI4H and MosMedData; the red bounding boxes (added by authors) enclose the COVID-19 diseased regions. (a) the COVID-19 positive class from UCSD-AI4H, (b) the COVID-19 negative class from UCSD-AI4H, (c) the COVID-19 positive class from MosMedData, (d) the COVID-19 negative class from MosMedData.

Figure 1.7 shows the COVID-19 CT images with radiologist-provided ground truth labels for segmentation of ground-glass opacities and consolidations.

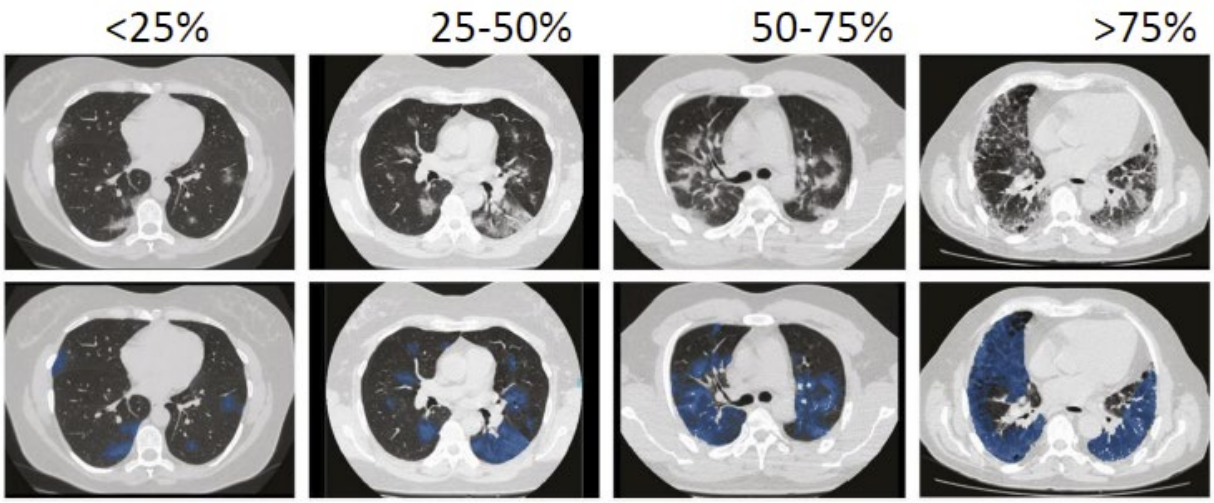

Figure 1.7: Examples of chest CT scans of patients with varying degrees of COVID-19 severity from the MosMedData. Left to right: mild to critical severity based on percentage of lung affected. Lower row shows the physician annotations added in blue color. [139]

1.9 Organization of this Dissertation

The remaining part of this dissertation is organized as follows: Chapter 2 is a discussion

on the related works in COVID-19 detection, highlighting the various methods that have been used in existing works. Chapter 3 explores the effectiveness of various deep learning models for COVID-19 detection, while also comparing them based on different performance metrics. Chapter 4 presents the related work on COVID-19 lesion segmentation. In Chapter 5, we present the COVID-19 lesion segmentation models whose performance metrics are compared with the state-of-the-art. Next, in Chapter 6, we present the ablation studies for COVID-19 detection and segmentation. Finally, in Chapter 7, we discuss the conclusions of this research and provide directions for future research.

CHAPTER 2

RELATED WORK: COVID-19 CLASSIFICATION*

2.1 Overview

Shoeibi et al. [86] reviewed studies on the application of deep learning (DL) techniques for COVID-19 diagnosis and automated segmentation of lungs using X-ray and CT images. Deep learning and traditional machine learning (ML) have been used to diagnose the COVID-19 accurately using many public databases. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), autoencoders (AEs), deep belief networks (DBNs), generative adversarial networks (GANs), and hybrid networks such as CNN-RNN and CNN-AE have been used for automated detection of COVID-19 as shown in Figure 2.1.

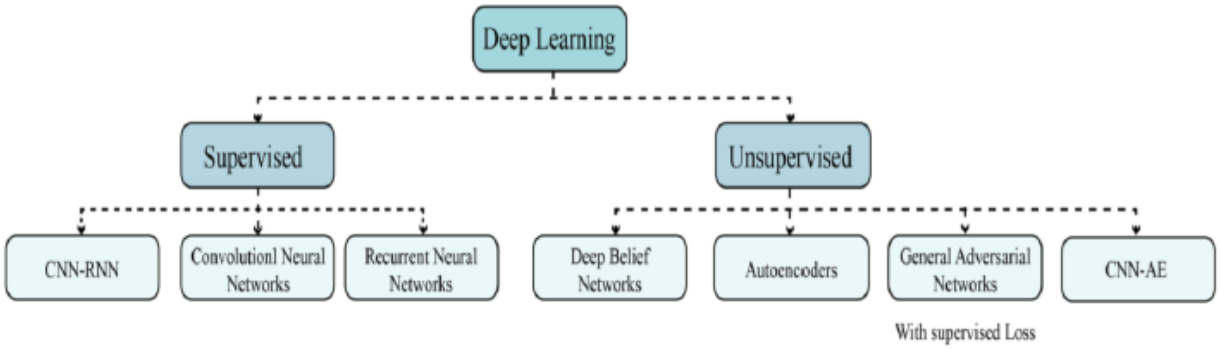

Figure 2.1: The various deep learning methods used for COVID-19 diagnosis.

Applying machine learning for segmentation is highly important to save radiologists' time. Fuzzy clustering methods [128], [129] and DL procedures such as Ronneberger et al.'s U-Net [105] are important.

* This entire chapter is reproduced from Mittal, B. and Oh, J. 2021. CoviNet: Covid-19 diagnosis using machine learning analyses for computerized tomography images. SPIE Proceedings Vol. 11878: Thirteenth International Conference on Digital Image Processing (ICDIP 2021), with permission from the Society of Photo-Optical Instrumentation Engineers (SPIE).

Various DL methods used for the automated detection of COVID-19 patients using X-ray and CT images are two-dimensional (2D) CNN, AlexNet, Visual Geometry Group (VGG) network, GoogLeNet, DenseNet, XceptionNet, MobileNet, SqueezeNet, Inception-ResNet, CapsNet, NasNetmobile, ShuffleNet, EfficientNet, and Generative Adversarial Networks (GAN). A generalized end-to-end framework for COVID-19 diagnosis is depicted in Figure 2.2.

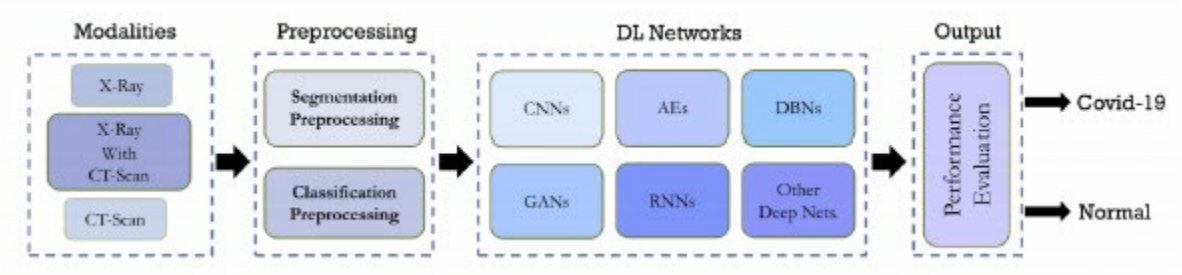

Figure 2.2: A generalized end-to-end framework for the COVID-19 classification.

A typical 2D CNN architecture is shown in Figure 10. CNNs comprise convolutional, pooling, and fully connected layers. The convolutional layers are usually followed by the pooling layers, and their output is fed to the fully connected layers as shown in Figure 2.3. Also, a variety of methods like dropout and batch normalization, originally proposed by LeCun et al. [21] help these networks learn better.

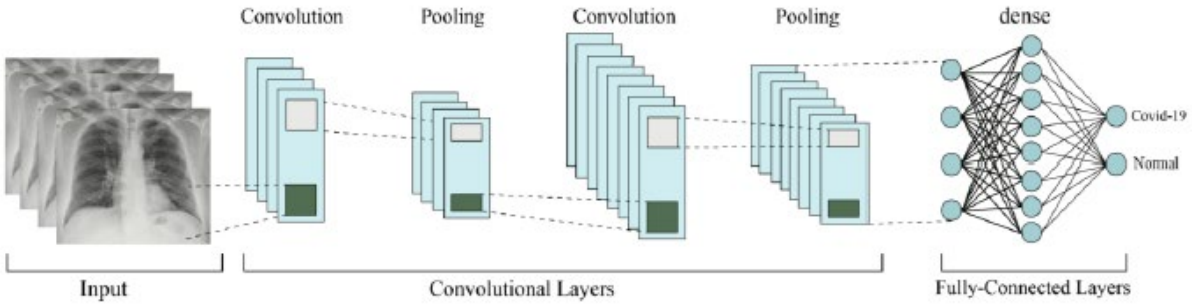

Figure 2.3: A typical 2D CNN architecture for COVID-19 classification

Most works on COVID-19 classification are based on chest CT or Chest X-Ray (CXR) images. We summarize the research based on chest CT and CXR in this section, which are relevant to our

COVID-19 classification work.

2.2 Related Work Review

Ko et al. [12] implemented a fast-track COVID-19 classification network (FCONet) that uses transfer learning on four deep models namely, VGG16, ResNet-50, Inception-v3, and Xception. In each of the four models, the ImageNet pre-trained convolutional base was followed by newly added classification layers comprising a flattening layer, two fully connected layers and a softmax layer. Their classifier has three classes: COVID-19, other pneumonia, and non-pneumonia. For the Xception and InceptionV3, the F1-score was 0.8806 and 0.5247 respectively, and the MCC was 0.8491 and 0.5281, respectively.

Similarly, Shah et al. [16] proposed CTnet10 based on CNN, which was compared with DenseNet-169, VGG-16, ResNet-50, InceptionV3, and VGG-19. Their CNN was trained from scratch, while the other deep models were ImageNet pre-trained models for which transfer learning was done using the COVID-19 dataset.

Goncharov et al. [27] proposed a modified U-Net model for the classification. Their modification to U-Net was that the classification layers were added to the most high-resolution upper part of U-Net rather than the bottom.

Al-Karawi et al. [5] used Generative Adversarial Network (GAN) with four different deep learning models for the data augmentation and reported accuracy improvement by several percentage points based on the InceptionV3 pre-trained model.

Ardakani et al. [6] employed an ensemble (Covidiag) classifier comprising five classifiers namely decision tree, K-nearest neighbor, naive Bayes, support vector machine, and ensemble on 20 radiological features from CT images.

Bridge et al. [7] proposed a generalized extreme value (GEV) distribution activation function in an Inception-v3 CNN model. The GEV helps to improve the performance over the traditional sigmoid activation function when one class significantly outweighs the other.

Fang et al. [8] diagnosed COVID-19 from radiomics features or hand-engineered features from medical images to diagnose COVID-19. The three selected radiomic features were fed to the support vector machine for the final classification.

Hassantabar [9] employed a 3-layer multilayer perceptron (MLP) on the statistical features of images. They also employed a CNN to learn from lung images directly, and then performed a comparison of the CNN versus the MLP.

Horry et al. [10] proposed a bidirectional long short-term memory network with mixture density network (DBM) for COVID-19 classification from chest CT images. The Memetic Adaptive Differential Evolution (MADE) algorithm was used for hyper-parameter tuning.

Hu et al. [11] proposed a weakly supervised deep learning to minimize manually labeled CT images needed while still accurately distinguishing COVID-19 from non-COVID-19 cases. A multi-scale learning scheme involves spatially aggregating via Global Max Pooling (GMP) of the feature maps from different convolutional layers. Integrated Gradients feature attribution method, joint saliency maps helped to extract the bounding box to predict the location of the lesions.

Liu et al. [13] aimed to distinguish COVID-19 from general pneumonia (GP) based on 34 statistical texture features including the gray-level-gradient co-occurrence matrix (GLGCM), ReliefF feature selection algorithm, and an ensemble classifier. First, 34 statistical texture features were extracted, including 13 gray-level co-occurrence matrix (GLCM) features, 15 gray-

level-gradient co-occurrence matrix (GLGCM) features, and 6 histogram features. Second, ReliefF algorithm selected those features whose average weights exceed an empirically set threshold T. Finally, the selected features were fed to an ensemble of the bagged tree (EBT) and four other machine learning classifiers including support vector machine (SVM), logistic regression (LR), decision tree (DT), and K-nearest neighbor with Minkowski distance equal weight (KNN).

Perumal [14] applied a transfer learning technique to detect COVID-19 from the pre-trained model on viral pneumonia trained on the NIH chest x-ray dataset [104] of 30,805 patients. It was reported that this model was superior to out-of-domain ImageNet pre-trained models. Further, Haralick texture-based features only focus on the area of interest to detect COVID-19.

Purohit et al. [15] presented a multi-image augmentation to increase the dataset size for training a CNN-based model for detecting COVID-19 in chest X-Ray and chest CT scan images. The multi-image augmentation obtains discontinuity information from the filtered images to effectively result in a larger training set to bolster model performance.

Salehi's [36] survey paper provided a high-level overview on epidemiology and pathogenesis of COVID-19, disease symptoms, reports recently published deep learning-based models like Convolutional Neural Networks' performance on lung imaging data.

Silva et al. [17] proposed an efficient deep learning technique with a voting-based approach to classify COVID-19 from CT images. Data augmentation was done using rotation, horizontal flip, and scaling, and transfer learning techniques were also employed. The 5-fold cross-validation F1-score is 0.8619. The cross-dataset analysis shows poor generalization since accuracy drops from 87.68% to 56.16% on the best evaluation scenario. Thus, COVID-19

detection through CT-scans must improve exponentially and be validated on more diverse datasets to evaluate the methods in a realistic scenario.

Song et al. [18] proposed a large-scale bi-directional generative adversarial network (BigBiGAN) architecture was used to extract the features. The extracted semantic features from the CT images were then used for classifying using three classifiers, namely, linear, support vector machine (SVM), and k-nearest neighbor (KNN).

Sun et al. [19] proposed an Adaptive Feature Selection guided Deep Forest (AFS-DF) algorithm with the Logistic regression classifier. The AFS-DF helped reduce the redundancy of features to help the model learn better from relatively small-scale data. Feature selection helps make features less redundant, and final classification is done via three methods: Logistic Regression (AFSDF-LR), Support Vector Machine (AFSDF-SVM), and Random Forest (AFSDF-RF) of which the AFSDF-LR achieved the best F1-score.

Xu et al. [20] distinguished influenza-A viral pneumonia (IAVP), COVID-19, and healthy cases in CT images using deep learning. Segmentation of infected candidate regions was done first using a 3D deep learning model. These segmented images were then fed to a location-attention classification model for the final classification prediction along with a confidence score. Finally, the Noisy-OR Bayesian function was used to categorize each CT case along with a confidence score.

Afshar et al. [62] proposed a COVID-CAPS framework based on Capsule Networks to detect COVID-19 from X-ray images. Advantage of COVID-CAPS over CNN-based models is that it is less complex, has a much lower number of trainable parameters in comparison to the more prevalent CNN models. On small COVID-19 data sets, this model shows a better performance

than Deep Neural Networks. COVID-CAPS comprises four convolutional layers and three capsule layers, which use a fast iterative Expectation Maximization procedure for "routing-by-agreement" that updates the probability with which a part is assigned as COVID-19 positive based on the proximity of the vote coming from that part to the votes coming from other parts that are assigned as COVID-19 positive. COVID-CAPS was pre-trained on the public NIH Chest X-ray dataset [104] for common thorax diseases. Thereafter, transfer learning is done to fine tune the capsule layers only. Since their dataset is highly imbalanced, the loss function was modified to assign a higher penalty to the false negatives. Pre-training was also done on another dataset of X-ray images and transfer learning was applied on COVID-19 X-ray dataset.

Albahli et al. [63] detected COVID-19 by using chest X-ray data using ECOVNet, an assortment of deep convolutional neural networks (CNN) based on EfficientNet. Data augmentation is done on the chest x-ray data prior to CNN input. Then, ImageNet pre-trained weights for EfficientNet are transferred with some customized training and fine-tuning on top layers, followed by an ensemble of CNNs.

Burgos-Artizzu [64] developed an Inception CNN-based COVID-19 classifier from publicly available CXR images downloaded from seven different online data sources. This is a pre-print and has not been peer-reviewed.

Hu et al. [67] designed a tool for diagnosing COVID-19, utilizing sixteen data augmentation operations, and performance was evaluated without noisy labels and with noisy labels. This is a pre-print and has not been peer-reviewed.

Huang et al. [66] present a quantitative analysis based on an ambispective observational cohort study of 125 patients with COVID-19 in Xiangyang, China.

In [68], Ilhan uses convolutional neural networks (CNNs) and ensemble learning with feature level fusion fed to multiple classifiers with majority voting scheme for classifying chest x-ray images into COVID-19, pneumonia, and no-finding classes. The classifiers used were VGG16, ResNet50, ResNet101, NasNet, Inception V3, and Xception. Five-fold cross validation accuracy values were 87.6%, 85.7% and 85.7% for InceptionV3, ResNet50 and VGG16, respectively. This paper is a pre-print and has not been peer-reviewed.

In [69], Jin et al. deployed an AI system for automatic analysis of CT images to detect COVID-19 pneumonia features.

Khan et al. [70] developed the CoroNet, a Deep Convolutional Neural Network (Xception architecture) pre-trained on ImageNet dataset is re-trained end-to-end on two different public databases on COVID-19 and other chest pneumonia X-ray images. The F-score is 89.8% for 4-classes of COVID-19 vs Pneumonia bacterial vs pneumonia viral vs normal and is 91% for 3-classes of COVID-19 vs Pneumonia bacterial vs normal.

Li et al. [71] developed COVNet which extracts visual features from volumetric chest CT scans for the detection of COVID-19 positive patients from negative patients. The dataset was collected from six hospitals with 4,352 chest CT scans from 3,322 patients and the non-Covid classes included community-acquired pneumonia (CAP) and other non-pneumonia abnormalities to ensure model robustness. On an independent test set, the sensitivity and specificity for detecting COVID-19 was 90% and 96%, and for CAP was 87% and 92%.

Mohammed et al. [72] proposed ResNext+, a weakly-supervised COVID-19 detection approach which provides slice level predictions from only volume level data labels. A lung segmentation mask was used for pre-processing and spatial and channel attention were to

extract spatial features. The Long Short-Term Memory (LSTM) helps capture the sequential dependency of the slices. A slice attention module precedes the final dense layer. At slice level, the proposed method has an F1 score of 81.4%, which is lower than the previous state-of-the-art with an F1 score 83.4%.

Chen et al. [75] propose the aggregated residual transformation to build a robust and expressive feature representation with the soft attention mechanism to improve the model's performance in segmenting the COVID-19 disease in chest CT.

Rahaman et al. [73] employs 15 different ImageNet pre-trained CNN models (VGG series, Xception, ResNetV1 series, ResNetV2 series, Inception series, DenseNet series, and MobileNet networks) with transfer learning due to limited number of CXR images in COVID-19 dataset. VGG19 obtains the highest classification F1 score of 0.90.

Somasekar et al. [74] reviewed the research done in machine learning, image analysis applications, datasets available, and challenges in the fight against the COVID-19 pandemic. They suggested three areas of research namely: Deep convolutional neural networks with transfer learning to assist in COVID-19 diagnosis from Chest X-Ray (CXR) images, disease prognosis based on patient characteristics, comorbidities, initial symptoms, vital signs to identify high risk patients; epidemiological studies using deep neural networks.

Ter-Sarkisov [107] presents COVID-CT-Mask-Net model for COVID-19 prediction from CT scans. First, it detects the ground glass opacities and consolidations in CT scans. Second, the ranked regional predictions (bounding boxes with scores) in Mask R-CNN are used to make accurate predictions of the image class. The goal is to predict the COVID, common pneumonia and control classes for a dataset of CT scans from China National Center for Bioinformation.

There is an opportunity to combine texture descriptors such as local binary patterns, edge detection histogram and local density features with deep learning features to improve model performance [13], [14]. Also, since noise negatively impacts model performance, denoising methods in pre-processing should help.

2.3 3D Deep Learning

In recent years, three-dimensional (3D) deep learning models have become prevalent in the broader field of computer vision and in medical imaging. The promise of these 3D learning techniques comes from utilizing and learning the spatial and temporal information inherent in the datasets.

3D learning increases the computational complexity exponentially because of the increased number of mathematical matrix operations that are performed on the numpy arrays [134] for each convolutional and pooling and other layers of the network. It may not always be a fair comparison to compare the 3D model performance with those of 2D models, as the 3D volume data may not be available which is required for training a 3D model. The image pixel count of typical 3D medical image datasets is 17M/50k = 340 times more pixels than 2D images from ImageNet, and time performance would get multiplied cubically for a 3D kernel as shown in Figure 2.4.

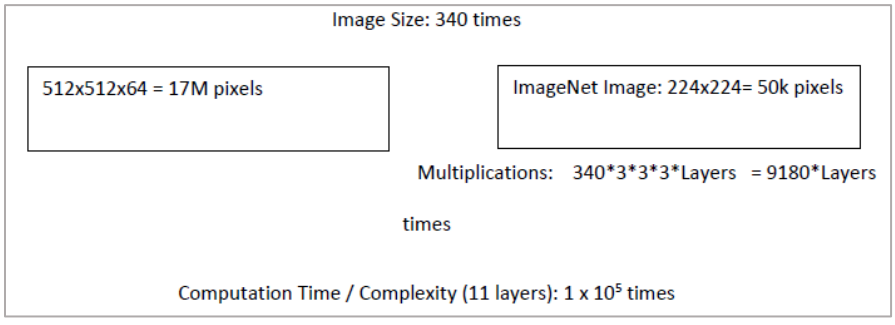

Figure 2.4: Image dimensionality of 2D versus 3D and Complexity

For training with a very small batch size of 2, our system memory utilization was reaching an average of 80% on the MosMedData 3D CT volumes. Additionally, the MosMedData is of higher resolution, but we needed to downsize the image from 512x512x64 to 224x224x64 which in effect leads to 40% to 75% information loss.

Thus, the 3D models would increase the computational complexity by several orders of magnitude. There is not a clear way to scale all the other vision algorithms to the 3D volumes in the medical domain. Each convolutional layer could take up to 16 GB of memory for a 512x512x512 cube for a 12-bit image where the pixel values range from -2048 to 2048. So, the memory of a large 128 GB machine would be completely utilized with just 8 convolutional layers prohibiting the training of a deeper model from scratch.

If the available dataset of 3D volumes is small, data augmentation can help increase the effective dataset size as shown in Figure 2.5. This larger dataset will enable the training of a deeper and more complex model which can exhibit superior performance and be more robust to noise.

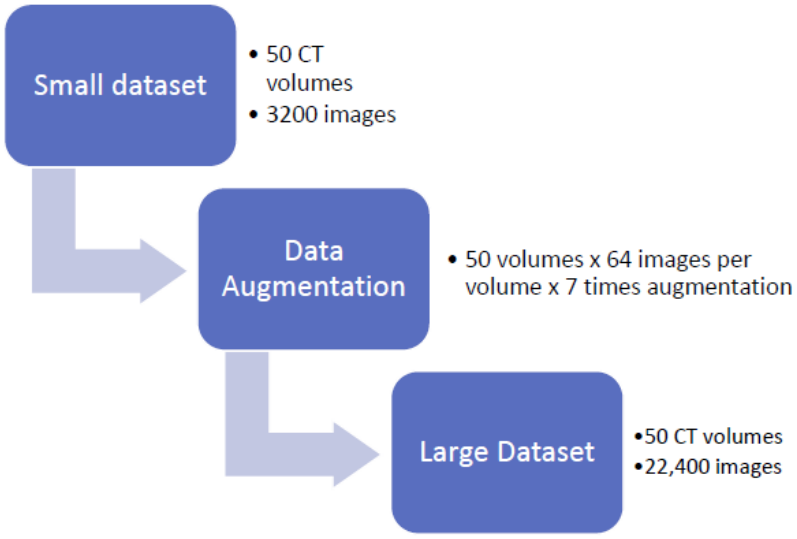

Figure 2.5: Why data augmentation

In addition to transformations (shifting, shearing, zooming etc.), generative adversarial networks have been used to augment the data by creation of synthetic images as shown in Figure 2.6.

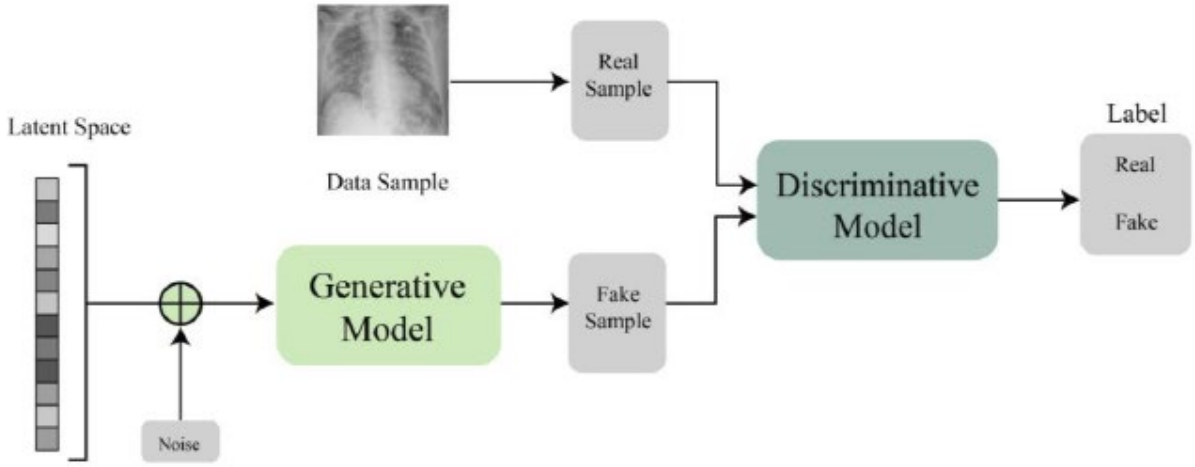

Figure 2.6: A simple GAN Architecture [86]

CHAPTER 3

METHODOLOGY AND EXPERIMENTS: COVID-19 CLASSIFICATION*

In this chapter, we describe the proposed CoviNet classifier and the various CNN-based deep learning models for which experimentation is performed.

3.1 Proposed CoviNet

In this subsection, we discuss our implementation of CoviNet, a deep 3D convolutional neural network trained from scratch to classify CT scans into one of the COVID-19 positive or COVID-19 negative classes. We were inspired by Pominova et al. [25] and Liu et al. [26] to come up with this novel approach which involves the use of 3D filters in convolutional layers to train a deep 3D CNN from scratch on the 3D CT scan volumes. This is different from the existing research for diagnosing COVID-19 which uses individual slice-level CT imaging data to come up with slice-level predictions. The 3×3×3 convolutional filter W moves over the feature map F of dimension k×k×k with a stride of 1, to perform repeated three-dimensional convolutions, which results in a new feature map H of dimension (k-2) ×(k-2)×(k-2). To keep the dimensions unchanged after the three-dimensional convolutions, we use the same padding for the outermost pixels.

The proposed CoviNet's architecture is shown in Figure 3.1. It has a network depth of 16 layers comprising four 3D convolutional layers, four 3D max-pooling layers, four 3D batch normalization layers, one global average 3D pooling layer, two fully connected dense layers, one dropout layer, and a final softmax layer. All four convolutional layers labeled in Figure 3.1 as

* Sections 3.1 to 3.4 and 3.6 to 3.8 are reproduced from Mittal, B. and Oh, J. 2021. CoviNet: Covid-19 diagnosis using machine learning analyses for computerized tomography images. SPIE Proceedings Vol. 11878: Thirteenth International Conference on Digital Image Processing (ICDIP 2021), with permission from the Society of Photo-Optical Instrumentation Engineers (SPIE).

Conv3D have a kernel size of 3×3×3 but use different numbers of kernels at 64, 128, and 256. The RELU activation function is used. The four 3D-max-pooling layers take a 2×2×2 sliding cube which subsamples the image length, width and depth dimensions and has a stride of 2. To speed up the training of CoviNet, 3D batch normalization layers are included after the 3D pooling layers and are labeled in Figure 3.1 as BN. Then, 3D global average pooling takes a 4D input of size length×width×depth×channels (=12×12×2×256) and outputs a one-dimensional output of size 256 channels. Next, the fully connected layer follows with a dimension of 512. After that, there is a dropout layer with a dropout factor of 0.3 which is introduced to make the model robust to noise. The final layer, as shown in Figure 3.1, is the softmax layer with a sigmoid activation which outputs the predicted probability of being COVID-19 positive. Adam optimizer is used with an initial learning rate of 0.001 with an exponential decay rate of 0.96 over 100,000 decay steps. The validation accuracy is maximized during training.

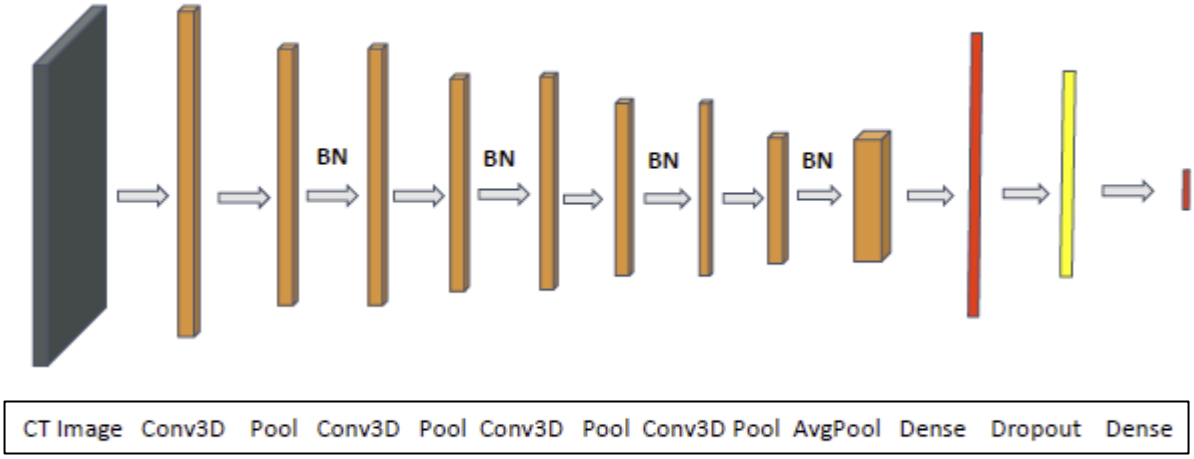

Figure 3.1: Architecture of Proposed CoviNet

Figure 3.2 shows the various steps involved in implementation of the CoviNet model. Table 3.1 shows the layers, dimensions and parameter counts of the proposed CoviNet model.

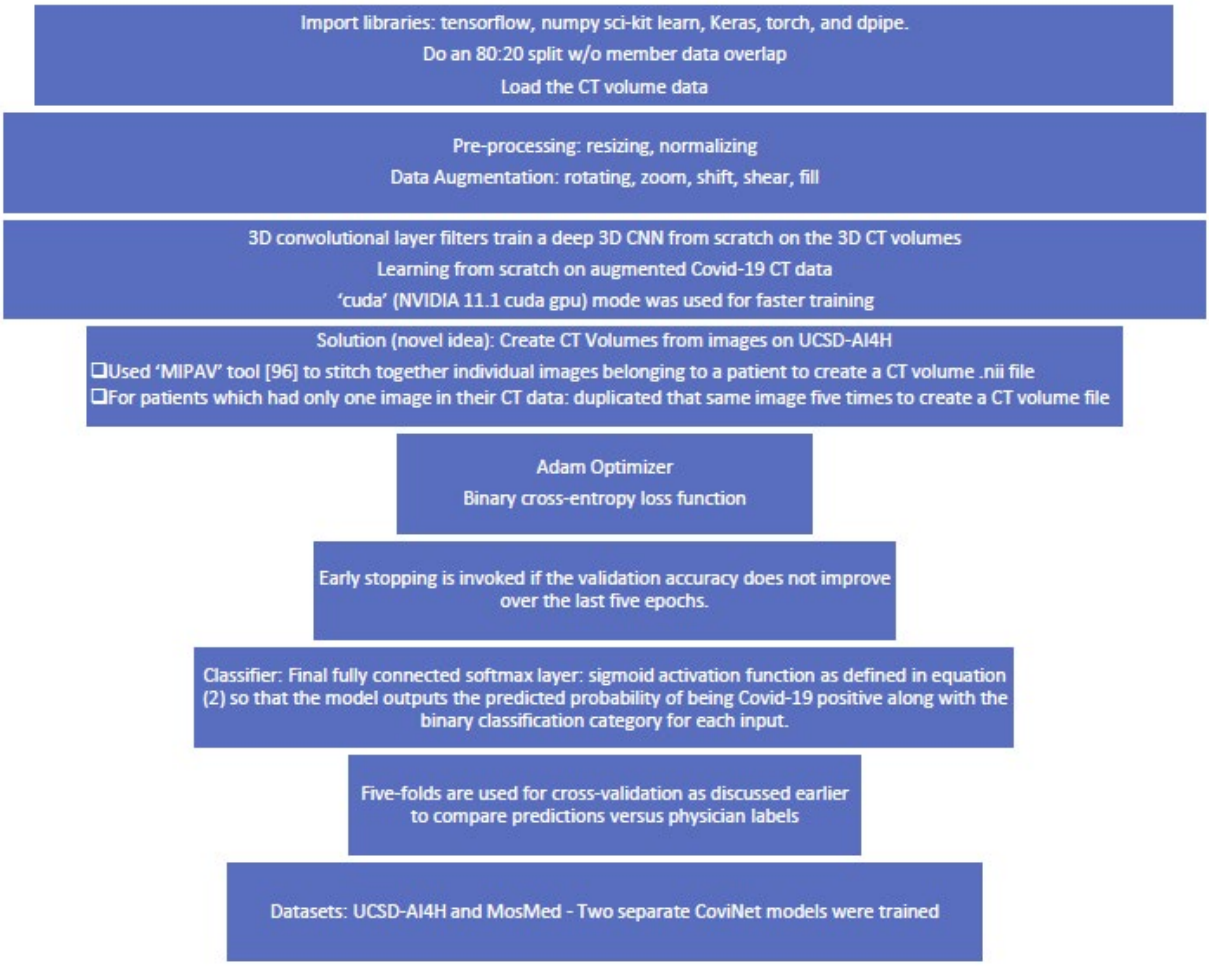

Figure 3.2: Steps involved in the implementation of the CoviNet model.

Table 3.1: Layers, dimensions and parameter counts of the CoviNet model.

Layer (type)	Output Shape	Param #
Input Layer	Count, 224, 224, 64, 1	0
3D conv	Count, 222, 222, 62, 64	1792
3D Max Pooling3d	Count, 111, 111, 31, 64	0
Batch Normalization	Count, 111, 111, 31, 64	256
3 D conv	Count, 109, 109, 29, 64	110656
3D Max Pooling3d	Count, 54, 54, 14, 64	0
Batch Normalization	Count, 54, 54, 14, 64	256
3D conv	Count, 52, 52, 12, 128	221312
3D Max Pooling	Count, 26, 26, 6, 128	0

Layer (type)	Output Shape	Param #
Batch Normalization	Count, 26, 26, 6, 128	512
3D conv	Count, 24, 24, 4, 256	884992
3D Max Pooling	Count, 12, 12, 2, 256	0
Batch Normalization	Count, 12, 12, 2, 256	1024
3D global average pooling	Count, 256	0
Dense	Count, 512	131584
Dropout	Count, 512	0
Dense	Count, 1	513
Total params: 1,352,897, Trainable params: 1,351,873, non-trainable params: 1,024		

3.2 Datasets

To evaluate the CoviNet's performance in comparison to recent studies, two open-source CT imaging datasets with physician-provided labels of COVID-19 positive or COVID-19 negative were used. These datasets are used in prior research studies and have a reasonable number and quality of CT images. The data was annotated by radiologists and classified into COVID-19 positive or negative classes. Figure 6 shown earlier showed the sample images for the COVID-19 positive and negative classes from both datasets. Note that the UCSD-AI4H has individual images in .png or .jpg format from various medical facilities with varying and low resolution. The MosMedData has one CT volume per patient with all CT slices for a patient in one .nii file format.

3.2.1 UCSD-AI4H Data [23]

The first dataset used is the UCSD-AI4H data [23] with 397 CT images of 278 COVID-19 negative patients and 349 CT images of 216 COVID-19 positive patients. The entire UCSD dataset was used in our experiments, and an 80:20 split was done for the train versus test set. The training set has 317 COVID-19 positive CT images and 279 COVID-19 negative CT image slices,

and the validation set has 70 COVID-19 positive CT images and 80 COVID-19 negative CT image slices. Of 384 patients, 229 only have one CT slice, 73 patients have two slices and only 82 patients have in between three to sixteen slices. Table 3.2 shows the patient and image counts for the UCSD-AI4H dataset.

Table 3.2: Patient and Image counts in the Dataset 1: UCSD-AI4H Dataset [23]

Class	Training	Validation
COVID-19 Positive	172 patients 279 images	41 patients 70 images
COVID-19 Negative	140 patients 317 images	31 patients 80 images

3.2.2 MosMedData [24]

The second dataset used is the MosMedData [24] with anonymized human lung CT scans from 1110 individuals with signs of COVID-19 (CT1-CT4) or without signs of COVID-19 (CT0). On the MosMedData, for this research, we performed random stratified sampling to select 6,642 normal CT image slices from 254 COVID-19 negative patients and 4,245 CT image slices from 172 COVID-19 positive patients. With an 80:20 split for training versus testing, the training set has 3,385 COVID-19 positive CT images and 5,305 COVID-19 negative CT image slices, and the test set has 860 COVID-19 positive CT images and 1,337 COVID-19 negative CT image slices. Table 3.3 shows the patient and image counts for the MosMedData.

Table 3.3: Patient and Image counts in Dataset 2: MosMedData [24]

Class	Training	Validation
COVID-19 Positive	138 patients 3,385 images	34 patients 860 images
COVID-19 Negative	203 patients 5,305 images	51 patients 1,337 images

*MosMed CT volumes have four dimensions as length × width × depth × channels (red, blue, and green) and have the COVID-19 positive or negative physician-provided label for each CT volume.

Almost all the CT volumes in MosMedData have multiple slices and most images have 30 or more slices with a maximum of 64 slices per patient. Figure 3.3 shows the first 40 slices from a CT volume from the MosMedData.

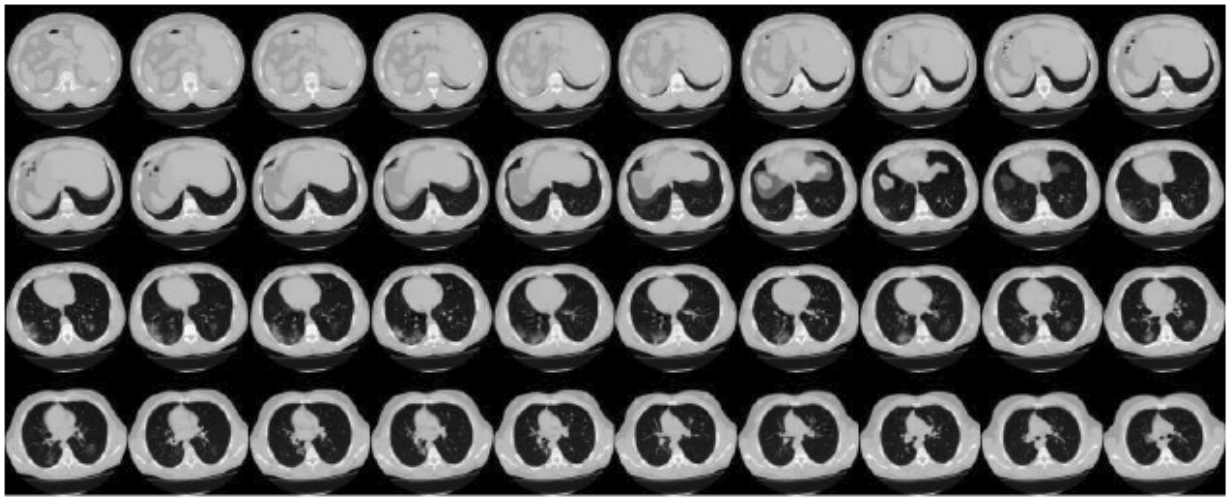

Figure 3.3: First 40 slices from a CT volume (MosMedData)

Table 3.4 shows a comparison of the UCSD-AI4H and the MosMed datasets based on imaging data format, type, and quantity.

Table 3.4: Comparison of UCSD and MosMedData included in our research.

UCSD-AI4H	MosMed
Individual .png images	CT volume data in .nii format
up to 16 CT slices / patient	1 CT volume with up to 64 slices / patient
Used entire dataset	Randomly sampled from this dataset
216 COVID-19 positive patients	172 COVID-19 positive patients
278 COVID-19 negative patients	254 COVID-19 negative patients
UCSD CT slices → .png or .jpg → length × width × channel (red, blue, and green)	MosMed CT volumes → .nii → length × width × depth × channel (red, blue, green)

3.3 Pre-Processing and Data Augmentation

3.3.1 Pre-Processing

The preprocessing was done on the entire dataset before feeding the data for classification. The images were normalized to have pixel values ranging from 0 to 255 to ensure that the images have adequate contrast as shown in Figure 3.4.

Figure 3.4: Normalization of CT Image

For the UCSD-AI4H dataset, images were resized to 224×224. For the MosMed dataset, pre-processing also included rotation of the volumes by 90 degrees as shown in Figure 19 because the data had the left lung and right lung arranged vertically. This rotation helped correct the orientation with the left and right lung placed horizontally side-by-side as shown in Figure 3.5. Then, the data was resized to get an image width, height, and depth of 224×224×64, and 128×128×64 for another experiment.

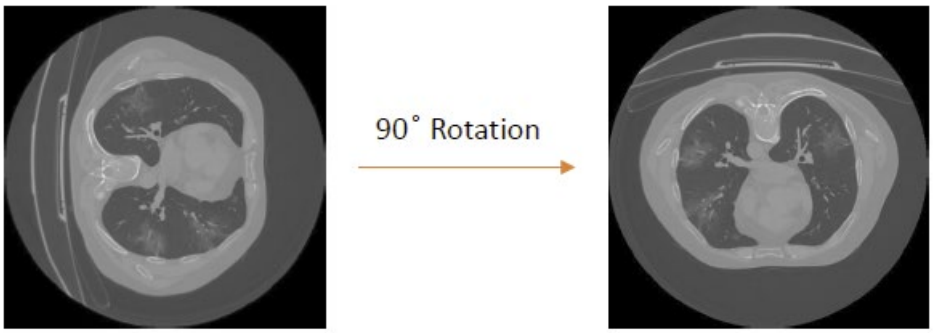

Figure 3.5: Rotation of the CT volume by 90 degrees on MosMedData [24]

3.3.2 Data Augmentation

Deep neural network approaches using convolutional neural networks exhibit great performance, but require a large, labeled training dataset. Data Augmentation effectively makes the dataset larger and more varied to reliably train a deeper model with a higher number of parameters. Using data augmentation, the model becomes more reliable and robust to noise and shows a superior performance as shown in Figure 3.6.

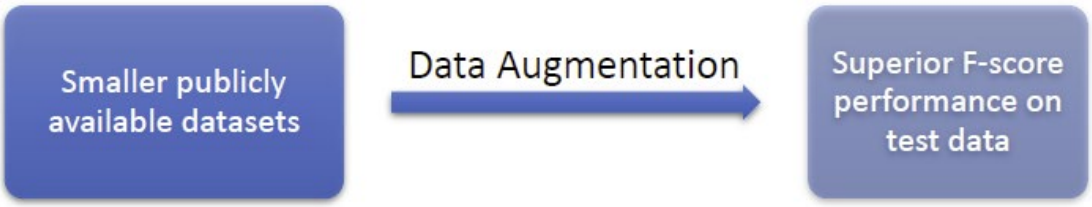

Figure 3.6: How data augmentation helps

With Data Augmentation for CoviNet, we increased the number of images by seven times, and increased the variety of images to enhance learning that is not overfitted to training. The 'Image Data Generator' library was used. The various augmentations applied were:

- Zoom factor of 0.2
- Rotation of +/- 15 degrees
- Horizontal shift of 0.1
- Vertical shift of 0.1
- Shear range of 0.2
- Fill mode of nearest

Then, image augmentation was done with a zoom factor of 0.2, a rotation of +/- 15 degrees, a horizontal shift of 0.1, a vertical shift of 0.1, a shear range of 0.2, and nearest fill mode. This image augmentation helped increase the number of images and a variety of images enhance learning and help the model generalize better and not be overfitted to the training dataset.

3.4 Implementation of CoviNet

CoviNet was implemented in 'jupyter-notebook' using python's 'tensorflow', 'keras', and other libraries [133] [134] [135] [137]. The pre-processing of the data was done using the 'nibabel' library for medical image processing which can read the CT volume data provided in .nii format [138]. Note that the datasets UCSD-AI4H and MosMed have different file types. For UCSD-AI4H, we have individual .png or .jpg image files as the input data with each image having the COVID-19 positive or negative label. For the MosMed data, the input CT volumes are .nii files, each of which is a sequence of image slices. The MosMed CT volumes have four dimensions as length × width × depth × channels (red, blue, and green) and have the COVID-19 positive or negative physician-provided label. This is the reason that the pre-processing for UCSD-AI4H, had an extra step to stitch together individual images belonging to a patient to create a CT volume .nii file using the 'MIPAV' tool [96]. For patients which had only one image in their CT data, we duplicated that same image five times to create a CT volume file, because CoviNet only intakes data in the form of CT volumes and not images.

Next, data augmentation was performed on the entire dataset using the 'scipy' and 'ndimage' libraries. Data augmentation included resizing, rotating by random angle measures from -20 to 20 degrees, and normalizing the image to show maximum contrast in lung tissue regions. The 'keras' library in 'tensorflow' was used to create the various custom layers including convolutional, activation, pooling, batch normalization, dense, dropout, and softmax layers. The dataset was split 80:20 for training and validation, and each patient's images are either entirely in the training dataset or entirely in the validation dataset. Five-folds are used for cross-validation as discussed earlier. Next, we trained the proposed CoviNet from scratch on the augmented data

which is fed in batches. The model learns the features over each successive batch by minimizing the binary cross-entropy loss between the training image labels and predictions during training/transfer learning which is defined in Eq. 3.1. The batch size selected is 30 images for all models, except CoviNet whose batch size is selected as 2 CT volumes. One full iteration over all batches covering the entire dataset makes for one training epoch, and the model trains over several epochs. Early stopping is invoked if the validation accuracy does not improve over the last five epochs.

$$\text{Binary Cross-Entropy Loss Function} = -\sum_{n=1}^{2} \log \frac{e^{y_{i,n}}}{\sum_{j=1}^{2} e^{y_{j,n}}} \quad \text{(Eq. 3.1)}$$

In each of the four 3D convolutional layers, we chose the activation function as RELU because it is computationally less expensive. For the final fully connected softmax layer, we chose a sigmoid activation function as defined in Eq. 3.2 so that the model outputs the predicted probability of being COVID-19 positive along with the binary classification category for each input.

$$\text{Sigmoid Activation Function} = \frac{e^{y_i}}{\sum_{j=1}^{2} e^{y_j}} \quad \text{(Eq. 3.2)}$$

The model compares the physician-provided ground truth labels against the predictions and the binary cross-entropy loss is minimized during training via backpropagation. Finally, the five-fold cross-validation is performed and the average of the cross-validation performance across the five folds is reported for both datasets. Two separate CoviNet models were trained and validated for the UCSD-AI4H and MosMed datasets, respectively. On the UCSD-AI4H dataset, CoviNet takes 10 minutes for pre-processing and data augmentation, and 25 minutes to train for each fold. Correspondingly, on the MosMed dataset, CoviNet takes 1 hour for pre-processing and

data augmentation, and 67 minutes to train for each fold.

3.5 CoviNet Enhanced: Our Improved Model for COVID-19 Classification

We first proposed the CoviNet, a 3D CNN for COVID-19 classification from lung CT in the initial draft of the thesis. After that, we improved the CoviNet using texture analysis in a hybrid approach using 3D CNN and texture features in an ensemble classifier approach named as CoviNet Enhanced.

3.5.1 CoviNet Enhanced Model

This ensemble classifier combines the outputs of the 3D deep convolutional neural network trained from scratch on the original CT images and Leung-Malik texture feature inputs-based Support Vector machine models.

This 3D CNN and texture analysis-based hybrid approach shows a superior performance than our original CoviNet. The ensemble of CNNs and SVMs help the model perform better because the CNNs and SVMs are complementary approaches.

We use a conditional majority voting approach for classification. For cases where the original 3D CNN predictions have probability between 46% and 54%, then the model uses the Leung Malik Texture-features' SVM model predictions along with 3D CNN via majority voting to make the final classification.

Our approach consists of five steps.

(1) 3D CNN

The 3D CNN having four convolutional layers is trained from scratch on the augmented CT data with early stopping is invoked if the validation accuracy does not improve over the last

five epochs.

(2) Feature Extraction

Leung-Malik texture features [130] are extracted. The filters used are shown in Figure 3.7.

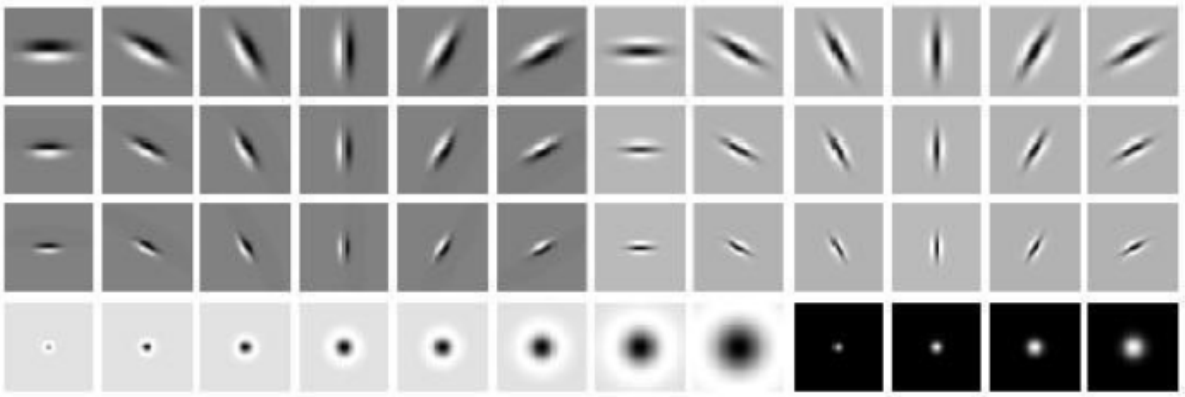

Figure 3.7: LM filter bank with 48 filters

For each image, the 48 Leung-Malik filters are convolved over the entire input image to extract 48 images, once for each Leung-Malik filter for each original CT image. The Leung-Malik features have textural, shape and intensity-based features. The first and second derivatives of Gaussian in different orientations and scales have nearly all black pixels since lungs do not have elongated objects which help predict COVID-19. The features which can identify the ground glass opacities and consolidations will be suited for our work. The 12 Laplacian of Gaussian filters show textural informational features, in that these could possibly be reliable features for COVID-19 classification. We show these 12 Laplacian of Gaussian features in Figure 3.8.

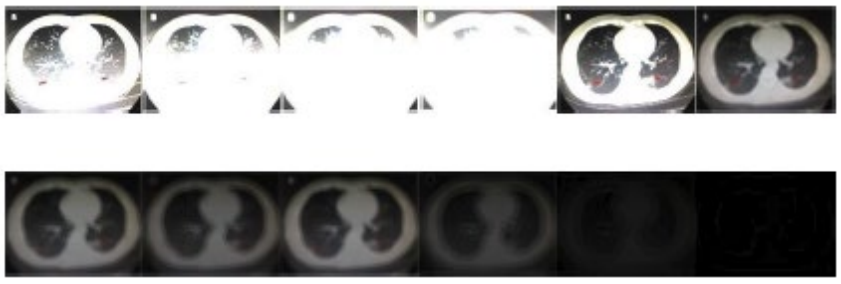

Figure 3.8: LM filter bank with the 12 Laplacian of Gaussian filters (LM 37 through LM48)

(3) Feature Selection via SVM models on Leung-Malik texture features

We create separate SVM models taking one feature at a time for all the 12 Laplacian of Gaussian Leung-Malik texture filters. Then, final feature selection is done to select the two top performing Leung-Malik texture features. The selected top-performing features are LM37 and LM41 and are shown in Figure 3.9.

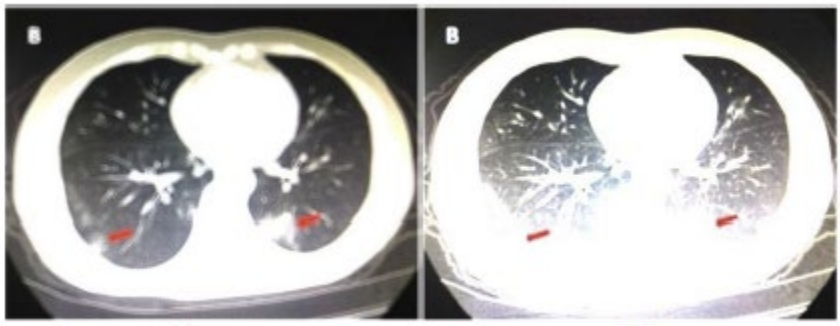

Figure 3.9: Leung-Malik features: (a) LM41, (b) LM37

(4) SVM Models

The selected Leung-Malik texture features, namely the LM37 and LM41, along with the original image are classified via three separate Support Vector Machine classifiers. We use the held-out test dataset approach [133]. The 80:20 split is used for training versus testing. We used the numpy, matplotlib and sklearn libraries to implement our model in python via the 'jupyter-notebook' application [134] [135] [136] [137].

First the image datasets and the texture feature inputs are resized to 224x224. The COVID-19 CT images in .jpg format with COVID-19 positive or negative physician-provided labels are fed into the first SVM model. The Leung-Malik feature LM41 is used to build the second SVM model. The Leung Malik feature LM37 is used to build the third SVM model. For the SVM, we perform a grid search for the parameter values of 'C' of 1, 10, 100, 1000, with a linear kernel for

efficiency, and with gamma values of 0.001 and 0.0001 with the Radial Basis Function kernel. We then get the predicted class for each image in the test dataset. This is repeated over the five folds.

(5) Ensemble of 3D CNNs and SVMs

Finally, we use conditional majority voting to combine the predicted outputs of the 3D CNN with the three support vector machines. This gives the final predicted classification of COVID-19 positive or Covid-10 negatives. The model performance is evaluated based on comparing this final prediction with the ground truth labels.

The detailed architecture of CoviNet Enhanced is shown in Figure 3.10.

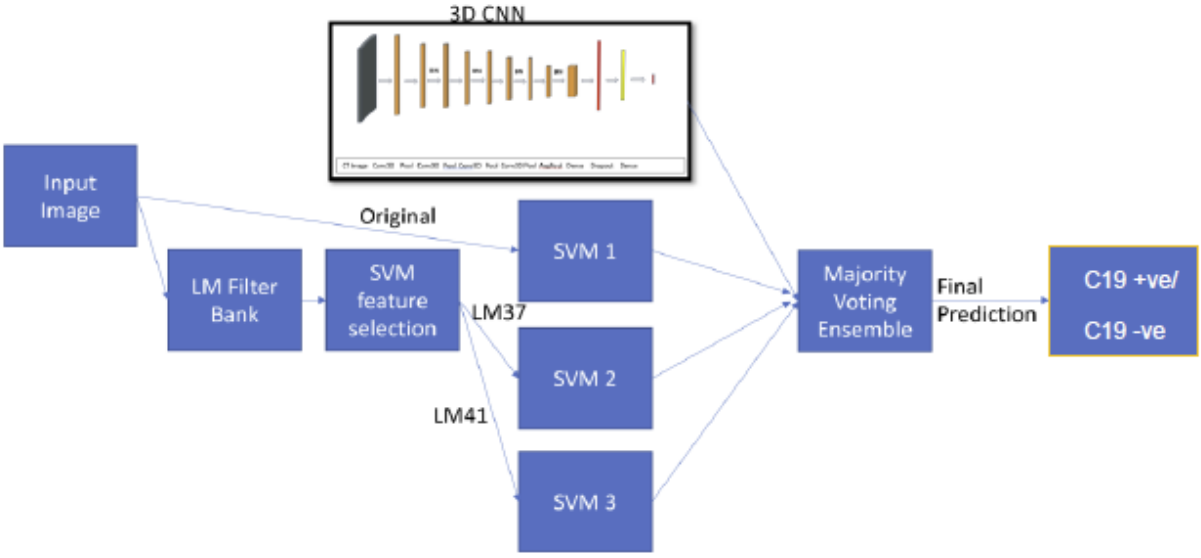

Figure 3.10: Ensemble classifier of 3D CNNs and Support Vector Machines based on texture features.

3.5.2 Implementation of CoviNet Enhanced

CoviNet Enhanced was implemented in 'jupyter-notebook' using python's 'tensorflow', 'keras', and other libraries [133] [134] [135] [136] [137]. The pre-processing of the data was done using the 'nibabel' library for medical image processing which can read the CT volume data provided in .nii format [138]. Data augmentation was performed on the entire dataset using the

'scipy' and 'ndimage' libraries. The dataset was split 80:20 for training and validation, and each patient's images are either entirely in the training dataset or entirely in the validation dataset. Five-folds are used for cross-validation as discussed earlier.

We trained and evaluated our model's performance on UCSD-AI4H and MosMed lung CT datasets described in Tables 3.5 and 3.6.

Table 3.5: Patient and Image counts in Dataset 1:UCSD-AI4H dataset [23]

Class	Training	Validation
COVID-19 Positive	172 patients 279 images	41 patients 70 images
COVID-19 Negative	140 patients 317 images	31 patients 80 images

Table 3.6: Patient and Image counts in Dataset 2: MosMedData [24]

Class	Training	Validation
COVID-19 Positive	138 patients 3,385 images	34 patients 860 images
COVID-19 Negative	203 patients 5,305 images	51 patients 1,337 images

3.5.3 Comparison between CoviNet and CoviNet Enhanced

The approach comparison between CoviNet and CoviNet Enhanced is shown in Table 3.7.

Table 3.7: Approach Comparison between CoviNet and CoviNet Enhanced

	CoviNet	CoviNet Enhanced
Approach	3D CNN	3D CNN + Texture Analysis
Classifier	3D CNN on original CT	Ensemble classifier comprising: 3D CNN on original CT SVM on original CT SVM on LM filter 37 SVM on LM filter 41 For cases where the original image CNN predictions have probability between 46% and 54%, then the model uses the Leung Malik Texture-features with majority voting to make the final classification.

The performance results comparison of CoviNet and CoviNet Enhanced are shown in Table 3.8.

Table 3.8: Performance Comparison between CoviNet and CoviNet Enhanced

Model/Data	Accuracy	Precision	Sensitivity	Specificity	F1-score	MCC
CoviNet/UCSD-AI4H	75.0%	68.7%	91.7%	58.3%	0.786	0.530
CoviNet/ MosMedData	94.1%	89.2%	97.1%	92.2%	0.930	0.882
CoviNet Enhanced/ UCSD-AI4H	92.2%	93.0%	93.0%	91.2%	0.930	0.842
CoviNet Enhanced/ MosMedData	96.4%	94.3%	97.1%	95.9%	0.957	0.926

3.6 Implementation of Others

We implement several ImageNet pre-trained classifiers such as DenseNet169, VGG16, ResNet-50, InceptionV3, and VGG19 which can be found in [12, 16].

Each of these models comprises two parts: a convolutional base and a classifier. The convolutional base has stacks of convolutional and pooling layers in varying configurations to generate deep features from the images. Whereas the classifier classifies the images based on the extracted features from the convolutional base. For our implementations for each of these models, we retained the convolutional base and removed the classifier part of the pre-trained ImageNet and replaced it with our classifier consisting of the dense layer with 1000 neurons, a dropout layer, another dense layer with 500 neurons, and finally a softmax layer which outputs the predicted probability of being COVID-19 positive or negative for each input image.

The first step is pre-processing and data augmentation on the entire dataset. We then performed an 80:20 split for training and cross-validation, and each patient's images are all either

solely in the training dataset or in the validation dataset as discussed earlier. Next, all the base convolutional ImageNet pre-trained weights were kept frozen and only the last four classification layer weights were kept unfrozen to facilitate transfer learning on the COVID-19 datasets. The models pre-trained on ImageNet were first loaded and transfer learning with COVID-19 datasets on the final four layers (two dense layers, one dropout, and one softmax layer) of the network is done.

Next, the classifiers output the predicted probability of the input being COVID-19 positive or negative. The physician-provided ground truth labels are compared against the predictions and the validation dataset loss is minimized during training. Finally, five-fold cross-validation and the average of the cross-validation performance across the five folds are reported for each of the 5 models for both datasets.

During transfer learning, the weights of the pre-trained ImageNet are kept frozen so that only the last four classification layers will be trained on the COVID-19 dataset. The RMSprop optimizer is used, and validation accuracy is maximized during training. This work will utilize the supervised learning framework with the radiologist-provided labels included in both open-source datasets: UCSD-AI4H and MosMed. The number of epochs in each model varied with the upper limit of 100 since early stopping was done when validation accuracy stopped improving over five successive epochs. The model parameters chosen were the batch size of 30, the learning rate of 0.001, a dropout factor of 0.3 in the dropout layer, and the binary cross-entropy loss function.

In this section, we assess the effectiveness of the proposed CoviNet model in classifying COVID-19 positive cases on two selected open-source datasets. All models were implemented in ‘python’ and ‘jupyter-notebook’ using various machine learning libraries including ‘keras’, ‘sci-kit

learn', 'nibabel' and 'matplotlib' [133] [134] [135] [136] [137] [138]. The machine used was Intel(R) Xeon(R) W-10885M CPU @ 2.40GHz, 2400 MHz, 8 Core(s), 16 Logical Processor(s), NVIDIA Quadro RTX 5000 with 128GB RAM.

3.6.1 DenseNet

DenseNet was first proposed in 2017 by Huang et al. [46]. It comprises dense blocks, followed by convolutional layers with batch normalization and RELU activation function. Figure 3.11 shows a 5-layer dense block having a growth rate of k=4 [46]. Each layer receives input from all the preceding layers' feature-maps. The DenseNet comprises densely connected CNN layers, in a dense block with the outputs of each layer connected with all descendant layers [86]. Due to the dense connectivity between the layers, it is termed as DenseNet. Network parameters are reduced dramatically by efficient utilization of feature reuse.

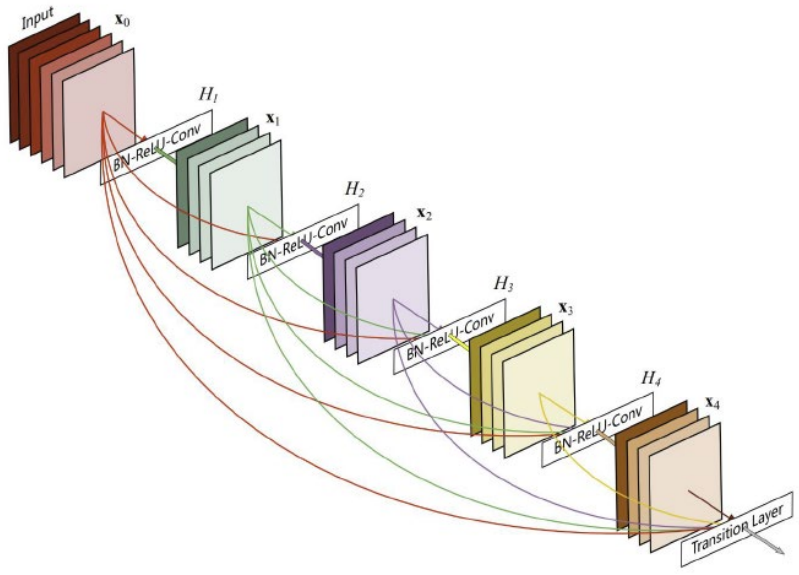

Figure 3.11: DenseNet169 architecture

3.6.2 VGG Models: VGG16 and VGG19

The Visual Geometry Group (VGG) architecture [86] comprises a few convolutional layers,

each of which utilizes the ReLU activation function. For classification, this network uses a softmax classifier in the final layer of the model. Filter size for convolutional layers is picked equal to 3x3, with a stride of 2 in VGG-E. VGG-11, VGG-16, and VGG-19 are three variants of the VGG-E model that have 11, 16, and 19 layers correspondingly. All variants of VGG architecture end with three FC layers. Nevertheless, the numbers of convolution layers are different; VGG-19 contains 16 convolution layers, VGG-16 has 13 convolution layers, and VGG-11 has eight convolution layers. Figure 3.12 depicts the building block of the VGG network used for COVID-19 detection [30], [65].

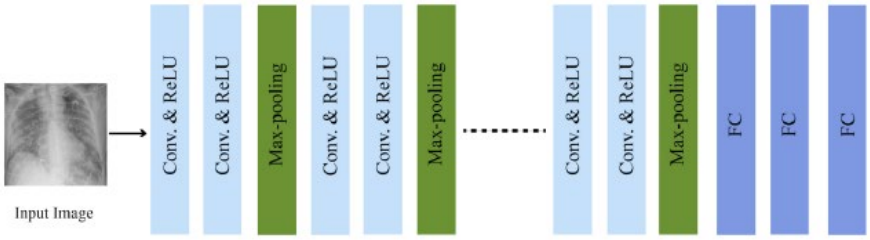

Figure 3.12: A typical VGG architecture used for COVID-19 detection.

3.6.3 Inception V3

InceptionV3, which integrates the inception architecture with residual connections, was first introduced by Szegedy et al. [106]. These residual connections significantly improve training speed. The filter concatenation stage in Inception architecture is replaced by residual connections [85] and this is called the Inception-ResNet and has multiple versions. Figure 3.13 shows the architecture of the inception layer. This is an efficient deep neural network architecture for computer vision. It was inspired by Arora et al.'s [112] theoretical work and Lin et al.'s [113]. It performed well on the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14). The architecture decisions were based on the intuition of multi-scale processing and the Hebbian principle that two neurons will have stronger weights when they activate together.

One specific implementation of the InceptionV3 is the GoogLeNet, a 22 layers deep network. The dense building blocks help this become a sparser architecture with increased depth and width of the network without much increase in the computational complexity.

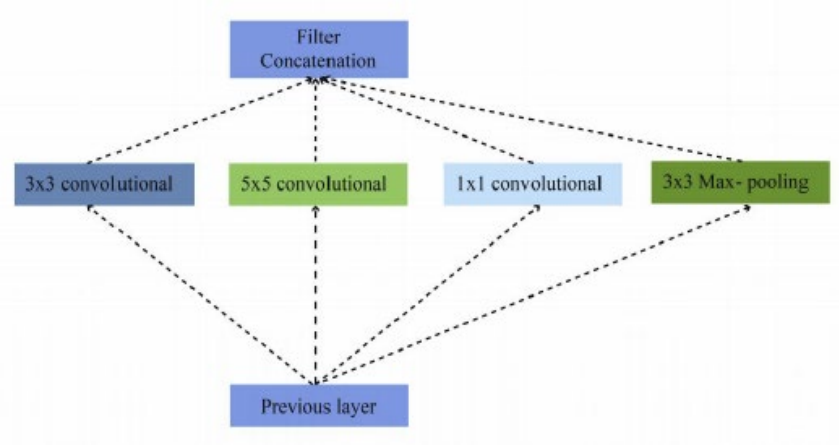

Figure 3.13: The Inception layer

3.6.4 ResNet

The Residual Network (ResNet) [86] is created with various numbers of layers; 1202,152, 101, 50, and 34. ResNet50 is one of the popular variants containing 49 convolution layers and 1 FC layer at the end of it. The total number of MACs and weights are 3.9M and 25.5M, respectively [30], [84], [85]. Figure 3.14 shows a typical ResNet architecture used for COVID-19 detection [97].

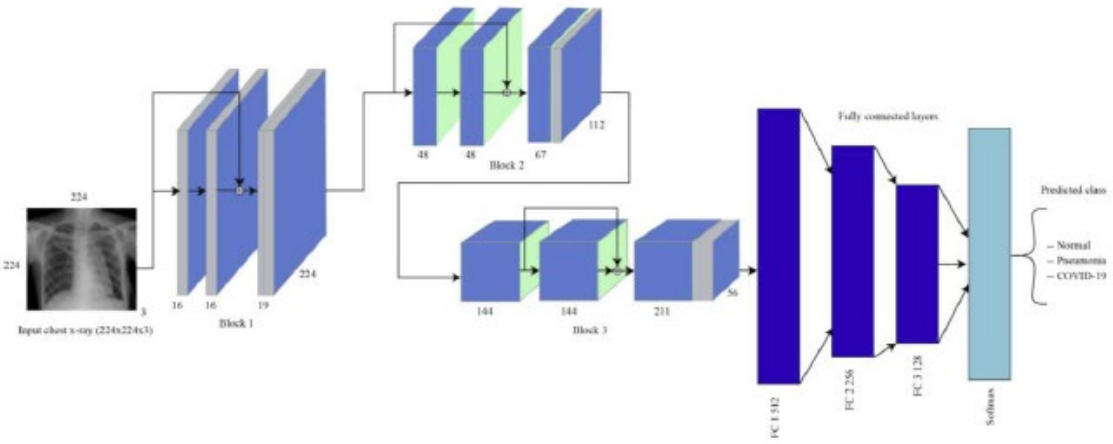

Figure 3.14: A typical ResNet architecture used for COVID-19 detection.

3.7 Metrics

The various performance metrics to evaluate model performance are accuracy, precision, recall, sensitivity, specificity, F-score, and Matthew's correlation coefficient (MCC) [22]. The definitions of these metrics are as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{(Eq. 3.3)}$$

$$\text{Sensitivity / Recall} = \frac{TP}{TP+FN} \quad \text{(Eq. 3.4)}$$

$$\text{Specificity} = \frac{TN}{TN+FP} \quad \text{(Eq. 3.5)}$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad \text{(Eq. 3.6)}$$

$$\text{F1} = \frac{2\times Precision\times Recall}{Precision+Recall} = \frac{2TP}{2TP+FP+FN} \quad \text{(Eq. 3.7)}$$

$$\text{MCC} = \frac{TP\times TN-FP\times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad \text{(Eq. 3.8)}$$

where TP, TN, FP, and FN stand for COVID-19 positive patients predicted as COVID-19 positive, COVID-19 negative patients predicted as COVID-19 negative, COVID-19 negative patients predicted as COVID-19 positive, and COVID-19 positive patients predicted as COVID-19 negative, respectively.

These are depicted in the confusion matrix shown in Table 3.9.

Table 3.9: Confusion Matrix

	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	**TP**	**FN**
Actual COVID-19 Negative	**FP**	**TN**

Although F1-score is a better choice than accuracy given the imbalanced class distribution, F1-score does not depend on the number of true negatives (TN) and only includes

true positives (TP), false positives (FP) and false negatives (FN). In fact, precision, recall and the F1-score (which is a function of precision and recall) only consider the positive class to be the class we are interested in. Regardless, of the value of TN, whether 0, 1, or very large, the precision, recall and F1-score would remain the same. Moreover, accuracy is not a good metric for imbalanced datasets.

Thus, an even more reliable metric is the Matthews correlation coefficient (MCC) or Pearson's phi coefficient [40]. The MCC gives a high score only if the prediction obtains good results in all the four confusion matrix categories (true positives, false negatives, true negatives, and false positives), proportionally both to the size of positive elements and the size of negative elements in the dataset. Thus, MCC is one of the best measurements to describe the confusion matrix and even with class imbalance. Intuitively, MCC treats the actual class and the predicted class as two binary variables and computes their true correlation coefficient. Higher values of the MCC will indicate a better performing model, as higher values of MCC indicate a greater correlation between the actual and predicted values. If the classifier is perfect, the FP=FN=0, and the MCC will be 1. If the classifier's prediction is always wrong, then the TN=TP=0, and the MCC will be -1. If the classifier is fully random, FP=FN=TP=TN=25%, then the MCC will be 0. Note that the MCC is perfectly symmetric in that it weights each of the positive and negative classes equally, regardless of the class imbalance [22].

As we identify more and more COVID-19 positive patients (higher sensitivity), we also increase our chances of mistakenly identifying normal individuals (lower specificity) and vice versa. Based on purposes for which we intend to use the test results, we can establish thresholds for sensitivity and specificity values since there is a tradeoff between sensitivity and specificity.

The field of medical diagnosis is intolerant to FN (type 2 errors) i.e., a high recall or sensitivity is a hard constraint. To get the lowest possible type 2 error, we first need to get the highest sensitivity possible, and then find the best specificity corresponding to that highest sensitivity.

Although the MCC is not derived from Sensitivity and Specificity directly, in essence, the MCC is weighting sensitivity and specificity equally, since it is weighting the FN and FP equally. However, in our case, sensitivity is more important than specificity, so we want to place more weight on sensitivity. Thus, we should evaluate the performance of models based not just on the MCC, but on both the MCC and sensitivity.

3.8 Performance Evaluation of CoviNet and Comparison with the Other Models

Based on the discussion in in earlier subsections of Section 3, we trained and tested the proposed CoviNet, CoviNet Enhanced and the other models, DenseNet169, VGG16, ResNet-50, InceptionV3, and VGG19 using the UCSD-AI4H dataset described in Section 3.2. Table 8 shows the performance comparison of the proposed CoviNet and CoviNet Enhanced classifiers compared with other five models. Among all 7 models, CoviNet and CoviNet Enhanced exhibited the highest F1-score of 0.786, and 0.930 and the MCC of 0.530, and 0.842 respectively on the UCSD-AI4H dataset. Further, CoviNet and CoviNet Enhanced show significantly superior performance than the next best model, VGG16, which had an F1-score of 0.750, and MCC of 0.541 as indicated in Table 3.10.

Table 3.10: Metrics for all models in our experiments using UCSD-AI4H dataset

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score	MCC
DenseNet169 [16] Shah et al.	73.1%	76.4%	67.9%	77.7%	0.702	0.481
VGG16 [12] Ko et al.	77.1%	76.7%	73.6%	80.1%	0.750	0.541

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score	MCC
ResNet-50 [12] Ko et al.	57.4%	57.1%	50.7%	63.3%	0.532	0.148
InceptionV3 [12] Ko et al.	69.7%	67.5%	71.9%	67.8%	0.688	0.408
VGG19 [16] Shah et al.	74.0%	73.5%	69.7%	77.9%	0.710	0.481
CoviNet (Ours)	75.0%	68.7%	91.7%	58.3%	0.786	0.530
CoviNet Enhanced (Ours)	92.2%	93.0%	93.0%	91.2%	0.930	0.842

The corresponding confusion matrices are shown in Table 3.11. In Table 3.11, the data counts for CoviNet represent the number of patients (since CoviNet takes volume data), and for all other models, the counts represent the number of images.

Table 3.11: Confusion matrices for all models in our experiments using UCSD-AI4H dataset. (a) DenseNet, (b) VGG16, (c) ResNet50, (d) InceptionV3, (e) VGG19, (f) CoviNet, (g) CoviNet Enhanced. CoviNet reports based on number of patients and all others based on number of images.

DenseNet	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	62	22
Actual COVID-19 Negative	18	47

(a)

VGG16	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	51	18
Actual COVID-19 Negative	16	64

(b)

ResNet50	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	35	34
Actual COVID-19 Negative	29	50

(c)

InceptionV3	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	50	20
Actual COVID-19 Negative	26	54

(d)

VGG19	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	49	21
Actual COVID-19 Negative	18	62

(e)

CoviNet	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	33	3
Actual COVID-19 Negative	15	21

(f)

CoviNet Enhanced	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	40	3
Actual COVID-19 Negative	3	31

(g)

Based on the discussion in Sections 3.1, 3.3, and 3.4, we trained and tested the proposed CoviNet, CoviNet Enhanced and the other models, DenseNet169, VGG16, ResNet-50, InceptionV3, and VGG19 using the MosMed dataset described in Section 3.2. Table 10 shows the performance comparison of the proposed CoviNet classifier compared with other models. Additionally, in Table 10, we also report results by Goncharov [27] who used the MosMed dataset also. Note that the MosMed data is organized as one CT volume per patient, we do not have any

concern about the same patient's data being included concurrently in training, validation, or test datasets in Goncharov et al.'s study [27].

Table 3.12 shows the performance results on CoviNet, CoviNet Enhanced and all Other models on the MosMedData, and Table 3.13 shows the corresponding confusion matrices. In Table 3.13, the CoviNet and CoviNet Enhanced counts represent the number of patients since it takes volume data, and for all other models, the counts represent the image counts. Comparing all the nine models as shown in Table 3.12, on MosMed data, CoviNet 224×224 and CoviNet Enhanced exhibited the highest F1-score of 0.930, 0.957 and the MCC of 0.881, 0.927 on the MosMed dataset. Further, CoviNet 224×224 and CoviNet Enhanced 224x224 shows significantly superior performance with an F1-score of 0.930, and 0.957 and MCC of 0.881, and 0.927 than the next best model, Goncharov [27], which had an F1-score of 0.827, and MCC of 0.770.

Table 3.12: Metrics for all models in our experiments and Goncharov [27] using MosMedData.

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score	MCC
DenseNet169 [16] Shah et al.	61.6%	52.2%	57.3%	64.3%	0.532	0.221
VGG16 [12] Ko et al.	65.2%	57.0%	52.8%	73.1%	0.541	0.268
ResNet-50 [12] Ko et al.	58.5%	46.5%	44.7%	67.3%	0.456	0.121
InceptionV3 [12] Ko et al.	64.5%	55.9%	46.4%	76.1%	0.505	0.236
VGG19 [16] Shah et al.	65.6%	59.0%	51.3%	74.8%	0.528	0.278
multitask-sp1 U-Net [27] Goncharov et al.	89.4%	72.1%	96.9%	86.8%	0.827	0.770
CoviNet (Proposed) 128×128	89.8%	89.6%	84.3%	93.4%	0.869	0.786
CoviNet (Proposed) 224×224	94.1%	89.2%	97.1%	92.2%	0.930	0.881
CoviNet Enhanced (Ours) 224x224	96.5%	94.3%	97.1%	96.1%	0.957	0.927

Table 3.13: **Confusion Matrices for all models in our experiments and Goncharov [27] using MosMedData. (a) DenseNet, (b) VGG16, (c) ResNet50, (d) InceptionV3, (e) VGG19, (f) CoviNet, (g) CoviNet Enhanced.**

DenseNet	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	486	363
Actual COVID-19 Negative	474	855

(a)

VGG16	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	449	400
Actual COVID-19 Negative	358	970

(b)

ResNet50	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	379	469
Actual COVID-19 Negative	435	894

(c)

InceptionV3	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	394	455
Actual COVID-19 Negative	318	1011

(d)

VGG19	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	435	414
Actual COVID-19 Negative	335	994

(e)

CoviNet	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	33	1
Actual COVID-19 Negative	4	47

(f)

CoviNet Enhanced	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	33	1
Actual COVID-19 Negative	2	49

(g)

The CoviNet and CoviNet Enhanced models comprise a 3-dimensional convolutional neural network (CNN) on three-dimensional CT volumes for COVID-19 detection. CoviNet and CoviNet Enhanced performed better than the other five CNNs with transfer learning since 3D CNNs learn not just from one frame per patient at a time but up to 64 frames of both lungs per patient which help provide greater clarity, detail, reduce the impact of slice-level noise and enable more precise 3D features to be learned. Another reason for CoviNet's and CoviNet Enhanced's superior performance is that we are training the CoviNet model from scratch. Another notable observation is that CoviNet's and CoviNet Enhanced's performance is much higher on MosMed than on UCSD because UCSD-AI4H dataset has fewer images and at a lower resolution, a lower number of slices per patient, and with a single image slice in most of the patients' data.

Finally, it is remarkable to see in Table 3.12 that the CoviNet which used 128×128 sized CT volume data outperforms the next highest performing model, Goncharov [27], which used 128×160 sized CT volumes. In other words, even with lower resolution input frames, CoviNet

exhibits a better F1-score and MCC than Goncharov [27]. Thus, CoviNet is truly a superior model as compared to other recently published works.

The Confusion Matrix for CoviNet and CoviNet Enhanced on MosMedData is shown in Table 3.14. CoviNet Enhanced model failed to diagnose one out of 34 COVID-19 positive patients, and falsely identified two patients as positive out of 51 COVID-19 negative patients. CoviNet showed a higher MCC, sensitivity and specificity score than those in prior published works. The model performs well on both Sensitivity (97.1%) and Specificity (96.1%). CoviNet Enhanced has a very high Sensitivity, NPV, Specificity and PPV, so it is very highly field of medical diagnosis which is intolerant to Type 2 errors. CoviNet Enhanced is an excellent model to rule-out disease if a patient tests as negative, and is suited for all populations with a low or high probability of disease.

Table 3.14: Confusion Matrix for (a) CoviNet, (b) CoviNet Enhanced on MosMedData

CoviNet	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	33	1
Actual COVID-19 Negative	4	47

(a)

CoviNet Enhanced	Predicted COVID-19 Positive	Predicted COVID-19 Negative
Actual COVID-19 Positive	33	1
Actual COVID-19 Negative	2	49

(b)

The CoviNet model failed to diagnose one out of 34 COVID-19 positive patients, and falsely identified four patients as positive out of 51 COVID-19 negative patients. CoviNet showed a higher MCC, sensitivity and specificity score than those in prior published works. The model performs well on both Sensitivity (97.1%) and Specificity (92.2%). CoviNet has both a very high

Sensitivity and NPV, so it is well suited in the field of medical diagnosis which is intolerant to Type 2 errors. This is indicated by its Positive Predictive Value (Precision) of 89.2%, and the Negative Predictive Value of 97.9%. CoviNet is an excellent model to rule-out disease if a patient tests as negative. It is most suited for populations in which there is a low to moderate probability of disease.

3.9 Randomly Split Experiment for UCSD-AI4H

In this experiment, we split the data randomly between the training, validation, and test datasets. We trained and tested all the other models, DenseNet169, VGG16, ResNet-50, InceptionV3, and VGG19 using the UCSD-AI4H dataset described in Section 3.2. This is done to prove the hypothesis that random split of data is not the correct way of doing the experiment and the metrics are 10%-15% higher than the metrics if the data split was patient-wise. Table 3.15 shows the performance results of all the other models. Any given member's image data is split randomly between the various data splits, allowing for data overlap between training and validation.

Table 3.15: Metrics for all pre-trained models using Randomly Split UCSD-AI4H.

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score	MCC
DenseNet169 [16] Shah et al.	85.38%	87.60%	81.67%	88.68%	0.8343	0.7171
VGG16 [12] Ko et al.	86.32%	83.17%	89.09%	83.87%	0.8583	0.7317
ResNet-50 [12] Ko et al.	62.33%	62.58%	54.47%	69.27%	0.5671	0.2496
InceptionV3 [12] Ko et al.	77.61%	72.16%	85.93%	70.30%	0.7820	0.5697
VGG19 [16] Shah et al.	82.56%	78.40%	89.95%	76.01%	0.8321	0.6722

Table 3.16 shows the timing performance of CoviNet on the UCSD-AI4H and MosMed.

Table 3.16: Timing performance of CoviNet model on the UCSD-AI4H and the MosMed dataset

	Training time (hours)	
Time	UCSD-AI4H	MosMed
Pre-process & Augment Data	0.17 hour	1 hour
Train time / Fold	0.42 hour	1.10 hour
Total	2.25 hours	6.58 hours

3.10 Comparison Studies

3.10.1 Comparison Study 1 for CoviNet Enhanced

CoviNet Enhanced is compared with Wang et al's DeCovNet [92] in terms of the sensitivity and specificity in Table 3.17.

Table 3.17: Comparison of CoviNet Enhanced with Wang et al's DeCovNet

Detail	CoviNet Enhanced	Wang et al.'s DeCovNet [92]
Dataset	Publicly available lung CT datasets used are UCSD-AI4H and MosMed.	The lung CT dataset utilized for Wang et al.'s DeCovNet [92] is a proprietary dataset.
Sensitivity / Specificity	CoviNet Enhanced's sensitivity is 93.0% and specificity is 91.2% (UCSD-AI4H dataset) and sensitivity is 79.1% and specificity is 82.0% (MosMed 1110).	We implemented this method, and its sensitivity is 80.95% and specificity is 91.18% (UCSD-AI4H dataset) and sensitivity is 58.3% and specificity is 76.7% (MosMed 1110).
CNN Structure	It is an ensemble classifier with separate 3D CNN and SVM models with conditional majority voting to make the final classification. It uses the Leung-Malik texture features along with original CTs as input data.	It is a single 3D CNN classifier taking only original CTs as input data.
Optimizer	Adam optimizer Exponentially decaying learning rate Initial learning rate 0.0001 Decay rate 0.96 Decay Steps 100,000	Adam optimizer Constant learning rate of $1x10^{-5}$
Early Stopping	Upper limit of 100 epochs with early stopping criterion based on validation accuracy not improving over the next 5 epochs.	Early stopping is not used 100 epochs

3.10.2 Comparison Study 2 for CoviNet Enhanced:

CoviNet Enhanced's performance is now compared with Imani's Gabor filter approach [131] in Table 3.18.

Table 3.18: Comparison of CoviNet Enhanced with Imani's Gabor filter approach [131].

Detail	CoviNet Enhanced	Imani's Gabor & Morphological filter approach [131]
Dataset	Publicly available lung CT datasets used are: UCSD-AI4H, and MosMed.	Two Public datasets: Cohen's COVID-19 Chest X-ray, and UCSD-AI4H chest CT
Accuracy F-score	UCSD-AI4H dataset: Accuracy: 92.2%, F-score: 0.930	UCSD-AI4H dataset: Random Forest classifier using Gabor filters with convolutional processing: Accuracy: 76.7%, F-score: 0.743 Random Forest using Morphological filter: Accuracy: 75.3%, F-score: 0.753
CNN Structure	It is an ensemble classifier with separate 3D CNN and SVM models with conditional majority voting to make the final classification. It uses the Leung-Malik texture features along with original CTs as input data.	First, Shape and structural characteristics extracted by morphological filters, without and with CNN processing are classified via two independent classifiers: Support Vector Machine and Random Forest. Second, Textural features extracted by Gabor filters without and with CNN processing classified via two independent classifiers: Support Vector Machine and Random Forest.

3.10.3 Additional Insights

Our classification results on MosMed are better than what Ko et al [12] also reported. Note that Ko et al.'s work is doing a different problem of diagnosing COVID-19, versus non-COVID-19 pneumonia versus normal. But we should not just look at the accuracy numbers, we need to look at the MCC score and the sensitivity metrics. Ko et al's model had an MCC score of 0.849 for Xception and 0.528 for the InceptionV3, and the Sensitivity of 78.7% for Xception and 35.6% for

InceptionV3. Our corresponding metrics for the MosMed dataset which is also a large dataset like Ko et al's proprietary data are much higher with a Sensitivity of 97.1% and an MCC of 0.882.

Goncharov et al. [27] used the entire MosMed dataset for classification, we only used a subset of the MosMed data for CoviNet's classification. We agree that doing the analysis on a larger dataset will help. For this reason, we will also repeat our experiment on the entire MosMed data comprising 1110 patients.

Our classification results on the MosMed dataset are having a Sensitivity of 97.1% and an MCC of 0.882, versus Goncharov et al's sensitivity of 96.9% and the MCC of 0.77. Goncharov et al's method of multitask learning was also very slow to train versus our relatively more efficient model. Goncharov et al. also just reported the AUC metric, but we performed a more extensive performance analysis and reported all performance metrics that help to understand the model performs better.

We can conclude that supervised learning-based deep-learning approaches are effective in automating the image analysis for COVID-19 detection via CT images.

CHAPTER 4

RELATED WORK: COVID-19 SEGMENTATION*

4.1 Overview

A systematic review of 415 published papers from January 2020 to October 2020 for the diagnosis or prognosis of COVID-19 from lung imaging revealed some methodical deficiencies [28], [29]. There is a high likelihood of duplicated images across datasets called "Frankenstein datasets", biases such as datasets not representative of populations, and low-quality data. Accurate segmentation of computerized tomography (CT) volumes is a challenge due to complex structures, pathological changes, individual differences in infection characteristics, and low image quality [34]. In disease segmentation from CT scans, an insufficient amount of data to train a deep model, and images having a low contrast between infections and normal tissues are some additional challenges [4]. Moreover, class imbalance and annotation errors also make the segmentation task more challenging [43]. To solve these issues and make the research reproducible, higher-quality datasets, heavily documented research, and external validation are needed.

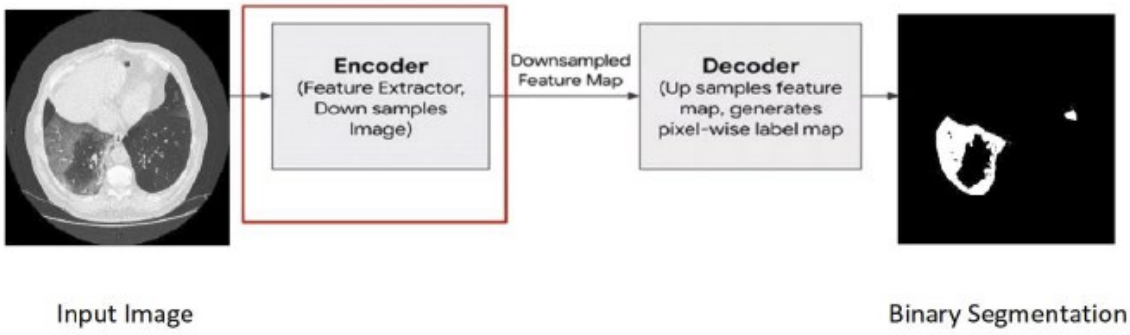

Figure 4.1: Neural network-based image segmentation

* Sections 4.1 to 4.3 are reproduced from Mittal, B. & Oh, J. CoviSegNet: Covid-19 Disease Area Segmentation using Machine Learning Analyses for Lung Imaging. ISPA 2021, 13-15 September 2021, Zagreb, Croatia [accepted for publication], with permission from IEEE.

Figure 4.1 showcases the Basic Architecture of neural network-based image segmentation of the input images to generate the binary predictions of the segmented mask.

The U-Net network, like SegNet, has the same layer counts of pooling and up-sampling layers, but it also has trainable deconvolution layers [86], [105]. Also, in this network, there is a corresponding skip connection between the up-sampling and down-sampling layers. Figure 4.2 shows a general form of U-Net architecture used to segment the lung in COVID-19 patients.

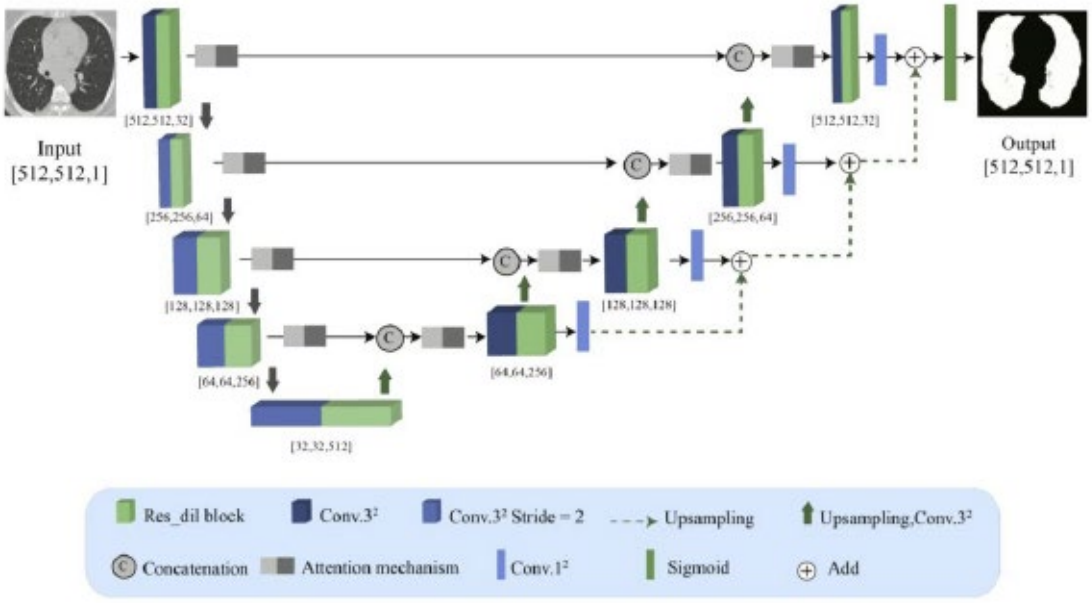

Figure 4.2: U-Net architecture

The U-Net family of models (U-Net, SegNet, PSPNet and Mask-RCNN etc.) are Fully Convolutional Neural Networks (FCNs) and are widely used in semantic segmentation including medical image segmentation. The fully Convolutional Networks for Semantic Segmentation [98] comprise several convolutional layers. The earlier convolutional layers are the encoder and perform feature extraction and downsampling by aggregating the low-level features to high-level features. The later convolutional layers are the decoder and perform up sampling to original size to generate the pixel-wise map.

4.2 Related Work Review

Shi et al. [42] reviewed the trends in artificial intelligence-based classification of COVID-19 disease in lung imaging to help medical specialists. Accurate delineation of infections in X-ray and CT images is highly important for subsequent quantification of disease severity. Such research aims to help radiologists make clinical decisions, i.e., for disease diagnosis, tracking, and prognosis. They discuss the entire pipeline of medical imaging and analysis techniques involved with COVID-19, including image acquisition, segmentation, diagnosis, and follow-up. Mertz [39] explained physician-established knowledge as to which X-ray or CT abnormalities strongly indicate a COVID-19 diagnosis. These abnormalities are the localization of lesions lower in the lungs, as well as hazy areas are known as ground-glass opacities (GGOs) and consolidations (the amount of the lung that is filled with fluid instead of air) that are peripheral and bilateral.

Zhou et al. [94] proposed the U-Net based segmentation network using a spatial attention and channel attention modules better feature representation of the rich contextual relationships. In addition, the focal Tversky loss is introduced to deal with small lesion segmentation. The obtained Dice Score and Hausdorff Distance are 83.1% and 18.8, respectively.

Zhou et al. [95] implemented the UNet++ with the encoder and decoder sub-networks connected through a series of nested, dense skip pathways which reduce the semantic gap between the feature maps of the encoder and decoder sub-networks. Nodule segmentation in the low-dose CT scans of chest, nuclei segmentation in the microscopy images, liver segmentation in abdominal CT scans, and polyp segmentation in colonoscopy videos were evaluated. Figure 4.3 showcases the UNet++ model architecture. The black connections in part (a) indicate the original U-Net, and the green and blue connections show dense convolution

blocks on the skip pathways, and red connections indicate deep supervision. Part (b) shows the first skip pathway of U-Net++ and part (c) shows that UNet++ can be pruned at inference time, if trained with deep supervision.

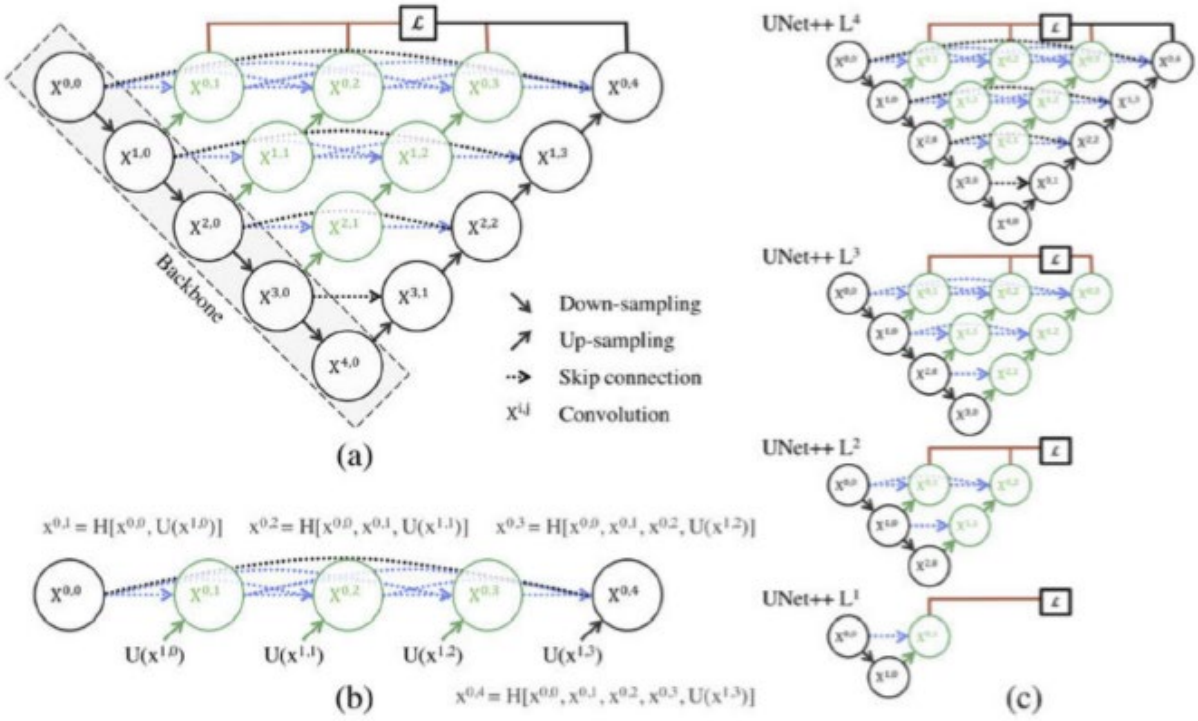

Figure 4.3: U-Net architecture (a) original with skip connections, (b) detailed analysis of the first skip pathway of UNet++, (c) UNet++ after pruning.

Goncharov et al. [27] proposed 2D and 3D U-Net models for segmentation of COVID-19 diseased areas including the consolidations and ground-glass opacities on CT imaging data. The MosMedData [49] was utilized, which is a publicly available dataset from Russia. It contains 50 physician-labeled masks and the corresponding 50 CT volumes from 50 patients with mild severity of COVID-19 disease. The dice score achieved was 65.0%. This is likely because the dataset is highly sparse and contains only mildly affected COVID-19 patients and the proportion of COVID-19 lesion area to the total lung volume is very small. Amyar et al. [31] proposed a new multitask deep learning model to jointly identify COVID-19 patients and segment COVID-19

lesions from chest CT images. Three learning tasks: segmentation, classification, and reconstruction were jointly performed with different datasets. A common encoder was used for disentangled feature representation with three tasks, and two decoders, and a multi-layer perceptron for reconstruction, segmentation, and classification, respectively. For segmentation, the dice coefficient achieved was 88%.

Anthimopoulos et al. [32] proposed a deep purely convolutional neural network for the semantic segmentation of Interstitial Lung Disease (ILD) patterns to detect and classify the ILD type from thoracic CT scans which is a challenging task even for experienced radiologists. Semi-supervised learning was used utilizing both labeled and unlabeled image regions and achieved 81.8% weighted (balanced) accuracy and an inference time of 58 ms.

Cai et al. [33] quantified COVID-19 pneumonia from CT image data from 99 patients using two U-Net models to segment lung and COVID-19 lesions. Lung volume, lesion volume, non-lesion lung volume (NLLV), and non-lesion lung volume percentage were quantified. Further, RF classifier was used to classify disease severity (Moderate, Severe, and Critical), and predict the length of ICU stay, oxygen support, hospital stay, sputum NAT-positive, and patient prognosis. The mean %NLLV in three severity groups were 92.18%, 82.94%, and 66.19% for moderate, severe, and critical cases, respectively. The AUC in classification of moderate vs (severe + critical) was 92.7%, and for severe vs critical was 92.9%.

Chen et al. [34] proposed a novel dictionary-based approach to segment lung tumors from CT images. Sparse shape composition is integrated with the eigenvector space shape prior model, called eigenspace sparse shape composition, to reduce local shape reconstruction error caused by the weak and misleading appearance prior information. Furthermore, a new vertex search

strategy based on the gradient vector flow field is also proposed to drive the shape deformation to the target boundary.

Elaziz et al. [35] proposed the MPAMFO multi-level thresholding approach for segmentation which combines the features of marine predators' algorithm (MPA) and moth-flame optimization (MFO). In this, the MFO was utilized as a local search method for MPA to avoid trapping at local optima. Thirteen COVID-19 images were segmented using this MPAMFO approach with five threshold levels on the image histograms and compared against eight other methods. The Peak Signal-to-Noise Ratio (PSNR) was highest for MPAMFO in 70% of images and the Structural Similarity Index (SSIM) was highest for MPAMFO in 61% of the images. A limitation is that a very small dataset with just 13 COVID-19 images and segmentation result was not compared against the ground truth.

Fan et al. [4] proposed an automatic deep network for segmentation called Inf-Net. First, a parallel partial decoder was used to aggregate the high-level features and generate a global map. Then, the implicit reverse attention and explicit edge attention are utilized to model the boundaries and enhance the representations. Next, semi-supervised segmentation is done with a randomly selected propagation strategy, which only requires a few labeled images. Semi-Inf-Net outperforms U-Net and U-Net++ and achieves a dice score of 59.7% with semi-supervised learning.

Koohbanani et al. [37] proposed NuClick, a CNN-based approach to segment nuclei and cells in histology and cytology images. Just one click inside each object is enough for NuClick to yield a precise annotation. For multicellular structures such as glands, authors proposed a novel approach to segment the glandular boundaries requiring the annotator to provide a squiggle as

a guiding signal. NuClick applies to a wide range of object scales, robust against variations in the user input, adaptable to new domains, and delivers reliable annotations for training machine learning models.

Liu et al. [38] proposed a novel U-Net with a deep residual attention module (DRANet) to segment and quantify ischemic stroke lesions and WMH lesions in the MRI images accurately and simultaneously. Accurate segmentation and quantification of ischemic stroke lesions and WMH lesions are important for the diagnosis and prognosis of ischemic stroke. However, radiologists have a difficult time distinguishing these two types of similar lesions. The dice coefficient of ischemic stroke segmentation is 76.39% and that of WMH segmentation is 72.83%.

Oulefki et al. [40] performed Lung CT image segmentation which is a necessary initial step for lung image analysis, but segmentation is difficult due to intensity in-homogeneity, presence of artifacts, and closeness in the gray level of different soft tissues. The accuracy, sensitivity, F-measure, precision, MCC, Dice, Jacquard, and specificity are 98%, 73%, 0.71, 0.73%, 0.71, 71%, 57%, 99% respectively.

Saeedizadeh et al. [41] proposed TV-UNet, which is a U-Net-based model, to detect ground-glass regions at a pixel level. In the TV-UNet model, 2D-anisotropic total variation is added to the binary-cross entropy loss function which serves as the regularization term to promote connectivity of the segmentation map for COVID-19 pixels. A dice score of 76.4% was achieved on their 70:30 train versus test data split on the MedSegData [48].

Voulodimos et al. [43] proposed two deep learning models for semantic segmentation of pneumonia infected area segmentation in CT images for the detection of COVID-19. The U-Net model has an F-score of 0.65 and the FCN model has an F-score of 0.58 on the test dataset. Wu

et al. [44] proposed a hybrid weak label-based approach to quantify disease severity via CT images based on the extent and type of pulmonary opacities. It uses both manually annotated pulmonary opacities from COVID-19 pneumonia and the patient-level disease-type data from the clinical report. First, a U-Net classifier was trained with semantic labels to segment the total infected region. It was used to initialize another U-Net, which was trained to segment the consolidations with patient-level information using the Expectation-Maximization (EM) algorithm. Finally, the segmentation model's performance was evaluated on several datasets from institutes in Iran, Italy, South Korea, and the United States, and a dice coefficient of 63.2% was achieved.

Zhang et al. [45] proposed a novel conditional generative model, called CoSinGAN for COVID-19 infection segmentation from a single radiological image given the annotation mask of the lungs and infected regions. They performed five-fold cross-validation on the COVID-19-CT-Seg dataset (20 CT cases) and the held-out MosMed dataset (50 CT cases). Both 2D U-Net and 3D U-Net have outperformed COVID-19-CT-Seg-Benchmark. Hassantabar [9] performed COVID-19 segmentation of COVID-19 affected areas in lung images using a CNN architecture with 3 convolutional layers but only achieved a Jaccard index of 0.4.

Safarov et al.'s [83] work in from the colonoscopy domain and they created a U-Net++ and DenseNet based model for segmenting polyps from two public colonoscopy datasets with physician-annotated binary masks for 1000 and 612 polyp images, respectively. CoviSegNet's performance on the MedSeg data is better than that of Safarov et al's A-DenseUNet. CoviSegNet's dice score and IoU on MedSeg data is 93.2% and 87.3% versus A-DenseUnet's (CVC-

612 dataset) is 79.55% and 79.62%. Additionally, CoviSegNet's dice score and IoU on MosMed data is 91.5% and 84.3% versus A-DenseUnet's (KVASIR-SEG dataset) is 81.33% and 79.27%.

In terms of the model architecture, the A-DenseUnet uses 164 layer DenseNet as the U-Net++ backbone, and CoviSegNet uses the 169 layer DenseNet as the U-Net backbone. A-DenseUnet's skip connections connect the different depths of the U-Net and have horizontal dense connections and connections between each depth, whereas CoviSegNet uses same depth skip connections only. Both A-DenseUnet and CoviSegNet use the spatial and channel attention mechanisms which produces excellent results by suppressing noise and focusing on the regions of interest. A-DenseUnet also employed the dilated convolution as shown in Figure 4.4 to systematically aggregate multi-scale contextual information without losing resolution.

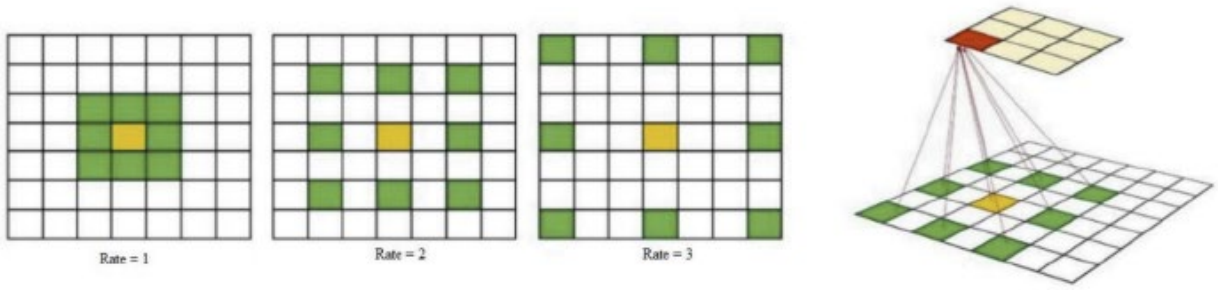

Figure 4.4: Dilation Convolution with different dilation rates

Xie et al. [88] proposed the RTSU-Net leveraging the structural relationships for pulmonary lobe segmentation between pulmonary lobes, vessels, airways, and the pleural wall by introducing a novel non-local neural network which learns the self-attention weights. RTSU-Net was trained and validated on a cohort of 5000 subjects from the COPDGene study.

Xie et al. [89] proposed a novel framework combining a Convolutional neural network and a Transformer (CoTr) for accurate 3D medical image segmentation of 11 major human organs. The CNN extracts the feature representations, and an efficient deformable Transformer

(DeTrans) models the long-range dependency on the extracted feature maps. The transformer, DeTrans, pays attention selectively via the deformable self-attention mechanism which reduces the computational complexity of the transformer. The 'MultiAtlas Labeling Beyond the Cranial Vault (BCV)' dataset covering 11 major human organs was used for validation.

Wu et al. [87] proposed a Joint Classification and Segmentation (JCS) system to perform real-time and explainable COVID-19 chest CT diagnosis on chest CT images of 400 COVID-19 patients and 350 uninfected cases. They reported an average sensitivity of 95.0% and a specificity of 93.0% on the classification test set, and 78.5% Dice score on the segmentation test set. They combined the segmentation and classification models by combining the encoder features of the segmentation model with the backbone features of the classification model. The loss function used was the sum of the standard binary cross-entropy loss and the Dice loss.

Li et al. [79] proposed a U-net Ensemble Model for Segmentation in Histopathology Images. Saood et al. [108] implemented two known deep learning networks, SegNet and U-NET, for COVID-19 segmentation of infected and healthy lung tissue in lung CT. SegNet is characterized as a scene segmentation network and U-NET as a medical segmentation tool. Authors perform experiments on a relatively small dataset with just 100 images, and SegNet shows 0.95 mean accuracy in classifying infected vs. healthy tissues, while the U-NET shows 0.91 mean accuracy in segmentation.

Han et al. [109] proposed the 2.5D Perpendicular-UNet to fuse the segmentation results of three perpendicular 2.5D Res-UNets in the task of liver and hepatic tumor segmentation which reduces computational complexity of 3D models and shows improved segmentation accuracy than 2D models.

Hou et al. [110] proposed spatial partitioning, which internally distributes the input and output of convolutional layers across GPUs/TPUs using the Mesh-TensorFlow framework by training a 3D U-Net on 512 by 512 by 512 high resolution data. This overcomes the memory limitations of a single GPU/TPU limitation and using lower resolution cropped 3D volumes of 256x256x64 which leads to information loss.

Jin et al. [111] proposed a prior knowledge driven domain adaptation and a dual-domain enhanced self-correction learning scheme model (DASC-Net) for COVID-19 infection segmentation on CT images. It consists of an attention and feature domain enhanced domain adaptation model (AFD-DA). Enhancements are an image-level activation feature extractor with attention to lung abnormalities and a multi-level discrimination module for hierarchical feature domain alignment.

4.3 Attention Mechanism

Attention helps to focus the model on one or a few key elements at a time. Content-based soft attention is convenient, and the system learns about where to attend using backpropagation. Attention was first discovered in the field of Neural Machine Translation by Bahadanau et al. [126] and Li et al. [127] and their work helped address the vanishing gradients problem of deep neural networks. Attention is an internal action, as it needs a learned attention policy.

Attention really works well for the highly sparse data seen in COVID-19 segmentation and detection. Some mild cases may only have less than 5% of the lung volume showing ground-glass opacities and to accurately segment these regions, the model should be able to learn well even from the highly sparse data. Mathematically, sparse data is harder to learn from as the gradients may not be large enough, and only occur in a small region of the images or volumes.

Moreover, the sequential nature of the COVID-19 images in CT volumes can be captured using the attention mechanism alone, without the use of Long Short-Term Memory (LSTM) networks or RNNs [90]. This attention approach is preferred as it is remarkably more time-efficient than the RNNs or LSTMs.

CHAPTER 5

METHODOLOGY AND EXPERIMENTS: COVID-19 SEGMENTATION*

5.1 Proposed CoviSegNet Model

Our proposed CoviSegNet is a novel approach for disease severity quantification of COVID-19. Our approach consisted of seven steps: (1) First, to enhance the U-Net, our encoder comprises an ImageNet-pre-trained DenseNet169 as the backbone which serves to learn the hierarchical features efficiently. (2) The decoder block follows and there are skip connections between the encoder and the decoder. (3) Thereafter, spatial and channel attention modules are applied, so that the model can learn to focus on the region of interest. (4) Next, transfer learning is applied on the encoder pre-trained DenseNet 169 block, and supervised learning from scratch is done for the decoder block with the skip connections and the attention modules. (5) Then, the pixel-wise predicted mask is generated for the whole image with each pixel outputting the probability of it being the COVID-19 disease-affected region. (6) Subsequently, we apply adaptive thresholding to get the finalized segmentation as the predicted binary mask, with white pixels showing COVID-19 disease-affected regions and black pixels denoting non-diseased areas. (7) Finally, performance evaluation is done by comparing the predicted masks with the physician-annotated ground truth masks on the held-out test dataset.

This is a novel approach and no previously published research in COVID-19 segmentation utilized the spatial-attention and channel-attention mechanism. These U-Net models are

* Sections 5.1 to 5.5, 5.8, 5.9 are reproduced from Mittal, B. & Oh, J. CoviSegNet: Covid-19 Disease Area Segmentation using Machine Learning Analyses for Lung Imaging. ISPA 2021, 13-15 September 2021, Zagreb, Croatia [accepted for publication], with permission from IEEE.

implemented using a supervised learning framework on three public, and physician-annotated datasets.

Our proposed CoviSegNet comprises the spatial and channel-attention-based model on the pre-trained DenseNet169 backbone U-Net model for COVID-19 lesions segmentation. This is an enhancement of the U-Net method in that it uses spatial-attention and channel-attention mechanisms to learn the hierarchical features of U-Net. Additionally, the attention U-Net model is built on an ImageNet-trained backbone, and the transfer learning paradigm is applied. This is a novel approach and no previously published research in COVID-19 segmentation utilized the spatial-attention and channel-attention mechanism. Additionally, previously published research on COVID-19 segmentation did not use the pre-trained DenseNet169.

The model architecture is shown in Figure 34. We will now discuss the architecture in detail reviewing each of the components from left to right and each of the layers in those components shown in Figure 5.1.

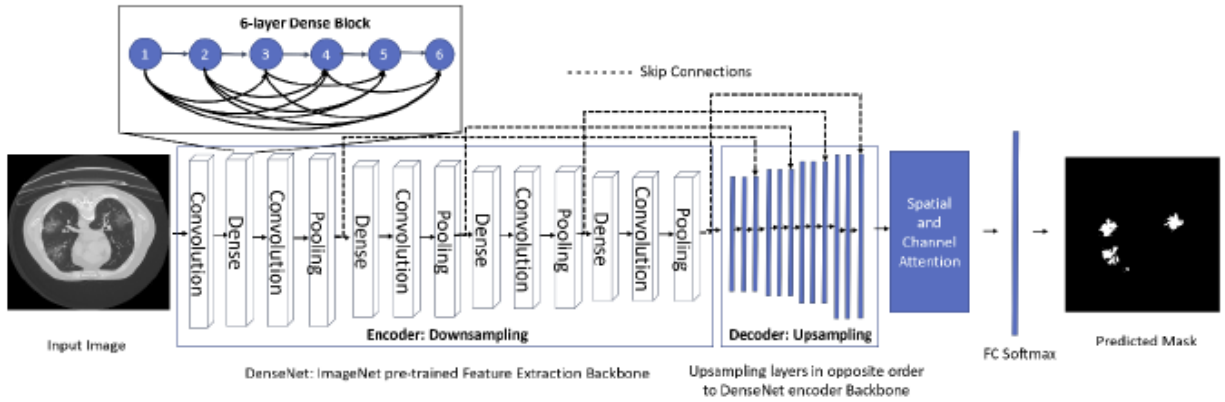

Figure 5.1: The architecture of the proposed CoviSegNet model.

The input data should be in individual image slices with the three channels (Red, Blue, and Green). Next, the encoder is the CNN part of the architecture, and it downsamples the image. In

general, the encoder comprises multiple sets of successive convolutional and pooling layers, with each successive pooling layer reducing the image height and width dimensions by a factor of 2. For the encoder block, any of the existing CNNs could be used, such as simple CNN with just the convolution and pooling layers, or VGG6, ResNet50, etc. Instead of training the encoder layers from scratch, ImageNet pre-trained convolutional neural networks such as DenseNet169, VGG16, and ResNet50 with a transfer learning paradigm are preferred. This is because the transfer learning strategy is expected to work well even with a limited amount of data that might be insufficient to train a deep model from scratch.

Two different encoder backbones, namely, the DenseNet169 and the VGG16, pre-trained on ImageNet were applied to identify the best-performing model. In the DenseNet169 architecture-based backbone, the feature maps of all preceding layers are used as inputs into each layer, which enables the various layers to share the collective feature maps. Each layer only adds only a small set of feature maps and allows the flow of information and gradients through the network facilitating quick training. DenseNet169 was originally proposed by Huang et. al. in [46]. The VGG16 architecture-based backbone utilizes same-padding in its convolution layers, which work better in extracting and learning the hierarchical features than the zero-padding used in some other models such as ResNet. These hierarchical features were used for further processing by the attention gates. Note that a dropout layer and a batch normalization layer is present after each pooling operation, to facilitate efficient training without overfitting.

DenseNet169 is finally selected as the preferred CNN to be used as the Encoder in the CoviSegNet model. DenseNet comprises an initial convolutional layer, which is then connected to the dense block. The first dense block has 6 layers, and each of those layers is connected to all

the following layers in the block. After that, there is another convolutional layer, followed by the pooling layer. Then, there were three more such successive sets each comprising the dense block, convolutional layer, and pooling layer. The output of the last pooling layer from the encoder is then fed to the decoder block. The decoder performs the up sampling and has the layers in reverse order than the encoder, with the pooling layers replaced by the inverse-pooling layers which serve to increase the length and width by a factor of two. There are skip connections between the encoder and decoder, and this is what makes this a U-Net model.

After that spatial and channel attention modules are applied which help the model focus on the regions of interest and learn from the highly sparse data efficiently and effectively. These attention gates learn the hierarchical feature representations from the decoder and encoder layers. This is followed by a fully connected softmax layer which outputs pixel-level probabilities of the image showing COVID-19 disease.

Then, 3x3 Median filtering is done followed by adaptive thresholding to generate the pixel-level predicted binary mask.

5.2 Median Filter and Adaptive Thresholding

The various noise types of noise that affects computerized tomography images include speckle noise, Gaussian noise, salt, and pepper noise [115]. Such noise causes the predicted probability map of segmentation to also exhibit salt and pepper noise as shown in part (b) of Figure 5.2.

We thus apply the median filter to remove the salt and pepper noise, followed by adaptive thresholding to get the final predicted binary mask. Median filtering is a widely used technique in image processing for noise reduction while preserving the edges. It is highly effective in

removing low to moderate levels of Gaussian noise, speckle noise and salt-and-pepper noise [116]. A 3x3 window slides over the input image, replacing each entry with the median of neighboring entries as shown in Figure 5.3.

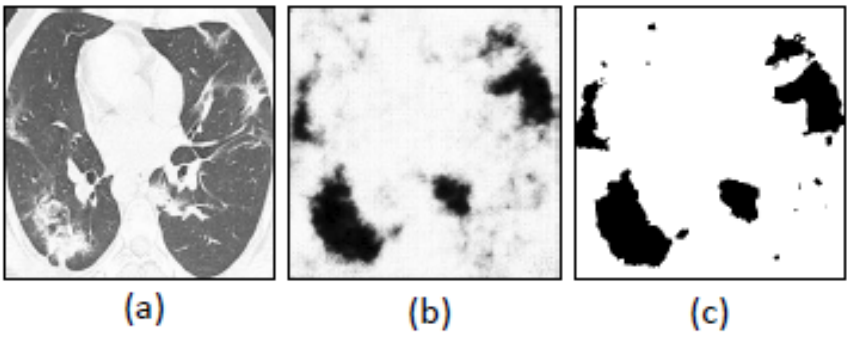

Figure 5.2: (a) Original CT Image, (b) Predicted probability of COVID-19 diseased area shown in black color, (c) Final predicted binary mask after applying a 3x3 median filter and adaptive thresholding.

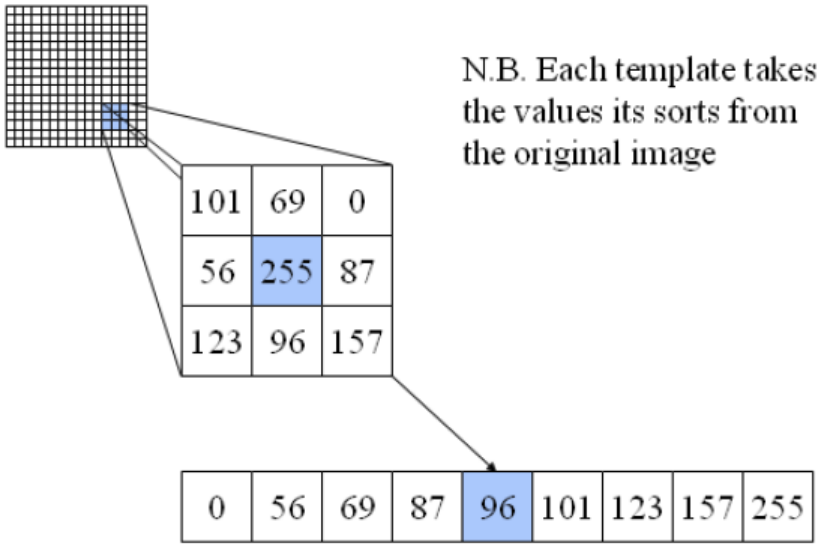

Figure 5.3: Median filter showing the 3x3 filter; the center pixel will be replaced by the median value of 96.

Otsu's method for adaptive thresholding technique [117] assumes a bimodal image with clearly defined foreground and background, as we have in the sky and cloud images. The algorithm iteratively tests thresholds to find which one produces the minimum variance in the foreground and background saturation values. The feature extraction process is the same as

described above for the fixed thresholds; the only difference is that an optimal threshold is found for every image instead of for the whole set of training images.

We first convert the RGB pixel probability image to a grayscale image. Then we apply the Otsu's method for Adaptive thresholding [117] to get the unique threshold for each image. That image specific threshold will then be used to get the binary prediction segmentation mask.

These enhanced U-Net model creation steps are explained in Figure 5.4.

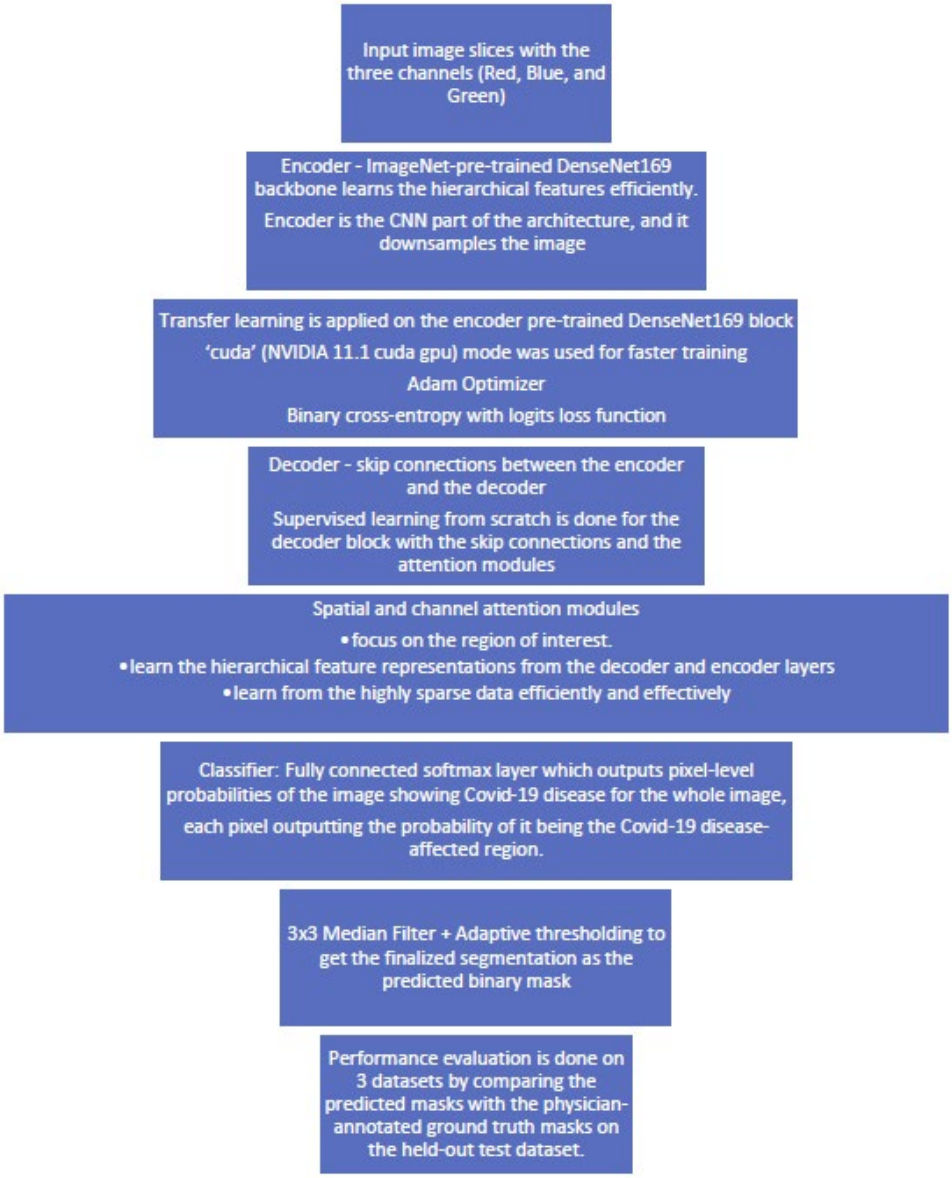

Figure 5.4: CoviSegNet, a novel enhanced U-Net for disease severity quantification of COVID-19.

5.3 Datasets

Three publicly available datasets with a CT image data are selected for this research. All these datasets have been used in prior research studies and have many CT segmentations. The data in [48] and [49] was annotated by radiologists with binary pixel masks of regions of interest (characteristic areas of consolidation and ground-glass opacities) showing COVID-19 diseased areas versus non-diseased areas.

5.3.1 MedSegData [48]

MosMedData comprises:

Part 1: Single CT volume with 100 slices and the corresponding ground truth masks showing COVID-19 diseased areas from 43 patients (more than 70% have only 1 or 2 slices per patient)

Part 2: 9 CT volumes with 350 total slices in Part 2 of the dataset.

MedSegData [48] has both normalized images with corresponding binary masks in the form of NIFTI-files. Part 1 of the dataset has 100 CT image slices of 43 COVID-19 positive patients' CT volumes, and Part 2 of the dataset has 350 CT image slices of 9 COVID-19 positive patient's CT volumes. [48] was annotated by two radiologists from Oslo, Norway for COVID-19 positive cases on data originally made public by SIRM (Italian Society of Medical & Interventional Radiology) [53].

The Hounsfield Unit scale was used for normalization (the air was normalized to -1000 and fat to -100), which is frequently used in volumetric CT medical images. After that, the images were resized and saved as a single NIFTI-file (512 x 512 x 110). The authors only shared the volumes for the slices that they annotated, and not the full volumes with all the annotated and unannotated slices.

The pixel-level ground truth masks have three different types of disease-affected regions, namely, ground-glass opacities, consolidation, and pleural effusion. The negative class is shown as black pixels to indicate COVID-19 disease-free regions. The pleural effusion class is shown in white, the consolidation class is shown in light gray, and the ground-glass opacities class is in dark gray. Figure 5.5 shows the CT image along with its corresponding mask. The Image and patient count for Covid SemiSeg dataset are shown in Table 5.1.

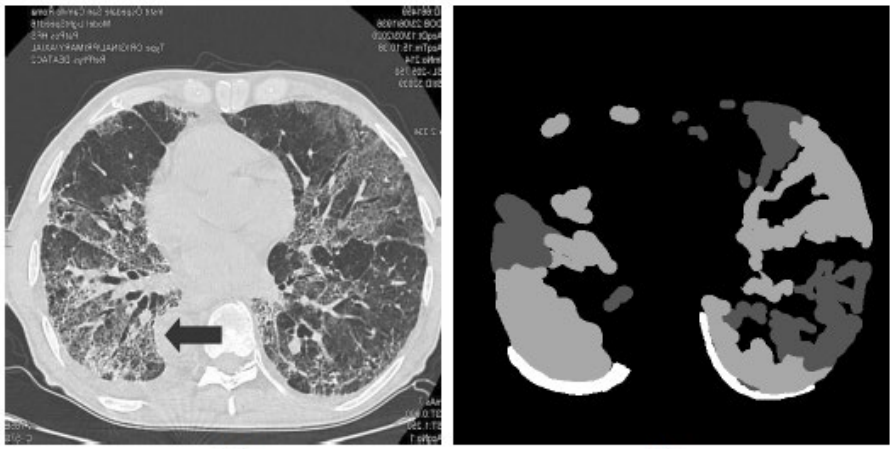

Figure 5.5: Sample CT from the MedSegData [48] (a) a CT slice (b) its corresponding ground truth mask showing three diseased classes, namely, pleural effusion in white, consolidation in light gray, and ground-glass opacities in dark gray.

Table 5.1: Dataset 1: MedSegData [48]

Counts	Training	Validation	Test
Images	342	39	69
Patients	37	5	10

5.3.2 MosMedData [49]

The MosMedData comprises 50 patients' CT volumes with a total of 2049 CT slices and corresponding labeled masks.

MosMedData [49] has CT volume data in .nii format which contain 2049 non-normalized images, and masks in the form of NIFTI-files for 50 COVID-19 positive patients. [49] was annotated by two radiologists independently. If the raters' contours did not align, another meta-rater requested annotation correction.

The pixel-level ground truth masks are binary. The negative class is shown as black pixels to indicate COVID-19 disease-free regions, and the positive class is shown as white pixels denoting both the ground-glass opacities and consolidations combined into one pixel-level label. Figure 5.6 shows the CT image along with its corresponding binary mask. The Image and patient counts for the MosMedData dataset are shown in Table 5.2.

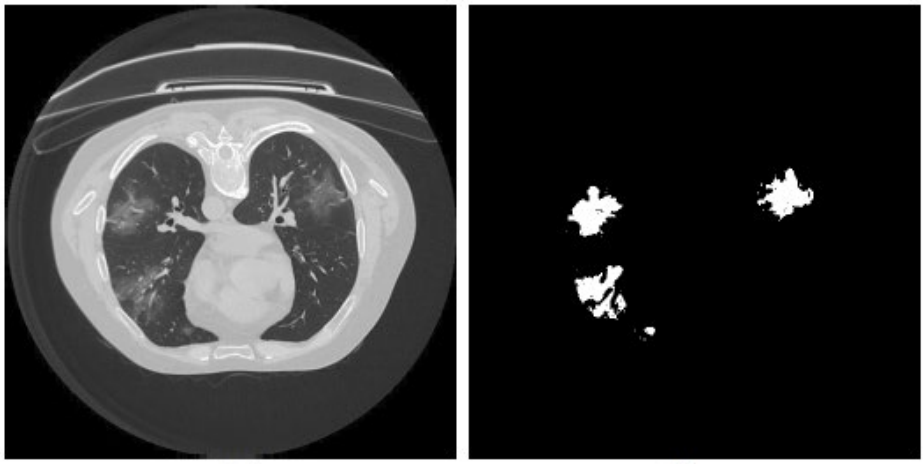

Figure 5.6: Another sample CT from the MosMedData [49] (a) a CT slice (b) its corresponding binary ground truth mask.

Table 5.2: Dataset 2: MosMedData dataset [49]

Counts	Training	Validation	Test
Images	1476	164	409
Patients	35	5	10

5.3.3 Covid SemiSeg [50]

This includes:

- the MedSegData [48] with 100 slices and labeled masks from 43 patients, and
- 1600 CT images of 45 patients from Cohen et al. [56] COVID-19 CT Collection dataset with corresponding masks.
- Labeled via semi-supervised learning by Fan et al. [50]

The 1600 images from COVID-19 CT Collection dataset by Cohen et al. [56] were pulled from various sources including but not limited to Radiopedia, SIRM and coronacases.org. These 1600 images do not have segmented labels from by a physician, as the labels were created via semi-supervised learning methods by Fan et al. [50]. But, since this dataset was made public with the exact training and testing data split used, and several other published works also utilized this dataset. Hence, we chose to utilize the dataset in this current research to serve as a comparison with recently published work.

The pixel-level ground truth masks are binary. The negative class is shown as black pixels to indicate COVID-19 disease-free regions, and the positive class is shown as white pixels denoting both the ground-glass opacities and consolidations combined into one pixel-level label. Figure 5.7 shows the CT image along with its corresponding binary mask. The Image and patient count for Covid SemiSeg dataset are shown in Table 5.3.

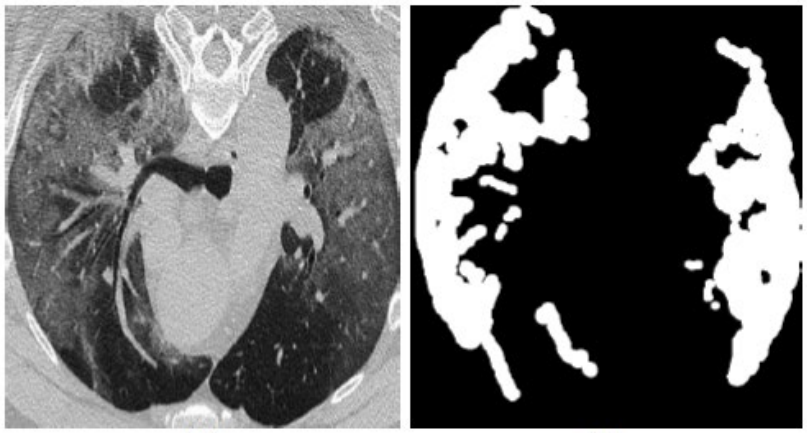

Figure 5.7: Sample CT from the Covid SemiSeg [50] (a) a CT slice (b) its corresponding ground truth mask.

Table 5.3: Dataset 3: SemiSeg dataset [50]

Counts	Training	Validation	Test
Images	1190	170	340
Patients	62	9	17

5.4 Pre-Processing

Pre-processing is done to prepare the datasets. First, the volume data with .nii files is converted to slice level .jpeg using the MIPAV tool [96] and the NIfTI-Image-Converter [77]. The dimensions of the original volume data are N×C×H×W×D, where N is the number of images, C denotes the channels, H is the height of the image, W is the width of the image and D is the depth corresponding to all image slices for a single CT. After this conversion, the slice level dimensions are N×C×H×W. Note that there is a single channel for the binary mask and the red, blue, and green channels for a CT image.

Then, the images and their corresponding lung mask images are loaded in batches. The images are rescaled to the fixed pixel spacing of 2,2, the image intensities are clipped to the fixed window and the input images were normalized to have pixel values ranging from (0, 1). Non-COVID-19 images that get tagged as 'negative' should ideally be excluded. Note that all datasets we used only had COVID-19 positive patients' data. Some slices show signs of COVID-19, but the physician-provided mask has all black pixels because not every slice of a positive patient will manifest disease. Those frames are the 'positive without a mask showing the COVID-19 positive class' frames and those frames which have the COVID-19 diseased areas are 'positive with a mask showing the COVID-19 positive class.' Both these categories will be included in training, validation, and test sets to mimic real-world data. While some published works discarded the frames with zero annotations i.e., blank annotations to make the model train better, we chose

not to do that because we want the model to process the real-world data which has both the closed and open lung frames. The pre-processing steps are shown in Figure 5.8.

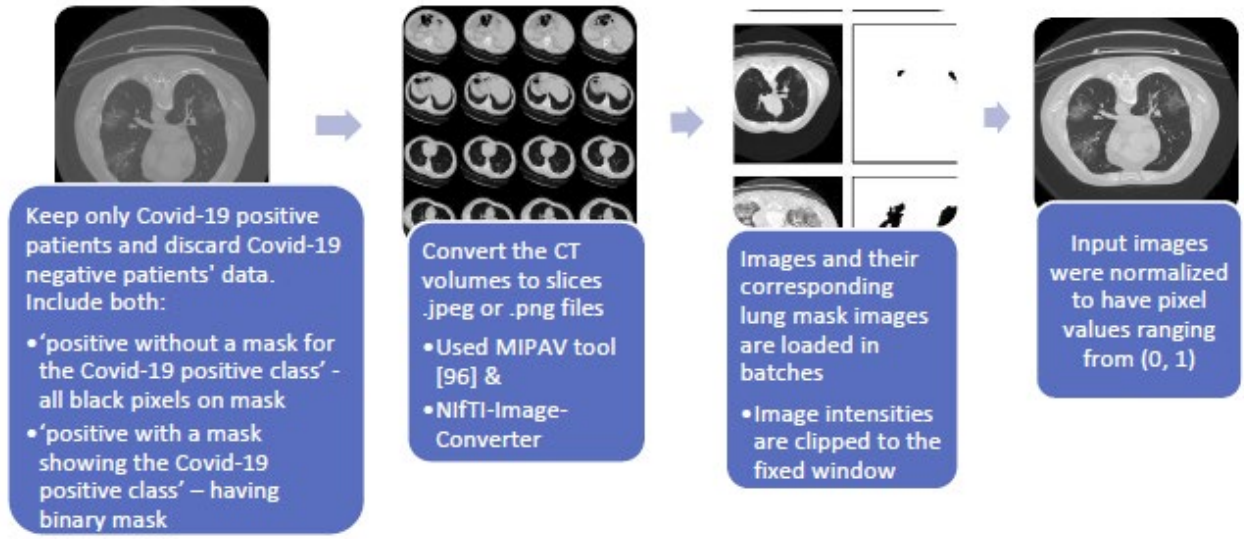

Figure 5.8: Pre-processing steps involved in CoviSegNet.

5.4.1 Additional Pre-Processing Needed for the MedSeg Dataset

Some of the labeled images had the pixel-level ground truth masks have three different types of disease-affected regions, namely, ground-glass opacities, consolidation, and pleural effusion. Per advice received from an expert radiologist, we omitted the pleural effusion positive class, and we combined the GGO and consolidation into a single positive segmentation mask. The rationale for leaving out the pleural effusion class is that the pleural effusion class occurs in the pleural cavity which is not a part of the lung; it is a space between the pleura lining of the lung and the chest wall. Additionally, pleural effusion is not a common symptom and is rarely seen in COVID-19 positive patients. Figure 5.9 shows the CT image along with the original three-class ground truth and the final single-class ground truth mask.

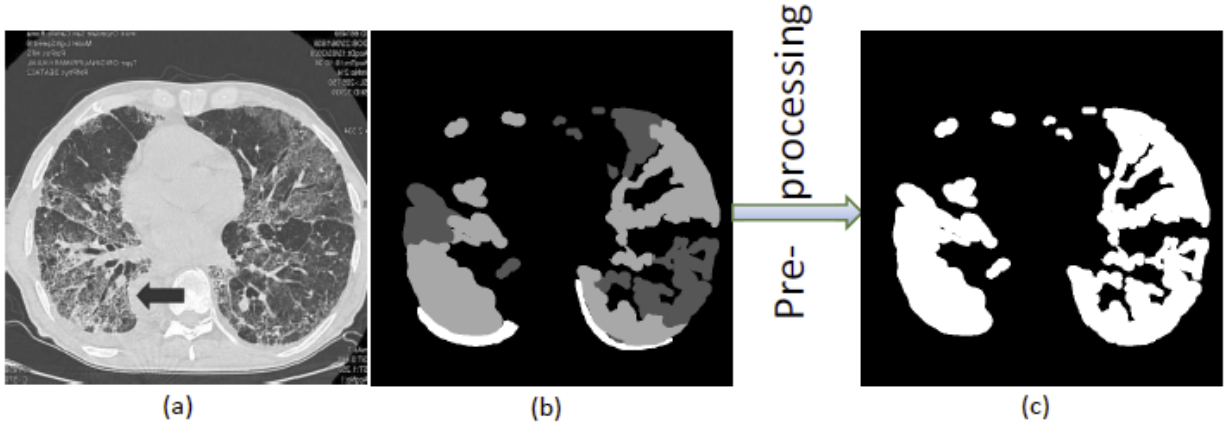

Figure 5.9: Sample CT from the MedSegData [48] (a) a CT slice, (b) the corresponding ground truth mask with three disease classes: pleural effusion (white), consolidation (light gray), GGOs (dark gray), (c) final ground truth mask with GGOs and consolidation in white

5.5 Implementation: Proposed CoviSegNet

As mentioned earlier, the proposed CoviSegNet is an Attention-based enhanced U-Net model which uses spatial-attention and channel-attention mechanisms to learn the hierarchical features of an enhanced U-Net. The encoder comprises an ImageNet pre-trained backbone, and the transfer learning paradigm is applied. This is a novel approach and no previously published research in COVID-19 segmentation has utilized the spatial-attention and channel-attention mechanism or the pre-trained DenseNet169 backbone.

The four dense blocks used in our model comprise 6, 12, 32, 32 layers as shown earlier in the architecture of CoviSegNet in Figure 34. We decided to use this transfer learning strategy since we have a limited amount of data which is insufficient to train a deep model from scratch. As expected, the model exhibited a superior performance with such a pre-trained backbone. The approach consisted of the below detailed five steps and was implemented in Python language using 'jupyter-notebook'.

First, we imported the various libraries such as tensorflow, numpy, sci-kit learn, Keras,

torch, and dpipe [134].

Second, pre-processing is done to prepare the datasets as explained in the pre-processing sub-section above. Third, the dataset is split into training, validation, and test datasets and we ensured that there is no same member's data overlap between the training, validation, and testing datasets. We use a ratio of 80:20 for training versus testing in all the experiments. Note that this is a held-out test dataset, and there is no patient's image data overlap between training and testing. During the process of training, the system auto-generated the validation dataset with 15% of images from training. The images from one patient are either all in training or validation or test. In other words, no overlap of data from a patient is allowed between the three splits.

Fourth, our novel proposed method, the Attention U-Net model is trained on the COVID-19 segmentation data. The 'cuda' (NVIDIA 11.1 cuda gpu) mode was used for faster training. The optimizer used was Adam. The loss function used was the binary cross-entropy with logits. The threshold chosen for the lung segmentation was 0.5 to generate the binary predicted mask. The predicted masks on the external test dataset are compared against the ground truth labels to report the dice score and Jaccard index. The attention-gates help to focus the model's attention on the learned hierarchical features and the final representations are passed through the final softmax classifier.

Lastly, the final softmax classifier uses the sigmoid activation function and generates the pixel-wise predicted mask of the COVID-19 pneumonia lesion regions. The 3x3 Median filtering was done to regularize the predicted output and then adaptive thresholding was done to generate the final predicted segmentation.

Note that for COVID-19 segmentation, the cross-validation hyperparameters were epoch

count of 200, number of batches per epoch of 100, and the number of samples per batch of 32. An early stopping criterion is used to stop training once the validation dataset loss did not reduce by at least 0.001 over the last five epochs.

5.6 Proposed CoviSegNet Enhanced: Our Improved Model for COVID-19 Segmentation

In the initial draft of the thesis, we first proposed the CoviSegNet, a COVID-19 diseased area (showing ground-glass opacities and consolidations) segmentation from lung CT of COVID-19 patients. After that, we improved the CoviSegNet using an enhancement to U-Net with EfficientNetB7 encoder backbone [99][100][101][102]. EfficientNetB7 is an ImageNet pre-trained encoder which learns the hierarchical features distinguishing COVID-19 from non-COVID-19 CT. EfficientNets [99] use a simple and highly effective and time-efficient compound scaling all the dimensions of network width, depth and resolution with a stable constant coefficient. There are B0 to B7 variants of this model; each successively higher model from B0 to B7 is more complex with a higher number of parameters. We chose the EfficientNetB7 since it achieved the highest performance. Supervised learning is used, and all the EfficientNetB7 layers' Image-Net pre-trained weights are kept frozen during the learning process. Additionally, we used the Nadam optimizer which Nesterov Accelerated Gradient (NAG) Momentum method [103]. This method applies the acceleration to the parameters before computing the gradients, and then updates with the gradients computed with the interim parameters. This helps to avoid the exploding gradients which occur in deep networks. Like CoviSegNet, our CoviSegNet Enhanced U-Net model also has spatial and channel attention [4], [38], [94], [111]. It also has horizontal skip connections at each depth. We also trained and tested the proposed CoviSegNet Enhanced using three public physician-annotated datasets with radiologist-labeled binary masks and the corresponding

computerized tomography (CT) images.

Our enhancements are: (1) First, to enhance CoviSegNet, our encoder comprises an ImageNet-pre-trained EfficientNetB7 as the encoder backbone which serves to learn the hierarchical features efficiently. (2) Second, we use an improved optimizer for training, namely, the Nadam optimizer.

5.7 Implementation of CoviSegNet Enhanced

We now explain the implementation details and architecture of CoviSegNet Enhanced. Our enhanced U-Net model has a highly complex encoder, the ImageNet-trained EfficientNetB7 with 813 layers with spatial and channel attention. No previously published research in COVID-19 segmentation utilized the EfficientNetB7 encoder backbone or the Nadam optimizer. The CoviSegNet Enhanced architecture is shown in Figure 5.10. The detailed architecture of the EfficientNetB7 encoder is shown in Figure 5.11. The Module 1, Module 2 and Module 3 from EfficientNetB7 are shown in Figure 5.12.

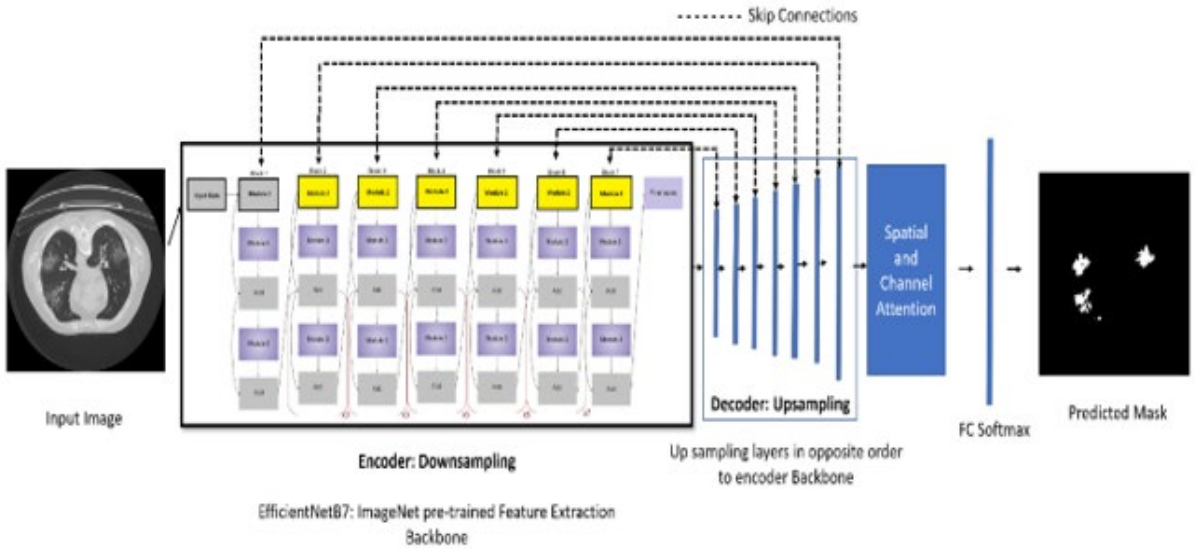

Figure 5.10: Architecture of CoviSegNet Enhanced classifier for Segmentation of COVID-19 diseased areas.

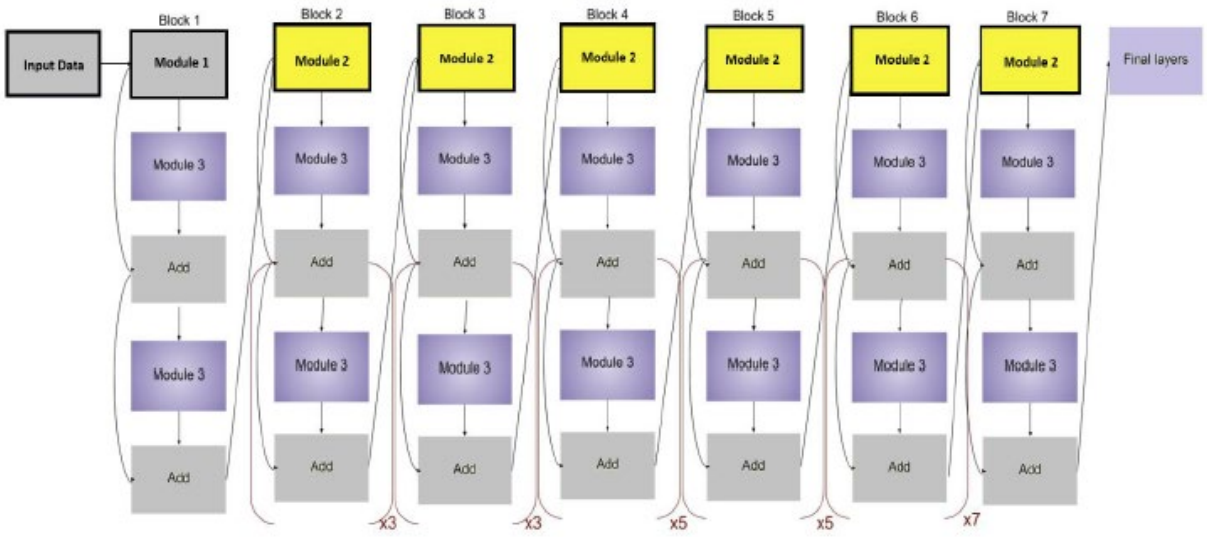

Figure 5.11: Architecture of EfficientNetB7.

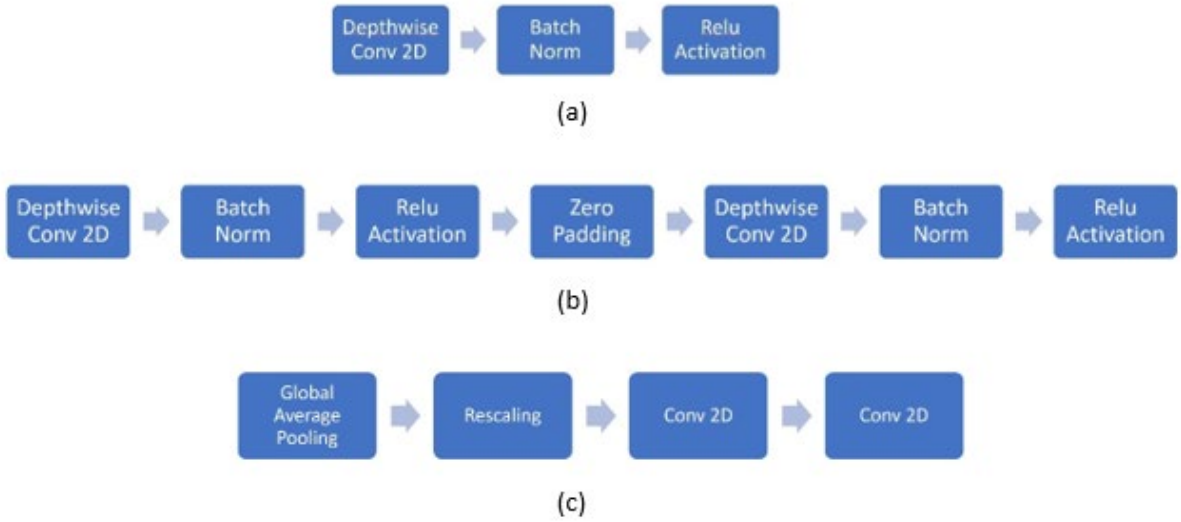

Figure 5.12: EfficientNetB7's component modules: (a) Module 1, (b) Module 2, and (c) Module 3

We will now discuss the implementation in detail. First, the various libraries including numpy, glob, 8tensorflow, and keras are imported [133] [134]. Then, the training and validation image data along with the physician-annotated binary masks are normalized to have pixel values in range (0, 255) and are resized to 256-by-256. The data involves multiple CT slices per patient, and the data is input in the form of 2-dimensional .jpg images and have three channels. We then split the data into train, validation and test in the ratio 68:12:20.

Second, the pre-processed input data is fed to an ImageNet-trained EfficientNetB7 encoder which is an 813-layer convolutional neural network that learns the hierarchical features efficiently. The Image-Net weights of EfficientNet are kept frozen for all the layers. The encoder is followed by the decoder which up-samples the data. Next, attention mechanisms in both spatial and channel dimensions are applied to the U-net by assigning attention gates on each up-sampling level. In our model, we use three stacks of up-sampling and down-sampling. The number of filters in the convolutional layers vary from 64, 128, 256, 512 and 1024. These attention gates use the decoded or up-sampled tensors as query and the encoded or down-sampled tensors as keys. Next, additive attention learning is used to get the self-attention values. The up-sampled tensor is concatenated with the attention gate output.

Third, the attention gate output is fed to the final softmax layer which outputs the prediction probability of each pixel being a COVID-19 diseased pixel. Finally, the prediction probability map is median filtered followed by adaptive thresholding to generate the binary predicted mask. The binary predictions on the held-out test dataset are compared against the physician-annotated masks to evaluate the model performance using the dice score and Jaccard Index / IOU.

We minimize the binary cross entropy loss between the binary predictions and the binary physician annotated labels during training. The Nadam optimizer's parameters are: a learning rate of 0.02, beta_1 of 0.9, beta_2 of 0.999, and an epsilon of $1x10^{-7}$. The segmentation model is trained for up to 200 epochs with early stopping once the model loss does not improve more than 0.001 over the next five epochs. The batch-size is 32 and each epoch has 100 batches. The CoviSegNet Enhanced's performance on various datasets is shown in Table 5.4

Table 5.4: CoviSegNet Enhanced's performance on MedSeg, MosMed and SemiSeg datasets.

Dataset - CoviSegNet Enhanced	Dice Score	IOU
MedSeg	98.4%	96.9%
MosMed	92.9%	86.7%
SemiSeg	81.8%	69.2%

Note that we tried using the EfficientNetB0 and finally chose the EfficientNetB7 as our enhanced model since EfficientNetB0's dice score was lower at 85.4% versus the EfficientNetB7 at 98.4%.

5.7.1 Comparison of CoviSegNet and CoviSegNet Enhanced:

The approach comparison between CoviSegNet and CoviSegNet Enhanced is shown in Table 5.5.

Table 5.5: Approach comparison between CoviSegNet and CoviSegNet Enhanced

	CoviSegNet	CoviSegNet Enhanced
Approach	DenseNet169 ImageNet pre-trained encoder-based U-Net model with spatial and channel attention	EfficientNetB7 [99] ImageNet pre-trained encoder-based U-Net model with spatial and channel attention
Optimizer	Adam	Nadam

The performance results comparison of CoviSegNet and CoviSegNet Enhanced on all three datasets, namely the MedSeg, MosMed and SemiSeg, are shown in Table 5.6.

Table 5.6: CoviSegNet's performance on MedSeg, MosMed and SemiSeg datasets.

	CoviSegNet		CoviSegNet Enhanced	
Dataset	Dice Score	IOU	Dice Score	IOU
MedSeg	93.2%	87.3%	98.4%	96.9%
MosMed	91.5%	84.3%	92.9%	86.7%
SemiSeg	82.3%	70.0%	81.8%	69.2%

5.8 Metrics

The metrics used for performance evaluation of the predicted segmentation masks are the dice score and the Intersection over Union (IoU). The Dice score is defined in Eq. 5.1:

$$\text{Dice Score} = \frac{2\times TP}{(TP+FP)+(TP+FN)} \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \quad \text{(Eq. 5.1)}$$

The Jaccard Index (also known as Intersection over Union, or IoU) is defined in Eq. 5.2:

$$\text{Jaccard Index / IoU} = \frac{TP}{(TP+FP+FN)} = \frac{Dice\ Score}{(2 - Dice\ Score)} \quad \text{(Eq. 5.2)}$$

where TP denotes the area of overlap between the COVID-19 positive ground truth and COVID-19 positive predicted mask, FN denotes the area of the COVID-19 positive ground truth which was missed in the predicted mask, and FP denotes the area of COVID-19 positive predicted mask which does not overlap with the ground truth.

5.8.1 Graphical Interpretation of Dice Score and IoU interpretation

The Dice Score is a measure of the average performance. It places equal emphasis on the False positives and False negatives. Figure 5.13 shows the dice score metric in graphical form. IOU is the worst-case performance. IoU indicates a score that is lower than the dice score. Figure 5.14 shows the Jaccard Index / IoU metric in graphical form.

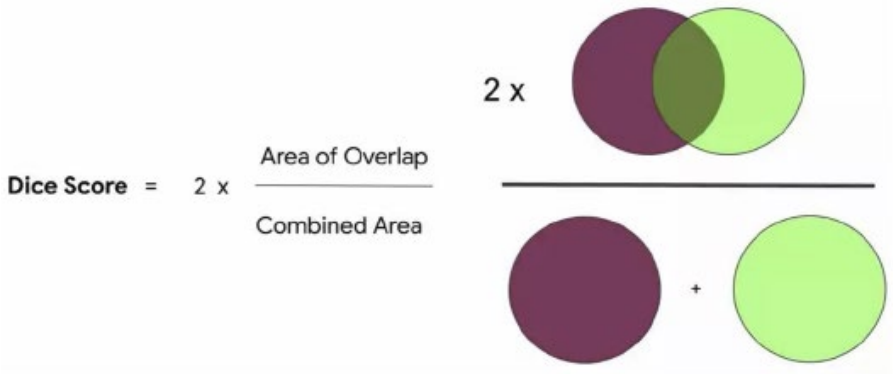

Figure 5.13: Dice Score metric in graphical form

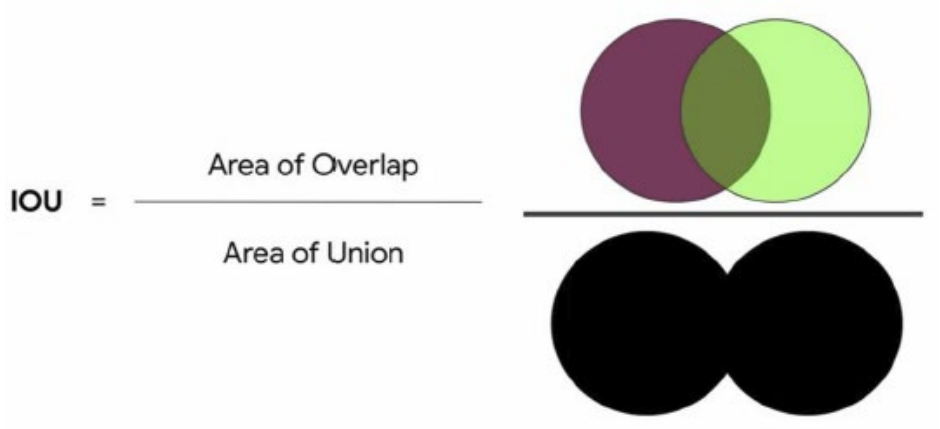

Figure 5.14: Jaccard Index / IoU metric in graphical form

5.9 Performance Evaluation of CoviSegNet, CoviSegNet Enhanced and Comparison with the Other Models

In this section, we assess the effectiveness of the proposed CoviSegNet, CoviSegNet Enhanced models in segmenting the COVID-19 diseased areas on three selected open-source CT datasets. All models were implemented in 'python' and 'jupyter-notebook' using various machine learning libraries including 'keras', 'sci-kit learn', 'nibabel' and 'matplotlib' [133] [134] [135] [136] [137] [138]. The machine used was Intel(R) Xeon(R) W-10885M CPU @ 2.40GHz, 2400 MHz, 8 Core(s), 16 Logical Processor(s), NVIDIA Quadro RTX 5000 with 128GB RAM.

Three different public datasets were used for the evaluation of CoviSegNet's and CoviSegNet Enhanced's dice score and IoU on held-out test data. We compared our proposed model's results with recently published work which also utilized the same dataset for a fair comparison.

5.9.1 Dataset 1 – MedSegData [48]:

Data was split such that a given patient's CT data was either all in training, or validation or test. Dataset comprises CT volumes of 52 patients with 450 image slices. The training set had

342 slices from 37 patients, the validation set had 39 slices from 5 patients, and the testing set had 69 slices from 10 patients. As shown in Table 5.7, our proposed CoviSegNet and CoviSegNet Enhanced models achieved the highest reported dice score of 93.2% and 98.4% respectively in comparison with recently published work.

Table 5.7: Metrics for the proposed CoviSegNet, CoviSegNet Enhanced and comparable recent research using MedSegData [48].

Model	Dice Score	Jaccard Index / IoU
Multitask [31] *	88.0%	78.6%
Dilated FCN [32]*	85.1%	74.1%
TV-UNet [41]*	76.4%	61.8%
CoviSegNet (Ours)	93.2%	87.3%
CoviSegNet Enhanced (Ours)	98.4%	96.9%

*Note that Multitask [31], Dilated FCN [32] and TV-UNet [41] results are just being mentioned as is from their published work. References [31] and [41] used the same MedSegData [48] that we used also, so their results are a good performance comparison baseline for our model. Although [32] used another dataset, namely, the Interstitial Lung Disease CT imaging dataset, we still believe it is a comparable work since it also segments the ground-glass opacities and consolidations as also done in our work.

5.9.2 Dataset 2 – MosMedData [49]:

Data was split such that a given patient's CT data was either all in training, or validation or test. The dataset comprises 50 patients' CT volumes with a total of 2049 CT slices. The training set had 1476 slices from 35 patients, the validation set had 164 slices from 5 patients, and the testing set had 409 slices from 10 patients. Our proposed CoviSegNet and CoviSegNet Enhanced models achieved the highest reported dice score of 91.5% and 92.9% respectively, which is higher

than Goncharov et al's [27] dice score of 65.0% on the MosMedData. These results are shown in Table 5.8.

Table 5.8: Comparison of metrics for the proposed CoviSegNet, CoviSegNet Enhanced with Goncharov [27] using MosMedData [49].

Model	Dice Score	Jaccard Index / IoU
Goncharov et al. 3D U-Net [27]*	65.0%	48.1%
CoviSegNet (Proposed)	91.5%	84.3%
CoviSegNet Enhanced (Proposed)	92.9%	86.7%

*Note that Goncharov et al.'s [27] results are just being mentioned as is from their published work since they used the same MosMedData [49].

5.9.3 Dataset 3 – SemiSeg dataset [50]:

Data was split such that a given patient's CT data was either all in training, or validation or test. The dataset includes the MedSegData [48] with 100 slices and labeled masks from 43 patients. It also includes 1600 CT images from the COVID-19 CT Collection dataset by Cohen et al. [56] of 45 patients with corresponding masks created via semi-supervised learning methods by Fan et al. [50]. The training set had 1190 slices from 62 patients, the validation set had 170 slices from 9 patients, and the testing set had 340 slices from 17 patients.

Table 5.9: Metrics for the proposed CoviSegNet, CoviSegNet Enhanced and comparable recent research using Covid SemiSeg dataset [50].

Model	Dice Score	Jaccard Index / IoU
U-Net++ [51]*	51.8%	35.0%
Inf-Net [4]*	68.2%	51.7%
Semi-Inf-Net [4]*	73.9%	58.6%
Semi-Inf-Net +FCN8s [4]*	47.4%	31.1%
TV-UNet [41]*	80.1%	68.8%
CoviSegNet (Proposed)	82.3%	70.0%
CoviSegNet Enhanced (Proposed)	81.8%	69.2%

*Note that U-Net++ [51], Inf-Net [4], Semi-Inf-Net [4], Semi-Inf-Net +FCN8s [4], and TV-UNet [41] results are just being reported as is from their published work since their work also utilized the same dataset, with the same training, and test split as provided in the SemiSeg dataset [50].

Our proposed CoviSegNet and CoviSegNet Enhanced models achieved the highest reported dice score of 82.3%, 81.8% as shown in Table 5.9. This is higher than the next best dice score of TV-UNet [41], and the other comparison studies recently published as shown in Table 25.

The predicted Segmentation results from the held-out test datasets are shown along with the CT image and its corresponding ground truth in Figure 5.15.

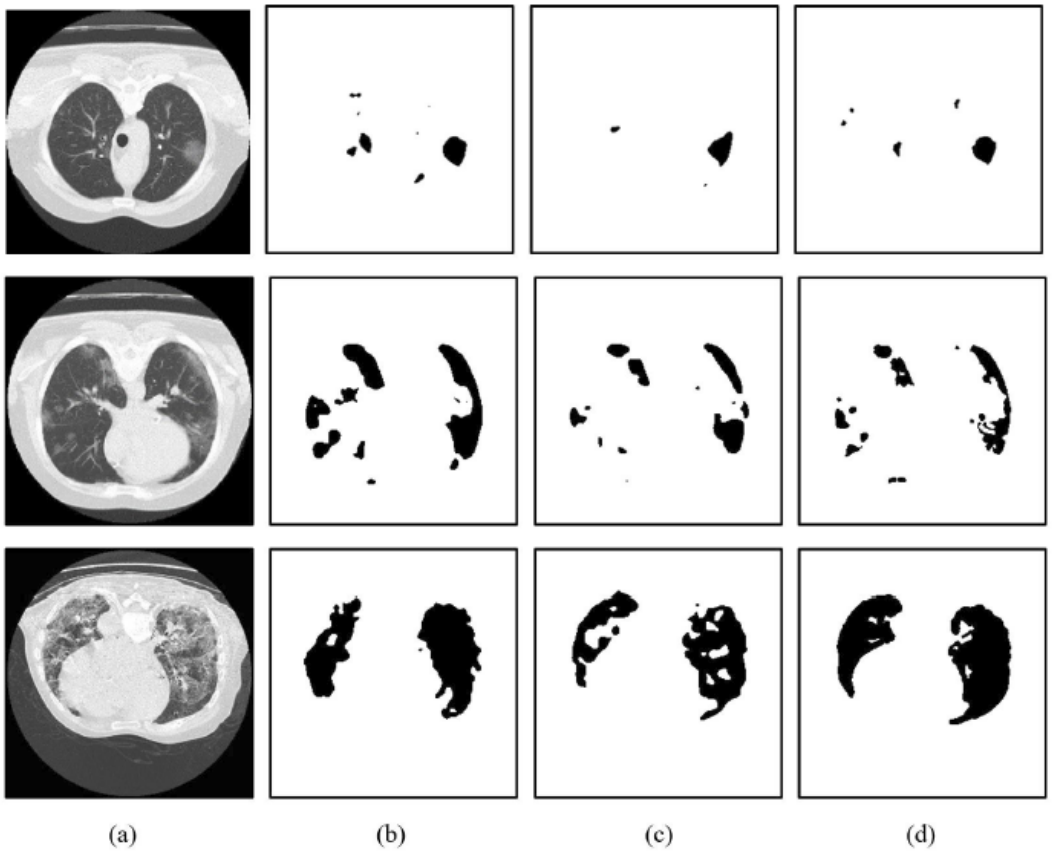

Figure 5.15: Segmentation Result: (a) CT image, (b) CoviSegNet Prediction, (c) CoviSegNet Enhanced Prediction, (d) Ground Truth mask with positive class (GGOs and consolidation) in black.

The CoviSegNet performance on the Covid SemiSeg dataset [50] is considerably lower than that on other datasets. We attribute this lower performance to two reasons. First, the Covid SemiSeg dataset [50] has machine annotated labels which are likely less accurate and have more noise than physician annotated labels. Second, the images Covid SemiSeg dataset [50] are of lower resolution at 96 dpi, varying sizes from 250×250 to 400×400, and 8 bit versus the MosMed which are of resolution 96 dpi, 512×512, and 24 bit and MedSeg which have a resolution of 96 dpi, 630×630, and 24 bit.

On all the experiments shown in Tables 14, 15, and 16, our Attention U-Net model is the highest performing as measured by the dice score and the IoU. The ImageNet pre-trained backbone, DenseNet169 showed a much better dice score and IoU than the VGG16 for our Attention U-Net model. This is because the dense connections substantially reduce the number of parameters and so the model can train better without overfitting even with the relatively small datasets available in this research. In comparison, the VGG16 architecture backbone was also notably slower to train. Additionally, the spatial and channel attention helps the model to learn from the sparse data efficiently and effectively.

5.9.4 Comparison Study: CoviSegNet Enhanced versus Safarov et al. [83]

Table 5.10 shows the comparison of CoviSegNet Enhanced with Safarov et al's colonoscopy polyp segmentation [83].

Table 5.10: Comparison of CoviSegNet Enhanced with Safarov et al [83]

Detail	CoviSegNet Enhanced	Safarov et al. [83]
Dataset	Publicly available, physician annotated lung CT datasets: MedSeg, MosMed, and CoviSemiSeg	Two public colonoscopy datasets with physician-annotated binary masks: 1000 polyp images, and 612 polyp images

Detail	CoviSegNet Enhanced	Safarov et al. [83]
Dice Score	Dice Score (On MedSeg): 98.4%	We implemented Safarov et al.'s model on MedSeg data. Dice Score (On MedSeg): 93.5%
CNN Structure	ImageNet pre-trained EfficientNetB7 encoder-based attention U-Net model with spatial and channel attention for COVID-19 diseased area segmentation in lung CT. 813-layer EfficientNetB7 with U-Net backbone with attention.	DenseNet-based attention U-Net++ model for segmenting polyps. 164-layer DenseNet as the U-Net++ backbone with attention.

5.9.5 Comparison Study: CoviSegNet Enhanced versus TV-UNet [41]

In Table 5.11, we compare the EfficientNetB7-based Enhanced attention U-Net with Saeedizadeh et al.'s TV-UNet [41].

Table 5.11: Comparison of CoviSegNet Enhanced with Saeedizadeh et al.'s TV-UNet [41].

Detail	CoviSegNet Enhanced	TV-UNet [41]
Dataset	Publicly available, physician annotated lung CT datasets: MedSeg, MosMed, and CoviSemiSeg	Publicly available, physician annotated lung CT datasets: MedSeg, and CoviSemiSeg
Dice Score	Dice Score: 98.4% (on MedSeg), 81.8% (on Semi-Seg; machine-labeled)	Dice Score: 76.4% (on MedSeg) 80.1% (on Semi-Seg; machine labeled)
CNN Structure	ImageNet pre-trained EfficientNetB7 encoder-based attention U-Net model with spatial and channel attention for COVID-19 diseased area segmentation in lung CT. 813-layer EfficientNetB7 with U-Net backbone with attention. Nadam Optimizer.	27-layer U-Net with total variation (TV) as loss function. Adam optimizer.

5.9.6 Comparison Study: CoviSegNet Enhanced versus Goncharov et al. [27]

Table 5.12 shows the comparison of CoviSegNet Enhanced (EfficientNet-based attention U-Net) with Goncharov et al. [27] for COVID-19 segmentation.

Table 5.12: Comparison of CoviSegNet Enhanced with Goncharov et al. [27]

Detail	CoviSegNet Enhanced	Goncharov et al. [27]
Dataset	Publicly available, physician annotated lung CT datasets: MedSeg, MosMed (50 CT volumes), CoviSemiSeg	Publicly available, physician annotated lung CT dataset: MosMedData[49](50 CT volumes)
Dice Score	Dice Score: 92.9% (on MosMed)	Dice Score: 65.0% (on MosMed)
CNN Structure	ImageNet pre-trained EfficientNetB7 encoder-based attention U-Net model with spatial and channel attention for COVID-19 diseased area segmentation in lung CT. 813-layer EfficientNetB7 with U-Net backbone with attention. Nadam Optimizer.	Training from scratch Basic U-Net Ronneberger et al. [105] architecture. 2D U-Net, and 3D U-Net Stochastic Gradient Descent optimizer.

CHAPTER 6

ABLATION STUDIES*

6.1 In-depth Ablation Studies

6.1.1 Ablation Study 1: CoviNet Enhanced – 3D versus 2D CNN

Table 6.1 shows the CoviNet Enhanced's performance on UCSD-AI4H on 3D CNN versus 2D CNN.

Table 6.1: CoviNet Enhanced's performance on UCSD-AI4H on 3D CNN versus 2D CNN

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score	MCC
2D CNN	77.10%	76.67%	73.64%	80.13%	0.7498	0.5409
3D CNN CoviNet Enhanced	92.2%	93.0%	93.0%	91.2%	0.930	0.842

6.1.2 Ablation Study 2: CoviSegNet and CoviSegNet Enhanced With and Without Original Backbone

We compare the performance of CoviSegNet and CoviSegNet Enhanced model for scenarios with the best performing backbone and the next best performing backbone.

1. CoviSegNet With DenseNet – This is the ImageNet pre-trained DenseNet169 as the encoder backbone. This is the original CoviSegNet model whose implementation was discussed in Section 5.6.1.

2. CoviSegNet Without DenseNet – In this scenario, we replace the DenseNet169 in the CoviSegNet model with ImageNet pre-trained VGG16.

3. CoviSegNet Enhanced With EfficientNetB7 – This uses the ImageNet pre-trained EfficientNetB7 as the encoder backbone. This is the original CoviSegNet Enhanced model whose implementation was discussed in Section 5.6.2.

* Sections 6.1 & 6.2 (Ablation Studies 2 & 3) are reproduced from Mittal, B. & Oh, J. CoviSegNet: Covid-19 Disease Area Segmentation using Machine Learning Analyses for Lung Imaging. ISPA 2021, 13-15 September 2021, Zagreb, Croatia [accepted for publication], with permission from IEEE.

4. CoviSegNet Enhanced Without EfficientNetB7 – In this scenario, we replace the EfficientNetB7 in the CoviSegNet Enhanced model with ImageNet pre-trained VGG16.

Table 6.2 shows CoviSegNet's performance on MedSegData [48] With DenseNet and Without DenseNet and CoviSegNet Enhanced's performance on MedSegData [48] With EfficientNetB7 and Without EfficientNetB7.

Table 6.2: CoviSegNet's and CoviSegNet Enhanced's performance on MedSegData [48] With Original and Different backbone.

	CoviSegNet (Original = DenseNet169) (Different = VGG16)		**CoviSegNet Enhanced (Original = EfficientNetB7) (Different = VGG16)**	
Model	Dice Score	Jaccard Index / IoU	Dice Score	Jaccard Index / IoU
Original Backbone	93.2%	87.3%	98.4%	96.9%
Different Backbone	94.4%	89.3%	94.1%	88.8%

6.1.3 Ablation Study 3: With and Without Attention

We compare the the performance of 'CoviSegNet' and 'CoviSegNet Enhanced' segmentation models on MedSegData [48] for two scenarios: "With Attention" i.e. with spatial and channel attention and "Without Attention" i.e. by removing the spatial and channel attention mechanisms. For the "Without Attention" scenario, only the attention was removed, and everything else including the backbones remained unchanged. Table 6.3 shows the performance of CoviSegNet and CoviSegNet Enhanced on MedSegData [48] "With Attention" and "Without Attention."

Table 6.3: CoviSegNet's and CoviSegNet Enhanced's performance with and without attention.

Model	**Dice Score**	**Jaccard Index / IoU**
CoviSegNet With Attention	93.2%	87.3%
CoviSegNet Without Attention	64.5%	47.6%
CoviSegNet Enhanced With Attention	98.4%	96.9%
CoviSegNet Enhanced Without Attention	71.8%	56.0%

These results demonstrate that attention helps both CoviSegNet and CoviSegNet models exhibit a much higher dice score, specifically an absolute increase of 28.7% and 26.6% in dice score respectively, than without attention.

6.2 MosMed on CoviNet Enhanced: All Patient Data

The CT volumes of 254 patients with normal lung tissue and the CT volumes of 856 patients with COVID-19 positive lung tissue are utilized. Number of samples in train, validation and test are 755, 133 and 223. The data was split proportion was 68% : 12% : 20% for the training, validation and testing sets. The data split is shown in Table 6.4. This dataset is referred to as MosMed 1110 to distinguish it from MosMed presented earlier.

Table 6.4: 1110 patients' MosMed data details.

	Train	Validation	Test	Total Patients
Overall	755 patients	132 patients	223 patients	**1110**
COVID-19 Positive	582 patients	102 patients	172 patients	**856**
COVID-19 Negative	173 patients	30 patients	51 patients	**254**

Table 6.5 shows results on the 1110 patients' MosMedData CT volumes for CoviNet Enhanced. The performance measure is on the held-out test dataset.

Table 6.5: CoviNet Enhanced's and Wang et al.'s DeCovNet [92] performance on 1110 patients' MosMed data.

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score	MCC
CoviNet Enhanced	80.8%	75.7%	79.1%	82.0%	0.774	0.608
DeCovNet [92]	66.8%	74.1%	58.3%	76.7%	0.652	0.353

6.3 CoviNet Enhanced Ablation Experiment on UCSD-AI4H with CT Volumes and With Just One Slice Per Patient

Table 6.6 shows CoviNet Enhanced's performance on UCSD-AI4H with CT volumes and

with just individual CT slices.

Table 6.6: CoviNet Enhanced's performance for UCSD-AI4H on CT Volumes versus single slice.

Metric	CT Volumes	CT Slices
Accuracy	92.2%	72.1%
Precision	93.0%	100.0%
Sensitivity	93.0%	60.0%
Specificity	91.2%	100.0%
F1-score	0.930	0.750
MCC	0.842	0.559

6.4 Classification Experiment with 3D Max Pooling Instead of 2D max Pooling

CoviNet Enhanced is a 3D CNN model and uses 3D max pooling layers, 3D batch normalization layers and 3D convolutional layers only. The 2D max pooling was used in the 2D CNN, and the 3D max pooling was used in the 3D CNN (CoviNet Enhanced). Table 6.7 shows 2D max pooling on 2D CNN versus 3D max pooling on CoviNet Enhanced.

Table 6.7: 2D max pooling via 2D CNN versus CoviNet Enhanced's 3D max pooling

Model	Accuracy	Precision	Sensitivity	Specificity	F1-score	MCC
2D max pooling (2D CNN)	77.1%	76.7%	73.6%	80.1%	0.750	0.541
3D max pooling (CoviNet Enhanced)	92.2%	93.0%	93.0%	91.2%	0.930	0.842

6.5 Statistical Analysis (Confidence Interval) on 5 folds for CoviNet Enhanced

Table 6.8 shows the statistical analysis (confidence interval) on 5 folds for CoviNet Enhanced on UCSD-AI4H dataset.

Table 6.9 shows the statistical analysis (confidence interval) on 5 folds for CoviNet Enhanced on MosMed dataset.

Table 6.8: Statistical Analysis on CoviNet Enhanced's performance on UCSD-AI4H dataset.

Statistic	Accuracy	Precision	Sensitivity	Specificity	F1-score	MCC
Mean	92.2%	93.0%	93.0%	91.2%	0.930	0.842
95% Confidence Interval	+/-2.51%	+/-9.88%	+/-7.53%	+/-13.57%	+/-0.02	0.07

Table 6.9: Statistical Analysis on CoviNet Enhanced's performance on MosMed dataset.

Statistic	Accuracy	Precision	Sensitivity	Specificity	F1-score	MCC
Mean	96.4%	94.3%	97.1%	95.9%	0.957	0.926
95% Confidence Interval	+/-1.0%	+/-1.2%	+/-0.5%	+/-1.1%	+/-0.01	+/-0.01

6.6 Image-Level and Patient-Level Performance Metrics

Note that the CoviNet Enhanced model takes only 3D CT volumes. It processes the entire 3D volume together and makes predictions at volume only. To get image-level results, we provided input as a single CT slice duplicated 5 times to make volume. Table 6.10 shows the image-level and patient-level metrics of the CoviNet Enhanced model on the UCSD-AI4H dataset.

Table 6.10: Image-level and patient-level metrics of the CoviNet Enhanced

CoviNet Enhanced	Accuracy	Precision	Sensitivity	Specificity	F1-score	MCC
Patient Level	92.2%	93.0%	93.0%	91.2%	0.930	0.842
Image Level	72.1%	100.0%	60.0%	100.0%	0.750	0.559

Table 6.11 shows the CoviSegNet Enhanced's performance which is possible at image-level only since we are segmenting images.

Table 6.11: CoviSegNet's performance at image-level for COVID-19 segmentation.

Model	Dice Score	Jaccard Index / IoU
CoviSegNet Enhanced With Attention	98.4%	96.9%

CHAPTER 7

CONCLUSIONS AND FUTURE DIRECTIONS

7.1 Conclusions

While this thesis explores the problems of disease diagnosis and segmentation, we can observe the general conclusion that end-to-end deep supervised learning is an excellent paradigm in automating the lung CT analysis for COVID-19 disease diagnosis and severity quantification via segmentation. Such end-to-end learning outperforms the hand-engineered approaches and produces superior performance results on the held-out test datasets. The transfer learning-based approaches perform well if only a small dataset is available, and the learning from scratch works better in case a large, labeled dataset is available.

7.1.1 Classification

In our work, deep supervised-learning-based approaches for COVID-19 diagnosis on CT image datasets have proven successful and achieved excellent results. This research takes a step closer towards the goal of minimizing human error while increasing radiologists' productivity by automating the image analysis. The authors are highly thankful for the large publicly available labeled datasets which made this research feasible [23], [24]. For the worldwide adoption of such diagnosis techniques. Training from scratch via 3D Convolutional Neural Network utilizing CT volumes with up to 64 slices per patient showed a superior generalizability than doing transfer learning via deep ImageNet pre-trained models like ResNet-50. CoviNet and showed a higher F1-score, sensitivity, and Matthew's correlation coefficient than those of prior published works.

CoviNet showed a higher MCC, sensitivity and specificity score than those in prior published works. On the MosMedData, CoviNet shows good Sensitivity of 97.06% and Specificity

of 92.16%. CoviNet has both a very high Sensitivity and NPV, so it is well suited in the field of medical diagnosis which is intolerant to Type 2 errors. It has a Positive Predictive Value (Precision) of 89.19% and a Negative Predictive Value of 97.92%. CoviNet is an excellent model to rule-out disease if a patient tests as negative and is most suited for populations in which there is a low to moderate probability of disease.

We can conclude based on our research that COVID-19 diagnosis using artificial intelligence on Lung CT is efficient, accurate, highly sensitive in detecting early pneumonic changes, and helps differentiate COVID-19 from other viral pneumonia. This automatic diagnosis brings positive patient outcomes via early detection and patient isolation prevents disease spread and helps save radiologists' time and effort. Thus, supervised learning-based deep-learning approaches are effective in automating the image analysis for COVID-19 detection via CT images.

7.1.2 Segmentation

The diseased areas of the lungs having ground-glass opacities and consolidations can be quantified with excellent performance using enhanced U-Net models with spatial and channel attention with a pre-trained encoder backbone. Additionally, our COVID-19 segmentation using deep-learning-based approaches on CT image datasets produced successful results with a good dice score and IoU metric in comparison with recent studies. We even performed external dataset validation and our model generalized well and showed good performance.

CoviSegNet's deep-learning-based approach on CT image datasets produced successful results with a dice score and Jaccard index / IoU than those of prior published works. We performed external dataset validation and our model generalized well and showed good

performance. It was five times faster than training a model from scratch.

7.2 Future Directions

7.2.1 Classification

For future work, semi-supervised learning methods could be leveraged to enhance model performance. Additionally, if available, a larger dataset with high-quality images having several CT-slices per patient should be leveraged. More complex problems of quantifying disease severity and distinguishing non-viral or bacterial pneumonia from COVID-19 also merit further research. This paper propels research towards developing a fully automated and efficient system for diagnosing COVID-19 and all lung diseases. It would greatly help to be able to run the experiments on multiple machines using some experimental frameworks such as Mesh Tensorflow.

7.2.2 Segmentation

For future work, this model performance can be further increased by leveraging semi-supervised learning methods since we have a relatively small, labeled dataset. Additional work would also be needed to compute the severity scoring based on percentage lung involvement in each of five lung lobes for CT volumes based on medical guidelines [58], [61]. In the future, even after the pandemic is over, such techniques could be applied to all lung scans done for any reason to diagnose and prognosticate all respiratory illnesses including new viruses that may arise in the future [39].

Tversky loss: The Dice metric places equal importance on False Positives and False Negatives. However, in medical imaging, minimizing the Type II errors or False Negatives is more critical than the False Positives. In segmentation, this translates to caring more about not missing

to diagnose a diseased area, even at the cost of more normal areas misdiagnosed as diseased. Thus, for future work on COVID-19 segmentation, an even better metric than dice score would be the Tversky similarity index [114] which is a generalization of the dice score, which allows for flexibility in weighting the False positives and False Negatives.

APPENDIX

SUPPLEMENTAL MATERIALS

A.1 Other Publications

1. Tejaswini S.V.L.L., Mittal B., Oh J., Tavanapong W., Wong J., de Groen P.C., Enhanced Approach for Classification of Ulcerative Colitis Severity in Colonoscopy Videos Using CNN, Advances in Visual Computing. ISVC 2019. https://doi.org/10.1007/978-3-030-33723-0_3, 2019, pp. Accepted.

2. Tavanapong W., Oh J., Khaleel M., Mittal B., de Groen P.C., Past, Present, and Future: Artificial Intelligence for Colonoscopy: A Survey, TBD, to be Submitted in 2021.

A.2 Publications from This Thesis

1. Mittal B., Oh J., CoviNet: COVID-19 Diagnosis using Machine Learning Analyses for Computerized Tomography Images, in ICDIP 2021, May 2021.

2. Mittal B., Oh J., CoviSegNet - COVID-19 Disease Area Segmentation using Machine Learning Analyses for Lung Imaging, in ISPA 2021, Submitted, May 2021.

3. Mittal B., Oh J., CoviSegNet - Ablation Studies on COVID-19 Classification and Segmentation using Machine Learning Analyses for Lung Imaging, in ISPA 2021, To be submitted by June 25, 2021.

A.3 Challenges Overcome in this Dissertation Work:

This project was a significant undertaking and the various issues that were overcome to complete this project successfully are listed below:

1. We need to ensure that all the libraries being used are compatible and if any coding issues arise due to versioning, we need to resolve them by either updating the libraries installed or updating the coding to match the library version being used.

2. Another challenge was coming up with an idea to convert single CT images into volumes since the 3D CNN used for classification only uses CT volumes.

3. It was also a learning curve to get familiar with how to process the .nii CT volumes and I identified that the MIPAV tool [96] was extremely helpful in visualizing the data and converting

from images to volumes and vice versa. The NIfTI-Image-Converter [77] was another tool used for converting the .nii files to .png files.

4. Also, we found that running the code in "GPU" mode with the NVIDIA CUDA 11.1 installed speeded up the processing speed by nearly 8 times.

5. The literature review was very vast due to many research publications in COVID-19 diagnosis and severity quantification using AI techniques. From the initial screening of 4000 papers, I finally reviewed and read around 200 papers a few times to develop a grasp of this entirely new subject domain.

6. Next, a challenge was to find and select good quality datasets that were used in prior works from the more than 50 different public datasets available. The datasets comprised X-rays, ultrasound, and CT images, and we brainstormed a lot, reviewed with an expert radiologist, performed a lot of literature review, and then finally chose the four computerized tomography datasets for this research.

7. Finally, countless experiments and techniques had to be implemented before we could arrive at our best-performing models. It was truly labor-intensive and time-intensive work to run the various machine learning algorithms with different datasets and pre-processing techniques.

A.4 Computerized Tomography (CT): Technology Advances and Evolution

Computerized tomography (CT) imaging or "CAT scanning" (Computerized Axial Tomography) produces cross-sectional images of the anatomy with multiple slices like the slices in a loaf of bread. The word tomography comprises two Greek words: "tomos" which means slice and "graphe" which means drawing. The cross-sectional images as shown in Figure A.1 help in disease diagnosis, triage, and prognosis [119].

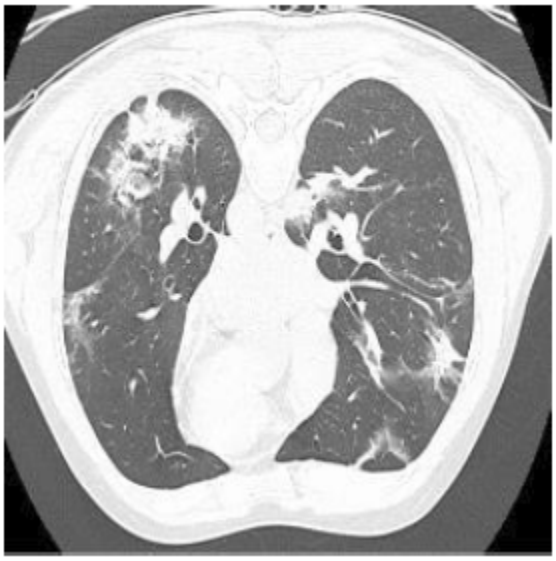

Figure A.1: Axial CT Image of Lung

In a CT System, the patient lays in a supine position on an automated table and is moved through a circular opening in the CT imaging system as shown in Figure A.2.

Figure A.2: Patient being scanned in CT machine.

The source of x-rays rotates around the inside of the circular opening. A narrow, fan-shaped beam of x-rays about 1-10 millimeter thick irradiates a section of the patient's body as shown in Figure A.3. A CT exam will have several phases of 10 to 50 rotations each. Detectors

collect several ‘snapshots’ of the radiation exiting the part of the patient’s body being irradiated during each rotation.

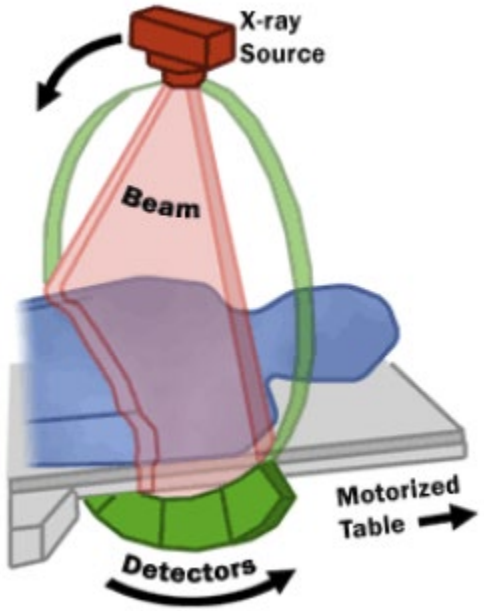

Figure A.3: CT Fan Beam

A computer then reconstructs the individual "snapshots" into one cross-sectional image (slice) of the internal anatomy and tissues for each complete rotation of the x-ray source. A "contrast material" might be injected into the patient’s body prior to doing the scan for better visualization [119].

Today’s advanced CT systems can produce not just "axial" but “spiral” scans as well. With spiral CT, the same examination with the same slice thickness as axial CT can be completed with 50% greater table speed using a pitch of 1.5; this results in a 33% reduction in patient dose [121]. Multiple slices are imaged simultaneously at higher resolution with a lower radiation dose and faster time. The Electron Beam CT (EBCT) does not have a moving part involved in generating the individual "snapshots" which allows for a quicker image acquisition than conventional CT scanners [119].

The first CT scanner was invented in 1967, by Sir Godfrey Hounsfield at EMI Central Research Laboratories using x-ray technology. Several hundred million CT scans are now done annually and comprise 64-320 slices with slice widths ranging from 1 to 5 mm [120]. CT is starting to replace radiography as the initial diagnostic exam and its use is quite widespread. However, the effective dose for a chest CT is approximately 100–1000 times larger than that for a corresponding chest x-ray examination [123]. This higher radiation exposure from CT is harmful to body tissues. Thus, the Food and Drug Administration (FDA) has brought awareness to reduce unnecessary radiation exposure from medical imaging [118], [120]. Notably, image reconstruction techniques have contributed to 70% - 80% radiation dose reduction over the past 4 years [125].

Recent advancements are faster speed of scanning taking a few seconds versus 30 minutes, lower CT patient radiation dose, and an improved image quality. A new generation CT scanner called the 'coronary CT angiogram' scans the heart and coronary arteries in less than one second, with a high specificity and high sensitivity for coronary artery disease. CT perfusion is a disruptive technology used to diagnose ischemic stroke and cancerous tumors. Fusion of functional imaging such as Positron Emission Tomography (PET) with CT is a great tool that merges data from two different types of exams and is highly useful in cancer treatment [122]. 3D CT imaging and Intensity-Modulated Radiotherapy (IMRT) helps in stereotactic radiosurgery and for radiation targeting around sensitive structures or organs. Additionally, low radiation dose CT has enabled physicians to increase the quality of pediatric CT examinations, especially for cervical spine procedures and chest CTs. Further, 3D printing of digital CT datasets helps to improve medical care to assist in surgical planning and prototyping implants. In summary, CT

enables rapid and accurate diagnosis in emergency departments for trauma, acute chest pain, or other urgent conditions [120].

Computer vision and deep learning are remarkably promising technologies and numerous tools are available and being developed to benefit the area of computerized tomography. CT scans are a treasure trove of information that remains untapped, and the goal is to extract more information from CT scans with lower radiation exposure [124].

A.3 Acronyms

2D CNN	Two-dimensional Convolutional Neural Network
AE	Auto-encoder
AFS-DF	Adaptive Feature Selection guided Deep Forest
BigBiGAN	Large-scale bi-directional Generative Adversarial Network
CoviNet	The model for COVID-19 diagnosis implemented in this dissertation
CoviSegNet	The model for COVID-19 segmentation implemented in this dissertation
CPU	Central Processing Unit
CT	Computerized Tomography
CXR	Chest X-ray
CNN	Convolutional Neural Network
CUDA	A deep learning framework that allows for Graphical Processing Unit processing and frequently used in machine learning work
DBN	Deep Belief Networks
DL	Deep Learning
DNN	Deep Neural Network
F-score	Harmonic mean of the model's precision and recall
GAN	Generative Adversarial Network
GGO	Ground Glass Opacity
GPU	Graphical Processing Unit
IAVP	Influenza-A Viral Pneumonia
jpg / .jpeg	Commonly used method of lossy compression for digital images. An image file format was standardized by the Joint Photographic Experts Group.
KNN	K Nearest Neighbors

LR	Logistic Regression
ML	Machine Learning
MCC	Matthew's Correlation Coefficient
.nii	Image file format that stores image volumes frequently used in 3D medical imaging such as Computerized Tomography
NMT	Neural Machine Translation
NVIDIA	A company that makes Graphical Processing Units
.png	An image file format. Stands for Portable Networks Graphics
RNN	Recurrent Neural Network
TV-UNet	Total Variation U-shaped Network
SVM	Support Vector Machine
VGG	Visual Geometry Group

REFERENCES

[1] World Health Organization, “Weekly epidemiological update on COVID-19 – 15 June 2021,” Jun. 2021.

[2] Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, Kanne JP, Raoof S, et al. The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. Radiology 2020 Jul;296 (1):172-180

[3] Bai, H. X., Hsieh, B., Xiong, Z., Halsey, K., Choi, J. W., Tran, T. M. L., Pan, I., Shi, L.-B., Wang, D.-C., Mei, J., Jiang, X.-L., Zeng, Q.-H., Egglin, T. K., Hu, P.-F., Agarwal, S., Xie, F.-F., Li, S., Healey, T., Atalay, M. K., & Liao, W.-H. (2020). Performance of Radiologists in Differentiating COVID-19 from non-COVID-19 Viral Pneumonia at Chest CT. Radiology, 296(2), E46–E54. https://doi.org/10.1148/radiol.2020200823

[4] D.-P. Fan, T. Zhou, G.-P. Ji, Y. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, “Inf-Net: Automatic COVID-19 Lung Infection Segmentation From CT Images,” IEEE Transactions on Medical Imaging, vol. 39, no. 8, pp. 2626–2637, Aug. 2020, number: 8.

[5] D. Al-Karawi, S. Al-Zaidi, N. Polus, and S. Jassim, “Machine Learning Analysis of Chest CT Scan Images as a Complementary Digital Test of Coronavirus (COVID-19) Patients,” Radiology and Imaging, preprint, Apr. 2020.

[6] A. A. Ardakani, U. R. Acharya, S. Habibollahi, and A. Mohammadi, “Covidiag: a clinical CAD system to diagnose COVID-19 pneumonia based on CT findings,” European Radiology, vol. 31, no. 1, pp. 121–130, Jan. 2021, number: 1.

[7] J. Bridge, Y. Meng, Y. Zhao, Y. Du, M. Zhao, R. Sun, and Y. Zheng, “Introducing the GEV Activation Function for Highly Unbalanced Data to Develop COVID-19 Diagnostic Models,” IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 10, pp. 2776–2786, Oct. 2020, number: 10.

[8] M. Fang, B. He, L. Li, D. Dong, X. Yang, C. Li, L. Meng, L. Zhong, H. Li, H. Li, and J. Tian, “CT radiomics can help screen the Coronavirus disease 2019 (COVID-19): a preliminary study,” Science China Information Sciences, vol. 63, no. 7, p. 172103, Jul. 2020, number: 7.

[9] S. Hassantabar, M. Ahmadi, and A. Sharifi, “Diagnosis and detection of infected tissue of COVID-19 patients based on lung x-ray image using convolutional neural network approaches,” Chaos, Solitons & Fractals, vol. 140, p. 110170, Nov. 2020.

[10] M. J. Horry, S. Chakraborty, M. Paul, A. Ulhaq, B. Pradhan, M. Saha, and N. Shukla, “COVID-19 Detection Through Transfer Learning Using Multimodal Imaging Data,” IEEE Access, vol. 8, pp. 149 808–149 824, 2020.

[11] S. Hu, Y. Gao, Z. Niu, Y. Jiang, L. Li, X. Xiao, M. Wang, E. F. Fang, W. Menpes-Smith, J. Xia, H. Ye, and G. Yang, “Weakly Supervised Deep Learning for COVID-19 Infection Detection and Classification From CT Images,” IEEE Access, vol. 8, pp. 118 869–118 883, 2020.

[12] H. Ko, H. Chung, W. S. Kang, K. W. Kim, Y. Shin, S. J. Kang, J. H. Lee, Y. J. Kim, N. Y. Kim, H. Jung, and J. Lee, “COVID-19 Pneumonia Diagnosis Using a Simple 2D Deep Learning Framework With a Single Chest CT Image: Model Development and Validation,” Journal of Medical Internet Research, vol. 22, no. 6, p. e19569, Jun. 2020, number: 6.

[13] C. Liu, X. Wang, C. Liu, Q. Sun, and W. Peng, “Differentiating novel coronavirus pneumonia from general pneumonia based on machine learning,” BioMedical Engineering OnLine, vol. 19, no. 1, p. 66, Dec. 2020, number: 1.

[14] V. Perumal, V. Narayanan, and S. J. S. Rajasekar, “Detection of COVID-19 using CXR and CT images using Transfer Learning and Haralick features,” Applied Intelligence, vol. 51, no. 1, pp. 341–358, Jan. 2021, number: 1.

[15] K. Purohit, A. Kesarwani, D. R. Kisku, and M. Dalui, “COVID-19 Detection on Chest X-Ray and CT Scan Images Using Multi-image Augmented Deep Learning Model,” Bioengineering, preprint, Jul. 2020.

[16] V. Shah, R. Keniya, A. Shridharani, M. Punjabi, J. Shah, and N. Mehendale, “Diagnosis of COVID-19 using CT scan images and deep learning techniques,” Emergency Radiology, Feb. 2021.

[17] P. Silva, E. Luz, G. Silva, G. Moreira, R. Silva, D. Lucio, and D. Menotti, “COVID-19 detection in CT images with deep learning: 13A voting-based scheme and cross-datasets analysis,” Informatics in Medicine Unlocked, vol. 20, p. 100427, 2020.

[18] J. Song, H. Wang, Y. Liu, W. Wu, G. Dai, Z. Wu, P. Zhu, W. Zhang, K. W. Yeom, and K. Deng, “End-to-end automatic differentiation of the coronavirus disease 2019 (COVID-19) from viral pneumonia based on chest CT,” European Journal of Nuclear Medicine and Molecular Imaging, vol. 47, no. 11, pp. 2516–2524, Oct. 2020, number: 11.

[19] L. Sun, Z. Mo, F. Yan, L. Xia, F. Shan, Z. Ding, B. Song, W. Gao, W. Shao, F. Shi, H. Yuan, H. Jiang, D. Wu, Y. Wei, Y. Gao, H. Sui, D. Zhang, and D. Shen, “Adaptive Feature Selection Guided Deep Forest for COVID-19 Classification With Chest CT,” IEEE Journal of Biomedical and Health Informatics, vol. 24, no. 10, pp. 2798–2805, Oct. 2020, number: 10.

[20] X. Xu, X. Jiang, C. Ma, P. Du, X. Li, S. Lv, L. Yu, Q. Ni, Y. Chen, J. Su, G. Lang, Y. Li, H. Zhao, J. Liu, K. Xu, L. Ruan, J. Sheng, Y. Qiu, W. Wu, T. Liang, and L. Li, “A Deep Learning System to Screen Novel Coronavirus Disease 2019 Pneumonia,” Engineering, vol. 6, no. 10, pp. 1122–1129, Oct. 2020, number: 10.

[21] Y. Lecun and Y. Bengio, “Convolutional Networks for Images, Speech, and Time-Series,” The Handbook of Brain Theory and Neural Networks, Jan.1995

[22] D. Chicco and G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” BMC Genomics, vol. 21, no. 1, p. 6, Dec. 2020.

[23] Yang, X., He, X., Zhao, J., Zhang, Y., Zhang, S., & Xie, P. (2020). Covid-CT-Dataset: A CT Scan Dataset about COVID-19. ArXiv:2003.13865 [Cs, Eess, Stat].

[24] Morozov S.P., Andreychenko A.E., Blokhin I.A., Gelezhe P.B., Gonchar A.P., Nikolaev A.E., Pavlov N.A., Chernina V.Y., Gombolevskiy V.A., (2020). MosMedData: Data set of 1110 chest CT scans performed during the COVID-19 epidemic. Digital Diagnostics, 1(1), 49–59.

[25] Pominova, M., Kondrateva, E., Sharaev, M., Bernstein, A., Pavlov, S., & Burnaev, E. (2019). 3D Deformable Convolutions for MRI Classification. 2019 18th IEEE International Conference on Machine Learning And Applications (ICMLA), 1710–1716.

[26] Liu, Y., Stojadinovic, S., Hrycushko, B., Wardak, Z., Lau, S., Lu, W., Yan, Y., Jiang, S. B., Zhen, X., Timmerman, R., Nedzi, L., & Gu, X. (2017). A deep convolutional neural network-based automatic delineation strategy for multiple brain metastases stereotactic radiosurgery. PLOS ONE, 12(10), 1–17. https://doi.org/10.1371/journal.pone.0185844

[27] Goncharov, M., Pisov, M., Shevtsov, A., Shirokikh, B., Kurmukov, A., Blokhin, I., Chernina, V., Solovev, A., Gombolevskiy, V., Morozov, S., & Belyaev, M. (2021). CT-based COVID-19 Triage: Deep Multitask Learning Improves Joint Identification and Severity Quantification. Medical Image Analysis, 102054.

[28] Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., Aviles-Rivero, A. I., Etmann, C., McCague, C., Beer, L., Weir-McCall, J. R., Teng, Z., Gkrania-Klotsas, E., Rudd, J. H. F., Sala, E., & Schönlieb, C.-B. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nature Machine Intelligence, 3(3), 199–217. https://doi.org/10.1038/s42256-021-00307-0

[29] Machine learning analyses of lung imaging for COVID-19 falls short, Minded launches to streamline psychiatric med refills and more digital health news briefs. (2021, March 17). MobiHealthNews. https://www.mobihealthnews.com/news/machine-learning-analyses-lung-imaging-covid-falls-short-minded-launches-streamline-psychiatric

[30] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, B. C. Van Esesn, A. A. S. Awwal, and V. K. Asari, “The history began from alexnet: A comprehensive survey on deep learning approaches,” arXiv preprint arXiv:1803.01164, 2018.

[31] Amyar, A., Modzelewski, R., Li, H., & Ruan, S. (2020). Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation. Computers in Biology and Medicine, 126, 104037. https://doi.org/10.1016/j.compbiomed.2020.104037

[32] Anthimopoulos, M., Christodoulidis, S., Ebner, L., Geiser, T., Christe, A., & Mougiakakou, S. (2019). Semantic Segmentation of Pathological Lung Tissue With Dilated Fully Convolutional Networks. IEEE Journal of Biomedical and Health Informatics, 23(2), 714–722. https://doi.org/10.1109/JBHI.2018.2818620

[33] Cai, W., Liu, T., Xue, X., Luo, G., Wang, X., Shen, Y., Fang, Q., Sheng, J., Chen, F., & Liang, T. (2020). CT Quantification and Machine-learning Models for Assessment of Disease Severity and Prognosis of COVID-19 Patients. Academic Radiology, 27(12), 1665–1678. https://doi.org/10.1016/j.acra.2020.09.004

[34] Chen, G., Xiang, D., Zhang, B., Tian, H., Yang, X., Shi, F., Zhu, W., Tian, B., & Chen, X. (2019). Automatic Pathological Lung Segmentation in Low-Dose CT Image Using Eigenspace Sparse Shape Composition. IEEE Transactions on Medical Imaging, 38(7), 1736–1749. https://doi.org/10.1109/TMI.2018.2890510

[35] Elaziz, M. A., Ewees, A. A., Yousri, D., Alwerfali, H. S. N., Awad, Q. A., Lu, S., & Al-Qaness, M. A. A. (2020). An Improved Marine Predators Algorithm With Fuzzy Entropy for Multi-Level Thresholding: Real World Example of COVID-19 CT Image Segmentation. IEEE Access, 8, 125306–125330. https://doi.org/10.1109/ACCESS.2020.3007928

[36] A. W. Salehi, P. Baglat, and G. Gupta, “Review on machine and deep learning models for the detection and prediction of Coronavirus,” Materials Today: Proceedings, vol. 33, pp. 3896–3901, 2020. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2214785320347131

[37] Koohbanani, N. A., Jahanifar, M., Zamani Tajadin, N., & Rajpoot, N. (2020). NuClick: A deep learning framework for interactive segmentation of microscopic images. Medical Image Analysis, 65, 101771. https://doi.org/10.1016/j.media.2020.101771

[38] Liu, L., Kurgan, L., Wu, F.-X., & Wang, J. (2020). Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease. Medical Image Analysis, 65, 101791. https://doi.org/10.1016/j.media.2020.101791

[39] Mertz, L. (2020). AI-Driven COVID-19 Tools to Interpret, Quantify Lung Images. IEEE Pulse, 11(4), 2–7. https://doi.org/10.1109/MPULS.2020.3008354

[40] Oulefki, A., Agaian, S., Trongtirakul, T., & Kassah Laouar, A. (2021). Automatic COVID-19 lung infected region segmentation and measurement using CT-scans images. Pattern Recognition, 114, 107747. https://doi.org/10.1016/j.patcog.2020.107747

[41] Saeedizadeh, N., Minaee, S., Kafieh, R., Yazdani, S., & Sonka, M. (2020). Covid TV-UNet: Segmenting COVID-19 Chest CT Images Using Connectivity Imposed U-Net. ArXiv:2007.12303 [Cs, Eess]. http://arxiv.org/abs/2007.12303

[42] Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., He, K., Shi, Y., & Shen, D. (2021). Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19. IEEE Reviews in Biomedical Engineering, 14, 4–15. https://doi.org/10.1109/RBME.2020.2987975

[43] Voulodimos, A., Protopapadakis, E., Katsamenis, I., Doulamis, A., & Doulamis, N. (2020). Deep learning models for COVID-19 infected area segmentation in CT images [Preprint]. Health Informatics. https://doi.org/10.1101/2020.05.08.20094664

[44] Wu, D., Gong, K., Arru, C. D., Homayounieh, F., Bizzo, B., Buch, V., Ren, H., Kim, K., Neumark, N., Xu, P., Liu, Z., Fang, W., Xie, N., Tak, W. Y., Park, S. Y., Lee, Y. R., Kang, M. K., Park, J. G., Carriero, A., et al. (2020). Severity and Consolidation Quantification of COVID-19 From CT Images Using Deep Learning Based on Hybrid Weak Labels. IEEE Journal of Biomedical and Health Informatics, 24(12), 3529–3538. https://doi.org/10.1109/JBHI.2020.3030224

[45] Zhang, P., Zhong, Y., Deng, Y., Tang, X., & Li, X. (2020a). CoSinGAN: Learning COVID-19 Infection Segmentation from a Single Radiological Image. Diagnostics, 10(11), 901. https://doi.org/10.3390/diagnostics10110901

[46] Huang, G., Liu, Z., van der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/CVPR.2017.243

[47] Pan, Y., Guan, H., Zhou, S., Wang, Y., Li, Q., Zhu, T., Hu, Q., & Xia, L. (2020). Initial CT findings and temporal changes in patients with the novel coronavirus pneumonia (2019-nCoV): A study of 63 patients in Wuhan, China. European Radiology, 30(6), 3306–3309. https://doi.org/10.1007/s00330-020-06731-x

[48] MedSegData, http://medicalsegmentation.com/covid19/

[49] Morozov S.P., Andreychenko A.E., Blokhin I.A., Gelezhe P.B., Gonchar A.P., Nikolaev A.E., Pavlov N.A., Chernina V.Y., Gombolevskiy V.A., (2020). MosMedData: Data set of 1110 chest CT scans performed during the COVID-19 epidemic. Digital Diagnostics, 1(1), 49–59

[50] Covid-Semi-Seg dataset: Fan, D.-P., Zhou, T., Ji, G.-P., Zhou, Y., Chen, G., Fu, H., Shen, J., & Shao, L. (2020). Inf-Net: Automatic COVID-19 Lung Infection Segmentation From CT Images. IEEE Transactions on Medical Imaging, 39(8), 2626–2637. https://doi.org/10.1109/TMI.2020.2996645

[51] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, S, 11045, 3–11. https://doi.org/10.1007/978-3-030-00889-5_1

[52] Pan, F., Ye, T., Sun, P., Gui, S., Liang, B., Li, L., Zheng, D., Wang, J., Hesketh, R. L., Yang, L., & Zheng, C. (2020). Time Course of Lung Changes at Chest CT during Recovery from Coronavirus Disease 2019 (COVID-19). Radiology, 295(3), 715–721. https://doi.org/10.1148/radiol.2020200370

[53] SIRM COVID-19 database, https://www.sirm.org/en/category/articles/covid-19-database

[54] Ng, M.-Y., Lee, E. Y. P., Yang, J., Yang, F., Li, X., Wang, H., Lui, M. M., Lo, C. S.-Y., Leung, B., Khong, P.-L., Hui, C. K.-M., Yuen, K., & Kuo, M. D. (2020). Imaging Profile of the COVID-19 Infection: Radiologic Findings and Literature Review. Radiology: Cardiothoracic Imaging, 2(1), e200034. https://doi.org/10.1148/ryct.2020200034

[55] Salehi, S., Abedi, A., Balakrishnan, S., & Gholamrezanezhad, A. (2020). Coronavirus Disease 2019 (COVID-19): A Systematic Review of Imaging Findings in 919 Patients. American Journal of Roentgenology, 215(1), 87–93. https://doi.org/10.2214/AJR.20.23034

[56] J. P. Cohen, P. Morrison, and L. Dao, “COVID-19 image data collection,” 2020, arXiv:2003.11597. [Online]. Available: http://arxiv. org/abs/2003.11597

[57] Hansell, D. M., Bankier, A. A., MacMahon, H., McLoud, T. C., Müller, N. L., & Remy, J. (2008). Fleischner Society: Glossary of terms for thoracic imaging. Radiology, 246(3), 697–722. https://doi.org/10.1148/radiol.2462070712

[58] https://radiologyassistant.nl/chest/covid-19/covid19-imaging-findings

[59] Saburi, A., Schoepf, U. J., Ulversoy, K. A., Jafari, R., Eghbal, F., & Ghanei, M. (2020). From Radiological Manifestations to Pulmonary Pathogenesis of COVID-19: A Bench to Bedside Review. Radiology Research and Practice, 2020, e8825761. https://doi.org/10.1155/2020/8825761

[60] Adams, H. J. A., Kwee, T. C., Yakar, D., Hope, M. D., & Kwee, R. M. (2020). Chest CT Imaging Signature of Coronavirus Disease 2019 Infection: In Pursuit of the Scientific Evidence. Chest, 158(5), 1885–1895. https://doi.org/10.1016/j.chest.2020.06.025

[61] Yang, R., Li, X., Liu, H., Zhen, Y., Zhang, X., Xiong, Q., Luo, Y., Gao, C., & Zeng, W. (2020). Chest CT Severity Score: An Imaging Tool for Assessing Severe COVID-19. Radiology: Cardiothoracic Imaging, 2(2), e200047. https://doi.org/10.1148/ryct.2020200047

[62] Afshar, P., Heidarian, S., Naderkhani, F., Oikonomou, A., Plataniotis, K. N., & Mohammadi, A. (2020). COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray images. Pattern Recognition Letters, 138, 638–643. https://doi.org/10.1016/j.patrec.2020.09.010

[63] Albahli, S., & Albattah, W. (2020). Detection of coronavirus disease from X-ray images using deep learning and transfer learning algorithms. Journal of X-Ray Science and Technology, 28(5), 841–850. https://doi.org/10.3233/XST-200720

[64] Burgos-Artizzu, X. P. (n.d.). Automated covid-19 detection from frontal Chest X-Ray images using Deep Learning: An online feasibility study. 19.

[65] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556, 2014.

[66] Huang, G., Wang, Y., Wu, X., Qu, G., Chen, J., Yu, H., Zhang, M., Wang, L., Ai, J., Zhu, H., Chen, L., & Pei, B. (2020). Staging and typing of chest CT images: A quantitative analysis based on an ambispective observational cohort study of 125 patients with COVID-19 in Xiangyang, China [Preprint]. Radiology and Imaging. https://doi.org/10.1101/2020.10.25.20219253

[67] Hu, R., Ruan, G., Xiang, S., Huang, M., Liang, Q., & Li, J. (2020). Automated Diagnosis of COVID-19 Using Deep Learning and Data Augmentation on Chest CT [Preprint]. Health Informatics. https://doi.org/10.1101/2020.04.24.20078998

[68] Ilhan, H. O., Serbes, G., & Aydin, N. (2020a). Decision and Feature Level Fusion of Deep Features Extracted from Public COVID-19 Data-sets.

[69] Jin, S., Wang, B., Xu, H., Luo, C., Wei, L., Zhao, W., Hou, X., Ma, W., Xu, Z., Zheng, Z., Sun, W., Lan, L., Zhang, W., Mu, X., Shi, C., Wang, Z., Lee, J., Jin, Z., Lin, M., … Xu, W. (2020). AI-assisted CT imaging analysis for COVID-19 screening: Building and deploying a medical AI system in four weeks [Preprint]. Health Informatics. https://doi.org/10.1101/2020.03.19.20039354

[70] Khan, A. I., Shah, J. L., & Bhat, M. M. (2020). CoroNet: A deep neural network for detection and diagnosis of COVID-19 from chest x-ray images. Computer Methods and Programs in Biomedicine, 196, 105581. https://doi.org/10.1016/j.cmpb.2020.105581

[71] Li, L., Qin, L., Xu, Z., Yin, Y., Wang, X., Kong, B., Bai, J., Lu, Y., Fang, Z., Song, Q., Cao, K., Liu, D., Wang, G., Xu, Q., Fang, X., Zhang, S., Xia, J., & Xia, J. (2020). Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy. Radiology, 296(2), E65–E71. https://doi.org/10.1148/radiol.2020200905

[72] Mohammed, A., Wang, C., Zhao, M., Ullah, M., Naseem, R., Wang, H., Pedersen, M., & Cheikh, F. A. (2020). Weakly-Supervised Network for Detection of COVID-19 in Chest CT Scans. IEEE Access, 8, 155987–156000. https://doi.org/10.1109/ACCESS.2020.3018498

[73] Rahaman, M. M., Li, C., Yao, Y., Kulwa, F., Rahman, M. A., Wang, Q., Qi, S., Kong, F., Zhu, X., & Zhao, X. (2020). Identification of COVID-19 samples from chest X-Ray images using deep learning: A comparison of transfer learning approaches. Journal of X-Ray Science and Technology, 28(5), 821–839. https://doi.org/10.3233/XST-200715

[74] Somasekar, J., Pavan Kumar, P., Sharma, A., & Ramesh, G. (2020). Machine learning and image analysis applications in the fight against COVID-19 pandemic: Datasets, research directions, challenges, and opportunities. Materials Today: Proceedings, S2214785320370620. https://doi.org/10.1016/j.matpr.2020.09.352

[75] Chen, X., Yao, L., & Zhang, Y. (2020). Residual Attention U-Net for Automated Multi-Class Segmentation of COVID-19 Chest CT Images. ArXiv:2004.05645 [Cs, Eess, q-Bio]. http://arxiv.org/abs/2004.05645

[76] Kwee, T. C., & Kwee, R. M. (2020). Chest CT in COVID-19: What the Radiologist Needs to Know. RadioGraphics, 40(7), 1848–1865. https://doi.org/10.1148/rg.2020200159

[77] Laurence, A. A. (2021). Alexlaurence/NIfTI-Image-Converter [MATLAB]. https://github.com/alexlaurence/NIfTI-Image-Converter (Original work published 2019)

[78] Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K. S. M., Lau, E. H. Y., Wong, J. Y., Xing, X., Xiang, N., Wu, Y., Li, C., Chen, Q., Li, D., Liu, T., Zhao, J., Liu, M., … Feng, Z. (2020). Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. New England Journal of Medicine, 382(13), 1199–1207. https://doi.org/10.1056/NEJMoa2001316

[79] Li, Y., Huang, X., Wang, Y., Xu, Z., Sun, Y., & Zhang, Q. (n.d.). U-net Ensemble Model for Segmentation in Histopathology Images. 8. Pre-print

[80] Ng, M.-Y., Lee, E. Y. P., Yang, J., Yang, F., Li, X., Wang, H., Lui, M. M., Lo, C. S.-Y., Leung, B., Khong, P.-L., Hui, C. K.-M., Yuen, K., & Kuo, M. D. (2020). Imaging Profile of the COVID-19 Infection: Radiologic Findings and Literature Review. Radiology: Cardiothoracic Imaging, 2(1), e200034. https://doi.org/10.1148/ryct.2020200034

[81] Rubin, G. D., Ryerson, C. J., Haramati, L. B., Sverzellati, N., Kanne, J. P., Raoof, S., Schluger, N. W., Volpi, A., Yim, J.-J., Martin, I. B. K., Anderson, D. J., Kong, C., Altes, T., Bush, A., Desai, S. R., Goldin, Jonathan, Goo, J. M., Humbert, M., Inoue, Y., … Leung, A. N. (2020). The Role of Chest Imaging in Patient Management during the COVID-19 Pandemic: A Multinational Consensus Statement from the Fleischner Society. Radiology, 296(1), 172–180. https://doi.org/10.1148/radiol.2020201365

[82] Saburi, A., Schoepf, U. J., Ulversoy, K. A., Jafari, R., Eghbal, F., & Ghanei, M. (2020). From Radiological Manifestations to Pulmonary Pathogenesis of COVID-19: A Bench to Bedside Review. Radiology Research and Practice, 2020, e8825761. https://doi.org/10.1155/2020/8825761

[83] Safarov, S., & Whangbo, T. K. (2021). A-DenseUNet: Adaptive Densely Connected UNet for Polyp Segmentation in Colonoscopy Images with Atrous Convolution. Sensors, 21(4), 1441. https://doi.org/10.3390/s21041441

[84] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in European conference on computer vision. Springer, 2016, pp. 630–645.

[85] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in Thirty-first AAAI conference on artificial intelligence, 2017.

[86] Shoeibi, A., Khodatars, M., Alizadehsani, R., Ghassemi, N., Jafari, M., Moridian, P., Khadem, A., Sadeghi, D., Hussain, S., Zare, A., Sani, Z. A., Bazeli, J., Khozeimeh, F., Khosravi, A., Nahavandi, S., Acharya, U. R., & Shi, P. (2020). Automated Detection and Forecasting of COVID-19 using Deep Learning Techniques: A Review. ArXiv:2007.10785 [Cs, Eess]. http://arxiv.org/abs/2007.10785

[87] Wu, Y.-H., Gao, S.-H., Mei, J., Xu, J., Fan, D.-P., Zhang, R.-G., & Cheng, M.-M. (2021). JCS: An Explainable COVID-19 Diagnosis System by Joint Classification and Segmentation. IEEE Transactions on Image Processing, 30, 3113–3126. https://doi.org/10.1109/TIP.2021.3058783

[88] Xie, W., Jacobs, C., Charbonnier, J.-P., & van Ginneken, B. (2020). Relational Modeling for Robust and Efficient Pulmonary Lobe Segmentation in CT Scans. ArXiv:2004.07443 [Cs, Eess]. https://doi.org/10.1109/TMI.2020.2995108

[89] Xie, Y., Zhang, J., Shen, C., & Xia, Y. (2021). CoTr: Efficiently Bridging CNN and Transformer for 3D Medical Image Segmentation. ArXiv:2103.03024 [Cs]. http://arxiv.org/abs/2103.03024

[90] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: Attention Is All You Need. CoRR abs/1706.03762, NIPS (2017)

[91] Zhao, S., Li, Z., Chen, Y., Zhao, W., Xie, X., Liu, J., Zhao, D., & Li, Y. (2020). SCOAT-Net: A Novel Network for Segmenting COVID-19 Lung Opacification from CT Images. MedRxiv, 2020.09.23.20191726. https://doi.org/10.1101/2020.09.23.20191726

[92] Wang, X., Deng, X., Fu, Q., Zhou, Q., Feng, J., Ma, H., Liu, W., & Zheng, C. (2020). A Weakly-Supervised Framework for COVID-19 Classification and Lesion Localization From

Chest CT. IEEE Transactions on Medical Imaging, 39(8), 2615–2625. https://doi.org/10.1109/TMI.2020.2995965

[93] Zhou, S., Wang, Y., Zhu, T., & Xia, L. (2020). CT Features of Coronavirus Disease 2019 (COVID-19) Pneumonia in 62 Patients in Wuhan, China. American Journal of Roentgenology, 214(6), 1287–1294.

[94] Zhou, T., Canu, S., & Ruan, S. (2021). Automatic COVID-19 CT segmentation using U-Net integrated spatial and channel attention mechanism. International Journal of Imaging Systems and Technology, 31(1), 16–27.

[95] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2018). UNet++: A Nested U-Net Architecture for Medical Image Segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, S, 11045, 3–11.

[96] MIPAV tool, https://mipav.cit.nih.gov/clickwrap.php

[97] M. Karim, T. Dohmen, D. Rebholz-Schuhmann, S. Decker, M. Cochez, ¨O. Beyan et al., “Deepcovidexplainer: Explainable covid-19 predictions based on chest x-ray images,” arXiv preprint arXiv:2004.04582, 2020.

[98] Long, J., Shelhamer, E. and Darrell, T. (2014) Fully Convolutional Networks for Semantic Segmentation. IEEE Transactions on Pattern Analysis & Machine Intelligence, 39, 640-651.

[99] Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. International Conference on Machine Learning, 6105–6114.

[100] L. Ilagan, R. Concepcion, M. Cabatuan and C. Roque, "Vision-Based Metal Oxide Semiconductor Transistor-Level Layout Error Classification Using EfficientNet Model," 2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2020, pp. 1-6

[101] Verma, P., Tripathi, V., & Pant, B. (2021). Comparison of different optimizers implemented on the deep learning architectures for COVID-19 classification. Materials Today: Proceedings.

[102] Marques, G., Agarwal, D., & de la Torre Díez, I. (2020). Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network. Applied Soft Computing, 96, 106691.

[103] T. Dozat, “Incorporating Nesterov Momentum into Adam,” ICLR 2016.

[104] X. Wang, Y. Peng, et al., ChestX-ray8: hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 3462–3471.

[105] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. ArXiv:1505.04597 [Cs].

[106] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014). Going Deeper with Convolutions. ArXiv:1409.4842 [Cs].

[107] Ter-Sarkisov, A. (n.d.). COVID-CT-Mask-Net: Prediction of COVID-19 from CT Scans Using Regional Features. 9.

[108] Saood, A., & Hatem, I. (2021). COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. BMC Medical Imaging, 21(1), 19.

[109] Han, L., Chen, Y., Li, J., Zhong, B., Lei, Y., & Sun, M. (2021). Liver segmentation with 2.5D perpendicular UNets. Computers & Electrical Engineering, 91, 107118.

[110] Hou, L., Cheng, Y., Shazeer, N., Parmar, N., Li, Y., Korfiatis, P., Drucker, T. M., Blezek, D. J., & Song, X. (2019). High Resolution Medical Image Analysis with Spatial Partitioning. ArXiv:1909.03108 [Cs, Eess].

[111] Jin, Q., Cui, H., Sun, C., Meng, Z., Wei, L., & Su, R. (2021). Domain adaptation based self-correction model for COVID-19 infection segmentation in CT images. Expert Systems with Applications, 176, 114848.

[112] S. Arora, A. Bhaskara, R. Ge, and T. Ma. Provable bounds for learning some deep representations. CoRR, abs/1310.6343, 2013.

[113] M. Lin, Q. Chen, and S. Yan. Network in network. CoRR, abs/1312.4400, 2013.

[114] Tversky, A. (1977). Features of similarity. Psychological Review, 84(4), 327–352.

[115] Senthilraja, S., Suresh, D. P., & Suganthi, D. M. (2014). Noise Reduction in Computerized Tomography Image Using WB – Filter. 5(3), 5.

[116] Arce, Gonzalo R. (2005). Nonlinear Signal Processing: A Statistical Approach. New Jersey, USA: Wiley. ISBN 0-471-67624-1.

[117] Greensted, Andrew. “Otsu Thresholding.” The Lab Book Pages, 17 June 2010, www.labbookpages.co.uk/software/imgProc/otsuThreshold.html.

[118] https://www.fda.gov/radiation-emitting-products/medical-x-ray-imaging/other-information-resources-related-whole-body-ct-screening

[119] https://www.fda.gov/radiation-emitting-products/medical-x-ray-imaging/what-computed-tomography

[120] Half A Century In CT: How Computerized Tomography Has Evolved — ISCT https://www.isct.org/computed-tomography-blog/2017/2/10/half-a-century-in-ct-how-computed-tomography-has-evolved

[121] https://radiologykey.com/imaging-principles-in-computed-tomography-2/

[122] Martino, S., Reid, J., & Odle, T. G. (2008). Changing Practice for Medical Imaging and Radiation Therapy Professionals. American Society of Radiologic Technologists, 41.

[123] Semelka, R. C., Armao, D. M., Elias, J., & Huda, W. (2007). Imaging strategies to reduce the risk of radiation in CT studies, including selective substitution with MRI. Journal of Magnetic Resonance Imaging, 25(5), 900–909.

[124] https://www.isct.org/computed-tomography-blog/2017/2/10/whats-in-store-for-the-future-of-ct-part-two

[125] Liu, W., Ding, X., Kong, B., Fan, B., & Chen, L. (2014). Reducing the radiation dose with the adaptive statistical iterative reconstruction technique for chest CT in adults: A parameter study. Chinese Medical Journal, 127(7), 1284–1288.

[126] Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural Machine Translation by Jointly Learning to Align and Translate. ICLR 2015.

[127] Li, K., Zhang, T., & Malik, J. (2019). Approximate Feature Collisions in Neural Nets. Advances in Neural Information Processing Systems, 32.

[128] T. M. Tuan, T. T. Ngan, C. N. Giap et al., “Semisupervised fuzzy clustering methods for x-ray image segmentation,” in Handbook of Data Science Approaches for Biomedical Engineering. Elsevier, 2020, pp. 251–289.

[129] A. Shoeibi, N. Ghassemi, H. Hosseini-Nejad, and M. Rouhani, “An efficient brain mr images segmentation hardware using kernel fuzzy c-means,” in 2019 26th National and 4th International Iranian Conference on Biomedical Engineering (ICBME). IEEE, 2019, pp. 93–99

[130] Varnousfaderani, E. S., Yousefi, S., Bowd, C., Belghith, A., & Goldbaum, M. H. (2015). Vessel Delineation in Retinal Images using Leung-Malik filters and Two Levels Hierarchical Learning. AMIA ... Annual Symposium proceedings. AMIA Symposium, 2015, 1140–1147.

[131] Imani, M. (2021). Automatic diagnosis of coronavirus (COVID-19) using shape and texture characteristics extracted from X-Ray and CT-Scan images. Biomedical Signal Processing and Control, 68, 102602.

[132] Gunasekaran, S., Rajan, S., Moses, L., Vikram, S., Subalakshmi, M., & Shudhersini, B. (2021). Wavelet Based CNN for Diagnosis of COVID 19 using Chest X Ray. 1084, 012015.

[133] Bernico, M. (2018). Deep learning quick reference : Useful hacks for training and optimizing deep neural networks with tensorflow and keras. Chapter 1, pp. 23.

[134] Harris, C. R., Millman, K. J., van der Walt, Stéfan J, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy. Nature (London), 585(7825), 357-362.

[135] Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in Science & Engineering, 9(3), 90-95.

[136] https://scikit-learn.org/stable/index.html

[137] https://jupyter.org

[138] https://libraries.io/pypi/nibabel

[139] Mittal, B. & Oh, J. CoviSegNet: Covid-19 Disease Area Segmentation using Machine Learning Analyses for Lung Imaging. ISPA 2021, 13-15 September 2021, Zagreb, Croatia. Accepted for publication.