Titre: Title:	Interpretable Learning Models for Acute Respiratory Distress Syndrome Detection from Chest Radiographs
Auteur: Author:	Mohammad Esmaeel Yahyatabar Arabi
Date:	2023
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Yahyatabar Arabi, M. E. (2023). Interpretable Learning Models for Acute Respiratory Distress Syndrome Detection from Chest Radiographs [Thèse de doctorat, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/56663/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/56663/
Directeurs de recherche: Advisors:	Farida Cheriet, & Philippe Jouvet
Programme: Program:	Génie informatique

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Interpretable learning models for Acute Respiratory Distress Syndrome Detection from Chest Radiographs

MOHAMMAD ESMAEEL YAHYATABAR ARABI
Département de génie informatique et génie logiciel

Thèse présentée en vue de l'obtention du diplôme de *Philosophiæ Doctor*
Génie informatique

Octobre 2023

POLYTECHNIQUE MONTRÉAL
affiliée à l'Université de Montréal

Cette thèse intitulée :

Interpretable learning models for Acute Respiratory Distress Syndrome Detection from Chest Radiographs

présentée par **Mohammad Esmaeel YAHYATABAR ARABI**
en vue de l'obtention du diplôme de *Philosophiæ Doctor*
a été dûment acceptée par le jury d'examen constitué de :

François GUIBAULT, président
Farida CHERIET, membre et directrice de recherche
Philippe JOUVET, membre et codirecteur de recherche
Sylvain DESCHÊNES, membre
Nelson SANCHEZ PINTO, membre externe

DEDICATION

To my loving family . . .

ACKNOWLEDGEMENTS

I am deeply grateful to Professor Farida Cheriet, my supervisor, for her trust, support, and invaluable guidance throughout my doctoral journey. Her guidance and mentorship have been invaluable, enabling me to gain a clear vision to tackle complex problems. I also extend my heartfelt appreciation to Dr. Philippe Jouvet, my co-director, for his steadfast support and valuable insights, which have led me to gain new perspectives on the challenges and facilitated my understanding of the medical domain. Special thanks to Philippe Debanné, our Research Associate, for fostering a collaborative and supportive environment and providing invaluable help writing the articles.

Sincere thanks to Nafiseh Erfan for her support at various stages of my doctorate program. I also express my appreciation to Mark Asmar for engaging in stimulating scientific discussions in the lab. Additionally, I am thankful to Maryam Khani for providing excellent ideas related to my project. Their valuable input has significantly enriched my research experience.

A special heartfelt thanks to my family: Sareh, Mohammad Ali, Zahra, Hadi, and my lovely nephew Mohammad, whose presence is a constant source of inspiration and motivation. I value their unconditional love and heartwarming encouragement. Their support has been a source of strength throughout my academic endeavors.

I am grateful to Bahman Khabiri and Hani Nemati for their warm welcome and support during my time in Canada. To my incredible circle of friends, including Javad Jalali, Mohammad Gorjian, Mohsen Jalali, Ali Yahyatabar, Rohollah Yahyatabar, Hamed Vavadi, Mahdi Aghajani, Madad Safari, Ghasem Bagheri, Hassan Gorji, Ehsan Azizpour, Komeil Zokaei, Ahmad Shahnejat, Fatemeh Firoozmandi, Hani Soleymani, Mahsa Shakeri, Mehdi Niakinezhad, Mohammad Reza Khakbaz, Shirin Shakeri, and Tooran Tavangar, I am deeply moved and appreciative of your constant presence and support during moments of both triumph and struggle.

To all of you, your enduring faith in me has played a pivotal role in shaping the person I have become today. This journey would have been neither feasible nor as fulfilling without the abundant love and encouragement you have showered upon me.

RÉSUMÉ

Le syndrome de détresse respiratoire aiguë (SDRA) est un problème médical potentiellement mortel affectant des individus de tous âges et posant de graves problèmes aux prestataires de soins de santé. Le diagnostic précis et rapide du SDRA est essentiel pour fournir une thérapie rapide et efficace. Cependant, le diagnostic du SDRA reste soumis à une variabilité significative entre les observateurs, ce qui peut entraîner des retards de traitement et compromettre les résultats des patients. Cette thèse vise à concevoir un modèle d'apprentissage en profondeur capable d'aider les professionnels de la santé en offrant un soutien diagnostique précis.

L'objectif principal de cette recherche est de développer et de tester un système de soutien décisionnel interprétable pour le diagnostic du SDRA, pouvant être intégré de manière transparente dans les services hospitaliers. L'accent est mis sur la fourniture d'informations cliniques essentielles et de références de prise de décision transparentes aux cliniciens. Pour atteindre cet objectif, une approche à deux volets est appliquée, utilisant les avantages de la segmentation d'image pour se concentrer sur les zones pulmonaires et fournir des étiquettes locales pour assurer l'interprétabilité de la décision du système. L'algorithme est conçu pour fournir des étiquettes localisées pour les images radiographiques thoraciques (CXR), permettant ainsi un processus de diagnostic plus intuitif et transparent. Par conséquent, le personnel médical aura accès à des références explicites, offrant l'interprétabilité des prédictions de l'algorithme. De plus, un outil basé sur le web est conçu pour assurer son applicabilité dans les contextes cliniques, accélérant le processus de validation et facilitant l'intégration dans les infrastructures cliniques actuelles. Étant donné la rareté des données pour ce travail spécifique, un nouvel ensemble de données est construit, qui répond aux exigences de la recherche et contribue aux progrès de la recherche médicale en tant que ressource utile pour la communauté scientifique.

L'approche proposée pour la détection du SDRA suit la procédure de diagnostic utilisée par le personnel médical expérimenté. La première étape consiste en la segmentation des poumons, qui se concentre uniquement sur les régions pulmonaires et réduit l'impact des bruits de fond non pertinents dans les images CXR. Pour pallier les limitations des modèles de segmentation conventionnels, une nouvelle architecture appelée Dense-Unet est présentée, combinant U-Net et DenseNet pour améliorer la circulation de l'information et éviter les problèmes de disparition de gradient, en particulier en présence de données rares. Afin d'éviter les conséquences négatives potentielles de l'utilisation de la segmentation avant la

classification, en particulier la propagation d'erreurs qui pourraient entraver l'identification efficace du SDRA, l'étude présente une architecture unifiée appelée Dense-YNet, qui fusionne les modèles de segmentation et de classification. Ce nouveau modèle en profondeur peut prédire avec précision des étiquettes locales au niveau des quadrants pulmonaires, attribuant ainsi des scores de gravité à chaque image CXR. Par conséquent, le modèle Dense-YNet peut accepter différentes définitions de la gravité du SDRA, fournissant aux professionnels de la santé une base solide pour les décisions de traitement. La validation externe des performances du modèle est essentielle pour déterminer son applicabilité dans le monde réel. Pour ce faire, un système web convivial pour l'analyse de validation est développé, où des experts médicaux sont invités à examiner et à fournir des commentaires sur les résultats de diagnostic du système. Les sorties des six instances des modèles proposés sont combinées à l'aide d'un processus de vote, permettant ainsi de développer des scores de gravité adaptés et améliorant l'applicabilité du système dans des situations cliniques.

Les résultats expérimentaux indiquent que le modèle Dense-YNet proposé est précis et utile pour diagnostiquer et évaluer la gravité du SDRA. La capacité de segmentation de l'algorithme permet une extraction précise des zones pulmonaires, même dans les images CXR présentant des régions pulmonaires masquées par des maladies, ce qui témoigne de performances de segmentation robustes. La plate-forme basée sur le web conçue autour du modèle Dense-Ynet permet une interaction et une validation en douceur, en en faisant un outil viable pour une utilisation par les cliniciens.

Cette étude de recherche contribue à la détection et à l'évaluation de la gravité du SDRA selon trois points de vue principaux : le développement d'un ensemble de données précieux pour la détection du SDRA, un modèle d'apprentissage en profondeur interprétable compatible avec des données limitées et une plate-forme conviviale pour l'interaction et la validation. Bien que le modèle Dense-YNet proposé montre un potentiel prometteur, une évaluation et une validation complètes sont nécessaires pour déterminer sa généralisabilité et son applicabilité dans le cadre réel. Des recherches supplémentaires devraient se concentrer sur l'expansion de l'ensemble de données, le développement d'une interface de programmation d'application pour une intégration en douceur, la priorisation des efforts de validation, la réalisation d'études statistiques approfondies et l'introduction d'un mécanisme de détection de distribution hors norme pour améliorer la fonctionnalité globale et garantir une utilisation clinique sécurisée.

ABSTRACT

Acute respiratory distress syndrome (ARDS) is a life-threatening medical problem affecting individuals of all ages and raising severe issues for healthcare providers. The accurate and prompt diagnosis of ARDS is essential in delivering fast and effective therapy. However, the diagnosis of ARDS remains subjected to significant inter and intraobserver variability, possibly leading to treatment delays and jeopardizing patient outcomes. This thesis aims to design a deep-learning model capable of assisting medical professionals by offering precise diagnostic support.

The main objective of this research is to develop and test an interpretable decision support system for the diagnosis of ARDS that can be seamlessly integrated into hospital wards. A key focus is on providing essential clinical insights and transparent decision-making references to clinicians. To achieve this goal, a two-pronged approach is applied, using the benefits of image segmentation to focus on the lung areas and providing local labels to ensure the interpretability of the system decision. The algorithm is designed to provide localized labeling for chest x-ray (CXR) images, allowing for a more intuitive and transparent diagnosis process. Consequently, the medical staff will have access to explicit references, offering the interpretability of the algorithm's predictions. Furthermore, a web-based tool is designed to ensure its applicability in clinical settings, accelerating the validation process and facilitating integration into current clinical infrastructures. Given the scarcity of data for this specific work, a new data set is built, which addresses the research requirements and contributes to the progress of medical research as a helpful resource for the scientific community.

The proposed approach for detecting ARDS follows the diagnostic procedure used by experienced medical personnel. The initial step is lung segmentation, which focuses solely on the lung regions and reduces the impact of irrelevant background noise in CXR images. To address the limitations in conventional segmentation models, a novel architecture known as Dense-Unet is presented, combining U-Net and DenseNet to improve information flow and avoid gradient vanishing issues, especially when data scarcity exists. To avoid the potential negative consequences of using segmentation ahead of classification, particularly error propagation that could impede the efficient identification of ARDS, the study presents a unified architecture called Dense-YNet, which fuses segmentation and classification models. This novel deep model can accurately predict local labels at the lung quadrant levels, thereby assigning severity scores to each CXR image. Consequently, the Dense-YNet model can accept various definitions of ARDS severity, providing health professionals with a firm basis for treatment decisions. External validation of the model's performance is crucial in determining

its applicability in the real world. To achieve this, a user-friendly web system for validation analysis is developed, where medical experts are invited to review and provide feedback on the system's diagnostic results. The outputs of six instances of proposed models are combined using a voting process, enabling the development of suitable severity scores and enhancing the system's applicability in clinical situations.

The experimental findings indicate that the proposed Dense-YNet model is accurate and helpful in diagnosing and assessing the severity of ARDS. The segmentation capability of the algorithm allows for accurate extraction of lung areas even in CXR images with concealed lung regions caused by illnesses, indicating robust segmentation performance. The web-based platform designed around the Dense-Ynet allows for smooth interaction and validation, making it a viable tool for clinician use.

This research study contributes to ARDS detection and severity assessment in three main viewpoints: developing a valuable dataset for ARDS detection, an interpretable deep learning model compatible with limited data, and a user-friendly platform for interaction and validation. While the proposed Dense-YNet model shows promising potential, comprehensive assessment and validation are required to determine its generalizability and applicability in the real-world setting. Further research should concentrate on expanding the dataset, developing an application programming interface for smooth integration, prioritizing validation efforts, performing thorough statistical studies, and introducing an Out-of-distribution detection mechanism to improve overall functionality and assure safe clinical usage.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ACRONYMS

AAM	Active Appearance Model
AECC	American-European Consensus Conference
AG-CNN	Attention-Guided Convolution Neural Network
AGM	Attention Gate Mechanism
AI	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
ARDS	Acute Respiratory Distress Syndrome
ASM	Active Shape Model
AUROC	Area Under the Receiver Operating Characteristic curve
CAD	Computer-Aided Diagnosis
cGAN	Conditional Generative Adversarial Network
CHUSJ	Centre Hospitalier Universitaire Sainte-Justine
CNN	Convolutional Neural Network
CRAL	Catego ry-wise Residual Attention Learning
CT	Computed Tomography
CXR	Chest X-Ray
DAL	Dual-path Adversarial Learning
DL	Deep Learning
DWT	Discrete Wavelet Transform
ELU	Exponential Linear Unit
GAN	Generative Adversarial Network
GCNN	Graph Convolutional Neural Network
GLCM	Gray Level Co-occurrence Matrix
GPU	Graphical Processing Unit
Grad-CAM	Gradient-weighted Class Activation Mapping
HOG	Histogram of Oriented Pattern
KNN	K Nearest Neighborhood
ILSVRC	ImageNet Large-Scale Visual Recognition Challenge
LBP	Linear Binary Pattern
IRMA	Image Retrieval in Medical Applications
LIME	Local Interpretable Model-Agnostic Explanations
LIV4D	Laboratoire d'Imagerie et de Vision 4D

MBW	Marker Based Watershedding
ML	Machine Learning
MRF	Markov Random Field
MRI	Magnetic resonance imaging
NLP	Natural Language Processing
PALICC	Pediatric Acute Lung Injury Consensus Conference
PARDIE	Paediatric Acute Respiratory Distress syndrome Incidence and Epidemiology
PET	Positron Emission Tomography
PARDS	Pediatric Acute Respiratory distress Syndrome
PCA	Principal Component Analysis
PDF	Portable Document Format
PICU	Pediatric Intensive Care Unit
R-CNN	Region-Based Convolutional Neural Network
ReLU	Rectified Linear Unit
RF	Random Forest
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic curve
ROI	Region of Interest
SDRA	syndrome de détresse respiratoire aiguë
SE	Squeeze and Excitation
SMF	Simple Median Filter
SVM	Support Vector Machine
TB	Tuberculosis
TMEMPR	Tridiagonal Matrix Enhanced Multivariance Products Representation
UI	User Interface
VGG	Visual Geometry Group

CHAPTER 1 INTRODUCTION

Respiratory failure occurs when the lungs cannot perform their primary task of maintaining normal gas exchange levels. This failure may result in hypoxemia[1], hypercapnia[2], or both. Respiratory failure not only affects the functionality of different organs but can also damage them if abnormal levels of carbon are present in the blood. According to a study published in 2020, the ratio of incidence of Acute Respiratory Distress Syndrome(ARDS) per ICU bed per four weeks was 0.39 in Calgary and Alberta, Canada, which shows that it is a common condition. The percentage may increase due to the outbreak of Covid-19 in recent years. The overall mortality rate for the syndrome is as high as 43%—notably, a third of patients with mild ARDS turn to moderate and severe ARDS. The statistics show the importance of rapid medications for ARDS cases. However, starting treatment has some clinical challenges since the early diagnosis of ARDS is not always straightforward.

Four major reasons are known as causes of occurring respiratory failure. First, it may result from dysfunction of the alveoli[3] in exchanging oxygen and carbon dioxide. The respiratory failure resulting from alveolar damage includes viral or bacterial infections, lung trauma, and exposure to harmful substances. The next cause is airway obstruction, which can manifest as various respiratory disorders, such as asthma or bronchopulmonary obstructive disease. The third reason is the lack of appropriate respiratory control commands from the brain to the associated muscles, which can occur in situations such as neurological disorders and drug overdose. In addition, mechanical lung muscle disorders expanding enough may be another main cause of respiratory failure [4].

The human body is complex and dynamic and, depending on age, genetics, lifestyle, and environmental factors, shows various reactions confronting illnesses. This makes the nature of medical diagnosis non-deterministic. Physicians must consider multiple possibilities that can partly explain the symptoms and signs of pathology. To narrow down, the observer needs to prescribe extra medical examinations to exclude possible diseases. The process is time-consuming and costly, and there is a risk of missing the golden time for starting delivery treatment and misdiagnosing. A vast range of conditions causes ARDS, and patients who develop ARDS exhibit significant heterogeneity. Therefore, there is no gold standard for ARDS since it is not a disease but a syndrome. High inter-[4] and intra- observer[5] variability

[1]Low oxygen levels
[2]High arterial carbon dioxide
[3]Small, hollow, balloon-like structures located at the ends of the bronchioles in the lungs
[4]Differences in diagnoses made by different observers examining the same medical test
[5]Differences in the diagnosis of the same patient by a single clinician on other occations

between healthcare professionals shows the complex nature of diagnosing ARDS. Moreover, an extra load of diseases on specific occasions increases the misdiagnosis rate among healthcare professionals. With the increase in the number of patients in health centers, the risk of false negative detection increases, threatening the lives of patients in health centers. In conclusion, being time-sensitive, high inter and intra-observer variability, and probable mistakes in the diagnosis of ARDS prove the need for a system to diagnose ARDS accurately. In addition to being fast and accurate, the system should have two significant specifications to be applicable in hospital wards. First, it should be automatic since it should perform as a component of the diagnosis pipeline to trigger the attention of clinicians even before their notice. Functioning as an alert system for doctors can reduce false negative rates, leading to improved detection of this life-threatening syndrome. Furthermore, the underlying system must provide clinical insight to perform as an assistant to doctors. Medical decision support systems' interpretability and explainability are crucial concerns that involve technological, legal, ethical, and social factors.

There are several reasons why interpretability in the decision-making system is a crucial clinical demand:

- Clinicians need to understand the rationale behind the decision to get informed about patient care. Without understanding the result, they may hesitate to rely on the system.
- The interpretability is necessary since it may reveal bias or error in the system. The feedback from the clinical part may be helpful for further development or modification of the system.
- Regulatory institutes and ethical standards explicitly state that digital systems should be transparent to ensure that the underlying system makes fair and ethical decisions.
- In case of an error that leads to harm or injury, there should be a clear explanation of how the system arrived at that decision.
- Patients may be more likely to engage with the treatment plan if they understand the system's explanation.

The uncertain nature of the diagnosis of a disease refers to the subjective sensation of inability to explain the patient's health issue precisely. This can result from several factors, including complexity or ambiguity of information, variability, or unpredictable signs of an abnormality of inconsistent data. Thus, the interpretability of a decision-making system is arguably

the most critical requirement in medical tools as it allows doctors to rely on the system, which is essential for it to function as a clinical assistant. It is crucial to investigate the capability of learning-based approaches in developing such systems in terms of their efficacy in interpretability.

Wilhelm Conrad Röntgen did not mention the use of X-rays in medicine when he discovered them in 1895. However, even in those days, the potential of X-rays was recognized by the media. It was stated: "We can foresee that one day these rays will be so perfect that they will only penetrate one layer of the body's soft tissue, while deeper layers will be visible. This could be of immeasurable help in diagnosing countless diseases other than those affecting the bones" [5]. Nowadays, chest radiographs are the most frequently prescribed examination, but they are known to be the most limiting factor in detecting ARDS. They are easy to take and available in every health center. Thus, they can be used as reliable and useful data for diagnosing the syndrome in our research work.

Traditional machine-learning (ML) algorithms have been employed in clinical decision-making systems for decades. Due to their high accuracy, quick computation, and interpretability, these models are reasonably straightforward and have gained widespread adoption until deep learning (DL) emerged.

Deep models have greatly improved the accuracy of computer vision tasks, including chest X-ray (CXR) image classification. The progress in DL gained momentum after the introduction of AlexNet, a convolutional neural network (CNN) model, in 2012 [6]. AlexNet's success in the ImageNet large-scale visual recognition challenge (ILSVRC) marked a significant turning point, as it demonstrated the potential of deep models in various machine-learning tasks. The advancements in DL have been facilitated by the availability of powerful computational resources such as GPUs and TPUs, along with the ever-growing availability of large datasets. Highly efficient software tools have also allowed computer scientists to take full advantage of deep learning. The models' performances in many fields dramatically increased, and automatic and efficient feature representation became available.

The potential of ML approaches, specifically DL methods, is acknowledged in medicine, but their limitations must be understood to make them applicable in hospital wards. The ability of ML methods is not assessed solely based on their recognition performances. One of the biggest issues in utilizing DL methods is their "black box" nature. Understanding how the algorithm makes decisions becomes more difficult as the models become more complex. This is because it is challenging for humans to comprehend the intermediate representations within the multiple layers of a deep model. To be valuable in health centers, a reliable tool must provide clear clinical insights into the underlying pathology [7].

To take full advantage of the efficiency of deep learning while enhancing their interpretability,

researchers have been working on ways to make deep models more transparent, such as presenting visual explanations or pinpointing crucial features that affect a decision. For example, in CNNs, posthoc methods such as gradient-weighted class activation mapping (Grad-CAM) and local interpretable model-agnostic explanations (LIME) provide heat maps to highlight the most effective areas associated with the predicted outputs. However, research studies prove that the provided heatmaps are not precise in medical images and specifically in the analysis of CXRs. In many pulmonary disease classification use cases reported in the literature, irrelevant areas are activated. It can be due to either the inability of the explaining algorithm to provide a precise map or the model's fault in discovering the relevant feature. In summary, while post-hoc explainable approaches are helpful tools, they require extra processing and should be used cautiously and in conjunction with other techniques to provide reliable heat maps. Furthermore, they cannot fully resolve the interpretability problem in CXR image analysis.

1.1 Overal research objective

The high mortality rate and the need for early diagnosis associated with ARDS underscore the importance of an accurate and reliable diagnostic protocol. Furthermore, early diagnosis of ARDS is essential for improving patient prognosis. Using CXR images as one of the most important criteria for diagnosing ARDS is potentially the best option due to its speed and ease of use. Therefore, we are motivated to develop a rapid and accurate deep-learning-based model to analyze CXR images for ARDS diagnosis and address critical aspects of an applicable system.

Moreover, one of the most crucial specifications of such systems is the ability to be interpretable and transparent. The goal is to propose a system that can generate precise results and articulate how it arrived at those conclusions. The model should be designed to handle small datasets since collecting and annotating medical data is costly in terms of time and human resources. Finally, the framework should be easily deployable in clinical infrastructure and meet the constraints of health centers. It should be integrated with existing tools in a hospital or be implemented as a standalone platform. Notably, the framework should allow validation by various experts and with different data sources. Therefore, it is necessary to create a web-based platform specifically designed to facilitate external validation.

To sum up, the overall objective of the thesis is to design and deploy a web-based interpretable platform to detect infiltrates in the lung fields of CXR images to aid in diagnosing ARDS. The approach should be quick, precise, and available for external validation through the web.

1.2 Organization

The organization of this thesis is designed to provide a comprehensive exploration of the research objectives. The thesis is divided into eight chapters, each addressing a specific aspect of the research.

Chapter 2 presents a literature review covering the existing methods and approaches for ARDS detection, chest X-ray classification, lung segmentation, and web-based platforms. This chapter establishes a foundation by discussing the current state of the art and identifying the research gaps the thesis aims to address.

Chapters 3, 4, 5, and 6 explore the specific research objectives in detail. Chapter 3 clearly defines the problem statements and presents the research objectives related to lung field segmentation, ARDS classification, and the development of a web-based platform. These objectives are then addressed individually in Chapters 4, 5, and 6 as separate papers within the thesis. Chapter 4 presents the Dense-UNet model, a novel network proposed for the segmentation of lung fields in chest radiographs [8]. Similarly, Chapter 5 focuses on the joint classification and segmentation for an interpretable diagnosis of ARDS from chest X-rays, detailing the materials, methods, experimental protocol, and results [9]. Chapter 6 presents the development of the PARDS-CxR, a web-based platform for the automatic stratification of ARDS severity, discussing its methodology, experimental design, and results [10].

Finally, Chapter 7 provides a general discussion that reflects on the advantages and limitations of the thesis contributions. It provides insights into the potential future directions for research. Lastly, Chapter 8 concludes the thesis by summarizing the key findings, highlighting their significance, and discussing the contributions made to the field.

CHAPTER 2 LITERATURE REVIEW

This chapter presents a literature review focusing on the detection and diagnosis of ARDS using chest radiographs. The review begins by discussing research studies related to ARDS detection and its clinical diagnosis. Next, it examines various techniques and approaches employed in classifying CXR images with a particular emphasis on DL algorithms. Furthermore, the chapter explores the crucial aspect of lung segmentation, which plays a vital role in accurately identifying lung diseases. The analysis encompasses traditional and DL-based lung segmentation methods, highlighting the extensive research to address this challenge. The chapter also examines the current web-based tools developed for the analysis of chest radiographs.

Through this review, we aim to gain insights into the current state-of-the-art methods and identify potential directions for further research and improvement in ARDS detection and diagnosis.

2.1 ARDS detection

Various diseases can lead to respiratory failures, such as asthma, chronic obstructive pulmonary disease (COPD), ARDS, and acute lung injury (ALI). ARDS can occur when the lungs undergo severe inflammation due to an infection or injury. This inflammation leads to fluid leakage from nearby blood vessels into the alveoli, making breathing progressively challenging. The lungs can become inflamed due to a variety of reasons, such as pneumonia or severe flu, sepsis, a severe chest injury, accidental inhalation of vomit, smoke, or toxic chemicals, near drowning, acute pancreatitis[1], and an adverse reaction to a blood transfusion [11].

The first independent description of ARDS was given by Ashbaugh et al. in 1967 [12]. Since then, numerous studies have focused on therapies, pathogenesis (underlying mechanisms and genetic predisposition), and other clinical aspects of the syndrome (risk factors and epidemiology). ARDS is a life-threatening lung disease that is associated with severe hypoxemia. Clinical trials face challenges due to the variability in identifying ARDS. The most widely accepted definition of ARDS, developed by the american-european consensus conference (AECC) in 1994 [13], is based on four key factors:

- Acute onset

[1] A serious condition where the pancreas becomes inflamed over a short time

- The class of hypoxemia severity must be in the ARDS range ($PaO_2/FiO_2 \leq 200$ mmHg)
- Bilateral infiltrates at the chest radiograph
- No clinical evidence of high left atrial pressure (to exclude cardiogenic cause of lung edema)

The recently introduced Berlin definition classifies ARDS into mild, moderate, or severe categories [1]. The criteria for each category and their corresponding mortality rates are presented in Table 2.1. The most frequent cause of mortality in ARDS is multi-organ failure. The primary source of reference for defining ARDS is the Berlin definition. This thesis will use the Berlin definition as the principal ARDS reference for adults.

Table 2.1 Categorizing ARDS based on Berlin definition and their mortality rates

Severity	PaO_2/FiO_2	**Mortality**
Mild	$200 - 300$	27%
Moderate	$100 - 200$	32%
Severe	≤ 100	45%

The criteria for pediatric acute respiratory distress syndrome (PARDS) are based on the modified Berlin definition, which takes into account the physiological and pathological patterns that are unique to children [14]. The Pediatric Acute Lung Injury Consensus Conference (PALICC) Group published recommendations for diagnosing PARDS in 2015 [15] and updated them in 2022 [16]. According to the ARDS in children, only new infiltrates on chest imaging meet the diagnostic criteria for PARDS, rather than bilateral infiltrates in adults. Patients under the age of 18 are recognized for pediatric care. Progressive infiltrates should be observed on chest imaging such as chest radiography, computed tomography (CT), or ultrasound, which are not fully explained by atelectasis, pleural effusions, or nodules.
ARDS can manifest in patients with normal lung function post-surgery but subsequently develop respiratory difficulties. Moreover, it can stem from non-pulmonary infections that damage the lung tissue, leading to ARDS. Despite extensive research on ARDS-related topics, with over 30,000 articles published on Pubmed since 2010[2], the disease's mortality rate remains high, at approximately 40%. Even among ARDS survivors, a significant number of them experience long-term lung damage, particularly older patients.

A study conducted in 2009 indicated that the mortality rate due to ARDS has not significantly decreased from 1994 despite the epidemiological common assumption [17]. Another

[2] `https://pubmed.ncbi.nlm.nih.gov/?term=ARDS&filter=years.2010-2023`

study highlighted that ARDS is frequently underrecognized and undertreated, which may contribute to its high mortality rate. There is a need for more effective strategies and interventions for the diagnosis, prevention, and management of ARDS in clinical practice [18]. The study found that among ICU admissions in 50 countries, the period prevalence of ARDS was 10.4%. These findings suggest a potential for improvement in ARDS patients' diagnosis and treatment management. Prompt treatment delivery during the initial stages of ARDS is crucial as therapy often fails in cases of advanced ARDS with acute processing and severe refractory hypoxemia. Therefore, it is imperative to diagnose ARDS early as it has a high mortality rate, as shown in Table 2.1.

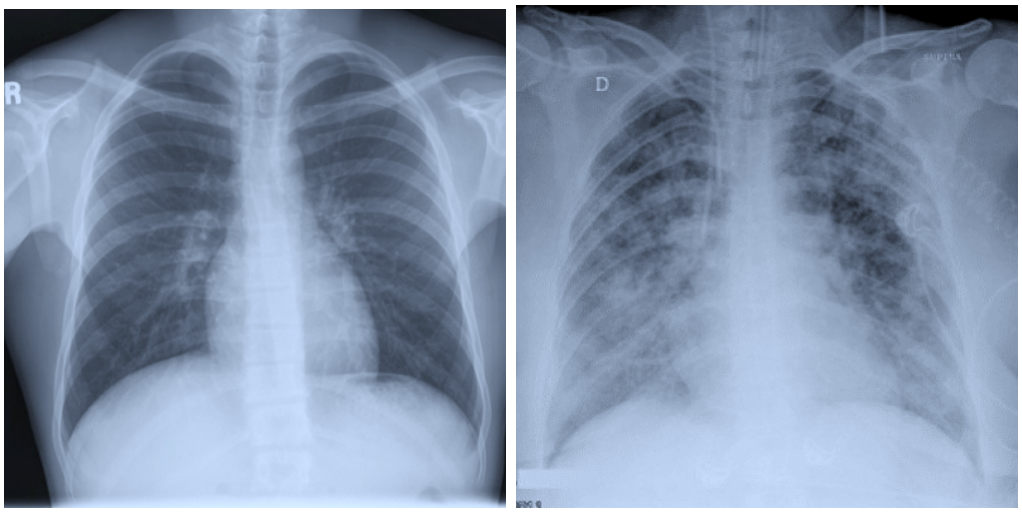
2.1.1 Clinical diagnosis of Acute Respiratory Distress Syndrome

Currently, there are no targeted drugs or therapies available to treat or prevent ARDS directly [19], and more importantly, early diagnosis of ARDS remains a significant challenge. This emphasizes the increased significance of early diagnosis of ARDS. According to the Berlin definition, the diagnosis process is based on three main criteria: timing, oxygenation, and chest imaging. Moreover, it is crucial to exclude cardiac failure as a potential cause. Regarding timing, ARDS should be diagnosed within one week of a known clinical insult or new/worsening respiratory symptoms [1]. Also, the syndrome is defined by hypoxemia, which can be characterized by the ratio of arterial oxygen tension to fractional inspired oxygen (PaO2/FiO2). Identifying the similarities and differences between cardiac failure and ARDS is important for diagnosing ARDS. While these conditions share some features, there are also notable distinctions that can help differentiate between them.

Chest image is the most limiting factor that plays the most distinctive role in ARDS diagnosis. Bilateral opacities for adults with ARDS and new infiltrates for PARDS must be present on chest imaging (chest x-ray or CT scan) that pleural effusions, lobar collapse, or nodules cannot fully explain. Despite the availability of a wide range of imaging modalities, including advanced techniques like positron emission tomography (PET) and magnetic resonance imaging (MRI), X-rays are the preferred imaging treatment due to their low cost, mobility, speed, and modest radiation exposure. Of 22.5 million imaging requests made by doctors in the UK in 2015–2016, 8 million were chest radiographs [20]. In addition to the advantages of X-ray images, CXR images can distinguish ARDS from other causes of acute hypoxemic respiratory failure, such as atelectasis, pleural effusions, cardiogenic pulmonary edema, or pneumonia. Although CXR offers several advantages in diagnosing ARDS, it has certain drawbacks in its ability to identify small or patchy areas of lung injury, detect early changes in lung pathology, and provide functional information regarding lung function, such as the degree of hypoxemia or severity of lung injury. Yet, due to its wide availability, relative price,

and capacity to offer a quick assessment of lung pathology, CXR remains the recommended initial imaging modality.

The fluid accumulation inside the lungs causes infiltration in the lung fields, increasing chest X-rays' opacity. According to the Berlin definition, ARDS usually presents with left-to-right lung involvement, resulting in bilateral infiltration on CXRss. Figure 2.1 compares CXR images for normal and ARDS cases.

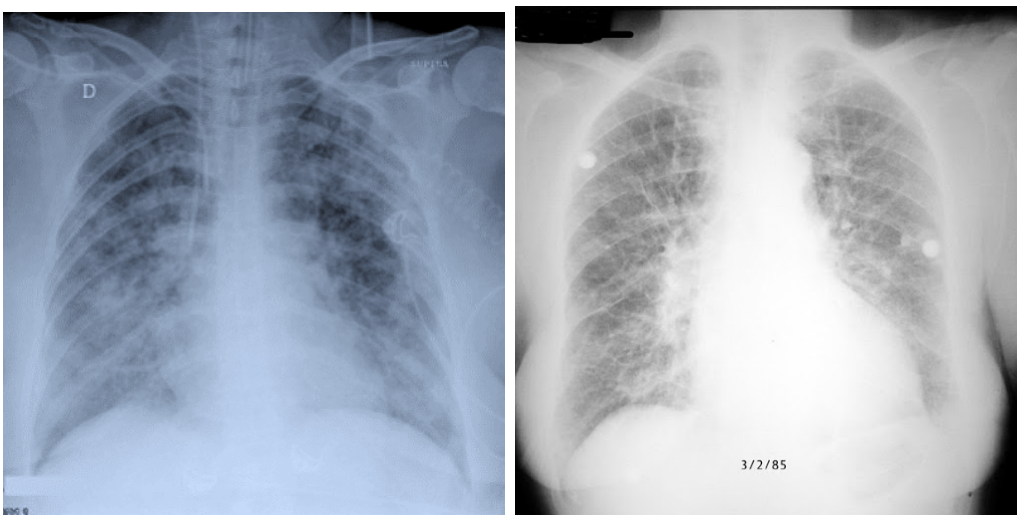


(a) (b)

Figure 2.1 (a) is the normal CXR and bilateral infiltration in CXR of ARDS patient is as (b)

CXR images are frequently utilized in the detection of ARDS, but the moderate inter-observer variability among clinicians (kappa = 0.55) [21] poses a challenge in achieving a consensus on disease identification. The delivery of treatment, which is crucial in cases of severe and pediatric ARDS, may be delayed due to this variability in agreement. Any strategy that reduces this uncertainty could result in a more reliable detection procedure. One of the main difficulties in diagnosing ARDS is ruling out other illnesses that show comparable symptoms in lung imaging, such as concurrent bacterial pneumonia or congestive heart failure. Two CXR images, one for congestive heart failure and the other for ARDS, are shown in Figure 2.2.

From the clinical point of view, the typical diagnosis procedure is a combination of medical history, physical examination, imaging studies, and laboratory tests. In diagnosing ARDS, doctors will first obtain a thorough medical history to fully comprehend the patient's symptoms, previous illnesses, and any recent surgeries or traumatic events that might have triggered ARDS. Then, they will apply the Berlin criteria, which provide clinical guidelines for



(a) (b)

Figure 2.2 (a) is the CXR image for an ARDS case, and (b) is the CXR of a patient with congestive heart failure

defining the severity and cause of ARDS based on the patient's clinical features and imaging tests. As noted previously, the radiographic findings, onset timing, and hypoxemia level are among the criteria. It is normal that during the procedure, clinicians rule out any other medical diseases that might be causing similar symptoms, such as pneumonia, pulmonary embolism, or heart failure.

The timing criterion plays an important role that must not be dismissed. ARDS must occur within one week of a recognized clinical insult or new/worsening respiratory symptoms like hypoxemia or lung infiltrates in CXRs. The first stage in clinical review for detecting ARDS is to investigate chest images. CXR images are the preferred diagnostic tool for clinicians, particularly when time is limited, as mentioned in Section 2.1.1. The radiologist analyses the chest X-ray image and looks for any indications of ARDS, such as fluid buildup, consolidation, and ground-glass opacities, in the lungs. Consolidation is the term used to describe the solidification of lung tissue brought on by inflammation or infection. In contrast, ground-glass opacities are hazy lung spots that indicate fluid buildup. Within the time, there must be bilateral opacities that effusions, lobar, or lung collapse cannot adequately explain or nodules on chest radiographs or CT scans. There should be no indication of heart failure or fluid overload that could explain the respiratory symptoms during the week of timing criterion of Berlin definition.

The Berlin hypoxemia criterion specifies a $PaO2/FiO2$ ratio of less than 300 mmHg in a patient with respiratory distress. Hypoxemia in ARDS is mainly measured by assessing

the partial pressure of arterial oxygen ($PaO2$) via arterial blood gas analysis. The inspired oxygen ($FiO2$) divided by the partial pressure of arterial oxygen ($PaO2$) determines the intended hypoxemia measure, $PaO2/FiO2$ ratio. ARDS is indicated by a $PaO2/FiO2$ ratio of less than 300 mmHg. According to the table, the more severe the syndrome, the lower the rate. Hypoxemia management in ARDS focuses on maintaining sufficient oxygenation while limiting the risk of lung damage from mechanical ventilation. PEEP, or positive end-expiratory pressure, is a mechanical ventilation technique that can assist ARDS patients to breathe more easily.

2.1.2 Machine learning in ARDS detection

Various ML methods are proposed to detect ARDS in its early stages regarding various ARDS criteria presented in ARDS definitions [22–24].

Due to data acquisition challenges, there are not many studies in the literature on detecting ARDS in CXR images. Zaglam et al. removed the rib cages in the lung areas based on [25] and then extracted patches in the intercostal regions [26]. They extracted texture features from the patches based on their statistical and spectral specifications and transformed them into a new space using linear discriminant analysis (LDA). Support vector machine (SVM) was used to classify the samples into ARDS and non-ARDS classes. The study achieved promising results and improved the kappa rate between intensivists. However, removing ribs requires human intervention and is not fully automatic. Few approaches are presented in the literature that use DL models for detecting ARDS in CXRs. Both had to create their dataset as the shortage of ARDS data is previously stated. Sjoding et al. used DenseNet-121 [27] pre-trained on 595,506 chest radiographs. They fine-tuned the model using 8,072 CXRs annotated for ARDS labels. They considered bilateral infiltrates for ARDS as defined in the Berlin definition [1]. The dataset is created at the University of Michigan, and the algorithm is validated by external data [28]. Pai et al. proposed a method based on both ML and DL models in two distinct branches. After pre-processing, they combined conventional ML models, including Random Forest, XGboost, and Logistic regression, to create a label for each CXR image [29]. In the parallel pathway, the pre-trained DenseNet-121 model provides new ARDS labels. In this branch, the images are first segmented to highlight the lung areas using region-based CNN (R-CNN) and U-Net models. By the combination of the probability scores, the ensemble model predicts the ARDS labels.

There are common points in DL-based approaches. Both methods applied pre-trained DenseNet-121 as the backbone of their classification pipelines. Their results confirm the effectiveness of this model in the classification of ARDS. The models perform well in the analysis of ARDS in terms of other pathologies. Moreover, new private datasets are also created to provide

fine-tuning feeds for their models. The need for a new dataset is feasible because first, the labels in public datasets are reported to have errors in labeling, e.g., a 10% error in Chest X-Ray14. This may put our task in trouble because some infiltration-marked samples are not approved by the clinicians from Centre Hospitalier Universitaire Sainte-Justine (CHUSJ), Montreal, Canada. Our clinical investigation acknowledged even more mislabeling in the public dataset. Figure 2.3 shows some wrong labeling of CXRs for infiltrates and non-infiltrates in ChestX-ray14 dataset [2]. Therefore, the data needs to be extended, and newly approved ARDS-labeled cases should be available for training during the study. Second, as stated, no ARDS label is provided in the public datasets.

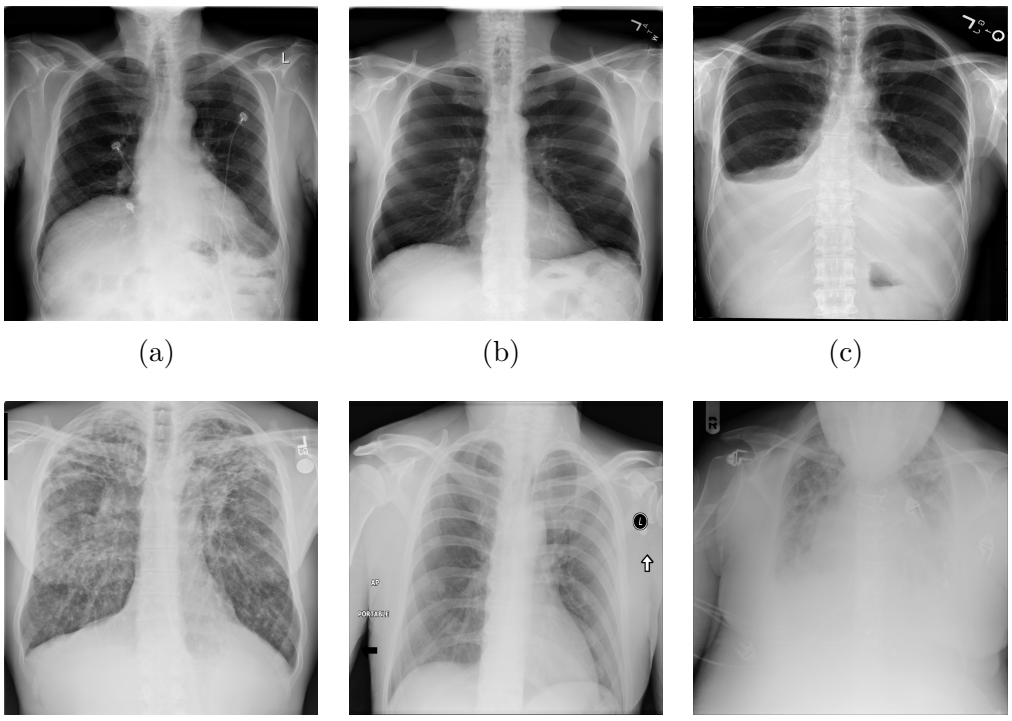
From another point of view, the proposed models can give ideas about interpreting CXRs, but for the detection of ARDS, a local infiltration pattern for sub-regions of the lungs needs to be discovered. No model is specifically presented to consider local features of images. The point is crucial since there are various definitions for ARDS, and not considering local information may put the dynamicity of the approaches in trouble. Thus, the architecture of the applied models or their dataset may need a change to be compatible with the specific task, e.g., ARDS or PARDS.

In summary, the interpretability of the model is a critical clinical demand. This is addressed in Zaglam's work, but the two deep models lack this critical feature. However, Zaglam's approach lacks automaticity since it needs interaction with doctors.

While there is limited research on ARDS, there is a significant body of work focused on the classification of other pathologies, and investigating these classifications could provide valuable insights into CXR analysis.

2.2 Chest X-Ray Classification

The debate in the literature about using computer-aided diagnosis (CAD) systems in the analysis of CXR images dates back to the mid-1950s. However, limited computer resources, the low quality of digitized images, and the lack of leading knowledge made it difficult to succeed in the field [30]. ML-based approaches have revolutionized the analysis of radiography images by using techniques. The potential of CAD systems was restricted by the human engineering of feature sets used in traditional machine-learning methods. A lack of data and resource limitations also hampered progress. Deep learning has been adopted, especially in the classification of CXRs, thanks to the development of strong graphical processor units (GPU) and the accessibility of big medical datasets. CXR analysis is only one of the many medical image applications where deep learning has shown to be quite successful. This section reviews the literature on pulmonary disease classification, including conventional and



(a) (b) (c)

(d) (e) (f)

Figure 2.3 Probable errors in labels of some samples of ChestX-ray14 dataset [2]. The top row shows some samples labeled as infiltration, and the bottom row contains images marked as non-infiltration CXRs [3]

DL approaches.

2.2.1 Conventional approaches

As mentioned, traditional machine learning approaches rely on hand-engineered features and a classifier that distinguishes between classes. The main process for streamlining traditional ML approaches comprises four main steps. After some preprocessing steps, features are extracted and selected to feed the classifier. Figure 2.4 shows these steps and the most common algorithms used in each step.

In a broad overview, preprocessing, feature extraction and selection, and classification are the main steps in machine learning approaches. The region of interest (ROI) is often discovered as a preprocessing phase, after which further processing techniques are applied to it [31–33]. The preprocessed images should undergo feature extraction steps to capture their critical features. Discrete wavelet transform (DWT) and local binary patterns (LBP) are widely utilized in traditional machine learning approaches for feature extraction [32–35].

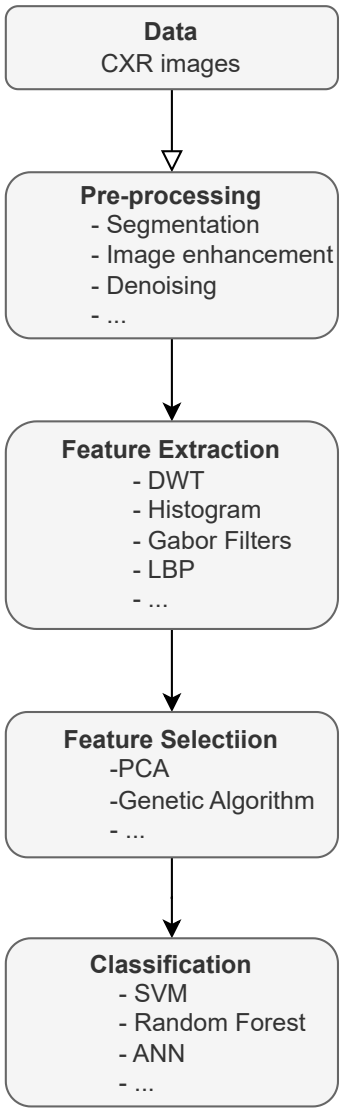


Figure 2.4 Flowchart of a machine learning pipeline for classification. The most commonly used approaches for classifying CXR images are separately mentioned in each block.

In many cases, reducing the dimensionality of features or selecting the most crucial ones is necessary for improving computational efficiency and mitigating the risk of overfitting in machine learning algorithms. One important option for feature selection is principal component analysis (PCA) [36], a statistical method that projects the features into a smaller dimensional space.

The prepared features should be fed into a classifier to detect the desired category within the given data. Among the traditional approaches for CXR classification, SVM has found its place in the classification tasks of CXR images [31–35] since it can handle a relatively high space of feature space and its robustness against outliers, which is a common occurrence in medical images.

In the rest of this section, some CXR analysis approaches in Machine Learning are elaborated to provide further details.

Kumar et al. utilized several pre-processing steps, including resampling, filtering, contrast enhancement, and segmentation, before extracting relevant features for classifying normal and pulmonary edema CXRs [31]. The segmentation map was manually created with the assistance of a computer program. Gabor filters were used to extract local features by capturing visual properties such as spatial localization, orientation selectivity, and spatial frequency of images. Eighteen images were obtained by filtering CXR images using Gabor filter instances. By capturing statistical features from the lung sides of the Gabor output, 216 features were extracted to feed the SVM classifier. The C-SV classifier algorithm was used in SVM training to tune the margin determined by hyper-parameters. This study uses twenty images of each class to train the SVM. Çamlica et al. used LBPs and SVM classifiers to classify various ranges of medical images, including chest radiography [32]. To focus on the ROI, they consider the average of saliency maps detected by the context-aware saliency algorithm. Then, the algorithm by image folding algorithm splits each image into patches and concatenates them into a single feature vector. The folded data is then used to extract LBP features, which are then used to train an SVM. They used image retrieval in medical applications (IRMA) 2009 dataset containing 14,410 x-ray images of different organs. In a similar study, Chan et al. used LBP and SVM to detect pneumothorax [33]. However, lung segmentation is used instead of a saliency map to concentrate on lung areas. The lung areas were distinguished by setting threshold values from the Otsu method. The method includes image multiscale intensity texture analysis and segmentation to overcome challenges caused by varied image quality among equipment. Singh et al. detected pneumonia in CXR images using a combination of DWT and gray level co-occurrence matrix (GLCM) with two classifiers: SVM and k-nearest neighbor (KNN) [34]. GLCM captures texture information of CXR images, such as contrast, energy, correlation, and homogeneity. The extracted feature

vectors were then used to train the SVM/KNN classifier. Their results showed the superior performance of SVM over KNN. Interestingly, Eren et al. employed conventional ML tools such as SVM and random forest(RF) to detect Covid-19, a highly contagious disease that emerged recently [35]. They compared seven combinations of tridiagonal matrix enhanced multivariance product representation (TMEMPR), Discrete Cosine Transform, and DWT, and two classifiers, RF and SVM. They indicated that using TMEMPR and SVM results in the best performance over all the 14 obtained combinations. The TMEMPR creates a structure including a triangular matrix that contains parameters representing image information. Rohmah et al. proposed an approach for capturing texture feature information using GLCM, LBP, and histogram of the oriented pattern (HOG) as feature extractors and PCA to reduce the dimension of the feature vectors [36]. They detected COVID-19 cases using the created feature space and SVM classifier. They showed that CXR images have better performance compared to CT images. Khobragade et al. detected critical pulmonary diseases such as lung cancer, pneumonia, and tuberculosis (TB) [37]. After some preprocessing steps, such as lung segmentation and contrast enhancement, they employed geometrical features such as area, lung field perimeter, equivalent diameter (assuming the lung is a circle, diameter is calculated from its area), and irregularity index, defined as:

$$I = \frac{4 \times \pi \times Area}{(perimeter)^2}. \tag{2.1}$$

They also used statistical features such as mean, standard deviation, and entropy as texture representatives. A feed-forward artificial neural network (ANN) classified the extracted features from the previous step. Reza et al. proposed an approach to extract a feature space comprising the image projection profile, histogram of oriented gradients, and a novel feature set called CXF30 [38]. CXF30 represents the signal properties of the vertical image profile using thirty specific signal characterizations, such as the positional value of the first peak and the positional difference between the first and second peaks. A feature selection method based on a genetic algorithm was utilized to obtain the best subset of features. The selected features were used to train a random forest machine learning model for image view classification.

The conventional approaches are fast and do not require much data for training. One of the most important advantages of these algorithms is their ability to explain the decisions made by the model. Since the models are mostly simple and separate, it is possible to trace the input within the pipeline. Additionally, the selected features are initially designed by experts, making it possible to interpret the feature spaces of the given input.

On the other hand, one of the most crucial steps in all the mentioned works is to leverage

the most optimal features of CXRs. In other words, it is necessary to identify the most distinguishing features to realize the potential of the algorithms fully. This often requires expert knowledge to engineer and recognize specific specifications. Moreover, there is always a trade-off between the size of the feature space and the capacity of the classifier. Increasing the size of the feature vectors (which may be necessary for complex tasks) is controversial since shallow classifiers cannot be well-trained and normally underfit. A small feature size is probably insufficient for a distinguishable representation of the input image.
Another challenge in most manual feature extraction is extracting the RoI prior to feature extraction. For instance, texture features are crucial for CXR classification, so extracting such features from the lung areas is vital, ignoring the background or other organs. Such a strategy requires an additional step to direct the model's attention to the desired object, which can increase the risk of propagating errors throughout the pipeline.

2.2.2 Deep learning approaches

The advance of deep learning has been made possible due to powerful computational resources and the availability of massive datasets. Extensive computations to train large models that require large datasets are indeed possible these days. Consequently, there is a wide availability of large datasets in various fields, including the analysis of CXRs.
In the analysis of CXR images, massive datasets containing hundreds of thousands of images labeled as various pathologies are now being used in the studies. In the CXR-analysis literature, Chest X-Ray14 [2], cheXpert [39], Mimic-CXR [40], and PadChest [41] are the most applied datasets for the classification of thoracic pathologies.
One of the main data sources widely used in studies is the Chest X-Ray14 dataset released in 2017 [2]. The dataset consists of 112,120 frontal-view CXR images taken from 30,805 unique subjects. The images have been labeled as associated with fourteen diseases, noted from clinical reports via natural language processing (NLP) techniques. The dataset is an expansion of Chest X-Ray8, including six extra pathologies. The labels annotated in this dataset are Emphysema, Infiltration, Mass, Pleural Thickening, Pneumonia, Pneumothorax, Atelectasis, Edema, Effusion, Hernia, Cardiomegaly, Pulmonary Fibrosis, Nodule, and Consolidation. The CheXpert (Chest eXpert) dataset comprises 224,316 chest radiographs from 65,240 individuals, annotated for fourteen pulmonary diseases as positive, negative, or uncertain [39]. The uncertain labels were assigned by an NLP labeler who extracted observations from radiology reports. MIMIC-CXR is a publicly available dataset that comprises 377,110 images from 227,835 radiographic studies of over 65,000 patients [40]. The CXRs in the dataset were obtained using different radiographic devices, resulting in variations in quality

and resolution. Moreover, each image is accompanied by a description of detected abnormalities provided by radiologists at the Beth Israel Deaconess Medical Center. The Padchest dataset comprises 116,868 images obtained from 69,882 patients and gathered at San Juan Hospital in Spain between 2009 and 2017 [41]. This dataset contains multiple projection views and supplementary metadata that can be utilized in research endeavors. Notably, 27% of the images have been manually annotated by trained physicians.

In the literature on CXR classification, two training strategies are applied. Firstly, deep models can be designed and trained using the dataset specific to the domain from scratch. These models start with randomly initialized weights and no prior domain knowledge. They learn everything from the task-specific dataset and are updated until they meet the stopping criteria, enriched by knowledge in the field. Secondly, models can be trained in another domain and customized for the specified problem, known as transfer learning. By leveraging prior knowledge from a large pre-training dataset, these models can achieve better generalization, particularly when training data is limited.

Many studies have focused on training newly designed models or other popular state-of-the-art architectures from scratch. Kieu et al. trained three instances of CNN models for detecting abnormalities in CXRs [42]. They investigated the left and right lung sides, as well as both together, each in a separate model. The model synthesized the results to obtain a comprehensive result. In another work, Islam et al. employed an ensemble of several models, including Alexnet [6], Resnet [43], and visual geometry group (VGG) [44]. The combined model outperformed the single classifiers for detecting Cardiomegaly and Tuberculosis [45]. Yao et al. used a two-stage model that took advantage of a DenseNet [27] encoder and a recurrent neural network decoder (RNN) for the diagnosis of fourteen diseases in the ChestX-Ray14 dataset [46]. The RNN exploits the dependencies between labels, which improves the performance of the model. The classification problem in these methods is defined for multi-pathology detection, and different steps focus on various features to aid in detecting different symptoms.

A relatively novel concept in deep learning, known as attention, is introduced to focus on the most relevant and essential parts of the input for the underlying ML task. It is inspired by humans who pay attention to specific details of an object and ignore non-relevant components. In the study by Quan et al., they presented a category-wise residual attention learning (CRAL) framework, in which an attention block is arranged to strengthen feature learning and get correlations among pathologies [47]. Their model consists of a high-level feature extraction module and an attention module, which provides the category-specific view for the model. Their attention model can be separately attached to other deep models. They classify fourteen diseases of the Chest X-Ray 14 dataset and achieve outstanding results. The

authors proposed an alternative approach for disease classification in CXR images using a deep model called attention-guided convolution neural network (AG-CNN) [48]. AG-CNN takes into account both global and local investigations of images. The model consists of two main branches: global and local. In the global branch, a sub-region of the CXR image is highlighted and used as the input for the local branch. The ROI is generated by thresholding the feature maps in the global branch, which drives the attention mechanism used in local investigations. The classification results of the AG-CNN outperformed the model with only the global branch, as evaluated on the ChestX-ray14 dataset. Peng et al. utilized two distinct pathways to consider both local and global information in their approach [49]. Their model comprises a mask branch and an image branch, which are fused. The mask branch extracts the relative location information of diseases occurring in a bottom-up and top-down structure. To segment the lung mask, they employed a U-net model, which led to generating location maps based on Euclidean distances from the lung and heart masks. These maps were combined with the original images to feed a Densenet-121 as the image branch.

Transformers, powerful models in NLP, have also found their role in image-based ML tasks by leveraging their attention mechanism and adaptability for capturing spatial relationships. Duong et al. merged EfficientNet [50] with the Vision Transformer [51], which is based on the self-attention to classify tuberculosis in CXR images and achieved an AUC of 100% [52]. They launched several experiments and found this combination the most appropriate for the task. A vision transformer partitions an image into patches fed into a set of transformer encoder layers. Each encoder layer consists of a feedforward neural network that implements non-linear transformations, as well as a self-attention mechanism that directs the model to focus on various parts of the image. Finally, a classification head predicts the class label using the output of the last transformer encoder layer. Vision transformers are used in many medical tasks, including detecting Covid-19 in CXR images [53]. Due to data scarcity on COVID-19 CXRs, Park et al. first used a public dataset to extract the low-level resolution CXR images. Then, the transformer used the feature corpus to detect COVID-19 abnormalities in chest radiographs.

During the COVID-19 pandemic, the demand for efficient medical diagnosis was substantial. However, privacy regulations imposed restrictions on the sharing of patient data for training purposes. To address this challenge, a collaborative federated learning approach was employed, enabling multiple medical institutions to collectively develop accurate COVID-19 screening models without the necessity of sharing patient data [54, 55]. Notably, this approach yielded competitive results compared to traditional centralized data-sharing models. Furthermore, it is worth highlighting that partitioning the identical dataset among separate federated learning clients, as observed by Slazyk et al., has the potential to enhance the

classification accuracy of chest X-ray (CXR) images for the models under investigation [56]. A strategy for learning, especially when the data is scarce, is transfer learning. Transfer learning is a technique in which a model trained on one task or domain is used as a starting point for training a model on a different task or domain. It involves pretraining a model on a large dataset, learning general patterns, and then fine-tuning it on a smaller, task-specific dataset. Several studies have shown that retraining models on CXR images can result in high performance for classifying thoracic diseases. For example, Crosby et al. combined two VGG-19 [44] models pre-trained on the Imagenet dataset to identify pneumothorax in CXRs [57]. Since the Pnemothorax needs a higher resolution than usual image sizes in DL (224×224 or 256×256), they extract patches of size 256×256 from the top third of each lung side. They fine-tuned the last two convolution layers and the fully connected classifier using patches and apex images extracted from coronal CXRs. Rasheed et al. exploited horizontal and vertical components of the wavelet transform to enhance images and facilitate feature extraction [58]. Then, a pretrained Resnet-50 [43] classifies the derived maps while almost half of the convolution layers and classifiers are retrained. Baltruschat et al. make a comprehensive comparison of approaches using Resnet [59]. They tried fine-tuning (retraining only the dense classifier), off-the-shelf (retraining all layers), and training from scratch. No unique approach is detected as the best strategy for classifying all fourteen considered diseases. Furthermore, the available meta-data (age, sex, and angle position) in ChestX-ray14 are also concatenated in an approach and aid the classification task. Rajpurkar et al. proposed the ChexNet [3] in which the pre-trained Densenet-121 [27] is retrained on the ChestX-Ray14 as one of the biggest publicly available annotated datasets. The initial network weights were obtained from the Densenet-121 trained on the common Imagenet. The CheXnet achieves state-of-the-art results with an AUC of 0.73 in classifying Infiltration.

Moreover, it is important to note that while classification models may perform well, using explainable artificial intelligence (AI) methods like Grad-CAM reveals that many do not derive conclusions from the accurate regions of the input CXR [60, 61]. An effective way to address this issue is to limit the computation area by employing segmentation methods to extract the RoI, which in our case is the lung. Teixeira et al. and Freire et al. have demonstrated that incorporating segmentation prior to classification improves performance and enhances generalization, especially when data is limited [61–63]. Hence, it is imperative to assess the available segmentation methods to incorporate them into our work effectively.

The table at 2.2 categorizes studies related to the classification of CXRs and includes specific instances from these studies.

Table 2.2 A summary of model structures for classification of CXR images

Model Structures	**Key Features**	**References**
CNN models	• Training from scratch and Transfer learning • Various models available	• Separate analyses of lung sides - Training from scratch [42] • DenseNet encoder and RNN decoder - Finding dependencies between labels [27]
Attention-based	• Local investigations of CXR images • strengthens feature learning	• Category-wise residual attention learning (CRAL) framework - Correlations among pathologies [47] • Global and local investigations - Improved detection of abnormalities [48]
Vision Transformer	• Self-attention mechanism • Handle various resolution	• Merges EfficientNet with Vision Transformer - AUC of 100% for tuberculosis [50]
Federated Learning	• Collaborating multiple institutes • Enable privacy-preserving	• Collaborative model development - Competitive results compared to centralized models [54]
Segmentation-Classification	• Limit computation area • Improve generalization	• Incorporates segmentation before classification - Enhances generalization and performance [62]

2.3 Lung Segmentation in Chest X-Ray images

Segmentation methods, like other image analysis techniques, have undergone significant transformations. Conventional methods heavily rely on domain-specific knowledge and human feature engineering, while deep learning models have emerged as highly accurate and powerful alternatives. Deep learning-based segmentation models have demonstrated superior performance compared to their previous-generation counterparts in various fields, including the segmentation of lung fields.

2.3.1 Conventional approaches

When deep learning is not taken into consideration, three primary approaches can be utilized to segment lung fields in CXR images: rule-based, model-based, and ML-based techniques. Rule-based techniques employ domain-specific expertise and prior knowledge to extract lung fields from images. These techniques usually involve a trial-and-error process that accounts for image characteristics such as opacity, texture, shape, and structure. Armato et al. analyzed gray-level histograms to extract the contours of the lung areas, and an iterative global and local thresholding approach is used [64]. The generated lung boundaries are subsequently smoothed using techniques such as rolling-ball techniques. Radiologists graded the contours' accuracy and completeness on a scale of one to five. The five-star evaluation of radiologists on 600 CXR images proves the effectiveness of this technique. Setting thresholds for the general purpose of segmentation is very common in image processing applications. Kasu et al. used Otsu to find the optimal threshold for highlighting lung regions in CXR images [65]. Li et al. simplified the previously proposed techniques for identifying edges based on derivatives using only the first derivative of the horizontal/vertical image profiles [66]. They determined the region of interest and lung boundaries using pattern classification and image feature analysis. Also, to create a smooth lung border, they performed an iterative contour-smoothing technique. In traditional approaches, it was crucial to use bone suppression techniques due to the strong edges of the rib cage and clavicle bones that cause the creation of local minima in hand-crafted feature extraction. Fredric et al. describe a semi-automatic approach for extracting scoliotic rib boundaries from CXR images [25]. Edge-following procedures are begun from user starting points along upper and lower rib edges, and the most parallel pair among detected edges is found to get the final rib boundary. Chattopadhyay et al. proposed a marker-based watershedding (MBW) segmentation to identify lung areas in CXR of COVID-19 patients [67]. The approach utilized a simple median filter (SMF) to denoise the raw images before applying MBW. Their results were validated by radiologists tested on 10 CXRs.

Although rule-based techniques can be successful, they require considerable human intervention and may not generalize well to new datasets. They are mostly task-specific and need human effort to tune the algorithm.

The second category in conventional methods is model-based techniques. Model-based approaches extract lung fields from CXR images and provide information on organ deformities using deformable models of the organs. These techniques can handle lung form and size variations and may be more accurate than rule-based techniques. Active appearance model (AAM) and active shape model (ASM) are model-based methods using the deformable models of organs of the CXR images and give suitable information about the organs' deformities.

The fundamental principle of ASM is to construct a statistical shape model from training images. This model represents the general variations in the shape and appearance of segmenting objects in the images. It can segment images by fitting the statistical model to the image and adjusting the model's shape accordingly. Similarly, AAM is a statistical method that extracts a model to fit the segmentation image. In AAM, a set of texture or color templates represents an object's appearance, while a set of reference points describes the object's shape. Wu et al. used ASM to segment lung fields in CXRs [68]. The approach uses an average active shape model, gray-scale projection, and affine registration to create the initial lung contours. The vertices of the active shape model have then pulled away from the stomach gas areas. A new objective function that contains edge and distance restrictions is developed to ensure a more evenly distributed distance distribution of vertices.

Wu et al. employ a multi-resolution representation and optimization technique to solve the objective function efficiently. Ginneken et al. proposed an advanced version of the ASM method based on optimal local features [69]. A nonlinear neural network classifier is used to find ideal landmark displacements and to establish a specific set of optimal features for each landmark at each resolution level. The computational complexity of the improved ASM method is approximately 20 times higher than the original one. Using AAM, Shao et al. employed shape and appearance models to separate the lung field area and build a reliable prior shape for the lung [70]. They constructed local shape and appearance composition models to address the high variance and lung ambiguity. Lastly, the hierarchical deformable segmentation produced reliable lung masks.

Candemir et al. presented a method for automatically segmenting the lungs in CXRs utilizing anatomical atlases and nonrigid registration techniques [71]. The authors proposed an approach that combines atlases to provide a priori information on the shape and location of the lungs, with deformable registration, which provides flexible deformation to better match the patient's particular anatomy.

Li et al. divided the image into small sub-regions using graph-based segmentation and then identified the salient value of each sub-region using a global contrast function. The lung area is calculated using this information, and cubic spline interpolation refines the lung boundaries [72].

Generally, model-based techniques require prior knowledge of the underlying anatomy, and they may not perform well if patients exhibit significant anatomical variation.

ML-based approaches for CXR image segmentation are gaining popularity rapidly. These approaches leverage traditional and modern deep neural networks to directly learn representations from data, eliminating the need for prior knowledge or explicit models. In the following section, we will review conventional ML approaches in this domain. Several studies

have been established using ML techniques to interpret the CXR images from segmentation tasks to the diagnosis of the diseases. The robustness of these methods highly depends on the hand-crafted features extracted from the data. McNitt-Gray used stepwise discriminant analysis to select features for CXR segmentation issues, which resulted in a smaller feature set [73]. Compared to the complete feature set, they observed similar accuracy in the performance of classifiers such as linear discriminators and neural networks while using this reduced feature set. This method resulted in increased enhanced computing cost efficiency. By unsupervised learning, clustering is a common method for highlighting regions of interest in CXR images. Ghosh et al. compare two common clustering algorithms, C-means and Fuzzy C-means (FCM), to solve the segmentation problem of chest radiology images [74]. The study was based on the quantity of data points and clusters, and the findings demonstrate that FCM gives equal results to C-Means clustering but takes longer to compute. Shi et al. proposed a novel approach for segmenting lung regions in chest X-rays [75]. It used Gaussian kernel-based fuzzy clustering with spatial restrictions, which enhances the traditional fuzzy c-means algorithm. This includes combining a distance metric produced by a Gaussian kernel with a penalty term considering the nearby pixels. Vittitoea et al. proposed a method for identifying lung areas in DCRs using Markov random field (MRF) modeling for computer-aided diagnosis (CAD) systems [76]. Textural and spatial properties were used to define lung and non-lung areas, and probability distributions were used to incorporate them into the MRF model.

These ML approaches have demonstrated considerable success compared to conventional tasks and can be easily generalized to new datasets. However, like other tasks, deep learning exhibits a new level of performance that surpasses previous machine learning methods.

2.3.2 Deep learning approaches

With deep learning, lung segmentation of CXR images has shown impressive progress. Convolutional neural networks (CNNs), a type of deep learning model, have been widely used as they can learn hierarchical representations of input images. U-Net, a popular CNN architecture proposed in 2015, has been adopted as a baseline model for lung segmentation [77]. The model as it is [60, 78] and its variations [79–81] are used in lung field segmentation in many studies. The network comprises two pathways: one for contraction and, symmetric to this, one for expansion, which are linked together by multiple connections. The contraction path contains convolutional and max pooling layers, which lower the image's spatial resolution and increase the depth of channels. On the other hand, the expansion path includes upsampling and convolutional layers that gradually increase the spatial resolution and decrease the number of channels to reach a map the same size as the image. The connections between

the paths allow the network to transform abstract information from the earlier layers and integrate it with high-level features in the later layers. The network structure is illustrated in Figure 2.5.

SegNet is an encoder-decoder network that contains five encoders with convolutional layers, batch normalization, and the rectified linear unit (ReLU) operation activation function [82]. The decoder network incrementally raises the resolution of feature maps to the input size dimensions using pooling indices instead of direct connections between the corresponding layers. LF-SegNet is proposed for the segmentation of lungs in CXR images. Its design is LF-SegNet, similar to SegNet's, but it incorporates a skip architecture, which assists in feature reuse and alleviates vanishing gradient [79]. LF-decoder SegNet's network contains five decoders and employs a basic upsampler variation that replicates feature maps to upsample them. The skip architecture and simple upsampler variation obtain better results and faster convergence. Mizuho et al. improved the robustness of the U-Net architecture for lung segmentation in CXRs of pulmonary illnesses [80]. They increased lung field segmentation performance by using the Bayesian optimizer. This optimizer was used to identify the optimal configuration of hyperparameters for the task. Several hyper-parameters from the original U-Net were changed, including batch size, learning rate, number of max-pooling layers, feature maps, drop-out probability, and the arrangement of batch normalization. Novikov et al. explored the challenge of balancing the size of neural networks and the available training data to avoid overfitting [81]. They describe an experiment in which they deactivated convolution kernels in a modified U-Net trained on CXR images. The authors discovered no significant difference in output, showing the potential for customization and tweaking to the domain's demands. They proposed three structures called All-drop-out, All-convolutional, and Inverted-Net. A drop-out layer is added after each convolutional layer in the All-drop-out network. In the All-convolutional network, strided convolutional layers are applied instead of pooling layers. Inverted-Net, derived from these architectures, is designed to be compatible with the nature of CXR data and has the best segmentation performance in segmenting multiple organs. Kholiavchenko used three deep architectures for organ segmentation in CXR images [83]. The utilized networks are UNet [77] augmented with the ResNeXt encoder, LinkNet [84] augmented with the ResNeXt encoder, and the Tiramisu [85] architecture augmented with the DenseNet [27]. LinkNet is a model identical to the original U-Net in that the transferred feature maps are summed rather than concatenated, reducing the number of training parameters. The Tiramisu architecture also employed dense blocks in the U-Net backbone network. Lastly, morphological procedures are used to smooth down the edges of the output masks of the deep networks.

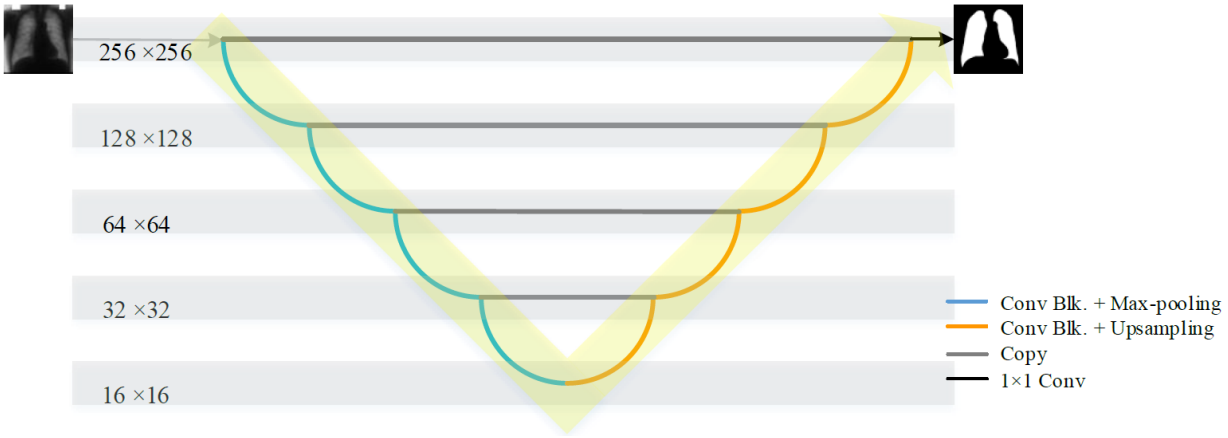


Figure 2.5 Original U-net model architecture.

R-CNN family of models has made major contributions to the field of object detection. Girshick et al. introduced the first model, R-CNN, in 2014 [86]. It generates region proposals by employing a selective search algorithm to identify potential object positions in the image. These suggestions are then fed into CNN, which extracts features and classifies the item in each one. A combination of classification and regression loss functions is used to train the model. Despite the model's strong performance, the computational costs of the mode are considerable, limiting the model's utility. One year later, the authors proposed Fast R-CNN to address the shortcomings of RCNN. Fast R-CNN distributes computation throughout the entire image instead of performing individual computations for each region proposal, thereby improving upon R-CNN [87]. Shaoqing et al. replaced the selective search algorithm with a Region Proposal Network (RPN), which creates region proposals directly from CNN features. This resulted in a more accurate and faster model, known as Faster R-CNN [88]. He et al. developed Mask R-CNN in 2017 by adding a parallel branch to predict a mask that generates a binary mask for each object instance, in addition to its bounding box, enabling classification at the pixel level [89] (Lung VS Non-Lung). Peng et al. suggested A-LugSeg framework for automatically segmenting the lungs in CXRs of multi-site CXR datasets [90]. Two subnetworks are included in the framework: Mask-RCNN for coarse segmentation and a hybrid refinement phase that includes enhanced ML techniques. The framework offers a coarse-to-fine cascade segmentation approach that precisely segments the lung contour while reducing the influence of surrounding organs. In another work with the same strategy, they fine-tune the coarse maps obtained by Mask R-CNN by refining the obtained boundaries using optimization algorithms [91]. Wang et al. utilized Mask R-CNN for the segmentation of lung, heart, and clavicles in CXR images [92]. The results demonstrate that the model is efficient and helps to reduce the false positive predicted regions.

Like classification, the concept of attention is also utilized in lung segmentation. Ullah proposed an image segmentation framework constructed with two encoders and two decoders [93]. The pre-trained VGG19 encoder is used for feature extraction, and the squeeze and excitation (SE) block is employed to provide the most important information to the first decoder, which generates an initial binary segmentation mask. The second encoder improves the original mask using the attention gate mechanism (AGM) to strengthen the feature map quality. AGMs focus on areas of interest while retaining spatial resolution. The second decoder generates the final enhanced segmentation output by using the output of both encoders. He et al. proposed a lung segmentor model, XLor, designed to capture long-range contextual information for more accurate lung segmentation in CXR images [94]. The input feature maps are divided into four directions: top-to-bottom, bottom-to-top, left-to-right, and right-to-left. Then, the module calculates attention maps for each direction, which weighs each pixel according to its similarity to pixels in the same direction. The final attention map is created by combining the attention maps in all directions. XLor uses the ResNet model [43] to extract features from input CXRs at the beginning of the process. Hwang et al. used Atrous convolutional layers to broaden the kernels' field of view [95]. Moreover, they used network-wise training as a multi-stage technique, which improved lung segmentation performance. In this technique, a pre-stage model is repeatedly trained and improved using both input images and the output of a prior stage in a coarse-to-fine strategy. This recurrent cascading technique helps the network to focus on details and learn precise lung field boundaries. Gaál presented a model that generates pixel-wise segmentation of the heart, left lung, and right lung [96]. The proposed network architecture included an Attention U-Net with Attention Gates (AGs) to focus on critical local features and minimize false positive rates. By using AGs, the Attention U-Net was trained to concentrate on the most important features. The segmentation procedure was designed to increase the precision and effectiveness of medical image analysis.

Gaggion introduced HybridGNet, a neural network architecture developed for anatomical segmentation in medical images, specifically CXR images [97]. HybridGNet represents anatomical structures as graphs, allowing it to consider anatomical restrictions while addressing topological defects and inconsistencies. HybridGNet is built on an encoder-decoder architecture, with the encoder using regular convolutions and the decoder decoding the features with graph convolutional neural networks (GCNNs). The image-to-graph skip connection layer lets convolutional block features flow into GCNN blocks, boosting segmentation accuracy. Liu et al. utilized a lung region generator before classification to eliminate irrelevant information in CXRs [60]. The generator comprises a lung segmentation and a post-processing step. A U-Net segmentation model highlights the lung areas in CXR images. The post-

processing step deals with issues such as diseases and poor image quality that can result in false-positive and false-negative regions. By utilizing lung anatomies, such as symmetry and size, false-positive errors are identified and corrected. The differences between the left and right lung sides reveal the false-negative regions misrecognized by U-Net.

In some specific applications, it is necessary to analyze the original-sized image, which can be problematic with regular deep models. Huynh presented a network and a training approach for processing large-sized CXR images [98]. The architecture consists of a CNN and a fully connected network. In a patch-based approach, the CNN extracts feature information from small input image patches and creates feature maps. These feature maps are then fed into the fully connected network, classifying each area as lung or non-lung. Finally, the created map is post-processed using morphological operations to achieve more accurate and smoother boundaries.

Mansoor et al. combined the model-based technique with deep learning. They used two deep learning architectures that learn space and shape parameters to partition deformable objects, specifically the lung field from CXR [99]. Unlike statistical shape models' traditional iterative convergence process, this approach separates the parameter space into linearly separable subspaces, and deep learning classifiers independently learn the shape parameters. This method is computationally viable and substantially more accurate than current statistical shape model methods. Chen et al. employed three deep learning models, namely FCN [98], U-Net [77], and SegNet [82], to accurately segment lung fields from CXR images [100]. The models were trained on pre-processed images that underwent histogram enhancement, threshold selection, and binarization, which significantly improved the segmentation performance, as demonstrated by the experimental results presented in the paper. Souza et al. proposed a two-phase approach for segmenting lung fields in CXR images [101]. The lung fields are initially segmented using AlexNet in a patch-based methodology. In an independent step, the lung fields are reconstructed in the second stage using a CNN model built on the ResNet architecture [43]. The final lung map is then produced by combining these two maps. The resulting map is then post-processed to highlight the locations of interest.

Generative adversarial networks (GANs) are also used as recent trending models for CXR organ segmentation. Figure 2.6 presents a schematic of how GAN models work. In [102], a novel dual-path adversarial learning (DAL) approach was proposed for learning features with various degrees of complexity using the GAN concept. The approach is evaluated on various medical images and illustrates high performances. Munawar et al. proposed a GAN-based approach for lung segmentation in CXRs [103]. A U-net architecture is used to build the lung map, and four discriminators are examined. Notably, even training on a tiny dataset, their model maintained a high level of accuracy. The authors concluded that their GAN-

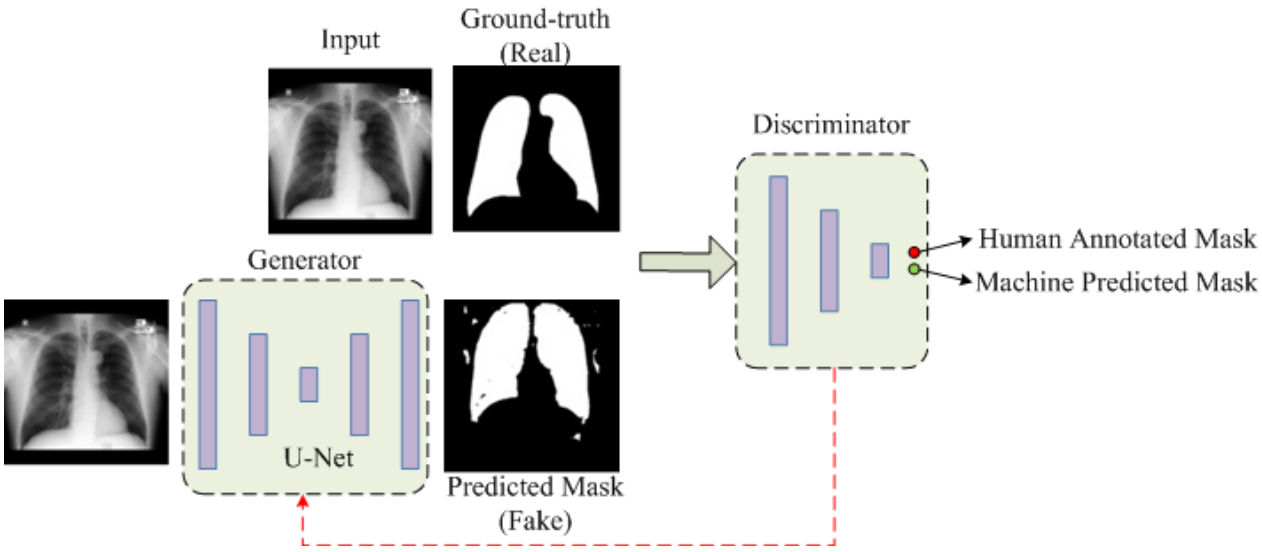


Figure 2.6 A GAN model to segment lung in CXR image

based method is effective for lung segmentation in CXR images. Eslami et al. proposed a deep learning-based approach for suppressing bone structures and segmenting organs in chest X-ray images using an image-to-image translation model. The method entails training a conditional generative adversarial network (cGAN) to convert an input x-ray image into a segmented output image that highlights the organs while suppressing bone structures [104]. The proposed method can potentially increase the precision and effectiveness of organ segmentation in CXR images and exhibit encouraging results.

Many studies have also focused on segmentation network modules. Liang et al. introduced an unpooling module, a small inception module that integrates distinct up-sampled feature maps [105]. Unlike upsampling, the unpooling module merges numerous up-sampled features, preventing error growth and propagation. On the other hand, Segnet [106] employs a max-unpooling technique, but due to the large number of zeros in the feature maps, it incurs more costs and takes longer to converge. LF-Segnet [79] replicates the elements during the upsampling process.

Cross entropy is the most commonly used loss function in medical image segmentation applications, including lung segmentation. Other measures, such as the dice metric and the Soft-max cross-entropy [106], are also employed. The most popular activation function is ReLU [107], but recent studies have found that the exponential linear unit (ELU) activation function [108] achieves a higher overlap score [43] and speeds up networks.

2.4 Web-based platform

The availability of various hardware and internet connections has revolutionized obtaining medical information. There is access to an enormous number of medical apps over the internet that can be used to check people's health, monitor their symptoms, and even detect diseases. These medical apps' accessibility has made it simpler for both patients and doctors to take control of their health and deliver individualized, effective care to their patients. Accessing medical apps from any place with an internet connection is the main advantage of such applications. Patients can use these apps to monitor their health while on the go, and clinicians can access patient information from anywhere. Doctors may now offer remote care because of this accessibility, which is crucial in regions with a shortage of medical staff.
The web applications available to the public aim to suit a wide range of users and provide health-related information. The health tracking app is one of the most often used categories of medical apps. Patients can use these apps to track their health information over time, including their weight, blood pressure, and heart rate. The applications can be installed on the smartphone and can notify predefined contacts in case some artifacts occur. Samsung Health [109] and Apple smart fall prediction [110] are pre-installed apps that are the most common apps for healthcare tracking. Some apps also analyze the patient's voice to detect potential illnesses [111]. By integrating smartwatches into mobile devices, applications can leverage the sensor to record blood pressure, heart rate, respiration rate, etc. On the other hand, some medical tools are specifically designed and utilized for particular diseases. We have come across two online web applications available to the public for analyzing CXR images. The first one called Chester [112], was designed by Cohen et al. They employed the state-of-the-art Densenet-121 model, pre-trained on ImageNet, and fine-tuned on the Chest X-Ray 14 dataset. The application is accessible through the web [3]. They used Grad-Cam technology to highlight the most involved areas in the image. However, it is recommended not to use this tool in healthcare centers. The Mobile Chest X-ray Analysis also used the same training set and model to detect 14 pathologies in chest radiographs [113]. The program, designed for smartphones, uses saliency maps acquired with GradCam to support the diseases that have been diagnosed. Therefore, although these systems offer activation maps highlighting the references for the decisions, they are not trustworthy enough to be used in clinics, as discussed previously.

[3] https://mlmed.org/tools/xray/

CHAPTER 3 RESEARCH OBJECTIVES

Although few studies on identifying ARDS CXR images have been conducted, a reliable and interpretable method for detecting ARDS that can be deployed in a hospital setting is still unavailable. The problem of developing a generalizable and understandable ARDS detection algorithm remains unsolved despite the massive studies in CXR analysis. In order to address the critical clinical need, we identify the limitations of the previous studies. To meet these constraints and completely realize our goals, we describe our objectives in this section.

3.1 Problem statements

We discussed the criticality of ARDS and emphasized the value of an early diagnosis in the previous sections. Moving forward, it is essential to consider the specifications needed for a medical device to be used in hospital wards and effectively deployed in clinical infrastructure settings.

Our primary focus is to detect ARDS in CXR images, as this imaging modality is easily accessible and cost-effective compared to other imaging techniques. While other diagnostic criteria are necessary for identifying ARDS, chest radiographs are crucial in detecting the syndrome. In addition to confirming the diagnosis of ARDS, chest radiographs can also play a role in determining complications associated with the condition. So, the main specification of the platform would be the capability of the tool to analyze chest X-rays in terms of ARDS manifestations.

The approach must accurately interpret CXR images for doctors, a crucial requirement for a tool to be accepted and deployed in clinics. The ability to interpret images equips the tool to assign a severity score, an essential aspect of clinical decision-making. Providing solely a binary label is inadequate for serving as an effective assistant to clinicians.

A model must function automatically without the need for doctors to intervene. Autonomy is essential because the model's performance must be independent and not reliant on human actions. Secondly, in case of deployment, the approach should operate independently and provide an appropriate alert message even before clinicians become aware of it. Anything that is not fully automatic may present challenges in implementation and real-time actions in critical cases. As discussed in 2.1.2, currently, there are no approaches in the literature that effectively address both challenges of automaticity and interoperability.

Additionally, data-driven approaches, particularly those based on deep learning, often require a large amount of data. For most medical applications, which often have problems with data

scarcity, this presents an extensive challenge. In the case of ARDS, there are no publicly available datasets that have been annotated with ARDS labels, which makes it necessary to create an internal dataset. However, the process of annotating data is costly and time-consuming. Therefore, the approach ought to perform well on the limited internal dataset that can be collected for this project.
The method should also have good generalizability for image analysis from different healthcare institutions. This is crucial because the imaging equipment and techniques utilized can impact the quality and features of medical images. As a result, the method should be robust enough to manage these variances and work effectively with images from various sources.
Additionally, it is essential to create a flexible platform that can work both independently and in cooperation with other medical devices to detect instances of ARDS. The platform should be able to be integrated into other AI systems that examine hypoxemia and cardiac failure, which, as was mentioned in earlier parts, are additional crucial factors for diagnosing ARDS. Additionally, the proposed platform should be available for external validation and go through further investigation by healthcare authorities in order to guarantee its reliability and performance. As a result, at this point, deploying the application as a web-based platform would be preferred since it is easy to access and utilize.
To sum up, our goal is to address the problem of ARDS identification in chest radiographs by proposing an approach with the following requirements:

- accurate and robust for CXR classification,
- interprets the syndrome complexity,
- operate independently and autonomously,
- robust in facing data scarcity,
- offers specifications to be used in prospective research and external validation to guarantee validity and reliability,
- deployed in a user-friendly web application.

By meeting these requirements, we aim to offer clinicians a helpful tool for the diagnosis and treatment of ARDS, thereby improving patient outcomes.

3.2 Reasearch Objectives

The primary aim of this thesis is to develop an algorithm for detecting ARDS that can be utilized in hospital wards. In order to meet the clinical requirements, the algorithm needs

to offer clinical insights regarding the conclusions it provides. To achieve this objective, we intend to leverage the benefits of segmentation in our work, ensuring that the model focuses on accurate regions. Since diagnosing ARDS involves identifying infiltration diffusion throughout the lung sides, the model should be capable of providing localized labels. This will result in the interpretability of CXR images, as the reference for its output decision will be evident. The algorithm should be designed as a web-based tool that can be easily validated and seamlessly integrated into existing clinical infrastructure.

3.2.1 Objective 1: Lung Field Segmentation in CXR images

As discussed in the previous section, it is common and helpful to highlight lung areas in CXR images for detecting pulmonary diseases [31, 33, 37, 49]. The rationale behind segmentation is to focus on regions of interest and avoid processing non-relevant areas that can mislead classification methods.

Among the various strategies used for lung segmentation in the literature, deep learning outperforms rule-based, model-based, and traditional ML-based algorithms. U-Net [77]-based approaches have drawn attention due to their stability and high performance. These models are used in a variety of applications, and several structures can be combined to customize them for specific tasks [79, 81], as explained in section 2.3.2. The main limitation of these models and similar ones is that they usually contain a large number of training parameters. This factor increases the risk of overfitting, especially when there are few training data available. Therefore, in terms of generalizability, it is crucial to have a model that is as lightweight as possible. Notably, since we need to deploy the model in clinics, heavy models can pose challenges in deployment.

It is discussed in 3.3 that the U-Net remains a promising model for tuning and adapting to CXR lung segmentation.

Therefore, our first objective is to:

Develop a light deep model for segmentation of lung regions in CXR images

3.2.2 Objective 2: ARDS Classification of CXR images

Extensive studies have been conducted in the community to analyze chest X-ray (CXR) images, leveraging large datasets. Deep learning models have proven to be highly effective at classifying pathologies in CXRs. However, since these datasets often lack ARDS labels, classification tasks for ARDS are limited. Moreover, deep models require significant amounts of data to be trained effectively, a major challenge in ARDS analysis of CXR images.

Although a few studies have focused on ARDS classification, they have primarily used transfer learning or trained on internal datasets annotated with ARDS labels. These studies have limitations in interpreting the results, a common issue with deep learning in various applications. The explainability tools like Grad-Cam have shown poor performance in CXR analysis, making it difficult to provide references for the model's decisions. One published study has provided insights into the ML pipeline results by leveraging traditional ML approaches. However, this model is not fully automatic and requires clinician actions to make a decision, which is against the intention of alerting doctors to suspicious cases.
We intend to utilize the previous objective, lung segmentation, in the current step of ARDS classification of CXRs. It has been noted in the literature that numerous studies have attempted to isolate the lung areas prior to classification.
The second objective is to:

Develop an interpretable system for automatically detecting ARDS in chest X-ray images.

3.2.3 Objective 3: Web-based platform

The current apps available for CXR analysis, such as Chester and Mobile Chest X-ray Analysis, cannot diagnose ARDS, nor are they designed for validation purposes. Furthermore, these apps are not easily integrable with other tools, which is essential for our goals. While these applications offer useful insights into user interface and design, they are insufficient for our objectives since they only detect pre-defined pathologies. (limited to 14 diseases labeled in the Chest X-Ray 14 dataset).
We aim to develop effective software that can be used at CHUSJ and customized for use in clinical settings. To do this, we must develop an approach that can work in conjunction with the pediatric intensive care unit (PICU) sector of the hospital's infrastructure and other relevant technologies. The approach must pass external validation before being put into practice. We recommend deploying the algorithm on a web server, allowing easy visualization of the model's output. Additionally, to add context, it should be possible to interpret images in terms of ARDS sign manifestation. The model will provide each image's severity score, and the web app's user interface (UI) should make the value readily visible. Furthermore, since different ARDS definitions are used for adults and children, the application must be capable of providing both modes to users.
Therefore, our third objective is to:

Develop a web-based platform for the severity analysis of ARDS in CXR images. This platform will be adaptable to the hospital's clinical settings, integrable into higher-level

medical tools, and capable of providing severity scores for adult and pediatric ARDS. The pediatric definition of ARDS is discussed in more detail in section 6.

3.3 General methodology

The ARDS diagnosis procedure can be divided into two steps: the first includes identifying the RoI, and the second comprises leveraging the extracted regions by looking for illness signs. The lungs host ARDS manifestation; hence, segmenting the CXR images is the starting point. The extracted masks can then detect ARDS infiltrates in CXR images.
The procedure discussed in this section is a commonly used approach that is also compatible with the way radiologists diagnose pathologies. It is important to exclude areas irrelevant to the specific abnormality being diagnosed. As most diseases manifested in CXR images are pulmonary illnesses, lung segmentation is a popular technique in the literature. U-Net [77], introduced in 2015, is a popular model used for biomedical image segmentation. This model has inspired several segmentation approaches, and various modified versions of the model are customized for different segmentation tasks. For lung segmentation, the efficiency of U-Net is investigated in this study.
In the context of lung segmentation, the U-Net's high number of feature maps in its high-level feature extraction layers adds significant complexity. However, our experiments demonstrate that many feature maps contain redundant information (see Figure 3.1). In several experiments, randomly occluding 20% to 35% of feature maps at various levels resulted in limited changes to the output mask in most executions. This observation suggests that the depth of feature maps can be reduced significantly for CXR lung field segmentation.

If we find a way to reduce the depth of feature maps, it can allow having a smaller model

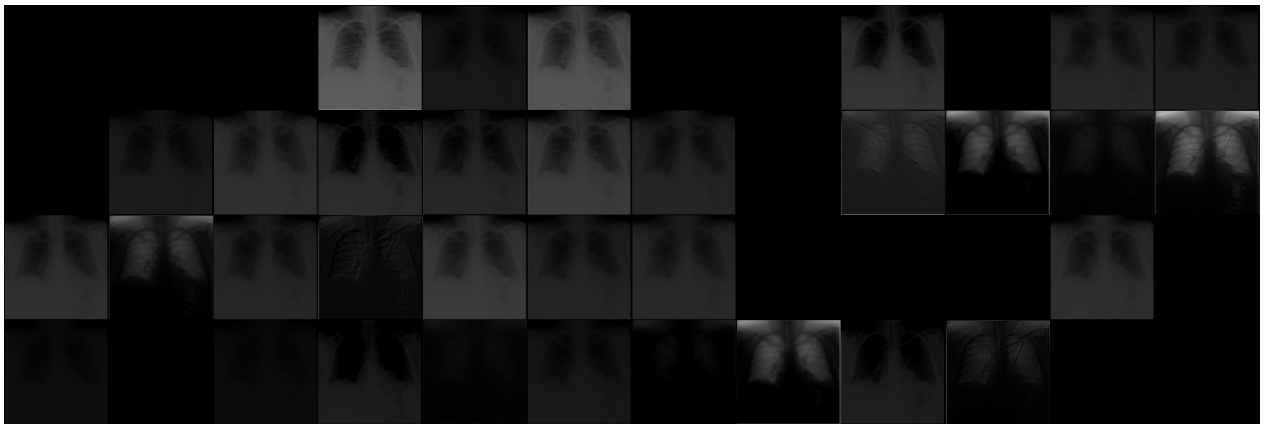


Figure 3.1 60 consequence feature maps extracted from an original U-net applied on CXR images for lung segmentation.

which can be beneficial from various viewpoints. Smaller models are less memory- and computationally-intensive, which makes them appropriate for real-time applications. This is an important factor since smaller models can be trained on lower-end hardware and require less time for execution, making them more practical for deploying algorithms in clinical settings. Although larger models may perform better in general, analysis of feature maps and numerical results indicate that segmentation models for lung fields do not require such complexity. A model overfits when it becomes too complex, which can result in poor generalization to unseen data. Simpler models are better at generalizing to new data because they have a lower capacity and are less likely to overfit. When a model becomes too complex, it may start to memorize the training data rather than learn the structure of the data. This is particularly important for lung segmentation tasks, as the model needs to perform effectively on new images from various machine settings and imaging conditions, which can result in significant variations.

Furthermore, another widespread challenge in training deep models is gradient vanishing. Gradient vanishing occurs when the gradients calculated during backpropagation become very small as they are propagated backward through the network, making it too small to update the weights of the earlier layers. This can also be problematic for U-Net models due to their deep encoder and decoder layers. However, skip connections can help mitigate this problem by allowing the gradients to flow directly from the encoder to the decoder layers, bypassing the intermediate layers and preserving the gradient information. Since we intend to reduce the number of feature maps in each layer, the gradient vanishing problem has intensified. The proposed model must consider information flow throughout the layers to address the gradient vanishing problem and enable the capturing of complex features in images.

DenseNet is a neural network design that connects each layer to the next in a feedforward path, resulting in a dense block. This architecture allows for the direct flow of information throughout the layers, which improves network information flow. Furthermore, layers have access to features extracted in earlier layers. This enables accessing multi-level features in the further layers, which provides a higher level of processing.

The way DenseNet tackles this challenge motivates us to use the same concept to increase information flow, allowing us to reduce the depth of feature maps in layers. The dense connectivity between the layers leads the gradient to flow directly through the model from the output layer to the earlier ones. By merging the design with the U-Net model, we customize the model to be appropriate for lung field segmentation. Chapter 4 provides a detailed exploration of the proposed model, Dense-Unet, including its design details and the results obtained.

For the next step, we aim to leverage lung segmentation before the classification algorithm,

as ARDS, a pulmonary abnormality, manifests in lung areas. Previous studies and our initial experiments have shown that incorporating lung segmentation improves the classification model's performance. In CXR images, lung infiltrates can be observed in areas affected by ARDS. The distribution of ARDS across lung regions varies depending on whether the patient is a child or an adult. Our initial plan for using segmentation involves creating a masked image instead of the original one. A segmentation model is used to acquire a lung mask, which is then used to occlude non-lung areas in the image. The masked image is then fed to a deep-learning model for classification.

However, lung field segmentation is not always the optimal solution since it can introduce errors into the system. One major issue is that lung boundary segmentation is typically performed on normal images, and the presence of certain diseases can cause dramatic changes in lung boundaries. ARDS, for instance, is a disease that can result in faded lung edges due to infiltrates in the lungs, which can make it challenging for segmentation models to isolate the lungs accurately. Moreover, there are no public datasets for lung segmentation that contain pathological images, so previously trained models on normal CXR images may exhibit significant errors in isolating lung fields in CXR images. Using segmentation and classification models sequentially can lead to error propagation throughout the model, especially if the segmentation mask is noisy or includes artifacts. In other words, the classification model assumes that the input images contain only pure lungs. If the lung is not accurately detected, the image may mislead the classification model by processing based on the flawed ROI. To address this, we aim to mitigate the effect of artifacts in segmentation while leveraging its advantages.

By integrating segmentation and classification models into a single unit model instead of using them serially, the effect of the error propagation phenomenon is reduced. In the strategy for designing the Y-Net, a head is taken from the bottleneck of a U-Net model. This enables the extracted features through the U-Net encoder to be used for classification. This strategy seems helpful for addressing the challenges of ARDS detection. In the previous section, we discussed the challenge of data shortage by designing a structure with a small number of parameters. However, data scarcity is still a problem in this step, as discussed previously. Therefore, we used the proposed segmentation model as the backbone of our model since it has already shown good performance. The Dense-Unet model is inspired by U-Net and DenseNet, hence its name. For this step, we are combining the model with the Y-Net concept, resulting in Dense-YNet as the name of the proposed detection model.

There are two commonly used definitions for ARDS. In one definition, the criterion is bilateral infiltrates, while in the other definition, the condition is progressive infiltration through time in CXR images. To design a model that can handle both definitions, we can detect

the locality information across the lung. Using this strategy, we can determine whether the infiltration is diffuse across both lung sides to analyze the criteria of the first definition. For the second definition, if the model can predict the local labels by analyzing two taken images that meet the timing criterion of ARDS, the model can determine if there are any new infiltrates in lung regions. This strategy also allows the assignment of severity scores to each image based on the local label predictions. We split each lung side vertically into two halves, called lung quadrants. The model is expected to classify the labels of each lung quadrant and then analyze the combination of affected quadrant locations. In addition, the model can do lung segmentation since the output of the U-shaped model is the lung mask. The model can segment pathological images, a unique advantage over most works in this field.

To train such a model, we need a dataset that contains both lung boundary ground truth and four labels corresponding to the lung quadrants. To create this dataset, we take a set of images from various data sources and annotate the CXR images with the help of clinicians at CHUSJ. Since the model is lightweight, it can be trained on a relatively small dataset.

The algorithm requires a few additional steps before being effectively implemented in clinical settings. First, while the model has been trained and validated on a diverse dataset, it is crucial to perform an external validation to assess its reliability and generalizability. To achieve this, we have developed a web application that enables people from around the world to contribute to the external validation of the model. This is an important step for the clinical tool since data can vary significantly based on the device and imaging protocol used in healthcare centers. The web application should be designed with a user-friendly UI and backend to facilitate ease of use, and observers' opinions about the decision support system should be recorded for validation analysis.

To report appropriate severity scores, six instances of the proposed models are used to predict labels, and a voting strategy determines the final result. A single model would provide a severity score for the entire CXR image, while quadrant labels are assigned a binary label (affected/unaffected). However, using ensemble models offers the opportunity to give greater precision to the severity assignment, as well as to generate more precise lung area masks. The third objective of the thesis is crucial since the platform must meet two critical expectations of doctors. The first is the ability to visualize the system's decision clearly, and the second is to provide external clinical validation to researchers and clinicians worldwide.

CHAPTER 4 ARTICLE 1: DENSE-UNET: A LIGHT MODEL FOR LUNG FIELDS SEGMENTATION IN CHEST X-RAY IMAGES

Conference Name: 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC2020)
Publication Date: July 24, 2020

Authors:
Mohammad Yahyatabar; Philippe Jouvet; Farida Cheriet

Abstract

Automatic and accurate lung segmentation in chest X-ray (CXR) images is fundamental for computer-aided diagnosis systems since the lung is the region of interest in many diseases and also it can reveal useful information by its contours. While deep learning models have reached high performances in the segmentation of anatomical structures, the large number of training parameters is a concern since it increases memory usage and reduces the generalization of the model. To address this, a deep CNN model called Dense-Unet is proposed in which, by dense connectivity between various layers, information flow increases throughout the network. This lets us design a network with significantly fewer parameters while keeping the segmentation robust. To the best of our knowledge, Dense-Unet is the lightest deep model proposed for the segmentation of lung fields in CXR images. The model is evaluated on the JSRT and Montgomery datasets and experiments show that the performance of the proposed model is comparable with state-of-the-art methods.

4.1 INTRODUCTION

Segmentation of thoracic organs in chest X-ray (CXR) images is the most important step in several tasks of computer-aided diagnosis (CAD). In CXRs, the lung is the region of interest (RoI) where several serious diseases commonly manifest. Therefore, the segmentation of the lungs in CXR can give vital indications regarding diseases such as acute respiratory distress syndrome (ARDS) and lung cancer [81]. The expensive and time-consuming nature of manual labelling by medical experts, with its large inter-observer and intra-observer variations, highlights the need for an automatic segmentation method [68].

Segmentation methods for the lung fields in CXR are categorized as rule-based [114], model-

based [115] and machine learning-based. As in most image analysis tasks, deep learning has outperformed the previous state of the art in CXR segmentation applications. U-Net, a well-known Convolution Neural Network (CNN) model [77], has served as a baseline architecture for most CXR segmentation models. It includes a contraction path and a symmetric expansion path that are linked by several connections in between. Since 2015, many studies have tried to modify the U-Net structure to extract RoIs in CXR images [79, 81, 116, 117]. In [79], the proposed network, called LF-Segnet, transfers only the pooling indices instead of using direct connections between the contraction and expansion paths. The work in [81] utilizes far fewer parameters in comparison with U-Net and as a result, the running time for training the network is quite short. They proposed three models called All-drop-out, All-convolutional and Inverted-Net. In the All-drop-out network, a drop-out layer is added after each convolutional layer, and in the All-convolutional network, strided convolutional layers are applied instead of using pooling layers. Inverted-Net, which is derived from these architectures, is designed to be compatible with the nature of CXR data and has state of the art performance in the segmentation of multiple organs. Inspired by Inverted-Net and U-Net, [118] proposed variations of these models by employing different techniques. In spite of the high accuracy rates of these models, they are quite costly in terms of computation and memory usage. Deep feature maps throughout these models induce a huge parameter space, while a high percentage of these contain redundant information without adding new distinguishing features. Moreover, a dual-path adversarial learning (DAL) approach was proposed in [102], in which features can be learned at different levels of complexity using the generative adversarial network (GAN) concept. Using generative models in segmentation has some advantages but because of their challenging training process, their generalization confronting some variation caused by probable diseases are relatively low. In a scenario presented in [119], after an initial segmentation, a procedure to reconstruct the missed lung regions is proposed to deal with abnormalities caused by disease. Although the method results in high performance in the segmentation of lung fields, the extra processing may increase the complexity of the model.

Providing annotated data in medical applications is always costly. Another important concern is the risk of over-fitting when increasing the complexity of the model while using a limited number of training samples. Based on the nature of CXRs, capturing the finest details in the RoI boundaries is not required and a simpler model leads to better generalization [81].

As the main contribution of our study, a novel U-Net based architecture is proposed for the semantic segmentation of lung fields in Chest X-ray images. This innovation leads to fast model convergence while significantly reducing the number of parameters and keeping the segmentation performance high.

4.2 Methodology

The central idea for our proposed Dense-Unet is that by increasing the information flow throughout the model, it can be designed with a smaller number of parameters.

Convolution layers produce many similarities between intermediate feature maps for CXR images. Therefore, a connectivity pattern is applied to utilize the whole capacity of the feature maps and to reduce the redundancies, leading to significantly decreased computation cost. DenseNet uses the concatenation of outputs from various intermediate layers as the input of the later layers [27]. The connections maintain all extracted features in the different layers and also increase information flow inside the model.

4.2.1 Dense-Unet Architecture

In the proposed version of the U-Net, using dense connectivity, generated feature maps in earlier layers are employed in all later layers in an end-to-end segmentation process (Fig. 4.1). In the feedforward passes, the layers have access to all previous maps directly. This brings

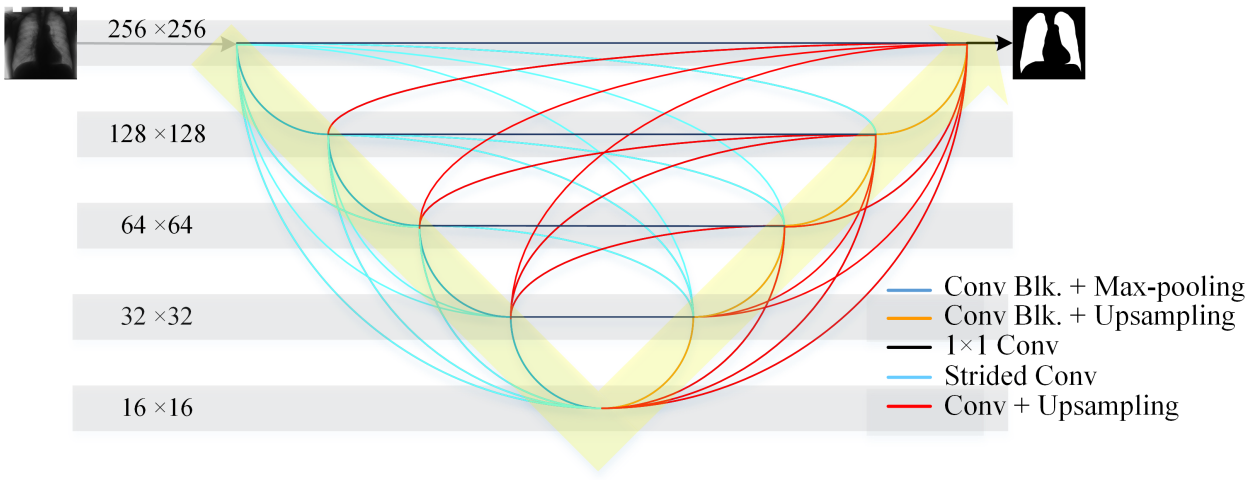


Figure 4.1 Overview of proposed Dense-Unet model

multi-level features to the layer and thus, a combination of different level maps are exploited. Moreover, in the backward gradient flow, learning is easier due to deep supervision; gradients can propagate through all the layers, even in primary ones.

Such a deep effect of the loss function on various layers in the model facilitates convergence and as a result of information flow, the model's design is lighter, with significantly fewer parameters while achieving good performance.

The connections shown in Fig. 4.1 equalize the size of transferred maps and extract features as well. There are $\frac{9\times(9-1)}{2} = 36$ inter-layer connections between the 9 layers of the network. Strided convolutions, max-pooling and upsampling operations are used to address the problem of unequal sizes within different layers. The size of the output feature maps is derived from equation (4.1):

$$n_{out} = \lfloor \frac{n_{in} + 2p - k}{s} \rfloor + 1 \tag{4.1}$$

where n_{in} and n_{out} are the sizes of the input and output tensors respectively, p shows the padding number around the map, k is the kernel size and s is the stride step.

The details of the architecture in Fig. 4.1 are as follows:

- The convolution block and max-pooling (down-sampling by 2) operations that are used in the main body of the Dense-UNet are shown as dark blue lines. A convolution block comprises two convolution layers followed by a batch normalization and a rectified linear activation function.
- The expansion path contains convolution blocks and up-sampling (instead of max-pooling in the contraction path), which are shown as orange lines. These generate feature maps $2n$ times bigger than the input map (n is the distance separating connected layers).
- The $1 \times 1 Conv$ connection (in black) does not change the map sizes; it only adjusts the depth of feature maps while extracting features. It is used to transfer feature maps in corresponding layers and also in the output layer to produce the final segmentation mask.
- The light blue lines are strided convolution layers that generate feature maps of smaller sizes. As observed in 4.1, these connections transfer maps from the encoder to the decoder and the map sizes are adjusted by tuning s.
- The last connection type is the red lines, which contain a convolution layer and upsampling operations with scaling factors equal to 2, 4, and 8.

4.3 Experiments and Results

4.3.1 Data

Two public datasets were employed to evaluate the proposed method. The first one, provided by the Japanese Society of Radiology Technology (JSRT), contains 247 chest X-rays images

with and without lung nodules. The size of the images is 2048 × 2048 pixels, with 4096 intensity levels (12 bits) [120]. In [121], reference standard masks of the thoracic organs in the JSRT images downsized to 1024 × 1024 are provided.

The Montgomery dataset was designed for tuberculosis CAD and created by the National Library of Medicine in collaboration with the Department of Health and Human Services, Maryland, USA. The dataset contains 138 CXRs as well as their reference standard masks at an image size of 4020 × 4892 pixels and 12-bit depth.

4.3.2 Training strategy

The network was trained using two-third of the images, in which 20% of the data were reserved for validating the training process and tuning the models. The performance of the models was evaluated using the remaining third of the images. To compare our work with the state-of-the-art methods, images were resized to 256 × 256.

A data augmentation technique was employed to generate new images to compensate for the limited size of the dataset. Cropping, rotation and brightness changes were the transformations applied to produce new images.

To apply non-linearity to the model, the Rectified Linear Unit (ReLU) activation function was utilized in intermediate layers and the Softmax function assigned the labels for every pixel in the output layer. The Adam optimizer was employed to minimize the loss function. Three common loss functions (Binary cross-entropy (BCE) and the Dice and Jaccard segmentation overlap scores) were considered as candidates for the optimization target. In addition, the Glorot method was used for weights initialization.

4.3.3 Performance evaluation and discussion

Table 4.1 shows the comparison of lung field segmentation performances between a human observer, state-of-the-art methods and our proposed method on the JSRT and Montgomery datasets. In this work, two complementary Dense-Unet configurations were employed to show the potential of our network design. The model architectures are the same in both configurations, but the "compact" version of the model is obtained by altering the depth of the intermediate feature maps. By using shallower feature maps, the compact model takes significantly fewer parameters while maintaining high segmentation performance. On the other hand, by increasing the number of channels in the intermediate feature maps, the "normal" model outperforms the human observer and reaches the state-of-the-art. Although the normal version is heavier than the compact one, it still needs far fewer parameters than

other reported methods.

Table 4.1 Comparison of Dense-Unet results (in percentages) with state-of-the-art methods and human observer.

	JSRT		Montgomery	
Architecture	**Dice**	**Jaccard**	**Dice**	**Jaccard**
Human observer [121]		94.6 ± 1.8		
LF-SegNet [79]		95.1		
ResNet-DAL [102]	97.5	95.1		
All-Dropout [81]	97.3	94.8		
Inverted-Net [81]	97.4	95.0		
SIFT-Flow [71]	96.7 ± 0.8	95.4 ± 1.5	96.0 ± 1.8	94.1 ± 3.4
Region Growing [114]	98.3 ± 0.7	96.3 ± 1.2	97.8 ± 0.5	96.6 ± 1.8
FASFCM [122]	97.6 ± 1.2	95.6 ± 1.5	95.8 ± 1.5	93.5 ± 2.1
Proposed (compact)	97.4 ± 0.1	94.9 ± 0.2	97.3 ± 0.4	94.7 ± 0.7
Proposed (normal)	97.6 ± 0.3	95.3 ± 0.5	97.9 ± 0.3	95.9 ± 0.5

Lighter models (those with fewer parameters) are a priority in many areas, especially for CNNs that are a part of a bigger system or that must implemented in hardware. Thus, a compact model that is robust for the segmentation of lungs in CXRs is desirable. In Table 4.2, the number of parameters and memory sizes for the original U-Net, two state-of-the-art models and the two Dense-Unet versions are listed. Clearly, the high flow of information through the network allows us to drastically reduce the number of parameters, yielding only $5,300$ for the compact model, while more than 34 million parameters are needed in the original U-Net.

Table 4.2 Comparison of Dense-Unet size with other CNN models.

Architecture	**# of parameters**	**Size**
Original U-Net [77]	34 512 388	131 MB
All-Dropout [81]	31 377 988	119 MB
Inverted-Net [81]	3 140 771	12 MB
Normal Dense-Unet	352 107	1.47 MB
Compact Dense-Unet	5 304	145 KB

Squeeze and Excitation (SE) blocks [123] can be added to CNN networks, and these are also taken into account in the proposed model. These additive blocks promote independence between the channels by adaptively recalibrating the extracted features. A variation of SE using spatial and channel-wise recalibration was introduced in [124]. These modules add a little complexity but attenuate the weaker feature maps and boost the more informative ones.

To evaluate the effect of SE blocks, Table 4.3 compares the experimented models in terms of segmentation mask Jaccard scores and number of parameters.

Table 4.3 Effect of SE blocks on Dense-Unet performance.

	Jaccard		
Architecture	**JSRT**	**Montgomery**	**#parameters**
Normal model with SE	0.953	0.959	352 107
Normal model without SE	0.953	0.959	336 713
Compact model with SE	0.949	0.947	5 304
Compact model without SE	0.934	0.931	5 051

The SE blocks increase the segmentation performances of the compact model while adding a small number of parameters. However, for the normal Dense-Unet, the Jaccard score remains unchanged while the number of parameters rises when using SE blocks.

We also studies the effect of different loss functions on our models. Networks were trained by optimizing three losses, BCE, Dice score and Jaccard score (Table 4.4). Utilizing the Jaccard score as the loss function results in the best overlap rates in both datasets.

Table 4.4 Effect of loss function on performances in both datasets. (Jc: Jaccard, Se: Sensitivity, Sp: Specificity)

Dataset	**Model**	**Loss Fcn.**	**Jc(%)**	**Se(%)**	**Sp(%)**
JSRT	Normal	BCE	94.4	97.4	98.6
		Dice	94.5	97.3	98.7
		Jaccard	95.3	97.9	98.8
	Compact	BCE	93.6	97.8	98.7
		Dice	94.6	97.8	98.0
		Jaccard	94.9	97.8	98.7
Montgomery	Normal	BCE	93.8	97.4	98.6
		Dice	94.6	97.0	99.1
		Jaccard	95.9	98.1	99.2
	Compact	BCE	94.2	96.8	99.1
		Dice	94.5	97.2	99.0
		Jaccard	94.7	97.3	99.1

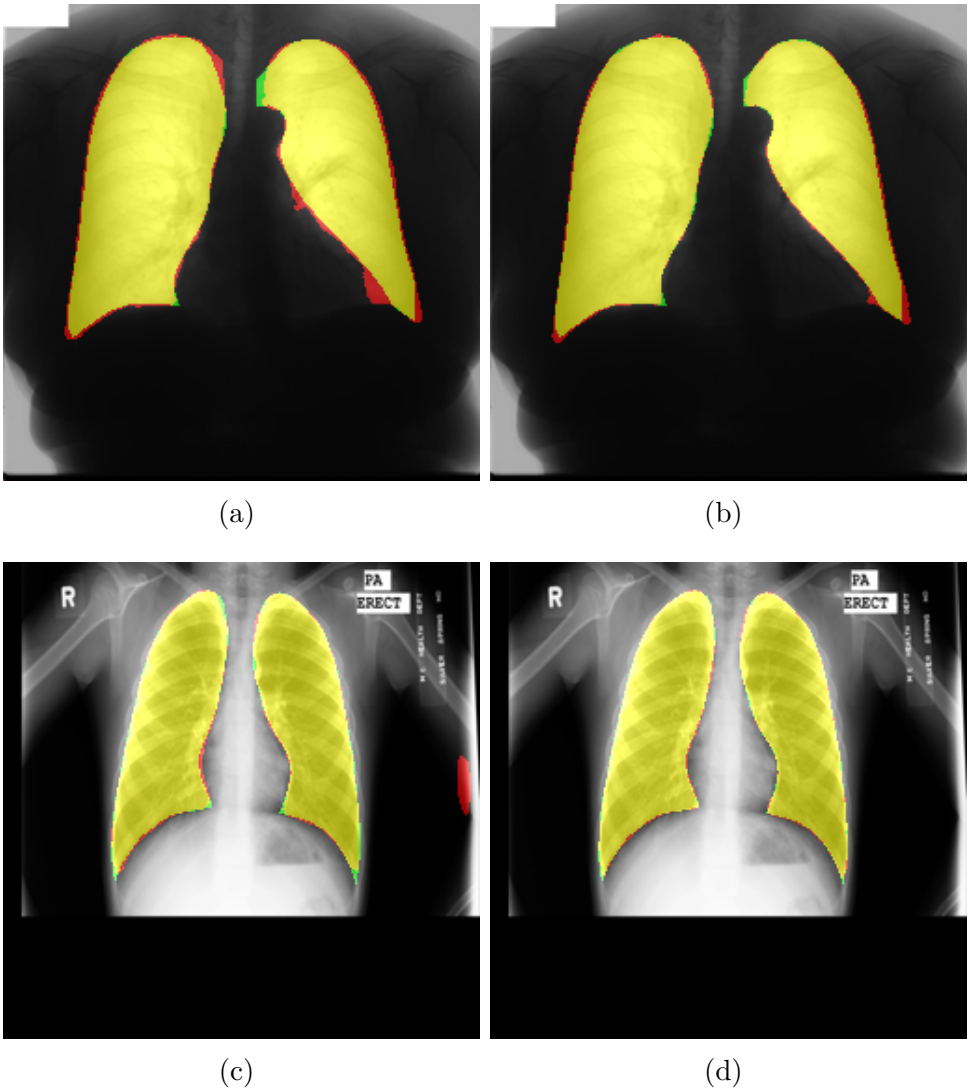
The experiments were performed using the PyTorch platform in Python 3.7 on a desktop PC with an Intel i7-4790 CPU and NVIDIA Geforce RTX 2070 GPU with 8GB RAM.
If the specified CAD application requires maximal precision for the organ boundaries, according to the observed results, the normal Dense-Unet would be the best choice. If however reducing memory usage is the priority, the compact model may be used while still achiev-

ing satisfactory accuracy. Fig. 4.2 shows two segmentation examples from the JSRT and Montgomery datasets by the compact and normal models.

As mentioned above, data augmentation provides synthetic images to generate new networks feeds. Generally, by increasing the augmentation probability, the performance of our model increases (Fig. 4.3). However, the overlap index rises for the Montgomery dataset more than for JSRT when applying data augmentation. In fact, in spite of the JSRT, the CXR images are more complex and heterogeneous in Montgomery. This can explain why adding more training variations has a greater impact on the generalization of the method for the Montgomery dataset.

4.4 Conclusion

In this paper, a novel CNN model inspired by DenseNet and U-Net is proposed for the segmentation of lung fields in CXRs. Taking advantage of high information flow, all layers in the proposed Dense-Unet are connected. This allows us to design a model with far fewer parameters and thus, a greatly reduced memory footprint. Two configurations of this novel architecture are introduced for different applications, depending on whether segmentation accuracy or small model size is the priority. Moreover, adding Squeeze and Excitation blocks brings the performance of the compact Dense-Unet close to that of the normal version. Using data augmentation, we compensate for the limited size of available training sets and enhance the models' generalization capacity on more heterogeneous image sets. In terms of loss functions, we observed the Jaccard overlap score to be a more appropriate optimization target than cross-entropy or Dice.



(a) (b)

(c) (d)

Figure 4.2 Visual results of lung field segmentations for JSRT (top row) and Montgomery (bottom row) images. Figures a) and c) (resp. b) and d)) are the masks produced by the compact (resp. normal) model. The areas overlapping with the reference masks (true positive) are shown in yellow, while the green areas are from the ground truth (false negative) and the red areas are from the predicted mask (false positive).

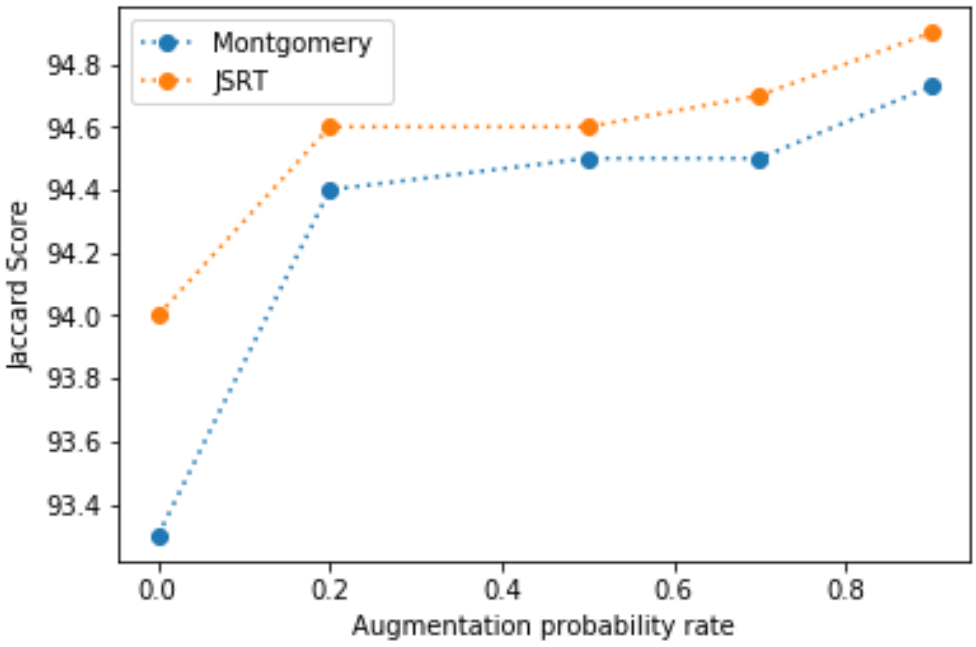


Figure 4.3 Jaccard scores vs. data augmentation probability.

CHAPTER 5 ARTICLE 2: JOINT CLASSIFICATION AND SEGMENTATION FOR AN INTERPRETABLE DIAGNOSIS OF ARDS FROM CHEST X-RAYS

Journal Name: Journal of Medical Imaging
Submission Date: March 2, 2023
Publication Date: October 17, 2023

Authors:
Mohammad Yahyatabar; Philippe Jouvet; Farida Cheriet

Abstract

Purpose: Acute respiratory distress syndrome (ARDS) is a life-threatening condition that can cause a dramatic drop in blood oxygen levels due to widespread lung inflammation. Chest radiography is widely used as a primary modality to detect ARDS due to its crucial role in diagnosing the syndrome, and the X-ray images can be obtained promptly. However, despite the extensive literature on chest X-Ray (CXR) image analysis, there is limited research on ARDS diagnosis due to the scarcity of ARDS-labeled datasets. Additionally, many machine learning-based approaches result in high performance in pulmonary disease diagnosis, but their decisions are often not easily interpretable, which can hinder their clinical acceptance. This paper aims to develop a method for detecting signs of ARDS in CXR images that can be clinically interpretable.

Approach: To achieve this goal, an ARDS-labeled dataset of chest radiography images is gathered and annotated for training and evaluation of the proposed approach. The proposed deep classification-segmentation model, Dense-Ynet, provides an interpretable framework for automatically diagnosing ARDS in CXR images. The model takes advantage of lung segmentation in diagnosing ARDS. By definition, ARDS causes bilateral diffuse infiltrates throughout the lungs. To consider the local involvement of lung areas, each lung is divided into upper and lower halves, and our model classifies the resulting lung quadrants.

Results: The quadrant-based classification strategy yields the area under the receiver operating characteristic curve of $\%95.1(95\%CI 93.5 - 96.1)$, allowing for providing a reference for the model's predictions. In terms of segmentation, the model accurately identifies lung regions in CXR images even when lung boundaries are unclear in abnormal images.

Conclusions: This study provides an interpretable decision system for diagnosing ARDS,

by following the definition used by clinicians for the diagnosis of ARDS from CXR images.

5.1 Introduction

Acute respiratory distress syndrome (ARDS) is a widespread inflammatory lung condition having various causes (mainly infections including COVID-19, aspiration, or trauma) and a high mortality rate. It can occur at any age and appears because of alveolar inflammation preventing the alveoli sacs from exchanging air with the bloodstream. The mortality rates corresponding to mild, moderate, and severe ARDSs are 27%, 32%, and 45% respectively, hence the need for quick and accurate management [1]. Late diagnosis and treatment delivery to such patients may complicate the treatment process and worsen the disease. Furthermore, accurate, fast and automated triaging can help to balance resource assignments, especially in high-demand clinical centers.

Based upon the most accepted definition of the syndrome [13], one of the most limiting and frequent factors in diagnosing ARDS is the presence of bi-lateral infiltrates in the chest X-ray (CXR) image. Considering the central role of CXR images in clinical management, an important issue emerges due to the high variability between physicians when reviewing such images ([26, 125]). Our group previously demonstrated that using a computer-aided diagnosis system to detect ARDS from CXRs can increase the accuracy of diagnosis [26].

Few works are found in the literature that address automated ARDS diagnosis from chest X-rays. Zaglam et al. [26] suppress the ribs visible in the lung areas based on the method in [25] and extract the patches in the inter-costal regions. Exploiting the statistical and spectral characteristics of the patches, texture features are extracted and transformed into a new space by linear discriminant analysis (LDA). A Support Vector Machine then classifies the samples as ARDS or non-ARDS. The authors achieve promising results and improved the agreement rate between intensivists. However, the patch-based nature of the system raises some challenges. A patch annotation process is required to provide the training data. Moreover, predicting the diagnosis from rows of patches raises challenges such as selecting their sizes and setting a threshold on the number of affected patches. Also, hand-engineered features are extracted from the intercostal spaces and the method involves an extra step for rib removal. Sjoding et al. exploit the capability of deep learning (DL) to achieve expert physician-level performance in ARDS detection [28]. First, the state-of-the-art DenseNet-121 model [27] is pretrained on two well-known public datasets, Mimic [126] and Chexpert [39]. Then, using the transfer learning strategy, their model is retrained on a University of Michigan (UM) ARDS dataset containing 8072 CXR images. The evaluation results confirm an accurate and robust diagnosis system. However, the system does not provide

any information to justify the decision, thus limiting its acceptability by clinicians.

Considerable research has been devoted to the interpretability of DL models' decisions. Two main paradigms have emerged in this area: learning an interpretable model (e.g. [127]), and post-hoc explainability of a model (e.g. [128]). In image classification, saliency maps are a common tool for highlighting the image areas contributing to the classification results. However, such maps are not reliable enough to be trusted in medical applications [129, 130]. Pixel-level segmentation prior to the classification stage limits the processing areas and assists the classification task [62], especially when training data is scarce. Besides, by limiting the region of interest, the effect of non-relevant background regions and image noise is compensated, thereby boosting the performance of the model.

In the present study, instead of simply taking the result of the segmentation process, we propose a joint segmentation and classification model that shares the weights and allows end-to-end training. Severe ARDS cases usually have high opacity in the lung areas, and such CXR images can mislead typical segmentation models. By sharing segmentation and classification weights in the integrated model, segmentation yields an improvement in classification. However, running segmentation and classification in series risks propagating any segmentation errors to the entire framework. Thus, the main rationale behind joining the segmentation and classification tasks is to take full advantage of both feature spaces to improve the performances interactively. In addition to the classification loss, the segmentation loss is also used which leads the model to focus on lung fields as the region of interest.

This work addresses the crucial clinical requirement of providing an interpretable framework for diagnosing ARDS. Indeed, clinical decision support systems must clearly show how decisions are taken, otherwise they won't be approved by government health authorities. To deal with this, we follow standard clinical practice in diagnosing pulmonary disorders from CXRs. Radiologists first exclude non-lung areas for their analysis, then explore signs of diseases in both lung sides. Visible infiltration effects diffused in both lung sides are an important sign of ARDS [13]. Thus, predicting local labels in each lung side not only facilitates the classification process, but also provides information for the local areas throughout the lung.

Lung quadrants are extracted by splitting each lung side into upper and lower lobes, and the final decision is based on the quadrants' predicted labels. Thus, a quadrant-based diagnosis system is proposed to deliver explainable decisions in line with clinical practice. The main contribution of this paper is to propose an interpretable decision support system for the diagnosis of ARDS. A joint segmentation and classification model provides a high-accuracy system that can identify each lung quadrant as containing ARDS infiltration or not. In parallel, the lung areas are accurately segmented by the model. Furthermore, a dataset is

provided that contains CXR images with labels of ARDS for each lung quadrant.

In the following section, we present related works on detecting pulmonary pathologies. In Section 5.2, data challenge and the provided ARDS dataset are justified; the proposed method for joint lung segmentation and disease classification is then explained. In Section 5.3, we describe the experimental setup and protocols. Section 5.4 discusses our experimental results in detail. Section 5.5 provides a conclusion to this work.

5.1.1 Related works

Deep learning has achieved remarkable outcomes across a range of applications, spanning from communication [131] and industry [132, 133] to healthcare systems [134]. In medical image processing, decision support systems assist doctors in various states, such as diagnosis. Analysis of chest radiographs is mainly linked to the classification of pathological images. Many studies focus on training new models or existing state-of-the-art architectures from scratch. Kieu et al. train three CNN models to detect abnormalities in CXRs [42]. They investigate the left lung, right lung, and both together, each in a separate model, and synthesize the results into final predictions. Islam et al. employed an ensemble of several models, including Alexnet, Resnet, and VGG [45]. The final result outperformed the merely use of classifiers for detecting Cardiomegaly and Tuberculosis. Li et al., in a two-stage model, take advantage of a DenseNet [27] encoder and recurrent neural network (RNN) decoder for diagnosis of fourteen diseases in the ChestX-Ray14 dataset [46]. The RNN exploits the dependencies between labels to improve model performance.

On the other hand, the presence of massive datasets, such as Chest X-Ray 14, MIMIC, and CheXpert, has encouraged researchers to take advantage of transfer learning. In this paradigm, the model discovers the underlying data structure and employs the acquired knowledge for specific downstream tasks. Transfer learning finds extensive use in the analysis of CXR images [3, 28, 49]. However, it often requires the sharing of data, leading to privacy concerns, especially when dealing with sensitive medical information. Here, federated learning steps as a solution, employing a central server and individual clients to facilitate private model training without exposing sensitive data [132]. During the COVID-19 pandemic, there was significant demand for efficient medical diagnosis. However, sharing patient data for training purposes was restricted by privacy regulations. A collaborative federated learning approach enabled multiple medical institutions to collaboratively develop accurate COVID-19 screening models without the need to share patient data [54, 55]. This resulted in competitive outcomes when compared to centralized data-sharing models. Slazyk et al. also noted that partitioning the identical dataset among separate federated learning clients could potentially

result in enhancements in the classification accuracy of CXR images for the models under examination [56].

Transformers, renowned for their achievements in NLP, are currently extending their capabilities to image-based ML problems. By combining EfficientNet with the Vision Transformer model, researchers achieved a remarkable 100% AUC in classifying tuberculosis from CXR images, demonstrating the adaptability of Transformers beyond their original NLP domain [52]. Vision Transformers partition images into patches for processing through encoder layers with self-attention, allowing them to effectively capture spatial relationships. This novel strategy has found success in medical applications, such as detecting Covid-19 in CXR images, where data scarcity was overcome by utilizing a public dataset for feature extraction and subsequent abnormality detection [53]. This expansion of Transformers into image-based tasks points out their potential to rise up various fields of machine learning and image analysis.

Some methods also search for features locally instead of in the full image. In a work published by Qingji et al. [59], a category-wise residual attention learning (CRAL) framework is presented in which attention blocks are arranged to strengthen the feature learning and find correlations among pathologies. In another study, global and local image features are exploited at the same time [48]. An ROI is generated by thresholding the feature maps detected in a global branch of the model; this drives the attention mechanism used in a local branch. Their classification results on ChestX-ray14 outperform the model with only a global branch. Xiao et al. also employed two separate pathways to consider both local and general information [49]. A U-net model is used to segment the lung masks. Then, location maps are derived based on the Euclidean distances from the masks and combined with the original images to feed a Densenet-121 for classification.

Madani et al. employ a Generative adversarial network (GAN) in a semi-supervised learning architecture to address the classification and data scarcity problems [135]. The model shows good generalization in facing unseen datasets. Yu-Xing et al. also propose a GAN to identify abnormal CXR samples [136]. Their model is trained to reconstruct normal CXR samples and, if the input is abnormal, the model poorly reconstructs the input which is considered as the measure for distinguishing CXRs.

Jennie et al. [57] believe that the normal sizes of CXR images used in deep models are not sufficient to classify pneumothorax. They separated two top third of each lung as apex images and extracted patches from them. The patch-based and apex-based fine-tunings of VGG-19 are proposed to address the size issue. Ensembling models based on soft voting improved the performance of pneumothorax detection. Resnet-50 and Densenet-121 are also popular models in the interpretation of CXRs. Ahmed et al. apply horizontal and vertical

components of the wavelet transform to enhance images and facilitate feature extraction [58]. Then, a pretrained Resnet-50 classifies the derived maps while almost half of the convolution layers and classifiers are retrained. Baltruschat et al. give a comprehensive comparison of approaches using Resnet [59]. The authors investigated the effect of non-image features(age, sex, and angle position) as well as various training strategies evaluated on ChestX-ray14 dataset [2].

There is, to date, no universal approach that is effective at classifying all the pathologies identified in public CXR datasets. ChestX-ray14 is one of the most extensive publicly available annotated datasets and includes fourteen common chest pathologies. As such, it has been employed in many studies. Rajpurkar et al. propose the ChexNet model, in which a pretrained Densenet-121 is retrained on ChestX-ray14 [3]. As noted above, several popular DL models have been employed to classify different pathologies in common datasets. However, infiltration, as a general symptom of several diseases, is targeted by several works. ChexNet reaches a state-of-the-art area under the ROC curve (AUC) of 0.73 on the classification of infiltration when tested on ChestX-ray14. Nevertheless, to diagnose ARDS, the diffusion of infiltrates throughout the lungs is challenging to detect, yet needs to be addressed in ARDS prediction.

In response to the COVID-19 pandemic, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the research community employed several AI techniques to automate diagnosis of the pathology from chest radiograph images. Unsurprisingly, DL methods play a critical role in the domain. Linda et al. propose a deep convolutional model called COVID-Net in which a machine-driven design exploration strategy is exploited to enhance the representational capacity of the model using a lightweight design pattern and selective long-range connectivity [137]. The model uses a sequence of lightweight residual projection-expansion-projection-extension (PEPX) blocks connected by residual connections throughout the model. The model obtains 92% accuracy in classifying normal, non-COVID pneumonia, and COVID-19 cases on their proposed COVIDx dataset. Ferhat et al. [138] fine-tune a SqueezeNet pre-trained network with Bayesian optimization to classify COVID-19 images, showing promising results on a small dataset. Muhammad et al. [139] evaluate COVID-ResNet on the COVIDx dataset and achieved better results than COVID-Net. They present a three-step training to tune a pre-trained ResNet-50 by progressively resizing the image to 128×128, 224×224 and 229×229 and fine-tune the model in each stage. Ezz El-Din et al. [140] propose COVIDX-Net, composed of seven different models. They reported that the VGG19 and DenseNet models yield the best performances for automated COVID-19 classification. Yujin et al. propose a framework containing segmentation and multiple patch classifiers [141]. After extracting the lung fields using the (FC)-DenseNet103, the masked

image is fed to a classification model using patch-by-patch training and inference. The final decision is based on majority voting with several ResNet models. Since their algorithm is patch-based, the quality of the segmentation is critical in their framework. Any segmentation error propagates through the classification models; this is especially problematic for COVID-19 cases where the lung boundaries can be difficult to identify.

In our approach, we aim to take advantage of the segmentation process while avoiding the limitations of serial usage of segmentation and classification models.

5.2 Materials and Methods

5.2.1 ARDS dataset

Most studies in the field of CXR image analysis employ existing publicly available datasets. MIMIC [126], Chexpert [39], and Chest X-Ray 14 [2] are of the most frequently used sources, and include between fourteen and eighteen pathologies. These datasets provide a total of one million CXR images covering several diseases [142]. However, none of them contain ARDS labels, an obvious limitation to analyzing this particular illness. Nevertheless, the datasets provide appropriate sources to pretrain deep networks to extract relevant CXR features.

In the present study, the specific dataset for ARDS classification is generated from four sources. In all cases, the lung is divided into upper and lower regions, thus four quadrants are labeled for this study. Our first dataset comes from a survey by Zaglam et al. [26] and contains 90 CXR samples annotated with labels for each lung quadrant by intensivists at Sainte-Justine University Hospital Center (SJUHC) in Montreal, Canada. Secondly, we use 100 CXR images from the Chest X-Ray 14 dataset [2] labeled as infiltrated/non-infiltrated at the quadrant level. Moreover, 183 CXR samples are derived from the PARDIE study [143]. In those two CXR sets, the quadrants are labeled by two intensivists at SJUHC; each image is included in the dataset only if both observers reach a consensus about its quadrants that results in 218 CXR images, of which 134 samples are labeled as ARDS. Additionally, we use 138 images from the Montgomery dataset [144] labeled as non-ARDS (no infiltration for all quadrants) with confirmation by SJUHC clinicians. Thus, the created ARDS dataset includes 356 CXRs of which 134 images meet the bilateral infiltrate criterion in the definition of ARDS [13].

In terms of segmentation data, the dataset also includes lung contours references. Lungs masks of 134 ARDS images are provided by two SJUCH intensivists. The images coming from Montgomery also consists of lung segmentation maps. For the other samples in our collected dataset, a Dense-Unet segmentation model [8] is used to segment the lungs, and

the resulting masks are included in the dataset after clinical validation and post-processing when necessary.

5.2.2 Joint segmentation-classification

Using two models in series, segmentation and classification, introduces limitations for diagnosis of ARDS. Indeed, any segmentation error in identifying the lung areas will propagate to the downstream network and cause classification errors as well. To address this, we want to design a model that is aware of the task of pathology localization. Moreover, limiting the computational area by segmentation ground-truth can also promote the classification function.

The proposed method in this paper is inspired by the Dense-Net [27], U-Net [77] and Y-Net [145] architectures. The dense connectivity pattern in DenseNet has reached remarkable results in analyzing CXR images in public datasets. Here, we use our previously proposed Dense-Unet [8], a light and robust segmentation model based on both U-Net and Dense-Net, as the backbone of a novel joint segmentation-classification model. As in Y-net [145], a head is branched from the bottleneck of the U-shaped model; we add to this design a second head from the decoder output to provide another classification feed. The features extracted from the two heads are concatenated, and by a sequence of convolution blocks, the output is created. Fig. 5.1 shows the new Dense-Ynet model architecture, indicating the layers and connections and the sizes of the feature maps.

In essence, the Dense-Ynet architecture is composed of a segmentation model in the form of a Dense-Unet backbone [8] (highlighted in gray and blue in Fig. 5.1), and convolutional layers sequences (highlighted in pink in Fig. 5.1) as the classification model. These two parts are interconnected with shortcut connection layers at multiple levels. The feature maps in each feature extraction level are accessible in the layers of the decoder path of the Unet. Feature maps are concatenated in every layer in the encoder-decoder pathways. Strided convolution and up-sampling+convolution operations are used to create maps of the same size. By configuring the s value in strided convolutions, the feature maps become accessible across various layers. While the input image maintains a size of 224×224, the intermediate feature maps take on square dimensions—112, 56, 28, and 14. For instance, connecting a layer of size 112 to another layer sized 28 involves utilizing strided convolution with an s value of $s = \frac{112}{28} = 4$. Similarly, applying the same principle, upsampling, tuned by the r value, enables the connection of a smaller-size layer to a larger one. This connectivity pattern increases the information flow and supports the design of a model with a lower number of training parameters and higher generalization capacity [8]. Such interconnections shorten the back-

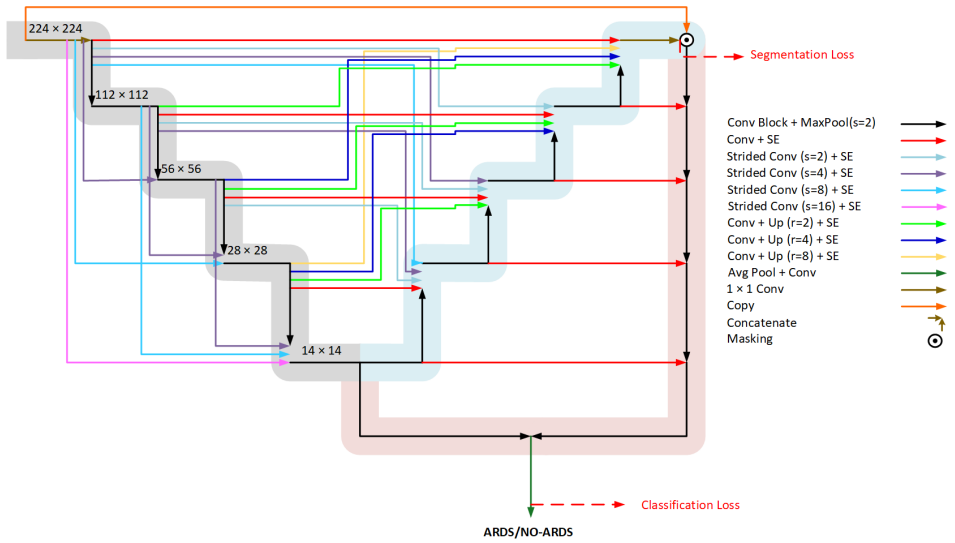


Figure 5.1 Architecture of our proposed Dense-Ynet, a joint segmentation-classification model. The colored arrows (legend at right) show the different operations done in the network layers and in the connections. Numbers indicate feature map sizes. SE: Squeeze & Excitation blocks.

propagation path and intensify supervision of loss values on trainable weights throughout the model. The input image is masked with the predicted segmentation maps at the end of the decoder and then, passes through a series of convolution blocks (right pathway of the model in Fig. 5.1). Label prediction relies on two sets of feature maps extracted from the original image and the masked one. A first branch connects to the end of the encoder path containing 14×14 feature maps. The other convolutional branch connects to the end of the decoder path, after creating a segmentation map. This branch is fed with the original CXR image masked with the lung fields extracted by the segmentation. To guide both the segmentation and classification processes, the entire model's loss function includes two weighted loss terms corresponding to both tasks:

$$L_{total} = w_{cls}L_{cls} + w_{seg}L_{seg} \tag{5.1}$$

where L_{cls} and L_{seg} refer to loss functions for classification and segmentation respectively. L_{Total} is the loss for the entire model. The weight values w_{seg} and w_{cls} are subject to:

$$w_{cls} + w_{seg} = 1 \tag{5.2}$$

The loss weights are considered as hyper-parameters and tuned using validation data. For

segmentation, the Jaccard index loss function is used as follows:

$$L_{seg} = Jaccard(s, m) = 1 - \frac{\sum_{i \in pixels}(s_i.m_i)}{\sum_{i \in pixels}(s_i^2 + m_i^2 - s_i.m_i) + \epsilon} \tag{5.3}$$

where the predicted segmentation and ground-truth maps are s and m respectively, and the summations are over all pixels i. In the denominator, ϵ is the stabilizer to avoid the term being undefined. Moreover, binary cross-entropy (BCE) for classification, as follows:

$$L_{cls} = BCE(y, p(y)) = -\frac{1}{N}\sum_{i=0}^{n} y_i.\log(p(y_i)) + (1 - y_i).\log(1 - p(y_i)) \tag{5.4}$$

where y and $p(y)$ are the groundtruth and predicted label values of the image.

Unlike in the U-Net, the number of feature maps does not increase in deeper layers of the encoder pathway in Dense-Ynet. The main pathway of Dense-Ynet employs feature maps with a depth of 100, which is different from the conventional depths of 64, 128, 256, 512, and 1024 in U-Net. By making this architectural choice during hyper-parameter tuning, Dense-Ynet achieves a lightweight profile featuring 3 million trainable parameters—markedly different from U-Net's 34 million. As illustrated in Fig. 5.1, 3×3 convoluion kernels are used for feature extracting during throughout the model except for the last layer that a 1×1 kernels are used. Squeeze and excitation (SE) blocks [124] are a type of attention-based block employed to recalibrate the feature maps in the depth and spatial dimensions. The idea is to boost the meaningful features and weaken the redundant or irrelevant ones. The channel-wise squeeze operation specifies a single value along with the depth of the map, and those values act as weights for the channel in the excitation step. The same process is done along with spatial units of the feature maps. The SE block used in our model is a combined version of channel- and spatial-wise blocks.

5.2.3 Quadrant-based classification

In the quadrant-based classification, each output passes through a *sigmoid* function and independently points to a quadrant. Quadrant labels are assigned after thresholding the four values, and the final label is based on the following logical operation:

$$P_{ARDS} = (P_{RL} \vee P_{RU}) \wedge (P_{LL} \vee P_{LU}) \tag{5.5}$$

in which P_{RL}, P_{RU}, P_{LL}, and P_{LU} are the predicted labels for right lower, right upper, left lower and left upper quadrants respectively. P_{ARDS} represents the final decision for the

presence of ARDS as manifested by bi-lateral infiltration in the CXR, where the $\vee$ and $\wedge$ symbols mean logical *OR* and *AND* operations. In other words, the image is an ARDS case if at least one quadrant in each lung is labeled as "infiltration". Our experimental ARDS dataset contains four labels associated with each lung quadrant.

5.3 Experimental Protocol

First, the proposed joint segmentation-classification method is evaluated in terms of the detection of ARDS. To this end, Dense-Ynet is evaluated with a five-fold cross-validation approach, and the experiments are repeated ten times to have reliable results. New images are also created from existing CXR images using basic data augmentation techniques such as random rotation, cropping and shifting. The rectified linear unit ($ReLu$) function provides non-linearity inside each convolution block, and at both outputs (classification and segmentation), the *sigmoid* function gives the output segmentation mask and labels values. The *ADAM* optimizer is used to train the model, and the model weights are initialized with the *Glorot* algorithm. The architecture was tested with various combinations of hyperparameters to determine the optimal configuration. In addition, observing the learning trend for a few epochs helped specify other factors, such as the probability rate for data augmentation and activation functions.

The training was configured with a segment weight parameter (w_{seg}) set to 0.6. This parameter is used to combine classification and segmentation loss values. The depth of feature maps within both main pathways (highlighted sections in Fig. 5.1) was set to 100, which directly influences the model's capacity. Among a variety of loss functions, Jaccard and BCE demonstrated the optimal configuration for training. The maximum number of epochs used during training was set to 300. Training was halted when no change in the total loss value on the validation set was observed for ten consecutive epochs. A learning rate of *3e-3* was employed to balance the convergence speed and fine-grained weight updates. For efficient data processing, a batch size of 12 was utilized, striking a balance between memory efficiency and computational throughput. The augmentation probability was also set to 0.95, enriching the training dataset and enhancing the model's robustness.

For the training and validation stages, 80% of the dataset is utilized, while the remaining 20% serves as the test set. To align with the model's specifications, the data is resized to dimensions of 224×224.

In addition, experiments are performed to justify the architecture of the backbone models used for segmentation and classification. The performance of existing state-of-the-art models

in CXR analysis is thus evaluated on ARDS detection. Then, the segmentation performance of the proposed model is assessed, and the results are compared to existing segmentation models. Finally, the impact of lung segmentation on the ARDS classification task is assessed.

5.3.1 Performance evaluation metrics

The model utilizes a sigmoid activation function for the output layer, yielding continuous outputs between zero and one. Choosing threshold values is context-sensitive and crucial for the task. Precision and Recall are key metrics used for threshold selection. In an ARDS detection system, detecting cases outweighs false positives. Thus, minimizing false negatives takes priority over decreasing false positives. This emphasizes Recall over Precision in our use case. Balancing these metrics involves a trade-off, as adjusting the threshold leads to different combinations. Another objective in threshold selection is to maximize the *F1* score, the harmonic mean of precisio n and recall.

Our system incorporates clinical considerations when selecting the threshold, accounting for both criteria. Specifically, we aim for a high *F1* score while maintaining a high recall, which inherently ensures a high precision as well. This equilibrium is achieved by choosing the peak *F1* score. If multiple *F1* peaks exist, the one aligned with a higher recall value is preferred.

Additionally, if changing the threshold results in a higher recall while keeping at least 95% of the maximum F1 score, the new threshold value is picked as the optimal one. Fig. 5.2 shows the evolution of the precision, recall, and F1 score by varying the decision threshold. The threshold selection procedure is applied to the model's predictions on the validation set, and the optimal threshold is then applied to the testing set for performance evaluation.

AUPR (Area Under the Precision-Recall Curve) serves as a fundamental metric for evaluating the performance of machine learning models. The Precision-Recall curve illustrates the trade-off between precision and recall as the classification threshold is adjusted. AUPR calculates the integral under this curve, offering a comprehensive assessment of the model's proficiency in accurately categorizing inputs. This metric proves particularly valuable in scenarios with imbalanced classes, especially when positive class samples are limited.

AUROC (Area Under the Receiver Operating Characteristic Curve) is another widely used evaluation metric, particularly in binary classification tasks, which is utilized to report the capability of the model. The ROC curve plots the true-positive rate against the false-positive rate as the threshold for classification changes. AUROC calculates the area under this curve, offering insight into a model's ability to distinguish between classes across various threshold settings.

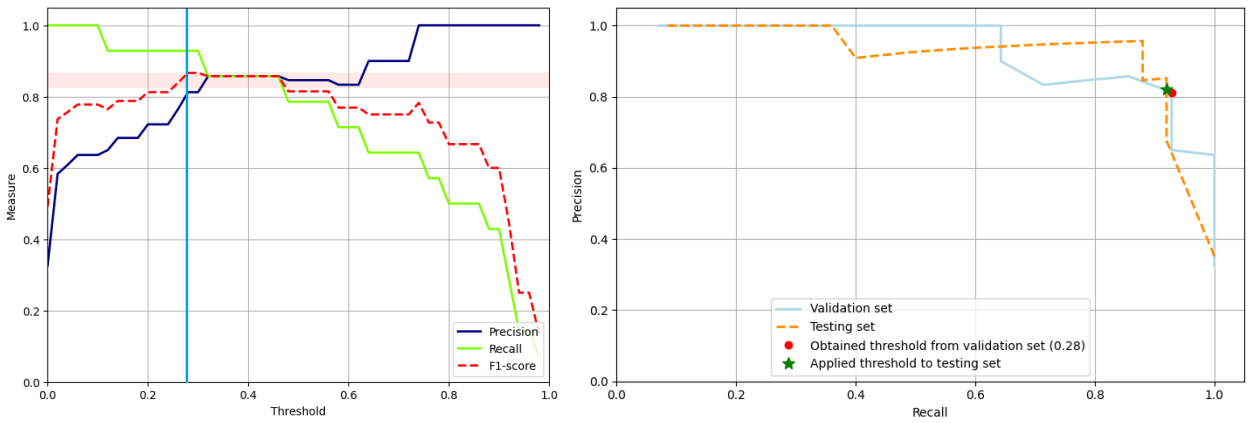


Figure 5.2 Decision threshold assignment for model prediction. (a) The dashed red line shows the F1 score; the selected threshold must keep it within 95% of its highest value (pink highlighted zone). The selected threshold is shown by a blue vertical line. (b) Precision-Recall curves for the validation and test sets, with selected threshold from (a) marked with symbols.

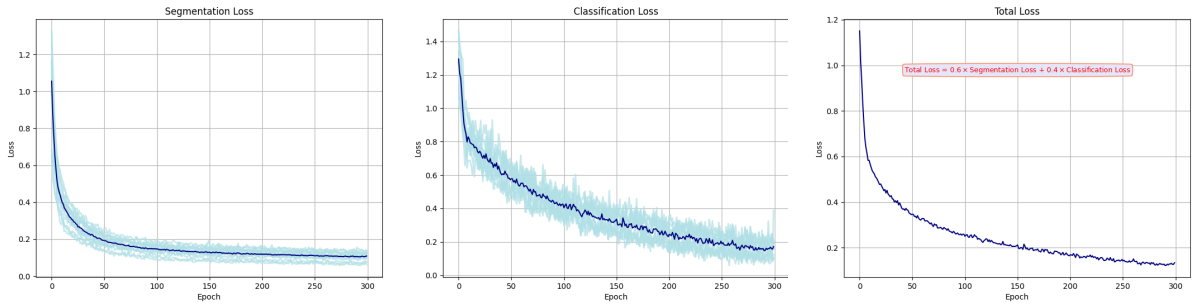
5.4 Results and Discussion

The model is trained for 300 epochs; however, the training process stops based on a specific stopping criterion: if the validation loss remains stagnant or deteriorates for ten consecutive epochs. The training process is depicted in Fig 5.3 illustrating classification loss, segmentation loss, and the total loss—a composite loss weighted by w_{seg}. As noted, a rationale for combining segmentation and classification is to leverage the impact of segmentation on classification. Upon comparing Fig. 5.3 (a) and Fig. 5.3(b), it becomes evident that the segmentation loss converges at a faster rate than the classification loss. As a result, the segmented lung information is seamlessly incorporated into the classification process, preserving the advantages of lung segmentation without the need for a consecutive segmentation-classification sequence.

The proposed model for joint segmentation-classification, Dense-Ynet, is assessed based on two strategies: image-level and quadrant-based classification. The same architecture is used in both cases, but with different fully-connected layers producing model output (four nodes for quadrant-based versus one node for image-level classification). However, selecting hyper-parameters is independently set by the validation data. Not only does the model discover ARDS in the CXR images, but a lung segmentation map is also produced, the latter being useful in many medical computer-based decision systems. Table 5.1 shows a summary of the results obtained on our ARDS-labeled dataset.

Table 5.1 Results depicting ARDS detection for both quadrant-based and image-level classification at a 95% confidence interval, presented in percentages. Precision and recall values are calculated employing the threshold selection methodology outlined in Section 5.3.1. RU, LU, RL, and LL denote the right upper, left upper, right lower, and left lower quadrants, respectively. In the "All" column, metrics are computed irrespective of quadrant positions. The concluding column presents overarching ARDS classification outcomes, based on the bilateral infiltration rule in the quadrant-based scenario. Notations: AUROC signifies Area Under the ROC Curve, and AUPR signifies Area Under the Precision-Recall Curve.

Classification Strategy	Metrics	RU	LU	RL	LL	All	ARDS
Image level	Precision						78.4(75.8-81.1)
	Recall						91.5(88.8-94.3)
	F1-score						84.1(82.2-86.1)
	AUPR						90.5(88.1-94.0)
	AUROC						93.9(93.0-94.7)
Quadrant based	Precision	85.3(82.6-88.1)	84.9(81.7-88.1)	85.0(81.6-87.5)	81.0(77.3-84.6)	84.1(81.3-86.8)	82.2(79.5-85.0)
	Recall	88.0(85.7-91.9)	89.6(86.7-92.7)	86.5(84.3-89.9)	83.6(80.1-87.0)	87.0(84.3-89.8)	89.7(86.4-93.1)
	F1-score	86.5(84.4-88.6)	86.9(84.4-89.3)	85.3(82.9-87.8)	81.6(79.2-84.1)	85.2(83.3-87.2)	85.0(82.8-86.9)
	AUPR	93.0(90.1-95.2)	91.7(89.5-94.5)	92.0(89.5-94.5)	90.3(87.7-92.7)	91.7(89.7-93.7)	91.3(89.5-93.4)
	AUROC	96.5(95.5-97.4)	96.4(95.4-97.3)	95.1(93.6-96.6)	95.2(93.9-96.5)	95.8(94.8-96.9)	95.1(93.5-96.1)



(a) (b) (c)

Figure 5.3 Training process within epochs. a) depicts the Jaccard values representing the segmentation loss over epochs. In b), the BCE loss indicative of the classification loss is presented. c) illustrates the comprehensive weighted loss, where w_{seg} is assigned a value of 0.6 as determined during hyperparameter tuning.

In the quadrant-based Dense-Ynet, the model is trained using four labels per CXR associated with each lung quadrant. This problem corresponds to multi-label classification, and the final decision is based on the quadrant labels as mentioned in Equation 5.5. The precision and recall measures are computed using the threshold selection strategy in Section 5.3.1. Quadrant-based classification demonstrates superiority across most of the metrics investigated in this study. Thus, not only does quadrant-based classification offer interpretability benefits, but it also leads to improved performance. Additionally, a DeLong statistical test was employed to compare the AUROC of our Dense-Ynet model with the popular Densenet, which has shown outstanding results in CXR analysis. The DeLong test helps in determining whether the observed differences in the AUCs of these models are statistically significant when comparing their ROC curves [146]. Dense-Ynet significantly enhances the performance of the Densenet model (AUC of 0.951 compared to 0.803; DeLong's test p-value = 0.01).

The ROC curves corresponding to the lung quadrants are illustrated in Fig. 5.4, where the shadowed areas represent the ROC curves for ten executions and the black solid line is the average curve over the executions.

Due to the threshold selection strategy, recall values are mostly higher than the precision measures. The classification of the left lower lung quadrant is more challenging for the model than the other quadrants. This might be due to the heart overlapping the lung fields mainly in the LL quadrant. Since the heart is normally denser than the air-filled lung, it causes brighter intensity, rendering decision-making difficult. It is similarly challenging to segment the lung in the LL quadrant in CXR images based on visual examination.
As observed in Table 5.1, The ARDS quadrant-based classification results in 95.1% area under

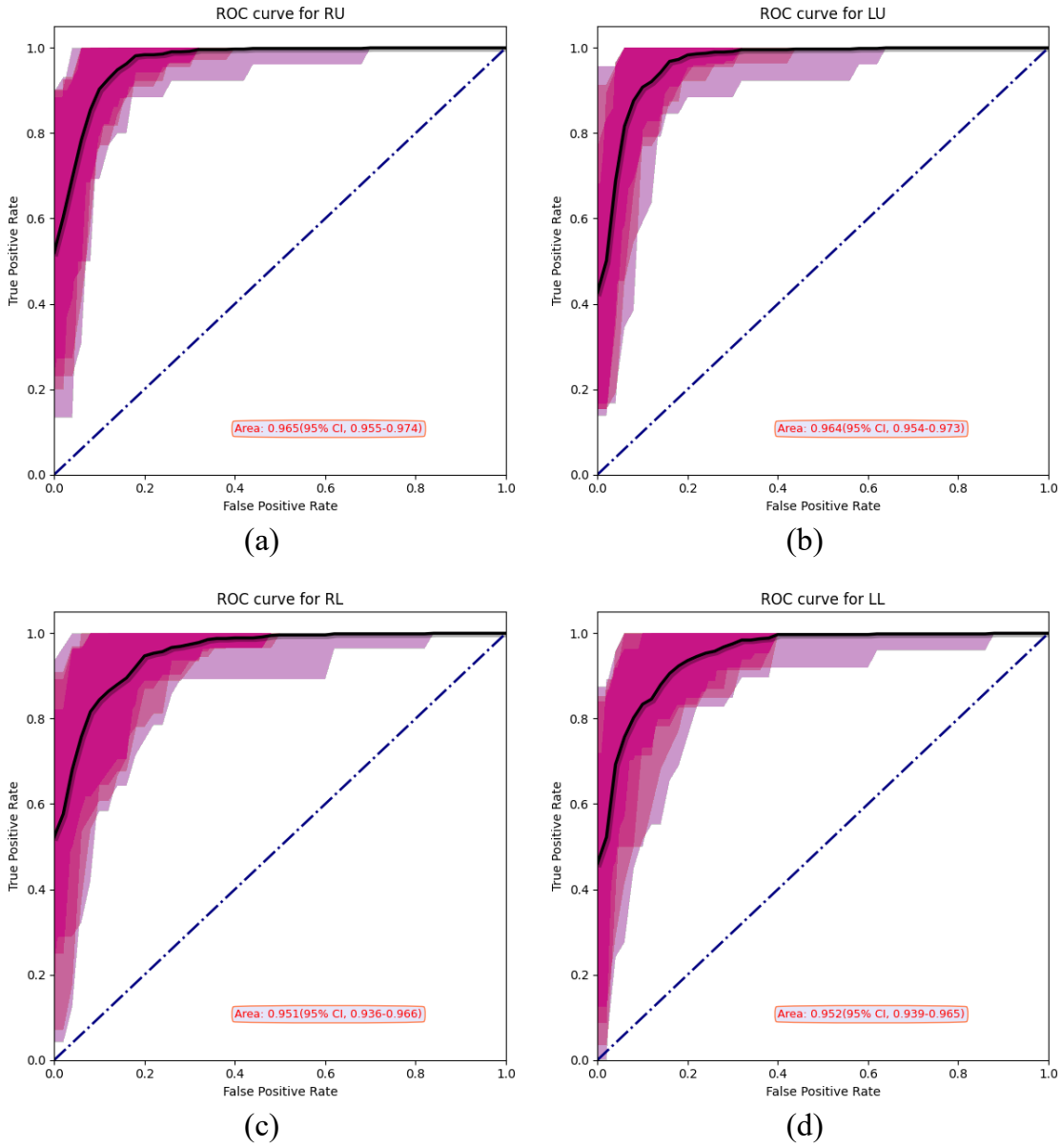


Figure 5.4 ROC curves for the four lung quadrants. The shadowed areas indicate the curves obtained for several executions and the black line shows the average value of the executions. RU: right upper quadrant, LU: left upper quadrant, RL: right lower quadrant, LL: left lower quadrant.

the ROC curve (AUROC) on previously unseen testing samples. The precision and recall scores are respectively 82.18% and 89.67%, based on the selected threshold value. Compared to the direct image-level classification, the quadrant-based approach improves classification by about 1% in terms of AUROC measure. The confusion matrix in Fig. 5.5 shows details of the classification predictions on the unseen data. In the matrix, precision and recall are shown in the first element of the last column and last row, respectively.

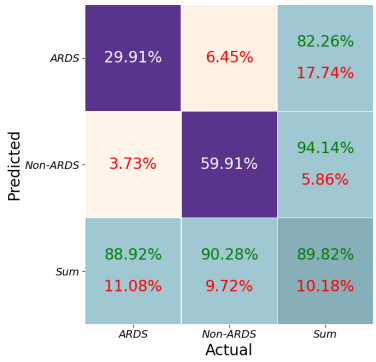


Figure 5.5 Confusion matrix of the quadrant-based classification using Dense-Ynet

Fig. 5.6 shows the ROC curves of execution for ARDS classification using the two classification strategies, namely quadrant-based and direct image-level. The curve for the image-level classification is obtained based on the one-node output of the model. In contrast, the curve for the quadrant-based one is reached by applying the threshold values determined for each quadrant and inferring the ARDS prediction using Equation 5.5.

Fig. 5.7 shows qualitative results for a few CXR images of pediatric and adult cases using our quadrant-based Dense-Ynet structure. In illustrated images, the lung segmentation results are acceptable even in images with highly diffused infiltrations. The segmentation works well in cases without infiltration or with moderate infiltration, thus accurately identifying the region of interest. Regarding classification, the ARDS labels are assigned based on the combination of quadrants, and correct classification of some quadrants does not guarantee accurate ARDS prediction. On the other hand, it is likely that the final decision will be correct despite misclassification of one or two quadrants. This is because correct classification of three lung quadrants guarantees that the global classification is also exact. Likewise, accurate classification of two quadrants on either side of the lungs could also result in a correct ARDS classification. This fact shows the importance of proposing an *interpretable* algorithm to use in clinics. By analyzing the lung quadrant predictions, clinicians can better understand the framework's decision-making references.

The idea behind the Dense-Ynet structure is to take advantage of (lung-) masked images in

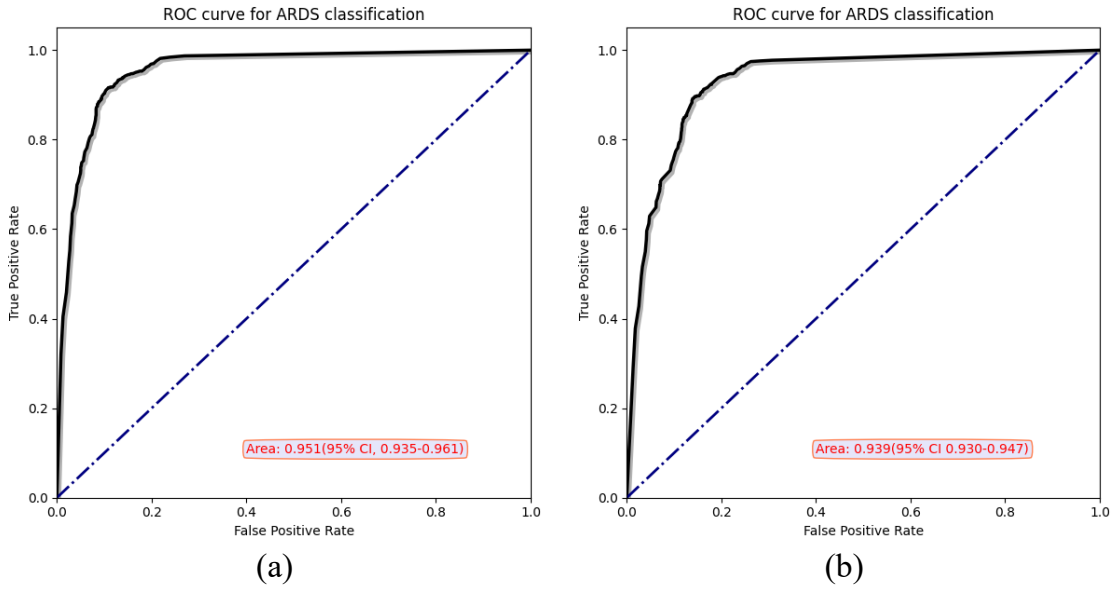


Figure 5.6 ROC curves for quadrant-based and image-level classification of ARDS

the weight sharing and interaction between the segmentation and classification tasks. Our experiments reveal that by limiting the exploration areas in the CXR image, the classification performance improves, either by adding a segmentation step before the classification or by using a unified model (Dense-Ynet). The whole idea behind focusing on the lung fields is that CXR images contain noise and misinformation from the background or can be biased towards device or pathology. The information from non-lung regions may fool the classification.

Although the quadrant-based diagnosis of ARDS provides an interpretable reference for the decision made by the framework, visualizing the activations that contribute most to the predictions can provide more insight into the capability of the unified segmentation-classification model. Fig. 5.8 gives two examples of the activation heatmaps obtained using the well-known GradCam method [128] for every quadrant using the quadrant-based classification.

5.4.1 ARDS classification by state-of-the-art models

As previously noted, multiple public datasets [2, 39, 126] are available in which hundreds of thousands of chest radiograph images are labeled in terms of multiple pathologies. These data provide the chance to use transfer learning for tasks with smaller datasets. Transfer learning enables DL approaches even for new pathologies with limited amounts of data [59]. The Chest-XRay14 dataset contains fourteen pathologies (not including ARDS) as stated previously.

Meanwhile, Densenet-121 is known as the most popular model for CXR analysis and has

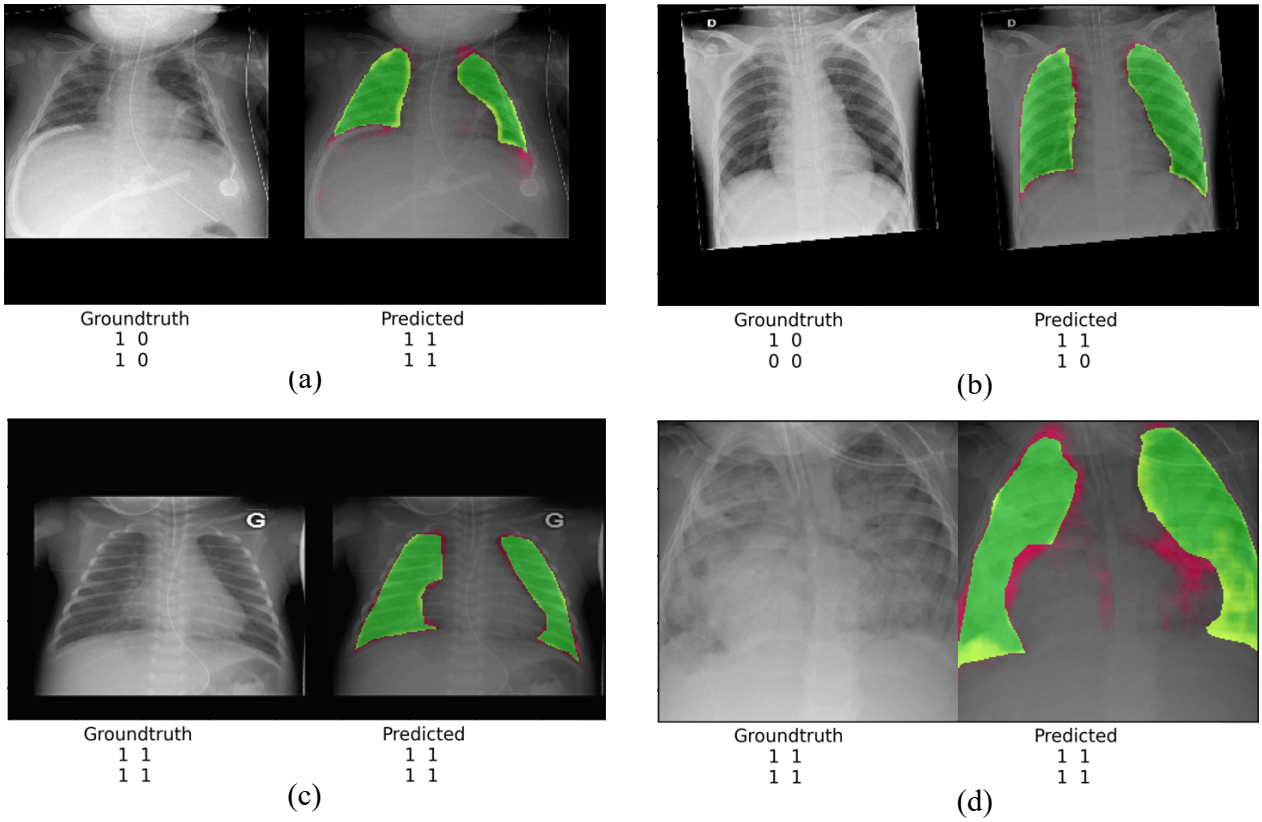


Figure 5.7 Four examples of lung field segmentation and classification results. Below each CXR image, quadrant labels and the network predictions are given. The RHS image in each row illustrates the lung segmentation ability of the algorithm. The green areas show the overlapping regions between the ground-truth and the model prediction. False-negative and false-positive areas are shown in yellow and red respectively.

Table 5.2 ARDS classification of CXR images using SOTA models with results presented at a 95% confidence interval.

Model	Average AUC
DenseNet-121	80.3% (77.1-81.9)
ResNet-50	73.3% (69.7-77.2)
EfficientNet-B1	78.4% (75.2-81.2)

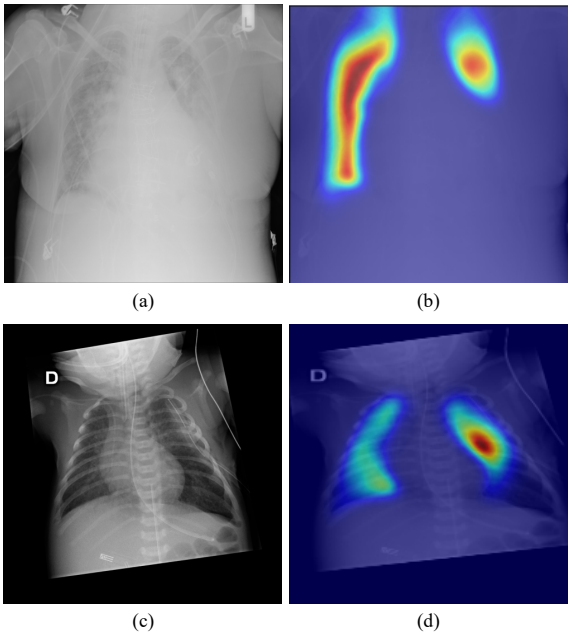


Figure 5.8 Activation heatmaps created using GradCam for the quadrant-based classification by the proposed Dense-Ynet. Red areas are the most activated ones.

achieved outstanding results [28]. To investigate the potential of different DL models in classification of ARDS, we experiment with Densenet-121 and two more state-of-the-art (SOTA) networks, ResNet-50 and EfficientNet-B1, pretrained on the large Chest-XRay14 [2] dataset. Then, the classifier part of each model is replaced with a raw fully connected layer and gets tuned by retraining on the ARDS dataset. 267 CXR images out of 356 total samples in the ARDS dataset are used to train the model. Besides, 35 images are used as a validation set to monitor the training process. The models are evaluated on the 54 remaining samples (test set). These sets are selected randomly, and the results are reported after ten times runs. Table 5.2 shows the average AUROC of ARDS classification using the three models.

Table 5.2 shows that DenseNet-121 results in the highest AUROC among the most common deep networks in CXR analysis. The potential of the Densenet model in the ARDS dataset is demonstrated in Table 5.2. There are several reasons supporting its effectiveness, including enhanced feature propagation, improved information flow, and mitigation of gradient vanishing. These factors contribute to a smaller number of required training parameters, aligning well with the constraints posed by our limited dataset. Furthermore, numerous studies in the literature have highlighted the efficacy of Densenet in analyzing CXR images compared to other commonly used models [28, 138, 147, 148]. This justifies the rationale behind using the Dense-Net architecture [28] as the backbone of our own model.

The pretrained Dense-Net performs very well for most of the pathologies in the literature. However, the model cannot replicate that level of results on the ARDS dataset. As seen in Table 5.2, it reaches 80.3% AUROC on an image-level classification, which may not be reliable enough in clinics. This is not surprising since the most similar label to ARDS in Chest-XRay14 is "infiltration", whose recognition rate is the lowest among all existing pathologies. Thus, DenseNet-121 does not guarantee a conclusive decision system for clinical usage.

We again use GradCam [128] to visualize the activation maps according to the classes predicted by DenseNet-121. Fig. 5.9 illustrates the heatmaps created by GradCam to localize the affected lung areas. We can see that the most active areas (in red) do not necessarily lie in the lung areas. Hence, even if the model predicts an accurate label, the result is not completely reliable since it can be based on irrelevant areas of the input image.

5.4.2 Lung segmentation performance

In terms of segmentation, Dense-Ynet performs well compared to segmentation backbone model used in this work. To assess the effectiveness of the proposed Dense-Ynet in lung segmentation, the model is evaluated on two common segmentation datasets, namely JSRT [120] and Montgomery [144]. Table 5.3 shows the segmentation results of the proposed Dense-Ynet and its baseline segmentation model, (Dense-Unet [8]) for three datasets. Both models perform well in lung segmentation for JSRT and Montgomery. These datasets contain images that are not pathological in terms of ARDS. Normally, images with infiltration are more challenging to segment since the lung boundaries can be partially blurred due to the disorder. This is where the proposed Dense-Ynet distinguishes itself by segmenting lung areas well in the ARDS dataset, while Dense-Unet cannot maintain its previous results.

Table 5.3 Comparative analysis of lung segmentation outcomes across public datasets and our ARDS dataset, reported at a 95% confidence interval. DS:Dataset.

Model	JSRT DS	Montgomery DS	ARDS DS
Dense-Unet	97.4% (96.5-98.1)	97.9% (97.2-98.4)	88.9% (85.4-90.6)
Dense-Ynet	96.2% (95.2-97.0)	96.7% (95.7-97.4)	94.1% (92.3-95.3)

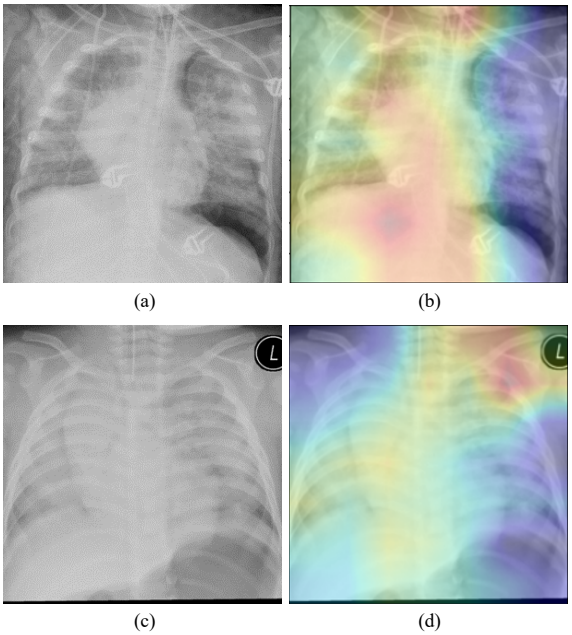


Figure 5.9 Generated heatmaps for ARDS identification by DenseNet-121 using the GradCam method. The left column contains original images and the right column shows the heatmaps. Both images are accurately classified as ARDS but the activation areas are not precise. Most activated areas are shown in red color.

5.4.3 Impact of lung segmentation on ARDS classification

In Section 5.4.2, we observed that although the DenseNet-121 model shows the best performance among the popular models, it does not provide sufficient reliability. Here, we evaluate the effect of a serial segmentation-classification procedure. As previously noted, radiologists focus on the lung fields as the region of interest and search for relevant signs in those areas. To emulate that workflow, a lung segmentation model can be used to identify the lung areas in CXR images prior to classification. For that purpose, we use the Dense-Unet [8] model to extract lung fields since that model performs well in lung segmentation in CXR images. Fig. 5.10 shows the overview of applying the segmentation and classification tasks in series. The segmentation models are pre-trained on the Montgomery dataset [144].

The input of the classification model is masked with the segmented binary image derived from the segmentation model. In this way, we limit the effective areas and facilitate the classification task. Table 5.4 confirms that using a segmentation model to mask CXR images before feeding the classification facilitates ARDS diagnosis. Therefore, if one can extract the lung fields in CXR images and remove background information, the pulmonary disease classification is likely to be more reliable.

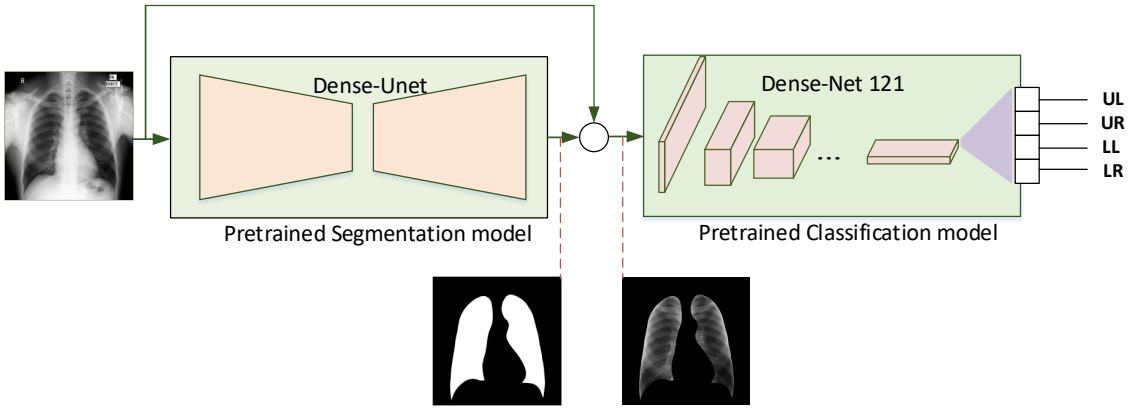


Figure 5.10 Lung masking before classification using pre-trained segmentation model.

Teixeira et al. [62] pointed out that establishing segmentation ahead of the classification increases the reliability of the diagnosis system even though the performance stays unchanged. However, boosting the classification performance is obviously another aim in our context.

Table 5.4 Impact of lung segmentation on the performance of the DenseNet-121 model in ARDS diagnosis, as assessed by the AUROC, with results presented at a 95% confidence interval.

Model	Without Segmentation	With Segmentation
DenseNet-121	80.3% (77.4-83.6)	82.5% (79.7-84.5)

Based on these observations, we conclude that the state-of-the-art Dense-Net121 might not perform confidently for the ARDS classification task. Besides, lung field segmentation prior to the classification (even if not the best segmentation performance) improves the reliability and boosts the detection performance. Indeed, by not considering the background, the algorithm omits the biases imposed on the task by non-specifying conditions like device-related noise and patient position. Qualitatively, GradCam maps show that a serial use of segmentation and classification improves the reliability of the predictions (see Fig. 5.11).

5.4.4 Discussion

We focus on ARDS within the context of pulmonary abnormalities, utilizing lung segmentation and subsequent classification on extracted RoI from CXR images. Lung segmentation is crucial for CXR diseases and is often performed using models like U-Net. However, U-Net's

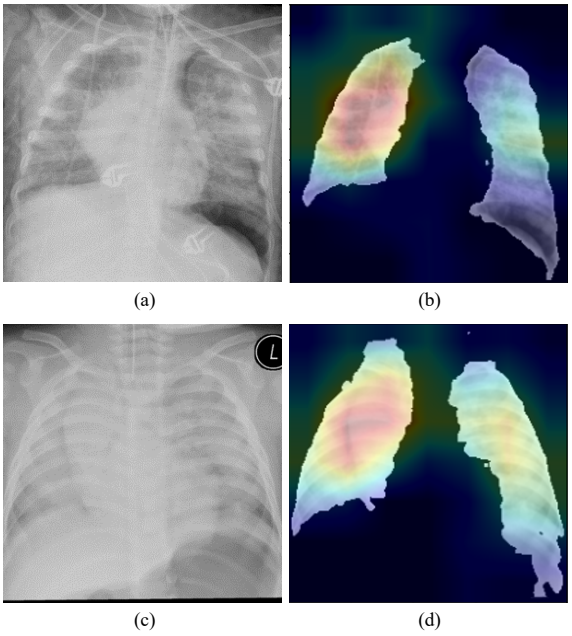


Figure 5.11 Activation maps for ARDS identification by DenseNet-121 with prior lung segmentation. Left column shows original images; right column shows heatmaps superimposed on masked images. Lung segmentation restricts the feature extraction areas; thus, GradCam points out more accurate areas in comparison with Fig. 5.9. We consider red zones as activated areas as signs of ARDS.

high-level feature maps can become excessively redundant, leading to over-processing and, in worse scenarios, challenges such as overfitting. To address this issue, we have reduced the depth of the feature maps for lung segmentation. This approach creates smaller, more efficient models suitable for clinical applications. Additionally, the sequential application of segmentation and classification introduces error propagation and can hinder classification, particularly in ARDS cases where lung boundaries are indistinct in CXR images. Taking inspiration from the Y-Net architecture, we have combined the segmentation and classification tasks into a Y-shaped model. The challenge of gradient vanishing, often encountered in deep models like U-Net, is mitigated in our design by utilizing DenseNet's direct information flow between layers. Our proposed Dense-Ynet model incorporates this connectivity to enhance information transfer while simultaneously reducing feature map depth. This model further divides the lung into quadrants for focused classification, aiding in ARDS detection and severity assessment. Given the limited availability of labeled ARDS datasets, we have developed a lightweight model to prevent overfitting. The increased information flow facilitated by dense connectivity in the Dense-Ynet architecture allows for a model design with fewer training parameters. Furthermore, our approach of performing ARDS classification at the quadrant level not only outperforms direct image-level classification but also supplies references for the final reported labels. In medical applications, these references are essential for validating decisions, offering transparency and a rationale for the conclusions drawn. This transparency also extends to the derivation of the Berlin criteria for defining ARDS, which is based on the local labels provided by our model.

5.5 Conclusions

A novel joint segmentation-classification network was proposed to identify accurate lung fields and discover ARDS cases in CXR images. Our DenseYnet outperforms the state-of-the-art method for classifying ARDS. ARDS causes bi-lateral diffuse infiltrations in the lungs, and its severity is directly associated with the extent of infiltration throughout the lungs. By classifying the lung quadrants instead of the entire image, we deal with a multi-label classification problem. Comparing quadrant-based and image-level classification, the quadrant-based classification results in better performances. Presenting labels for the four lung quadrants means the clinician can review the reference for the algorithm's decision. Interpretability is a crucial aspect of such clinical assistance systems since it makes the framework more trustable, in this case by providing local information to support the decision. Although clinical trials will be necessary to validate the model, the present study demonstrates the effectiveness of Dense-Ynet. By warning intensivists to ARDS cases, the model can play an important role in

the patients' survival since current explainable algorithms do not work well on CXR images.

Declaration of competing interest

The authors declare that they have no relevant conflicts of interest.

Data availability statement

A request for access to the data and implementation code can be made to Dr. Philippe Jouvet. The provision of code and data will be subject to the regulations established by the institutional review board governing the ethical oversight of research activities.

Acknowledgments

The authors sincerely acknowledge Philippe Debanné for revising this manuscript. This study was funded by the Artificial Intelligence Research and Transfer Institute (IVADO), the Ministry of Health of Quebec and Ste-Justine Hospital. The authors are also financially supported by Fonds de Recherche en Santé du Québec (FRQS).

CHAPTER 6 ARTICLE 3: A WEB-BASED PLATFORM FOR THE AUTOMATIC STRATIFICATION OF ARDS SEVERITY

Jounal Name: Diagnostics
Submission Date: January 15, 2023
Publication Date: March 1, 2023

Authors:
Mohammad Yahyatabar; Philippe Jouvet; Donatien Fily; Jérome Rambaud; Michaël Levy; Robinder G Khemani; Farida Cheriet

Abstract

Acute respiratory distress syndrome (ARDS), including severe pulmonary COVID infection, is associated with a high mortality rate. It is crucial to detect ARDS early, as a late diagnosis may lead to serious complications in treatment. One of the challenges in ARDS diagnosis is chest X-Ray (CXR) interpretation. ARDS causes diffuse infiltrates through the lungs that must be identified using chest radiography. In this paper, we present a web-based platform leveraging artificial intelligence (AI) to automatically assess pediatric ARDS (PARDS) using CXR images. Our system computes a severity score to identify and grade ARDS in CXR images. Moreover, the platform provides an image highlighting the lung fields, which can be utilized for prospective AI-based systems. A deep learning (DL) approach is employed to analyze the input data. A novel DL model, named Dense-Ynet, is trained using a CXR dataset in which clinical specialists previously labeled the two halves (upper and lower) of each lung. The assessment results show that our platform achieves a recall rate of 95.25% and a precision of 88.02%. The web platform, named PARDS-CxR, assigns severity scores to input CXR images that are compatible with current definitions of ARDS and PARDS. Once it has undergone external validation, PARDS-CxR will serve as an essential component in a clinical AI framework for diagnosing ARDS.

6.1 Introduction

Acute respiratory distress syndrome (ARDS) is a severe, even life-threatening condition, associated with respiratory failure, i.e., the inability of the lungs to fulfill their basic function of exchanging gases in the body. ARDS occurs in children and adults; its main causes include

respiratory infection, aspiration, or trauma. The first description of ARDS as a separate disease was provided in 1967. Variability in the ability to identify ARDS causes difficulty in clinical trials. The Berlin definition introduced diagnostic criteria, such as acute onset, severe hypoxemia (lack of oxygen in the blood), bilateral diffuse infiltrates visible in chest radiography, and absence of any evidence of cardiac failure or fluid overload [1]. Despite intensive studies investigating ARDS (60,000+ articles found in PubMed), its mortality rate is still as high as 43% [149]. Among the survivors of ARDS, a significant portion experienced lasting damage to the lungs, especially in older patients. The Berlin definition grades the severity of ARDS as being mild, moderate, or severe. Table 6.1 illustrates the oxygenation criteria and mortality rates associated with these severity levels.

Table 6.1 ARDS severities in the Berlin definition and associated oxygenation levels and mortality rates [1].

Severity	**PaO_2/FiO_2**	**Mortality**
Mild	$200 - 300$	27%
Moderate	$100 - 200$	32%
Severe	≤ 100	45%

As seen in Table 6.1, considering the high mortality rate of ARDS and its rapid progression, early diagnosis of ARDS is vital. Furthermore, the mortality rate is directly associated with the severity of the syndrome. The risk benefit profile of therapies depends on ARDS severity, making early stratification of ARDS severity crucial for management. The Pediatric Acute Lung Injury Consensus Conferences (PALICC) [14–16] were organized to address pediatric ARDS (PARDS) specifications and give treatment and diagnosis recommendations. According to the most recent definition of PARDS, PALICC-2 [16], the criteria allow for new infiltrates in chest radiography, even if only a region within a single lung is affected. One of the main reasons for this change in diagnostic criteria was the lack of agreement in the interpretation of chest images between radiologists or between radiologists and intensive care practitioners on the presence of bilateral infiltrates, which are required in the Berlin standard. López-Fernández et al. showed that interobserver agreement for bilateral infiltrates and quadrants of consolidation in PARDS was "slight" (kappa 0.31 and 0.33) [143]. Sjoding et al. reported similar results, with interobserver reliability of ARDS diagnosis being "moderate" (kappa = 0.50; 95CI, 0.40–0.59). Hence, there is an urgent need to improve the reliability of Chest X-Ray (CXR) intepretation in ARDS and PARDS to allow earlier diagnosis of the syndrome [150].

Several studies have applied machine learning (ML) and artificial intelligence (AI) approaches to analyze CXR images. One of the most common tasks reported in the literature is diag-

nosing pulmonary pathologies using chest radiography. Thanks to massive publicly available datasets, deep learning (DL) approaches have been broadly applied in medical pathology detection. However, there is as yet no dataset annotated with ARDS labels. Thus, few studies are found in the literature addressing the diagnosis of the syndrome.

To our knowledge, two papers present ML-based systems to identify ARDS in CXR images. The first one [26] proposed a method for detecting ARDS using a traditional ML approach based on hand-crafted features. The texture of intercostal image regions is considered as a discriminative feature for classifying samples. To highlight intercostal areas, a semi-automatic approach proposed by Plourde is utilized [25]. They succeed in reducing the inter-observer variability between clinicians in diagnosing PARDS. However, their approach is not automatic, and the rib segmentation step requires operator intervention. In the second work, an automatic ARDS detection and grading approach was proposed using a state-of-the-art DL model (Densenet) [28]. The authors first pretrained the model on public datasets (not containing ARDS samples) and then refined the model with a custom dataset consisting of ARDS-labeled images. Their approach performs well in diagnosing ARDS, but the model provides no evidence for the support system's decisions. Thus, although it works well in analyzing ARDS cases, the model lacks interpretability, which is essential for an ML system to be used in clinical settings.

Recently, due to the COVID-19 outbreak, the research community has gotten more involved in computer-based analysis of chest X-ray images as one of the easiest and fastest ways to check for signs of the disease. Mobile Chest X-Ray Analysis [113] and Chester [112] are prototype systems for CXR assessment developed using the aforementioned Densenet model, trained on the public Chest-XRay14 dataset [2]. Both systems provide evidence for the detected pathologies by means of saliency maps obtained using GradCam [128]. However, this can reveal areas that are irrelevant to the pathology being detected [129, 130]. Thus, although these systems provide activation maps pointing out the references for the decisions, they are not sufficiently reliable to be used in clinics.

The main contributions of this paper are to create a tool for stratifying the severity of ARDS in CXR images and to build a web-based platform for external validation. The platform uses local information to classify X-rays based on the distribution of infiltrates in the different lung quadrants, and it provides a global severity score for the image that is applicable in both children and adults. The web-based platform, PARDS-CxR, can be used as a standalone tool, or it can be integrated with other ARDS analysis tools to offer a comprehensive approach for clinical use.

The following section first explains the details of the data collection used to train our DL

model. Then, we describe the proposed DL model and its evaluation process, and we present the development of the Web platform. Section 6.3 presents the results of testing the ARDS assessment tool, and in Section 6.4, the strengths and drawbacks of our platform are discussed. Section 6.5 provides concluding remarks for this paper.

6.2 Materials and Methods

6.2.1 Methodology

This study contains four main phases, as illustrated in Figure 6.1. The end product is PARDS-CxR, the web-based application to detect ARDS. First, a substantial set of data is required to train the model. Existing public datasets do not include ARDS-labeled CXR images, so we created a new one. This data collection process is summarized in Section 6.2.2. Then, the proposed model must be trained on the CXR images. The model has two outputs associated with lung segmentation and ARDS classification, as explained in Section 6.2.3. The trained model is then tested on unseen data to be evaluated. Sections 6.2.4–6.2.6 detail the validation process. Finally, the model is uploaded to a server, and an interface is designed so the user can easily access it. The web application is addressed in Section 6.2.7.

6.2.2 ARDS Dataset

Collectively, the main publicly available CXR datasets provide around a million images with pathology labels [142]. This data motivated many researchers to employ AI techniques in this domain. However, no such datasets assign ARDS-specific labels to images. As our first step, we collected and annotated a dataset at Sainte-Justine Hospital, Montreal, Canada (CHUSJ), to address the lack of appropriate data. Our dataset comprises three data sources containing 373 CXR images. Ninety images and their corresponding labels came from a previous study by our team [26]. A further 100 images were taken from the Chest X-Ray14 dataset [2] and relabeled by clinical experts (JR, ML) in the hospital. Another 183 images were provided by the PARDIE study, a multi-national study that prospectively gathered chest X-ray images of children with ARDS [143]. For each image, labels were associated with the four lung quadrants obtained by splitting each lung into upper and lower portions. We refer to each quadrant by its position: left upper (LR), left lower (LL), right upper (RU), and right lower (RL). According to the Berlin definition [1], visible bilateral infiltrates are a mandatory criterion for a case to be categorized as ARDS. Two intensivists from CHUSJ assessed the presence of infiltrates in each quadrant. A sample was included in the dataset only if the clinical observers reached a consensus on the labels. In addition, 138 CXR images were taken

from the Montgomery dataset to represent the normal class [144]. These samples were labeled as non-ARDS if agreed by the clinical experts. By dropping images with disagreements in their labeling, our final ARDS dataset consisted of 356 images, of which 134 meet the bilateral infiltrates criteria in the Berlin definition [1].

6.2.3 Joint Segmentation and Classification Model

In computer-based diagnosis approaches, it is common to use segmentation ahead of classification to determine the region of interest. Lung segmentation separates the lung areas from the thoracic tissues surrounding them and is the primary image analysis step in many clinical decision support systems. Generalization to new datasets is a difficult challenge in the analysis of chest radiography. In that respect, segmentation is considered a strategy to limit the impact of specific imaging devices and settings, since it restricts the feature extraction to the lung fields and removes the effect of the image background [62, 63]. However, serial usage of segmentation and classification propagates the segmentation error into the classification network. Dense-Ynet is a convolutional network that takes advantage of Densenet, Y-net, and U-net models to do both tasks simultaneously in a joint segmentation–classification model. The backbone of the network used in this study is our previously developed Dense-Unet [8]. Dense-Unet is a segmentation model in which dense connections between the feature maps in various layers facilitate the information flow throughout the model, letting designers choose a configuration with a small number of training parameters. Our proposed Dense-Ynet takes advantage of automatic feature extraction from both the original and segmented images (Figure 6.2). The model has two outputs and is trained using two loss functions: the lung segmentation loss and quadrant classification loss. The model works based on the convolution operation. A convolution is a mathematical operation that filters the information of its input and creates feature maps. An inevitable effect of the convolution operation is to change the dimensions of the feature maps. To tackle this issue, upsampling and strided convolution operations are used to ensure that feature maps coming from different layers can be concatenated. Squeeze and excitation (SE) blocks [123] are also used after each convolution layer to improve the representational power of the blocks by recalibrating the features. The key strengths of Dense-Ynet are use of lung segmentation in its architecture, specialized connectivity, which enable better generalization, and prediction of local labels for each image.

To reach the final decision based on the Berlin definition, we must test for existing bilateral infiltrates. To that end, a simple logical operation in Equation (6.1) combines the predictions of each quadrant to check this condition:

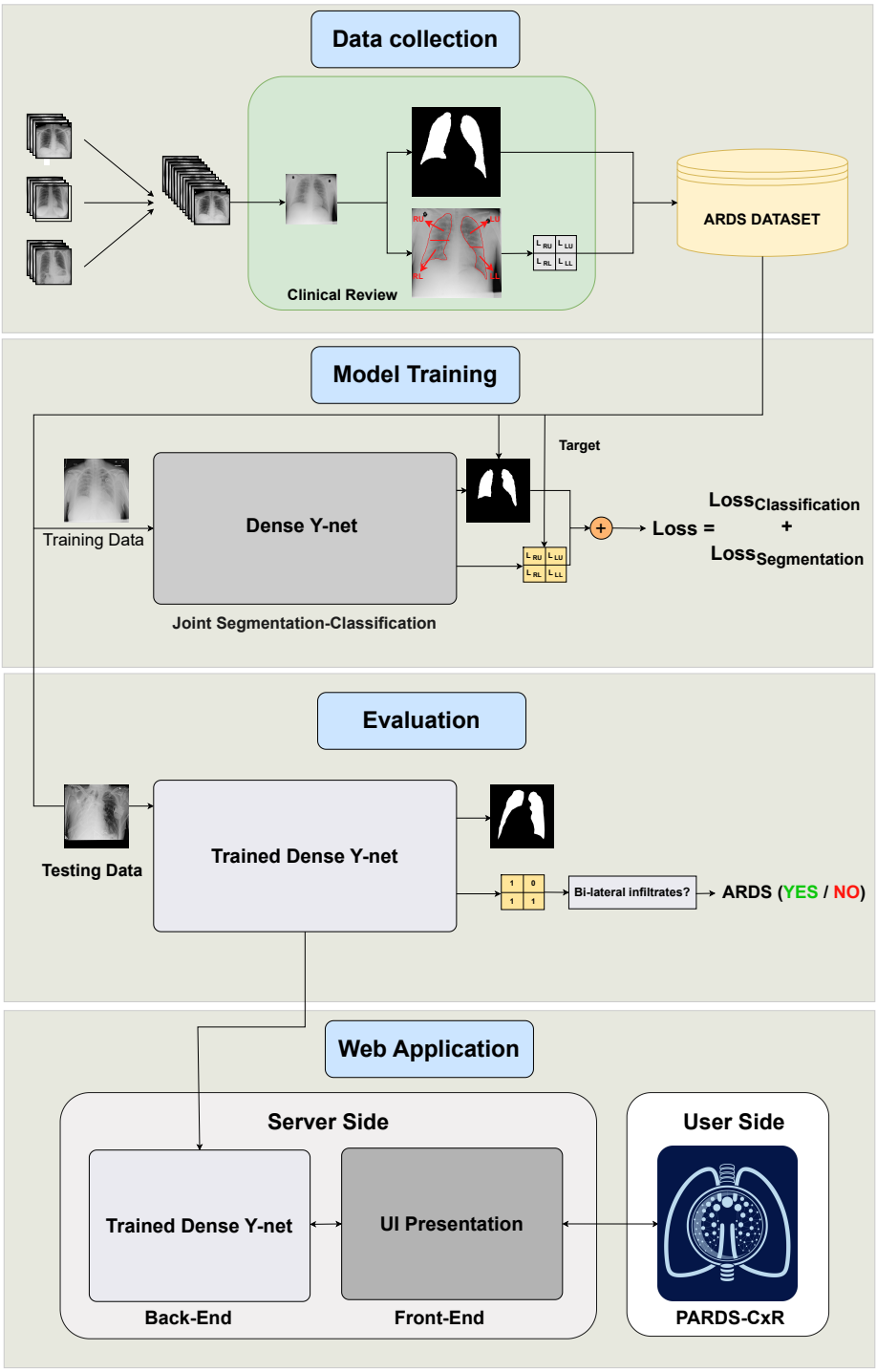


Figure 6.1 Organization of our study into four main phases. The data are collected from several data sources and annotated at Saint-Justine Hospital, Montreal, Canada. The DL model is trained using quadrant-level labels and lung segmentation maps. It is then evaluated on a set of previously unseen images; both the classification and segmentation performances are assessed. Finally, a web-based platform is designed and made available through the internet.

$$P_{ARDS} = (P_{RL} \vee P_{RU}) \wedge (P_{LL} \vee P_{LU}) \tag{6.1}$$

P_{RL}, P_{RU}, P_{LL}, and P_{LU} are the prediction labels for the right lower, right upper, left lower, and left upper quadrants, respectively. P_{ARDS} is the inferred ARDS label, and $\vee$ and $\wedge$ are logical *or* and *and* operations. The equation states that, if at least one quadrant is involved on each side, the case is recognized as (P)ARDS.

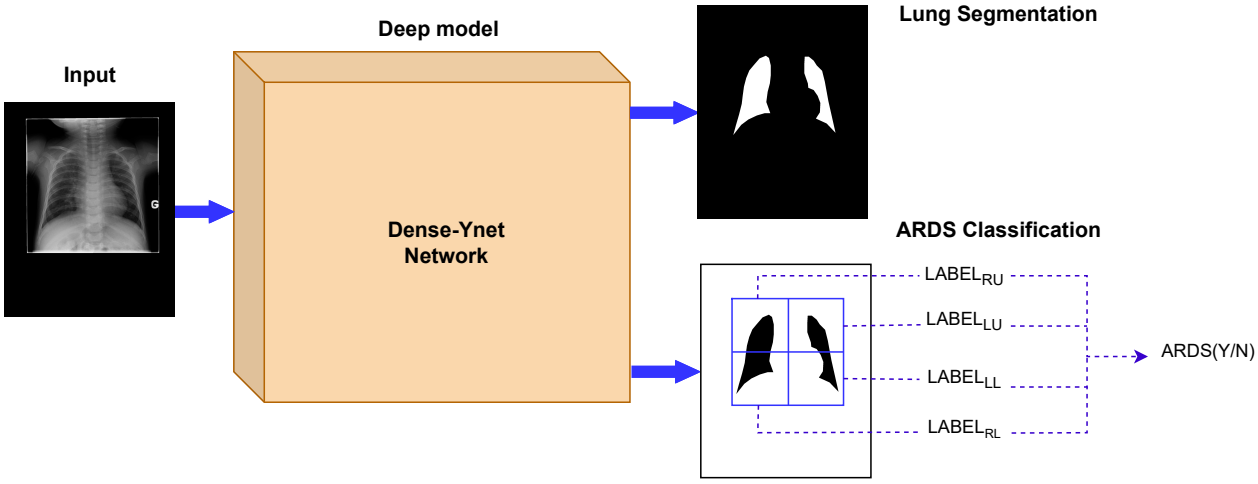


Figure 6.2 The Dense-Ynet model takes advantage of the interaction between the segmentation and classification tasks by performing them simultaneously. The features from the original and lung-segmented images are concatenated and utilized to classify ARDS cases.

6.2.4 Experimental Design

In this work, 267 images of the ARDS dataset are used to train the Dense-Ynet model. In addition, 35 images are used to validate the training process. For the testing stage, 54 images previously unseen by the network are used. The algorithm is evaluated with the five-fold cross-validation strategy. Cross-validation is a method that tries various training and testing data combinations to confirm the reported results' reliability. Data augmentation is a technique to enrich the training data by generating new images from the current training set. For this purpose, we use basic image processing techniques, such as random rotation, cropping, shifting, horizontal flipping, and intensity changing. The rectified linear unit ($ReLu$) activation function introduces non-linearity to network blocks. The *Sigmoid* function provides valid labels between zero and one in both the segmentation and classification output layers. *Adam* is the optimizer used for updating the model weights during training. To reach the optimal configuration, a set of hyperparameters must be explored to find the best model

structure and training policy. The Web platform (see Section 6.2.7) employs six Dense-Ynet instances, corresponding to the best hyper-parameters sets. Using an ensemble approach, the final result presented to the user combines the values received from the individual models.

The PARDS-CxR application detects lung quadrants consolidation, and the final ARDS label is derived from the quadrant predictions using Equation (6.1).

6.2.5 Scoring Scheme

To analyze the severity of ARDS in CXR images, a scoring scheme is proposed based on the number and the position of affected lung quadrants (see Table 6.2). The scheme is compatible with the Berlin definition, in which existing bilateral infiltrates are an essential criterion for ARDS diagnosis in chest radiography.

Table 6.2 Severity scoring scheme based on affected lung quadrants.

Affected Quadrants	**Score**	**Severity**
4 quadrants	5	Severe
3 quadrants	4	
2 quadrants (Different sides)	3	Mild
2 quadrants (Same side)	2	Non-ARDS
1 quadrant	1	
No affected quadrant	0	

Giving scores is important from two points of view. First, the score represents the severity of the diffused infiltrates throughout the lungs. Second, reporting disease severity helps clinicians follow appropriate treatment protocols or triaging. This type of system has been proposed for the Murray Lung Injury Score, as well as as part of the recently proposed RALE score in adult patients with ARDS.

6.2.6 Evaluation Metrics

Evaluation metrics are measured from the algorithm's performance on unseen test data to assess the approach. There is no metric representing the total capacity of the PARDS-CxR platform. However, we use a set of performance metrics to provide a complete overview of the model's operation. A confusion matrix quantifies the ability of the classifier to detect each class separately. It gives detailed measures comparing the actual and predicted labels, as shown in Figure 6.3.

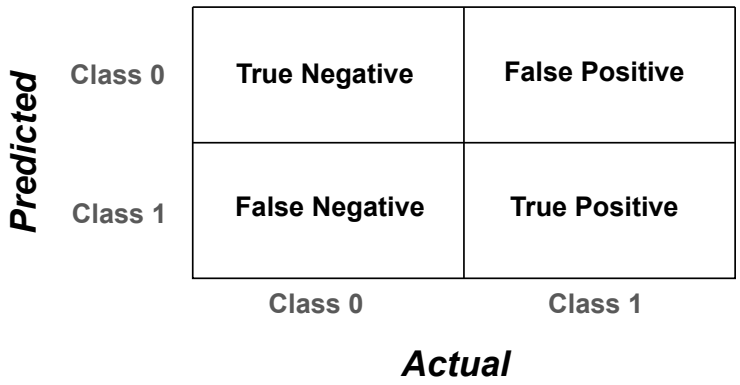


Figure 6.3 Confusion matrix for a binary classification problem. The matrix contains four elements that, together, evaluate the system's predictions versus the real labels.

The elements of the confusion matrix, namely, the true positive (TP), true negative (TN), false positive (FP), and false negative (FN) values, serve to calculate several assessment metrics as follows:

$$Accuracy = \frac{TP + TN}{FP + FN + TP + TN} \tag{6.2}$$

$$Precision = \frac{TP}{FP + TP} \tag{6.3}$$

$$Recall = \frac{TP}{FN + TP} \tag{6.4}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}. \tag{6.5}$$

The *Accuracy* metric represents the overall correctness of a classification algorithm. It cannot fully express the model performance, however, especially in the case of unbalanced testing data. *Precision* and *Recall* reveal the model's performance in discriminating between the different classes. *Precision* represents how precise the model is in identifying the target (positive) class. Specifically, it points out what portion of cases predicted as positive are really ARDS cases. On the other hand, the *Recall* value shows what proportion of predicted ARDS cases are actually labeled as ARDS. These two metrics have a complementary role in describing the model's behavior. The $F1$ score, derived from *Precision* and *Recall* values, is a single metric to quantify the algorithm's performance.

The receiver operating characteristic (ROC) curve illustrates the diagnostic capacity of a system by comparing true positive and false positive rates as the discrimination threshold (applied at the network's output layer to decide between the two classes) varies. The area under the ROC curve (AUROC) represents the discriminatory power of the classifier.

6.2.7 Web-Based Platform

We designed a web-based platform to facilitate the diagnosis of ARDS in CXR images by medical professionals. The platform is intended as a tool to provide a second opinion to clinicians, but no direct medical use is recommended until medical professionals validate the tool using external data. The PARDS-CxR platform takes advantage of six Dense-Ynet instances to provide scores for each input image. The scores are given based on the number and combination of affected lung quadrants as explained in Section 6.2.5. A global score is assigned by combining the outputs from the model instances. In addition, the application provides accurate lung segmentation maps, which are helpful in AI-based analysis of CXR images.

The web application utilizes the *React* library to create a user-friendly and interactive user interface (UI) for delivering the specified services. The library enables efficient code writing and makes it easier to manage, refine, and integrate the application with other tools. The platform supports both English and French languages and has two main modes for ARDS definitions for adults (Berlin) and children (PALICC-2). The difference between the modes is that, when using PALICC-2 mode, the platform requires two input images. The application response includes segmentation maps, severity scores (local and global), and an interpretation based on the definition.

Although the deep models are trained using graphical processing units (GPUs), the evaluation model does not require a GPU and can process the results in 2-3 s. Thus, the running bottleneck could be the network connection speed. The application is capable of storing data and providing log files, but this feature is currently disabled and will be activated when the validation protocol is approved. The PARDS-CxR platform is detailed further in Section 6.3.3.

6.3 Results

6.3.1 Quadrant-Based Classification

The PARDS-CxR web-based platform uses Dense-Ynet as the joint segmentation and classification model. In classification, the model predicts four labels associated with lung quadrants, as explained in Section 6.2.3. The platform uses an ensemble of six Dense-Ynet model instances with different training and model structure configurations. Regarding model structures, we experimented with different channel depths in convolution blocks, loss functions, weights for merging loss functions, activation functions, and initial network weights. For the training configurations, we varied several hyperparameters, namely, the learning rate, training batch size, augmentation probability, and stopping criterion.

Figure 6.4 shows the confusion matrix of the ensemble of models. To merge the results from the model instances, a hard voting strategy is employed based on the labels predicted independently by the models. To be precise, each model is trained separately with its specific configuration. The testing is also done independently, and if at least three models decide that an image is an ARDS case, the combined result is positive. By combining models with various configurations, the intrinsic biases of each one to accept or reject an image as ARDS are balanced in the ensemble output. Thus, the final performance improves compared to any individual model.

Table 6.3 compares the classification performances of the Dense-Ynet instances in terms of the four assessment metrics seen previously. Some of the listed models achieve higher precision, while others reach better recall values. By combining the predicted labels provided by these models, the ensemble algorithm achieves the highest $F1$ score, representing the best compromise between precision and recall. Indeed, ensembling the models does not outperform every one in terms of *Precision* and *Recall*, but the final $F1$ and accuracy values improve.

In this paper, the problem of ARDS diagnosis is based on the classification of lung quadrants. Thus, the task can also be considered as a multi-label classification problem. Figure 6.5 shows the ROC curves of all quadrants' predictions for the Dense-Ynet instances, i.e., the ROC curves associated with the binary classification of the lung quadrants, regardless of their positions. The AUROC metric is not directly related to the system's performance in ARDS diagnosis, but the misclassification of one lung quadrant may cause an error in classifying the image as a whole.

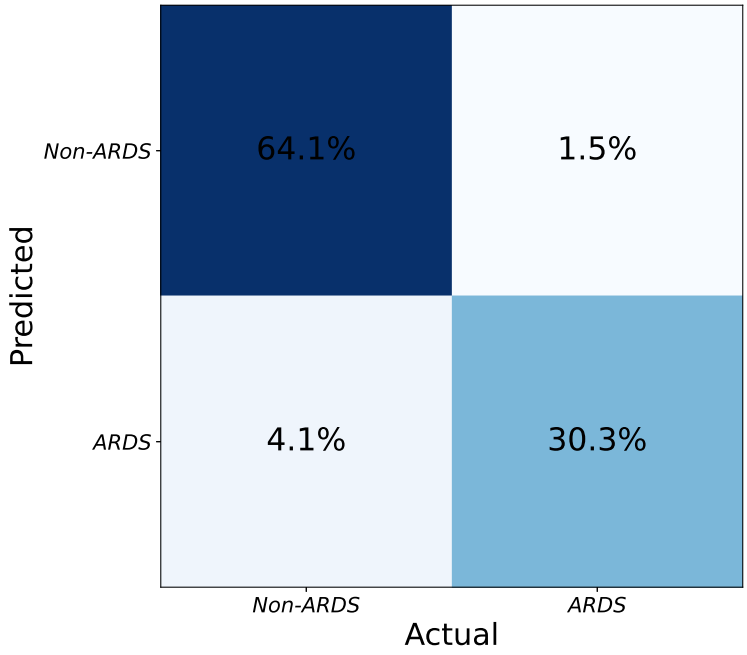


Figure 6.4 Final confusion matrix obtained from the combination of network instances using hard voting. The numbers (percentages) are obtained by taking the average of several tests (five-fold cross-validation).

Table 6.3 Evaluation of the six models and the result of their combination (ensemble model) for classification.

Model	**Accuracy**	**Recall**	**Precision**	***F*1**
Network 1	92.95%	88.45%	91.99%	90.19%
Network 2	93.54%	96.41%	84.37%	89.99%
Network 3	92.04%	94.42%	87.89%	91.03%
Network 4	92.96%	100.0%	83.33%	90.91%
Network 5	87.32%	100.0%	74.29%	85.25%
Network 6	88.74%	80.01%	80.01%	80.02%
Ensemble model	94.35%	95.25%	88.02%	91.49%

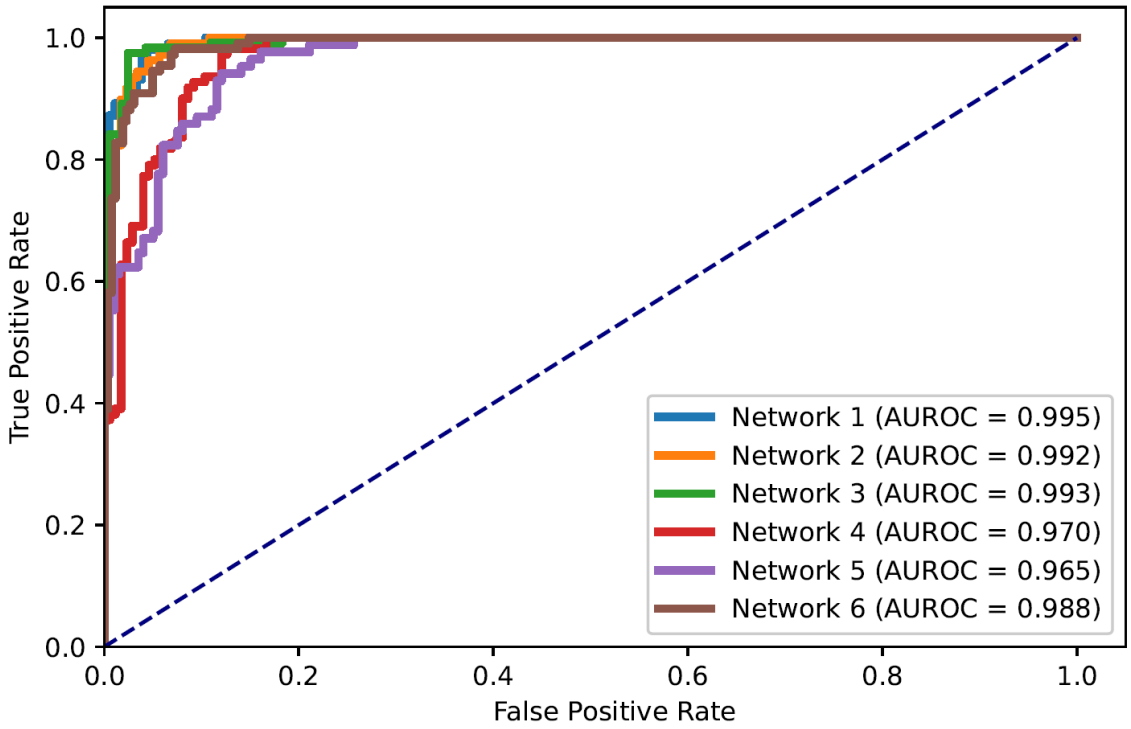


Figure 6.5 ROC curves for classification of lung quadrants regardless of their position in the lungs.

6.3.2 ARDS Severity Prediction

As seen in Table 6.2, the application determines the severity of ARDS in CXR images based on the number and combination of affected lung quadrants. The platform provides a global score for each input image by taking the average of the scores from each model. CXR images are then categorized into one of three severity grades based on the predicted scores: non-ARDS, mild ARDS, and severe ARDS. The platform's effectiveness in determining ARDS severity is illustrated in Figure 6.6. The three-class confusion matrix shows that the approach can detect ARDS and discriminate between mild and severe states of the syndrome.

6.3.3 PARDS-CxR, the Web-Based Platform

Our web application is currently loaded on a web server at CHUSJ and is accessible at the address (`https://chestxray-reader.chusj-sip-ia.ca`, accessed on 15 January 2023). The process of training and testing the deep model was programmed in Python using the PyTorch library [151]. The training process and hyperparameter search were executed on GPU, as they required intensive parallel computing. The trained model was then transferred to CPU to evaluate new images; thus, no graphical processor is necessary on the server to

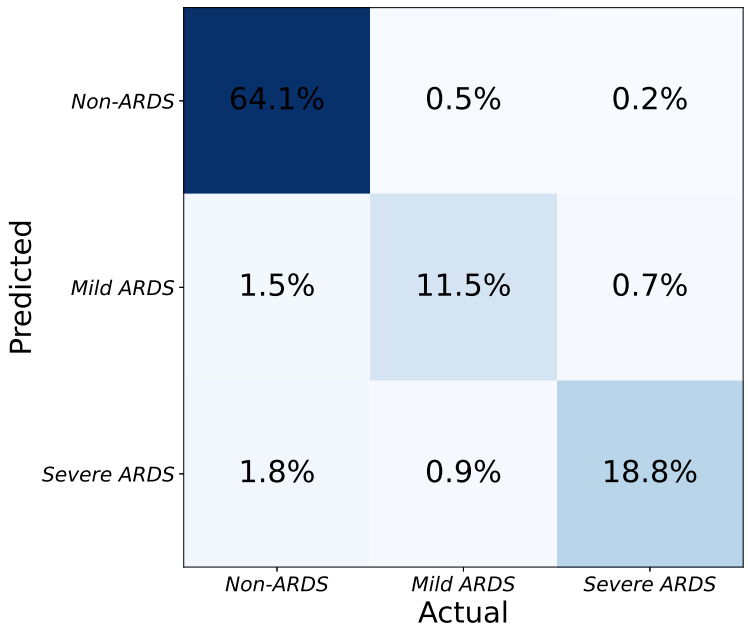


Figure 6.6 Confusion matrix for classification of ARDS severity with three levels (none, mild, severe).

run the application. The graphical user interface was written in JavaScript and is compatible with various internet browsers on the client side. No data are kept on the server side, and the application output image is available to store in the user's local storage. The user interface works in English and French, and CXR images can be uploaded using the menu option or drag-and-drop (see Figure 6.7).

The application bases itself on the most accepted definitions for ARDS and PARDS. Based on the Berlin definition, the presence of bilateral infiltrates in chest radiography is a criterion manifesting the existence of ARDS [1]. The platform processes the image and displays its decision by providing a percentage associated with the level of infiltration in each quadrant (Figure 6.7). A global percentage is also given based on infiltrate levels of infiltrates in quadrants and their combination as in Table 6.2. This value represents the severity of ARDS in the input image. An image with a global percentage above 60% is interpreted as an ARDS case, since, based on the proposed severity scoring system, infiltrates should be diffused through both lungs. Reporting each quadrant's involvement is necessary, since it gives the rationale behind the global severity measure. As seen in Figure 6.7, a segmentation map highlighting the lung segments is also provided.

Identifying progression of ARDS is also possible, as two images taken at different times can be compared by the system. Additionally, an example of CXR image comparison is displayed

in Figure 6.8.

Figure 6.7 Main interface of the PARDS-CxR web application. In the standard mode, a single CXR image is analyzed according to the Berlin definition.

6.4 Discussion

The proposed DenseY-net is a joint segmentation–classification model that diagnoses (P)ARDS based on lung quadrant-level classification. The results show that the model can accurately classify quadrants and, consequently, the entire input image. This labeling strategy offers a reasoning framework for decision-making and incorporates an interpretability feature into the platform. Ensemble modeling is used to combine the outcomes from six model instances. PARDS-CxR can also do lung field segmentation, which is a necessary element in many decision support systems. Our approach performs well in detecting the severity of ARDS by giving a score to each input determined by the number and position

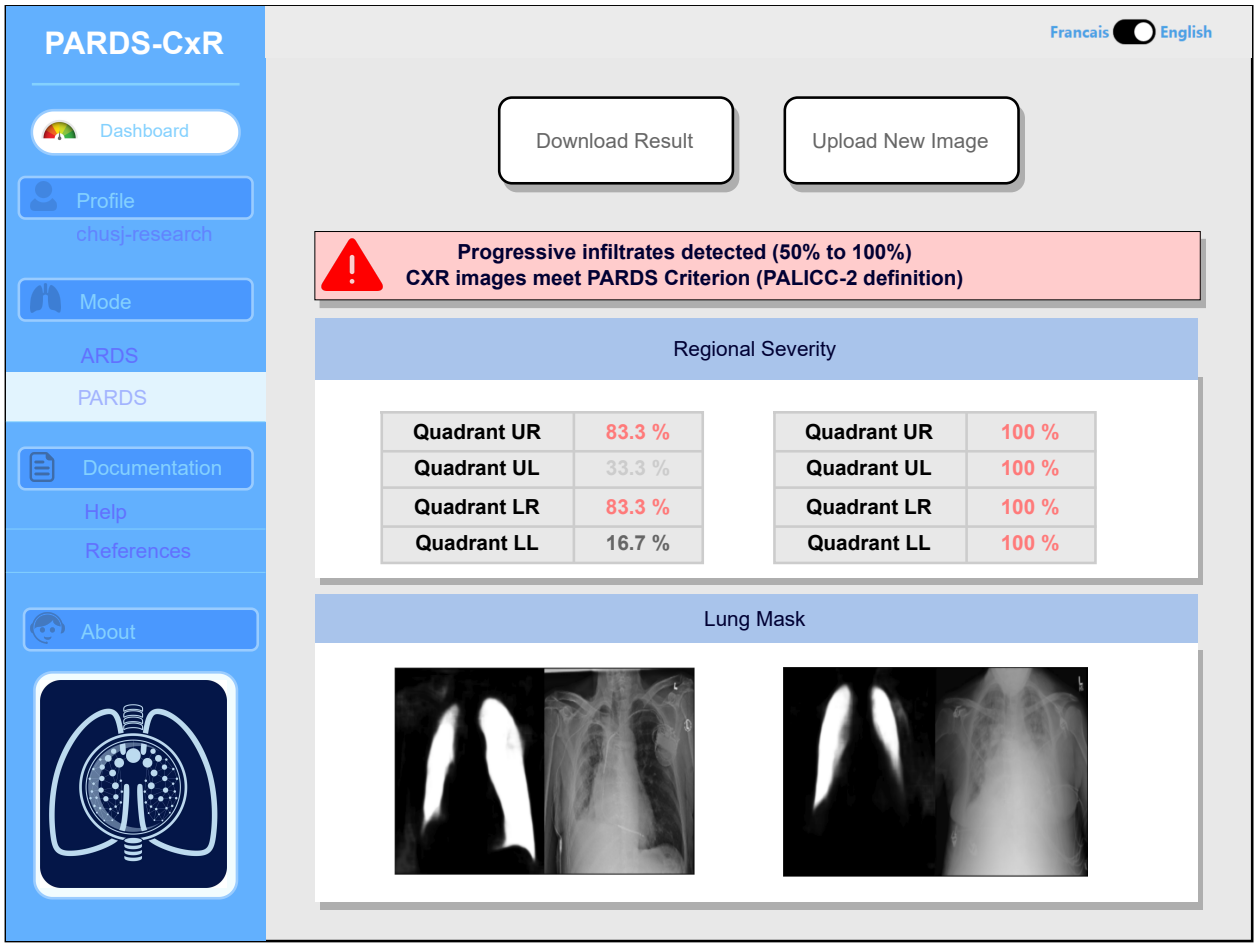


Figure 6.8 PARDS-CxR interface in image comparison mode. The platform can analyze two CXR images to detect ARDS progression based on the PALICC-2 definition.

of affected lung quadrants. This makes the model compatible with both ARDS and PARDS definitions.

A few large chest radiography datasets are publicly available for the research community [2, 3, 126]. A key benefit of deep learning is its capacity to analyze and learn features from a substantial amount of data. Therefore, it is unsurprising that several ML researchers have investigated CXR image analysis in various contexts. However, important limitations of these datasets make them unsuitable for developing dependable systems for the hospital setting. Indeed, most of the data are annotated using clinicians' notes processed by natural language processing (NLP) techniques [152]. This leads to erroneous labeling of a portion of the images. For example, a 10% error is reported for Chest X-Ray 14 [2], even though it is one of the most frequently used CXR datasets. The clinical review in [153] reveals an even higher rate of data labeling errors in that dataset.

Although adding some level of noise to the training inputs can improve a deep model's performance, biases and extensive labeling errors will decrease the model's accuracy. This could be a reason for the relatively poor generalization ability of deep models when confronting new samples from other data sources. Furthermore, available samples are annotated for a limited number of pathologies. Public CXR datasets cover between 14 and 18 chest pathologies, but these do not include ARDS or PARDS. To address this constraint, we collected our own CXR dataset from three different sources and annotated it for PARDS at CHUSJ. This dataset was labeled at the lung quadrant level, and the lung fields were manually identified in each image to establish a segmentation ground truth. The resulting dataset contains 356 CXR images, including 134 that meet the bilateral criteria for ARDS. Annotating data is costly in the clinical field, even more so considering that the DenseY-net model needs lung maps and quadrant-level ground-truth labels. Consequently, our ARDS dataset is relatively small. Nonetheless, our model is designed in such a way as to train adequately on small datasets. The specialized connectivity within the model allows for the creation of a lighter model with shallower intermediate feature maps, resulting in a smaller number of training parameters. A model with fewer parameters is more appropriate for training with small datasets. The algorithm was assessed on our own dataset, as explained in Section 6.2.4. A bigger dataset could increase the generalization capacity of the model ensemble. Moreover, external validation of the platform using data from various health centers will make it more reliable as a tool for prospective clinical research. Thus, as next steps in the web application's development, external validation and improving interpretability are two major points, since both are necessary to turn the platform into a practical tool in clinics.

Moreover, according to the (P)ARDS definition, co-occurrence of detectable infiltrates in

CXR and hypoxemia is necessary when no evidence of cardiogenic pulmonary edema is observed. Thus, although the presence of infiltrates in chest radiography is known as the most limiting factor for diagnosing ARDS, meeting other criteria is a requisite. The Clinical Decision Support System (CDSS) lab at CHUSJ has the capacity to investigate other ARDS diagnosis criteria, including cardiac failure and hypoxemia. Le et al. have employed NLP techniques and ML algorithms to detect cardiac failure in children [24]. Sauthier et al. have developed a method to accurately estimate Pao2 levels using noninvasive data [23]. Integrating the tool proposed in this study with other works will lead to a system giving comprehensive ARDS diagnoses. Sufficient electronic medical infrastructure is available in the PICU of CHUSJ to facilitate the flow of data from various sources [154]. By accessing data from clinical narrative analysis, measuring oxygenation indices, and detecting infiltrates in CXR images, it will be possible to make clinical decisions in real time. Therefore, an important objective for our team will be to implement an ARDS diagnosis package at CHUSJ, integrating all these criteria and data sources.

The strength of this study lies in the development of an algorithm that, in comparison to existing approaches, is more interpretable and automated and is compatible with existing ARDS definitions. Unlike an earlier ARDS diagnosis method proposed by our research team [26], the DL-based approach used in this application does not need any interaction from clinicians or operators to guide the algorithm. The novel model provides an end-to-end process that is simple for the user and provides the diagnotic outputs instantaneously. Recently, Sjoding et al. [28] proposed another automatic algorithm for detecting ARDS in CXR images. However, their approach lacks explainability, i.e., the system's decisions are not supported by further information. By contrast, since PARDS-CxR detects infiltrates in each lung quadrant, the basis for the decision is integral to our method. This strengthens the platform's reliability, since the user can reject or accept the decision by observing the delivered explanation. In addition, the proposed approach is compatible with both PARDS and ARDS definitions [1,14], as the scoring scheme used translates to a disease severity level. At present, the main limitation of our algorithm is its lack of external validation. Indeed, its development relied on a limited number of CXR images with a single team annotating them. For this reason, we have implemented the algorithm on a web platform to allow researchers to conduct validation studies.

6.5 Conclusions

This work has described a deep learning method and web-based platform for diagnosing acute respiratory distress syndrome (ARDS) from chest X-ray (CXR) images. The platform

uses an ensemble of novel Dense-Ynet networks that can accurately detect lung infiltrates in different quadrants and combine this information to detect ARDS and grade its severity. This approach ensures that our tool is compatible with various ARDS definitions in both adults and children. Following feedback from clinical researchers during a validation phase, the platform will be integrated into a complete clinical decision system for ARDS. The tool presented here will serve as the CXR analysis component within an AI-based framework that will monitor other factors, such as hypoxemia and occurence of cardiac arrest.

Author Contributions: F.C., M.Y., and R.K. conceptualized and designed the study. M.Y., P.J., F.C., D.F., M.L., and J.R. developed the study protocol. M.Y., D.F., M.L., and J.R. conducted the algorithm development. M.Y., F.C., and P.J. drafted the initial manuscript. All authors approved the final manuscript as submitted. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by grants from IVADO (Artificial Intelligence Research and Transfer Institute), the Quebec Ministry of Health and Sainte-Justine Hospital. M.Y. is financed by an award from the *Fonds de Recherche en Santé du Québec (FRQS)* Chair in Artificial Intelligence and Health. P.J. earns a research salary from FRQS.

Institutional Review Board Statement: The study was approved by the Institutional Review Board of Sainte-Justine Hospital (approval number: 2023-5124).

Informed Consent Statement: The study was carried out on a research database and the Institutional Review Board did not require informed consent.

Data Availability Statement: Access to data can be requested from Philippe Jouvet. Specific institutional review board rules will apply.

Acknowledgments: The authors gratefully acknowledge Philippe Debanné for his assistance in reviewing the manuscript. Furthermore, the authors thank all the investigators (pediatric intensivists/Radiologists) who participated in the PARDIE V3 study (Country, site and investigator list): *Argentina.* Hospital De Ninos Ricardo Gutierrez: Rossana Poterala; Hospital de Ninos sor Maria Ludovica: Pablo Castellani/Martin Giampieri/Claudia Pedraza; Hospital Nacional Prof. Alejandro Posadas: Nilda Agueda Vidal/Deheza Rosemary/Gonzalo Turon/Cecilia Monjes; Hospital Pediatrico Juan Pablo II: Segundo F. Espanol; Hospital Universitario Austral: Alejandro Siaba Serrate/Thomas Iolster/Silvio Torres; Sanatorio de Ninos de Rosario: Fernando Paziencia. *Australia.* Princess Margaret Hospital for Children: Simon Erickson/Samantha Barr/Sara Shea. Bolivia. Hospital del Nino Manuel Ascencio Villaroel: Alejandro F. Martinez Leon/Gustavo A. Guzman Rivera. *Canada.* CHU Sainte-Justine: Philippe Jouvet/Guillaume Emeriaud/Mariana Dumitrascu/Mary Ellen French. *Chile.* Hos-

pital Base de Valdivia: Daniel Caro I/Andrés A Retamal Caro; Hospital El Carmen de Maipu: Pablo Cruces Romero/Tania Medina; Hospital Luis Calvo Mackenna: Carlos Acuna; Hospital Padre Hurtado: Franco Diaz/Maria Jose Nunez. *China*. Children´s Hospital of Fudan Univ: Yang Chen. *Colombia*. Clinica Infantil de Colsubsidio: Rosalba Pardo Carrero; Hospital General de Medellin: Yurika P. Lopez Alarcon; Hospital Militar Central: Ledys María Izquierdo; Hospital Pablo Tobon Uribe (HPTU): Byron E. Piñeres Olave. *France*. CHU de Nantes: Pierre Bourgoin; Hopital d'enfants de Brabois–CHU de Nancy: Matthieu Maria. *Greece*. University of Crete, University Hospital PICU: George Briassoulis/Stavroula Ilia. *Italy*. Children's Hospital Bambino Gesu: Matteo Di Nardo/Fabrizio Chiusolo/Ilaria Erba/Orsola Gawronski; Children's Hospital Vittore Buzzi: Anna Camporesi. *Japan*. Hiroshima University: Nobuaki Shime/Shinichiro Ohshimo/Yoshiko Kida/Michihito Kyo. *Malaysia*. Universiti Kebangsaan Malaysia: Swee Fong Tang/Chian Wern Tai; University Malaya Medical Center: Lucy Chai See Lum/Ismail Elghuwael. *Mexico*. Hospital Espanol De Mexico: Nestor J. Jimenez Rivera. *Peru*. Hospital de Emergencias Pediatricas: Daniel Vasquez Miranda/Grimaldo Ramirez Cortez; Instituto Nacional de Salud del Nino: Jose Tantalean. *Portugal*. Hospital Santa Maria–Centro Hospitalar Lisboa Norte: Cristina Camilo. *Saudi Arabia*. King Abdullah Specialist Children's Hospital, King Abdulaziz Medical City: Tarek Hazwani/Nedaa Aldairi/Ahmed Al Amoudi/Ahmad Alahmadti. *Spain*. Cruces University Hospital: Yolanda Lopez Fernandez/Juan Ramon Valle/Lidia Martinez/Javier Pilar Orive; Hospital Regional Universitario de Malaga: Jose Manuel Gonzalez Gomez/Antonio Morales Martinez; Hospital Universitari I Politecnic La Fe: Vicent Modesto I Alapont; Sant Joan de Deu University Hospital: Marti Pons Odena; Hospital Universitario Central De Asturias: Alberto Medina; Virgen de la Arrixaca University Hospital: Susana Reyes Dominguez. *Turkey*. Akdeniz University School of Medicine: Oguz Dursun/Ebru Atike Ongun; Izmir Katip Celebi University Medical School and Tepecik Research and Training Hospital: Fulya Kamit Can/Ayse Berna Anil. *UK*. Evelina London Children's Hospital: Jon Lillie/Shane Tibby/Paul Wellman/Holly Belfield/Claire Lloyd; Great Ormond St. Children's Hospital: Joe Brierley/Troy E. Dominguez/Eugenia Abaleke/Yael Feinstein; Noah's Ark Children's Hospital for Wales: Siva Oruganti/Sara Harrison; Nottingham University Hospitals: Catarina Silvestre; Oxford Radcliffe Hospitals NHS Foundation Trust: James Weitz; Royal Manchester Children's Hospital: Peter-Marc Fortune/Gayathri Subramanian/Claire Jennings; St. Mary's Hospital: David Inwald/Calandra Feather/May-Ai Seah/Joanna Danin. *USA*. Arkansas Children's Hospital: Ron Sanders/Glenda Hefley/Katherine Irby/Lauren Edwards/Robert F Buchmann; Children's Hospital and Medical Center: Sidharth Mahapatra/Edward Truemper/Lucinda Kustka; Children's Hospital at Dartmouth: Sholeen T. Nett/Marcy Singleton/J. Dean Jarvis; Children's Hospi-

tal Colorado: Aline B. Maddux/Peter M. Mourani/Kimberly Ralston/Yamila Sierra/Jason Weinman/Zach VanRheen/Christopher Newman; Children's Hospital Los Angeles: Robinder Khemani/Christopher Newth/Jeni Kwok/Rica Morzov/Natalie Mahieu; Children's Hospital of Philadelphia: Nadir Yehya/Natalie Napolitano/Marie Murphy/Laurie Ronan/Ryan Morgan/Sherri Kubis/Elizabeth Broden; Children's Hospital of Wisconsin: Rainer Gedeit/Kathy Murkowski/Katherine Woods/Mary Kasch; Children's Mercy Hospital and Clinics: Yong Y. Han/Jeremy T. Affolter/Kelly S. Tieves/Amber Hughes-Schalk; Cincinnati Children's Hospital Medical Center: Ranjit S. Chima/Kelli Krallman/Erin Stoneman/Laura Benken/Toni Yunger; Connecticut Children's Medical Center: Christopher L Carroll/James Santanelli; Inova Children's Hospital: W. Keith Dockery/Shirin Jafari-Namin/Dana Barry/Keary Jane't; Joseph M Sanzari Children's Hospital at Hackensack University Medical Center: Shira Gertz; Nicklaus Children's Hospital: Fernando Beltramo/Balagangadhar Totapally/Beatriz Govantes; Northwestern University, Ann & Robert H Lurie Children's Hospital of Chicago: Bria Coates/Lawren Wellisch/Kiona Allen/Avani Shukla; Penn State Hershey Children's Hospital: Neal J. Thomas/Debbie Spear; Rainbow Babies and Children's Hospital, Steven L. Shein/Pauravi Vasavada; Saint Barnabas Medical Center: Shira Gertz; Stony Brook Children's Hospital: Margaret M. Parker/Daniel Sloniewsky; The Children's Hospital of Oklahoma; Christine Allen/Amy Harrell; UCSF Benioff Children's Hospital Oakland: Natalie Cvijanovich; University of Miami/Holtz Children's Hospital: Asumthia S. Jeyapalan/Alvaro Coronado-Munoz; University of Michigan–C.S. Mott Children's Hospital: Heidi Flori/Mary K. Dahmer/Chaandini Jayachandran/Joseph Kohne; University of Minnesota Masonic Children's Hospital: Janet Hume/Dan Nerheim/Kelly Dietz; University of WA/Seattle Children's Hospital: Lincoln Smith/Silvia Hartmann/Erin Sullivan/Courtney Merritt; Weill Cornell Medical College: Deyin D. Hsing/Steve Pon/Jim Brian Estil/Richa Gautam; Yale School of Medicine: John S. Giuliano Jr./Joana Ta.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ARDS	Acute Respiratory Distress Syndrome
AUROC	Area Under the ROC Curve
CDSS	Clinical Decision Support System
CHUSJ	Centre Hospitalier Universitaire Sainte-Justine (Sainte-Justine Hospital)
CXR	Chest X-Ray
CPU	Central Processing Unit
DL	Deep Learning
GPU	Graphical Processing Unit
LL	Left Lower
LU	Left Upper
ML	Machine Learning
NLP	Natural Language Processing
PALICC	Pediatric Acute Lung Injury Consensus Conference
PARDS	Pediatric Acute Respiratory Distress Syndrome
PICU	Pediatric Intensive Care Unit
ReLU	Rectified Linear Unit
RL	Right Lower
ROC	Receiver Operating Characteristic
RU	Right Upper
UI	User Interface

CHAPTER 7 GENERAL DISCUSSION

This chapter presents an overview of the main contributions of this study, emphasizing their advantages and limitations. It begins by exploring the significance of providing the annotated ARDS dataset in section 7.1. This dataset fills a crucial gap in the research community, as there is currently no available dataset with ARDS labels. In section 7.2, the contribution of the proposed algorithm for joint segmentation-classification is discussed in detail. The algorithm's performance is examined, highlighting key aspects and considerations. Finally, section 7.3 focuses on the platform and its interpretability. This section provides insights into the platform's capabilities and discusses its ability to provide meaningful explanations.

7.1 Annoated ARDS Dataset

Numerous CXRs have been utilized in pulmonary analysis through large datasets for research purposes. However, a significant gap exists in the analysis of ARDS, hindering both research and clinical applications, as there is a lack of properly labeled datasets specifically created for ARDS studies. Although a private organization has recently developed an ARDS-labeled dataset, unfortunately, it is not accessible to academics [29].

In this project, we aimed to alleviate this restriction by collecting and labeling a comprehensive dataset accessible for research purposes. The created dataset includes global labels for the entire image and localized labels for each quadrant of the CXR image. Furthermore, we have enriched the dataset with lung segmentation maps. To our knowledge, this dataset represents the first endeavor to make ARDS labels available for research purposes. Additionally, it is the first segmentation dataset to provide ground-truth lung fields for abnormal images affected by infiltrates. As a result, this study provides valuable data for the research community, which can enhance ARDS analysis and facilitate the development of associated models.

The data acquisition process for this thesis project adheres to the protocol submitted to IVADO, considering ethical and legal restrictions. As previously mentioned, the dataset was constructed using multiple data sources. The previous work by Nesrine et al. serves as the primary source of the dataset [21]. Although the dataset includes quadrant-level labels, it does not contain explicitly lung area segmentation maps. The Montgomery dataset, originally generated for segmentation purposes, served as the second data source for this study [144]. After obtaining confirmation from medical professionals at CHUSJ, the lung quadrants in this dataset were presumed to be free of ARDS. The third data source was the Chest X-Ray

14 dataset, which did not include ARDS labels or segmentation ground-truth [2]. To align the images from this dataset with ours, four medical professionals from CHUSJ provided labels and maps for each image. The labeling process involved each doctor assigning labels to the lung quadrants while two other clinicians extracted the lung areas using a provided segmentation tool.

Since diagnosing ARDS is naturally challenging and often results in moderate agreement among clinicians, several images were excluded from the dataset at this stage due to disagreements among expert observers. For the segmentation process, a research fellow extracted the lung fields from the chest X-rays, which another clinician reviewed and confirmed. Finally, the last data source consisted of data provided by the PARDIE group. Before proceeding with further processing, this data had to be anonymized. Subsequently, the procedures used for the third data source were followed.

In the data preparation phase, the anonymized data went through a strict security protocol. It was securely stored at the Laboratoire d'Imagerie et de Vision 4D (LIV4D) card-access lab on encrypted hard disks. To ensure confidentiality, the data transformation process was carried out using a password-based cloud platform that only granted access to individuals involved in the data collection process.

The final version of the dataset is currently stored solely on a cloud platform and is accessible only by Dr. Jouvet. He can provide it to interested parties for research purposes, but access to the dataset is subject to specific data usage instructions. Researchers can utilize this dataset for various tasks, such as segmentation and classification.

Having an ARDS-labeled dataset is crucial for researchers as it provides important annotated data for training and evaluating algorithms related to ARDS. However, it is critical to understand that the dataset's size might be a limitation, especially for specific deep-learning models trained from scratch. Compared to publicly available datasets with many CXRs, the ARDS dataset is relatively small despite its high quality and accurate labeling. The small size of the ARDS-labeled dataset can affect the performance of complex deep-learning algorithms that require large-scale datasets to learn intricate patterns and generalizations effectively. When utilizing a small dataset, there is an increased risk of overfitting, indicating that the training set may not fully represent the broader real-world data.

7.2 Joint Segmentation-Classification

The complete system must demonstrate satisfactory classification accuracy to be accepted for use in clinics. According to the context of this project, the analysis of CXRs and ARDS is characterized by the diffusion of lung infiltrates. However, accurately classifying infiltrates

remains the most challenging task among the various anomalies commonly mentioned in scientific literature.

Recent research publications have attempted to address this problem using advanced technologies. One notable approach is utilizing a version of vision transformers specifically designed for classification tasks [155]. However, despite their promising performance in thorax abnormalities, achieving a high AUROC score has proven challenging. The best AUROC score obtained in these experiments was only 72%, despite the average performance of 81%. The state-of-the-art deep model known as Densenet-121, renowned for its ability to classify thoracic diseases, yielded similar performance. This suggests that effectively classifying ARDS infiltrates remains a challenging task to overcome, even with the use of the most cutting-edge technology. Furthermore, the efficacy of these models may not align with the clinical requirements to be used in hospital wards.

On the other hand, the proposed approach has a significant AUROC of 94%, indicating promising outcomes. While it is still necessary to validate the model on external data and evaluate its performance in near-realistic scenarios, it is evident that the model has been thoroughly designed and fine-tuned, surpassing other existing models.

It is important to note that many of the models in this literature were evaluated using a multi-label task, which involves the classification of infiltrates along with several other abnormalities. Sometimes, this may lead to lower classification results, as label dependencies can potentially mislead the model and adversely affect the network's performance during the validation phase. Therefore, the pre-trained Densenet is initially used in this project to evaluate the ARDS dataset. The Densenet-121 model, pre-trained on CXR-14, is also evaluated on our ARDS dataset to assess its capabilities, and the results were as low as 80%, as shown in Table 5.4.

To maximize the potential of the model within the provided platform, a strategy involving an ensemble model consisting of six trained model instances has been implemented. The results, as presented in Table 6.3, demonstrate an improvement in accuracy, with the average value increasing from 91% to 94%. We maintain an optimistic outlook regarding the overall value that can be achieved while acknowledging the potential for a drop in performance when applied to external datasets.

Connection between segmentation and classification: In the proposed Dense-Ynet, the two tasks are interconnected in a way that shares a significant portion of their weights. Consequently, any deficiencies in lung segmentation may indicate potential classification errors.

If the resulting segmentation map is of low quality, it is an indicator for the end-user to evaluate the provided classification labels prudently. While there is no guarantee that correct

segmentation will lead to accurate classification, it emphasizes the importance of proper lung segmentation in achieving trustworthy diagnostic results.

Although lung segmentation is not the primary focus of this work, the obtained segmentation result is remarkably promising. Accurate lung segmentation is essential for various medical imaging applications because it enables precise analysis and detection of lung disorders. By using normal and abnormal images as inputs for the models, the model becomes capable of accurate segmentation, even for pathological CXRs. Most of the previous research on lung field segmentation has primarily focused on normal images, including the Montgomery [144] and JSRT datasets [120]. Comparing methods becomes complicated due to the lack of publicly available datasets that provide segmentation maps of diseased images, particularly those affected by anomalies like infiltrations. However, visual results show the model's success when facing abnormal X-ray images. In our work, obtaining precise segmentation maps of lungs significantly improves ARDS segmentation accuracy.

In the PARDS-CxR platform, the proposed model can segment critical images by providing segmentation maps alongside classification results. This simultaneous training of the algorithm with both ground-truths enhances performance for both tasks by leveraging the complementary information offered by classification labels and detailed segmentation maps. Yu et al. also confirmed the advantages of concurrently training a segmentation network and a classification network for lung nodule classification in CT images [156]. Therefore, not only do we provide a reliable reference for lung segmentation, which can be helpful for pathological CXRs, but utilizing segmentation reference maps also improves the overall diagnostic model.

On the negative side, it is necessary to acknowledge that segmentation maps are a key component of the proposed ARDS classification process. While requiring two sets of ground-truths (Segmentation pixel-level and classification quadrant-level) is reasonable when both segmentation and classification tasks are targeted, it becomes a constraint when we are exclusively aimed at ARDS classification. So, for the further development and enhancement of our models, any attempt to fine-tune or retrain the proposed Dense-Ynet necessitates both datasets. In other words, unlike many classification models, relying solely on image-level labels is insufficient in developing the proposed model. Fortunately, the current network shows the potential to provide near-ground-truth maps for segmenting faded lungs in CXR images. This capability allows it to enrich existing datasets with lung segmentation maps. As a result, in creating suitable datasets, refining these maps becomes a viable option instead of creating them from scratch, making it a time and resource-saving approach that aligns well with models like ours.

Generalization When it comes to ML problems, one always encounters a significant concern regarding the performance of the model when applied to unseen external data. This

concern is particularly heightened in the field of CXR analysis due to the various situations of image taking and many devices that capture CXR images. When considering the generalization ability of an ML-based decision-making system, it is crucial to investigate three key concepts: data, model, and the nature of the present problem.

Accessing the right data was one of the primary challenges our project encountered. The quantity, quality, and representativeness of the data strongly influence the system's capacity for accurate generalization. Since there were no accessible datasets with ARDS labels, we created our dataset that includes the necessary labels to ensure appropriateness. However, the prevalence of chest X-rays in medical practice raises a more significant concern. Chest X-rays are commonly recommended for patients who visit their doctor or the emergency department with symptoms such as chest pain, injury, or shortness of breath. This requires a wide market for various X-ray image-capturing devices, leading to the utilization of numerous types of image-taking tools and settings. Consequently, a wide range of CXR image types exist, each exhibiting distinct characteristics and detailed specifications.

The amount of training data alone cannot determine the effectiveness of ML models in achieving generalization. This is because it is not feasible to encompass data from all existing devices. ML models may encounter difficulties in generalizing successfully when they encounter data from new sources. Therefore, although considering the size of our dataset as a limitation in this work is reasonable, models should prioritize understanding the underlying structure rather than relying solely on training data. As explained, the ARDS-labeled dataset includes four various data sources, but achieving maximum generalization is not guaranteed. Thus, the data challenge in generalizing CXR analysis models cannot be resolved by simply increasing the size of the datasets. Instead, a high-quality model is necessary.

Occam's Razor[1] is a fundamental notion with scientific and philosophical roots that aid in achieving better generalization [157]. According to Occam's Razor, the one with the fewest assumptions should be preferred when faced with two competing hypotheses. This principle suggests that simpler models make fewer assumptions about the training data and are thus more likely to capture genuine patterns, thereby improving generalization. Simpler models offer several advantages in terms of mitigating overfitting. These models are less prone to overfitting because they have a limited capacity to fine-tune and adapt to noise in the training data. Instead, they focus on identifying crucial connections and patterns to enhance generalization to new, unseen data. Moreover, the simplicity of models facilitates a more comprehensive exploration of hyperparameters, such as the learning rate or regularization terms. As the complexity of the model decreases, tuning these hyperparameters becomes

[1]Occam's (or Ockham's) razor is a principle attributed to the 14th-century logician and Franciscan friar William of Ockham.

more straightforward, thereby enhancing generalization performance. Additionally, simpler models are generally easier to interpret, making it easier to understand and articulate arguments in favor of such decisions.

The Dense-Ynet model has been specifically developed to reduce the need for many training parameters. This is achieved by incorporating dense connections in the model, facilitating a high information flow. By utilizing the structures of U-net, Densenet, and Y-net, the model can maintain its lightweight nature while still extracting relevant features. In addition to its ability to capture informative features, the compact architecture of the Dense-Ynet model makes it well-suited for achieving good generalization. Due to its fewer parameters and effective information flow, the model can focus on identifying essential patterns and relationships in the data, thereby enhancing generalization performance. Thanks to this advantage, the Dense-Ynet model is appropriate for various decision-making tasks while retaining its capacity to generalize to unseen data.

Another rationale for developing a lightweight model is to prevent overfitting, which, given our dataset, may occur fast with a larger model. The structure of the model and hyperparameters have been carefully adjusted to strike a balance between the dataset size and the number of training parameters in the model. This helps the model learn specific patterns in CXR images and identify ARDS signs without becoming excessively specialized to the training data, thereby promoting better generalization and enhancing decision-making capabilities.

Another crucial aspect considered in designing the Dense-Ynet model was using segmentation advantages for the classification task. Numerous studies in the literature have demonstrated the benefits of segmentation in addressing generalization issues. When analyzing CXR images, lung segmentation plays a vital role in ignoring the background, which can introduce bias and variability in the images. By incorporating segmentation techniques, the model aimed to enhance its generalizability. Dense-Ynet leveraged the benefits of segmentation to improve its overall performance and generalization capabilities. Therefore, the model is designed to achieve better generalization by considering both the size of the model and incorporating segmentation in its structure.

The inherent complexity of the problem impacts the generalization of underlying ML-based decision-making systems. Some tasks are more challenging to generalize than others, particularly those with high dimensionality, noisy data, or ambiguous patterns. Generalization can be greatly enhanced by understanding the features of the problem and adjusting the data preparation, feature engineering, and model selection accordingly.

In the study by Maguolo et al., the researchers utilized AlexNet to classify images based on the datasets they belong to [61]. Surprisingly, the model achieved great success in accurately

identifying the data source, thereby highlighting the inherent data challenges associated with this task. However, this success also brought to light the complexity of CXR analysis, as the model's performance may vary between different datasets. Additionally, their study revealed that biased results might be reported even in highly cited publications. The researchers successfully trained a model using CXR images with occluded lungs, challenging the notion that specific features are necessary for achieving high-performance classification.
The nature of the issue gets even more challenging when considering CXR images collected under various circumstances, such as the patient's posture, breathing condition, age, etc. For instance, individuals with severe COVID-19 might be unable to stand during image acquisition. Suppose a decision-making system is designed to identify a specific disease using CXRs. In that case, there is a risk that the diagnosis model tends to prioritize recognizing the positions of patients rather than focusing on the pulmonary signs of the disease. In less severe cases, the model could exhibit bias towards the illness based on the unintended external circumstances of the patient. Similarly, suppose a patient is not upright or fails to take a deep breath for various reasons. In that case, it can result in distortions in the image, making it more challenging to identify abnormalities accurately. Additionally, there might be correlations between diseases and patients' sex, age, or even the imaging center. Suppose an image is taken in an urgent care or COVID-specific center. In that case, the CXR might exhibit biases towards lung trauma or COVID-19, respectively, and the model can recognize the anomaly from the dataset. These aspects highlight the inherent complexity of the model, which often leads to poor generalization across various research efforts in the literature. However, there is hope that the proposed model will perform well during external validation because it was created expressly to capture the structural elements of input images.

Validation There are several ways to validate decision-making systems. Empirical approaches involve comparing the outcomes reached by the model to expert views or assessing the results against real-world data through experiments. Analytical approaches rely on mathematical or logical reasoning to prove the correctness and soundness of the model. Hybrid methods combine empirical and analytical approaches to enhance validity and reliability.
This project employs a hybrid validation approach, commencing with analytical validation. The algorithm and application underwent verification using internal data and established evaluation metrics to ensure accurate results. However, empirical validation is necessary to confirm its suitability for hospital wards. The web-based platform, PARDS-CxR, features a robust review system designed to gather detailed feedback from researchers. This valuable input will greatly facilitate the comprehensive evaluation of the platform and contribute to creating an in-depth report. The ultimate goal is to obtain the necessary approvals for de-

ploying the application in hospitals.
Validation of the CXR analysis model with actual data is crucial to ensure its reliability. In this project, the model's settings are designed for validation by clinicians and researchers globally. The algorithm cannot be utilized until it successfully passes the rigorous external validation process, guaranteeing its accuracy and performance in real-world scenarios.

7.3 Interpretable platform

As discussed earlier, for online medical applications, the interpretability of AI models is essential. It enhances confidence and trust, provides information about underlying processes, and ensures accuracy and dependability. Additionally, it contributes to avoiding legal and ethical issues related to the application of AI.
According to a study by Richens et al., existing ML techniques for diagnoses mostly look for links between patients' symptoms and their medical history [158]. These techniques only reveal disorders directly connected to the symptoms rather than searching for the underlying reasons for the symptoms. When there are multiple possible explanations for a patient's symptoms, this might result in a less precise or even dangerous diagnosis. The initial impact of these findings shows that lacking a strong justification, doctors are unlikely to rely entirely on the outputs of machine learning computations for diagnosis.
Methods like GradCam are methods of local explainability that highlight the attention to the pixel locations that match a certain prediction. The model's last convolution layer generates gradients used to construct the heatmaps. However, they do not provide interpretability of the underlying models.
To enhance the interpretability of AI models, two main strategies are employed:

- Ante-hoc methods: These involve using inherently interpretable models like decision trees or logistic regression. Techniques such as fuzzy logic or decision rules construct models with built-in interpretability features.
- Post-hoc approaches: These techniques utilize external methods to explain the behavior or output of a black-box model. They include sensitivity analysis, visualizing model behavior, analyzing feature importance, or employing activation maximization.

In our task, the model's interpretability arises from its adaptability to established criteria for ARDS in at least two definitions: ARDS [1] and PARDS [14]. By the very nature

of deep learning models, they make it challenging to achieve straightforward, fine-detailed interpretability. However, adding a higher level of interpretability enhances the model's capability.

While it may not provide complete interpretability through a direct connection between input and output in the complex deep model, the interpretability strategy employed in this method offers clinicians transparency in ARDS detection. Doctors can determine whether the model's conclusions are supported by sound reasoning. Due to the capacity of the PARDS-CxR platform to align with diverse definitions of ARDS, there is no need to retrain the model or acquire new datasets to provide accurate diagnoses under different definitions.

The final results of the model can be equipped with activation map techniques such as Grad-Cam or LIME. However, it is controversial to use them to resolve interpretability issues [2].

Doctors can analyze the model from two viewpoints due to the platform's resources for providing local information.

Firstly, they observe that the ultimate conclusion is supported by clinical rationale. The Berlin definition, for instance, offers a logical basis where at least one quadrant on each side of the lung must be engaged. This criterion ensures the validity of the procedure. If physicians disagree with the quadrant labels assigned to the lung side, they can contest the decision. While the level of detail regarding lung locality may not be granular, determining the degree of engagement, such as identifying infiltrations, provides a reference point at the quadrant level.

However, the platform does not offer pixel-level details or small patches of lung areas in the CXR image, limiting the model's granularity. As a result, the finest level of detail provided is represented by lung quadrant scales. While the highly detailed level has its benefits, it also has drawbacks. On the positive side, having access to more detailed information about the patient's CXR imaging can aid in the validation process and instill confidence in doctors. This additional information can assist in obtaining more accurate readings and a better understanding of the image.

In the second viewpoint, it is important to consider the drawbacks of providing an excessively detailed reference. One negative aspect is that excessive information may complicate the validation process, potentially leading to confusion and consuming more time. When there is an overwhelming amount of information accessible, it becomes challenging to validate and interpret the results. Striking the right balance between detail and interpretability is crucial to ensure efficient and accurate validation.

We believe that offering interpretability at quadrant levels, as the platform presents, is suf-

[2]It is crucial to note that these techniques have certain drawbacks, particularly in clinical settings. Therefore, the deployable version of the platform does not include them.

ficient for clinical purposes. It can still be a useful tool for diagnosis in medical settings without overwhelming details.

From the technical perspective, if an Ante-hoc mode is desired for such a problem, a patch-based design would be the appropriate solution. However, there are certain challenges with its implementation and design. First, obtaining patched data might be costly and challenging to acquire. Second, establishing a patch-based approach on web servers or other computational infrastructure may be complicated due to the increased processing resources required. Therefore, finer details and patch-based models may offer particular advantages. However, carefully weigh their implementation's trade-offs and practical implications.

System Integration When deploying the platform on the web, it is crucial to prioritize the integration of PARDS-CxR with existing workflows and tools used in clinical settings. In this regard, there are primarily two viewpoints to consider. Firstly, the app can be deployed easily within hospital wards, offering simple accessibility and efficient data transfer between the image-capturing equipment and the application. This integration is crucial for efficient workflow and effective utilization of the app in real-time medical scenarios.

Secondly, the integration aspect is important due to the nature of ARDS diagnosis, which requires a multi-factor analysis. While analyzing CXR is vital, accurately investigating other criteria outlined in ARDS definitions is similarly essential. The app enables healthcare professionals to evaluate all pertinent factors for a thorough and precise diagnosis by smoothly integrating with other clinical tools. PARDS-CxR provides a simple severity score for ARDS in patients, making it easy to consider the app's assessment within an integrated system of ARDS detection that incorporates the full criteria of ARDS.

A key feature of a web application is input compatibility, which guarantees a seamless connection with established protocols like DICOM. This compatibility provides quick data transformation through the program to access data from the imaging device or clinic archives. In addition, the output of the app can be presented in CSV or standard JSON formats, both of which are readable and integrable. This adaptability makes it easier to integrate the app with other tools because it allows it to conveniently share information about the severity of ARDS in the given radiograph, including data for each quadrant, overall scores, final labels, and the segmentation map. So, the application can seamlessly collaborate with various clinical systems and software by adhering to widely accepted standards such as DICOM and offering output formats like CSV and JSON. The app's usability and usefulness in real-world medical settings are affected by its simplicity of integration inside a healthcare environment.

In the validation version of the app, the output of the app can be available in a portable document format (PDF). This format ensures clarity for end-users and allows them to access and evaluate the web application consistently.

The web app can generate a log file that reports the app's usage received by any other tool. This log file serves multiple purposes, including ensuring data transformations' validity and enhancing inter-applicational connectivity. The log file provides the overall efficiency of the web app within a widespread network of healthcare apps by giving insights into usage patterns and interactions with other tools. While the app can be a valuable tool for collecting data, it functions as a "black box," guaranteeing simple and smooth data translation without leaving any unnecessary or undesirable traces. While preserving the app's usability and effectiveness, this strategy emphasizes user privacy and confidentiality. The web app can be a reliable component in the network by functioning as a black box with streamlined data transformation and minimal footprint. It offers a smooth integration experience without adding extra data storage or tracking to the entire system. This specification enhances the efficiency and usability of the app, allowing healthcare professionals to focus on utilizing its core functionalities without concerns about data leakage or unwanted data accumulation.

APIs (Application Programming Interfaces) are essential in easily integrating online applications into networks, providing a straightforward and authorized approach to accessing the application and facilitating integration within the healthcare workflow. Although an application programming interface (API)–based application may not be necessary for the validation version, it is vital to recognize that the PARDS-CxR web app's lack of an API may affect its ability to integrate with the rest of the healthcare infrastructure. Therefore, to support the web application's future development, it becomes vital to consider creating API-based access.

When implementing an API, including thorough support documentation to enable efficient use is crucial. Although a video illustrating how to use the application has already been created, it is crucial to include an entire manual to provide a complete grasp of the capabilities of the API.

Out of Distribution DetectionA safe medical diagnosis system must be able to recognize instances that go outside of its predetermined knowledge limits. This makes it possible to assess the system's performance using comparable samples. Differentiating between samples that are inside the training distribution (known as "In data") and those that are outside of it (known as "Out data") is the goal of out-of-distribution detection (OoDD).

CHAPTER 8 CONCLUSION

The high-level contribution of this project includes three main domains. Firstly, the release of data for ARDS detection is of great significance. Secondly, developing a deep model compatible with the limited data and yielding interpretable results is crucial. Thirdly, creating a platform that organizes the interaction between users and the model, as well as serves the purpose of validation.

The research addresses a critical research gap by providing a dataset for analyzing CXRs for ARDS diagnosis. To fill this gap, an ARDS-labeled dataset was created using data from four sources and various image-capturing equipment. However, annotating medical data, particularly CXR images, poses several challenges. As a result, the dataset size is insufficient to train data-hungry DL models adequately. This project presents a unique opportunity to collect data from multiple image-capturing sources. Researchers and physicians can contribute by providing input and making initial annotations for the data through the online application. This distributed annotation strategy has the potential to gather a significant dataset. Before including an image in the current dataset, it should undergo a review procedure conducted by medical experts within the application's host environment to ensure quality control. Notably, the data collection strictly adheres to approved data collection protocols and privacy policies. Data is not saved on the server unless authorized explicitly by the user. Therefore, it is highly recommended to utilize PARDS-CxR to its maximum capacity to acquire a sufficiently large dataset. This dataset will not only improve the functionality of the current model but will also benefit other researchers in the scientific community. Another key benefit of acquiring a distributed dataset is the opportunity to conduct comprehensive numerical statistical analysis. The abundance of data collected from multiple sources allows for an extensive examination of various patients and CXRs on a large scale. This analysis enables valuable insights to be obtained regarding the characteristics, trends, and variations present within the dataset. In summary, by leveraging the capabilities of distributed annotation and ensuring compliance with privacy protocols, the PARDS-CxR application has the potential to enhance the existing dataset greatly. This enables further statistical analysis of the large-scale dataset acquired through user utilization.

The proposed Dense-Ynet model demonstrates promising performance in ARDS detection using CXR images. However, it is essential to note that the model has only been tested on a subset of our internal dataset. Therefore, it is crucial to assess the generalizability of the model for practical applications. To overcome this limitation, we have taken measures to reduce biases and ensure a thorough evaluation of the model's generalizability. We have en-

listed the expertise of specialists from various healthcare facilities to assess the effectiveness of the model. Section 6.5 contains a list of the PARDIE group researchers who are welcome to participate in evaluating this study. We recognize the need to conduct a rigorous evaluation process to confirm the model's generalizability. Consequently, we recommend using a thorough methodology that covers the entire procedure, ensuring the dependability and credibility of the decision-making system for subsequent phases of this project. By doing so, we can maximize the ability of the Dense-Ynet model to detect ARDS using CXR images and establish a strong foundation for its application in real-world scenarios.

The platform includes a user-friendly UI that facilitates user interactions, especially during validation. It generates a PDF report for each session and creates a log file that can be utilized in other applications. However, two restrictions have been identified in the previous sections. Firstly, supporting API-based queries is essential for an app like PARDS-CxR. It can function as a standalone tool or be easily integrated into any system, even without direct access to the program's source code. Secondly, in the current platform, OoDD yields some limitations. While uploading irrelevant images may not pose a significant challenge, problems arise when users mistakenly select the wrong analysis platform for processing a different modality. Receiving a notice stating that the image is not included in the distribution of CXR images is crucial. The issue can be effectively resolved by adding an OoDD functionality to the platform.

Based on the general discussion in this section, there are several recommendations for future work that can enhance the functionality and effectiveness of the platform. Firstly, it is crucial to focus on providing more data through the platform. Expanding the dataset will improve the accuracy and reliability of the system and offer a helpful training set for other research in this domain. Additionally, creating an API for system integration would enable smooth interaction with other programs and services, promoting interoperability and boosting the platform's usefulness. Furthermore, validation of the platform should be prioritized to ensure its effectiveness and address any potential issues or limitations for real applications. Conducting statistical analysis of the uploaded data will provide valuable insights, helping to identify patterns, trends, and areas for improvement. Lastly, implementing an OoDD mechanism will enhance the platform's robustness by identifying instances where the input data falls outside CXR image distributions, thereby enabling appropriate actions. If incorporated into future work, these suggestions will significantly contribute to the platform's overall functionality and performance.

REFERENCES

[1] A. D. T. Force, V. Ranieri, G. Rubenfeld, B. Thompson, N. Ferguson, E. Caldwell *et al.*, "Acute respiratory distress syndrome," *Jama*, vol. 307, no. 23, pp. 2526–2533, 2012.

[2] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2097–2106.

[3] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya *et al.*, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[4] M. A. Matthay, R. L. Zemans, G. A. Zimmerman, Y. M. Arabi, J. R. Beitler, A. Mercat, M. Herridge, A. G. Randolph, and C. S. Calfee, "Acute respiratory distress syndrome," *Nature reviews Disease primers*, vol. 5, no. 1, p. 18, 2019.

[5] E. Posner, "The early years of chest radiology in britain." *Thorax*, vol. 26, no. 3, p. 233, 1971.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[7] B. McNicholas, M. G. Madden, and J. G. Laffey, "Machine learning classifier models: the future for acute respiratory distress syndrome phenotyping?" pp. 919–920, 2020.

[8] M. Yahyatabar, P. Jouvet, and F. Cheriet, "Dense-unet: a light model for lung fields segmentation in chest x-ray images," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 1242–1245.

[9] ——, "Joint classification and segmentation for an interpretable diagnosis of acute respiratory distress syndrome from chest x-rays," *Journal of Medical Imaging*, vol. 10, no. 5, pp. 054 504–054 504, 2023.

[10] M. Yahyatabar, P. Jouvet, D. Fily, J. Rambaud, M. Levy, R. G. Khemani, and F. Cheriet, "A web-based platform for the automatic stratification of ards severity," *Diagnostics*, vol. 13, no. 5, p. 933, 2023.

[11] E. Fan, D. Brodie, and A. S. Slutsky, "Acute respiratory distress syndrome: advances in diagnosis and treatment," *Jama*, vol. 319, no. 7, pp. 698–710, 2018.

[12] D. Ashbaugh, D. B. Bigelow, T. Petty, and B. Levine, "Acute respiratory distress in adults," *The Lancet*, vol. 290, no. 7511, pp. 319–323, 1967.

[13] G. R. Bernard, A. Artigas, K. L. Brigham, J. Carlet, K. Falke, L. Hudson, M. Lamy, J. R. Legall, A. Morris, and R. Spragg, "The american-european consensus conference on ards. definitions, mechanisms, relevant outcomes, and clinical trial coordination." *American journal of respiratory and critical care medicine*, vol. 149, no. 3, pp. 818–824, 1994.

[14] P. A. L. I. C. C. Group *et al.*, "Pediatric acute respiratory distress syndrome: consensus recommendations from the pediatric acute lung injury consensus conference," *Pediatric critical care medicine: a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies*, vol. 16, no. 5, p. 428, 2015.

[15] R. G. Khemani, L. Smith, Y. M. Lopez-Fernandez, J. Kwok, R. Morzov, M. J. Klein, N. Yehya, D. Willson, M. C. Kneyber, J. Lillie *et al.*, "Paediatric acute respiratory distress syndrome incidence and epidemiology (pardie): an international, observational study," *The Lancet Respiratory Medicine*, vol. 7, no. 2, pp. 115–128, 2019.

[16] G. Emeriaud, Y. M. López-Fernández, N. P. Iyer, M. M. Bembea, A. Agulnik, R. P. Barbaro, F. Baudin, A. Bhalla, W. B. de Carvalho, C. L. Carroll *et al.*, "Executive summary of the second international guidelines for the diagnosis and management of pediatric acute respiratory distress syndrome (palicc-2)," *Pediatric Critical Care Medicine*, vol. 24, no. 2, pp. 143–168, 2023.

[17] J. Phua, J. R. Badia, N. K. Adhikari, J. O. Friedrich, R. A. Fowler, J. M. Singh, D. C. Scales, D. R. Stather, A. Li, A. Jones *et al.*, "Has mortality from acute respiratory distress syndrome decreased over time? a systematic review," *American journal of respiratory and critical care medicine*, vol. 179, no. 3, pp. 220–227, 2009.

[18] G. Bellani, J. G. Laffey, T. Pham, E. Fan, L. Brochard, A. Esteban, L. Gattinoni, F. Van Haren, A. Larsson, D. F. McAuley *et al.*, "Epidemiology, patterns of care, and

mortality for patients with acute respiratory distress syndrome in intensive care units in 50 countries," *Jama*, vol. 315, no. 8, pp. 788–800, 2016.

[19] H. Banavasi, P. Nguyen, H. Osman, and A. O. Soubani, "Management of ards – what works and what does not," *The American Journal of the Medical Sciences*, vol. 362, no. 1, pp. 13–23, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0002962920305577

[20] W. Dai, J. Doyle, X. Liang, H. Zhang, N. Dong, Y. Li, and E. P. Xing, "SCAN: Structure Correcting Adversarial Network for Organ Segmentation in Chest X-rays," *arXiv e-prints*, p. arXiv:1703.08770, Mar 2017.

[21] N. Zaglam, S. Essouri, O. Fléchelles, G. Emeriaud, F. Cheriet, and P. Jouvet, "Inter-observer variability for radiography in pediatric acute respiratory distress syndrome and improvement with a computer-aided diagnosis system," *Austin J Emergency & Crit Care Med*, vol. 2, no. 3, p. 1020, 2015.

[22] W. Zhang, Y. Chang, Y. Ding, Y. Zhu, Y. Zhao, and R. Shi, "To establish an early prediction model for acute respiratory distress syndrome in severe acute pancreatitis using machine learning algorithm," *Journal of Clinical Medicine*, vol. 12, no. 5, p. 1718, 2023.

[23] M. Sauthier, G. Tuli, P. A. Jouvet, J. S. Brownstein, and A. G. Randolph, "Estimated pao2: a continuous and noninvasive method to estimate pao2 and oxygenation index," *Critical care explorations*, vol. 3, no. 10, 2021.

[24] T.-D. Le, R. Noumeir, J. Rambaud, G. Sans, and P. Jouvet, "Detecting of a patient's condition from clinical narratives using natural language representation," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 3, pp. 142–149, 2022.

[25] F. Plourde, F. Cheriet, and J. Dansereau, "Semiautomatic detection of scoliotic rib borders from posteroanterior chest radiographs," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 4, pp. 909–919, April 2012.

[26] N. Zaglam, P. Jouvet, O. Flechelles, G. Emeriaud, and F. Cheriet, "Computer-aided diagnosis system for the acute respiratory distress syndrome from chest radiographs," *Computers in Biology and Medicine*, vol. 52, pp. 41 – 48, 2014.

[27] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

[28] M. W. Sjoding, D. Taylor, J. Motyka, E. Lee, D. Claar, J. I. McSparron, S. Ansari, M. P. Kerlin, J. P. Reilly, M. G. Shashaty *et al.*, "Deep learning to detect acute respiratory distress syndrome on chest radiographs: a retrospective study with external validation," *The Lancet Digital Health*, vol. 3, no. 6, pp. e340–e348, 2021.

[29] K.-C. Pai, W.-C. Chao, Y.-L. Huang, R.-K. Sheu, L.-C. Chen, M.-S. Wang, S.-H. Lin, Y.-Y. Yu, C.-L. Wu, and M.-C. Chan, "Artificial intelligence–aided diagnosis model for acute respiratory distress syndrome combining clinical data and chest radiographs," *Digital Health*, vol. 8, p. 20552076221120317, 2022.

[30] M. L. Giger, H.-P. Chan, and J. Boone, "Anniversary paper: history and status of cad and quantitative image analysis: the role of medical physics and aapm," *Medical physics*, vol. 35, no. 12, pp. 5799–5820, 2008.

[31] A. Kumar, Y.-Y. Wang, K.-C. Liu, I.-C. Tsai, C.-C. Huang, and N. Hung, "Distinguishing normal and pulmonary edema chest x-ray using gabor filter and svm," in *2014 IEEE international symposium on bioelectronics and bioinformatics (IEEE ISBB 2014)*. IEEE, 2014, pp. 1–4.

[32] Z. Camlica, H. R. Tizhoosh, and F. Khalvati, "Medical image classification via svm using lbp features from saliency-based folded data," in *2015 IEEE 14th international conference on machine learning and applications (ICMLA)*. IEEE, 2015, pp. 128–132.

[33] Y.-H. Chan, Y.-Z. Zeng, H.-C. Wu, M.-C. Wu, H.-M. Sun *et al.*, "Effective pneumothorax detection for chest x-ray images using local binary pattern and support vector machine," *Journal of healthcare engineering*, vol. 2018, 2018.

[34] N. Singh, R. Sharma, and A. Kukker, "Wavelet transform based pneumonia classification of chest x-ray images," in *2019 International Conference on Computing, Power and Communication Technologies (GUCON)*. IEEE, 2019, pp. 540–545.

[35] F. Eren and Z. Gündoğar, "Classification of covid-19 x-ray images using tridiagonal matrix enhanced multivariance products representation (tmempr)," in *2021 6th International Conference on Computer Science and Engineering (UBMK)*. IEEE, 2021, pp. 221–226.

[36] L. N. Rohmah and A. Bustamam, "Improved classification of coronavirus disease (covid-19) based on combination of texture features using ct scan and x-ray images," in *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*, 2020, pp. 105–109.

[37] S. Khobragade, A. Tiwari, C. Patil, and V. Narke, "Automatic detection of major lung diseases using chest radiographs and classification by feed-forward artificial neural network," in *2016 IEEE 1st international conference on power electronics, intelligent control and energy systems (ICPEICES)*. IEEE, 2016, pp. 1–5.

[38] S. Reza, O. B. Amin, and M. Hashem, "A novel feature extraction and selection technique for chest x-ray image view classification," in *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*. IEEE, 2019, pp. 189–194.

[39] J. Irvin, P. Rajpurkar, M. Ko, Y. Yu, S. Ciurea-Ilcus, C. Chute, H. Marklund, B. Haghgoo, R. Ball, K. Shpanskaya *et al.*, "Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 590–597.

[40] A. E. Johnson, T. J. Pollard, L. Shen, L.-w. H. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L. Anthony Celi, and R. G. Mark, "Mimic-iii, a freely accessible critical care database," *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.

[41] A. Bustos, A. Pertusa, J.-M. Salinas, and M. de la Iglesia-Vayá, "Padchest: A large chest x-ray image dataset with multi-label annotated reports," *Medical image analysis*, vol. 66, p. 101797, 2020.

[42] P. N. Kieu, H. S. Tran, T. H. Le, T. Le, and T. T. Nguyen, "Applying multi-cnns model for detecting abnormal problem on chest x-ray images," in *2018 10th International Conference on Knowledge and Systems Engineering (KSE)*. IEEE, 2018, pp. 300–305.

[43] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[44] M. Frid-Adar, A. Ben-Cohen, R. Amer, and H. Greenspan, "Improving the segmentation of anatomical structures in chest radiographs using u-net with an imagenet pre-trained encoder," *CoRR*, vol. abs/1810.02113, 2018. [Online]. Available: http://arxiv.org/abs/1810.02113

[45] M. T. Islam, M. A. Aowal, A. T. Minhaz, and K. Ashraf, "Abnormality detection and localization in chest x-rays using deep convolutional neural networks," *arXiv preprint arXiv:1705.09850*, 2017.

[46] L. Yao, E. Poblenz, D. Dagunts, B. Covington, D. Bernard, and K. Lyman, "Learning to diagnose from scratch by exploiting dependencies among labels," *arXiv preprint arXiv:1710.10501*, 2017.

[47] Q. Guan and Y. Huang, "Multi-label chest x-ray image classification via category-wise residual attention learning," *Pattern Recognition Letters*, 2018.

[48] Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, and Y. Yang, "Thorax disease classification with attention guided convolutional neural network," *Pattern Recognition Letters*, vol. 131, pp. 38–45, 2020.

[49] X. Liang, C. Peng, B. Qiu, and B. Li, "Dense networks with relative location awareness for thorax disease identification," *Medical physics*, vol. 46, no. 5, pp. 2064–2073, 2019.

[50] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*. PMLR, 2019, pp. 6105–6114.

[51] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.

[52] L. T. Duong, N. H. Le, T. B. Tran, V. M. Ngo, and P. T. Nguyen, "Detection of tuberculosis from chest x-ray images: Boosting the performance with vision transformer and transfer learning," *Expert Systems with Applications*, vol. 184, p. 115519, 2021.

[53] S. Park, G. Kim, Y. Oh, J. B. Seo, S. M. Lee, J. H. Kim, S. Moon, J.-K. Lim, and J. C. Ye, "Vision transformer for covid-19 cxr diagnosis using chest x-ray feature corpus," *arXiv preprint arXiv:2103.07055*, 2021.

[54] I. Feki, S. Ammar, Y. Kessentini, and K. Muhammad, "Federated learning for covid-19 screening from chest x-ray images," *Applied Soft Computing*, vol. 106, p. 107330, 2021.

[55] A. Makkar and K. Santosh, "Securefed: federated learning empowered medical imaging technique to analyze lung abnormalities in chest x-rays," *International Journal of Machine Learning and Cybernetics*, vol. 14, no. 8, pp. 2659–2670, 2023.

[56] F. Ślazyk, P. Jabłecki, A. Lisowska, M. Malawski, and S. Płotka, "Cxr-fl: Deep learning-based chest x-ray image analysis using federated learning," in *International Conference on Computational Science*. Springer, 2022, pp. 433–440.

[57] J. Crosby, T. Rhines, F. Li, H. MacMahon, and M. Giger, "Deep learning for pneumothorax detection and localization using networks fine-tuned with multiple institutional datasets," in *Medical Imaging 2020: Computer-Aided Diagnosis*, vol. 11314. International Society for Optics and Photonics, 2020, p. 113140C.

[58] M. Rasheed, "Classification of chest diseases using wavelet transforms and transfer learning," *Medical Imaging and Computer-Aided Diagnosis*, p. 158, 2020.

[59] I. M. Baltruschat, H. Nickisch, M. Grass, T. Knopp, and A. Saalbach, "Comparison of deep learning approaches for multi-label chest x-ray classification," *Scientific reports*, vol. 9, no. 1, pp. 1–10, 2019.

[60] H. Liu, L. Wang, Y. Nan, F. Jin, Q. Wang, and J. Pu, "Sdfn: Segmentation-based deep fusion network for thoracic disease classification in chest x-ray images," *Computerized Medical Imaging and Graphics*, vol. 75, pp. 66–73, 2019.

[61] G. Maguolo and L. Nanni, "A critic evaluation of methods for covid-19 automatic detection from x-ray images," *Information Fusion*, vol. 76, pp. 1–7, 2021.

[62] L. O. Teixeira, R. M. Pereira, D. Bertolini, L. S. Oliveira, L. Nanni, G. D. Cavalcanti, and Y. M. Costa, "Impact of lung segmentation on the diagnosis and explanation of covid-19 in chest x-ray images," *Sensors*, vol. 21, no. 21, p. 7116, 2021.

[63] N. de Sousa Freire, P. P. de Souza Leão, L. A. Tiago, A. d. A. C. Gonçalves, R. A. Pinto, E. M. dos Santos, and E. Souto, "Generalizability of cnn on predicting covid-19 from chest x-ray images," in *Anais do XXII Simpósio Brasileiro de Computação Aplicada à Saúde*. SBC, 2022, pp. 36–47.

[64] S. G. Armato III, M. L. Giger, and H. MacMahon, "Automated lung segmentation in digitized posteroanterior chest radiographs," *Academic radiology*, vol. 5, no. 4, pp. 245–255, 1998.

[65] N. R. Kasu and C. Saravanan, "Segmentation on chest radiographs using otsu's and k-means clustering methods," in *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2018, pp. 210–213.

[66] L. Li, Y. Zheng, M. Kallergi, and R. A. Clark, "Improved method for automatic identification of lung regions on chest radiographs," *Academic Radiology*, vol. 8, no. 7, pp. 629–638, 2001.

[67] S. Chattopadhyay, "An approach to identify regions of interest in chest x-ray images of covid-19 patients and its clinical validation: An indian study," *Artificial Intelligence Evolution*, pp. 41–55, 2022.

[68] G. Wu, X. Zhang, S. Luo, and Q. Hu, "Lung segmentation based on customized active shape model from digital radiography chest images," *Journal of Medical Imaging and Health Informatics*, vol. 5, no. 2, pp. 184–191, 2015.

[69] B. Van Ginneken, A. F. Frangi, J. J. Staal, B. M. ter Haar Romeny, and M. A. Viergever, "Active shape model segmentation with optimal features," *IEEE transactions on medical imaging*, vol. 21, no. 8, pp. 924–933, 2002.

[70] Y. Shao, Y. Gao, Y. Guo, Y. Shi, X. Yang, and D. Shen, "Hierarchical lung field segmentation with joint shape and appearance sparse learning," *IEEE transactions on medical imaging*, vol. 33, no. 9, pp. 1761–1780, 2014.

[71] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE transactions on medical imaging*, vol. 33, no. 2, pp. 577–590, 2013.

[72] X. Li, L. Chen, and J. Chen, "A visual saliency-based method for automatic lung regions extraction in chest radiographs," in *2017 14th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. IEEE, 2017, pp. 162–165.

[73] M. F. McNitt-Gray, H. Huang, and J. W. Sayre, "Feature selection in the pattern classification problem of digital chest radiograph segmentation," *IEEE Transactions on Medical Imaging*, vol. 14, no. 3, pp. 537–547, 1995.

[74] S. Ghosh and S. K. Dubey, "Comparative analysis of k-means and fuzzy c-means algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 4, no. 4, 2013.

[75] Z. Shi, P. Zhou, L. He, T. Nakamura, Q. Yao, and H. Itoh, "Lung segmentation in chest radiographs by means of gaussian kernel-based fcm with spatial constraints," in *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, vol. 3. IEEE, 2009, pp. 428–432.

[76] N. F. Vittitoe, R. Vargas-Voracek, and C. E. Floyd Jr, "Identification of lung regions in chest radiographs using markov random field modeling," *Medical Physics*, vol. 25, no. 6, pp. 976–985, 1998.

[77] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv e-prints*, p. arXiv:1505.04597, May 2015.

[78] Q. Que, Z. Tang, R. Wang, Z. Zeng, J. Wang, M. Chua, T. S. Gee, X. Yang, and B. Veeravalli, "Cardioxnet: Automated detection for cardiomegaly based on deep learning," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2018, pp. 612–615.

[79] A. Mittal, R. Hooda, and S. Sofat, "Lf-segnet: A fully convolutional encoder–decoder network for segmenting lung fields from chest radiographs," *Wireless Personal Communications*, vol. 101, no. 1, pp. 511–529, Jul 2018. [Online]. Available: https://doi.org/10.1007/s11277-018-5702-9

[80] M. Nishio, K. Fujimoto, and K. Togashi, "Lung segmentation on chest x-ray images in patients with severe abnormal findings using deep learning," *arXiv preprint arXiv:1908.07704*, 2019.

[81] A. A. Novikov, D. Lenis, D. Major, J. Hladůvka, M. Wimmer, and K. Bühler, "Fully convolutional architectures for multiclass segmentation in chest radiographs," *IEEE Transactions on Medical Imaging*, vol. 37, no. 8, pp. 1865–1876, Aug 2018.

[82] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[83] M. Kholiavchenko, I. Sirazitdinov, K. Kubrak, R. Badrutdinova, R. Kuleev, Y. Yuan, T. Vrtovec, and B. Ibragimov, "Contour-aware multi-label chest x-ray organ segmentation," *International Journal of Computer Assisted Radiology and Surgery*, vol. 15, no. 3, pp. 425–436, 2020.

[84] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," in *2017 IEEE visual communications and image processing (VCIP)*. IEEE, 2017, pp. 1–4.

[85] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional densenets for semantic segmentation," in *Proceedings of*

the IEEE conference on computer vision and pattern recognition workshops, 2017, pp. 11–19.

[86] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[87] R. Girshick, "Fast r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1440–1448.

[88] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[89] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[90] T. Peng, Y. Gu, Z. Ye, X. Cheng, and J. Wang, "A-lugseg: Automatic and explainability-guided multi-site lung detection in chest x-ray images," *Expert Systems with Applications*, vol. 198, p. 116873, 2022.

[91] T. Peng, C. Wang, Y. Zhang, and J. Wang, "H-segnet: hybrid segmentation network for lung segmentation in chest radiographs using mask region-based convolutional neural network and adaptive closed polyline searching method," *Physics in Medicine & Biology*, vol. 67, no. 7, p. 075006, 2022.

[92] J. Wang, Z. Li, R. Jiang, and Z. Xie, "Instance segmentation of anatomical structures in chest radiographs," in *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 2019, pp. 441–446.

[93] I. Ullah, F. Ali, B. Shah, S. El-Sappagh, T. Abuhmed, and S. H. Park, "A deep learning based dual encoder–decoder framework for anatomical structure segmentation in chest x-ray images," *Scientific Reports*, vol. 13, no. 1, p. 791, 2023.

[94] Y.-B. Tang, Y.-X. Tang, J. Xiao, and R. M. Summers, "Xlsor: A robust and accurate lung segmentor on chest x-rays using criss-cross attention and customized radiorealistic abnormalities generation," in *International Conference on Medical Imaging with Deep Learning*. PMLR, 2019, pp. 457–467.

[95] S. Hwang and S. Park, "Accurate lung segmentation via network-wise training of convolutional networks," in *Deep Learning in Medical Image Analysis and Multimodal*

Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3. Springer, 2017, pp. 92–99.

[96] G. Gaál, B. Maga, and A. Lukács, "Attention u-net based adversarial architectures for chest x-ray lung segmentation," *arXiv preprint arXiv:2003.10304*, 2020.

[97] N. Gaggion, L. Mansilla, C. Mosquera, D. H. Milone, and E. Ferrante, "Improving anatomical plausibility in medical image segmentation via hybrid graph neural networks: applications to chest x-ray analysis," *IEEE Transactions on Medical Imaging*, 2022.

[98] H. T. Huynh and V. N. N. Anh, "A deep learning method for lung segmentation on large size chest x-ray image," in *2019 IEEE-RIVF International Conference on Computing and Communication Technologies (RIVF)*. IEEE, 2019, pp. 1–5.

[99] A. Mansoor, J. J. Cerrolaza, G. Perez, E. Biggs, K. Okada, G. Nino, and M. G. Linguraru, "A generic approach to lung field segmentation from chest radiographs using deep space and shape learning," *IEEE Transactions on Biomedical Engineering*, vol. 67, no. 4, pp. 1206–1220, 2019.

[100] H.-J. Chen, S.-J. Ruan, S.-W. Huang, and Y.-T. Peng, "Lung x-ray segmentation using deep convolutional neural networks on contrast-enhanced binarized images," *Mathematics*, vol. 8, no. 4, p. 545, 2020.

[101] J. C. Souza, J. O. B. Diniz, J. L. Ferreira, G. L. F. da Silva, A. C. Silva, and A. C. de Paiva, "An automatic method for lung segmentation and reconstruction in chest x-ray using deep neural networks," *Computer methods and programs in biomedicine*, vol. 177, pp. 285–296, 2019.

[102] L. Bi, D. Feng, and J. Kim, "Dual-path adversarial learning for fully convolutional network (fcn)-based medical image segmentation," *The Visual Computer*, vol. 34, no. 6, pp. 1043–1052, Jun 2018.

[103] F. Munawar, S. Azmat, T. Iqbal, C. Grönlund, and H. Ali, "Segmentation of lungs in chest x-ray image using generative adversarial networks," *Ieee Access*, vol. 8, pp. 153 535–153 545, 2020.

[104] M. Eslami, S. Tabarestani, S. Albarqouni, E. Adeli, N. Navab, and M. Adjouadi, "Image-to-images translation for multi-task organ segmentation and bone suppression

in chest x-ray radiography," *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2553–2565, 2020.

[105] L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, and D. Rueckert, "Drinet for medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2453–2462, Nov 2018.

[106] L. Saidy and C.-C. Lee, "Chest x-ray image segmentation using encoder-decoder convolutional network," in *2018 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW)*. IEEE, 2018, pp. 1–2.

[107] S. Peng, C. Zheng, F. Xu, H. Xiao, H. Nam, and Y. Wu, "Blood vessels segmentation by using cdnet," in *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, June 2018, pp. 305–310.

[108] C. Xiao, J. Liu, X. Chen, H. Han, C. Shu, and Q. Xie, "Deep contextual residual network for electron microscopy image segmentation in connectomics," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, April 2018, pp. 378–381.

[109] V. J. Beltrán-Carrillo, A. Jiménez-Loaisa, M. Alarcón-López, and J. L. Elvira, "Validity of the "samsung health" application to measure steps: A study with two different samsung smartphones," *Journal of sports sciences*, vol. 37, no. 7, pp. 788–794, 2019.

[110] K.-Y. Tai, D.-L. Chiang, T.-S. Chen, V. R. Shen, F. Lai, and F. Y.-S. Lin, "Smart fall prediction for elderly care using iphone and apple watch," *Wireless Personal Communications*, vol. 114, pp. 347–365, 2020.

[111] R. Williams, "Insight health apps," https://www.insighthealthapps.com/pages/support-central-software-downloads, 2023, accessed: 2023-04-07.

[112] J. P. Cohen, P. Bertin, and V. Frappier, "Chester: A web delivered locally computed chest x-ray disease prediction system," *arXiv preprint arXiv:1901.11210*, 2019.

[113] Kexugit, "Using microsoft ai to build a lung-disease prediction model using chest x-ray images," Jul 2018. [Online]. Available: https://learn.microsoft.com/en-ca/archive/blogs/machinelearning/using-microsoft-ai-to-build-a-lung-disease-prediction-model-using-chest-x-ray-images

[114] P. Chondro, C.-Y. Yao, S.-J. Ruan, and L.-C. Chien, "Low order adaptive region growing for lung segmentation on plain chest radiographs," *Neurocomputing*, vol. 275, pp. 1002 – 1011, 2018.

[115] J. Lee, H. Wu, and M. Yuan, "Lung segmentation for chest radiograph by using adaptive active shape models," in *2009 Fifth International Conference on Information Assurance and Security*, vol. 1, Aug 2009, pp. 383–386.

[116] A. Harouni, A. Karargyris, M. Negahdar, D. Beymer, and T. Syeda-Mahmood, "Universal multi-modal deep network for classification and segmentation of medical images," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, April 2018, pp. 872–876.

[117] H. Lu, H. Wang, Q. Zhang, D. Won, and S. W. Yoon, "A dual-tree complex wavelet transform based convolutional neural network for human thyroid medical image segmentation," in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, June 2018, pp. 191–198.

[118] O. Gómez, P. Mesejo, O. Ibáñez, A. Valsecchi, and O. Cordón, "Deep architectures for high-resolution multi-organ chest x-ray image segmentation," *Neural Computing and Applications*, Oct 2019.

[119] J. C. Souza, J. O. B. Diniz, J. L. Ferreira, G. L. F. da Silva, A. C. Silva, and A. C. de Paiva, "An automatic method for lung segmentation and reconstruction in chest x-ray using deep neural networks," *Computer Methods and Programs in Biomedicine*, vol. 177, pp. 285 – 296, 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0169260719303517

[120] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi, "Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules." *AJR. American journal of roentgenology*, vol. 174 1, pp. 71–4, 2000.

[121] B. van Ginneken, M. B. Stegmann, and M. Loog, "Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database," *Medical Image Analysis*, vol. 10, no. 1, pp. 19 – 40, 2006.

[122] W. Yang, Y. Liu, L. Lin, Z. Yun, Z. Lu, Q. Feng, and W. Chen, "Lung field segmentation in chest radiographs from boundary maps by a structured edge detector," *IEEE Journal of Biomedical and Health Informatics*, vol. 22, pp. 842–851, 2018.

[123] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 2018, pp. 7132–7141.

[124] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation'in fully convolutional networks," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 421–429.

[125] M. W. Sjoding, C. R. Cooke, T. J. Iwashyna, and T. P. Hofer, "Acute respiratory distress syndrome measurement error. potential effect on clinical study results," *Annals of the American Thoracic Society*, vol. 13, no. 7, pp. 1123–1128, 2016.

[126] A. E. Johnson, T. J. Pollard, S. J. Berkowitz, N. R. Greenbaum, M. P. Lungren, C.-y. Deng, R. G. Mark, and S. Horng, "Mimic-cxr, a de-identified publicly available database of chest radiographs with free-text reports," *Scientific data*, vol. 6, no. 1, pp. 1–8, 2019.

[127] Z. Chen, Y. Bei, and C. Rudin, "Concept whitening for interpretable image recognition," *Nature Machine Intelligence*, vol. 2, no. 12, pp. 772–782, 2020.

[128] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 618–626.

[129] N. Arun, N. Gaw, P. Singh, K. Chang, M. Aggarwal, B. Chen, K. Hoebel, S. Gupta, J. Patel, M. Gidwani *et al.*, "Assessing the trustworthiness of saliency maps for localizing abnormalities in medical imaging," *Radiology: Artificial Intelligence*, vol. 3, no. 6, 2021.

[130] K. B. Ahmed, G. M. Goldgof, R. Paul, D. B. Goldgof, and L. O. Hall, "Discovery of a generalization gap of convolutional neural networks on covid-19 x-rays classification," *Ieee Access*, vol. 9, pp. 72 970–72 979, 2021.

[131] Q. Zheng, P. Zhao, H. Wang, A. Elhanashi, and S. Saponara, "Fine-grained modulation classification using multi-scale radio transformer with dual-channel representation," *IEEE Communications Letters*, vol. 26, no. 6, pp. 1298–1302, 2022.

[132] K. Zhao, J. Hu, H. Shao, and J. Hu, "Federated multi-source domain adversarial adaptation framework for machinery fault diagnosis with data privacy," *Reliability Engineering & System Safety*, vol. 236, p. 109246, 2023.

[133] K. Zhao, F. Jia, and H. Shao, "A novel conditional weighting transfer wasserstein auto-encoder for rolling bearing fault diagnosis with multi-source domains," *Knowledge-Based Systems*, vol. 262, p. 110203, 2023.

[134] B. Jin, L. Cruz, and N. Gonçalves, "Deep facial diagnosis: deep transfer learning from face recognition to facial diagnosis," *IEEE Access*, vol. 8, pp. 123 649–123 661, 2020.

[135] A. Madani, M. Moradi, A. Karargyris, and T. Syeda-Mahmood, "Semi-supervised learning with generative adversarial networks for chest x-ray classification with ability of data domain adaptation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 1038–1042.

[136] Y.-X. Tang, Y.-B. Tang, M. Han, J. Xiao, and R. M. Summers, "Abnormal chest x-ray identification with generative adversarial one-class classifier," in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 1358–1361.

[137] L. Wang, Z. Q. Lin, and A. 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. 1–12, 2020.

[138] F. Ucar and D. Korkmaz, "Covidiagnosis-net: Deep bayes-squeezenet based diagnosis of the coronavirus disease 2019 (covid-19) from x-ray images," *Medical hypotheses*, vol. 140, p. 109761, 2020.

[139] M. Farooq and A. Hafeez, "Covid-resnet: A deep learning framework for screening of covid19 from radiographs," *arXiv preprint arXiv:2003.14395*, 2020.

[140] E. E.-D. Hemdan, M. A. Shouman, and M. E. Karar, "Covidx-net: A framework of deep learning classifiers to diagnose covid-19 in x-ray images," *arXiv preprint arXiv:2003.11055*, 2020.

[141] Y. Oh, S. Park, and J. C. Ye, "Deep learning covid-19 features on cxr using limited training data sets," *IEEE transactions on medical imaging*, vol. 39, no. 8, pp. 2688–2700, 2020.

[142] E. Çallı, E. Sogancioglu, B. van Ginneken, K. G. van Leeuwen, and K. Murphy, "Deep learning for chest x-ray analysis: A survey," *Medical Image Analysis*, vol. 72, p. 102125, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1361841521001717

[143] Y. M. López-Fernández, L. S. Smith, J. G. Kohne, J. P. Weinman, V. Modesto-Alapont, S. B. Reyes-Dominguez, A. Medina, B. E. Piñeres-Olave, N. Mahieu, M. J. Klein *et al.*, "Prognostic relevance and inter-observer reliability of chest-imaging in pediatric

ards: a pediatric acute respiratory distress incidence and epidemiology (pardie) study," *Intensive care medicine*, vol. 46, no. 7, pp. 1382–1393, 2020.

[144] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, "Two public chest x-ray datasets for computer-aided screening of pulmonary diseases," *Quantitative imaging in medicine and surgery*, vol. 4, no. 6, p. 475, 2014.

[145] S. Mehta, E. Mercan, J. Bartlett, D. Weaver, J. G. Elmore, and L. Shapiro, "Y-net: joint segmentation and classification for diagnosis of breast biopsy images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 893–901.

[146] X. Sun and W. Xu, "Fast implementation of delongs algorithm for comparing the areas under correlated receiver operating characteristic curves," *IEEE Signal Processing Letters*, vol. 21, no. 11, pp. 1389–1393, 2014.

[147] M. M. Islam, T. Hannan, L. Sarker, and Z. Ahmed, "Covid-densenet: A deep learning architecture to detect covid-19 from chest radiology images," in *Proceedings of International Conference on Data Science and Applications*. Singapore: Springer Nature Singapore, 2023, pp. 397–415.

[148] H. Tabrizchi, A. Mosavi, Z. Vamossy, and A. R. Varkonyi-Koczy, "Densely connected convolutional networks (densenet) for diagnosing coronavirus disease (covid-19) from chest x-ray imaging," in *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE, 2021, pp. 1–5.

[149] Y. R. Sedhai, M. Yuan, S. W. Ketcham, I. Co, D. D. Claar, J. I. McSparron, H. C. Prescott, and M. W. Sjoding, "Validating measures of disease severity in acute respiratory distress syndrome," *Annals of the American Thoracic Society*, vol. 18, no. 7, pp. 1211–1218, 2021.

[150] M. W. Sjoding, T. P. Hofer, I. Co, A. Courey, C. R. Cooke, and T. J. Iwashyna, "Interobserver reliability of the berlin ards definition and strategies to improve the reliability of ards diagnosis," *Chest*, vol. 153, no. 2, pp. 361–367, 2018.

[151] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.

[152] S. Towfighi, A. Agarwal, D. Mak, and A. Verma, "Labelling chest x-ray reports using an open-source nlp and ml tool for text data binary classification," 2019. [Online]. Available: https://doi.org/10.1101/19012518

[153] L. Oakden-Rayner, "Exploring the chestxray14 dataset: problems," https://laurenoakdenrayner.com/2017/12/18/the-chestxray14-dataset-problems/, December 2017.

[154] D. Brossier, R. El Taani, M. Sauthier, N. Roumeliotis, G. Emeriaud, and P. Jouvet, "Creating a high-frequency electronic database in the picu: the perpetual patient," *Pediatric critical care medicine*, vol. 19, no. 4, pp. e189–e198, 2018.

[155] S. Taslimi, S. Taslimi, N. Fathi, M. Salehi, and M. H. Rohban, "Swinchex: Multi-label classification on chest x-ray images with transformers," *arXiv preprint arXiv:2206.04246*, 2022.

[156] H. Yu, J. Li, L. Zhang, Y. Cao, X. Yu, and J. Sun, "Design of lung nodules segmentation and recognition algorithm based on deep learning," *BMC bioinformatics*, vol. 22, pp. 1–21, 2021.

[157] C. Rasmussen and Z. Ghahramani, "Occam's razor," *Advances in neural information processing systems*, vol. 13, 2000.

[158] J. G. Richens, C. M. Lee, and S. Johri, "Improving the accuracy of medical diagnosis with causal machine learning," *Nature communications*, vol. 11, no. 1, p. 3923, 2020.