Capsule Network-based COVID-19 Diagnosis and Transformer-based Lung Cancer Invasiveness Prediction via Computerized Tomography (CT) Images

Shahin Heidarian

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of

Master of Applied Science (Electrical and Computer Engineering) at

Concordia University

Montréal, Québec, Canada

December 2021

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Shahin Heidarian**

Entitled: **Capsule Network-based COVID-19 Diagnosis and Transformer-based Lung Cancer Invasiveness Prediction via Computerized Tomography (CT) Images**

and submitted in partial fulfillment of the requirements for the degree of

Master of Applied Science (Electrical and Computer Engineering)

complies with the regulations of this University and meets the accepted standards with respect to originality and quality.

Signed by the Final Examining Committee:

_____ Chair
Dr. Habib Benali

_____ External Examiner
Dr. Abdessamad Ben Hamza

_____ Examiner
Dr. Wei-Ping Zhu

_____ Supervisor
Dr. Arash Mohammadi

Approved by _____
Dr.Yousef R. Shayan, Chair
Department of Electrical and Computer Engineering

_____ 2021

_____
Dr. Mourad Debbabi, Dean
Faculty of Engineering and Computer Science

Abstract

Capsule Network-based COVID-19 Diagnosis and Transformer-based Lung Cancer Invasiveness Prediction via Computerized Tomography (CT) Images

Shahin Heidarian

Early diagnosis and prognosis of life-threatening diseases such as the novel coronavirus infection (COVID-19) and Lung Cancer (LC), involves tackling critical challenges including but not limited to their undisclosed characteristics, non-stationary nature, significant inter-disease similarities, and intra-disease variations. In particular, within the context of a highly contagious disease such as COVID-19, early and reliable diagnosis is of significant importance. On the other hand, when it comes to diagnosis and prognosis of LC, an accurate prediction of the disease invasiveness becomes of primary importance. Recent advancements of Artificial Intelligence (AI) and Deep Learning (DL)-based architectures have resulted in a surge of interest in the utilization of medical images to develop decision support and stand-alone models to address the aforementioned challenges. In this context, the focus of the thesis is on the utilization of volumetric chest CT images to develop robust and fully-automated diagnostic frameworks for COVID-19 diagnosis and LC invasiveness prediction. In particular, Capsule Network (CapsNet) and Transformer-based architectures are developed to expand the application of AI in this domain. More specifically, first, CT-CAPS [1] and COVID-FACT [2] frameworks are proposed to analyze CT images, identify slices demonstrating infection, and perform patient-level classification of COVID-19. The proposed frameworks are developed based on the CapsNet architecture, which unlike the widely-used Convolutional Neural Networks (CNNs), is capable of capturing spatial relations among instances in an image and being trained on small datasets. These characteristics are of utmost importance when analyzing a newly emerged disease with specific spatial patterns in its images. Furthermore, following the recent and ever-increasing interest in using Low-Dose and Ultra-Low-Dose CT scans (LDCT and ULDCT)

for COVID-19 screening, the WSO-CAPS framework [3] is proposed to enhance performance of the proposed models to deal with noisy and low-quality CT scans. In addition, given that CT scans acquired from multiple centers and cohorts mainly show different qualities and characteristics, which negatively affect the generalizability of DL-based models, a unique multi-center dataset of CT scans, referred to as the "SPGC-COVID Dataset" [4], is constructed, which incorporates CT scans of COVID-19, Community Acquired Pneumonia (CAP), and normal cases, obtained using standard and low-dose imaging protocols. An enhancement approach is then proposed to boost the performance of the developed classification frameworks when being tested on varied CT scans in the SPGC-COVID dataset. With respect to the second objective of this thesis (i.e., Lung Cancer invasiveness prediction), the CAE-Transformer framework is proposed, which utilizes image-driven features to predict the invasiveness of Lung Adenocarcinomas (LUACs) from non-thin 3D CT scans. The proposed framework introduces a new viewpoint in CT scan analysis, which relies on the sequential nature of the volumetric CT scans. More specifically, the CAE-Transformer [5] adopts the transformer architecture, which was initially designed for sequential data, to capture inter-slice dependencies in an efficient and non-complex fashion.

List of Abbreviations

Abbreviation	Description
2D	2 Dimensional
3D	3 Dimensional
AAH	Atypical Adenomatous Hyperplasia
AIS	Adenocarcinoma In Situ
AUC	Area Under the Curve
BiLSTM	Bi-directional Long Short term Memory
CAE	Convolutional Auto-Encoder
CAP	Community Acquired Pneumonia
CNN	Convolutional Neural Network
CR	Chest Radiographs
CT	Computed Tomography
CapsNets	Capsule Networks
CvT	Convolutional Vision Transformer
DICOM	Digital Imaging and Communications in Medicine
DL	Deep Learning
FC	Fully-Connected
FNR	False Negative Rate
FPCA	Functional Principal Component Analysis
GAN	Generative Adversarial Network
GAP	Global Average Pooling

GGN	Ground Glass Nodules
GGO	Ground Glass Opacity
GMP	Global Max Pooling
Grad-CAM	Gradient-weighted Class Activation Mapping
HU	Hounsfield Units
ICU	Intensive Care Unite
IDRI	Image Database Resource Initiative
IPA	Invasive Pulmonary Adenocarcinoma
LC	Lung Cancer
LDCT	Low Dose CT scans
LIDC	Lung Image Database Consortium
LN	Layer Normalization
LSTM	Long Short term Memory
LUAC	Lung Adenocarcinoma
MHA	Multi-Head Attention
MIA	Minimally Invasive Adenocarcinoma
ML	Machine Learning
MLP	Multi-Layer Perceptron
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
MoE	Mixture of Expert
NLP	Natural Language Processing
PE	Positional Embedding
PSN	Part-Solid Nodules
RNN	Recurrent Neural Networks
ROC	Receiver Operating Characteristic
ROI	Region of Interest
RT-PCR	Reverse Transcription Polymerase Chain Reaction

ReLU	Rectified Linear Unit
SPGC	Signal Processing Grand Challenge
SSN	Sub-Solid Nodule
SVM	Support Vector Machine
TE	Token Embedding
UID	User Identifier
ULDCT	Ultra-Low Dose CT scans
ViT	Vision Transformer
WEM	Window Estimator Module
WL	Window Level
WSO	Window Setting Optimization
WW	Window Width
kVP	kilo Voltage Peak
pGGN	pure Ground Glass Nodules

Acknowledgments

I would like to extend my sincere gratitude to my supervisor, Dr.Arash Mohammadi, whose endless support, thoughtful advice, and ambitious spirit of research made this project possible. Dr. Arash's passion for research, sympathy during challenging times in my life, and commitment to giving recommendations, even late at night and early in the morning, changed my entire life. He acquainted me with new perspectives in constructive collaboration and teamwork, and inspired me to believe in myself, without which I would have struggled a lot towards the completion of my MASc. program.

I would also like to thank my fellow labmates, whose friendship and support helped me thrive in a positive and supportive atmosphere. In particular, I would like to express my heartfelt gratitude to Dr. Parnian Afshar, my intelligent friend and research partner, whose insightful comments, encouragement, and productive feedbacks during the past two years were invaluable in achieving my research and academic goals.

In addition, I would like to thank Dr. Moezedin Javad Rafiee for his supportive manner and great hospitality during our group meetings. I really appreciate his professional attitude and patience in answering my questions. His support was crucial in conducting this research project, and armed me with profound knowledge in the field of radiology. I would also like to thank Dr. Anastasia Oikonomou, whose technical comments and interpretations, besides her prompt and friendly responses to my inquiries, helped me enhance the quality of this research.

Finally, a warm and special thanks to my family, whose unconditional support and sacrifices made a better future for me, and showed me that distance only separates our bodies but never our hearts. I cannot thank you enough for all the love and encouragement you have given me.

Contents

List of Figures

List of Tables

Chapter 1

Thesis Overview

Automatic and accurate diagnosis of critical diseases such as contagious lung infections and Lung Cancer (LC) has recently attracted huge attentions among researchers and professionals in the field of Artificial Intelligence (AI). Especially, considering recent advances in fields of Machine Learning (ML) and Deep Learning (DL), developing an accurate and fast automated framework to assist healthcare professionals with diagnosis and treatment planning has become more promising than ever. Such frameworks can utilize medical images, biological signals, and/or clinical data to compensate for the limitations of the current diagnostic and prognostic solutions. Recently, the global outbreak due to the novel coronavirus disease (COVID-19) has sparked an unforeseeable global crisis since its emergence in late 2019.

The COVID-19 pandemic has reshaped our societies and people's lives in many lasting ways, and caused millions of deaths so far. In spite of the global enterprise to prevent the rapid outbreak of the disease and flatten the epidemic curve, there are still thousands of reported cases around the world on daily bases, which consistently raise the concern of facing a major epidemic wave or a new fatal and contagious variant of the virus such as the Omicron variant. The emergence of the COVID-19 pandemic further emphasizes the necessity and benefits of developing fast and reliable AI-based diagnostic and prognostic solutions.

In addition, besides the direct consequences of a pandemic, there are a wide variety of side-effects targeting the healthcare system. In particular, during a pandemic, all the resources are directed towards fighting the fast-spreading disease, which in turn results in late diagnoses and inaccurate

treatments of other vital diseases such as LC. These side-effects put a heavy burden on the healthcare system and professionals and once again highlight the importance of developing automated stand-alone and decision support frameworks to provide timely and accurate diagnosis as well as additional information about the disease.

1.1 Thesis Objectives

In brief, this thesis mainly focuses on the development of automated DL-based frameworks for the following two major applications.

1.1.1 Diagnosis of COVID-19 Disease from Volumetric Chest CT Scans

The early diagnosis of COVID-19 is of paramount importance to assist health and governmental authorities with developing efficient resource allocation plans and breaking the transmission chain. In the case of a pandemic such as the current COVID-19 global outbreak, healthcare professionals experience unexpected heavy workloads, caused by the abrupt increase in the number of people in the need of examination, which reduces their concentration and efficiency to properly identify cases and confirm the results. Such an increasing workload points out the need to distinguish normal cases and non-COVID infections from COVID-19 cases in a timely fashion. Reverse Transcription Polymerase Chain Reaction (RT-PCR), which is currently the gold standard in diagnosis of COVID-19, is time-consuming and prone to high false-negative rate [8]. Recent studies have demonstrated the strong capability of Chest Radiographs (CR) and Computed Tomography (CT) scans in providing distinctive patterns associated with the COVID-19 infection [9, 10]. Studies also show specific distribution of the disease imaging manifestations in the lung [11–14]. It is worth noting that chest radiograph acquisition is relatively simple with less radiation exposure than CT scans. A single CR image, however, fails to incorporate details of infections in the lung and cannot provide a comprehensive view for the lung infection diagnosis. CT scan, on the other hand, is an alternative imaging modality that incorporates the detailed structure of the lung and infected areas by generating cross-sectional images (slices) to create a 3D representation of the body. Such scans are highly sensitive to the diagnosis of COVID-19 infection, particularly based on its specific abnormality pattern and infection distribution

in the lung [8]. Obtaining several images per patient (often more than 100 slices), however, makes the CT scan analysis challenging and time-consuming as radiologists should carefully review all images before making a decision. In addition, chest CT is widely used as a primary diagnostic imaging modality in many countries, especially in those where RT-PCR test resources are limited. Therefore, there is an unmet need to develop advanced automated frameworks based on CT images to speed up the diagnosis procedure. In particular, the following challenges need to be targeted:

- Generally speaking, COVID-19 lung imaging manifestations are highly overlapped with those of the Community Acquired Pneumonia (CAP), leading to mis-classification even by experienced radiologists. This has motivated development of DL-based frameworks for identification of COVID-19 patients based on medical images. Most DL-based algorithms proposed to analyze medical images and identify COVID-19 cases are mainly developed based on Convolutional Neural Networks (CNNs). CNNs, however, require extensive data augmentation and large datasets to identify detailed spatial relations between image instances. In other words, CNNs commonly fail to recognize an object when it is rotated or transformed. In the case of a relatively new disease such as COVID-19, or next probable pandemics caused by a new unforeseeable phenomenon, large annotated/labeled datasets are not easily accessible. Moreover, finding spatial relations in CT images is highly important as most COVID-19 cases have been reported with a specific infection distribution in their images [11–14]. Capsule Networks (CapsNets) [15], in contrast to the CNNs, are equipped with a routing by agreement process enabling them to capture spatial patterns between instances of an image. More specifically, by reaching a mutual agreement, higher-level objects are constructed from lower-level ones. Even without a large dataset, capsules interpret the object instantiation parameters as well as its existence, which jointly determine the object's characteristics and spatial relations with other instances. The superiority of Capsule Networks over their CNN-based counterparts has been recently shown in different medial image processing problems [16–21].

- Aside from the utilized DL architecture, such models commonly achieve lower performances when there is heterogeneity in the data characteristics between the train and test sets, which is common when acquiring data from multiple imaging centers [22]. Therefore, the necessity

of developing a robust framework is of utmost importance to minimize the effect of the gap between the train and test sets and provide acceptable results on varied external datasets. In the case of CT scans, there are several factors contributing to the characteristics of the images among which, type of scanners, scanner manufacturers, and scanning protocols have the most influence on the quality and characteristics of the scans [23, 24]. Furthermore, the patients' clinical and surgical history can add more complexity and undesired artifacts to the CT scans that might have been blind to the trained model [25].

- Finally, it is worth noting that CT scans in their standard form expose patients to a high level of harmful radiation causing devastating effects on the body. Alternatively, Low and Ultra-Low Dose CT scans, commonly known as LDCT and ULDCT respectively, have been recently used in many diagnostic applications and proved to be effective in providing informative details of the disease imaging manifestation [26–28]. Recently, in the course the current COVID-19 pandemic, the increasing number of suspected COVID-19 cases in need of being scanned led the radiologists to move from standard dose CT scans to acquiring LDCT and ULDCT, aiming to decrease the detrimental effects of the scans caused by the radiation on the patients [29, 30]. Decreasing the radiation dose is usually performed by using lower x-ray tube currents, which in turn will impose a high level of noise on the acquired images, requiring more time and attention to accurately analyze the images.

Capitalizing on the above discussion, multiple Capsule Network-based frameworks are proposed in this thesis to address the following diagnostic tasks using chest CT scans:

(1) Classifying patients into COVID-19 and non-COVID (normal and CAP) cases using standard dose chest CT scans. Such automated framework provides healthcare professionals with immediate and significant assistance to exclude non-COVID cases quickly in the first step, helping them to pay more attention and allocate more medical resources to COVID-19 identified cases.

(2) Distinguishing normal, CAP, and COVID-19 cases in a three-way classification from a multi-center dataset of heterogeneous CT scans including LDCT, ULDCT, and scans of patients with the history of heart disease or surgery. An efficient and robust framework capable of

accomplishing this task will further optimize the treatment plan and resource allocation in healtcare systems and is one step forward towards the ultimate goal of development of clinically applicable AI-based frameworks.

(3) Automatically adjusting CT window setting parameters, finding the optimized setting to view LDCT and ULDCT, and identifying slices with the evidence of infection. The proposed model is capable of providing additional information on the disease image manifestations and locating the lung areas with the evidence of infection from LDCT and ULDCT, as well as standard dose CT scans.

1.1.2 Lung Adenocarcinoma invasiveness prediction from non-thin section volumetric CT scans

The second main focus of this thesis is on the diagnosis of LC, which is the deadliest and least funded cancer worldwide [31, 32]. Non-small-cell LC is the major type of LC, and Lung Adenocarcinoma (LUAC) is the most prevalent histologic sub-type [33]. A timely and accurate attempt to differentiate the LUACs is of utmost importance to guide a proper treatment plan, as in some cases, a pre-invasive or minimally invasive case can be monitored with regular follow up CTs, whereas invasive lesions should undergo immediate surgical resection if they are deemed eligible. Lung nodules manifesting as Ground Glass Nodules (GGN) or Subsolid Nodules (SSNs) on CT have a higher risk of malignancy than other incidentally detected small solid nodules, and SSNs are often diagnosed as lung adenocarcinoma [34, 35]. LUACs are categorized according to their histology into three categories: pre-invasive lesions including Atypical Adenomatous Hyperplasia (AAH) and Adenocarcinoma In Situ (AIS), Minimally Invasive Adenocarcinoma (MIA), and Invasive Pulmonary Adenocarcinoma (IPA) [35]. Most often, the SSN's types are diagnosed based on their pathological findings performed after surgical resections, which is not desired for treatment planning. Currently, radiologists use chest CT scans to assess the invasiveness of the SSNs based on their imaging findings and patterns prior to determining the proper treatment. Such visual approaches, however, are time-consuming, subjective, and error-prone. So far, several studies have used high-resolution and thin-slice ($< 1.5mm$) CT images for the SSN classification, which require longer analysis times, as

well as more storage capacity and reconstruction time [36,37]. Recent LC screening recommendation, however, suggests using Low Dose CT scans with thicker slice-thicknesses (up to $2.5mm$) [38, 39]. Moreover, lung nodules are mostly identified from CT scans performed for varied clinical purposes acquired using routine standard or low dose scanning protocols with non-thin slice thicknesses (up to $5mm$) [40]. The above discussion implies the necessity of developing an accurate automated classification framework that performs well regardless of the underlying technical settings.

1.2 Contributions

The main objective of this thesis is the development of automated DL-based frameworks to analyze chest CT scans. More specifically, this thesis proposes automated frameworks to identify the COVID-19 infection and predict the invasiveness of Lung Adenocarcinoma from chest CT scans and attempts to take one step forward towards reaching the ultimate goal of using DL-based solutions in clinical practice. Besides the model development, this thesis investigates various aspects of models, which play important roles in the healthcare systems. In particular, this thesis evaluates the interpretability of the models, the ability to adjust the functionality based on expert's preferences, and the capability of the trained models to be generalized over other medical centers and cohorts. The main contribution of this thesis research work are briefly outlined below:

(1) **CT-CAPS and COVID-FACT Frameworks [1,2]:** These proposed frameworks are developed to automatically distinguish COVID-19 cases from non-COVID (CAP and normal) ones. Both frameworks are developed based on Capsule Networks and take the volumetric CT scan (all slices) as the input and provide the classification probability scores as the output. Both proposed frameworks utilize a two-stage approach, which is designed to facilitate the translation from 2D slice-level domain to the patient-level diagnosis. This is of paramount importance in COVID-19 detection using CT scans, as a CT examination is typically associated with hundreds of slices that cannot be analyzed at once. It is also worth noting that the first stage of both frameworks is trained on a dataset, which does not require any infection annotation or a very precise slice labeling. This leads to a fast and timely design process, which is highly valuable when we are faced with early emergence of a new type of data. In what follows, the two

proposed frameworks are briefly described:

(a) CT-CAPS is a fully-automated framework based on Capsule Networks, which represents each slice of a CT scan by a small feature map in the first stage and utilizes the generated feature maps to distinguish COVID-19 cases from non-COVID (CAP and normal) cases in the next stage. The proposed CT-CAPS copes with the development difficulties caused by the large number of CT slices per patient and emphasizes the capability of capsules to represent a large volumetric CT scan by a very small matrix. In this framework, a Capsule Network-based feature extractor is proposed to detect specific characteristics of CT slices, followed by a Max Pooling Layer to convert slice-level feature maps into patient-level ones. Finally, a stack of fully connected layers are added to provide the final decision. Furthermore, to improve on the explainability of the model, the Grad-CAM localization mapping approach [41] is incorporated to determine lung regions contributing the most to the final decision.

(b) COVID-FACT is the extension of the CT-CAPS framework in which the first stage detects slices demonstrating infection in a volumetric CT scan to be analyzed and classified in the next stage. At the second stage, candidate slices detected at the previous stage are classified into COVID and non-COVID cases and a voting mechanism is applied to generate the patient-level classification scores. COVID-FACT's two-stage architecture has the advantage of being trained on even a weakly labeled dataset, as errors at the first stage can be compensated at the second stage. In addition, manual infection annotation is completely removed from the COVID-FACT. The only information required from the radiologists to train the first stage is the slices containing evidence of infection and the radiologist's input is not required in the test phase of the COVID-FACT and the trained framework is fully automated.

In this thesis, two variants of the aforementioned frameworks are also developed, one of which is fed with the whole chest CT image, while the other one utilizes the segmented lung area as the input. In the latter case, instead of using an original chest CT image, first a pre-trained segmentation model [42] is applied to extract the lung region, which is then provided as the

input. This will be further clarified in Chapter 3. Experimental results show that the model coupled with the lung area segmentation achieves relatively higher performances compared to the other variation working with original images. As a final note, it is worth mentioning that the pre-trained lung segmentation model mentioned above is related to the well-studied lung segmentation task, which is totally different from the infection segmentation.

(2) **Robust COVID-19 Identification from a Multi-center Dataset of Chest CT Scans [4]:** First, an automated two-stage classification framework based on Capsule Networks is introduced as an extension of the COVID-FACT framework, which is tailored to robustly classify volumetric chest CT scans into one of the three target classes (COVID-19, CAP, or normal). The proposed framework integrates a scalable unsupervised enhancement approach to boost its performance and robustness in the presence of gaps between the train and test sets regarding types of scanners, imaging protocols, and technical parameters. More specifically, an enhancement approach is proposed to update the model's parameters by extracting confident predictions from different test sets and utilize them to re-train the model in order to increase its capability and robustness in the presence of gaps between the imaging protocols and patients' clinical history. In the proposed framework, different versions of the model are trained based on different test sets and outputs are combined to generate the final predictions, which are more accurate and robust. On the other hand, a unique test dataset, referred to as the SPGC-COVID dataset, is introduced to facilitate training and evaluation purposes. SPGC-COVID dataset consists of COVID-19, CAP, and normal cases acquired with various imaging settings from different medical centers, including images with different slice thickness, radiation dose, and noise level. In addition to different technical parameters, the dataset consists of CT scans of patients who have heart diseases or have undergone heart surgery, besides having COVID-19 or CAP infections.

(3) **WSO-CAPS [3]:** Generally speaking, ROI or slice selection module plays an important role in most automated diagnostic and prognostic frameworks (including COVID-FACT) as it facilitates the translation from the slice-level to the patient-level domain by detecting the candidate slices or ROIs demonstrating infection at the first step, and passing them to the

subsequent modules. In the proposed WSO-CAPS, a Window Setting Optimization (WSO) mechanism is introduced and incorporated into a slice-level Capsule Network-based classifier to boost the model's performance in detecting slices with the evidence of infection. The main focus of the proposed WSO-CAPS is to enhance the performance on the noisy chest LDCT and ULDCT images. However, the introduced technique has also been beneficial in the analysis of standard CT scans. Basically, to deal with low quality CT images, radiologists manually adjust the screen setting using some specific windowing functions to narrow down the displayed components and adjust the image contrast as some manifestations are only visible in a specific window depending on their tissue density which is commonly distributed from $HU_{air}(-1000)$ to > 4000 in the Hounsfield Units (HU) [6, 43]. This approach will also remove the undesired noises and artifacts in the image, facilitating its interpretation. Most windowing functions utilize mapping functions based on two parameters of Window Level (WL) and Window Width (WW) by which the function is determined. The WSO-CAPS framework is equipped with a mechanism to automatically identify the best (WL,WW) pairs to resemble the radiologists' efforts in reviewing such low quality scans. It is also hypothesized that due to the similarities in the imaging modality and reconstruction technology between various CT scans, augmenting other CNN and Capsule Network-based frameworks by the WSO module would have a high potential to improve the performance of diagnostic/prognostic frameworks which are working with other types of CT scans.

(4) **CAE-Transformer [5]:** This framework is proposed to predict the invasiveness of lung adenocarcinomas using volumetric non-thin CT scans. The building block of the CAE-Transformer is the novel self-attention mechanism and the transformer encoder. In addition, unlike current vision transformers which consider different patches in an image as a sequence of data [44, 45], the proposed CAE-Transformer uses a Convolutional Auto-Encoder (CAE) model [46] to extract informative features from CT slices and stack them to form a sequential feature map. The CAE is first pre-trained on the public LIDC-IDRI dataset, then fine-tuned on an in-house dataset. The obtained sequential feature maps are then fed to a transformer model containing multiple multi-head self-attention layers, followed by a stack of fully connected

layers to provide the final predictions. It is also worth noting that, unlike most existing studies which rely on the nodule patches as the model's input, the CAE-Transformer does not require a detailed annotation of the nodules and takes the whole CT image as the input. The only required information from the radiologists/experts is the set of slices with the evidence of a nodule without further details. Experimental results show that DL-based models improve the result achieved by the study performed in Reference [40] based on the histogram-based and radiomics features while the CAE-Transformer provided the highest improvement among its DL-based counterparts.

1.3 Thesis Organization

The rest of the thesis is organized as follows:

- Chapter 2 provides a literature review on the chest CT scan analysis. In addition, this chapter provides the background material required to follow developments presented in the reminder of the thesis. Furthermore, the detailed description of the datasets used in this thesis are presented in this chapter.

- Chapter 3 presents the proposed CT-CAPS and COVID-FACT frameworks developed for automatic identification of COVID-19 cases from standard dose CT scans.

- Chapter 4 presents the proposed robust three-way classification framework for diagnosis of COVID-19, CAP, and normal cases from a varied and multi-center dataset of CT scans. In addition, the proposed WSO-CAPS framework, which aims to enhance the slice-level predictions, is described in this chapter.

- Chapter 5 provides a detailed description of the proposed CAE-Transformer designed to predict the invasiveness of lung adenocarcinomas using non-thin CT scans and radiomics features.

- Chapter 6 concludes the thesis and explains some directions for future research studies.

Chapter 2

Literature Review and Background

As stated previously, there has been a recent surge of interest in development of Deep Learning-based diagnostic and prognostic tools based on chest CT scans. In this chapter, recent related research works proposed in literature are presented. Background materials, which are widely used throughout this thesis and required to follow the subsequent chapters are also provided. Finally, an overview of the datasets used in this thesis is presented.

2.1 COVID-19 Diagnosis

Recently, Convolutional Neural Networks (CNNs) have been widely used in several studies to develop DL-based frameworks based on medical images (e.g., CR and CT scans) in order to account for the human-centered weaknesses in detecting COVID-19. CNNs are powerful models in image-related tasks and are capable of extracting distinguishing features from CT scans and chest radiographs [47]. The study performed by Reference [48] is an example of the application of CNN in COVID-19 detection, where CNN is first pre-trained on the ImageNet dataset [49]. Fine-tuning is then performed using a CR dataset to distinguish normal, non-COVID pneumonia (viral and bacterial), and COVID-19 infections. Reference [50] has also explored the same problem, with the difference that the CNN is followed by a Support Vector Machine (SVM) to identify positive COVID-19 cases. Another study [51] proposed a CNN-based model utilizing depth-wise convolutions with varying dilation rates to extract more diversified features from chest radiographs. They used a pre-trained model

on a dataset of normal, viral, and bacterial pneumonia patients, followed by additional fine-tuned layers on a dataset of COVID-19 and other pneumonia patients. As stated previously, besides the studies based on the CR images, there has been a surge of interest in utilizing 2D and 3D CT images to identify COVID-19 infection. For instance, Reference [52] proposed a DenseNet-based model to classify manually selected slices with COVID-19 manifestations and pulmonary parenchyma into COVID-19 and normal classes. The underlying study achieved a satisfactory accuracy for the patient-level classification by averaging slice-level probabilities, followed by a threshold of 0.8 on the averaged values. However, the dataset used to train and test the model does not include other types of pneumonia in this study. In general, such methods require manual selection of slices demonstrating infection to feed the model, which makes the overall process time-consuming and only partially automated. To extract features from all CT slices, the proposed framework in Reference [7] first segmented the lung regions using a U-net based segmentation method [53], and then used them to fine-tune a ResNet50 model, which was pre-trained on natural images from the ImageNet dataset [54]. Extracted features are then combined using a max-pooling operation followed by a fully connected layer to generate probability scores for each disease type. Indeed, these types of methods combine extracted features from all slices of a patient, with or without infection, which potentially results in lower accuracy as there are numerous slices without evidence of infection in a volumetric CT scan of an infected patient. In another study [55], segmented lungs are fed into a multi-scale CNN-based classification model, which utilizes intermediate CNN layers to obtain three-way classification scores, and aggregates those scores generated by intermediate layers to make the final prediction. The model proposed in Reference [22] uses a two-stage method consisting of a Deeplabv3-based lung lesion segmentation model [56], followed by a 3D ResNet18 classification model [57] to identify lung lesions and abnormalities and use them to classify patients into COVID-19, CAP, and normal findings. They manually annotated chest CT scans into seven regions to train their lung segmentation model, which is a time-consuming and sophisticated task requiring a high level of thoracic radiology expertise to accomplish. It is worth noting that CT imaging is superior for COVID-19 detection and diagnosis purposes when compared to chest radiographs. However, as in the case of CT imaging, we are dealing with 3D inputs and several slices per patient (compared to one chest radiograph per patient), the learning process is significantly more challenging. As such, DL-based models trained

over CT scans cannot be directly compared with those developed based on chest radiographs.

From another viewpoint, existing diagnostic methods developed based on chest CT scans are generally divided into slice-level and patient-level methods. These studies can further be classified into segmentation-based or feature extraction-based approaches. Segmentation-based methods [22, 58, 59] aim to train a model on a large dataset of annotated lung lesions to detect regions of infection and determine the disease severity and type. Although the lung segmentation task has been well-studied [42], infection segmentation requires extensive collaboration with radiologists to perform the sophisticated infection and abnormality annotation task, making the training process too complicated and time-consuming. Moreover, in some cases [59], the overall performance is low for scenarios with mild lung infections. As an example of segmentation-based methods, Reference [58] used a semi-supervised method based on pre-trained existing segmentation models to detect lung infected regions to be incorporated into a CNN-based classifier via an attention mechanism to increase the classification accuracy. In another study, Reference [59] proposed a model which extracts handcrafted radiomics features from the segmented lung and infected regions, followed by a feature selection mechanism to feed multi-stage random forest classifiers to classify patients into four groups based on their infection size obtained from the first step. Then, a random forest model is trained for each group as the final classifier. The model developed in the aforementioned Reference [22] is also a case of segmentation-based models.

With regard to the feature extraction-based approaches, different frameworks have recently been introduced, commonly utilizing a CNN-based model. Such methods either use a 3D CNN to analyze the whole CT volume in a single stage or apply 2D CNNs on CT slices and aggregate slice-level results via an aggregation mechanism. As an example, the model proposed in Reference [60] fed a 3D CNN-based classifier with lung regions, segmented by a pre-trained U-Net [53] to classify COVID-19 and normal cases. Reference [55] extended patient-level labels into slice-level and used the same label for all slices in a CT scan to train a deep model, utilizing the intermediate CNN layers to obtain classification features. These features are then combined to make the final decision. It is worth mentioning that using patient-level labels for all slices in a CT scan is not reasonable and will add errors into the system as each volume of CT scan contains many slices without any visible infection area. As another example, the model proposed in the mentioned Reference [7] is indeed a

case of feature extraction-based approaches. The aforementioned methods either require a carefully annotated data to segment regions of infection or extend patient-level labels to all slices, resulting in unexplainable and potentially lower results. Moreover, some of the aforementioned works have only proposed slice-level classifiers, which makes such methods partially automated.

As mentioned before, CNN, which is widely adopted in COVID-19 studies, suffers from an important drawback that reduces its reliability in clinical practice. CNNs are required to be trained on different variations of the same object to fully capture the spatial relations and patterns. In other words, CNNs, commonly, fail to recognize an object when it is rotated or transformed. In practice, extensive data augmentation and/or adoption of huge data resources are needed to compensate for the lack of spatial interpretation. As COVID-19 is a relatively new phenomenon, large datasets are not easily accessible, especially due to strict privacy preserving constraints. Furthermore, most COVID-19 cases have been reported with a specific infection distribution in their image [11–14], which makes capturing spatial relations in the image highly important. As such, Capsule Networks, which is an alternative model capable of capturing spatial relations and being trained on small datasets, is used as the building block of the models proposed in this thesis. It is worth noting that the superiority of Capsule Networks over their counterparts has been recently shown in different medial image processing problems [16–21]. In the case of COVID-19, a Capsule Network-based framework [61], referred to as the COVID-CAPS, has been recently proposed to identify COVID-19 cases using CR images, which achieved an accuracy of 98.3%, a specificity of 98.6%, and a sensitivity of 80%. As stated previously, CT imaging is superior for COVID-19 detection and diagnosis purposes when compared to chest radiographs. However, as in the case of CT imaging, we are dealing with 3D inputs and several slices per patient (compared to one chest radiograph per patient), the learning process is significantly more challenging. As such, accuracies of deep models trained over CT scans cannot be directly compared with those obtained based on chest radiographs. The structure and mathematical representation of caspules are presented in Sub-section 2.1.1.

2.1.1 Capsule Networks

A Capsule Network (CapsNet) is an alternative architecture for CNNs with the advantage of capturing hierarchical and spatial relations between image instances. Each capsule layer utilizes

several capsules to determine existence probability and pose of image instances using an instantiation vector. The length of the vector represents the existence probability, and the orientation determines the pose. Each capsule i is made up of a set of neurons, which collectively create the instantiation vector $\boldsymbol{u}_i$ for the associated instance. Capsules in lower layers try to predict the output of capsules in higher levels using a trainable weight matrix $\boldsymbol{W}_{ij}$ as follows

$$\hat{\boldsymbol{u}}_{j|i} = \boldsymbol{W}_{ij}\boldsymbol{u}_i, \tag{2.1}$$

where $\hat{\boldsymbol{u}}_{j|i}$ is the predicted output of capsule j in the next layer by the capsule i in the lower layer. The association between the prediction $\hat{\boldsymbol{u}}_{j|i}$ and the actual output of capsule j, denoted by $\boldsymbol{v}_j$, is determined by taking the inner product of $\hat{\boldsymbol{u}}_{j|i}$ and $\boldsymbol{v}_j$. The higher the inner product, the more contribution of the lower level capsules to the higher level one. The contribution of capsule i to the output of the capsule j in the next layer is determined by a coupling coefficient c_{ij}, trained over a course of few iterations known as the "Routing by Agreement" given by

$$\begin{aligned}
a_{ij} &= \boldsymbol{v}_j.\hat{\boldsymbol{u}}_{j|i}, && (2.2)\\
b_{ij} &= b_{ij} + a_{ij}, && (2.3)\\
c_{ij} &= \frac{\exp(b_{ij})}{\sum_k \exp(b_{ik})}, && (2.4)\\
\boldsymbol{s}_j &= \sum_i c_{ij}\hat{\boldsymbol{u}}_{j|i}, && (2.5)\\
\text{and} \quad \boldsymbol{v}_j &= \frac{\|\boldsymbol{s}_j\|^2}{1+\|\boldsymbol{s}_j\|^2}\frac{\boldsymbol{s}_j}{\|\boldsymbol{s}_j\|}, && (2.6)
\end{aligned}$$

where a_{ij} is referred to as the agreement coefficient between the prediction and actual output, and b_{ij} denotes the log prior of the coupling coefficient c_{ij}. Vector $\boldsymbol{s}_j$ denotes the capsule output before applying the squashing function. As the length of output vectors represents probabilities, the ultimate output of capsule j ($\boldsymbol{v}_j$) is obtained by squashing $\boldsymbol{s}_j$ between 0 and 1 using the squashing function defined in Eq. (2.6). In order to update weight matrix $\boldsymbol{W}_{ij}$ through a backward training process, the loss function is calculated for each capsule k as follows

$$l_k = T_k \max(0, m^+ - ||\boldsymbol{v}_k||)^2 + \lambda(1-T_k)\max(0, ||\boldsymbol{v}_k|| - m^-)^2, \tag{2.7}$$

where T_k is 1 when the class k is present and 0 otherwise. m^+, m^-, and λ are hyper parameters of the model and are originally set to 0.9, 0.1, and 0.5, respectively. The overall loss is the summation of all losses calculated for all capsules.

In addition to the Capsule Networks, the Gradient-weighted Class Activation Mapping (Grad-CAM) localization approach [41] is widely utilized in this thesis to visualize the distinctive patterns in a chest CT scan recognized by the intermediate deep layers within the proposed frameworks. Using the Grad-CAM approach, the relation between the model's prediction and the features extracted by the intermediate layers (mainly convolutional layers) can be visually verified, which ultimately leads to a higher level of interpretability of the developed models. The detailed description of the Grad-CAM approach is presented in Sub-section 2.1.2.

2.1.2 Grad-CAM

Basically, the Grad-CAM's outcome is a weighted average of the feature maps of a convolutional layer, followed by a Rectified Linear Unit (ReLU) activation function, i.e.,

$$L^c_{Grad-CAM} = ReLU\left(\sum_k \alpha^c_k A^k\right), \tag{2.8}$$

where $L^c_{Grad-CAM}$ refers to the Grad-CAM's output for the target class c; α^c_k is the importance weight for the feature map k and the target class c, and; A^k refers to the feature map k of a convolutional layer. The weights α^c_k are obtained based on the gradients of the probability score of the target class with respect to an intermediate convolutional layer followed by a global average pooling function as follows

$$\alpha^c_k = \frac{1}{Z}\sum_i\sum_j \frac{\partial y^c}{\partial A^k_{ij}}, \tag{2.9}$$

where y^c is the prediction value (probability) for target class c, and Z refers to the total number of feature maps in the convolutional layer.

2.2 CT Window Setting Optimization

The majority of state-of-the-art frameworks are trained and evaluated using only standard-dose CT images, and few models have been developed based on LDCT and ULDCT so far. As an example of such models, an end-to-end framework is proposed in Reference [62] to predict the risk of lung cancer using 3D LDCT images. As another example, the framework developed in Reference [63] utilizes CNN and gradient boosting decision trees to predict the risk of lung cancer using LDCT images. However, no specific measure is considered in these studies to deal with the noisy and low-quality LDCT images. Recently, some research studies have incorporated a window setting optimization mechanism into the automated diagnostic/prognostic frameworks to improve the performance of the models [64, 65]. More specifically, the method proposed in Reference [64] utilizes a stochastic window tissue normalization mechanism that randomly samples windowing parameters (WL,WW) from two Gaussian distributions in the training phase to segment abdominal CT images. This method, however, does not consider an optimized setting and merely normalizes the windows using randomly sampled (WL,WW). In another study [65], the proposed model uses a stack of four CNN followed by two fully connected layers as the Window Estimator Module (WEM) along with an Inception-ResNet-v2 model [6] as the lesion classifier to detect the best window setting parameter for each 2D input image. It then considers the average of all obtained (WL,WW) values from the entire dataset as the final setting. Their proposed method using a combination of several window settings could improve the accuracy of the multi-class intracranial hemorrhage detection from 87.65% to 88.35% and the binary classification (normal and abnormal) from 95.59% to 96.43% using brain CT images. The WEM mechanism proposed in this study resulted in a wide distribution of (WL,WW) values calculated for each slice, and an average function over the entire dataset might not be the optimized value. It is worthy of note that none of the aforementioned algorithms, which aim to adjust windowing parameters, were developed based on the LDCT and ULDCT scans. In a recent study [66], a Window Setting Optimization (WSO) mechanism is proposed which uses a single convolutional layer at the beginning of the pipeline to map the full-range DICOM images to the range of interest using specific windowing functions. In this thesis, the WSO mechanism introduced in [66] is adapted to detect slices demonstrating infection in an in-house dataset of LDCT

and ULDCT acquired from COVID-19 and normal cases, as well as simulated low dose images of CAP cases. More specifically, a multi-window framework, referred to as the WSO-CAPS, is proposed which applies a windowing function similar to those used by radiologist's monitors on the full-range DICOM images and passes the modified images to a classifier based on the Capsule Networks [15].

2.3 Lung Cancer Invasiveness Prediction

In general, existing publications on the SSN classification and invasiveness assessment can be categorized into two main groups: (1) Radiomics-based and (2) Deep Learning-based frameworks [67]. In the former, a set of histogram-based, morphological, and clinical features are extracted from the CT images which are then analyzed using statistical or machine learning techniques such as the studies conducted in References [68,69]. As another example of such frameworks, a histogram-based model is developed in Reference [40] to predict the invasiveness of primary adenocarcinoma SSNs from non-thin CT scans of 109 pathologically labeled SSNs. In this study, a set of histogram-based and morphological features along with additional features extracted via the Functional Principal Component Analysis (FPCA) is fed to a linear logistic regression, achieving the accuracy of 81.0% and Area Under the ROC Curve (AUC) of 0.91. Deep learning-based frameworks, on the other hand, extract informative and discriminative features in an automated fashion. Existing deep models working with volumetric CT scans can be classified into two main groups: (*i*) The first approach is to feed the whole volume of images (i.e., all 2D slices) or stack of all nodule patches (cropped images including nodules) into a 3D model (e.g., 3D CNN) to provide a patient-level prediction [70, 71]. Processing a large 3D CT scan at once, however, demands more complex models, more computational resources, and larger training datasets. (*ii*) The second approach, on the other hand, analyzes individual 2D CT slices or Regions of Interest (ROIs) in the first step and aggregates the results through a sequential model such as Recurrent Neural Networks (RNN) or LSTM or via another aggregation mechanism based on pooling or fully connected layers [1, 2, 72]. It is also worth noting that most of the published studies are developed and evaluated based on the public LIDC-IDRI [73] dataset which does not have pathologically proven labels and focuses more on nodule detection than

classification.

Due to the nature of the volumetric CT scans, which utilize a sequence of 2D images (slices) to provide a detailed representation of the body, there have been recently a surge of interest in application of sequential deep models for diagnostic/prognostic tasks based on CT scans. Recently, a new sequential deep model based on a novel self-attention mechanism, commonly known as "Transformer" [74], has been proposed which shows superior performances in the tasks related to the sequential data. Transformer models benefit from a novel self-attention mechanism which is capable of capturing global context and dependencies between instances in sequential data while requiring far less computational resources compared to conventional LSTM and RNN architectures. Transformers are also superior to their counterparts in terms of parallelization and dynamic attention. Although the transformer model was initially designed for Natural Language Processing, there have been recently significant attempts to adopt the self-attention mechanism for image processing applications. Vision Transformer (ViT) [44] and Convolutional Vision Transformer (CvT) [45] are two popular types of transformers designed to address image processing tasks. Both models, however, apply the self-attention to the small patches in a 2D image. Analyzing a series of CT slices, however, requires a framework capable of capturing inter-slice relations. Although the development of transformers for sequential medical images is currently in its nascent stage, recent models proposed for COVID-19 disease identification and image segmentation [75–77] have shown promising results and potentials. In this thesis, the self-attention mechanism and the transformer encoder utilized in References [44,74] are modified to be compatible with the task at hand. In particular, the sequential input is provided by concatenating feature maps generated from each slice.

It is also worth noting that, unlike most existing studies which rely on the nodule patches as the model's input, the CAE-Transformer does not require a detailed annotation of the nodules and takes the whole CT image as the input. The only required information from the radiologists/experts is the set of slices with the evidence of a nodule without further details.

The multi-head self-attention mechanism, which is the building block of the proposed transformer-based framework, is described in Sub-section 2.3.1.

2.3.1 Multi-Head Self-Attention Mechanism

The transformer model is the building block of the CAE-Transformer framework which uses a novel self-attention mechanism to capture global dependencies among various instances in the input sequence with a high parallelization capability, reducing the computational complexity and memory allocation of other recurrent-based architectures such as RNN and LSTM. The self-attention mechanism is based on a Scaled Dot-Product Attention function, mapping a query and a set of key-value pairs to an output, where the query (Q), keys (K), values (V), are learnable representative vectors for the instances in the input sequence with dimensions d_k, d_k, and d_v, respectively. The output of a self-attention module is computed as a weighted average of the values, where the weight assigned to each value is computed by a similarity function of the query and the corresponding key after applying a softmax function [74]. More specifically, the attention values on a set of queries are computed simultaneously, packed together into a matrix Q. The keys and values are similarly represented by matrices K and V. The output of the attention Scaled Dot-Product Attention function is computed as

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{2.10}$$

where K^T is the transpose of the matrix K. It is also beneficial to linearly project the queries, keys, and values h times with various learnable linear projections to vectors with d_k, d_k and d_v dimensions, respectively, before applying the attention function. On each of the projected versions of queries, keys, and values, the attention function is performed in parallel, resulting in $d_v - dimensional$ output values. These values are then concatenated and once again linearly projected via a fully-connected layer. This process is called “Multi-Head Attention (MHA)" which helps the model to jointly attend to information from different representation sub-spaces at different positions [74]. The output of the MHA module is

$$\begin{aligned} MHA(Q, K, V) &= Concat(head_1, \cdots, head_h)W^O, \\ head_i &= Attention(QW_i^Q, KW_i^K, VW_i^V), \end{aligned} \tag{2.11}$$

where the projections are achieved by parameter matrices $W_i^Q \in \mathbb{R}^{d_{model} \times d_k}$, $W_i^K \in \mathbb{R}^{d_{model} \times d_k}$,$W_i^V \in \mathbb{R}^{d_{model} \times d_v}$, and $W^O \in \mathbb{R}^{hd_v \times d_{model}}$.

2.4 Datasets

In this section, the datasets used for model development and evaluation in this thesis are described in detail. Furthermore, additional information on the imaging protocol, de-identification, and labeling process is provided.

2.4.1 COVID-CT-MD Dataset

COVID-CT-MD [78] includes volumetric chest CT scans of 171 patients tested positive for COVID-19 infection, 60 CAP patients, and 76 normal patients, acquired from April 2018 to May 2020. The average age of patients is 50 ± 16 including 183 men and 124 women. A subset of 55 COVID-19 and 25 CAP cases in the COVID-CT-MD is analyzed by three experienced radiologists to detect slices with a distinctive evidence of infection. More specifically, the patient-level labeling has been performed by all three radiologists, and majority voting is adopted for the final assignment. For the purpose of slice-level labeling, given the limited time and complexities of the slice-level annotation, one radiologist has provided the slice-level labels. However, an initial analysis is performed by other radiologists over a subset of 14 patients to confirm the inter-reader agreement. The labeled subset of the data contains $4,993$ slices demonstrating infection and $18,416$ slices without evidence of infection. Sample CT slices with and without an evidence of infection in one COVID-19 and one CAP case are shown in Figure 2.1. This labeled data is then randomly divided into three groups, including 60%, 10%, and 30% independent parts of the data, to train, validate, and test the developed slice-level models. The remaining data is split with the same proportion and used along with the labeled data to train and evaluate the patient-level classifiers. This data leakage between the train and test sets has been prevented. In other words, all slices related to a patient are included either in the train or the test dataset. The data collection work is performed based on the policy certification number 30013394 of Ethical acceptability for secondary use of medical data approved by Concordia University. The COVID-CT-MD dataset is available online through

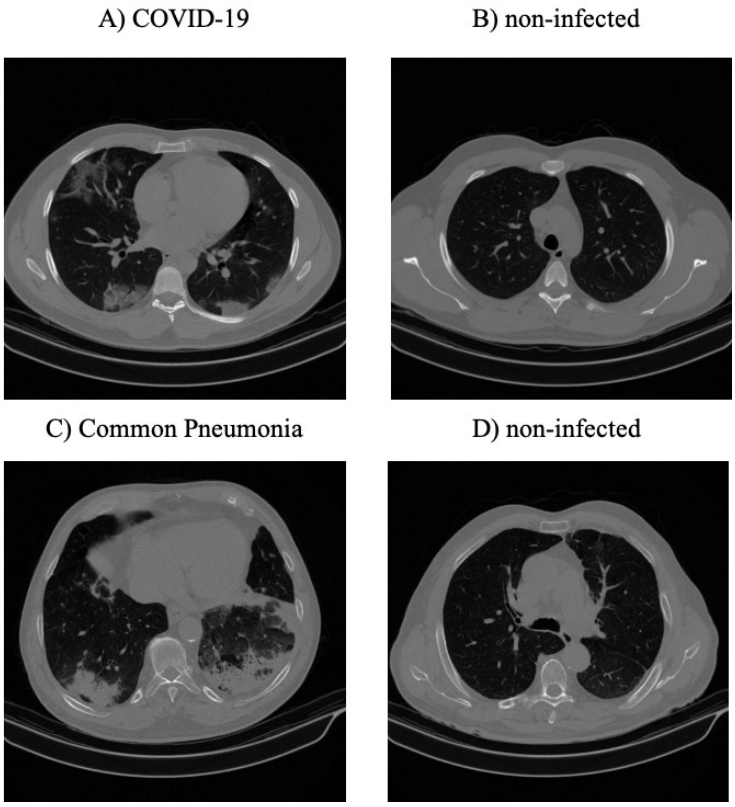


Figure 2.1: A, B: Infected and non-infected sample CT slices in a COVID-19 case; C, D: Infected and non-infected sample CT slices in a non-COVID Pneumonia (CAP) case.

Figshare [1]. Furthermore, informed consent is obtained from all the patients. Finally, the dataset is complied with the Digital Imaging and Communications in Medicine (DICOM) supplement 142 (Clinical Trial De-identification Profiles) [79], indicating that all CT studies are de-identified by either removing or obfuscating the patient and center-related information such as names, User Identifiers (UIDs), dates, times, and comments based on the directions specified in DICOM supplement 142 [79].

Labelling Process

Diagnosis of COVID-19 infection is based on positive RT-PCR test results, clinical findings, epidemiology data, and CT scan COVID-19 manifestations by three experienced thoracic radiologist. CAP and normal cases were included from another study, and the diagnosis was confirmed using clinical parameters, and CT scans. The labeling process aims to specify slices with distinctive disease

[1] https://figshare.com/s/c20215f3d42c98f09ad0

Table 2.1: Imaging device and acquisition settings used to acquire the COVID-CT-MD dataset.

Scanner Manufacturer and Model	Slice Thickness (mm)	Image Type	kVP (kV)	Exposure Time (ms)	Reconstruction Matrix	Radiation Level
SIEMENS, SOMATOM Scope	2	Axial	110	600	512×512	Standard

manifestations in a timely manner rather than those with minimal findings.

Imaging Protocol

All CT examinations have been acquired using a single CT scanner with the same acquisition setting and technical parameters, which are presented in Table 2.1, where kVP (kiloVoltage Peak) and Exposure Time affect the radiation exposure dose, while Slice Thickness and Reconstruction Matrix represent the axial resolution and output size of the images, respectively [80].

2.4.2 SPGC-COVID Dataset

In what follows, different datasets used to construct the SPGC-COVID dataset [81] are described individually, followed by supplementary information about the demographic data, imaging protocols, acquisition settings, de-identification, and the labeling process. The so-called SPGC-COVID dataset is comprised of four different sets, each with specific characteristics to evaluate the robustness and generalizability of DL-based models from different aspects. As the SPGC-COVID dataset is primarily used as the test set in the experiments conducted in this research work, the four components are referred to as different test sets, which contain 51 COVID-19, 28 CAP, and 51 normal cases in total. The SPGC-COVID dataset is publicly available on Figshare [2]. An overview of different datasets and imaging centers is visualized in Fig. 2.2 and different components of this dataset are described as follows

- **Test Set 1:** Low-Dose and Ultra-Low-Dose CT scans of COVID-19 and normal cases acquired from the same imaging center as that of the COVID-CT-MD dataset. This dataset is a subset of an in-house dataset of LDCT and ULDCT [82] and is publicly available.

- **Test Set 2:** CT scans of COVID-19, CAP, and normal cases acquired in a different imaging

[2] https://figshare.com/articles/dataset/SPGC-COVID_Dataset/16632397

Table 2.2: Number of cases, demographic data, and acquisition information for train and test sets in the SPGC-COVID dataset.

Dataset	COVID-19	CAP	Normal	Age (Mean± SD)	Gender	Imaging Center
Train	171	60	76	50.78 ± 16.84	$183M/124F$	1
Test 1	15	0	15	40.97 ± 14.38	$19M/11F$	1
Test 2	10	10	10	61.00 ± 13.39	$25M/5F$	2
Test 3	10	10	10	46.77 ± 20.89	$15M/15F$	1
Test 4	16	8	16	46.23 ± 14.74	$25M/15F$	Both

center (Tehran Heart Center, Iran) using the "SIEMENS SOMATOM Emotion 16" scanner and different scanning parameters. Some cases in this dataset have additional history of cardiovascular disease/surgeries with specific CT imaging findings, which are not available in the other datasets used to train or evaluate the proposed models.

- **Test Set 3:** CT scans of COVID-19, CAP, and normal cases obtained by the same scanner and scanning protocol used to acquire the COVID-CT-MD dataset. Cases in this test set are not included in the COVID-CT-MD dataset.

- **Test Set 4:** A combination of new CT scans of all three categories (i.e., COVID-19, CAP, Normal) obtained from the same centers as those of Test set 1 and 2, using the same acquisition settings and scanners.

Additional statistical and demographic information about different test sets in the SPGC-COVID dataset is provided in Table 2.2. In Table 2.2, Center 1 represents the Babak Imaging Center and Center 2 is the Tehran Heart Center. Both imaging centers are located in Tehran, Iran and use the Filtered Back Projection reconstruction method [83] to obtain the CT images. Some sample CT slices from the first three test sets are shown in Fig. 2.3. The important technical parameters that contribute the most to the image quality and characteristics of the acquired CT scans are presented in Table 2.3.

Labeling Process

Diagnosis of the cases scanned in Center 1 is obtained by finding the consensus between three experienced radiologists who have considered the following three main criteria (similar to those of

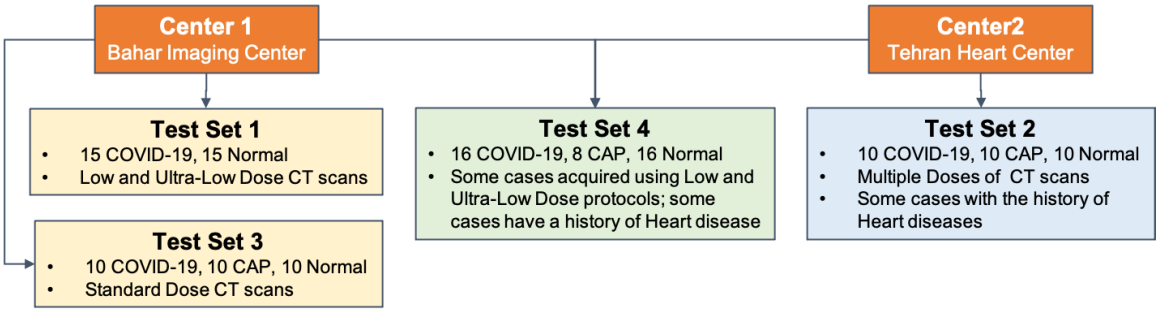


Figure 2.2: Overview of the SPGC-COVID dataset.

Table 2.3: Acquisition parameters used to obtain each test set of the SPGC-COVID dataset.

Dataset	Slice Thickness (mm)	Reference Exposure (mAs)	kVp (kV)	Radiation (mSv)	Number of Slices (per patient)
Test 1	2	$15-20$	110	$\sim 0.3-1.5$	$126-169$
Test 2	$1.5-5$	25	$110-130$	~ 2	$53-221$
Test 3	2	50	$100-110$	~ 7	$115-183$
Test 4	$1.5-6$	$15-25$	$110-130$	$\sim 0.3-2$	$52-224$

the COVID-CT-MD dataset) to label the data:

(i) RT-PCR test (if available);

(ii) Imaging findings including Ground Glass Opacities (GGOs), consolidations, crazy paving pattern, bilateral and multifocal lung involvement, peripheral distribution, and lower lobe predominance of findings;

(iii) Clinical symptoms of the COVID-19 infection, and;

(iv) Epidemiology.

For the cases acquired from Center 2, $(13/18)$ COVID-19 cases have positive RT-PCR test results, and the remaining cases have been labeled by one experienced radiologist following the same aforementioned criteria. The SPGC-COVID dataset complies with the DICOM supplement 142 (Clinical Trial De-identification Profiles), which ensures that all personal information is removed or obfuscated. Some demographic and acquisition attributes related to the patients' gender and age, scanner type, and image acquisition settings have been preserved to provide useful information about the dataset. It is also worth noting that the SPGC-COVID dataset was used as the test set in the 2021 Signal Processing Grand Challenge (SPGC) on COVID-19 diagnosis, which was organized as part of

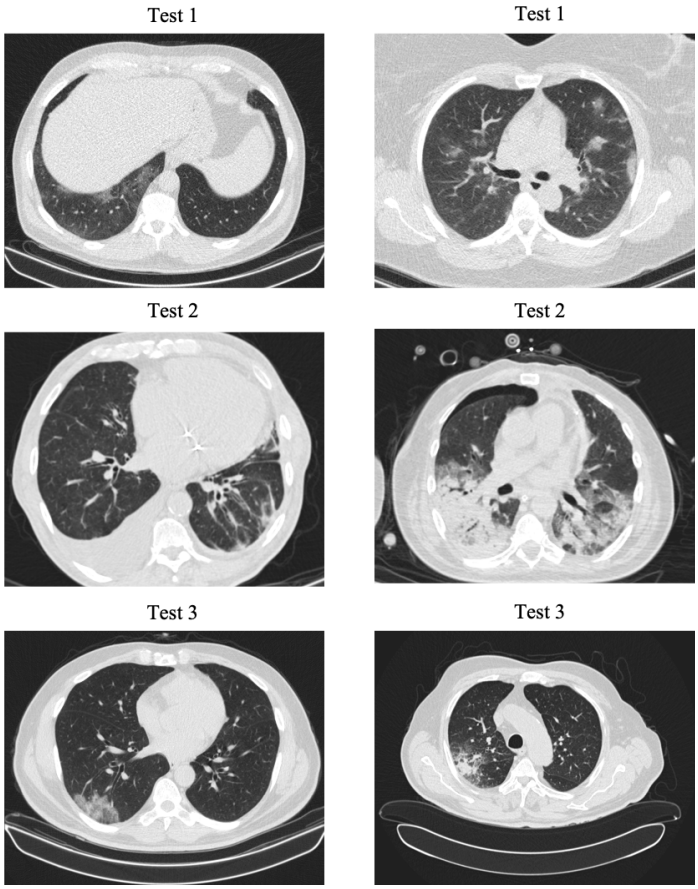


Figure 2.3: Sample CT slices from the first three test sets of the SPGC-COVID dataset.

the 2021 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).

2.4.3 WSO Dataset

This dataset is used for training and evaluation of the WSO-CAPS framework, introduced in Chapter 4. In the following subsections, the three subsets of the dataset along with a brief description of the acquisition protocols and annotation process by three experienced thoracic radiologists are described.

- **Low-Dose and Ultra-Low-Dose CT scans**

  100 COVID-19 and 60 normal volumetric CT scans are collected from the imaging center used to acquire the COVID-MD-CT dataset. The acquired scans are reconstructed using the Filtered Back Projection method [83]. The radiation dose in standard chest CT scans is estimated at

$7mSv$, which is reduced to $1-1.5mSv$ in LDCT scans and as low as $0.3mSv$ in the ULDCT ones. LDCT images are acquired from patients with $>60kg$ bodyweight using the mAs value (X-ray tube current $\times$ slice scanning time) of 20, kVp value (X-ray tube kilovoltage peak) of $110v$, and the slice thickness of $2mm$, while the ULDCT images are obtained from patients with the bodyweight of less than $60kg$, and the $15mAs$ has been used to acquire the scans. This subset contains $7,703$ slices demonstrating infection and $15,464$ slices without the evidence of infection. Similar to the aforementioned datasets, the labeling process was performed by three experienced thoracic radiologists and the majority voting was adopted to determine the final label.

- **Simulated Low Dose CT scans**

  Since low-dose CAP CT scans were not easily accessible, they are simulated using the standard dose ones available in the COVID-CT-MD dataset. Most of the image simulation techniques are based on paired images, which in this case means paired standard and low dose images that exactly correspond to each other. As collecting paired CT scans is not feasible for the problems at hand, an unsupervised image-to-image translation technique, referred to as CycleGAN [84] was adopted. This model, essentially, consists of two sets of generators and discriminators, where the first set converts standard-dose images to low-dose ones. Consequently, the second set transfers the generated low dose images back to standard dose ones, and the output is compared with the original source image, forming the main term of the loss function. Using this technique and taking the output of the first generator, standard dose CAP images are converted to low dose ones. The dataset contains 60 simulated low dose CAP cases and is used along with the COVID-19 and normal ones to train and evaluate the WSO-CAPS model. This dataset contains $3,359$ slices demonstrating infection and $5,768$ slices without evidence of infection.

- **Standard Dose CT scan**

  This subset is basically the same as the COVID-CT-MD dataset. However, slice-level labels for 35 more CAP cases are also provided to expand the labeled slices, providing $7,138$ slices demonstrating infection and $21,442$ slices without evidence of infection.

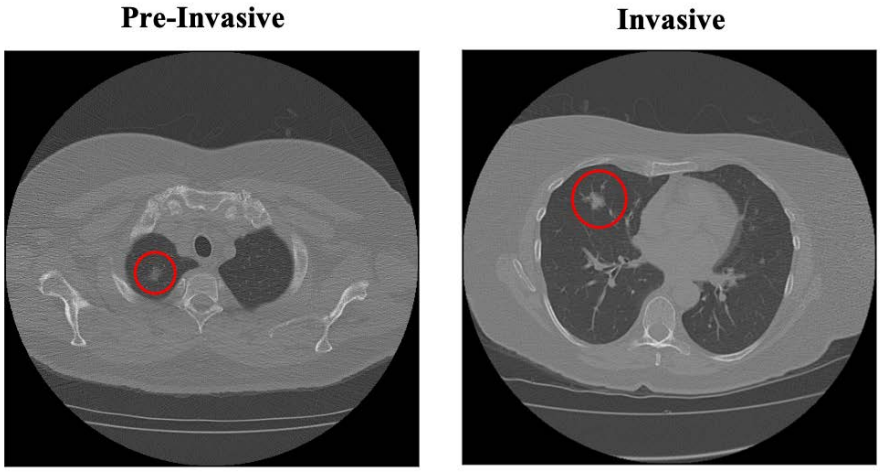


Figure 2.4: Sample pre-invasive and invasive lung adenocarcinomas.

2.4.4 LUAC Dataset

We have utilized the same dataset used in [40] with an additional five nodules from the same institution to train and evaluate the model.

This dataset contains the volumetric chest CT scans initially introduced in Reference [40]. In addition, to further balance the dataset, five additional cases which are acquired from the same institution are added to the original dataset. This dataset contains volumetric CT scans of 114 pathologically proven SSNs, segmented and reviewed by 2 experienced thoracic radiologists. All SSN labels are provided after surgical resections. SSNs are initially classified into three categories of pre-invasive lesions including Atypical Adenomatous Hyperplasia (AAH) and Adenocarcinoma In Situ (AIS), Minimally Invasive Adenocarcinoma (MIA), and Invasive Pulmonary Adenocarcinoma (IPA). Following the original study [40], the first two categories are grouped to represent the pre-invasive and minimally invasive class with 58 cases, and the invasive nodules are kept as the other class including 56 cases. In addition to the nodule labels, the CT slices with the evidence of a nodule are also determined by the radiologists, facilitating the development of DL-based frameworks. Fig. 2.4 shows two sample lung adenocarcinomas from the LUAC dataset.

2.5 Summary

In this chapter, an overview of the existing solutions proposed in the literature is provided along with a detailed description of the background material and datasets used in this thesis. The limitations and potentials of the existing AI-based solutions for the tasks at hand are also discussed in this chapter. In summary, the existing solutions either require a sophisticated data collection and training process or are incapable of extracting informative and robust features from a small dataset. Furthermore, this chapter highlights the recent surges of interest in the context of CT scan acquisition and analysis, as well as the areas for which few comprehensive studies have been performed. Besides the literature review, the details of the Capsule Networks and Multi-Head Self-Attention Mechanism, which are the building blocks of the frameworks proposed in this thesis, are provided. Finally, an overview of four in-house datasets used to train and test the proposed frameworks is presented.

Chapter 3

Fully Automated COVID-19 Identification from Chest CT Scans

The newly discovered coronavirus disease (COVID-19) has been globally spreading and causing hundreds of thousands of deaths around the world as of its first emergence in late 2019. The rapid outbreak of this disease has overwhelmed healthcare infrastructures and arisen the need to allocate medical equipment and resources more efficiently. The early diagnosis of this disease will lead to the rapid separation of COVID-19 and non-COVID cases and helps healthcare authorities devise efficient resource allocation plans. In this regard, a growing number of studies are investigating the potential of DL-based approaches in the early diagnosis of COVID-19 from medical images. CT scans have recently shown distinctive features and higher sensitivity compared to other diagnostic COVID-19 tests, in particular the current gold standard, i.e., the RT-PCR test. Current DL-based models are mainly developed based on CNNs to identify COVID-19 pneumonia cases from medical images. CNNs, however, require extensive data augmentation and large datasets to identify detailed spatial relations between image instances. Furthermore, most existing algorithms developed based on CT scans either extend slice-level predictions to patient-level ones using a simple thresholding mechanism or rely on a sophisticated infection segmentation to identify the disease.

In this chapter, two Capsule Network-based frameworks, referred to as the “CT-CAPS" and “COVID-FACT” respectively, are proposed to automatically diagnose COVID-19 positive cases

from volumetric CT scans. Both proposed frameworks utilize Capsule Networks, as their main building block. Therefore, they are capable of addressing the failure of the commonly used CNN architectures [47] in recognizing spatial relations between objects in an image and thus eliminating the need for large labeled datasets to reach a satisfying result. CT-CAPS is a fully-automated framework for the identification of COVID-19 positive cases from volumetric chest CT scans. In particular, to be independent of sophisticated segmentation of the area of infection, it automatically extracts distinguishing features from 2D chest CT images in its first stage, which are then leveraged to differentiate COVID-19 from non-COVID cases in the second stage. More specifically, the obtained slice-level features are extracted from the penultimate capsule layer in the model of the first stage. The experiments on the COVID-CT-MD dataset, described in Sub-section 2.4.1, show the state-of-the-art performance with the accuracy of 89.8%, sensitivity of 94.5%, specificity of 83.7%, and Area Under the ROC Curve (AUC) of 0.93. The second automated framework, COVID-FACT, has a modified and improved pipeline, particularly in the second stage. Similar to the CT-CAPS, it is a two-stage fully-automated Capsule Network-based framework for identification of COVID-19 from chest CT scans. Unlike the CT-CAPS' approach for translation from slice-level to patient-level diagnosis, in COVID-FACT, slices demonstrating infection are detected at the first stage and are fed to the second stage, which is responsible for classifying patients into COVID-19 and non-COVID cases. Based on the experiments, COVID-FACT achieves an accuracy of 90.82%, a sensitivity of 94.55%, a specificity of 86.04%, and an AUC of 0.98 on the same COVID-CT-MD dataset. It is worth mentioning that the development of the CT-CAPS and COVID-FACT frameworks depends on far less supervision and annotation in comparison to their counterparts.

The remainder of the chapter is organized as follows: First, the details of the proposed CT-CAPS framework and the obtained results are presented. Next, the structure of the COVID-FACT framework is explained, and the related experimental results and model evaluations are provided. Then, the obtained results and possible sources of the error are investigated. Finally, a brief summary of the chapter is presented.

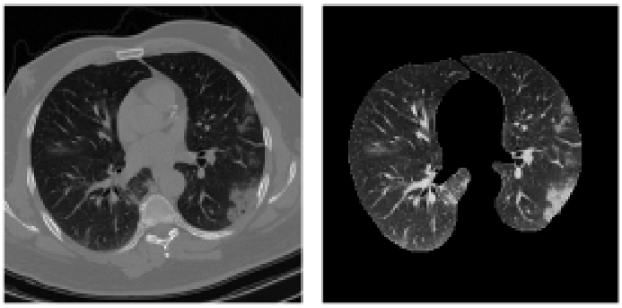


Figure 3.1: The original CT image and the corresponding lung region segmented by the R231CovidWeb model.

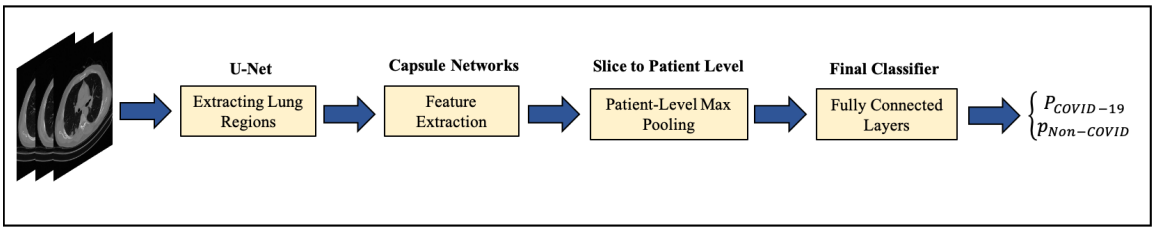


Figure 3.2: The pipeline of the CT-CAPS framework proposed to identify COVID-19 and non-COVID cases from chest CT scans.

3.1 CT-CAPS Framework

In this section, the detailed description of the CT-CAPS's main components and the processing pipeline is presented.

3.1.1 Lung Segmentation

In order to remove uninformative components and unwanted artifacts (e.g., metallic artifacts) in a CT scan, a pre-trained U-Net-based lung region segmentation model [42], referred to as the "U-net (R231CovidWeb)", is utilized which has been fine-tuned specifically on the COVID-19 images. A sample of lung region extracted by this model is illustrated in Fig. 3.1. It is worth mentioning that unlike segmenting infected regions, lung region segmentation is a well-studied topic and highly efficient models have been introduced so far. The input of the R231CovidWeb model is a CT scan with the original slice size of 512×512. The model returns the extracted lung tissues, which will further go through some normalization and resizing steps. More specifically, the output images will be normalized between 0 and 1 to help the generalizability and effective convergence of the

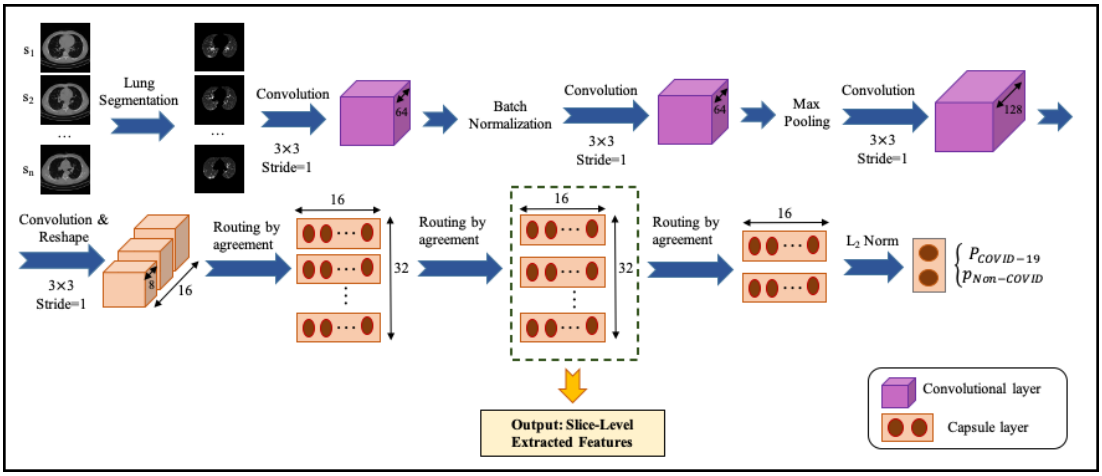


Figure 3.3: The Capsule Network-based model developed to extract slice-level features from chest CT scans.

model. Following the literature [22, 55], the output images are down-sampled from the original 512×512 size to 256×256 to reduce the complexity and memory requirements with negligible loss of information. Finally, slices without visible lung tissues are excluded and the remaining ones are saved to be used in the CT-CAPS framework.

3.1.2 CT-CAPS Architecture

The CT-CAPS' pipeline is illustrated in Fig. 3.2. The first stage of the CT-CAPS framework consists of a 2D Capsule Network, which aims to classify 2D CT slices into COVID-19 and non-COVID images and provide compressed feature maps for each image to be used in the next stage. As shown in Fig. 3.3, the model in the first stage consists of four convolutional and three capsule layers. The first and second layers are convolutional ones, followed by a batch-normalization layer. Similarly, the third and fourth layers are convolutional followed by a max-pooling layer. The fourth layer, referred to as the primary capsule layer, is reshaped to form the desired primary capsules. Afterwards, three capsule layers perform sequential routing by agreement processes, defined in Sub-section 2.1.1, to extract deeper features and spatial patterns. Finally, the two capsule in the last layer represent the two classes of infected and non-infected slices. More specifically, the length of each capsule represents the probability of the input image belonging to the corresponding target class. In the next step, slice-level features extracted from intermediate capsule layers of the described network are aggregated to move on to the patient-level domain. In this regard, the penultimate capsule layer

is used as the representative feature map of the CT slices, and a global max pooling layer is then applied to the set of feature maps generated for a patient (each corresponding to a single slice). Using this feature selection and max pooling mechanism, each 3D volumetric CT scan is represented by a small 32×16 matrix. Experimental results, presented in Section 3.2, demonstrate the ability of the obtained feature maps to efficiently distinguish between COVID-19 and non-COVID images. The results obtained in the first stage (i.e., the output of the max pooling layer) are then fed to a stack of four fully connected layers with the size of 256, 128, 32, and 2 respectively. In addition, the loss function is modified to compensate for the relatively imbalanced training dataset. More specifically, a weighted version of the loss function is used such that a higher penalty rate is given to the less frequent class, which is COVID-19 in this case. For the fully connected layers, however, the class weights are equal. The loss function of the Capsule Network model is modified as follows:

$$loss = \frac{N^+}{N^+ + N^-} \times loss^- + \frac{N^-}{N^+ + N^-} \times loss^+, \tag{3.1}$$

where N^+ represents the number of COVID-19 samples, N^- is the number of non-COVID samples, $loss^+$ denotes the loss value associated with COVID-19 samples, and $loss^-$ is the loss value associated with non-COVID cases.

3.2 CT-CAPS' Experimental Results

The feature extraction part of the CT-CAPS is trained on a subset of the COVID-CT-MD dataset, described earlier in Sub-section 2.4.1, for which slice-level labels are available. This subset contains 55 COVID-19, 25 CAP, and the entire 78 normal cases. The Adam optimizer with an initial learning rate of $1e-4$, batch size of 16, and 100 epochs is used to train the model in stage 1. For the fully connected patient-level classifier, the initial learning rate of $1e-3$, and 500 epochs are used. In each stage, the model with the lowest loss value on the validation set is considered as the final model for evaluation. The evaluation results on the COVID-CT-MD dataset are presented in Table 3.1. The testing set used in this study contains 53 COVID-19 and 43 non-COVID cases (including 19 CAP and 24 normal cases). CT-CAPS is compared with its duplicate but using the whole CT images without extracting the lung tissues. In another experiment, capsule layers are replaced by two fully

Table 3.1: CT-CAPS's Patient-Level Classification Results.

Performance	CT-CAPS	CT-CAPS (no lung)	CT-CNN	CT-Res50
Accuracy	**89.8%**	82.6%	78.6%	81.6%
Sensitivity	94.5%	87.3%	87.3%	**96.4%**
Specificity	**83.7%**	76.6%	67.4%	62.8%
AUC	**0.93**	0.86	0.79	0.82
# Params.	**0.5M**	**0.5M**	$243.9M$	$24M$

connected layers with the size of 128, while the rest of the architecture and parameters are kept the same. The fully connected dense layer before the last layer is then taken as the new feature map to make a CNN-based alternative model for the comparison, referred to as the CT-CNN. In a similar experiment, Resnet50, which is the backbone of many similar works such as the model proposed in Reference [7], is used in the feature extraction stage. In this case, similar to Reference [7], the fully connected layer with $2,048$ neurons before the last layer is taken as the feature map, followed by the same max pooling aggregation mechanism. The comparison results are presented in Table 3.1. As shown in this table, the CT-CAPS framework achieves the accuracy of 89.8%, high sensitivity of 94.5%, specificity of 83.7%, and AUC of 0.93 using the default probability threshold of 0.5. Table 3.1 also implies that due to the lower complexity of the CT-CAPS, fewer training parameters are required, which in turn significantly improve the training time.

It is worth mentioning that the main concern in clinical practice is to have a high sensitivity in identifying COVID-19 positive patients, even if the specificity is not very high. As such, the classification cut-off probability can be modified by physicians using the Receiver Operating Characteristic (ROC) curve in order to provide a desired balance between the sensitivity and the specificity (e.g., having a high sensitivity while the specificity is also satisfying). In other words, physicians can decide how much certainty is required to consider a CT scan as a COVID-19 positive case. By choosing a cut-off value greater than 0.5, the CAP cases that contain highly overlapped features with COVID-19 cases can be excluded. On the other hand, by selecting a lower cut-off value, more cases will be allowed to be identified as a COVID-19 case. Experimental results show that increasing the probability threshold from 0.5 to 0.6 improves the accuracy to 90.8%, and the specificity to 86.0% while the sensitivity remains the same. Table 3.2 presents the performance of the proposed CT-CAPS using different cut-off probabilities.

Table 3.2: Performance of the CT-CAPS using different cut-off probabilities.

Cut-off Probability	Accuracy	Sensitivity	Specificity
0.3	86.7%	94.5%	76.7%
0.4	88.8%	94.5%	81.4%
0.5	89.8%	94.5%	83.7%
0.6	90.8%	94.5%	86.0%
0.7	89.8%	90.9%	88.4%

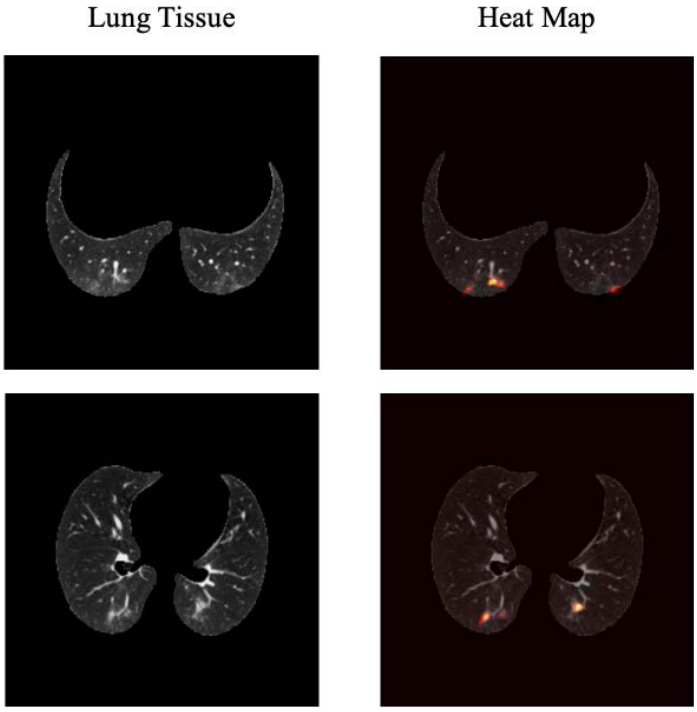


Figure 3.4: The heat maps generated by the GRAD-CAM localization approach from the last convolutional layer of the CT-CAPS framework for two sample images with COVID-19-related evidences of infection.

In addition to the aforementioned numerical results, the Grad-CAM localization approach, described in Sub-section 2.1.2, is utilized to visualize the distinctive patterns in a chest CT scan recognized by the last convolutional layer in the CT-CAPS. Fig. 3.2 illustrates the recognized abnormal regions for two lung samples containing small evidences of COVID-19 infection. In these two examples, it can be observed that the model correctly identified the regions of infection that had the highest contribution to the final decision.

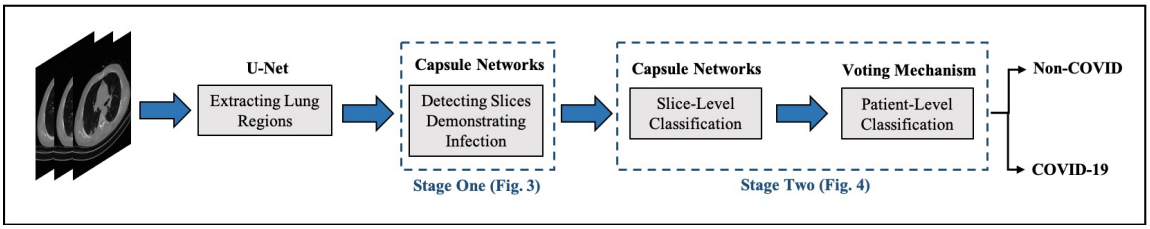


Figure 3.5: The two-stage architecture of the proposed COVID-FACT.

3.3 COVID-FACT Framework

In this section, different components of the proposed COVID-FACT are explained. The overall architecture of the COVID-FACT is illustrated in Fig. 3.5, which consists of a lung segmentation model at the beginning, followed by two capsule network-based models and an average voting mechanism coupled with a thresholding approach to generate patient-level classification results. The lung segmentation model used in COVID-FACT is the same as that of the CT-CAPS described in Sub-section 3.1.1. In addition, similar to the proposed CT-CAPS, the Grad-CAM localization approach is incorporated into the model to highlight important components of a chest CT scan that contribute the most to the final decision.

3.3.1 COVID-FACT's Stage One:

The first stage of the COVID-FACT, shown in Fig. 3.3.1, is adopted from stage 1 in CT-CAPS which is essentially responsible for identifying slices demonstrating infection (caused by COVID-19 or CAP). Using this stage, slices without any evidence of infection are discarded, and only the ones demonstrating infection are focused. Intuitively speaking, this process is similar in nature to the way that radiologists analyze a CT scan. When radiologists review a CT scan containing numerous consecutive cross-sectional slices of the body, they identify the slices with an abnormality in the first step, and analyze the abnormal ones to diagnose the disease in the next step. Existing DL-based CT scan processing methods either use all slices as a 3D input to a classifier or classify individual slices and transform slice-level predictions to patient-level ones using a threshold on the entire slices [85]. Determining a threshold on the number of slices or on the ratio of slices demonstrating infection over the entire slices is not precise, as most pulmonary infections have different stages which may reveal different involvement levels in various lung regions [86]. Furthermore, a CT scan may contain

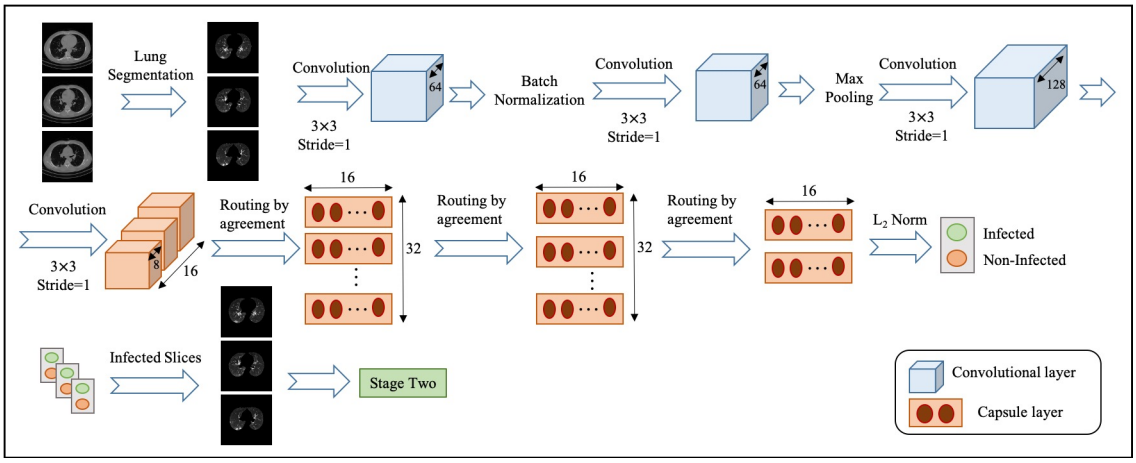


Figure 3.6: Architecture of the COVID-FACT at stage one.

a different number of slices depending on the acquisition settings, which makes it challenging to find such a threshold. In most methods which pass all slices as a 3D input to the model, the input size is fixed and the model is trained to assign higher scores to slices demonstrating infection. However, the performance of such models will be reduced when being tested on a dataset other than the dataset on which they are originally trained [22]. In the COVID-FACT framework, the output of stage one may vary in size for each patient due to different areas of lung involvement and phase of infection, unlike the CT-CAPS which generates a single $(32, 16)$ matrix in the first stage. It is worth mentioning that the modified loss function introduced in Eq. (3.1) is also used in the training phase of the COVID-FACT, so that a higher penalty rate is given to the false negatives (i.e., mis-classified infectious slices).

It is also worth noting that the lung region segmentation, described in Sub-section 3.1.1, is performed in one of the variants of the COVID-FACT as a preprocessing step. The first stage of the COVID-FACT, on the other hand, is tasked with this specific issue of extracting slices demonstrating infections.

3.3.2 COVID-FACT's Stage Two:

As mentioned earlier, COVID-FACT intends to apply classification methods on a subset of slices demonstrating infection rather than the entire slices in a CT scan. As such, the second stage of the COVID-FACT takes candidate slices of a patient detected in stage one as the input, and classifies

them into one of the COVID-19 or non-COVID (including normal and CAP) classes, i.e., a binary classification problem is considered. Stage two is a stack of four convolutional layers and two capsule layers, as shown in Fig. 3.3.2. The length of the last two capsules indicates classification probabilities. An average voting function is then applied to the classification probabilities, in order to aggregate slice-level values and find the patient-level predictions as follows:

$$P(p_k \in c) = \frac{1}{L_k} \sum_{i=1}^{L_k} P(s_i^k \in c), \tag{3.2}$$

where $P(p_k \in c)$ refers to the probability that patient k belongs to the target class c (e.g., COVID-19), L_k is the total number of slices detected in stage one for patient k, and $P(s_i^k \in c)$ refers to the probability that the i^{th} slice detected for patient k belongs to the target class c. It is worth noting that while, initially, the COVID-FACT performs slice-level classification in its second stage, the output is patient-level classification (through its voting mechanism). In addition, similar to the stage one, in the training phase of stage two, the weighted loss function proposed in Eq. (3.1) is used, and the default cut-off probability of 0.5 is chosen to distinguish COVID-19 and non-COVID cases. As another note to this discussion, I would like to add that the coronavirus infection is, typically, distributed across the lung volume and, as such, manifests itself in several CT slices. Therefore, having a single slice identified as a COVID-19 infection can not necessarily lead to a positive COVID-19 detection.

To further improve the ability of the proposed COVID-FACT to distinguish COVID-19 and non-COVID cases and attenuate the effects of errors in the first stage, all patients with less than 3% of slices demonstrating infection in their entire volume are classified as non-COVID cases. These cases are more likely normal cases without any slices with evidence of infection. The few slices with infection identified for these cases might be due to the model error in the first stage, non-infectious abnormalities such as pulmonary fibrosis, or motion artifacts in the original images, which will be covered by this threshold. Based on the study in Reference [86], it can be interpreted that 4% lung involvement is the minimum percentage for COVID-19 positive cases. In addition, the minimum percentage of slices demonstrating infection detected by the radiologist in the COVID-CT-MD dataset is 7%, and therefore 3% would be a safe threshold to prevent mis-classifying infected cases as normal.

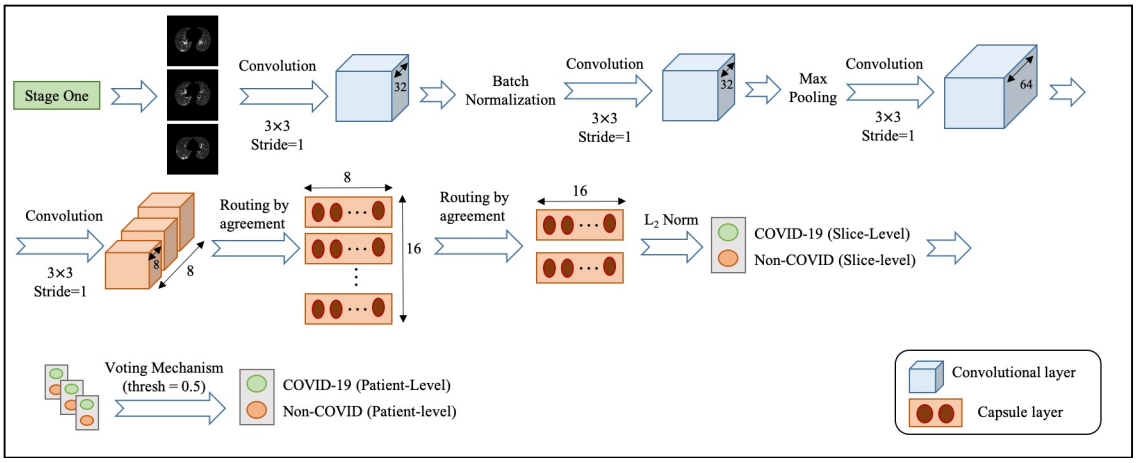


Figure 3.7: Architecture of the COVID-FACT at stage two.

As a final note, it is worth mentioning that, for the proposed COVID-FACT framework, the role of stage one is critical to achieving a fully-automated framework, which does not require any input from the radiologists, especially when an early and fast diagnosis is desired. However, the COVID-FACT framework is completely flexible and stage one can be skipped if the slices demonstrating infections have already been identified by the radiologists, meaning that the normal cases are already identified in this case and stage two merely separates COVID-19 and CAP cases.

3.4 COVID-FACT's Experimental Results

Similar to the CT-CAPS framework, the proposed COVID-FACT is trained and evaluated based on the COVID-CT-MD dataset. The testing set is also the same as that of the CT-CAPS. To train the model, the Adam optimizer with an initial learning rate of $1e-4$, batch size of 16, and 100 epochs are used. The model with the minimum loss value on the validation set was selected to evaluate the performance of the model on the test set. The COVID-FACT framework achieved an accuracy of 90.82%, sensitivity of 94.55%, specificity of 86.04%, and AUC of 0.98. The associated ROC curve is shown in Fig. 3.8. The training and validation loss curves are also illustrated in Fig. 3.9.

In a second experiment, the proposed model is trained using the complete CT images without segmenting the lung regions. The obtained model reached an accuracy of 90.82%, sensitivity of 92.72%, specificity of 88.37%, and AUC of 0.95. The corresponding ROC curve is shown in Fig. 3.8. This experiment shows that segmenting lung regions in the first step will increase the sensitivity

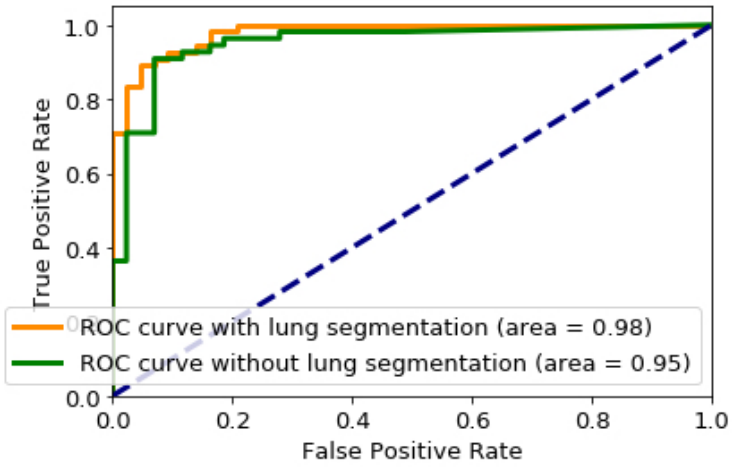


Figure 3.8: ROC curve of the proposed COVID-FACT.

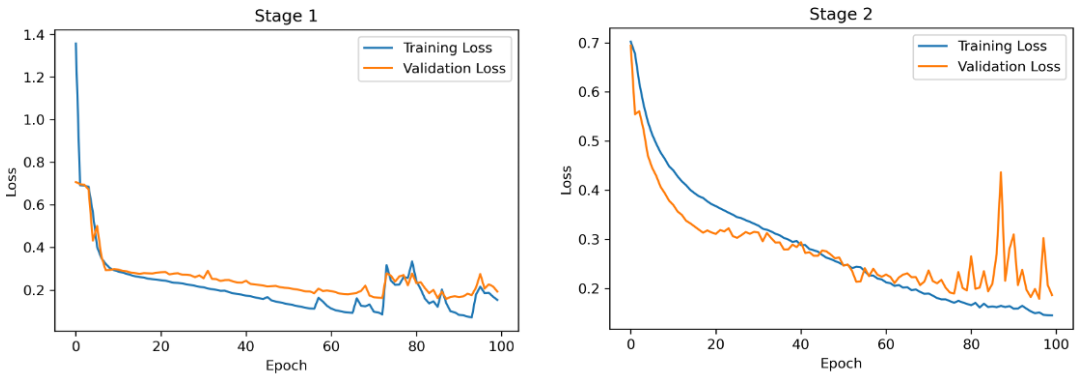


Figure 3.9: Training and Validation loss curves obtained for the COVID-FACT stage one and stage two.

from 92.72% to 94.55% and the AUC from 0.95 to 0.98, while slightly decreases the specificity from 88.37% to 86.04%. Although the numerical results show a slight improvement achieved by segmenting the lung regions, further investigating the sources of errors demonstrates the superiority of using segmented lung regions over the original CT images. In the COVID-FACT framework using lung region segmentation, none of COVID-19 and CAP cases have been mis-classified as a normal case by the 3% thresholding after the first stage, and 95.84% (23/24) of normal cases have been identified correctly using this threshold, while for the model without the lung segmentation, there is one mis-classification of a COVID-19 case by the 3% thresholding, and 91.66% (22/24) of normal cases were identified correctly by this threshold.

The performance of the COVID-FACT is further compared with a CNN-based alternative to

demonstrate the effectiveness of the Capsule Networks and their superiority over CNN in terms of number of trainable parameters and accuracy. In other words, the CNN-based alternative model has the same front-end (convolutional layers) as that of COVID-FACT in both stages. However, the capsule layers are replaced by fully connected layers including 128 neurons for intermediate layers and 2 neurons for the last layer at each stage. The last fully connected layer in each stage is followed by a softmax activation function and the remaining modifications and hyper-parameters are kept the same as used in COVID-FACT. The CNN-based COVID-FACT achieved an accuracy of 71.43%, sensitivity of 81.82%, and specificity of 58.14%. The COVID-FACT's performance and number of trainable parameters for examined models are presented in Table 3.3. It is worth noting that in designing the CNN-based COVID-FACT described above, the complexity and structure have been kept similar to its Capsule Network-based version. The goal is to illustrate potential advantages of Capsule Network-based design over its CNN-based counterpart. Alternative models using CNN architecture and fully connected layers such as the DenseNet model [52], however, consist of several convolutional layers and a high degree of complexity. As such, such complex models are expected to outperform the CNN-based COVID-FACT.

As mentioned earlier, the ROC curve provides physicians with a precious tool to modify the sensitivity/specificity balance based on their preference by changing the classification cut-off probability. To elaborate on this point, the default cut-off probability is changed from 0.5 to 0.75 and reached an accuracy of 91.83%, a sensitivity of 90.91%, and a specificity of 93.02%. Further increasing the cut-off probability to 0.8 results in the same accuracy of 91.83%, a lower sensitivity of 89.01%, and a higher specificity of 95.34%. On the other hand, decreasing the cut-off probability from 0.5 to 0.35 will increase the accuracy and the sensitivity to 91.83% and 98.18% respectively, while slightly decreases the specificity to 83.72%. The performance of the COVID-FACT for different values of cut-off probability is presented in Table 3.4.

While the performance of COVID-FACT is evaluated by its final decision made in the second stage, the first stage plays a crucial role in the overall accuracy of the model. As such, the performance of the COVID-FACT in the first stage is also reported in Table 3.5. As shown in Table 3.5, $\sim 91\%$ of the slices demonstrating infection are identified correctly by the COVID-FACT at the first stage, while there are some mis-classified slices that will be passed to the next stage as the infectious slices.

Table 3.3: Results obtained by the COVID-FACT framework and its alternative CNN-based counterpart.

Method	Accuracy	Sensitivity	Specificity	AUC	Trainable Parameters
COVID-FACT with Lung Segmentation	**90.82%**	**94.55%**	86.04%	**0.98**	**406,880**
COVID-FACT without Lung Segmentation	**90.82%**	92.72%	**88.37%**	0.95	**406,880**
CNN-based COVID-FACT	71.43%	81.82%	58.14%	0.67	365,806,660

It is also evident that the CNN-based model cannot properly identify infectious slices, which in turn led to the low performance of the second stage. It is worth mentioning that stage one is only responsible for detecting candidate slices, while stage two classifies the slices into COVID-19 and non-COVID categories. The second stage is followed by an aggregation mechanism, which takes all the slices of a patient into account and consequently decreases the impact of mis-classified slices at the first stage.

In another experiment, the performance of the model when the commonly used focal loss function [87] is utilized to train the model is investigated. The COVID-FACT framework trained by the focal loss function ($\gamma = 2$, $\alpha = 0.25$) achieved the same patient-level performance compared to the proposed model while the performance of the first stage is lower by achieving the accuracy of 92.79%, sensitivity of 87.69%, and the specificity of 97.03%. The lower sensitivity in the first stage shows benefits of using the modified loss function, described in Eq. (3.1), as the role of the first stage in the pipeline is to detect slices with the evidence of infection to be analyzed in the second stage. Therefore, the model that was trained using the modified loss function has been selected as the final model due to its higher accuracy and sensitivity in detecting slices demonstrating infection.

As another experiment, performance of stage two is evaluated without applying the first stage to provide a better comparison of the models used in the second stage. More specifically, the stage two model is trained based on the infectious slices identified by the radiologist and evaluated on the labeled test set including 17 COVID-19 and 8 CAP cases. The numbers of correctly predicted cases in this experiment are presented in Table 3.6. The experimental results obtained by the COVID-FACT framework using the lung segmentation achieved quite a similar performance compared to the case in which the model was trained based on the outputs of stage one. This result further demonstrates

Table 3.4: Performance of the COVID-FACT for different values of cut-off probability.

Cut-off Probability	**0.35**	**0.5**	**0.6**	**0.7**	**0.75**	**0.8**
Accuracy (%)	**91.83**	90.82	**91.83**	90.82	**91.83**	**91.83**
Sensitivity (%)	**98.18**	94.55	92.73	90.91	90.91	89.01
Specificity (%)	83.72	86.04	90.70	90.70	93.02	**95.34**

Table 3.5: The performance of the COVID-FACT's stage one in diagnosis of slices demonstrating infection.

Method (Stage 1)	**Accuracy**	**Sensitivity**	**Specificity**	**AUC**
COVID-FACT with Lung Segmentation	**93.14%**	**90.75%**	94.01%	**0.96**
COVID-FACT without Lung Segmentation	92.78%	87.59%	**94.36%**	**0.96**
CNN-based COVID-FACT	79.74%	33.00%	91.28%	0.64

that the Capsule Network and the aggregation mechanism used in stage two can cope with errors in the previous stage and achieve desirable performance. It is worth mentioning that this experiment was performed using only the labeled dataset, which consequently provided a smaller dataset to train the model.

The localization maps generated by the Grad-CAM method are also illustrated in Fig. 3.10 for the second and fourth convolutional layers in the first stage of the COVID-FACT. It is evident in

Table 3.6: Correctly predicted cases using only the COVID-FACT's stage two without applying the first stage

Model	**COVID-19**	**CAP**
Stage 2 (with Lung Segmentation)	**94.1% (16/17)**	**87.5% (7/8)**
Stage 2 (without Lung Segmentation)	88.2%(15/17)	62.5%(5/8)
Stage 2 (CNN-based)	82.4%(14/17)	25%(2/8)

Fig. 3.10 that the COVID-FACT model is looking at the right infectious areas of the lung to make the final decision. Due to the inherent structure of the capsule layers, which represent image instances separately, their outputs cannot be superimposed over the input image. Consequently, in this study, the Grad-CAM localization maps are only presented for convolutional layers.

3.4.1 K-Fold Cross-Validation

The performance of the COVID-FACT and its variants have also been evaluated based on the 5-fold cross-validation [88] to provide more objective assessments. In this experiment, the COVID-FACT achieves the accuracy of $87.61 \pm 2.00\%$, the sensitivity of $88.30 \pm 3.22\%$, and specificity of $86.75 \pm 1.91\%$. Using the same 5-fold cross-validation technique, the COVID-FACT without using the segmented lung regions achieves the accuracy of $87.31 \pm 3.37\%$, sensitivity of $88.32 \pm 5.00\%$, and specificity of $86.03 \pm 3.18\%$. Finally, the CNN-based COVID-FACT achieves the accuracy of $64.49 \pm 1.61\%$, sensitivity of $79.58 \pm 6.61\%$, and specificity of $46.67 \pm 8.48\%$. The results confirm the superiority of the COVID-FACT using the segmented lung areas over its variants which is in line with the previous experiments based on the randomly selected test dataset. Moreover, similar to the previous experiments, modifying the cut-off probability is beneficial in the cross-validation case to adjust the capability of the model to focus on COVID-19 or non-COVID cases depending on radiologists' priorities. More specifically, in the aforementioned 5-fold cross-validation, decreasing the cut-off probability to 0.35 increases the sensitivity to $92.97 \pm 2.96\%$ while the overall accuracy remains the same. Increasing the cut-off probability to 0.6, on the other hand, increases the specificity to $91.16 \pm 3.73\%$ and provides the same accuracy similar to the previous case.

3.5 Discussion

In order to further determine the limitations and possible improvements, mis-classified cases are investigated. Table 3.7 shows the number of mis-classified cases for each type of input disease (COVID-19, CAP, normal) obtained at stage two, as well as the number of normal cases that were not identified correctly by the 3% threshold after the first stage. The low rate of errors obtained by the 3% threshold in the first stage demonstrates the capability of COVID-FACT to identify normal

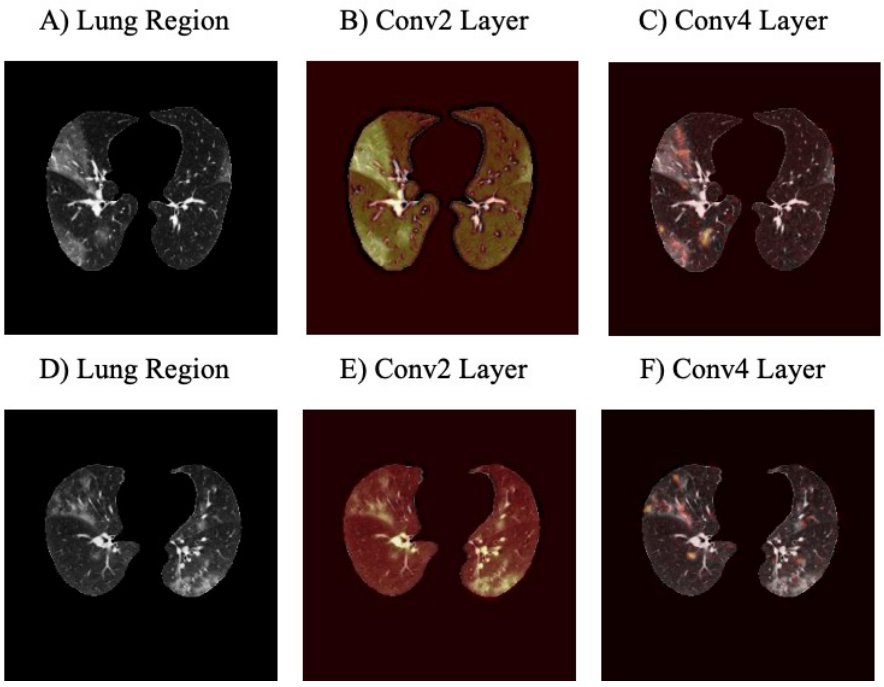


Figure 3.10: Localization heat maps generated by the Grad-CAM approach for two sample CT slices, based on the second and forth convolutional layers in the COVID-FACT's first stage.

cases in the first stage, which is very helpful for physicians and radiologists to exclude normal cases at the very beginning of their study.

As in the case of highly contagious diseases such as COVID-19, the False Negative Rate (FNR) is of utmost importance, such errors are further analyzed to explore the possible sources of the mis-classification. As shown in Table 3.7 there are $3/55$ COVID-19 cases that are mis-classified by the COVID-FACT framework. Further reviews revealed that one mis-classified COVID-19 case contains unifocal infection manifestation with consolidation predominance rather than GGO, which are more common in CAP cases rather than COVID-19 ones. One other case of error was identified as an incomplete CT scan with missing slices, which has consequently made the correct identification difficult for the framework. In addition, the aforementioned errors are reviewed in the case of image quality and lung segmentation as other potential causes of the errors. The assessment results showed that the image qualities are adequate and the segmentation model performed well without removing or cropping the infection manifestations. Therefore, some errors are likely to be caused by the similarities between the infection patterns in CAP and COVID-19 cases. It is worth noting that decreasing the cut-off probability from 0.5 to 0.35, as shown in Table 3.4, will result in the correct

Table 3.7: The number of the mis-classified cases for each type of the input disease and the number of cases that were not identified correctly by the 3% threshold.

Input	**Errors (Thresholding)**	**Errors (Stage 2)**
COVID-19	$0/55$	$3/55$
CAP	$0/19$	$5/19$
normal	$1/24$	$1/24$

classification of the two false-negative cases, which contain similar characteristics to other infections. This can be considered as a remedy, when FNR is of the main concern.

It is also identified that errors in stage one are mainly caused by non-infectious abnormalities such as pulmonary fibrosis and artifacts. In this regard, slices with the evidence of artifact with no signs of infection manifestation have been explored. In some cases, the motion artifact or the artifacts caused by the presence of metallic objects inside the body have generated some components in the image that were mis-classified as infectious slices. Fig. 3.11 illustrates 4 samples of such slices in which images (A) and (B) belong to a mis-classified normal case while images (C) and (D) are related to two CAP cases, which are classified correctly in the second stage. It is worth mentioning that, the number of such slices is negligible, especially when they appear in cases that have multiple infectious slices (caused by CAP or COVID-19). In those cases, the influence of such slices with the evidence of artifact will be diminished by the second stage and the following aggregation mechanism. Motion artifact reduction algorithms can be investigated as a future work to cope with undesired impacts of the artifacts on the final result.

It is worth mentioning that during the labeling process accomplished by the radiologist to detect slices demonstrating infection, it was noticed that in some cases the abnormalities are barely visible with the standard visualization setting (window level and window width). Those abnormalities have been detected by changing the image contrast (by adjusting the window level and window width) manually by the radiologist. This limitation demonstrated the need to research on finding the optimal contrast and window setting parameters, which is addressed in Chapter 4. Another limitation can be considered as the retrospective study used in the data collection part of this research. Although

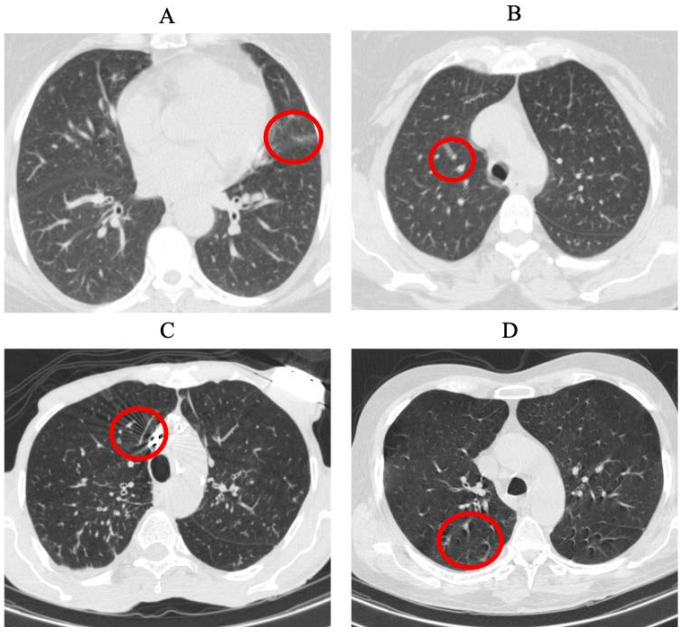


Figure 3.11: Examples of chest CT slices with the evidence of artifact where no infection manifestation is observed.

the COVID-CT-MD dataset is acquired with the utmost caution and inspection, a retrospective data collection might add inappropriate cases to the study at hand. The potential improvement to address this limitation could be the collaboration of more radiologists in analyzing and labeling the data to assess if the interobserver agreement is satisfactory or not.

As a side note to this discussion, I would like to mention that while both CT and CR can decrease the false negative rate at the admission and discharge times, the CR is less sensitive, and less specific compared to CT. Some studies, such as Reference [89], report that CR often shows no lung infection in COVID-19 patients at early stages resulting in a low sensitivity of 69% for the diagnosis of COVID-19. Therefore, chest CT has a key role in the diagnosis of COVID-19 in the early stages of the infection and also in setting up a prognosis. Consequently, CT is considered as the preferred modality for grading and evaluation of imaging manifestations for COVID-19 diagnosis. It is worth adding that as CT scans are 3D images, as opposed to 2D chest radiographs, and more difficult to be processed using ML and DL techniques, as the currently available resources cannot efficiently process the whole volume at once. As such, slice-level and thresholding techniques are utilized to cope with such limitations, leading to a reduced performance compared to the models working with CR (e.g., the COVID-CAPS [20]). The focus of this research is to further enhance the performance

of CT-based COVID-19 diagnosis models to fill the gap between the radiologists' performance and that of volumetric-based DL techniques.

3.6 Summary

In this chapter, two fully automated Capsule Network-based frameworks, referred to as the "CT-CAPS" and "COVID-FACT" are proposed, which utilize the advantages of Capsule Networks to identify the COVID-19 disease in a coarsely-labeled dataset of COVID-19, CAP, and normal cases. The experimental results indicate the capability of both frameworks to automatically analyze volumetric chest CT scans and distinguish different cases while far fewer parameters and a less sophisticated labeling process compared to their existing counterparts are utilized. While both frameworks achieve acceptable and desired results, experiments demonstrate the superiority of the COVID-FACT over CT-CAPS by achieving an accuracy of 90.82%, sensitivity of 94.55%, specificity of 86.04%, and AUC of 0.97. In summary, CT-CAPS demonstrates that the penultimate capsule layer can be a proper compact feature representative of CT scans to be used for classification tasks. COVID-FACT, which is the extension of the proposed CT-CAPS, extracts candidate slices with the evidence of infection and passes them to a Capsule Network-based classifier followed by an averaging mechanism to provide patient-level labels. Moreover, the benefits of extracting lung tissues in the proposed frameworks, and the flexibility of the models to be adjusted based on radiologists' preferences to achieve desired results have been demonstrated by the experimental results. As a final note, the multi-center SPGC-COVID dataset, introduced in Sub-section 2.4.2, is used to further enhance the performance of the COVID-FACT framework when being tested on a varied dataset of chest CT scans. The model development and evaluation details are provided in Chapter 4.

Chapter 4

Robust COVID-19 Identification from Multi-center and Heterogeneous Datasets

The main objective of this chapter is to enhance the performance and robustness of the diagnostic frameworks previously introduced in Chapter 3. In addition, this chapter proposes a modified slice-level infection diagnosis model to further enhance the performance of the DL-based models working with CT scans.

With regard to the second objective, an extension of the proposed automated COVID-FACT framework is developed, which is tailored to robustly classify volumetric chest CT scans into one of the three target classes (COVID-19, CAP, or normal). The proposed framework integrates a scalable enhancement approach to boost its performance and robustness in the presence of gaps between the train and test sets regarding types of scanners, imaging protocols, and technical acquisition parameters. The proposed framework can also be generalized on varied external datasets with high flexibility to update itself upon receiving new external datasets in an unsupervised fashion. More specifically, the subset of the unlabeled test images for which the model generated a confident prediction is extracted and used along with the training set to re-train and update the benchmark model (the model trained on the initial train set). Finally, an ensemble architecture is adopted to aggregate the predictions from multiple versions of the model. For initial training and development purposes, the COVID-CT-MD dataset is used, which is described in Sub-section 2.4.1 and contains

volumetric CT scans acquired from one imaging center using a constant standard radiation dose scanning protocol. To evaluate the model, the SPGC-COVID dataset, introduced in Sub-section 2.4.2, was utilized. Among the test cases, there are CT scans with similar characteristics to the train set, as well as noisy low-dose and ultra-low-dose CT scans. In addition, some test CT scans were obtained from patients with a history of cardiovascular diseases or surgeries. The obtained results show that while the proposed model is trained on a relatively small dataset acquired from only one imaging center using a specific scanning protocol, it performs well on heterogeneous test sets obtained by multiple scanners using different technical parameters. The experimental results also demonstrate the capability of the proposed unsupervised enhancement approach to improve the performance and robustness of the model when being evaluated on varied external test sets. The performance of the proposed framework is compared with state-of-the-art approaches proposed at the COVID-19 grand challenge mentioned in refSREP-sec:spgc-covid. The results demonstrate that the proposed framework performs well on all test sets and outperforms all the submitted models by achieving the overall accuracy of 96.15% ($95\%CI : [91.25 - 98.74]$), COVID-19 sensitivity of 96.08% ($95\%CI : [86.54 - 99.5]$), CAP sensitivity of 92.86% ($95\%CI : [76.50 - 99.19]$), normal sensitivity of 98.04% ($95\%CI : [89.55 - 99.95]$), and the AUC of 0.992.

With regard to the second objective, an automated framework based on the Capsule Networks, referred to as the "WSO-CAPS", is proposed to efficiently detect slices demonstrating infection using LDCT and ULDCT. The WSO-CAPS framework is essentially an extension of the slice-level classifiers proposed in the previously described frameworks and is equipped with a Window Setting Optimization (WSO) mechanism to jointly identify slices with the evidence of infection and find the best window setting parameters to resemble the radiologists' efforts in reviewing LDCT and ULDCT. The experimental results on the WSO dataset, described in Sub-section 2.4.3 show that the WSO-CAPS improves the capability of the Capsule Network and its counterparts to identify slices demonstrating infection.

The remainder of this chapter is organized as follows: First, the components of the proposed classification framework are briefly described, followed by a detailed description of the proposed unsupervised enhancement approach. Then, the experimental results and a brief discussion are presented. Next, the details of the proposed WSO-CAPS are described, and the associated results are

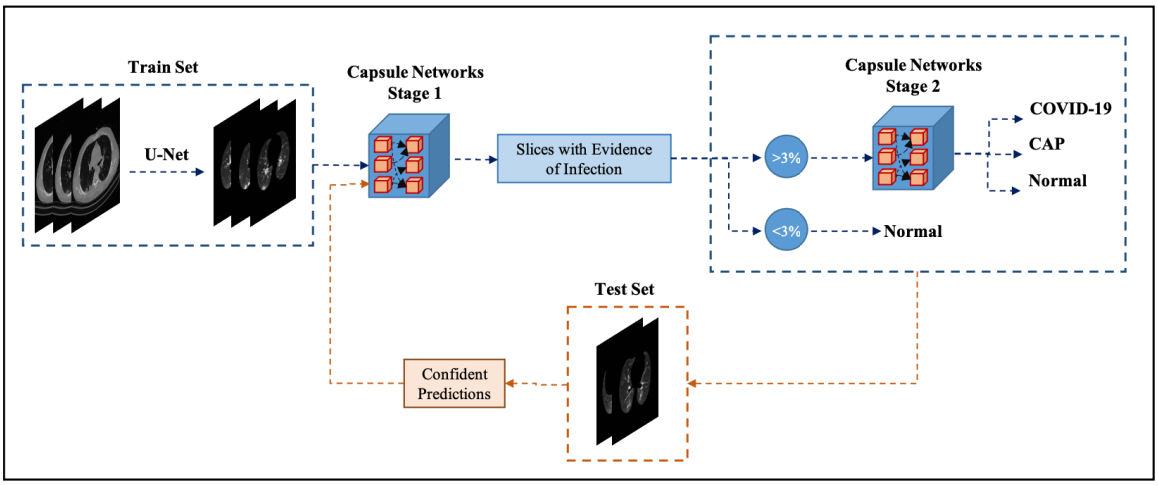


Figure 4.1: The pipeline of the proposed robust three-way classification framework and the associated enhancement approach.

presented. Finally, a brief summary of the chapter is provided.

4.1 Boosted and Robust COVID-FACT Framework

In this section, we develop a two-stage framework similar to the “COVID-FACT" to classify volumetric CT scans into three target classes of COVID-19, CAP, and normal. We then use the unlabeled data from the test sets to boost the performance and robustness of the framework on the unseen cases. The pipeline of the proposed boosted and robust COVID-FACT framework is shown in Fig. 4.1. Different components of the proposed framework are described below:

4.1.1 Preprocessing

Following the preprocessing step of the CT-CAPS and COVID-FACT frameworks, the lung areas are first extracted from the CT images by the well-trained R231CovidWeb segmentation model to remove the insignificant and distracting components. In addition, similar to the proposed frameworks, all images are down-sampled into the (256×256) size and normalized into the $[0, 1]$ interval. This step is crucial as image sizes may vary and pixel intensities may be in different ranges when the images are acquired by different scanners.

4.1.2 Stage One

The first stage performs the infection identification task, which aims to find slices with the evidence of infection (caused by CAP or COVID-19) for each patient. The identified slices will then be classified into one of the three target classes in the second stage. In this regard, the same architecture as the first stage of the COVID-FACT framework is adopted for this task. In addition to the original architecture (i.e. COVID-FACT's stage 1), residual connections are added between the convolution layers in this modified version to transfer low-level features to the deeper layers. This modification further assists the model in identifying informative features. Additionally, a dropout layer is added before the capsule layers to overcome the overfitting problems during the training phase. The detailed structure of the classification model used in the first stage is shown in Fig. 4.2(a).

The labeled subset of the training dataset (i.e. COVID-CT-MD) has been used to train this stage over 100 epochs using the Adam optimizer with a learning rate of $1e-4$. To account for the imbalanced number of slices in each class, the same weighted loss function as that of the COVID-FACT is used. The original test set in the COVID-CT-MD dataset has been used as the validation set to select the final model.

4.1.3 Stage Two

The second stage takes the candidate slices from the previous stage and classifies them into one of the COVID-19, CAP, or normal cases. More specifically, the slices demonstrating infection recognized by the first stage are used to train a Capsule Network-based three-way classification model. The architecture of stage two is shown in Fig. 4.2(b). Similar to the first stage, a weighted loss function is used to cope with the imbalanced number of samples in some categories. At this stage, however, the loss weights associated with normal and CAP classes are set to 5 and the weight for the COVID-19 class is set to 1. Note that as the normal cases are extremely rare at this stage, the weights are set differently compared to those calculated by Eq. (3.1), to maintain the stability of the training process, while forcing the model to pay more attention to the minority classes. In addition, the binary cross-entropy loss function is used which translates the three-way classification problem at hand into three binary classification tasks. In fact, the loss value is calculated separately for each

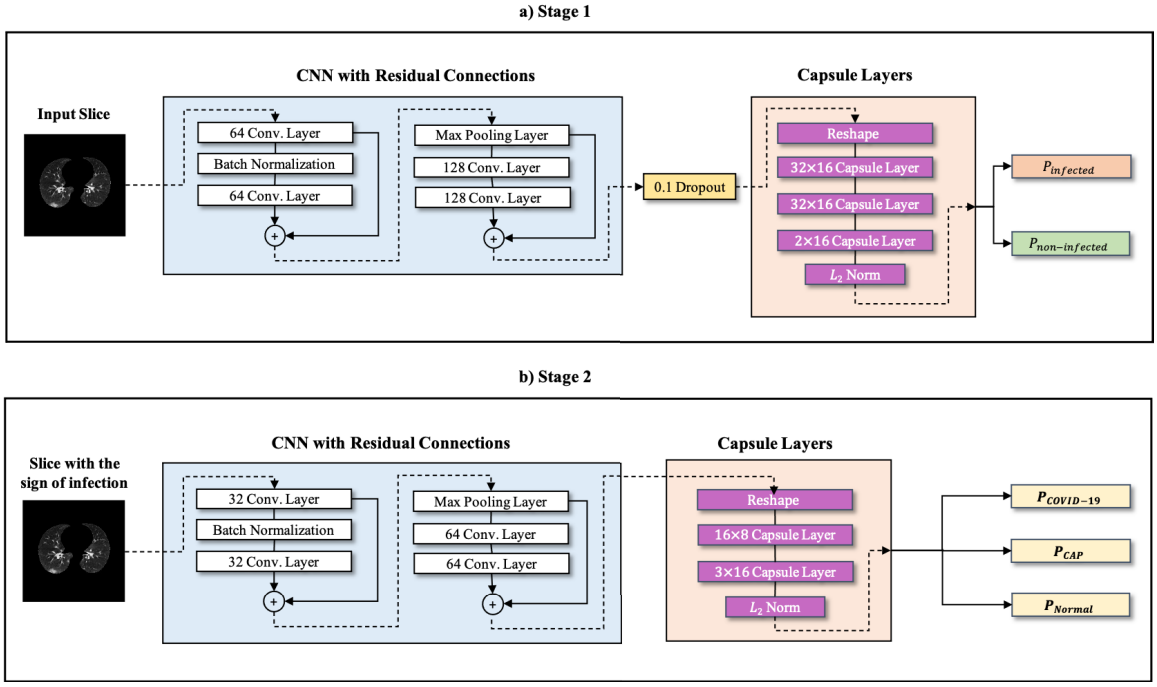


Figure 4.2: a) The structure of the CapsNet binary classifier in stage 1. b) The structure of the three-way classifier in stage 2. + sign denotes the residual addition.

binary label associated with a target class (i.e., COVID-19, CAP, normal). Finally, a majority voting mechanism is adopted to transfer slice-level predictions into patient-level ones and determine the final label. It is worth noting that an accurate model in the first stage detects only a few candidate slices from normal cases. The 3% thresholding mechanism similar to the one incorporated into the COVID-FACT can then be applied to the output of the first stage to identify those cases with only a few identified infectious slices in the first stage and label them as normal. In other words, if less than 3% of the slices in a volumetric CT scan are classified as infectious, the corresponding CT scan is classified as a normal case.

4.1.4 Unsupervised Enhancement

Unseen CT scans acquired by different scanners and scanning protocols contain heterogeneous characteristics, leading to lower performance of a pre-trained model. To increase the robustness, it is possible to take advantage of the extra unlabeled samples that are available via the various test cases, and utilize this extra set of CT scans in an unsupervised fashion. In other words, inspired by the ideas of "Active Learning [90–92]", where different data samples are extracted to train the

model in different stages, and "Semi-Supervised Learning [93, 94]", where a label is assigned to unlabeled cases based on a pre-defined metric, an autonomous mechanism is developed to extract and label a part of test dataset using a probabilistic selection criteria with reduced complexity. The selected sample and the assigned labels are then used to re-train and boost the initially trained model. More specifically, those test cases for which the model generated the most confident results (i.e., high probability) are selected. Similarly, among the selected cases, those with a high confidence in slice-level predictions are picked. To specify the confidence level for the obtained results, the probability that a volumetric CT scan belongs to a specific category is defined as the ratio of the slices predicted as the target class over the total number of slices (all slices containing the lung lesion), which can be written as follows:

$$P(\mathbb{X} \in \mathcal{C}_i) = \frac{n_{\mathcal{C}_i}}{\sum_{i=1}^{C} n_{\mathcal{C}_i}}, \tag{4.1}$$

where $\mathbb{X}$ represents the input volumetric CT scan, C represents the number of target classes, and $n_{\mathcal{C}_i}$ denotes the number of slices belonging to the target class $\mathcal{C}_i$. Then, a confidence threshold is specified and a prediction is considered confident if the probability of the input CT scan belonging to any of the target classes is greater than the pre-set threshold. In this study, 80% is used as the confidence threshold. A similar approach is used to extract confident slices and their corresponding labels. In this case, the probability of a slice belonging to a target class is determined by the output of the CapsNet classifier in stage 2, which is the length (L_2Norm) of capsules in the last layer. It is worth mentioning that for those normal cases, which are identified in the first stage using the described thresholding mechanism, only the slices which are misclassified as infectious with a high probability (e.g., more than the confidence threshold) are selected. Such slices will be labeled as normal in the enhancement phase. Following the aforementioned steps, a set of slices and their corresponding labels will be obtained to augment the training dataset aiming to make the model more aware of the new features available in the unseen datasets and achieve more robust feature maps. Therefore, for each test set, a set of confident slices and their associated labels are obtained, which have been added to the train set to re-train the model of the second stage. It is worth noting that the first stage has been kept unchanged in this approach. Finally, after re-training the benchmark model

based on the confident slices acquired from each test set, several enhanced models (each related to one test set) are achieved and the associated patient-level probability scores are averaged to provide the final prediction. This aggregation mechanism depends on the target test set. More specifically, to apply the model on each test set, the predictions obtained by the models enhanced over the other test sets are averaged. For instance, the model developed for the diagnosis of cases in test set 1 takes the average of probability scores provided by the models enhanced on test set 2 and 3. The main reason for using such an aggregation mechanism is that the enhancement based on a specific test set will further boost the probability scores of confidently predicted slices while having limited influence on other cases in the same set. As such, incorporating the model enhanced on a test set will not bring in any further improvement to the evaluation process of the same set. The results presented in Table 4.2 further support this discussion. It is worth noting that the first three test sets are used to enhance the benchmark model and the fourth test set is kept aside for only evaluation purposes. As such, upon receiving new test datasets, the results of the enhanced models on the individual test sets (each representing a specific center or scanning protocol) can be aggregated to provide the classification results for the new cases. The unsupervised model enhancement described above along with the subsequent ensemble averaging mechanism make the entire pipeline a robust automated framework that can be easily improved and updated upon receiving new datasets from different imaging centers.

4.1.5 Experimental Results

As previously stated, to evaluate the performance of the proposed framework and the effectiveness of its unsupervised enhancement approach, the first three test sets are used to enhance the benchmark model, and the fourth test set is kept aside only for evaluation purposes. The results obtained by applying the enhanced ensemble model on all of the test sets are shown in Table 4.1. In this table, the AUC value is calculated based on the micro average of the values obtained for each class. In addition, to further validate the obtained results, confidence intervals for the total accuracy and sensitivity are provided using the method introduced in Reference [95].

To elaborate on the effect of the proposed unsupervised enhancement approach, the performance of the benchmark model (i.e., before enhancement) and the models enhanced by individual test sets (i.e., before averaging the outputs) is presented in Table 4.2. Results shown in Table 4.2 imply that

Table 4.1: Results obtained by the proposed robust framework for different test sets of the SPGC-COVID dataset. 95% Confidence Intervals obtained for the total performance using the significance level of 0.05 are presented in parentheses.

Test Set	**Accuracy(%)**	**COVID-19 Sensitivity(%)**	**CAP Sensitivity(%)**	**Normal Sensitivity(%)**	**AUC (micro)**
Test 1	100	100	NA	100	1.000
Test 2	86.67	80	90	90	0.952
Test 3	100	100	100	100	1.000
Test 4	97.50	100	87.50	100	0.999
Total	96.15 (CI: [91.25-98.74])	96.08 (CI: [86.54-99.5])	92.86 (CI: [76.50-99.19])	98.04 (CI: [89.55-99.95])	0.992

Table 4.2: The ratio of correctly classified cases over total cases in the each class obtained by the proposed robust model, the benchmark model, and three partially enhanced models.

Test Set	Sensitivity	Proposed	Enhanced #1	Enhanced #2	Enhanced #3	Benchmark
Test1	COVID-19	**15/15**	15/15	15/15	15/15	15/15
	Normal	**15/15**	15/15	15/15	15/15	15/15
Test2	COVID-19	**8/10**	8/10	8/10	8/10	8/10
	CAP	**9/10**	9/10	8/10	9/10	8/10
	Normal	**9/10**	9/10	9/10	9/10	9/10
Test3	COVID-19	**10/10**	10/10	9/10	9/10	9/10
	CAP	**10/10**	10/10	10/10	10/10	10/10
	Normal	**10/10**	10/10	10/10	10/10	10/10
Test4	COVID-19	**16/16**	16/16	16/16	15/16	15/16
	CAP	7/8	6/8	7/8	**8/8**	7/8
	Normal	**16/16**	16/16	16/16	16/16	16/16

the probability of the input CT scan belonging to the target class in some misclassified cases has been on the thresholding edge (close to 0.5) and could be corrected after incorporating the models enhanced over other test sets.

In addition to the final patient-level predictions, the performance of the first stage on the validation set in detecting slices demonstrating infection is evaluated to provide a clearer insight into the internal components of the framework. The first stage achieved an accuracy of 93.41%, sensitivity of 91.04%, and specificity of 94.26% in the binary (infectious & non-infectious) classification task. As slice-level labels (i.e., binary labels indicating the existence of infection in a CT slice) are not available for test sets, the result on the validation set is only reported. Moreover, as mentioned earlier, the output of the first stage can be used to identify most normal cases before entering the next stage. The results

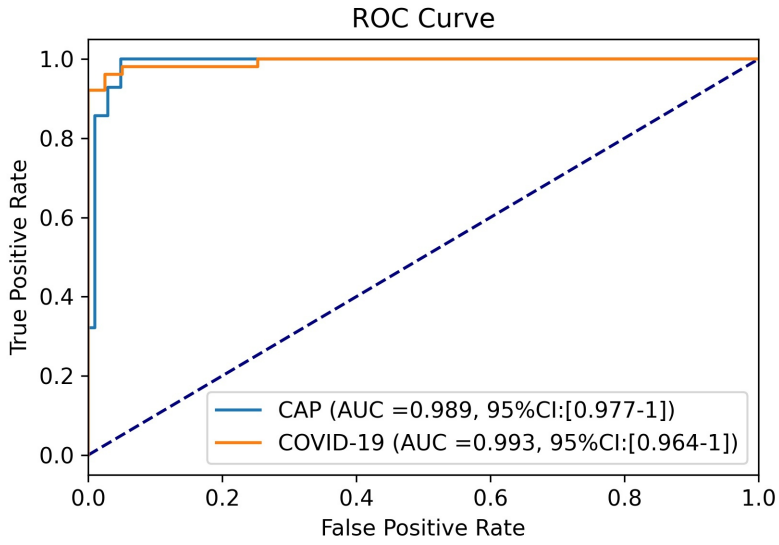


Figure 4.3: ROC curves for COVID-19 vs. others and CAP vs. others.

also show that nearly all of the normal cases in the four test sets ($45/46$ cases) have been identified correctly by the thresholding mechanism applied to the output of the first stage, while none of the COVID-19 and CAP cases have been misclassified as normal using this thresholding approach. In Fig. 4.3, the ROC curves for COVID-19 and CAP cases against other classes (e.g., COVID-19 vs. CAP and Normal) are plotted. The associated AUC values are also provided.

Comparison

The proposed framework is compared with top six models [96–101] developed following the Signal Processing Grand Challenge (SPGC) on COVID-19 diagnosis, which was organized as part of the 2021 IEEE International Conference on Acoustics, Speech, & Signal Processing (ICASSP). In the first phase of this SPGC, participants had access to the same train and validation sets as those used in this study to develop and evaluate their models. In the second phase, they were provided with the first three test sets and had two weeks to submit their final models. Finally, the best-performing models based on the first three test sets have been evaluated on the fourth test set to determine the overall performances. Experimental results demonstrate that the robust framework proposed in this chapter outperforms its counterparts proposed in the SPGC. Furthermore, it benefits from a scalable enhancement approach that can be integrated into most of the state-of-the-art models

to improve their performance when being tested on a heterogeneous dataset. In what follows, the six best-performing models from the SPGC COVID-19 are briefly described, followed by their corresponding performances on the entire test sets presented in Table 4.3.

- **Reference [96]:** In this model, slice-level predictions are acquired from an EfficientNet-based classifier [102] and a weighted majority voting is proposed to obtain the final patient-level labels. To train this classifier, the authors first trained two separate binary classifiers to detect slices demonstrating infection from COVID-19 and CAP cases. Then, they fed these models with unlabelled cases to provide the training set for the main classifier. Additionally, they only considered the middle slices (e.g., 80 middle slices) in a volumetric CT scan at the training phase.

- **Reference [97]:** This model aggregates the output of six classifiers developed based on the 3D ResNet101 model [103]. One model in this proposed framework is a three-way classifier trained over all of the cases, while the other five models are binary classifiers independently trained over COVID-19 and CAP cases using different combinations of train and validation sets.

- **Reference [98]:** This model presents a feature extraction-based approach in which a modified pre-trained ResNet50 model classifies each slice into the target classes and the penultimate fully connected layer is extracted as the feature map. Next, a max-pooling layer followed by two fully connected layers is used to generate patient-level prediction from slice-level feature maps. The output of this model is then aggregated with two BiLSTM patient-level classifiers, which are fed by the same slice-level feature maps to provide the final patient-level labels.

- **Reference [99]:** The pre-trained 3D Resnet50 [104] is the backbone of this model. The authors first doubled the number of slices for each case using a 3D cubic interpolation method. Then, they extracted the lung area using a pixel-based segmentation approach, followed by classical image processing techniques such as pixel filling and border cleaning. Finally, a subset of slices is selected from each volumetric CT scan based on their lung area and an experimentally-set threshold, which is then resized into a $(224, 224, 224)$ data, using a 3D cubic interpolation method, providing the patient-level input for training and evaluation purposes.

- **Reference [100]:** This model utilizes a two-stage framework in which the first stage is responsible for performing a multi-task classification to classify 2D slices into one of the target groups and identify the location of the slice in the sequence of CT images at the same time. The model at the first stage uses an ensemble of four popular CNN-based classifiers (i.e., ResneXt50 [105], DenseNet161 [106], Inception-V3 [107], and Wide-Resnet [108]), followed by an aggregation mechanism that divides the whole volumetric CT scan into 20 groups of slices and calculates the percentage of infected slices related to COVID-19 and CAP classes in each group. The values obtained for all groups are then concatenated and fed into an XG-boost classifier [109] in the second stage to generate patient-level predictions.

- **Reference [101]:** The model proposed in this work initiates with a slice-level EfficientNet-B1 classifier [102] aiming to classify slices and generate feature maps (intermediate layers) to be used in the subsequent sequence classifier. In the sequence classifier, several weak classifiers are trained and the outputs are aggregated using an adaptive weighting mechanism to obtain the final patient-level results. To further enhance the performance of the model and cope with the imbalanced training set, a combination of weak and strong data augmentations is applied to the training cases, forcing the model to produce similar labels for both types of augmented images. Furthermore, to improve the robustness of the model when being tested on varied datasets, the K-Means clustering method ($K = 3$) [110] is adopted to develop a single classifier for each cluster of the data and aggregate the results via a majority voting approach.

In addition to the aforementioned models, the proposed framework is further compared with another model which utilizes the same train and test sets (excluding the 4th test set) to target the same classification task [111]. A brief description of this model is as follows:

- **Reference [111]:** This model aims to introduce a robust training algorithm and classification framework, which is capable of being updated upon receiving new datasets to deal with the characteristic shifts in different test sets. First, it adopts a two-stage architecture similar to the COVID-FACT model proposed in Reference [2] and trains the benchmark model in a self-supervised fashion [112] and the majority voting is adopted to obtain patient-level labels. The backbone model used in this study is DenseNet169 [106] and strict slice preprocessing

Table 4.3: Performance of the DL-based counterparts of the proposed framework. P-values related to the McNemar's test with the significance level of 0.05 are presented in the last column, comparing the proportion of errors on the entire test set caused by the proposed framework and its counterparts.

Model	Accuracy(%)	COVID-19 Sensitivity(%)	CAP Sensitivity(%)	Normal Sensitivity(%)	McNemar's test p-value
Ref. [96]	90	86.27	89.28	94.11	0.07681
Ref. [97]	88.46	86.27	89.28	90.19	0.03088
Ref. [98]	87.69	88.23	78.57	92.15	0.00739
Ref. [99]	85.38	84.31	82.14	88.23	0.00052
Ref. [100]	84.61	90.19	60.71	92.15	0.00073
Ref. [101]	80.00	88.23	35.71	96.07	0.00005
Ref. [111]	72.22	65.71	85.00	71.43	0.00002
Proposed	**96.15**	**96.08**	**92.86**	**98.04**	–

and sampling methods are applied to the training set. Such methods contain pixel-based approaches with some fixed thresholds used to extract lung areas and select the slices with the most visible lung area. Next, each test set is divided into four quarters, which are then used in an unsupervised updating process, in which quarters are passed to the model sequentially and confident predictions are selected to fine-tune the slice-level classifiers. A slice-level prediction is considered confident in this study if it achieves a probability of at least 0.9 in agreement with the patient-level label.

Table 4.3 illustrates the performance of seven automated models developed to tackle the same three-way classification task using the same train and test datasets. The statistical McNemar's test [113] with the significance level of 0.05 is also used to compare the overall performance of the proposed framework with the aforementioned models. This comparison aims to test the hypothesis that the models have the same proportion of errors on the entire test sets. The corresponding p-values are reported in Table 4.3 and indicate that the hypothesis is rejected for almost all the models except the first one as the corresponding p-value is slightly more than 0.05. In other words, there is a significant difference in the proportion of errors between the proposed framework and six of the aforementioned models, while such a difference is not significant in the case of the model proposed in Reference [96].

Table 4.4: The number of CT slices extracted from each test set of the SPGC-COVID dataset by the proposed enhancement approach to augment the training set.

	COVID-19	CAP	Normal
Test 1	595	0	4
Test 2	382	563	3
Test 3	427	341	2

4.1.6 Discussion

In Table 4.4, the numbers of slices extracted from each test set to augment the train set are presented. The low number of normal slices demonstrates the high performance of the first stage in identifying slices with and without the evidence of infection.

I would like to highlight the effect of the suggested 3% threshold used to identify normal cases based on the outcome of the first stage. As mentioned earlier, 3% is a safe threshold to identify normal cases as it is extremely rare to observe less than 3% involvement of the lung parenchyma in COVID-19 cases. However, it is possible that the number of slices identified as infectious in a normal case exceeds this 3% threshold. This could happen mainly in those CT scans with a large slice-thickness and fewer slices (e.g., less than 100 slices). In such cases, a minor error (a small number of misclassified slices by the first stage) will mistakenly indicate a large involvement of the lung parenchyma. Such errors can be avoided by increasing the 3% threshold or using an adaptive threshold (e.g., based on the slice-thickness and number of slices) when we are dealing with a fewer number of slices per patient. In this study, only one normal case has been misclassified and increasing the threshold to 6% could remove the error while the other cases were not affected. The promising results and benefits of the first stage in identifying slices demonstrating infection, once again, indicate its significant potential to be used in other CT scan-related models to help identify normal cases and concentrate only on a subset of slices rather than the whole volume.

Furthermore, I would like to highlight that the results shown in Table 4.2 demonstrate the incapability of the model enhanced based on a test set to improve the performance of the model on the same set. This is mainly because of the fact that the additional data used to update the benchmark model is constructed by the cases with the highest probability scores (whether correct or not) and incorporating them into the train set will force the model to further increase the corresponding probability scores while does not have much effect on other slices. As such, in the test phase, it is

more reasonable to aggregate the outputs obtained by all enhanced models except the one associated with the target test set.

Finally, it is worth noting that it is possible to design more advanced techniques to select the cases and images from the new test sets using the metrics introduced in the field of Active Learning [90,91] through which the cases which bring more diversity to the training set and the associated feature maps are detected and used for training purposes. In addition to the enhancement techniques in the field of Active Learning, there have been several recent studies on using Generative Adversarial Networks (GANs) to cope with the data and domain shift in medical images [114, 115] where the labeled data is not available in the target domain. The main goal in such frameworks is to achieve a domain invariant image representation which can efficiently embed the important features of the image regardless of the imaging modality or imaging technique. Similarly in Reference [116], an auto-encoder and feature augmentation-based approach is proposed to adapt the model to various imaging modalities obtained by different scanners. In this study, however, we are dealing with only one imaging modality (i.e., CT scan) and the level of characteristic shift between the images is lower compared to the images investigated in the aforementioned studies. Moreover, high performances could be achieved using a far less complicated mechanism.

4.2 The WSO-CAPS Framework

As previously stated, thoracic radiologists have recently tended to use low and ultra-low dose scanning protocols, especially after the emergence of the COVID-19 disease. LDCT and ULDCT, however, suffer from a high noise level which makes them difficult and time-consuming to interpret even by expert radiologists. In addition, some abnormalities are only visible using specific window settings on the radiologists' monitor. Currently, manual adjustment of the windowing settings is the common approach to analyze such low-quality images. In this section, a Window Setting Optimization (WSO) mechanism is embedded into the slice-level classifier proposed in the CT-CAPS and COVID-FACT frameworks to tackle the aforementioned issues and improve the accuracy and interpretability of the proposed DL-models. The resulting model is referred to as the "WSO-CAPS". It is also shown that the performance of the classifier can be improved by using an ensemble

architecture to train multiple WSO modules in parallel and obtain several windowing settings at the same time. Using this mechanism, the WSO-CAPS identifies optimized (WL,WW) pairs that are best suited for the detection of slices demonstrating infection from LDCT and ULDCT images. The experimental results demonstrate that the WSO-CAPS outperforms the original Capsule Network-based model by improving the binary (normal/abnormal) accuracy from 89.4% to 92.0%, sensitivity from 85.4% to 90.3%, and specificity from 92.2% to 93.3%. The superiority of the WSO-CAPS is also demonstrated when it is working with standard dose CT scans.

In what follows, the main idea behind the proposed WSO-CAPS which is the Window Setting Optimization mechanism is explained in detail, followed by the description of the main components in the WSO-CAPS pipeline.

4.2.1 Window Setting Optimization

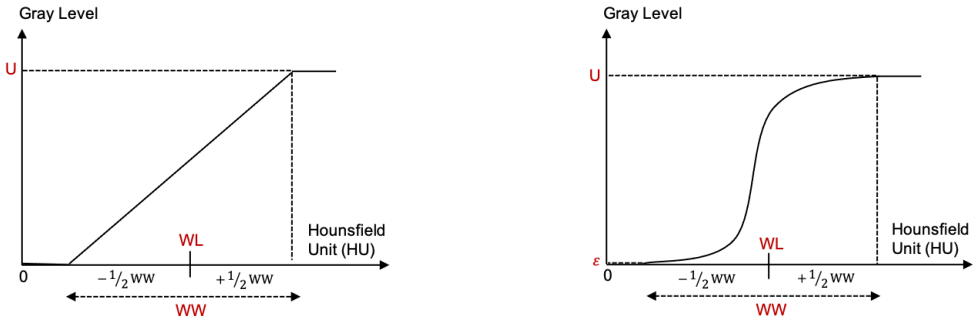
The windowing functions similar to the ones incorporated in radiologist's monitors are adopted to restrict the pixel values in a specific window ranging from 0 to the upper bound of U based on the setting parameters (WL,WW). As shown in Fig. 4.4, linear and sigmoid mappings can be utilized as the windowing function to map all the values inside the window specified by the (WL,WW) to the $[0, U]$, and assign all the values outside the window range to 0 or U. The linear windowing function can be formulated by the Eq. (4.2).

$$F_{lin}(x) = min(max(Wx + b, U), 0), \tag{4.2}$$

where $W = \frac{U}{WW}$ and $b = -\frac{U}{WW}(WL - \frac{WW}{2})$. The sigmoid windowing function can be formulated by the Eq. 4.3.

$$F_{sig}(x) = \frac{U}{1 + \exp -(Wx + b)}, \tag{4.3}$$

where $W = \frac{2}{WW} log(\frac{U}{\epsilon} - 1)$ and $b = \frac{-2WL}{WW} log(\frac{U}{\epsilon} - 1)$. Eq. (4.2) and Eq. (4.3) indicate that the windowing function can be achieved by a convolutional layer with a filter size of 1×1 and a stride of 1, followed by a custom activation layer which is an upper-bounded rectified linear unit (ReLU), or sigmoid function multiplied by U, for the linear or sigmoid windowing function, respectively [66]. The proposed WSO convolutional layer can be used immediately after the input layer to display



(a) Linear Windowing Function (b) Sigmoid Windowing Function

Figure 4.4: Different Windowing Functions, Figure from Reference [6]

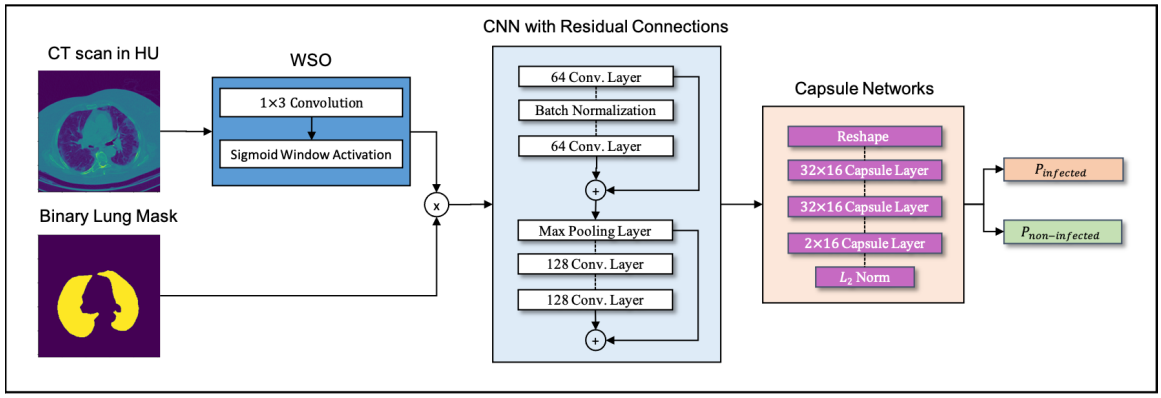


Figure 4.5: WSO-CAPS Pipeline, $\times$ sign represents the element-wise multiplication, $+$ sign denotes the residual addition.

full-range DICOM images in the associated window. This implementation method facilitates finding the optimized window settings on the fly as the weight and bias of the WSO convolutional layer can be trained jointly with the rest of the model to provide the best (WL,WW) pairs.

4.2.2 Proposed Model

The WSO convolutional layer initiates the classification pipeline by converting the input DICOM image from the full-range Hounsfield Unit (i.e. ranges from -1024 to > 4000) into the specific window ranges from 0 to the upper-bound U, which is 1 in this case. Three convolution channels followed by the sigmoid windowing function are used in the proposed WSO-CAPS framework. Following the preprocessing steps of the previously proposed frameworks, the same lung area segmentation, down-sampling, and normalization methods are used as the preprocessing steps. The

Table 4.5: Binary classification results obtained by the Capsule Networks and single channel WSO-CAPS.

Performance	CapsNet	CapsNet with Residual Connection	WSO-CAPS (ReLU)	WSO-CAPS (sigmoid)	WSO-CAPS no lung segmentation (sigmoid)
Accuracy(%)	89.4	89.5	91.4	**91.6**	90.3
Sensitivity(%)	85.5	86.3	**91.7**	89.1	85.7
Specificity(%)	92.2	91.9	91.2	93.5	**93.9**

Table 4.6: Results obtained by different architectures of the WSO-CAPS framework using the sigmoid window activation and the model proposed in Reference [7].

Performance	WSO-CAPS (3 channels)	WSO-CAPS (3 Branches)	WSO-CAPS (3 Channels - 3 Branches)	ResNet50 (Ref [7])
Accuracy(%)	**92.0**	91.0	91.5	83.1
Sensitivity(%)	**90.3**	88.5	88.4	76.4
Specificity(%)	93.3	92.8	**93.7**	88.0

preprocessed images are then fed into the Capsule Network-based classifier which is adopted from the architecture proposed in the previous section (i.e. the robust and enhanced classification framework) which demonstrated a high performance on heterogeneous chest CT images including LDCT and ULDCT.

As shown in Fig. 4.5, the inputs of the WSO-CAPS are the original CT images and the corresponding lung mask generated by the R231CovidWeb segmentation model, which are then followed by the WSO convolution layer with the size of 1×3 to detect 3 pairs of (WL,WW) at the same time. The output of the WSO layer is then fed to a stack of four convolutional layers with the sizes of $64, 64, 128, 128$, respectively; followed by three capsule layers such that the amplitude of the last layer represents the probability of the input image belonging to each target class. It is also worth mentioning that the same weighted loss function as that of the other proposed models in this thesis is used to train the WSO-CAPS.

4.2.3 Experimental Results

The WSO-CAPS model is trained based on the Low-Dose, Ultra-Low Dose, and simulated Low-Dose CT scans from the WSO Dataset described in Sub-section 2.4.3. 70% of the cases (i.e.,

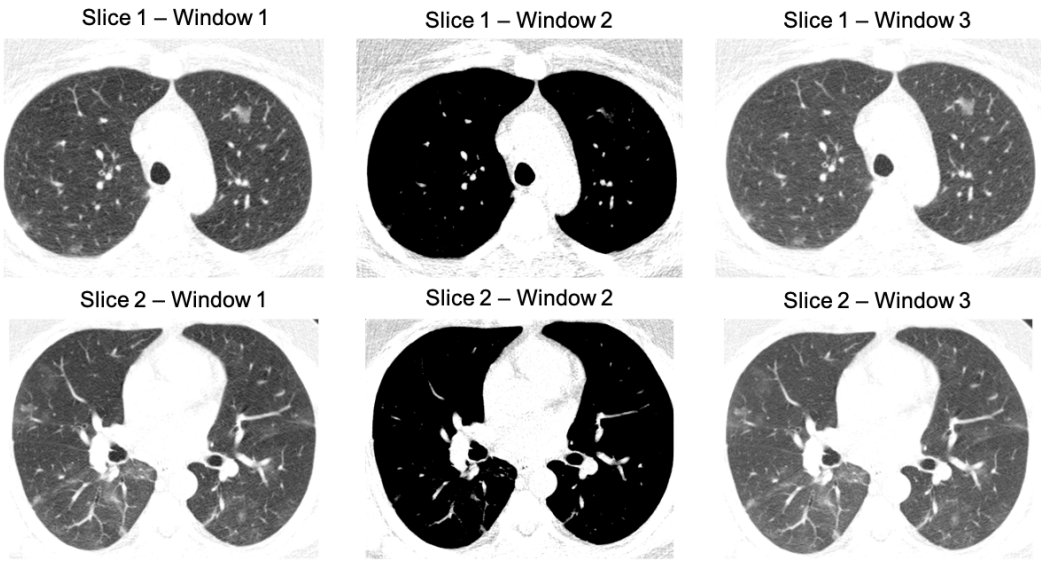


Figure 4.6: Effects of the three optimized window settings identified by the WSO-CAPS on two sample chest CT slices.

154 cases) is randomly selected as the train set from which 10% is randomly selected as the validation set to determine the best model during the training phase. The remaining cases are used for evaluation purposes. It is worth mentioning that the data leakage is strictly prevented between the train and test sets. A batch size of 32, a learning rate of $1e-4$, and 100 epochs were used as the training hyperparameters. Moreover, the weight and bias of the WSO convolution layer were initiated based on the values that corresponded to the standard windowing parameters for the lung CT scans (i.e., $WL=-500, WW=1400$). As the first experiment, the models developed based on sigmoid and ReLU activation functions are investigated. The performance of the WSO-CAPS framework without using the lung segmentation model is also evaluated. The corresponding results are provided in Table 4.5. Based on the obtained results, the sigmoid windowing function using lung segmentation is selected as the best model. In the next step, to further improve the capability of the WSO-CAPS in detecting abnormality CT manifestations through different windows, two experiments are conducted. In the first experiment, the size of the WSO convolution layer is increased to 3, while in the second experiment, an ensemble architecture is adopted, which trains three branches of the WSO-CAPS model in parallel which are merged in the intermediate capsule layer using a concatenation layer.

The performance of the WSO-CAPS is also compared with the ResNet50 model used in Reference [7]. The related results are presented in Table 4.6. Tables 4.5 and 4.6 indicate that the

WSO-CAPS framework with 1 branch and 3 WSO convolution channels using the sigmoid windowing function outperforms its counterparts. The results also demonstrate that all the models equipped with the WSO mechanism outperform the same models without using the window adaptation layer. It is worth mentioning that increasing the complexity of the framework by adding more convolution channels and branches could not further improve the performance. In the last step, to provide a better insight into the proposed window setting optimization module,the identified (WL,WW) pairs are investigated and the CT images are reviewed through the obtained window settings, considering the $\epsilon = 0.01$ in Eq. (4.3). The optimized setting parameters obtained by the WSO-CAPS framework using 1 channel is $(-555.9, 1032.0)$, which is quite similar to the standard setting but adds more contrast and noise reduction power to the model. The WSO-CAPS using 3 channels obtained $(-592.4, 1095.7)$, $(-277.1, 517.8)$, and $(-630.4, 1165, 4)$ as the optimized parameters. The first identified window setting in this case, is also close to the standard one which helps the model not to miss the details evident through the standard window. To better visualize the effects of the obtained parameters, Fig. 4.6 illustrates two sample CT images displayed by the optimized settings obtained by the WSO-CAPS using 3 channels. The capability of the WSO-CAPS to view the lung entities through different windows is evident in Fig. 4.6. It can also be concluded that the second window focuses more on the structure of the lung and vessels and removes the noisy and infectious components, while the first and third windows visualize the infection manifestations at different contrast levels. In another experiment, the WSO-CAPS framework is trained using standard-dose CT scans to further investigate the generalizability of the model. In this case, the WSO-CAPS achieved the accuracy of 91.6%, sensitivity of 92.0%, and specificity of 91.4% while the CapsNet model without incorporating the WSO module showed a lower performance by achieving the accuracy of 90.5%, sensitivity of 89.8%, and specificity of 90.7%. Therefore, similar to the Low-Dose CT scans, the superiority of the WSO-CAPS over its counterparts is evident when dealing with standard-dose CT scans.

4.3 Summary

In the chapter, the COVID-FACT framework is first extended to tackle the three-way classification task (i.e., identification of COVID-19, CAP, and normal cases) based on volumetric CT scans acquired from multiple centers using different imaging protocols. An unsupervised enhancement approach is also proposed, which can enable a DL-based framework to be adapted to the heterogeneity in different test sets. This enhancement approach updates the model's parameters by extracting confident predictions from the test sets and utilize them to re-train the model in order to increase its capability and robustness in the presence of gaps between the imaging protocols and patients' clinical history. It is shown that different versions of the model can be trained based on different test sets and their outputs can be combined to generate the final predictions, which are more accurate and robust.

In addition, this chapter proposed the WSO-CAPS framework to identify slices demonstrating infection from low and ultra-low-dose volumetric CT scans. The WSO-CAPS framework benefits from a Window Setting Optimization module, which is implemented by a 1×3 convolution layer followed by a sigmoid-based window activation function. The experimental results on the in-house WSO dataset indicate that incorporation of the WSO module into the classification models will improve the performance. The proposed WSO-CAPS improved the accuracy of the slice-level classifier by 2.6%, and achieved the accuracy of 92.0%, the sensitivity of 90.3%, and the specificity of 93.3%. It is also showed that the WSO-CASP using 3 WSO convolution channels will provide better results compared its variant using a single channel. It is worth mentioning that detecting infectious slices in a volumetric CT scan is an integral component of many state-of-the-art frameworks dealing with volumetric medical images. As such, the WSO-CAPS is expected to be beneficial to other DL-based frameworks working with 3D CT scans.

Chapter 5

Invasiveness Prediction of Lung Adenocarcinoma Subsolid Nodules from Non-Thin Section 3D CT Scans

Lung cancer is the leading cause of mortality from cancer worldwide and has various histologic types, among which Lung Adenocarcinoma (LUAC) has recently been the most prevalent. Lung adenocarcinomas are classified as pre-invasive, minimally invasive, and invasive adenocarcinomas. Timely and accurate knowledge of the invasiveness of lung nodules leads to a proper treatment plan and reduces the risk of unnecessary or late surgeries. Currently, the primary imaging modality to assess and predict the invasiveness of LUACs is the chest CT. The results based on CT images, however, are subjective and suffer from a low accuracy compared to the ground truth pathological reviews provided after surgical resections. In this chapter, a predictive transformer-based framework, referred to as the “CAE-Transformer", is developed to classify LUACs. The CAE-Transformer utilizes a Convolutional Auto-Encoder (CAE) to automatically extract informative features from CT slices, which are then fed to a modified transformer model to capture global inter-slice relations. Experimental results on the in-house dataset of 114 pathologically proven Sub-Solid Nodules (SSNs) demonstrate the superiority of the CAE-Transformer over the histogram/radiomics-based models and its DL-based counterparts, achieving an accuracy of 87.58%, sensitivity of 86.67%, specificity of

88.0%, and AUC of 0.88, using a 10-fold cross-validation. The influence of Positional Embedding (PE), Global Max Pooling (GMP), Global Average Pooling (GAP) layers, and Feature Concatenation, which are commonly used in transformer-based models to aggregate the encoded instances and generate the final output, are also investigated in this study. The results show that using the PE in conjunction with GMP achieves the highest accuracy for the task at hand.

The LUAC dataset, described in Sub-section 2.4.4, is used to train and evaluate the CAE-Transformer model. Experimental results shows that DL-based models (including the CAE-Transformer) improve the results achieved by the ML-based models proposed in Reference [40], which are essentially using histogram-based and radiomics features to predict the SSN's invasiveness. More specifically, the CAE-Transformer improved the accuracy from 81.0% to 87.58%, sensitivity from 80.0% to 86.67%, and specificity from 81.8% to 88.0%, while achieving a slightly lower AUC value of 0.88 compared to the original AUC of 0.91.

5.1 CAE-Transformer Framework

In this section, different components of the proposed CAE-Transformer framework are described in detail. Fig. 5.1 shows an overview of the CAE-Transformer framework, along with the architecture of the associated transformer encoder.

5.1.1 Preprocessing

Following the preprocessing steps introduced in previous chapters, lung regions are first extracted, using the well-trained U-Net-based lung segmentation model developed in Reference [42], and the output images are down-sampled from $(512, 512)$ to $(256, 256)$ to reduce the complexity and memory allocation without significant loss of information. In this study, however, the CT images are not normalized in the $[0, 1]$ interval to preserve the original pixel intensity distributions for each nodule.

5.1.2 Convolutional Auto-Encoder (CAE)

In order to represent CT images by compressed and informative feature maps, to be used as the input of the subsequent modules, a Convolutional Auto-Encoder (CAE) is initially pre-trained based

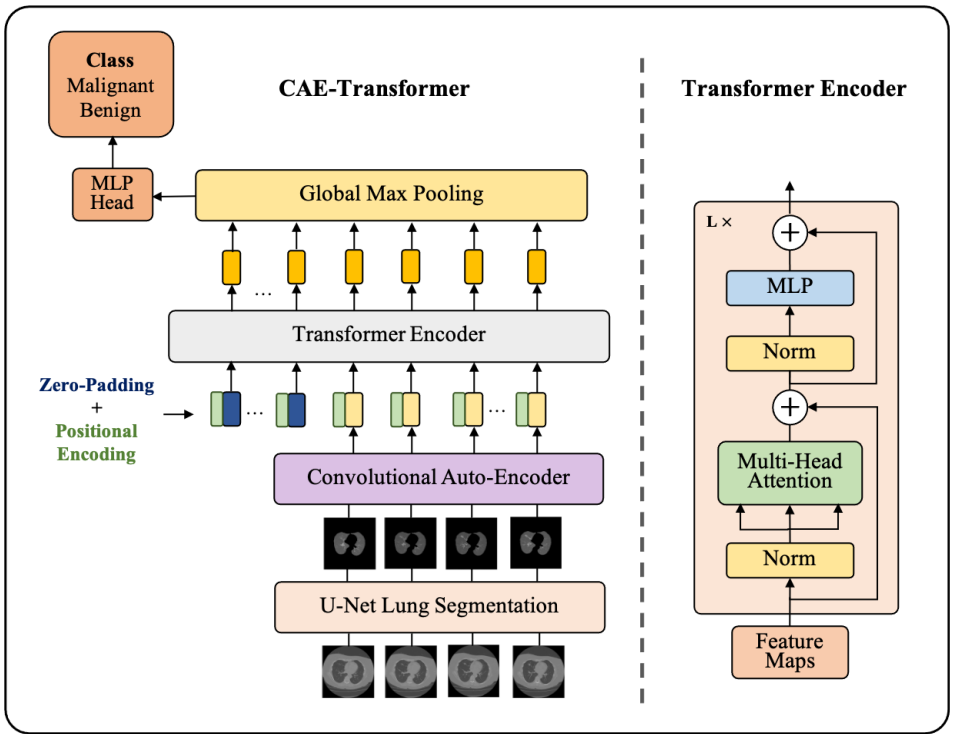


Figure 5.1: Left: Overview of the proposed CAE-Transformer framework, Right: Architecture of the Transformer Encoder

on the public LIDC-IDRI dataset, which contains $244,527$ chest CT images with or without the evidence of a nodule. The CAE model consists of an encoder and a decoder part. The encoder is responsible for generating a compressed representation of the input image via a stack of 5 convolution and 5 max-pooling layers, followed by a fully-connected layer with the size of 256, while the decoder part attempts to reconstruct the original image using the compressed feature representation generated by the encoder. By minimizing the Mean Squared Error (MSE) between the original and the reconstructed image, the CAE learns to produce highly informative feature representations for the input images. Finally, the pre-trained model is fine-tuned on the CT images in the LUAC dataset.

5.1.3 Proposed Transformer

The transformer model used in the proposed CAE-Transformer framework is adopted from the transformer encoder proposed in References [44, 74], and modified for the task of interest. More specifically, a transformer encoder is initialized by applying the MHA module, described in Subsection 2.3.1 , on the normalized CAE-generated feature maps corresponding to input instances

(i.e., CT slices), followed by a residual connection which adds low-level features of the input to the output of the MHA module. A Layer Normalization (LN) is then applied to the results. The normalized values are then passed to the next module which contains a Multi-Layer Perceptron (MLP), followed by another residual connection as shown in Fig. 5.1. The CAE-Transformer is constructed by stacking 3 transformer encoder blocks on top of each other with projection dimensions of 256 , key dimensions of 128, and 5 number of heads in each MHA module. Finally, the features obtained by the stack of transformer encoders from all input instances are passed to a GMP layer, and a fully-connected layer with 32 neurons is applied to generate the final binary classification results. The final fully-connected layer uses a softmax activation function to produce probability scores. Dropout layers are also incorporated to prevent the model from getting over-fitted.

It is also worth mentioning that in conventional transformers and their modified versions (e.g., ViT and CvT), some information about the position of instances in the input sequence (e.g., relative or absolute positions) are added into the model in different encoded forms such as Positional Embeddings (PE) or Token Embeddings (TE). Following the literature, the CAE-Transformer incorporates the PE layer in its pipeline to embed more information about the position of each slice in a series of CT images. To further investigate the effects of this layer on the final results, PE is removed in separate experiments and the obtained results are reported for comparison. In addition, as the number of slices with the evidence of a nodule varies between different subjects (from 2 to 25 slices per nodule), the maximum number of such slices in the LUAC dataset (i.e., 25 slices) is taken, and the input sequences are zero-padded based on this number, so that all sequences have the same dimension of $(25, 256)$. The following equations describe how the CAE-Transformer's output is obtained.

$$
\begin{aligned}
&(s_1{}', s_2{}', \cdots, s_{c_i}{}') = \textit{U-Net}(s_1, s_2, \cdots, s_{c_i}), && i = 1 \cdots N \\
&(f_1, f_2, \cdots, f_{c_i}) = CAE((s_1{}', s_2{}', \cdots, s_{c_i}{}'), && i = 1 \cdots N \\
&z_0 = ZeroPad(f_1, f_2, \cdots, f_{c_i}) + PE, && i = 1 \cdots N \\
&z_l{}' = MSA(LN(z_{l-1})) + z_{l-1}, && l = 1 \cdots L \\
&z_l = MLP(LN(z_l{}')) + z_l{}', && l = 1 \cdots L \\
&o = LN(z_L), \\
&x = GMP(o_1, o_2, \cdots, o_{25}), \\
&y = MLP(x), && (5.1)
\end{aligned}
$$

where s denotes the original CT slices, s' represents the segmented CT images, c_i signifies the number of slices with the evidence of a nodule in the case i, f represents the CAE-generated feature maps corresponding to the CT images, MSA denotes the Multi-Head Self-Attention module, and l shows the l_{th} MSA layer. The number 25 indicates the maximum number of slices with the evidence of a nodule per subject in this study, and y is the final prediction.

5.2 Experimental Results

The performance of the proposed CAE-Transformer framework is evaluated using the 10-fold cross-validation method. The CAE model is pre-trained using a batch size of 128, learning rate of $1e-4$ and 200 epochs. The best model on the randomly sampled 20% of the dataset was selected as the best model. The model was then fine-tuned on the LUAC dataset using a lower learning rate of $1e-6$ and 50 epochs. To fine tune the final CAE, only the middle fully-connected layer and its previous and next convolution layers were trained while the other layers have been kept unchanged. The CAE-generated features are then used to train the transformer encoder. The transformer is trained using a learning rate of $1e-3$, batch size of 64, and 200 epochs. The results of the CAE-Transformer are presented in Table 5.1.

The performance of the proposed CAE-Transformer framework is compared with the results

Table 5.1: Results obtained by the CAE-Transformer and its counterparts. Concat. refers to the Feature Concatenation aggregation function.

Model	Accuracy (%)	Sensitivity (%)	Specificity (%)	AUC
Ref. [40]	81.0	80.0	81.80	0.91
GMP-FC	84.02	**87.0**	80.67	0.90
GAP-FC	83.18	85.33	80.67	0.90
CAE-LSTM	84.92	85.0	84.33	0.84
CAE-Transformer (No PE & Concat.)	85.0	83.0	86.0	0.92
CAE-Transformer (No PE & GAP)	81.52	80.0	83.0	0.88
CAE-Transformer (No PE & GMP)	85.0	**87.0**	83.0	**0.93**
CAE-Transformer (Concat.)	85.83	83.0	88.33	0.92
CAE-Transformer (GAP)	85.83	**87.0**	84.67	0.88
CAE-Transformer (GMP)	**87.58**	86.67	**88.0**	0.88

obtained by the ML-based model proposed in Reference [40]. As the other models proposed in the literature are not trained over the same dataset, they are not considered for the comparison. The CAE-Transformer is further compared with non-transformer alternative models by aggregating the CAE-generated feature maps using GMP and GAP, followed by a stack of fully connected and batch normalization layers. The best experimental results for such models were obtained by utilizing 4 fully connected layers with 128, 128, 32, and 2 neurons, respectively. The performance of the CAE-Transformer is also compared with its LSTM-based counterpart, referred to as the "CAE-LSTM", obtained by replacing the transformer blocks with a stack of LSTM layers while using the same hyper-parameters and complexity.

It is worth mentioning that in the transformer architecture, an aggregation mechanism is commonly used to aggregate all the sequential features generated by the last transformer encoder. The proposed CAE-Transformer utilizes a GMP layer in this regard. In addition to the GMP, different aggregation mechanisms (PE, GAP, and Feature Concatenation) are investigated in separate experiments and the results are reported in Table 5.1. In other words, the final GMP layer in the CAE-Transformer is replaced by other aggregation functions to evaluate their influence on the model, while the rest of the model remained the same.

The experimental results provided in Table 5.1 show that most DL-based models outperform the original radiomics and ML-based model, while the CAE-Transformer using PE and GMP achieves the highest accuracy and specificity among the developed frameworks. It is worthy of note that increasing the complexity of the model and changing hyper-parameters could not improve the performance when GAP and Feature Concatenation were included or when the PE was removed.

5.3 Summary

In this chapter, an automated transformer-based framework, referred to as the “CAE-Transformer", is proposed to enhance the existing radiomics and ML-based models aiming to predict the invasiveness of lung adenocarcinoma subsolid nodules from 3D CT scans. The proposed CAE-Transformer framework significantly improved the performance of the previously developed models by increasing the accuracy by 6.58%, sensitivity by 6.67%, and specificity by 6.2%. The CAE-Transformer is also capable of capturing global inter-slice relations in a volumetric CT scan while requiring less computational resources compared to RNN and LSTM-based frameworks.

Chapter 6

Summary and Future Research Directions

This chapter concludes the thesis with a list of main contributions made in this dissertation and some proposed directions for future works.

6.1 Summary of Thesis Contributions

The research works presented in this thesis are motivated by recent advances in the design and implementation of AI and DL-based models for image and data processing, aiming to develop decision support and stand-alone models assisting healthcare professionals in diagnose and prognosis of critical illnesses. Considering recent progress in development of innovative DL-based architectures, particularly Capsule Networks, Convolutional Auto-Encoder (CAE), U-Net, and Transformers, this thesis aimed to tackle the limitations and drawbacks of the existing solutions, focusing on two particular tasks, i.e., COVID-19 diagnosis and Lung Cancer invasiveness prediction. In this regard, the thesis made a number of contributions, as briefly outlined below:

(1) **CT-CAPS and COVID-FACT Frameworks:** Two fully automated frameworks are proposed to automatically diagnose COVID-19 positive cases from volumetric CT scans. Both proposed frameworks utilize Capsule Networks, as their main building block. Therefore, they are capable of addressing the failure of the commonly used CNN architectures [47] in recognizing

spatial relations between objects in an image and thus eliminating the need for large labeled datasets to reach a satisfying result. The CT-CAPS deals with the development difficulties caused by the large number of CT slices per patient and emphasizes the capability of capsules to represent a large volumetric CT scan by a very small matrix. COVID-FACT is the extension of the CT-CAPS framework, in which the first stage detects slices demonstrating infection in a volumetric CT scan to be analyzed and classified in the next stage. At the second stage, candidate slices detected at the previous stage are classified into COVID and non-COVID cases, and a voting mechanism is applied to generate the patient-level classification scores. COVID-FACT's two-stage architecture has the advantage of being trained on even a weakly labeled dataset, as errors at the first stage can be compensated at the second stage. The proposed frameworks do not require any infection annotation or a very precise slice labeling, which is a valuable asset due to the limited knowledge and experience of the novel COVID-19 disease. In fact, manual infection annotation is completely removed, and the radiologist's input is not required in the test phase. Furthermore, the Grad-CAM localization mapping approach [41] is incorporated into the models to determine lung regions contributing the most to the final decision, aiming to improve the interpretability of the models. The experiments on the in-house COVID-CT-MD dataset showed that both proposed frameworks achieved satisfactory results. CT-CAPS achieved the accuracy of 89.8%, sensitivity of 94.5%, specificity of 83.7%, and AUC of 0.93. COVID-FACT demonstrated a superior performance over the CT-CAPS by achieving an accuracy of 90.82%, sensitivity of 94.55%, specificity of 86.04%, and AUC of 0.97.

(2) **Robust COVID-19 Identification and SPGC-COVID Dataset:** At one hand, this thesis proposed an extension of the COVID-FACT framework, which is tailored to robustly classify volumetric chest CT scans into one of the three target classes (COVID-19, CAP, or normal). The proposed framework integrates a scalable enhancement approach to boost the model's performance and robustness in the presence of gaps between the training and test sets regarding types of scanners, imaging protocols, and technical acquisition parameters. The proposed framework can be generalized on varied external datasets with high flexibility to update itself

upon receiving new external datasets. On the other hand, the thesis introduced a unique test dataset, referred to as the SPGC-COVID dataset. This dataset contains new CT scans with similar characteristics as the COVID-CT-MD dataset as well as noisy low-dose and ultra-low-dose CT scans. In addition, some test CT scans were obtained from patients with a history of cardiovascular diseases or surgeries, which can further challenge the DL-based frameworks. The obtained results showed that while the proposed model is trained on a relatively small dataset acquired from only one imaging center using a specific scanning protocol, it performs well on heterogeneous test sets obtained by multiple scanners using different technical parameters. It is also shown that the model can be updated via an unsupervised approach to cope with the data shift between the train and test sets and enhance the robustness of the model upon receiving new datasets from different centers. More specifically, the subset of the unlabeled test images for which the model generated a confident prediction are extracted and used along with the training set to update the model's parameters. Finally, an ensemble architecture is adopted to aggregate the predictions from multiple versions of the enhanced model. The experimental results on the SPGC-COVID dataset demonstrated that the proposed framework performs well on all test sets and outperforms its counterparts by achieving the overall accuracy of 96.15%, COVID-19 sensitivity of 96.08%, CAP sensitivity of 92.86%, normal sensitivity of 98.04%, and the AUC of 0.992.

(3) **WSO-CAPS:** This model is proposed to efficiently detect slices demonstrating infection from LDCT and ULDCT. The WSO-CAPS framework is essentially an extension of the slice-level classifiers proposed in this thesis and is equipped with a Window Setting Optimization (WSO) mechanism to jointly identify slices with the evidence of infection and find the best window setting parameters to resemble the radiologists' efforts in reviewing LDCT and ULDCT. The experimental results on the WSO dataset show that the WSO-CAPS improves the capability of the Capsule Network and its CNN-based counterparts to identify slices demonstrating infection. It is also shown that the performance of the slice-level classifier can be improved by using an ensemble architecture to train multiple WSO modules in parallel and obtain several windowing settings at the same time. Using this mechanism, the WSO-CAPS identified

optimized (WL,WW) pairs that are best suited for the detection of slices demonstrating infection from LDCT and ULDCT images. The experimental results demonstrated that the WSO-CAPS outperforms the original Capsule Network-based model by improving the binary (normal/abnormal) accuracy from 89.4% to 92.0%, sensitivity from 85.4% to 90.3%, and specificity from 92.2% to 93.3%. The superiority of the WSO-CAPS is also demonstrated when it is working with standard dose CT scans.

(4) **CAE-Transformer:** This framework is proposed to predict the invasiveness of lung adenocarcinomas using volumetric non-thin CT scans. The building block of the CAE-Transformer is the novel multi-head self-attention mechanism and the transformer encoder, which is capable of capturing global inter-slice relations. In addition, unlike current vision transformers which consider different patches in an image as a sequence of data [44, 45], the proposed CAE-Transformer uses a CAE to extract informative features from CT slices and stack them to form a sequential feature map. It is also worth noting that, unlike most existing studies which rely on the nodule patches as the model's input, the CAE-Transformer does not require a detailed annotation of the nodules and takes the whole CT image as the input. The only required information from the radiologists/experts is the set of slices with the evidence of a nodule. Experimental results on the in-house dataset of pathologically proven Sub-Solid Nodules (SSNs) showed that the CAE-Transformer outperforms the ML-based models proposed in Reference [40], which are developed based on histogram-based and radiomics features, by achieving an accuracy of 87.58%, sensitivity of 86.67%, specificity of 88.0%, and AUC of 0.88, using a 10-fold cross-validation.

6.2 Future Research

(1) For the purpose of enhancing the capability of the proposed frameworks in detecting CT slices with an evidence of abnormality, motion artifact reduction algorithms can be utilized to reduce the undesired impacts of such artifacts on the final result.

(2) The idea of utilizing window setting optimization, described in Chapter 4, is still in its early stages. It is hypothesized that the WSO module has a high potential to be incorporated into

the frameworks working with other types of CT scans, particularly the frameworks with an infection detection model in their pipeline. As a direction for future research, the WSO module can be embedded into other DL architectures to determine its effects and validate the aforementioned hypothesis. In addition, it is worth modifying the pipeline to detect the optimized windowing parameters for each slice, instead of the entire cases. This could be achieved by training a simple CNN or Capsule Network-based model focusing on finding best parameters for each slice on the fly.

(3) Another future research direction would be the adoption of the Transformer architecture for COVID-19 identification.

(4) Due to the nature of the datasets used in this thesis (i.e., medical images), obtaining a large and diversified dataset from different countries and cohorts in a short time is challenging. The diversity of the dataset can be expanded in future studies to perform more comprehensive investigations on the generalizability of the proposed frameworks, as well as determining the maximum level of the shift in image characteristics that can be compensated by the unsupervised enhancement approach proposed in Chapter 4.

(5) In the case of the LUAC classification, the size and diversity of the dataset can be increased in future studies to target the three-way SSN classification task. In addition, an ROI/slice selection model can be embedded at the beginning of the pipeline to make the entire framework fully automated.

(6) In the context of LUAC classification, further improvements could possibly be achieved by extracting radiomics features from each SSN, as well as each CT slice, to be added to the CAE-generated feature maps. In addition, GAN-based or feature augmentation-based approaches can be used to obtain a domain invariant image representation to efficiently extract important features, regardless of the imaging modality or acquisition settings. This can potentially improve the performance of the proposed DL-based models in the presence of data and domain shift in the images, similar to the models developed References [114–116].

(7) Incorporating the CNN and/or Capsule Network architectures into the transformer encoder,

similar to the idea proposed in Reference [45], can also be a direction for future studies.

(8) Segmenting lung areas demonstrating infection or abnormality in a CT scan reveals highly beneficial information about the location, size, shape, and distribution of the disease of interest, which ultimately results in valuable assessments. In this regard, it is worth developing a U-net based segmentation model, and modifying the internal layers and activation functions to jointly segment the abnormal areas and classify them into the desired categories.

Bibliography

[1] Shahin Heidarian, Parnian Afshar, Arash Mohammadi, Moezedin Javad Rafiee MD, Anastasia Oikonomou MD, Konstantinos N. Plataniotis, and Farnoosh Naderkhani, "Ct-Caps: Feature Extraction-Based Automated Framework for Covid-19 Disease Identification From Chest Ct Scans Using Capsule Networks," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. jun 2021, pp. 1040–1044, IEEE.

[2] Shahin Heidarian, Parnian Afshar, Nastaran Enshaei, Farnoosh Naderkhani, Moezedin Javad Rafiee, Faranak Babaki Fard, Kaveh Samimi, S. Farokh Atashzar, Anastasia Oikonomou, Konstantinos N. Plataniotis, and Arash Mohammadi, "COVID-FACT: A Fully-Automated Capsule Network-Based Framework for Identification of COVID-19 Cases from Chest CT Scans," *Frontiers in Artificial Intelligence*, vol. 4, may 2021.

[3] Shahin Heidarian, Parnian Afshar, Nastaran Enshaei, Farnoosh Naderkhani, Moezedin Javad Rafiee, Anastasia Oikonomou, Faranak Babaki Fard, Akbar Shafiee, Konstantinos N. Plataniotis, and Arash Mohammadi, "Wso-Caps: Diagnosis Of Lung Infection From Low And Ultra-Lowdose CT Scans Using Capsule Networks And Windowsetting Optimization," in *2021 IEEE International Conference on Autonomous Systems (ICAS)*. aug 2021, pp. 1–5, IEEE.

[4] Shahin Heidarian, Parnian Afshar, Nastaran Enshaei, Farnoosh Naderkhani, Moezedin Javad Rafiee, Anastasia Oikonomou, Akbar Shafiee, Faranak Babaki Fard, Konstantinos N. Plataniotis, and Arash Mohammadi, "Robust Automated Framework for COVID-19 Disease

Identification from a Multicenter Dataset of Chest CT Scans," *ArXiv*, sep 2021.

[5] Shahin Heidarian, Parnian Afshar, Anastasia Oikonomou, Konstantinos N. Plataniotis, and Arash Mohammadi, "CAE-Transformer: Transformer-based Model to Predict Invasiveness of Lung Adenocarcinoma Subsolid Nodules from Non-thin Section 3D CT Scans," *ArXiv*, oct 2021.

[6] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alex Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, Feb. 2017.

[7] Lin Li, Lixin Qin, Zeguo Xu, Youbing Yin, Xin Wang, Bin Kong, Junjie Bai, Yi Lu, Zhenghan Fang, Qi Song, Kunlin Cao, Daliang Liu, Guisheng Wang, Qizhong Xu, Xisheng Fang, Shiqin Zhang, Juan Xia, and Jun Xia, "Using Artificial Intelligence to Detect COVID-19 and Community-acquired Pneumonia Based on Pulmonary CT: Evaluation of the Diagnostic Accuracy," *Radiology*, vol. 296, no. 2, pp. E65–E71, aug 2020.

[8] Yicheng Fang, Huangqi Zhang, Jicheng Xie, Minjie Lin, Lingjun Ying, Peipei Pang, and Wenbin Ji, "Sensitivity of Chest CT for COVID-19: Comparison to RT-PCR," *Radiology*, vol. 296, no. 2, pp. E115–E117, aug 2020.

[9] C. Hani, N.H. Trieu, I. Saab, S. Dangeard, S. Bennani, G. Chassagnon, and M.-P. Revel, "COVID-19 pneumonia: A review of typical CT findings and differential diagnosis," *Diagnostic and Interventional Imaging*, vol. 101, no. 5, pp. 263–268, may 2020.

[10] Marina Carotti, Fausto Salaffi, Piercarlo Sarzi-Puttini, Andrea Agostini, Alessandra Borgheresi, Davide Minorati, Massimo Galli, Daniela Marotto, and Andrea Giovagnoni, "Chest CT features of coronavirus disease 2019 (COVID-19) pneumonia: key points for radiologists," *La radiologia medica*, vol. 125, no. 7, pp. 636–646, jul 2020.

[11] Harrison X. Bai, Ben Hsieh, Zeng Xiong, Kasey Halsey, Ji Whae Choi, Thi My Linh Tran, Ian Pan, Lin-Bo Shi, Dong-Cui Wang, Ji Mei, Xiao-Long Jiang, Qiu-Hua Zeng, Thomas K. Egglin, Ping-Feng Hu, Saurabh Agarwal, Fang-Fang Xie, Sha Li, Terrance Healey, Michael K.

Atalay, and Wei-Hua Liao, "Performance of Radiologists in Differentiating COVID-19 from Non-COVID-19 Viral Pneumonia at Chest CT," *Radiology*, vol. 296, no. 2, pp. E46–E54, aug 2020.

[12] Michael Chung, Adam Bernheim, Xueyan Mei, Ning Zhang, Mingqian Huang, Xianjun Zeng, Jiufa Cui, Wenjian Xu, Yang Yang, Zahi A. Fayad, Adam Jacobi, Kunwei Li, Shaolin Li, and Hong Shan, "CT Imaging Features of 2019 Novel Coronavirus (2019-nCoV)," *Radiology*, vol. 295, no. 1, pp. 202–207, apr 2020.

[13] Heshui Shi, Xiaoyu Han, Nanchuan Jiang, Yukun Cao, Osamah Alwalid, Jin Gu, Yanqing Fan, and Chuansheng Zheng, "Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study," *The Lancet Infectious Diseases*, vol. 20, no. 4, pp. 425–434, apr 2020.

[14] Ming-Yen Ng, Elaine Y. P. Lee, Jin Yang, Fangfang Yang, Xia Li, Hongxia Wang, Macy Mei-sze Lui, Christine Shing-Yen Lo, Barry Leung, Pek-Lan Khong, Christopher Kim-Ming Hui, Kwok-yung Yuen, and Michael D. Kuo, "Imaging Profile of the COVID-19 Infection: Radiologic Findings and Literature Review," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, pp. e200034, feb 2020.

[15] Geoffrey Hinton, Sara Sabour, and Nicholas Frosst, "Matrix capsules with EM routing," *6th International Conference on Learning Representations, ICLR 2018 - Conference Track Proceedings*, pp. 1–29, 2018.

[16] Parnian Afshar, Arash Mohammadi, and Konstantinos N. Plataniotis, "Brain Tumor Type Classification via Capsule Networks," in *2018 25th IEEE International Conference on Image Processing (ICIP)*. oct 2018, pp. 3129–3133, IEEE.

[17] Parnian Afshar, Konstantinos N. Plataniotis, and Arash Mohammadi, "Capsule Networks for Brain Tumor Classification Based on MRI Images and Coarse Tumor Boundaries," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. may 2019, pp. 1368–1372, IEEE.

[18] Parnian Afshar, Konstantinos N. Plataniotis, and Arash Mohammadi, "Capsule Networks for Brain Tumor Classification Based on MRI Images and Coarse Tumor Boundaries," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. may 2019, pp. 1368–1372, IEEE.

[19] Parnian Afshar, Anastasia Oikonomou, Farnoosh Naderkhani, Pascal N. Tyrrell, Konstantinos N. Plataniotis, Keyvan Farahani, and Arash Mohammadi, "3D-MCN: A 3D Multi-scale Capsule Network for Lung Nodule Malignancy Prediction," *Scientific Reports*, vol. 10, no. 1, pp. 7948, dec 2020.

[20] Parnian Afshar, Konstantinos N. Plataniotis, and Arash Mohammadi, "BoostCaps: A Boosted Capsule Network for Brain Tumor Classification," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. jul 2020, pp. 1075–1079, IEEE.

[21] Parnian Afshar, Arash Mohammadi, and Konstantinos N. Plataniotis, "BayesCap: A Bayesian Approach to Brain Tumor Classification Using Capsule Networks," *IEEE Signal Processing Letters*, vol. 27, pp. 2024–2028, 2020.

[22] Kang Zhang, Xiaohong Liu, Jun Shen, Zhihuan Li, Ye Sang, Xingwang Wu, Yunfei Zha, Wenhua Liang, Chengdi Wang, Ke Wang, Linsen Ye, Ming Gao, Zhongguo Zhou, Liang Li, Jin Wang, Zehong Yang, Huimin Cai, Jie Xu, Lei Yang, Wenjia Cai, Wenqin Xu, Shaoxu Wu, Wei Zhang, Shanping Jiang, Lianghong Zheng, Xuan Zhang, Li Wang, Liu Lu, Jiaming Li, Haiping Yin, Winston Wang, Oulan Li, Charlotte Zhang, Liang Liang, Tao Wu, Ruiyun Deng, Kang Wei, Yong Zhou, Ting Chen, Johnson Yiu-Nam Lau, Manson Fok, Jianxing He, Tianxin Lin, Weimin Li, and Guangyu Wang, "Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography," *Cell*, vol. 181, no. 6, pp. 1423–1433.e11, jun 2020.

[23] Mathias Meyer, James Ronald, Federica Vernuccio, Rendon C. Nelson, Juan Carlos Ramirez-Giraldo, Justin Solomon, Bhavik N. Patel, Ehsan Samei, and Daniele Marin, "Reproducibility

of CT Radiomic Features within the Same Patient: Influence of Radiation Dose and CT Reconstruction Settings," *Radiology*, vol. 293, no. 3, pp. 583–591, dec 2019.

[24] Lan He, Yanqi Huang, Zelan Ma, Cuishan Liang, Changhong Liang, and Zaiyi Liu, "Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule," *Scientific Reports*, vol. 6, no. 1, pp. 34921, dec 2016.

[25] Mingyue Li, Yalan Dong, Haijun Wang, Weina Guo, Haifeng Zhou, Zili Zhang, Chunxia Tian, Keye Du, Rui Zhu, Li Wang, Lei Zhao, Heng Fan, Shanshan Luo, and Desheng Hu, "Cardiovascular disease potentially contributes to the progression and poor prognosis of COVID-19," *Nutrition, Metabolism and Cardiovascular Diseases*, vol. 30, no. 7, pp. 1061–1067, jun 2020.

[26] Cristina Manera Dorneles, Gabriel Sartori Pacini, Matheus Zanon, Stephan Altmayer, Guilherme Watte, Marcelo C. Barros, Edson Marchiori, Matteo Baldisserotto, and Bruno Hochhegger, "Ultra-low-dose chest computed tomography without anesthesia in the assessment of pediatric pulmonary diseases," *Jornal de Pediatria*, vol. 96, no. 1, pp. 92–99, jan 2020.

[27] Michael Messerli, Thomas Kluckert, Meinhard Knitel, Stephan Wälti, Lotus Desbiolles, Fabian Rengier, René Warschkow, Ralf W. Bauer, Hatem Alkadhi, Sebastian Leschka, and Simon Wildermuth, "Ultralow dose CT for pulmonary nodule detection with chest x-ray equivalent dose – a prospective intra-individual comparative study," *European Radiology*, vol. 27, no. 8, pp. 3290–3299, aug 2017.

[28] Lucia J.M. Kroft, Levinia van der Velden, Irene Hernández Girón, Joost J.H. Roelofs, Albert de Roos, and Jacob Geleijns, "Added Value of Ultra–low-dose Computed Tomography, Dose Equivalent to Chest X-Ray Radiography, for Diagnosing Chest Pathology," *Journal of Thoracic Imaging*, vol. 34, no. 3, pp. 179–186, may 2019.

[29] Seyed Mohammad Hossein Tabatabaei, Hamidreza Talari, Ali Gholamrezanezhad, Bagher Farhood, Habibollah Rahimi, Reza Razzaghi, Narges Mehri, and Hamid Rajebi, "A low-dose chest CT protocol for the diagnosis of COVID-19 pneumonia: a prospective study,"

Emergency Radiology, vol. 27, no. 6, pp. 607–615, dec 2020.

[30] Salar Tofighi, Saeideh Najafi, Sean K. Johnston, and Ali Gholamrezanezhad, "Low-dose CT in COVID-19 outbreak: radiation safety, image wisely, and image gently pledge," *Emergency Radiology*, vol. 27, no. 6, pp. 601–605, dec 2020.

[31] Suneel D. Kamath, Sheetal M. Kircher, and Al B. Benson, "Comparison of Cancer Burden and Nonprofit Organization Funding Reveals Disparities in Funding Across Cancer Types," *Journal of the National Comprehensive Cancer Network*, vol. 17, no. 7, pp. 849–854, jul 2019.

[32] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal, "Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, nov 2018.

[33] Roy S. Herbst, Daniel Morgensztern, and Chris Boshoff, "The biology and management of non-small cell lung cancer," *Nature*, vol. 553, no. 7689, pp. 446–454, jan 2018.

[34] Ha Young Kim, Young Mog Shim, Kyung Soo Lee, Joungho Han, Chin A Yi, and Yoon Kyung Kim, "Persistent Pulmonary Nodular Ground-Glass Opacity at Thin-Section CT: Histopathologic Comparisons," *Radiology*, vol. 245, no. 1, pp. 267–275, oct 2007.

[35] Jinglei Lai, Qiao Li, Fangqiu Fu, Yang Zhang, Yuan Li, Quan Liu, and Haiquan Chen, "Subsolid Lung Adenocarcinomas: Radiological, Clinical and Pathological Features and Outcomes," *Seminars in Thoracic and Cardiovascular Surgery*, jun 2021.

[36] Xiaonan Cui, Marjolein A. Heuvelmans, Shuxuan Fan, Daiwei Han, Sunyi Zheng, Yihui Du, Yingru Zhao, Grigory Sidorenkov, Harry J.M. Groen, Monique D. Dorrius, Matthijs Oudkerk, Geertruida H. de Bock, Rozemarijn Vliegenthart, and Zhaoxiang Ye, "A Subsolid Nodules Imaging Reporting System (SSN-IRS) for Classifying 3 Subtypes of Pulmonary Adenocarcinoma," *Clinical Lung Cancer*, vol. 21, no. 4, pp. 314–325.e4, jul 2020.

[37] Xiaoliang Shao, Rong Niu, Zhenxing Jiang, Xiaonan Shao, and Yuetao Wang, "Role of PET/CT in Management of Early Lung Adenocarcinoma," *American Journal of Roentgenology*, vol. 214, no. 2, pp. 437–445, feb 2020.

[38] Ella A. Kazerooni, John H.M. Austin, William C. Black, Debra S. Dyer, Todd R. Hazelton, Ann N. Leung, Michael F. McNitt-Gray, Reginald F. Munden, and Sudhakar Pipavath, "ACR–STR Practice Parameter for the Performance and Reporting of Lung Cancer Screening Thoracic Computed Tomography (CT)," *Journal of Thoracic Imaging*, vol. 29, no. 5, pp. 310–316, sep 2014.

[39] Keisuke Fujii, Kyle McMillan, Maryam Bostani, Christopher Cagnon, and Michael McNitt-Gray, "Patient Size–Specific Analysis of Dose Indexes From CT Lung Cancer Screening," *American Journal of Roentgenology*, vol. 208, no. 1, pp. 144–149, jan 2017.

[40] Anastasia Oikonomou, Pascal Salazar, Yuchen Zhang, David M. Hwang, Alexander Petersen, Adam A. Dmytriw, Narinder S. Paul, and Elsie T. Nguyen, "Histogram-based models on non-thin section chest CT predict invasiveness of primary lung adenocarcinoma subsolid nodules," *Scientific Reports*, vol. 9, no. 1, pp. 6009, dec 2019.

[41] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *2017 IEEE International Conference on Computer Vision (ICCV)*. oct 2017, pp. 618–626, IEEE.

[42] Johannes Hofmanninger, Forian Prayer, Jeanny Pan, Sebastian Röhrich, Helmut Prosch, and Georg Langs, "Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem," *European Radiology Experimental*, vol. 4, no. 1, pp. 50, dec 2020.

[43] Kyongtae T. Bae, Gita N. Mody, Dennis M. Balfe, Sanjeev Bhalla, David S. Gierada, Fernando R. Gutierrez, Christine O. Menias, Pamela K. Woodard, Jin Mo Goo, and Charles F. Hildebolt, "CT Depiction of Pulmonary Emboli: Display Window Settings," *Radiology*, vol. 236, no. 2, pp. 677–684, aug 2005.

[44] Dosovitskiy Alexey, Beyer Lucas, Kolesnikov Alexander, Weissenborn Dirk, Zhai Xiaohua, Unterthiner Thomas, Dehghani Mostafa, Minderer Matthias, Heigold Georg, Gelly Sylvain, Uszkoreit Jakob, and Houlsby Neil, ," *ArXiv*.

[45] Wu Haiping, Bin Xiao, Codella Noel, Liu Mengchen, Dai Xiyang, Yuan Lu, and Zhang Lei, "CvT: Introducing Convolutions to Vision Transformers," *ArXiv*, mar 2021.

[46] Jonathan Masci, Ueli Meier, Dan Ciresan, and Jurgen Schmidhuber, "Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction," pp. 52–59. 2011.

[47] Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do, and Kaori Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights into Imaging*, vol. 9, no. 4, pp. 611–629, aug 2018.

[48] Linda Wang, Zhong Qiu Lin, and Alexander Wong, "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images," *Scientific Reports*, vol. 10, no. 1, pp. 19549, dec 2020.

[49] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, may 2017.

[50] Prabira Kumar Sethy, Santi Kumari Behera, Pradyumna Kumar Ratha, and Preesat Biswas, "Detection of coronavirus Disease (COVID-19) based on Deep Features and Support Vector Machine," *International Journal of Mathematical, Engineering and Management Sciences*, vol. 5, no. 4, pp. 643–651, aug 2020.

[51] Tanvir Mahmud, Md Awsafur Rahman, and Shaikh Anowarul Fattah, "CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization," *Computers in Biology and Medicine*, vol. 122, pp. 103869, jul 2020.

[52] Shuyi Yang, Longquan Jiang, Zhuoqun Cao, Liya Wang, Jiawang Cao, Rui Feng, Zhiyong Zhang, Xiangyang Xue, Yuxin Shi, and Fei Shan, "Deep learning for detecting corona virus

disease 2019 (COVID-19) on high-resolution computed tomography: a pilot study," *Annals of Translational Medicine*, vol. 8, no. 7, pp. 450–450, apr 2020.

[53] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," pp. 234–241. 2015.

[54] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. jun 2009, pp. 248–255, IEEE.

[55] Shaoping Hu, Yuan Gao, Zhangming Niu, Yinghui Jiang, Lao Li, Xianglu Xiao, Minhao Wang, Evandro Fei Fang, Wade Menpes-Smith, Jun Xia, Hui Ye, and Guang Yang, "Weakly Supervised Deep Learning for COVID-19 Infection Detection and Classification From CT Images," *IEEE Access*, vol. 8, pp. 118869–118883, 2020.

[56] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam, "Rethinking Atrous Convolution for Semantic Image Segmentation," *ArXiv*, jun 2017.

[57] Kensho Hara, Hirokatsu Kataoka, and Yutaka Satoh, "Learning Spatio-Temporal Features with 3D Residual Networks for Action Recognition," in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*. oct 2017, pp. 3154–3160, IEEE.

[58] Duy M. H. Nguyen, Duy M. Nguyen, Huong Vu, Binh T. Nguyen, Fabrizio Nunnari, and Daniel Sonntag, "An Attention Mechanism with Multiple Knowledge Sources for COVID-19 Detection from CT Images," *ArXiv*, sep 2020.

[59] Feng Shi, Liming Xia, Fei Shan, Bin Song, Dijia Wu, Ying Wei, Huan Yuan, Huiting Jiang, Yichu He, Yaozong Gao, He Sui, and Dinggang Shen, "Large-scale screening to distinguish between COVID-19 and community-acquired pneumonia using infection size-aware classification," *Physics in Medicine and Biology*, vol. 66, no. 6, pp. 065031, mar 2021.

[60] Xinggang Wang, Xianbo Deng, Qing Fu, Qiang Zhou, Jiapei Feng, Hui Ma, Wenyu Liu, and Chuansheng Zheng, "A Weakly-Supervised Framework for COVID-19 Classification and

Lesion Localization From Chest CT," *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2615–2625, aug 2020.

[61] Parnian Afshar, Shahin Heidarian, Farnoosh Naderkhani, Anastasia Oikonomou, Konstantinos N. Plataniotis, and Arash Mohammadi, "COVID-CAPS: A capsule network-based framework for identification of COVID-19 cases from X-ray images," *Pattern Recognition Letters*, vol. 138, pp. 638–643, oct 2020.

[62] Diego Ardila, Atilla P. Kiraly, Sujeeth Bharadwaj, Bokyung Choi, Joshua J. Reicher, Lily Peng, Daniel Tse, Mozziyar Etemadi, Wenxing Ye, Greg Corrado, David P. Naidich, and Shravya Shetty, "End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography," *Nature Medicine*, vol. 25, no. 6, pp. 954–961, jun 2019.

[63] Jason L. Causey, Yuanfang Guan, Wei Dong, Karl Walker, Jake A. Qualls, Fred Prior, and Xiuzhen Huang, "Lung cancer screening with low-dose CT scans using a deep learning approach," *ArXiv*, jun 2019.

[64] Yuankai Huo, Yucheng Tang, Yunqiang Chen, Dashan Gao, Shizhong Han, Shunxing Bao, Smita De, James G. Terry, Jeffrey J. Carr, Richard G. Abramson, and Bennett A. Landman, "Stochastic tissue window normalization of deep learning on computed tomography," *Journal of Medical Imaging*, vol. 6, no. 04, pp. 1, nov 2019.

[65] Manohar Karki, Junghwan Cho, Eunmi Lee, Myong-Hun Hahm, Sang-Youl Yoon, Myungsoo Kim, Jae-Yun Ahn, Jeongwoo Son, Shin-Hyung Park, Ki-Hong Kim, and Sinyoul Park, "CT window trainable neural network for improving intracranial hemorrhage detection by combining multiple settings," *Artificial Intelligence in Medicine*, vol. 106, pp. 101850, jun 2020.

[66] Hyunkwang Lee, Myeongchan Kim, and Synho Do, "Practical Window Setting Optimization for Medical Image Deep Learning," *ArXiv*, dec 2018.

[67] Dongdong Gu, Guocai Liu, and Zhong Xue, "On the performance of lung nodule detection, segmentation and classification," *Computerized Medical Imaging and Graphics*, vol. 89, pp. 101886, apr 2021.

[68] Chen Gao, Ping Xiang, Jianfeng Ye, Peipei Pang, Shiwei Wang, and Maosheng Xu, "Can texture features improve the differentiation of infiltrative lung adenocarcinoma appearing as ground glass nodules in contrast-enhanced CT?," *European Journal of Radiology*, vol. 117, pp. 126–131, aug 2019.

[69] Johanna Uthoff, Matthew J. Stephens, John D. Newell, Eric A. Hoffman, Jared Larson, Nicholas Koehn, Frank A. De Stefano, Chrissy M. Lusk, Angela S. Wenzlaff, Donovan Watza, Christine Neslund-Dudas, Laurie L. Carr, David A. Lynch, Ann G. Schwartz, and Jessica C. Sieren, "Machine learning approach for distinguishing malignant and benign lung nodules utilizing standardized perinodular parenchymal features from CT," *Medical Physics*, vol. 46, no. 7, pp. 3207–3216, jul 2019.

[70] Guixia Kang, Kui Liu, Beibei Hou, and Ningbo Zhang, "3D multi-view convolutional neural networks for lung nodule classification," *PLOS ONE*, vol. 12, no. 11, pp. e0188290, nov 2017.

[71] Shuang Liu, Yiting Xie, Artit Jirapatnakul, and Anthony P. Reeves, "Pulmonary nodule classification in lung cancer screening with three-dimensional convolutional neural networks," *Journal of Medical Imaging*, vol. 4, no. 04, pp. 1, nov 2017.

[72] M. Mehdi Farhangi, Nicholas Petrick, Berkman Sahiner, Hichem Frigui, Amir A. Amini, and Aria Pezeshk, "Recurrent attention network for false positive reduction in the detection of pulmonary nodules in thoracic CT scans," *Medical Physics*, vol. 47, no. 5, pp. 2150–2160, may 2020.

[73] Samuel G. Armato, Geoffrey McLennan, Luc Bidaut, Michael F. McNitt-Gray, Charles R. Meyer, Anthony P. Reeves, Binsheng Zhao, Denise R. Aberle, Claudia I. Henschke, Eric A. Hoffman, Ella A. Kazerooni, Heber MacMahon, Edwin J. R. van Beek, David Yankelevitz, Alberto M. Biancardi, Peyton H. Bland, Matthew S. Brown, Roger M. Engelmann, Gary E. Laderach, Daniel Max, Richard C. Pais, David P.-Y. Qing, Rachael Y. Roberts, Amanda R.

Smith, Adam Starkey, Poonam Batra, Philip Caligiuri, Ali Farooqi, Gregory W. Gladish, C. Matilda Jude, Reginald F. Munden, Iva Petkovska, Leslie E. Quint, Lawrence H. Schwartz, Baskaran Sundaram, Lori E. Dodd, Charles Fenimore, David Gur, Nicholas Petrick, John Freymann, Justin Kirby, Brian Hughes, Alessi Vande Casteele, Sangeeta Gupte, Maha Sallam, Michael D. Heath, Michael H. Kuhn, Ekta Dharaiya, Richard Burns, David S. Fryd, Marcos Salganicoff, Vikram Anand, Uri Shreter, Stephen Vastagh, Barbara Y. Croft, and Laurence P. Clarke, "The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI): A Completed Reference Database of Lung Nodules on CT Scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931, jan 2011.

[74] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, L ukasz Kaiser, and Illia Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. 2017, vol. 30, Curran Associates, Inc.

[75] Gao Xiaohong, Qian Yu, and Gao Alice, "COVID-VIT: Classification of COVID-19 from CT chest images based on vision transformer models," *ArXiv*, jul 2021.

[76] Ara Abigail E. Ambita, Eujene Nikka V. Boquio, and Prospero C. Naval, "COViT-GAN: Vision Transformer forCOVID-19 Detection in CT Scan Imageswith Self-Attention GAN forDataAugmentation," pp. 587–598. 2021.

[77] Hsu Chih-Chung, Chen Guan-Lin, and Wu Mei-Hsuan, "Visual Transformer with Statistical Test for COVID-19 Classification," *ArXiv*, jul 2021.

[78] Parnian Afshar, Shahin Heidarian, Nastaran Enshaei, Farnoosh Naderkhani, Moezedin Javad Rafiee, Anastasia Oikonomou, Faranak Babaki Fard, Kaveh Samimi, Konstantinos N. Plataniotis, and Arash Mohammadi, "COVID-CT-MD, COVID-19 computed tomography scan dataset applicable in machine learning and deep learning," *Scientific Data*, vol. 8, no. 1, pp. 121, dec 2021.

[79] Working Group 18 Clinical Trials DICOM Standards Committee, "Supplement 142: Clinical Trial De-identification Profiles," *DICOM Standard*, pp. 1–44, 2011.

[80] Siva P. Raman, Mahadevappa Mahesh, Robert V. Blasko, and Elliot K. Fishman, "CT Scan Parameters and Radiation Dose: Practical Advice for Radiologists," *Journal of the American College of Radiology*, vol. 10, no. 11, pp. 840–846, nov 2013.

[81] Shahin Heidarian, Parnian Afshar, Nastaran Enshaei, Farnoosh Naderkhani, Moezedin Javad Rafiee, Anastasia Oikonomou, Akbar Shafiee, Faranak Babaki Fard, Konstantinos N. Plataniotis, and Arash Mohammadi, "SPGC-COVID Dataset," 2021.

[82] Parnian Afshar, Moezedin Javad Rafiee, Farnoosh Naderkhani, Shahin Heidarian, Nastaran Enshaei, Anastasia Oikonomou, Faranak Babaki Fard, Reut Anconina, Keyvan Farahani, Konstantinos N. Plataniotis, and Arash Mohammadi, "Human-level COVID-19 Diagnosis from Low-dose CT Scans Using a Two-stage Time-distributed Capsule Network," *ArXiv*, may 2021.

[83] Francois Pontana, Julien Pagniez, Thomas Flohr, Jean-Baptiste Faivre, Alain Duhamel, Jacques Remy, and Martine Remy Jardin, "Chest computed tomography using iterative reconstruction vs filtered back projection (Part 1): evaluation of image noise reduction in 32 patients," *European Radiology*, vol. 21, no. 3, pp. 627–635, mar 2011.

[84] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*. oct 2017, pp. 2242–2251, IEEE.

[85] Mohammad Rahimzadeh, Abolfazl Attar, and Seyed Mohammad Sakhaei, "A fully automated deep learning-based network for detecting COVID-19 from a new and large lung CT scan dataset," *Biomedical Signal Processing and Control*, vol. 68, pp. 102588, jul 2021.

[86] Nan Yu, Cong Shen, Yong Yu, Minghai Dang, Shubo Cai, and Youmin Guo, "Lung involvement in patients with coronavirus disease-19 (COVID-19): a retrospective study based on quantitative CT findings," *Chinese Journal of Academic Radiology*, vol. 3, no. 2, pp. 102–107, jun 2020.

[87] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar, "Focal Loss for Dense Object Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 2, pp. 318–327, feb 2020.

[88] M. Stone, "Cross-Validatory Choice and Assessment of Statistical Predictions," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 36, no. 2, pp. 111–133, jan 1974.

[89] Ho Yuen Frank Wong, Hiu Yin Sonia Lam, Ambrose Ho-Tung Fong, Siu Ting Leung, Thomas Wing-Yan Chin, Christine Shing Yen Lo, Macy Mei-Sze Lui, Jonan Chun Yin Lee, Keith Wan-Hang Chiu, Tom Wai-Hin Chung, Elaine Yuen Phin Lee, Eric Yuk Fai Wan, Ivan Fan Ngai Hung, Tina Poy Wing Lam, Michael D. Kuo, and Ming-Yen Ng, "Frequency and Distribution of Chest Radiographic Findings in Patients Positive for COVID-19," *Radiology*, vol. 296, no. 2, pp. E72–E78, aug 2020.

[90] Asim Smailagic, Pedro Costa, Hae Young Noh, Devesh Walawalkar, Kartik Khandelwal, Adrian Galdran, Mostafa Mirshekari, Jonathon Fagert, Susu Xu, Pei Zhang, and Aurelio Campilho, "MedAL: Accurate and Robust Deep Active Learning for Medical Image Analysis," in *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. dec 2018, pp. 481–488, IEEE.

[91] Asim Smailagic, Pedro Costa, Alex Gaudio, Kartik Khandelwal, Mostafa Mirshekari, Jonathon Fagert, Devesh Walawalkar, Susu Xu, Adrian Galdran, Pei Zhang, Aurélio Campilho, and Hae Young Noh, "O-MedAL: Online active deep learning for medical image analysis," *WIREs Data Mining and Knowledge Discovery*, vol. 10, no. 4, jul 2020.

[92] Samuel Budd, Emma C. Robinson, and Bernhard Kainz, "A survey on active learning and human-in-the-loop deep learning for medical image analysis," *Medical Image Analysis*, vol. 71, pp. 102062, jul 2021.

[93] Xiao Cai, Feiping Nie, Weidong Cai, and Heng Huang, "Heterogeneous Image Features Integration via Multi-modal Semi-supervised Learning Model," in *2013 IEEE International Conference on Computer Vision*. dec 2013, pp. 1737–1744, IEEE.

[94] Lars Schmarje, Monty Santarossa, Simon-Martin Schroder, and Reinhard Koch, "A Survey on Semi-, Self- and Unsupervised Learning for Image Classification," *IEEE Access*, vol. 9, pp. 82146–82168, 2021.

[95] Alan Agresti and Brent A. Coull, "Approximate is Better than "Exact" for Interval Estimation of Binomial Proportions," *The American Statistician*, vol. 52, no. 2, pp. 119–126, may 1998.

[96] Shubham Chaudhary, Sadbhawna Sadbhawna, Vinit Jakhetiya, Badri N Subudhi, Ujjwal Baid, and Sharath Chandra Guntuku, "Detecting Covid-19 and Community Acquired Pneumonia Using Chest CT Scan Images With Deep Learning," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. jun 2021, pp. 8583–8587, IEEE.

[97] Zaifeng Yang, Yubo Hou, Zhenghua Chen, Le Zhang, and Jie Chen, "A Multi-Stage Progressive Learning Strategy for Covid-19 Diagnosis Using Chest Computed Tomography with Imbalanced Data," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. jun 2021, pp. 8578–8582, IEEE.

[98] Pratyush Garg, Rishabh Ranjan, Kamini Upadhyay, Monika Agrawal, and Desh Deepak, "Multi-Scale Residual Network for Covid-19 Diagnosis Using Ct-Scans," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. jun 2021, pp. 8558–8562, IEEE.

[99] Shuohan Xue and Charith Abhayaratne, "Covid-19 Diagnostic Using 3d Deep Transfer Learning for Classification of Volumetric Computerised Tomography Chest Scans," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. jun 2021, pp. 8573–8577, IEEE.

[100] Fares Bougourzi, Riccardo Contino, Cosimo Distante, and Abdelmalik Taleb-Ahmed, "CNR-IEMN: A Deep Learning Based Approach to Recognise Covid-19 from CT-Scan," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. jun 2021, pp. 8568–8572, IEEE.

[101] Bingyang Li, Qi Zhang, Yinan Song, Zhicheng Zhao, Zhu Meng, and Fei Su, "Diagnosing Covid-19 from CT Images Based on an Ensemble Learning Framework," in *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. jun 2021, pp. 8563–8567, IEEE.

[102] Mingxing Tan and Quoc V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *ArXiv*, may 2019.

[103] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep Residual Learning for Image Recognition," *ArXiv*, dec 2015.

[104] Amir Ebrahimi, Suhuai Luo, and Raymond Chiong, "Introducing Transfer Learning to 3D ResNet-18 for Alzheimer's Disease Detection on MRI Images," in *2020 35th International Conference on Image and Vision Computing New Zealand (IVCNZ)*. nov 2020, pp. 1–6, IEEE.

[105] Saining Xie, Ross Girshick, Piotr Dollar, Zhuowen Tu, and Kaiming He, "Aggregated Residual Transformations for Deep Neural Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. jul 2017, pp. 5987–5995, IEEE.

[106] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q. Weinberger, "Densely Connected Convolutional Networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. jul 2017, pp. 2261–2269, IEEE.

[107] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna, "Rethinking the Inception Architecture for Computer Vision," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. jun 2016, pp. 2818–2826, IEEE.

[108] Sergey Zagoruyko and Nikos Komodakis, "Wide Residual Networks," *ArXiv*, may 2016.

[109] Tianqi Chen and Carlos Guestrin, "XGBoost," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, aug 2016, pp. 785–794, ACM.

[110] James MacQueen et al., "Some methods for classification and analysis of multivariate observations," in *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Oakland, CA, USA, 1967, vol. 1, pp. 281–297.

[111] Nicolas Ewen and Naimul Khan, "Online Unsupervised Learning For Domain Shift In Covid-19 CT Scan Datasets," in *2021 IEEE International Conference on Autonomous Systems (ICAS)*. aug 2021, pp. 1–5, IEEE.

[112] Longlong Jing and Yingli Tian, "Self-Supervised Visual Feature Learning With Deep Neural Networks: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 11, pp. 4037–4058, nov 2021.

[113] Quinn McNemar, "Note on the sampling error of the difference between correlated proportions or percentages," *Psychometrika*, vol. 12, no. 2, pp. 153–157, jun 1947.

[114] Eric Tzeng, Judy Hoffman, Kate Saenko, and Trevor Darrell, "Adversarial Discriminative Domain Adaptation," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. jul 2017, pp. 2962–2971, IEEE.

[115] Prashant Pandey, Prathosh A. P, Vinay Kyatham, Deepak Mishra, and Tathagato Rai Dastidar, "Target-Independent Domain Adaptation for WBC Classification Using Generative Latent Search," *IEEE Transactions on Medical Imaging*, vol. 39, no. 12, pp. 3979–3991, dec 2020.

[116] Euijoon Ahn, Ashnil Kumar, Michael Fulham, Dagan Feng, and Jinman Kim, "Unsupervised Domain Adaptation to Classify Medical Images Using Zero-Bias Convolutional Auto-Encoders and Context-Based Feature Augmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2385–2394, jul 2020.