LUNG DISEASES CLASSIFICATION BASED ON MICRO PATTERNS IDENTIFICATION THROUGH ATTENTION GUIDED MECHANISM FROM X-RAY IMAGES

A Thesis submitted to the

UPES

For the Award of
Doctor of Philosophy
in
Computer Science & Engineering
By
Shiva Prasad Koyyada

Sept. 2024

Supervisor
Dr. Hitesh Kumar Sharma

External Supervisor
Dr. Thipendra P Singh

Computer Science & Engineering
School of Computer Science (SOCS)
UPES
Dehradun- 248007: Uttarakhand

LUNG DISEASES CLASSIFICATION BASED ON MICRO PATTERNS IDENTIFICATION THROUGH ATTENTION GUIDED MECHANISM FROM X-RAY IMAGES

A Thesis submitted to the

UPES

For the Award of
Doctor of Philosophy
in
Computer Science & Engineering
By

Shiva Prasad Koyyada

(SAP ID: 500073479)

Sept. 2024

Supervisor
Dr. Hitesh Kumar Sharma
Professor
School of Computer Science
UPES

External Supervisor
Dr. Thipendra P Singh
Professor
School of Computer Science Engineering and Technology
Bennett University

Computer Science & Engineering
School of Computer Science (SOCS)
UPES
Dehradun - 248007: Uttarakhand

DECLARATION

I declare that the thesis entitled Lung Diseases Classification Based on Micro Patterns Identification Through Attention Guided Mechanism From X-Ray Images has been prepared by me under the guidance of Dr. Hitesh Kumar Sharma, Professor of School of Computer Science, UPES. No part of this abstract has formed the basis for the award of any degree or fellowship previously.

Shiva Prasad Koyyada
School of Computer Science (SOCS),
UPES, Dehradun, India
DATE: 19th June 2024

CERTIFICATE

I certify that Shiva Prasad Koyyada has prepared his thesis entitled **"Lung Diseases Classification Based on Micro Patterns Identification Through Attention Guided Mechanism From X-Ray Images"**, for the award of PhD degree of UPES, under my guidance. He has carried out the work at the School of Computer Science, UPES.

Internal Supervisor
Dr. Hitesh Kumar Sharma
School of Computer Science,
UPES, Dehradun, India
Date: 19th June 2024

BENNETT UNIVERSITY
THE TIMES GROUP

June 12, 2024

CERTIFICATE

I certify that Mr. **Shiva Prasad Koyyada** has prepared his thesis entitled **"Lung Diseases Classification Based on Micro Patterns Identification Through Attention Guided Mechanism from X-Ray Images"**, for the award of PhD degree from **University of Petroleum & Energy Studies, Dehradun, Uttarakhand** under my guidance. He has carried out this work at the **School of Computer Science, UPES, Dehradun, Uttarakhand.**

Thipendra.

Dr. Thipendra P Singh

External Supervisor
Professor
School of Computer Science Engineering and Technology,
Bennett University, Greater Noida, UP, India

ABSTRACT

The traditional chest radiograph should be re-evaluated, given the essential role that chest X-ray images (CXRs) play in diagnosing lung diseases. This reassessment is crucial because CXRs provide vital insights into various lung conditions, enabling earlier detection and improved patient outcomes. As advancements in imaging techniques and AI interpretation emerge, updating our understanding of CXRs can enhance diagnostic accuracy and clinical decision-making. Interpreting radiographs presents inherent challenges, such as overlapping anatomical structures and subtle abnormalities that can obscure critical findings. Consequently, highly trained radiologists are essential for minimizing false positives and negatives. Their expertise is vital in navigating these complexities, considering the variability in image quality and the similarities between different conditions.

Radiologists typically rely on their experience to make informed diagnoses by carefully examining CXR images for signs of infection. However, the explanations for their conclusions can often be vague or inconsistent due to the complexities involved in interpreting CXR images. From medical images, convolutional neural networks (CNNs) have proven highly effective in classifying and identifying diseases. While deep learning (DL) models show significant promise, their lack of explainability presents a major challenge to clinical application in the highly regulated healthcare environment. To assist radiologists, a system architecture was developed while addressing DL model lack of explianability. It consit of multiple phases. One phase employs a deep network architecture to identify relevant diseases, while another phase utilizes Explainable Artificial Intelligence (XAI) methods to identify local indicators within the CXR images. This approach not only facilitates the classification of lung diseases, including COVID-19, but also effectively addresses issues related to data imbalance.

By employing weakly supervised learning, we aim to identify significant regions within CXR images, derive interpretation rules, and explain the reasoning behind the DL model's outputs to replicate the decision-making process of radiol-

ogists. The study is structured into six major sections: (1) an examination of the evolution from conventional feature engineering to advanced deep learning techniques; (2) a study of these techniques applied to identify lung diseases such as COVID-19, pneumonia, and tuberculosis; (3) exploration of a custom deep learning network to extract lung features; (4) employing a localization network to extract local indicators from CXR images to mitigate the black box nature of the model; (5) creating an ensemble of localization models; and (6) handling data imbalance through a multi-stage approach while utilizing transfer learning models.

In this work, we contribute by extracting discriminant features using XAI methods. A fusion of regions of interest and other features was performed, leading to our fusion models achieving a mean accuracy of 99.29% and a mean recall of 99.33% on training data, and 97.81%, 98.39% on test data, respectively. We also implemented XAI ensembles that utilize saliency maps, gradient-weighted class activation mapping (Grad-CAM), and local interpretation model-agnostic explanations (LIME) to generate visual explanations for DL models predicting COVID-19 infections. The experimental results demonstrate the robustness of the ensemble approach, with the XAI ensemble achieving an accuracy of 98.85%, compared to 99.62% for individual LIME. Among transfer learning approaches, the XAI-Xception net yielded the best test result with 99.89% accuracy and 99.85% recall. By addressing class imbalance, our custom network achieved 99.61% accuracy on test data.

However, the study does present certain limitations. The dependency on specific imaging datasets may limit the generalizability of the findings to broader clinical populations. Furthermore, the complexity of the proposed models could hinder their adoption in clinical settings without sufficient explainability and validation. Our contributions encompass the generation of XAI annotations and the extraction of features that aid radiologists in their assessments. These XAI annotations will be made available to the research community.

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude for the divine guidance and inner strength that have supported me through this challenging journey Ph.D. Such endeavors require the support and guidance of many, and I am immensely thankful to all who have contributed.

I am deeply grateful for the guidance and mentorship provided by both my supervisors, Dr. Thipendra P. Singh and Dr. Hitesh Kumar Sharma. Dr. Singh's simplicity, honesty, ethics, respect, and love, alongside his inspirational nature to all, have greatly influenced me. His guidance, support, and encouragement throughout my research were instrumental in shaping my work. Equally, Dr. Sharma's fresh perspective and unwavering support in the final stages were invaluable in bringing this thesis to completion. Both of your contributions have been crucial to my success.

I extend my heartfelt appreciation to the faculty and staff of the School of Computer Science at UPES for fostering a supportive academic atmosphere and for their ongoing assistance. I am especially grateful to the administrative staff for their aid in various academic and administrative tasks.

I express my eternal gratitude to my parents, Satyanarayana and Venkatamma, for their enduring love, patience, and unwavering belief in me. To my beloved spouse, Srujana, I am deeply thankful for her boundless love, support, and understanding. Her patience and resilience have been my pillar of strength, especially during the most challenging moments of this journey. I also want to extend my appreciation to my children, Rigved (9 years) and Rishi Kashyap (3 years), whose presence brings immense joy and inspiration to my life.

I express my sincere appreciation to all who have supported me in various roles, contributing to the successful completion of my work.

SHIVA PRASAD KOYYADA

TABLE OF CONTENTS

DECLARATION ii
CERTIFICATE(Internal Supervisor) iii
CERTIFICATE(External Supervisor) iv
ABSTRACT v
ACKNOWLEDGEMENTS vii
LIST OF FIGURES xiii
LIST OF TABLES xv
LIST OF ABBREVIATIONS xvi
Chapter 1: INTRODUCTION **1**
1.1 TYPES OF LUNG DISEASES 1
1.2 MEDICAL IMAGING 2
1.2.1 X-RAY TRANSMISSION 3
1.2.2 COMPUTED TOMOGRAPHY(CT) 3
1.2.3 MAGNETIC RESONANCE IMAGING (MRI) 4
1.2.4 ULTRASOUND IMAGING 4
1.3 EVOLUTION OF TECHNIQUES FOR CHEST DISEASE DETECTION 5
Chapter 2: LITERATURE REVIEW **10**
2.1 LUNG DISEASE DETECTION 10
2.2 IMAGE PREPROCESSING METHODS 12
2.3 LUNG DISEASES 15
2.3.1 COVID-19 15
2.3.2 PNEUMONIA 21

2.3.3 TUBERCULOSIS 24
2.4 RESEARCH GAP ANALYSIS 27
2.5 MATERIALS AND METHODS 29
2.5.1 CONVOLUTIONAL NEURAL NETS (CNN), MACHINE LEARNING, SEGMENTATION 29
2.5.2 TRANSFER LEARNING(TL) METHODS 33
2.5.3 VISION TRANSFORMERS 45
2.5.4 EXPLAINABLE ARTIFICIAL INTELLIGENCE(XAI) IN LUNG DISEASE DETECTION 47
Chapter 3: PROBLEM FORMULATION **55**
3.1 BACKGROUND 55
3.2 MOTIVATION 55
3.3 PROBLEM STATEMENT 56
3.4 OBJECTIVE 56
3.5 DATA 57
Chapter 4: METHODOLOGY **62**
4.1 SUB OBJECTIVE 1 64
4.2 SUB OBJECTIVE 2 68
4.3 SUB OBJECTIVE 3 70
Chapter 5: EXPERIMENTS AND RESULTS **74**
5.1 ARCHITECTURE DETAILS TO GENERATE LOCAL FEATURES 74
5.2 ENSEMBLE METHODS, HANDLING CLASS IMBALANCE . . 78
Chapter 6: DISCUSSION **88**
6.1 INTRODUCTION 88
6.2 TIME COMPLEXITY ANALYSIS 91
6.3 LIMITATIONS OF THE STUDY 93
Chapter 7: CONCLUSION AND FUTURE SCOPE **94**
7.1 CONCLUSION 94
7.2 FUTURE SCOPE 95

REFERENCES 97

List of Figures

1.1 Basic elements of a projection X-ray system 3
1.2 Fourth generation of CT scanner design. 4
1.3 Evolution of CAD and related methods. [5, 6, 7, 8, 9, 10, 11, 12]. . . . 6

2.1 Worldwide research on various diseases from 2000 to 2023 (Date last access: on 25th October 2023, keywords- COVID-19, Pneumonia, Tuberculosis. https://www.sciencedirect.com/search) 11
2.2 Disease classification: various phases and methods 13
2.3 High level classification of image segmentation approaches 13
2.4 Literature review process 27
2.5 Image classification pipeline (ensemble method) 29
2.6 Building blocks of CNN 31
2.7 Lung disease classification-CNNs [133] 32
2.8 Various CNN architectures 33
2.9 Transfer learning method: vgg16 model summary 34
2.10 CNN architecture: AlexNet [12] 35
2.11 ResNet Architecture [32] 37
2.12 DenseNet Architecture [34] 38
2.13 DenseNet Architecture- different versions [34] 38

2.14 EfficientNet Architecture:(a) is a baseline network example; (b)-(d) are conventional scaling that only increases one dimension of network width, depth, or resolution. (e) Efficient net compound scaling method that uniformly scales all three dimensions with a fixed ratio. [138] 39
2.15 Mobilenet Architecture [139] 40
2.16 Inception architecture:Inception modules [30] 42
2.17 CNN architecture:Googlenet [30] 43
2.18 DarkNet Architectures [61, 140] 44
2.19 The Xception architecture [53] 45
2.20 Vision Transformer architecture [142] 46
2.21 Categorization of XAI methods[148] 48
2.22 Examples: Grad-CAM activation maps [150] 50
2.23 Usage of Grad CAM to explain infections[55] 53

3.1 COVID-19 and Normal images from the dataset 58

4.1 Overall system architecture for the proposed approach 63
4.2 Proposed method to extract local indicators 65
4.3 fusion of local and global features 66
4.4 COVID-19 images and their corresponding localization maps along with masks 67
4.5 Proposed approach: A two stage approach to handle class imbalance. 69
4.6 Block diagram for proposed method-Ensemble XAI; DCNN-Deep convolutional neural net;F-DCNN-Fusion DCNN 72

5.1 Experiment set up 75
5.2 Accuracy and loss curves for various image resolutions trained on COVID-19 data set used from table 3.1. 76

5.3 Confusion matrices for various resolution images - global data ((a) to (d) are train, (e) to (h) are test), COVID-19 data set used from Table 3.1. 77
5.4 Layer-wise detail of DCNN model 79
5.5 Error curves-a,b,c are generated while training on the GradCAM fusion dataset; d,e,f are generated while training on the Saliency fusion dataset. 84
5.6 Local features observed from CXR images of Nodule class. 86
5.7 Error curves observed during training lung nodule detection. 87
5.8 Error curve obtained during training-VIT 87
5.9 Confusion matrix recorded on test data-VIT 87

6.1 NORMAL images and their corresponding localization maps along with masks 89
6.2 Decision tree built on top of three PCA components with a depth of three 90
6.3 XAI generated images - a,b,c are original images; d,e,f are saliency maps g,h,i are LIME generated masks;j,k,l are GradCAM output . . 92

7.1 COVID images and their corresponding localization maps along with masks 96

List of Tables

2.1 Different techniques in image pre-processing phase 15

2.2 Overview of papers using Deep learning models from the selected studies.C=COVID-19, P=Pneumonia, Pv=viral Pneumonia, Pb=bacterial Pneumonia, N=Normal. 23

2.3 Overview of selected studies for Tuberculosis detection (TB=Tuberculosis, N=Normal) 26

2.4 Research gap analysis from the existing methods 28

3.1 Data distribution proposed approach (Dataset created by sampling COVID-19 images from [150,171], Normal images from [172]) . . . 57

3.2 Data distribution in COVIDx Dataset 58

3.3 Data sets available for public-COVID-19,Pneumonia,Normal 59

3.4 Datasets available for public for the detection of Tuberculosis 61

5.1 Predictions for various resolution images- TP-True Positive, TN-True Negative, FP-False Positive, FN-False Negative, Acc-Accuracy(%), in train data column: 1-train data,0-test data (Data used: COVID-19,Normal). 78

5.2 Metrics from state-of-the-art methods,Acc-Accuracy,Spec-Specificity 78

5.3 Metrics captured from the models built on XAI generated annotations when combined with original data(COVID-19); In train column, one indicates train data, zero indicates test data,Acc-Accuracy,Prc-Precision 80

5.4 Comparison of metrics - handling class imbalance through an ensemble method which includes customized model and TL methods;In train column one indicates train data, zero indicates test data,(COVID-19 data). 82
5.5 Metrics recorded after applying various TL methods on different XAI generated fusion datasets; In train column one indicates train data, zero indicates test data - An ensemble model(COVID19 data). 83
5.6 Metrics generated from TL methods- on original images of 224*224 resolution; In the train column, one indicates train data, and zero indicates test data(COVID-19 data. 84
5.7 Metrics recorded on fusion dataset with various weight compositions of local and global features trained with deep learning model to predict lung nodules. 85

List of Abbreviations

ACROYNIM	ABBREVIATION
AHE	Adaptive Histogram Equalization
ALE	Accumulated Local Effects
ANNs	Artificial Neural Networks
CAD	Computer-Aided Detection
CAM	Class Activation Mappings
CBAM	Convolutional Block Attention Module
CBIR	Content Based Image Retrieval
ChOA	Chimp Optimization Algorithm
CLAHE	Contrast Limited Adaptive Histogram Equalization
CNN	Convolution Neural Network
COPD	Chronic Obstructive Pulmonary Disease
CT	Computed Tomography
CXR	Chest X-ray
DCNN	Deep Convolution Neural Network
DL	Deep Learning
DT	Decision Tree
ELMs	Extreme Learning Machines
FC	Fully Connected
FD	Fusion Dataset
FDCNN	Fusion Deep Convolution Neural Network
FM	Fusion Model
GRAD-CAM	Gradient Class Activation Mappings
HE	Histogram Equalization
ICE	Individual Conditional Expectation
JPEG	Joint Pixel Export Group
LD	Local Dataset
LIME	Local Interpreatable Model Agnostic Explanations
LRG	Local Region Generator
MDS	Mutually Disjoint Data Sets
ML	Machine Learning
MLP	Multi Layer Perceptron
MRI	Magnetic Resonance Imaging
PCA	Principle Component Analysis
PDP	Partial Dependency Plots
PNG	Portable Network Graphics
RELU	Rectified Linear Unit
RoI	Region of Interest
SARS	Severe Acute Respiratory Syndrome

ACROYNIM	ABBREVIATION
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
TB	Tuberculosis
TL	Transfer Learning
VGG	Visual Geometry Group
ViTs	Vision Transformers
VQA	Visual Question Answering
VT	Vision Transformers
XAI	Explainable Artificial Intelligence
YOLO	You Only Look Once

Chapter 1

INTRODUCTION

Humans are the most precious living beings on earth, and they suffer from various diseases. Major diseases occur due to various conditions. The main reason to suffer from diseases is due to the failure or malfunction of an internal organ in the body. These can be the brain, heart, kidney, liver, and lungs.

Lung diseases are infectious diseases. Any condition with the lungs that prevents them from performing properly is referred as a lung disease. There are three main types of lung diseases

1.1 TYPES OF LUNG DISEASES

Airway diseases: These conditions affect the tubes (airways) that carry gases like oxygen into and out of the lungs. As a result, the airways are commonly constricted or closed. Asthma, chronic obstructive pulmonary disease (COPD), and bronchiectasis are examples of illnesses of the airways.

Lung tissue diseases: These conditions have an impact on the lung tissue's structure. Scarring or tissue inflammation (restrictive lung disease) prevents the lungs from fully expanding. As a result, it is challenging for the lungs to absorb oxygen and release carbon dioxide. They are consequently unable to breathe fully. These diseases include sarcoidosis and pulmonary fibrosis.

Lung circulation diseases: The blood arteries in the lungs become clotted, scarred, or inflamed, which results in several disorders. The condition of the heart is impacted by certain illnesses. Among them is pulmonary hypertension.

The most common lung diseases include:

- Asthma, Collapse of a part or all of the lung (pneumothorax or atelectasis)
- Swelling and inflammation in the main passages (bronchial tubes),
- Chronic obstructive pulmonary disease(COPD), Lung cancer and Lung infection(pneumonia),
- Pulmonary edema (abnormal build-up of fluid in the lungs), pulmonary embolus (Blocked lung artery).

In all these types, patients have breathing problems such as shortness of breath, inability to breathe deeply, and difficult to exhale. Along the same lines, a related pneumonia kind of disease "severe acute respiratory syndrome" (SARS) reported in Guangdong Province, China in late 2002 and actually named as SARS in 2003 [1]. Another severe respiratory disease was reported in Wuhan, Hubei province, China in 2019 and it is termed as 'WH-Human 1' coronavirus is also referred to as '2019-nCoV' [2] later named COVID-19.

With the development of COVID-19, researchers' focus has shifted to different lung conditions, among which include COVID-19, Pneumonia and Lung cancer are few conditions. The reason for COVID-19 and Pneumonia is lung infection. Lung cancer results in the formation of lung nodules, some of which remain unchanged over the time until they develop into malignant lesions. AI system's goal is to identify these diseases by looking at different diagnostic reports. Here the focus is on identifying COVID-19 from Chest X-ray (CXR) images. The further sections talk about different modalities.

1.2 MEDICAL IMAGING

There are various medical imaging techniques available to capture the human body structure and organ tissues in order to determine the abnormalities. The following are the different modalities available [3].

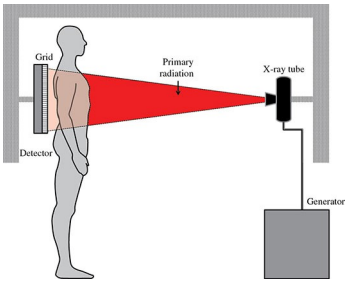


Figure 1.1: Basic elements of a projection X-ray system

1.2.1 X-RAY TRANSMISSION

X-ray transmission imaging is based on measuring the quantity of X-ray radiation that reaches a detector after the initial X-rays have been diminished by objects in their path [4]. The amount of diminishing is the result of the number of X-rays received by objects and scattering caused by the object itself. Since the amount of diminishing is different for various structures, these can be distinguished easily except for the ones that are in the single line of projection. Based on the intensity of X-rays and the proper one to utilize based on the type of body part to be diagnosed. There are different modalities in X-ray transmission such as Radiography, Mammography, and Fluoroscopy. In radiology, the differences in attenuation from body parts or tissues are captured on the detector which was placed just behind the patient's body as shown in the Figure 1.1. It shows the front view of the CXR image [1].

According to the patient's posture and orientation in relation to the X-ray source and detector panel, CXRs are broadly categorized into posteroanterior, anteroposterior, and lateral. In the posteroanterior (PA) and anteroposterior (AP) views, the X-ray source is positioned to the rear or front of the patient, respectively.

1.2.2 COMPUTED TOMOGRAPHY(CT)

With three different orientations for the X-ray beam source and detector, different projection views can be obtained by manipulating the position and direction of the beam in C-arm X-ray imaging, which improves upon the plain film approach. Two-

[1] https://radiologykey.com/projection-x-ray-imaging

dimensional images will be captured at different angles which help to construct a full image using a technique called filtered back projection. Filtered backpropagation and different computer algorithms are used to capture the three-dimensional view and projected it into two-dimensional images of different slices. Essentially CT uses X-rays to project the images onto a film in 360 degrees by rotating clockwise around the patient's body as shown in the Figure 1.2.

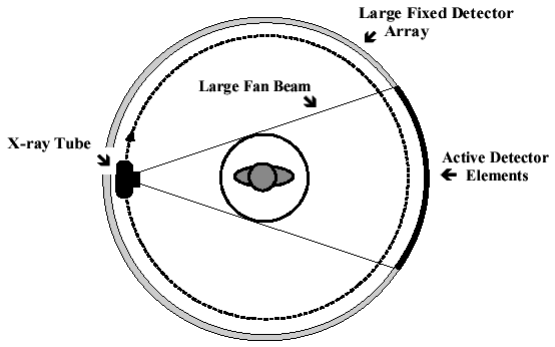


Figure 1.2: Fourth generation of CT scanner design.

1.2.3 MAGNETIC RESONANCE IMAGING (MRI)

Magnetic Resonance Imaging (MRI) is well-known for its ability to provide high-resolution images of soft tissues using magnetic fields and radio waves [4]. Although MRI is less frequently applied to lung disease detection due to inherent challenges like low proton density and motion artifacts, it remains valuable in certain scenarios. Techniques such as hyperpolarized gas MRI and oxygen-enhanced MRI offer detailed assessments of pulmonary conditions, including fibrosis and COPD. However, limitations such as high costs, extended scanning times, and sensitivity to patient movement hinder its regular use for lung disease screening .

1.2.4 ULTRASOUND IMAGING

Ultrasound imaging, commonly known as sonography, is a valuable diagnostic tool that utilizes high-frequency sound waves to generate real-time images of internal structures. In the context of lung examination, sonography is employed primarily

for assessing pleural conditions, such as pleural effusions or thickening, as it can provide information about the presence and extent of fluid accumulation. However, its application in lung pathology is limited compared to other modalities. Ultrasound is less effective for evaluating lung parenchyma due to the difficulty of visualizing air-filled structures. Additionally, operator dependency and limited field of view can hinder its diagnostic accuracy. Thus, while ultrasound has specific applications in lung assessment, it is not a primary imaging modality for comprehensive lung disease evaluation.

In this study, X-ray images are preferred due to their widespread availability, cost-effectiveness, and rapid acquisition, making them a practical choice in clinical settings. Unlike other imaging modalities, such as MRI or CT scans, which can be expensive and time-consuming, X-rays provide immediate insights into lung conditions. They are particularly valuable for diagnosing diseases like COVID-19, where timely intervention is crucial. Furthermore, the existing infrastructure for X-ray imaging in healthcare settings facilitates easier implementation of AI-driven diagnostic tools. This accessibility allows for a broader application of AI techniques in analyzing X-ray images, ultimately enhancing the ability to detect and monitor lung diseases efficiently.

1.3 EVOLUTION OF TECHNIQUES FOR CHEST DISEASE DETECTION

The objective of this research is to study the techniques that help to detect chest disease using the idea of localization. In the process, we are examining and investigating techniques for recognizing chest ailments such as COVID-19, Pneumonia, and Tuberculosis (TB) and how technology has evolved. How the research has evolved in finding the patterns from images over the decades starting from traditional methods to recent state-of-the-art techniques is shown in the Figure 1.3.

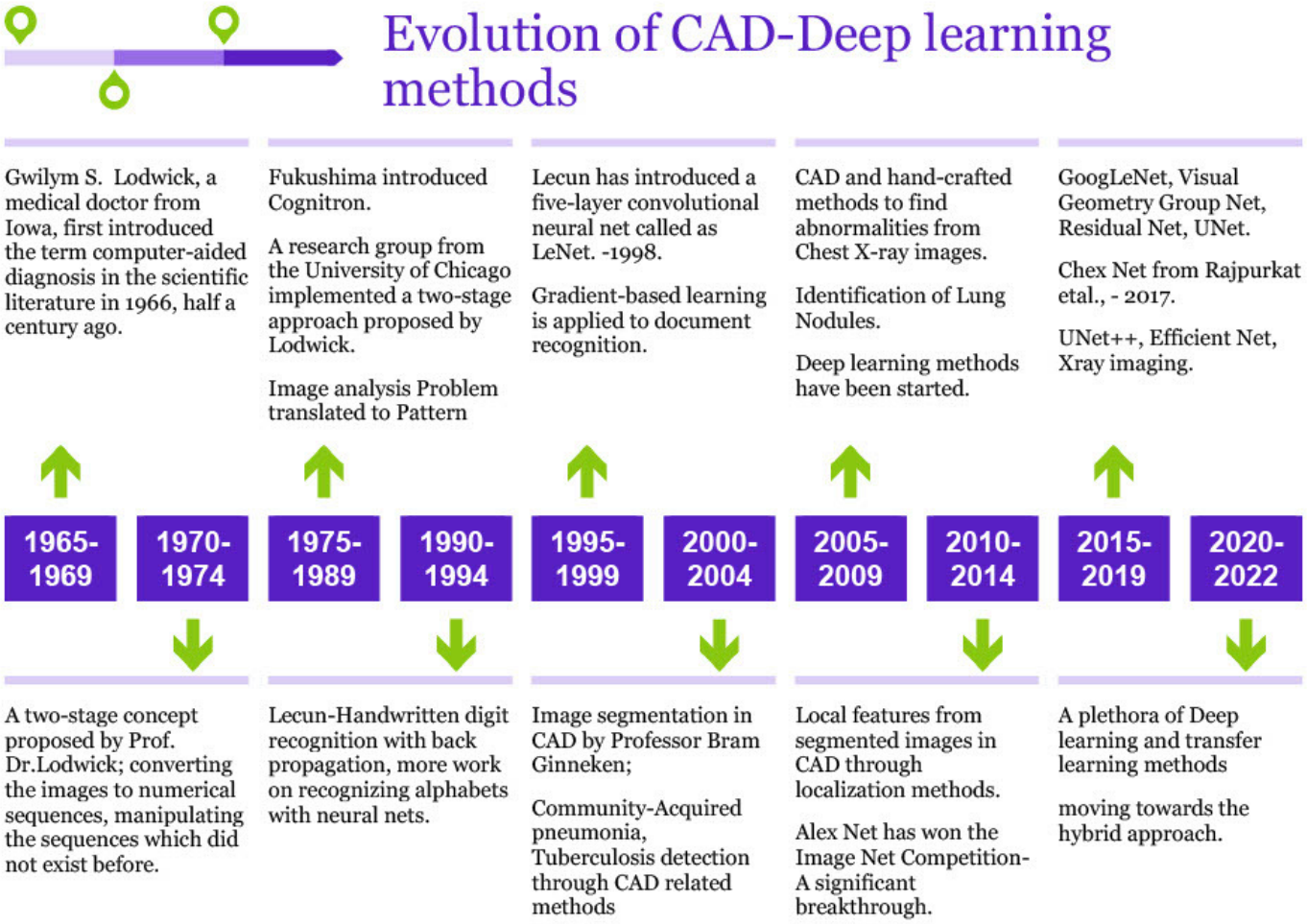


Figure 1.3: Evolution of CAD and related methods. [5, 6, 7, 8, 9, 10, 11, 12]

Despite the use of Artificial Neural Networks (ANNs) to diagnose interstitial lung disorders in 1999 [13], ANNs had never been used to predict pneumonia until 2003 [14]. CXR images were used to predict pneumonia in infants in 2004 [15]; researchers believed that a mix of pattern recognition and clinical experience is the best way to diagnose lung conditions [16]. Image wavelet transform coefficients were predominantly used to generate feature vectors [17], a 15-nearest neighbor algorithm, and distance-dependent weighting were able to identify pneumonia from CXR images. The authors [18] employed ANNs to diagnose pneumonia, and TB from patient epicrisis reports, whereas [19] identified it by decoding the region of interest from CXR images. Texture and shape features from segmented lung fields derived from CXR images used support vector machine (SVM) to classify TB and pneumonia on top of it [20]. In contrast, a fuzzy inference system was introduced to detect TB from CXR images by [9]. A machine learning-based framework where Computer-Aided Detection(CAD) scores from CXR images and clinical features of each subject are considered to identify TB [21]. In contrast, a neural net is employed to identify TB and pneumonia by deriving geometrical features from segmented CXR images [22].

A customized CNN was developed to detect interstitial lung diseases using lung image patches [23] with Adam optimizer, learning rate of 0.001. Feature extraction through Transfer Learning(TL) techniques has began [24, 25, 26, 27, 28, 29] to classify lung diseases. Various TL methods such as AlexNet [12], GoogleNet, InceptionNet [30, 31], ResNet [32], VGG16 [33] and DenseNet [34] are different feature extractors. ChexNet [11] is a 121-dense layer CNN used to identify pneumonia that became the state-of-the-art method and a benchmark model by outperforming average radiologist performance. However, [35, 36] tailored CNN cannot be ignored. The authors [37] have extracted region of interest features through an attention-based mechanism to identify TB from CXR images. Several review papers [38, 39, 40, 41, 42] have given detailed reviews with various deep learning methods from their perspective and compared with human readers, adding more value to state-of-the-art transfer learning techniques. Despite these studies, the dis-

ease has mysterious hidden secrets that need further attention. The studies which detect COVID-19 from various authors include COVIDNet [43], Dark COVID Net [44] inspired by Dark Net-19 [45], COVIDX-Net [46], COVID-ResNet [47] with different scales of input images, Deep COVID [48] employed with SqueezNet [49] and DenseNet, CoVNet-19 [50] is an ensemble of VGG19 and DenseNet121, CoroDet [51] is a custom CNN trained from scratch. Similarly a modified version of Extreme Inception [52, 53], a modified Efficient-Net [54], A VGG-19 with five additional max pooling layers [55], a patch-based CNN with limited set of trainable parameters [56], fuzzy tree-based feature extraction with ensembles by [57, 58] proposed Distance Based Naive Bayes (DBNB), which uses a computational method called Advanced Particle Swarm Optimization (APSO), which extracts the most compelling and relevant features from clinical findings and then employs Naive Bayes, pre-trained models InceptionV3, DenseNet121, and VGG19 were used to extract features and then combined individual predictions using Choquet fuzzy integral to get final labels [59], Apostolopoulos [60] experimented with VGG19, MobileNetv2 [61], Inception, Xceptionv2 [52], InceptionResNet-v2 with different parameters.

CXR images were pre-processed through the fuzzy color technique, stacked with original images to avoid noise [62], fed to a Squeeznet, MobileNet, and classified through SVM. A state-of-the-art fully connected CNN in conjunction with the adversarial critic model Attention U-Net to segment the lungs [63]. At the same time, Rajaraman has focused on reducing network complexity by iterative pruning on an ensemble of deep learning techniques to classify COVID-19 [64]. Fractal methods were used [65] to extract features and then a deep convolution neural net was defined to classify COVID-19.

Lung nodules are tissue lumps in the lungs that are tiny, round, or oval in shape. If you have a lung nodule, you have a more than 40 percent chance of developing lung cancer. Early detection and treatment of lung nodules can greatly enhance the quality of life for patients. Lung nodules are difficult to identify using various medical imaging modalities due to their small size and the interwoven nature of

chest architecture. Lung cancer is the most common disease, and the majority of the early literature focuses on lung nodules. However, there has been a shift in research towards tuberculosis, community-acquired pneumonia, and COVID-19, as the world has experienced or is experiencing these diseases at different times depending on their severity. Computer assisted methods(CAM) are used to determine lung nodules from digital chest images using difference image approach digital morphological processing, feature extraction approaches that use non-linear filters erosion and dilation to reduce the camouflage effect of ribs and vessels [66]. To detect lung cancer, [67] created the directional contrast filter for nodules (DCF-N), which consists of three concentric circles. An automated method developed by [68] for detecting Pneumothorax by detecting curved line patterns using the hough transform. Artificial neural networks were used in the diagnosis of pulmonary nodules from CXR images [69] using thresholding, profile matching analysis, and a three-layer feed-forward network, and [70] double matching method neocognitron [71] as a backbone with different convolutional kernels.

In this work, we focus on identifying COVID-19 from CXR images through the localization process, addressing a significant gap in the application of AI in medical imaging. While AI has shown promise in detecting various lung diseases, there is a lack of methodologies that effectively utilize localization techniques to enhance diagnostic accuracy. Our research objectives aim to explore how these methods can identify crucial regions of interest in CXR images for more accurate COVID-19 diagnosis. The report is structured to support these objectives, starting with a literature review that highlights the limitations of current AI approaches in radiology. This leads to the problem formulation, which details the challenges of diagnosing COVID-19 using traditional methods. Methodology outlines the multi-phase system architecture we developed, integrating deep learning and Explainable AI (XAI) methods then the experiments and results sections present research findings, while the discussion contextualizes these results within the broader AI landscape. Finally, the conclusion and future scope will summarize our contributions and propose directions for further research to proceed in healthcare along with the limitations.

Chapter 2

LITERATURE REVIEW

One of the pioneer experts in the medical field Dr.Ginneken [10] described how machine and deep learning surpassed prior rule-based approaches to become the most common ways to identify lung illness. Deep learning is based on the Convolution Neural Network (CNN), which was initially presented as "Cognitron" by Fukushima [72, 71] and later improved as "Neocognitron", a more vigilant way for finding visual patterns based on geometrical similarities. We can observe from the Figure 2.1 the number of publications over the years and how COVID-19 has pushed the research onto other lung diseases.

2.1 LUNG DISEASE DETECTION

It is known that non-medical researchers' knowledge will not match that of radiologist experts however one doing algorithm analysis on CXR images should have a basic knowledge of Chest anatomy and various abnormalities. These can be studied from chapter one of [73]. Disease detection is primarily defined as a classification problem that consists of various stages as shown below:

i image acquisition.

ii image preprocessing and segmentation.

iii feature extraction.

iv training classifier.

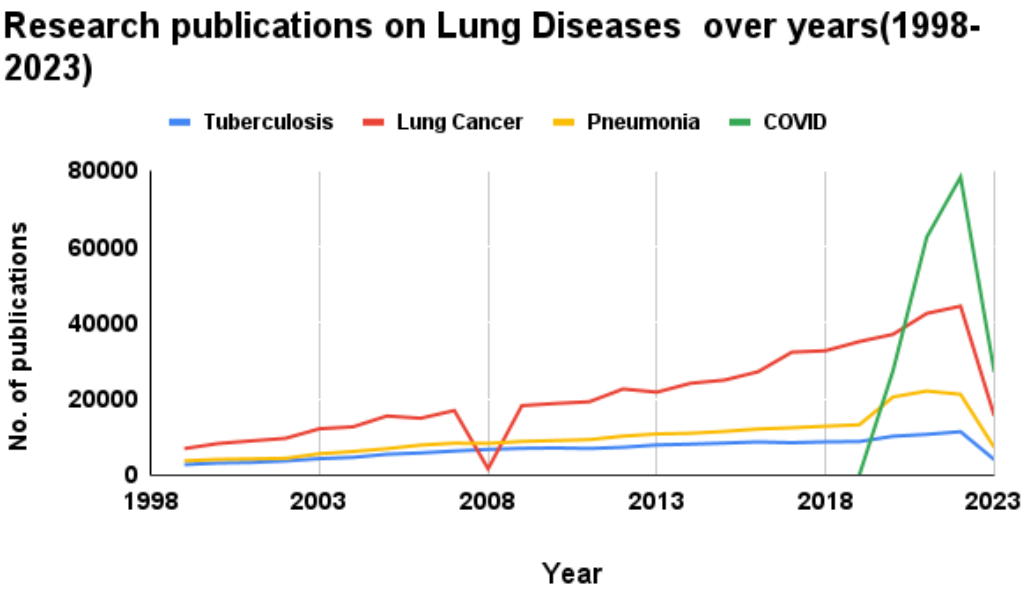


Figure 2.1: Worldwide research on various diseases from 2000 to 2023 (Date last access: on 25th October 2023, keywords- COVID-19, Pneumonia, Tuberculosis. `https://www.sciencedirect.com/search`)

Image acquisition: The electromagnetic spectrum is the primary phenomenon that underpins image acquisition. The most familiar images are those based on radiation from the electromagnetic spectrum, particularly visible light images used in photography. Other electromagnetic spectrum images include radiofrequency (radioastronomy, MRI), microwaves (radar imaging), infrared wavelengths (thermography), X-rays (medical, astronomical, or industrial imaging), and even gamma rays (nuclear medicine, astronomical observations). In addition to electromagnetic imaging, several other modalities are used. Acoustic imaging (using infrasound in geological exploration or ultrasound for echography), electron microscopy, and synthetic (computer-generated) imaging are some of these methods.

Image preprocessing and segmentation: Image preprocessing is much needed step in the pipleine, where all the images may not be in the same size, images may not be captured properly, noise gets added during transmission, sometimes enhancement may improve the results. Segmentation is another preprocessing step that is used to focus on perticular part of image i.e set of pixel values that are

necessary to focus on. Basic segmentation methods are thresholding to advanced methods like U-net++ segmentation. Lung images are segmented commonly by researchers for any lung disease study to capture lung regions.

Feature extraction: After preprocessing features gets extracted from images i.e geometrical features or CNN based features or content based image retrieval methods and transfer learing methods. The number of channels/resolution of images plays major role in extraction of useful features. Features are more useful when there is less noise in the images.

Training classifier: Once feature extraction is done, fully connected layers are used to derive further features and a sigmoid/softmax used to classify the diseases. Support vector machines are the usefull classifiers when more features present in the data.

The methods applied in various phases has been shown in the Figure 2.2.

2.2 IMAGE PREPROCESSING METHODS

Image preprocessing is crucial step before feeding the images to the CNNs. Real time image capturing techniques may lead to various resolutions that need resizing without loosing the quality of pixels, there may be occlusions or noise that needs to addressed with noise removal methods. In additon, image normalization to be done to brought the pixel values to the same range. Most of the literature has included the steps such as resizing, normalization, image enchancement [11, 74, 75, 76] and image augmentation as the preprocessing steps [77] to classify the images further. Mean normalization or z score normalization, image enhancement includes Histogram Equalization(HE) methods. In this work, a global thresholding method Otsu thresholding [78] has been used. Thresholding methods comes under image segmentation as shown in the the Figure 2.3[1].

The Otsu's thresholding algorithm is an automatic global thresholding algorithm identifies thresholding values with the following process:

- It processes the input image 'I'

[1]https://learnopencv.com/otsu-thresholding-with-opencv/

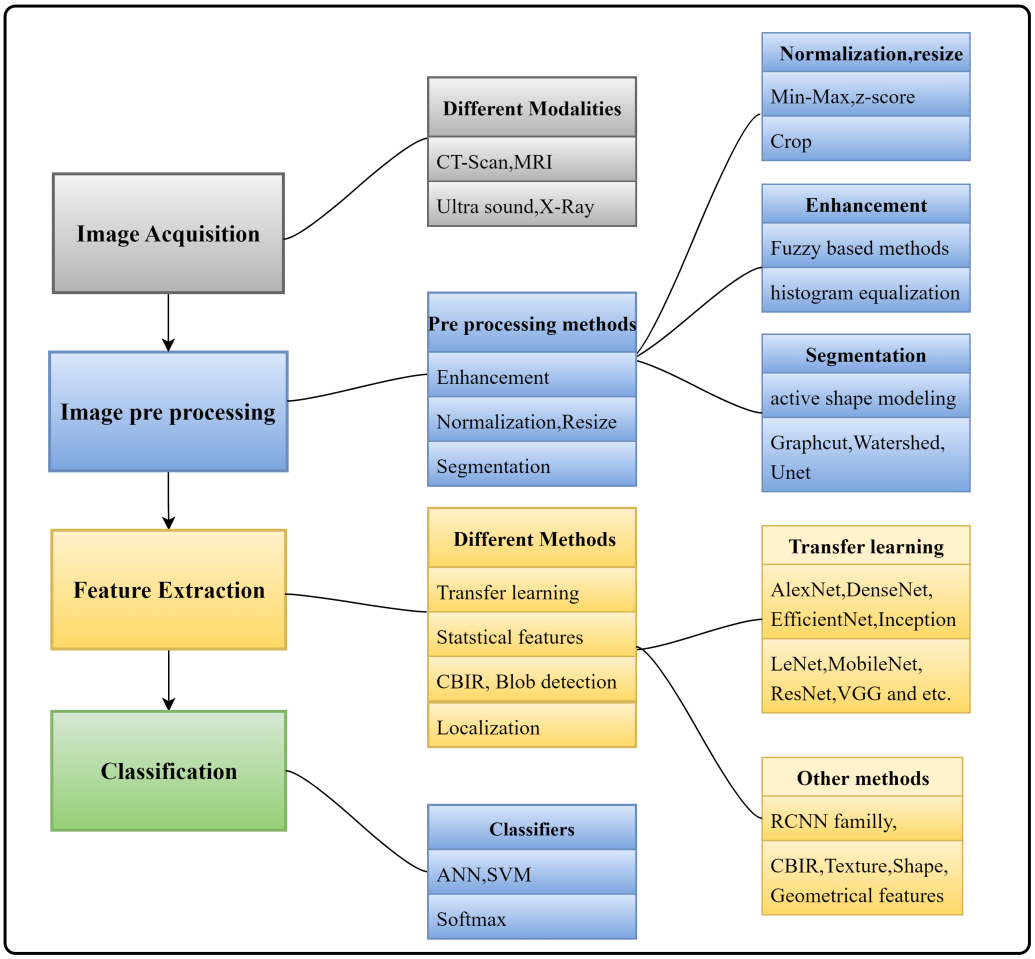


Figure 2.2: Disease classification: various phases and methods

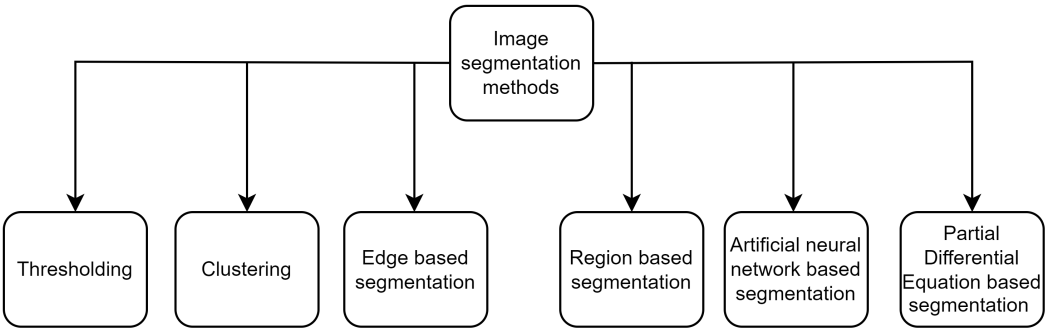


Figure 2.3: High level classification of image segmentation approaches

- Obtain image histogram (distribution of pixels)
- Compute the threshold value T using Otsus' method
- Replace image pixels into white in those regions, where saturation is greater than T and into the black in the opposite cases.

The technique analyzes the histogram of the image and segments the objects by minimizing the variance for every class. For bimodal images, this strategy typically yields the desired outcomes. Two distinct peaks, signifying varying intensity value ranges, can be seen in the histogram of this photograph. Otsu method has two options to find the threshold. The first is to minimize within-class variance defined above $\sigma_w^2(t)$, the second is to maximize between-class variance using the expression below: $\sigma_b^2(t) = w_1(t)w_2(t)[\mu_1(t) - \mu_2(t)]^2$, where μ_i is a mean of class i. w1,w2 are the weight of background and foreground intencity values for the corresponding variances [78]. The between-class variance maximization is done in the following way:

- calculate the histogram and intensity level probabilities
- initialize $w_i(0), \mu_i(0)$
- iterate over possible thresholds: t = 0,..., max_intensity
- update the values of w_i, μ_i, where w_i is a probability and μ_i is a mean of class i
- calculate the between-class variance value $[\sigma_b^2(t) = \sum_{i=1}^{t} P(i) \sum_{i=t+1}^{I} P(i)[\mu_1(t) - \mu_2(t)]^2]$ where probability P is calculated for each pixel value in two separated clusters C_1, C_2 using the cluster probability functions.
- the final threshold is the maximum $\sigma_b^2(t)$ value

Importance of image processing methods is better explained in [79]. Various types of noise artifacts typically seen in radiographs include Poisson, Gaussian, salt, and pepper noise, especially when acquired in huge quantities from public

domain sources such as the internet [80]. The need for such images is growing in part because eliminating one form of noise can sometimes affect the other. General pre-processing involves enhancement, and subtraction [81]. A brief description of the pre-processing stage is shown in the Table 2.1.

Preprocessing	**Most popular methods**
Lung Filed segmentation	Active shape modeling, Graph cut algorithm, Intensity-based threshold methods, Image Matching, Rule-based methods to detect Lung contours, Water shed method, Separation of background and foreground, Unet, UNet++ Architectures
Image enhancement	Histogram based methods (AHE, CLAHE), Fuzzy color techniques, and Filtering with Morphological operators.
Resizing, Normalization	Min-Max, cropping
Extracting RoI from Lung Fields	blob detection/ localization attention-based methods

Table 2.1: Different techniques in image pre-processing phase

Preprocessing steps such as image resizing, contrast enhancement, segmentation using Otsu's thresholding are applied to improve the quality of input images and extracting roi from lung fields using attention based methods are employed in this study.

2.3 LUNG DISEASES

2.3.1 COVID-19

Since the first case of COVID-19 in 2019, so much research has taken place worldwide. We have selected heterogeneous articles published by Elsevier, Google Scholar, IEEE, and other leading publishers. Researchers used a variety of state-of-the-art image processing approaches, ranging from hand-crafted techniques to the most cutting-edge transfer learning, and adversarial networks used to create synthesized images.

COVID-Net [82] is presented as one of the first AI architectures for detecting COVID-19. Thereafter a slew of studies has sprung up based on the COVIDx data set and diverse sampling methodologies. The COVID-Net architecture em-

ploys lightweight residual projection-expansion-projection-extension (PEPX) design patterns composed of many stages of projections with 1*1 convolutions and 3*3 depth-wise convolutions. According to the authors' understanding, such a tailored, lightweight, machine-driven design pattern has never been seen before. COVID-Net also has architectural diversity in terms of long-range connectivity, with kernel sizes ranging from 1*1 to 7*7. COVID-Net trained for 22 epochs with a batch size of 64, a learning rate of 0.0002, and a patience of five.

Several architectures were trained to detect COVID-19. For example [44], implemented Dark COVIDNet, a 17-layer architecture inspired by DarkNet-19 [61] with a flattened and a dense layer, softmax to classify three classes of COVID-19, pneumonia, and healthy lungs. The network trained for 100 epochs to improve the results and observed higher loss at the start of training due to fewer samples in the COVID-19 class. Five-fold cross-validation was done to avoid the problem of overfitting. Similarly, [83] applied a variety of transfer learning networks and concluded that DenseNet201 [34] produced the best results.

Apostolopulos employed VGG19 and MobileNet-V2 as feature extractors with the rectified linear unit as an activation function at hidden layers [60]. Khan has experimented with the Xception network [52] to classify COVID-19, pneumonia bacterial, pneumonia viral, and normal with a batch size of 10, trained for 80 epochs generated an output vector of shape 5*5*2048 for each sample of CXR image [53]. The authors of [47] leveraged the usage of ResNet-50 by introducing the network in three different stages with input images of shapes 128*128*3, 224*224*3, 229*229*3, where the first two stages divided into head and body trained for (3,5), (3,5) epochs, and last stage for 25 epochs respectively with discriminating learning rate [84]. The authors have aimed for a better generalization using progressive resizing with different input size images and reported a hundred percent recall on the COVID-19 class. COVID-CX Net has been developed by [74] with a backbone network DenseNet 121; it applied various image enhancement techniques such as Histogram Equalization (HE), Adaptive Histogram Equalization (AHE), and Contrast-Limited AHEs(CLAHE). It is inspired by ChexNet [11] and modified

as per the requirement of binary classification with a fully connected layer of 10 nodes followed by a drop out of 0.2 to prevent overfitting with sigmoid as an activation function in the last layer. An SVM classifier has been used by extracting features from multiple CNN models to classify the disease [85]. The ResNet50 model provided discriminant features that helped COVID-19 detection as per their study.

The Authors [62] have worked on three classes, namely coronavirus, pneumonia, and normal X-ray imagery; enhanced the images through affine transformations, i.e., a Fuzzy Colour technique [86]. After that, stacked the output image with the original image through Yotam's code[2], extracted features through SqueezNet [49], MobileNetV2 and used linear SVM classifier.

Efficient Net family [45] is well known to take care of depth, width, and resolution dimensions while training the model. The family of models takes the inputs from 224*224*3, 240*240*3, 260*260*3, 300*300*3, 380*380*3, and 456*456*3 by adding a fully connected layer and an output layer with softmax. Perhaps the authors [54] utilized swish [87] as an activation function, resulting in a smooth curve throughout the minimization loss process using a gradient descent technique.

FC Dense Net101 has been used to segment the lung regions from CXR images [56] and proposed a patch-based network, say 224*224 pixels cropped randomly from the image at various instances and fed to the network ResNet-18 with Adam optimizer [88], the initial learning rate of 0.0001 and trained for 100 epochs. Authors have investigated potential biomarkers in the CXR and found that the globally distributed localized intensity variation can be discriminatory for COVID-19.

The two-dimensional curvelet transformation used by [89] subdivides the input using a linear and weighted mixture of fundamental functions called curvelets. A curvelet is a set of radial and angular windows specified in the polar coordinate system in the frequency domain. A Fast Digital Curvelet Transform(FDCT) is implemented via wrapping to synthesize data from RGB image to grayscale and then fed to Efficient Net-B0 [45]. Leverage the usage of different transfer learning

[2]https://github.com/yig/imagestack/blob/master/imagestack.py

techniques including ResNet18,ResNet50,SqueezNet and DenseNet-121 on smaller datasets to identify COVID19 [48] . Models were trained for 100 epochs, with a batch size of 20, Adam optimizer with a learning rate of 0.0001, and image input of 224*224. As per the author's knowledge, CovidGAN [90] is the first paper so far to generate synthesized CXR images using Generative Adversarial Networks. With CovidGAN, authors could generate normal CXR and COVID +ve images by training with a batch size of 64, a learning rate of 0.0002, and a beta of 0.5 for 2000 epochs.

Another network experimented with DenseNet-121 by [91] with various combinations of parameters such as loss functions, optimizers, several epochs, learning rates, and reported good metrics with the Ada max optimizer with Cross-Entropy loss function and Step LR scheduler. An ensemble of Inception[31], VGG19, and DenseNet121 has been implemented by the authors of [59]. Further, the classifiers' predictions are aggregated using the Choquet Fuzzy integral method [92], which will give weight to individual classifiers based on the calculated fuzzy scores. Similarly, The authors of [50] proposed CoVNet-19, an ensemble of VGG19 and DenseNet-121 feature extractors, and SVM as a meta-learner. In contrast, [93] employed various multi-kernel-size spatial channel attention(MKSC) modules to extract feature maps further to classify COVID-19. COVID-RENet-1 and COVID-RENet-2 [94] introduced an averaging layer, ensuring region and edge-based operations between convolution blocks. These two networks differ in their architecture; however, they achieved the same score. A hybrid learning approach [95] consists of CNN and Recurrent Neural Nets(RNNs). Each step has a length of 64 in RNN, totaling 700. However, there is a lack of information on how RNNs have been used to extract features and fuse with CNN to call it a hybrid model. Interestingly [96] authors have focused on real-time detection by leveraging LeNet-5, Extreme Learning Machines (ELMs) [97], Chimp optimization Algorithm(ChoA) [98]. LeNet-5 is used as a feature extractor after being trained on a large dataset and provided to ELM input, and ChoA will take care of the rest. ChoA is designed to overcome the problem of slow convergence speed and getting trapped in local minima. This

makes the ELM network stable and reliable to operate in real-time operation.

Tahir proposed a systematic and consistent approach [99] for lung segmentation and COVID-19 localization with infection quantification. Lung segmentation is done by training 3000 annotated images through U-Net [100] , U-Net++. Image segmentation through watershed approach, i.e., separating foreground from background and fed to a transfer learning algorithm to extract features and for further classification using the weighted sum metric of various models [101]. Authors have given [102] importance to preprocessing of images before actually using several transfer learning techniques from VGG16 to SqueezNet. Image enhancement methods such as AHE and CLAHE made the image tidier. A content-based image retrieval (CBIR) model, as presented by [103], employs multi-similarity loss [104] along with an attention mechanism and a sophisticated mining sampling technique to find the best embedding space. Optimized embedding space is nothing but the low dimensional feature space learned through ResNet50 as a backbone and, in addition, a spatial attention module employed to extract local embeddings to provide additional guidance. A lightweight convolutional model is trained on Gaussian blurred images proposed by [105]. The architecture consists of four convolutional layers with 32,32,64,64 filters respectively, and each kernel size is 3*3. A Fully connected layer of 64 neurons with a drop out of 0.5 and a final layer with a softmax activation function.

In contrast to all the approaches proposed [106] has shown similar accuracy by completely masking lungs from the CXR image, which poses questions on the deep learning approaches. However, it is worth mentioning that the authors have proposed new protocols to automate the detection process of COVID-19. Some of the papers' observations and feature extractors are shown in the Table 2.2.

The authors have employed [107] the DarkCOVID Net model, formulated as binary and three class problem, trained on more than 10,000 CXR images, and achieved an average accuracy of 99.53,94.18 for binary and multi-classification respectively. DarkCOVID Net uses DarkNet-19 as a base model, which is constructed using YOLO(you look only once) real-time object detection system. Images are of

256*256 pixels fed to the network and trained for 100 epochs for binary and 50 epochs for the multiclass problem; A Concat CNN [75] was developed to detect COVID-19, viral pneumonia, and regular infections. The authors have used four CNN models as feature extractors and concatenated the feature maps for better efficiency of the network. Class imbalance is addressed by sampling an equal number(500) of images from three classes. The feature extractor is defined in terms of a number of filters in two convolutional layers with 32,64 combinations with max pooling of 2*2, dropout of 0.5, and activation of relu is used to bridge the layers. These results were compared with state-of-the-art models VGG16, Inceptionv3, ResNet50, and DenseNet121 with five-fold cross-validation. The authors [108] focused on three classes, COVID-19, Viral pneumonia, and Healthy images with 219,1341,1345 images in each class repectively. The authors have trained four pre-trained deep neural networks to find the best net. AlexNet gave good results among the other applied models, VGG16, MobileNetv2, and ResNet18.

A Deep CNN based technique (Focus Covid) [109] is proposed for COVID-19 prediction using chest radiographs. It is a modified version of Focus Net where some layers were removed, reducing the number of parameters (16,32,64,128,256). It has helped to reduce training time by reducing the no. of parameters. A lightweight CNN architecture [110] is for detecting COVID-19 disease, which is robust—a comparison study performed between transfer learning and shallow CNN. A total of 2541 samples were considered from two public databases consisting of Normal and COVID-infected images. Other transfer learning models are Inception-v2, Xception, MobileNet, and DenseNet201. The shallow six-layer network trained for different batch sizes 16,20,25,32,50,64 has a learning of 0.001 for 20 epochs.

A trained output-based transfer learning(TOTL) [111] approach for COVID-19 detection from CXRs. Pre-processing methods such as noising, contrasting, and segmentation were applied before feeding the image into pre-trained transfer learning models such as Inception, Xception, MobileNet, ResNet, and VGG. Here the features extracted from pre-trained models are fed to a shallow model consisting of 64,128,256,128,64 neurons with relu activation function and a dropout rate of 0.2.

A total of 18 models were implemented, [112] and their performance was evaluated. Major voting built on top of 18 models, including the top four models with above 93 percent accuracy. Two certified radiologist analyzed the image outputs generated by Grad-CAM, and their decision resemblance with Sqeeznet output. The performance of the CNN classifier can be improved using the nature-inspired optimization algorithm Hill Climbing(CNN-HCA) [113] by enhancing the CNN model's parameters. After evaluating the present state, the hill climbing algorithm is a local search optimization technique exploring superior solutions among neighborhoods. However, this algorithm works for two hyperparameters, such as kernel size and the number of neurons in the first dense layer, which certainly adds more parameters as it evolves. A multi-level image segmentation method [114] based on the swarm intelligence algorithm (SIA) to enhance the image segmentation of COVID-19 X-rays. Ant colony optimization was introduced later, direction crossover was used to enhance the convergence speed of the algorithm. Directional mutation strategy helps to jump out of local optima. This helps to find the right threshold value to segment the CXR image. A deep two-step learning (DL) architecture Multi COVIDNet [115] to detect COVID-19. The uniqueness of this paper introduces an optimization algorithm called "Multi-Objective Grasshopper Optimization Algorithm(MOGOA)" to optimize the DL network layers; the Grasshopper optimization Algorithm(GOA) can balance between exploration and exploitation. The nature-inspired swarming nature of GrossHooper inspires it. It has generated multiple solutions, picking the best one using the Pareto Optimality(PO) operator.

2.3.2 PNEUMONIA

U-Net Architecture was utilized to segment the lung field from CXR by [116] and employed ResNet-50, InceptionV3, InceptionResNetV2 architectures using Adam, SGD optimizers and a batch size of 16,32 to produce the best results to detect pneumonia. Some of the papers' observations and feature extractors are shown in Table 2.2.

Author	Feature Extractor	Observation
[47] COVID-ResNet	ResNet-50	Progressive resizing(initially trained with 128*128 and then larger size images) with discriminative learning rate - 4(C,Pb,Pv,N)-
[46] COVIDX-Net	VGG-19,Dense Net	Input images of size 224*224 without image augmentation trained with Stochastic gradient optimizer-two class (C,N)
[43] COVID Net	DenseNet-121, ResNet-18, SqueezNet	A machine driven design exploration strategy to generate deep learning architecture - 3 class(C,P,N)
[60]	MobileNet V2, VGG-19	Parameters used ReLU,Dropout,Adam optimizer with batch size 64- 4 class(224-C,700-P,504-N)
[62]	MobileNetv2, SqueezeNet	Fuzzy color technique applied to input image, Social Mimic optimization on output vector; Support Vector Machine with stochastic gradient descent optimization. 3 class(295-C,97-P,65-N)
[44] Dark COVID Net	inspired from DarkNet	Three Dark layers each consist of 1 convolution followed by Batch Norm and Leaky ReLU ; every Dark layer is separated with three convolutional layers. - 3 class(125-C,500-P,500-N)
[56]	FC DenseNet67,FC DenseNet103	Segmentation, random patch cropping; sailency maps for interpretability. - 5 class(180-C,54-P(b),20-P(v),57-TB,191-N)
[74]	DenseNet-121	Image enhancement techniques(HE,AHE,AHEA)- 3 class(428-C,3200-N)
[51]	New architecture- 22 layer CNN	Training from scratch leaky RELU – 50 epochs. 3 class(2843-C,1439-P,3108-N)

Table 2.2 (Continued on next page)

Author	Feature Extractor	Observation
[95]	Hybrid approach (CNN+RNN)	RNN designed input signal divided into 700 steps- 64 units per step. 3 class(2000-C,2000-P,2000-N).
[117]	Ensemble of VGG-16, VGG-19, and Inception-v3	Ensemble of transfer learning models with Majority voting Averaging weighted averaging on segmented CXRs(Unet). 3 class(27-C,242-P,234-N).
[118]	Custom architecture based on residual units	Two Convolutions followed by Residual and Shifter units and stacked for different resolution images. 4 class (305-C, 2780 P(b),1493 P(v),1583 N).
[103]	New architecture CBIR(ResNet-50 as backbone)	Multi smilarity loss + Spatial attention module. 3 class(3746-C,5641-P,8064-N)
[119]	ResNet-152,DenseNet-169	Discriminative fine-tuning, cyclical learning rate, and momentum- accuracy achieved in 20 epochs.- 2 class(107-C,112-P).
[54]	Efficient Net Family	Images of different resolutios to the family of models; activation function : swish[87] - 3 class(183-C , 5521-P ,8066-N)

Table 2.2: Overview of papers using Deep learning models from the selected studies.C=COVID-19, P=Pneumonia, Pv=viral Pneumonia, Pb=bacterial Pneumonia, N=Normal.

Experimentation has been done [120] with an extreme version of inception (Xception), VGG16 as feature extractors and concatenated a dense output layer with two neurons, softmax activation function, and trained for 50 epochs. Discriminative localization [35] exhibited using Class Activation Maps and solved the block-box nature of deep learning models to an extent using modified VGG16 on pediatric CXR images. Rajaraman et. al employed an algorithm [35] based on Anatomical atlases for the auto-detection of lung borders. Parameters tuned using the grid search method stopped at a learning rate of 0.0004, momentum 0.99, L2 regularization of 0.000001 in VGG16. Using characteristics extracted from [121], an unsupervised Fuzzy c-means classification learning method was applied to classify pneumonia disease into five groupings. In contrast, [19] encoded the region of interest into a vector of wavelet texture measures and derived statistical-based features to detect pneumonia from CXR images, and [17] have used other wavelet texture measures to extract features for the same task. Thorax disease classification [122] approach is a three-branch attention-guided CNN (AG-CNN). Pneumonia detection was achieved through ResNet-50 and DenseNet121. CheXLocNet [43] segmented CXR images using Mask R-CNN [123] and achieved better localization.

2.3.3 TUBERCULOSIS

ResNet101, VGG19, and DenseNet201 were employed [124] independently to extract features from the given data set without further fine-tuning. These features were extracted from the last convolutional block before the global average pooling layer. After the last convolutional block, the output shape of ResNet101,VGG19, DenseNet201 for a sample image are (7,7,2048) , (7,7,512),(7,7,1920) respectively. These vectors further flattened and passed to the XGBoost classifier to classify the TB. An ensemble of pre-trained models put forth by [117] having Inception-V3, InceptionResNet-V2, DenseNet-121 with an average stacking have shown better results. In addition, a three-layer custom CNN was built with 64,128,256 neurons in respective layers. It is trained for 100 epochs with a kernel size of 5*5 in each layer. A custom architecture of four CNN layers with 32,32,32,64 neurons, respec-

tively is designed by [125]. It is followed by a flattened and fully connected layer of 128 neurons. The model was trained for 200 epochs with a kernel of 3*3, activation function relu, and batch size of 32. A Convolutional Block Attention Module (CBAM) was proposed by [37], placed at every bottleneck of the residual network. CBAM will give more weight to the essential features and extract complex, in-depth features as training progresses. CBAM aims to ensure nonlinear interaction with an emphasis on multiple channel features. For the first time, a mix of demographic data and deep learning characteristics is applied by [28]. A deep CNN architecture [126] consists of 1,1,3,1 layers in each block of CNN-Max pooling, two fully connected layers, and a dropout layer towards the end. The authors stated that getting optimized weights with the said architecture is challenging, and they moved to transfer learning approaches for the better results. On the other hand, [21] has combined the CAD score and clinical features of patients to detect TB through ML. The authors have used AlexNet, Google Net, VGGNet, and ResNet as feature extractors with a meta learner for classification [24, 25, 26]. AlexNet, and GoogleNet are trained with a learning rate of 0.01,0.001 using a stochastic gradient descent optimizer, momentum of 0.9, and exponential decay of 0.002.

A reliable automated CXR image-based screening system [127] for detecting pulmonary diseases with a significant focus on TB. The approach mainly focused on localization to see lung boundaries from CXR. The feature sets were derived from segmented lung fields through object-detected methods, CBIR methods, and standard MATLAB region proposals.

The Lung fields were segmented from CXR images using a graph cut technique [128]. Object detection-inspired features such as shape and texture descriptors were extracted using various histogram methods such as the Intensity Histogram (IH), Gradient Magnitude Histogram(GM), Shape Descriptor Histogram(SD), and Curvature Descriptor Histogram(CD).Using CBIR algorithms, low-level characteristics like texture, intensity, edge, and form moment were retrieved, yielding a total of 594 features. An SVM classifier is trained to classify feature sets into normal and abnormal. Table 2.3 gives an overview of selected papers.

Author	Feature extractor	Observation
[129]	ResNet101,VGG19, DenseNet201(Best)	XGBoost classifier on extracted features. 2 classes (3500 -TB,3000 - Normal).
[117]	Inception-V3,InceptionResNet-V2, DenseNet-121	Ensemble & Stacking on top of learned features, 2 classes (17833 -TB,8851 - N).
[24]	Alexnet, GoogleNet	Ensemble of probability scores by a different weighted average of the probability scores. 2 class(DS1:58-TB,80-Normal,DS2: 336-TB,326-Normal).
[25]	GoogleNet,VGG,ResNet	Three proposals; individual feature extractor+SVM classifier; Bag of CNN features and Ensemble of all predictions. DS1:58-TB,80-Normal,DS2:336-TB,326-Normal.
[9]	structured data(chest pain, age, fever, loss of appetite, loss of weight, smoke addiction, bcg)	An Adaptive Neuro-fuzzy Inference System (ANFIS) is an Artificial Neural Network (ANN) approach that is functionally equivalent to a first-order Sugeno-style Fuzzy Inference System (FIS); a six-layer network.
[126]	Custom Deep CNN inspired by AlexNet with fewer hidden nodes.	AlexNet and custom CNN were used with a softmax at the end with two classes.
[130]	Gabor, gist, Histogram of gradient, Pyramid histogram of gradient features(PHOG) features	Feature extraction with out segmentation; Localized features frequency range, block, and an SVM classifier. 2 class(78-TB,78-Normal).
[131]	Statistical features through image histograms: mean, standard deviation, skewness, kurtosis, and entropy.	Principal component analysis on extracted features and Minimal distance classifier to differentiate TB and Non-TB. 2 class(30-TB,30-Normal)

Table 2.3: Overview of selected studies for Tuberculosis detection (TB=Tuberculosis, N=Normal)

A CXR image enhancement algorithm is proposed by [132]. It consists of segmentation and enhancement as sub-steps. Image segmentation is achieved by applying a sequence of membership and fuzzy distance-based operations on one-dimensional function and image enhancement through the fuzzy-based contrast enhancement technique.

2.4 RESEARCH GAP ANALYSIS

As discussed in the previous section, the literature review focuses on three diseases, although our work concentrates on one lung disease: COVID-19. The literature review has been conducted in a manner that emphasizes lung diseases and the methodologies used, as shown in Figure 2.4.

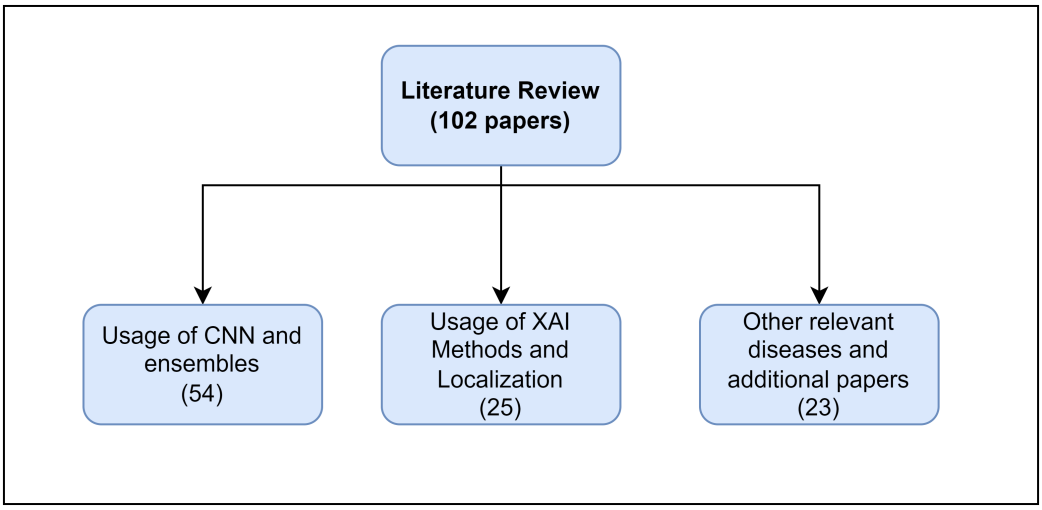


Figure 2.4: Literature review process

Since the research focus is on COVID-19, a thorough gap analysi study was conducted and consolidated the summary and gaps in the Table 2.4 from the selected studies.

Reference	Summary	Research gap
[43]	The authors introduced COVID-Net, a deep convolutional neural network design tailored for the detection of COVID-19 cases from CXR images that is open source and available to the general public and also introduced COVIDx, an open-access benchmark dataset that authors generated comprising of 13,975 CXR images across 13,870 patient patient cases,	Improve sensitivity to COVID-19 infections as new data is getting collected, scope to explore local lesions in continuation.
[46]	COVIDX-Net: explored seven models; no custom network was built explored the transfer learning techniques individually. As author mentioned MobileNet can be improved further to get better results.	scope to include localization
[47]	Authors have taken COVIDx Net dataset and worked on ResNet model fine tuned to get good recall in the datasets. Preprocessing methods used are data augmentation and compared results with COVID Net and COVIDx net.	Data sampling methods and preprocessing methods are limited.
[50]	The proposed model is a 2-leveled stacked ensemble ML model. A stacked ensemble model combines multiple classification models to form a heterogeneous combination that can interpret the same data in different ways.VGG19 and Dense121.	Models were trained on 2416 images only, train 100% , test 99.71% accuracy. Model needs to generalized. need explainability.
[51]	A 22 layer custom architecture was built and trained on two, three,four class of diseases. Leaky Relu is the activation function used.	Traditional model, highly dependent on CNN architecture to extract features. It has scored 99.12 average accuracy and 95.3 recall.
[53]	CoroNet is an Xception architecture to classify the COVID19 images with few extra layers and named it as Exception net.	Pre trained Transfer Learning Architecture. It can not be generalized. As per authors CoroNet still needs clinical study and testing but with higher accuracy and sensitivity for COVID-19 cases.
[59]	Pre-trained models InceptionV3, DenseNet121, and VGG19 were used to extract features and then combined individual predictions using Choquet fuzzy integral to get final labels.	Three models are ensemble producing 99% accuracy however it has not focused on local features.
[60]	Authors have mainly focused on transfer learning techniques such as VGG19,MobileNet V2, Inception, Xception, Inception ResNet v2.	Networks are trained on less than 1000 images. Focused only on global features, no localization was done .98.75% highest accuracy for VGG16.
[105]	A lightweight architecture that classifies COVID-19 from normal images with 4 cnn layers, 32,32,64,64 filters, each filter of 3*3 size	It has scored 92.7 % accuracy on test data when compared to 98.78% on train data - model is overfitting - there is a scope to improve further.
[107]	DarkCOVIDNet was built inspired from DarkNet. Replaced top two layers and fine tuned the model	Traditional object detection model achieved 99.53% average accuracy.
[109]	Focus COVID is a deep convolutional neural network-proposed for the COVID-19 detection using chest radiographs- An extension of Focus Net- Segmentation.	It has scored 96% average accuracy on multi fold data. No of images used only 500.
[110]	Robust deep learning-based system for reliably detecting COVID-19 from chest X-ray images. First, we evaluate the performance of various pre-trained deep learning models (InceptionV3, Xception, MobileNetV2, NasNet and DenseNet201) recently proposed for medical image classification. Second, a lightweight shallow convolutional neural network (CNN) architecture is proposed for classifying X-ray images of a patient with a low false-negative rate.	Shallow CNN provides a maximum accuracy of 99.68, 2541 images were used for training.
[115]	Multi COVIDNet is a two-step Deep learning (DL) architecture has been proposed for COVID-19 diagnosis using CXR. The proposed DL architecture consists of two stages, “feature extraction and classification”.	98.21% accuracy on 900 images of training, no explanation for images
[133]	The authors have used squeezenet with few modifications and named it as COV Diagnosis Net. It is a three class classifier which has three stages. a.offline augmentation of images to prepare the dataset with a uniform distribution. b. training with Bayse Squeezenet c. Decision making of the net with the testing phase.	Though the class imbalance is handled through data augmentation. In general data augmentation is a solution to overfitting of networks instead of increasing the size of data.

Table 2.4: Research gap analysis from the existing methods

2.5 MATERIALS AND METHODS

2.5.1 CONVOLUTIONAL NEURAL NETS (CNN), MACHINE LEARNING, SEGMENTATION

A CNN is employed to collect feature vectors during feature extraction phase, as shown in the Figure 2.5. Deep CNNs comprise a convolutional, max pooling, dense layer with parameters such as the number of kernels, kernel size, hidden and output activation functions, dropout rate, and the number of neurons. Different model architectures have been formed with various combinations. Finding the proper architecture is a laborious process that demands more data and optimization of hyperparameters such as batch size, learning rate, momentum, number of epochs, and batch normalization.

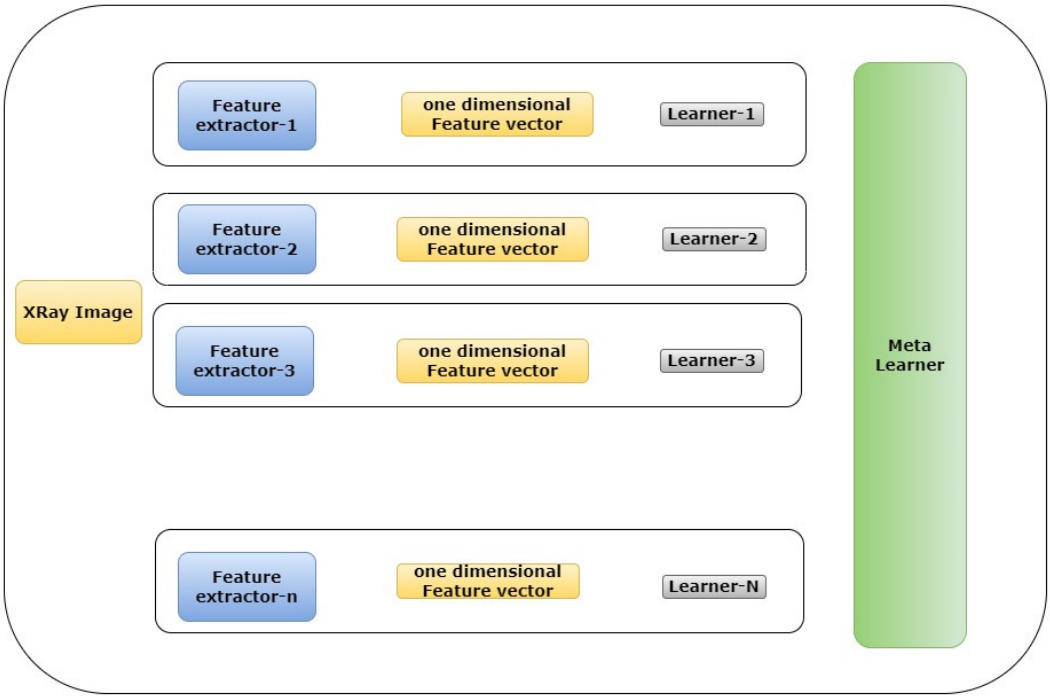


Figure 2.5: Image classification pipeline (ensemble method)

A CNN network is categorized as shallow or deep based on the number of layers. A customizable CNN is a deep network trained on a specific task. A CNN network consist of various building blocks. These are connected sequentially and

passes the image from one block to other block and finally a feature vector gets created for every image. These blocks consist of convolutional, activation, max-pooling, drop out layers. These gets repeated depending on the type of architecture, a flattened layer comes towards the end.

- Convolutional block consist of certain number of filters increases as network goes in deep. A filter is a two dimensional matrix of numbers that does element wise multiplication with the pixel values of image and slides through the image, produces spatial features. These are called feature maps. The feature map size may be same as input feature map size or it may reduces depending on the user's choice of padding.

- Activation block converts the feature values to non linear form by using various activation functions. For example, rectified linear unit that makes any values less than zero as zero and no change if it is positive number. This overcomes the problem of vanishing gradients. Other choices are tanh, leaky relu and etc.

- Maxpooling block does subsampling, i.e it reduces the feature map size by two. For example given a 16*16 feature map, a max pool of two uses 2*2 window and gets the maximum value out of all four values and slides for two pixel values towards the right and repeats the same task. It produces 8*8 feature map. It solves the problem of high number of parameters.

- Drop out is an optonal block uses to solve the problem of overfitting. Dropout rate determines how many neurons to fire while training and it make sures only a certain number of neurons active for every epoch.

- Flatten is layer that converts two dimension feature map to one dimensional feature vector. For example a 8*8 two dimensional map is converted to 1*16 feature vector.

The Figure 2.6 illustrates how this feature extraction process differs from typical neural net construction. These are elements of a standard layer of CNN. These

levels are referred to by two standard term sets. The convolutional net is seen in Figure 2.6(a) as a limited number of relatively sophisticated layers, with numerous ”stages” in each layer. In Figure 2.6(b), the convolutional net is perceived as a multiplicity of basic layers, with each processing stage being considered a separate layer. It follows that not all ”layers” have parameters.

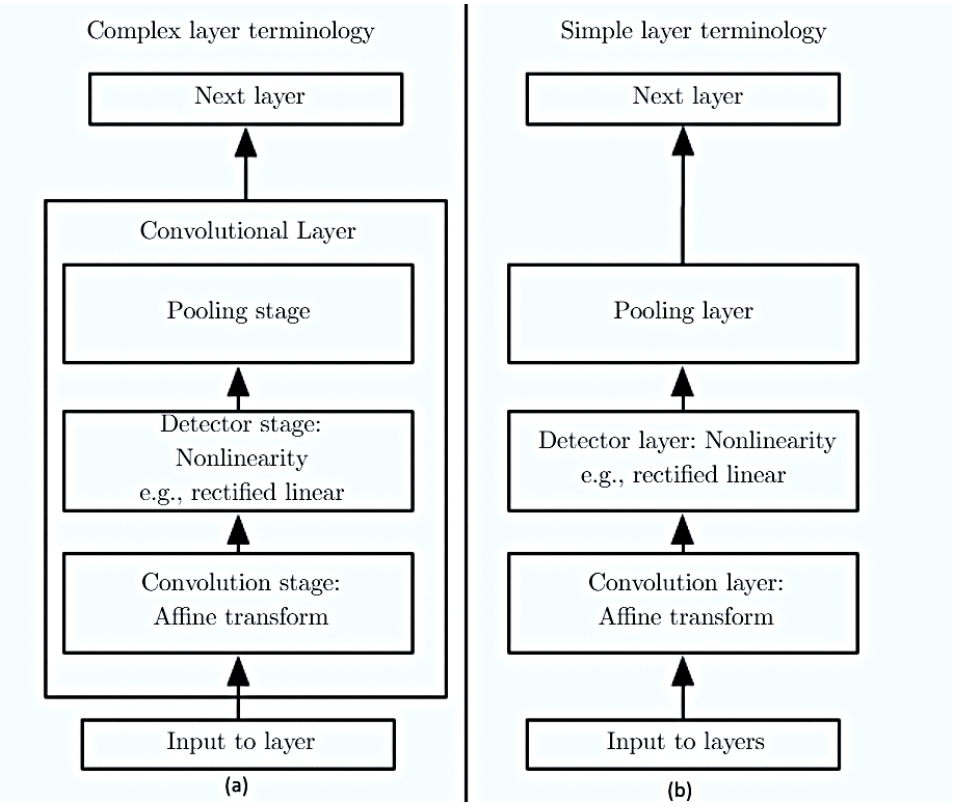


Figure 2.6: Building blocks of CNN

Pooling contributes to the representation being roughly invariant to modest translations of the input in both variants. When an input is translated by a modest amount, the majority of the pooled outputs remain unchanged, a property known as invariance to translation. These blocks let one to create several CNN architectures, as Figure 2.8 illustrates. Unlike the chain topologies presented here for simplicity, real convolutional networks also frequently feature large degrees of branching. A fixed image size is processed by a convolutional network, as seen in Figure 2.8(a). The tensor for the convolutional feature map is altered to smooth out the spatial dimensions after a few layers of switching between convolutional and pooling. A

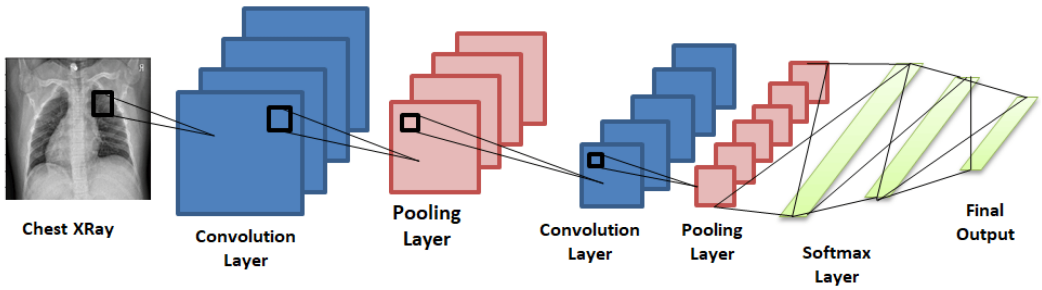


Figure 2.7: Lung disease classification-CNNs [134]

standard feedforward network classifier makes up the remainder of the network. CNNs play a crucial role in lung infection detection using CXR images. The basic architecture of CNN is shown in the Figure 2.7.

A convolutional network retains a completely connected portion even after processing an image of varying sizes, as seen in Figure 2.8(b). This network provides a fixed-size vector of 576 units to the completely connected component of the network by using a pooling operation with variable pool sizes but a fixed number of pools. As depicted in Figure 2.8(c), a convolutional network lacks a fully connected weight layer. Instead, one feature map for each class is produced by the final convolutional layer. Presumably, the model learns a map representing the likelihood of each class occurring at each spatial point. A feature map's average reduced to a single value serves as the uppermost softmax classifier's argument.

Machine learning models such as logistic regression, k-nearest neighbor, SVM and ensembles such as bagging,boosting (Adaboost, XGboost) can be trained on structured data. Structured data could be clinical features such as age, patient habits, smoking (y/n), diabetes (y/n), or texture, shape, intensity, color, statistical, and other features of CXR images. Different geometrical, machine learning, and ensemble methodologies utilized to classify CXRs were discussed in detail in [80]. Various CXR segmentation methods edge detection [135], Active shape modeling [136] , and modern methods such as U-Net [100], U-Net++ [137, 138, 63, 101] could be used for segmentation.

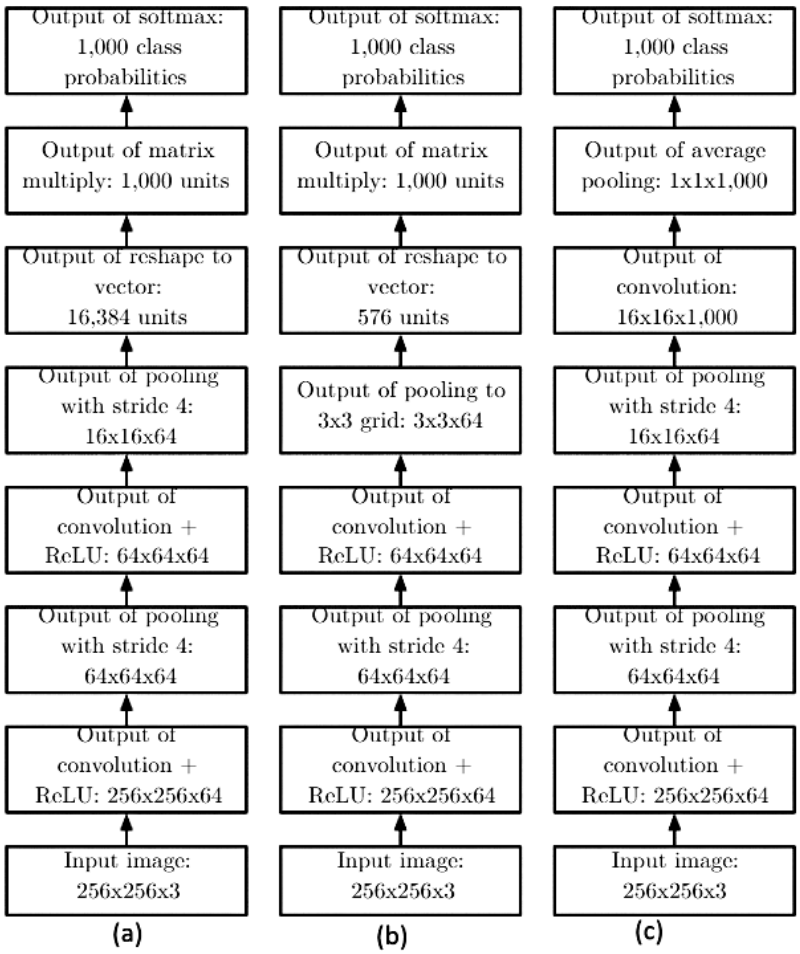


Figure 2.8: Various CNN architectures

2.5.2 TRANSFER LEARNING(TL) METHODS

A Deep CNN constructed and trained on an extensive data set and achieved excellent results could be reused. The caveat here is that the original network may have been trained on different data. These networks extract feature vectors and then build a meta-learner. These networks are known as Transfer Learning techniques (TL). Since 2015, many transfer learning methods have been developed, including Alex Net, Dense Net, Efficient Net, GoogleNet, Inception, Mobile Net, ResNet, Squeeze Net, VGGNet, and Xception.

AlexNet, ResNet, VGGNet, and GoogleNet are the most often used TL models in medical image analysis [42], while SVM is the most commonly used meta-

```
Model: "vgg16"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_1 (InputLayer)        [(None, 224, 224, 3)]     0

 block1_conv1 (Conv2D)       (None, 224, 224, 64)      1792

 block1_conv2 (Conv2D)       (None, 224, 224, 64)      36928

 block1_pool (MaxPooling2D)  (None, 112, 112, 64)      0

 block2_conv1 (Conv2D)       (None, 112, 112, 128)     73856

 block2_conv2 (Conv2D)       (None, 112, 112, 128)     147584

 block2_pool (MaxPooling2D)  (None, 56, 56, 128)       0

 block3_conv1 (Conv2D)       (None, 56, 56, 256)       295168

 block3_conv2 (Conv2D)       (None, 56, 56, 256)       590080

 block3_conv3 (Conv2D)       (None, 56, 56, 256)       590080

 block3_pool (MaxPooling2D)  (None, 28, 28, 256)       0

 block4_conv1 (Conv2D)       (None, 28, 28, 512)       1180160

 block4_conv2 (Conv2D)       (None, 28, 28, 512)       2359808

 block4_conv3 (Conv2D)       (None, 28, 28, 512)       2359808

 block4_pool (MaxPooling2D)  (None, 14, 14, 512)       0

 block5_conv1 (Conv2D)       (None, 14, 14, 512)       2359808

 block5_conv2 (Conv2D)       (None, 14, 14, 512)       2359808

 block5_conv3 (Conv2D)       (None, 14, 14, 512)       2359808

 block5_pool (MaxPooling2D)  (None, 7, 7, 512)         0

 flatten (Flatten)           (None, 25088)             0

 fc1 (Dense)                 (None, 4096)              102764544

 fc2 (Dense)                 (None, 4096)              16781312

 predictions (Dense)         (None, 1000)              4097000

=================================================================
Total params: 138,357,544
Trainable params: 138,357,544
Non-trainable params: 0
```

Figure 2.9: Transfer learning method: vgg16 model summary

learner. For example, feature vectors are taken from the flattened layer (None, 25088) or the fully connected layers(fc1 or fc2) (None, 4096) as illustrated in the Figure 2.9 might be fed into a softmax or SVM.

AlexNet [12] contains eight learned layers, five convolutional and three fully-connected as shown in the Figure 2.10. A 1000-way softmax receives the output from the final fully-connected layer and generates a distribution across the 1000 class labels. By maximizing the average of training samples of accurate labels under the prediction distribution, the net maximizes the multinomial logistic regression objective. Only the kernel maps from the preceding layer are connected to the kernels of the second, fourth, and fifth convolutional layers. Every kernel map in the second layer is connected to the kernels of the third convolutional layer. Every neuron in the layer above is connected to every other neuron in the completely connected layers. The first and second convolutional layers are followed by response-normalization layers. Max-pooling layers come after the fifth convolutional layer and both response-normalization layers. Every convolutional and fully connected layer's output is subjected to the relu non-linearity. It has 60 million parameters that maximizes the problem of overfitting. To handle such cases data augmentation and drop out are the two mechanisms introduced in the network.

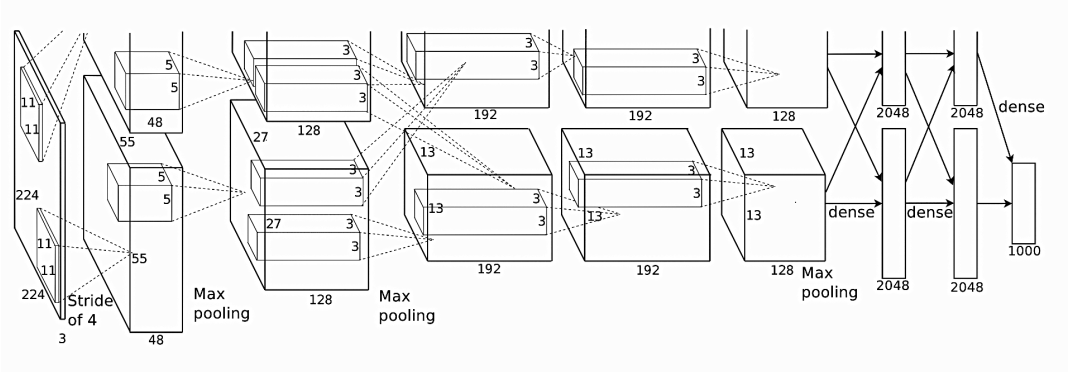


Figure 2.10: CNN architecture: AlexNet [12]

The VGG nets concept served as inspiration for ResNet [32]. Originally intended to be a simple network, as seen in Figure 2.11(b) convolutional layers typically use 3x3 filters and adhere to two straightforward design principles:

i The number of filters in each layer is the same for the size of the output feature map.

ii If the size of the feature map is half, then the number of filters must be doubled in order to preserve the temporal complexity per layer.

It directly uses convolutional layers with a stride of two to conduct down sampling. A global average pooling layer and a 1000-way fully-connected layer with softmax complete the network.. There are 34 weighted layers in all. As demonstrated in Figure 2.11(c), shortcut connections are added based on the plain net, converting the network into its corresponding residual version. In cases where the dimensions of the input and output are same, the identity shortcuts y = F(x, W_i) + x can be utilized directly (solid line shortcuts in 2.11(c)). It has taken into consideration two possibilities when the dimensions expand (dotted line shortcuts in 2.11(c)):

i Identity mapping is still carried out by the shortcut, with additional zero entries padded for larger dimensions. This option adds no additional parameters;

ii 1×1 convolutions are employed to match dimensions by using the projection shortcut in y = F(x, W_i + W_s * x).

For both scenarios, the shortcuts are executed across feature maps of varying sizes using a stride of two.

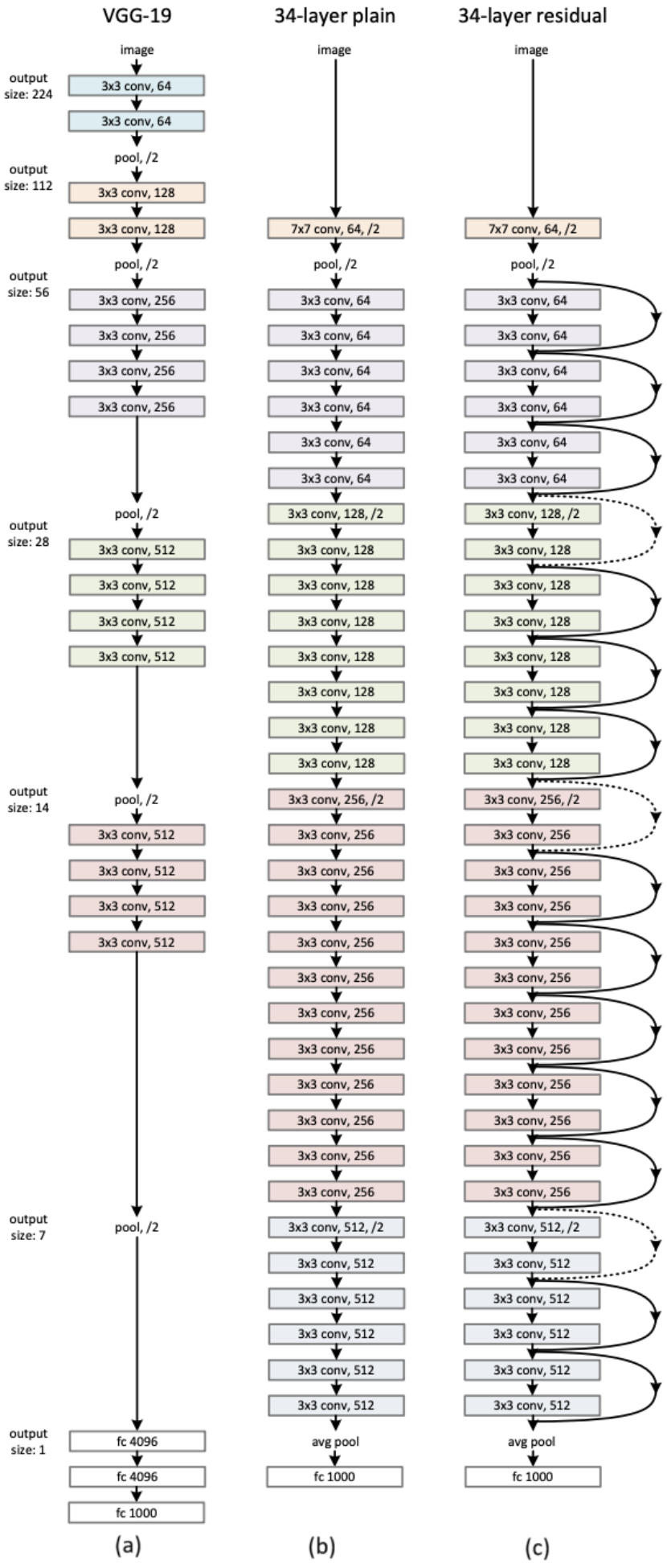


Figure 2.11: ResNet Architecture [32]

DenseNet [34] is again another TL model that is trained on imagenet by showing difference in the connection between layers. It has introduced direct connections from any layer to all subsequent layers as shown in the Figure 2.12 illustrates the layout of consequently, the n^{th} layer recieves the the feature-maps of all preceding layers, $[x_0, ..., x_{n-1}]$, as input: $x_n = H_n(x_0, x_1.....x_{n-1})$ where $[x_0, x_1, ...x_{n-1}]$ refers to the concatenation of feature maps produced in layers 0....n-1.

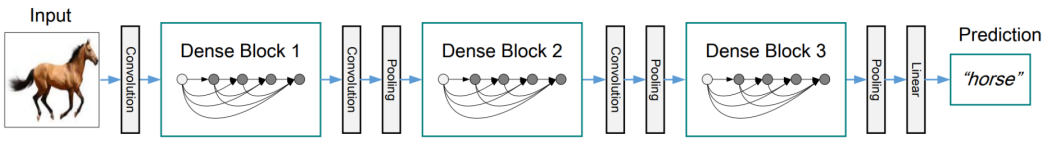


Figure 2.12: DenseNet Architecture [34]

In the Figure 2.12 Dense blocks refers to the sequence of convolutional blocks connected with varying filter sizes. It is available in diffferent versions(121,169,201 and 264 layers) as shown in the Figure 2.13.

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6	[1 × 1 conv; 3 × 3 conv] × 6
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12	[1 × 1 conv; 3 × 3 conv] × 12
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	[1 × 1 conv; 3 × 3 conv] × 24	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 48	[1 × 1 conv; 3 × 3 conv] × 64
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	[1 × 1 conv; 3 × 3 conv] × 16	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 32	[1 × 1 conv; 3 × 3 conv] × 48
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

Figure 2.13: DenseNet Architecture- different versions [34]

Efficient Net [139] is an idea of increasing/decreasing the depth of the existing network. i.e scaling up and down, wider scaling, deeper,resolution scaling,compund scaling as shown in the Figure 2.14.

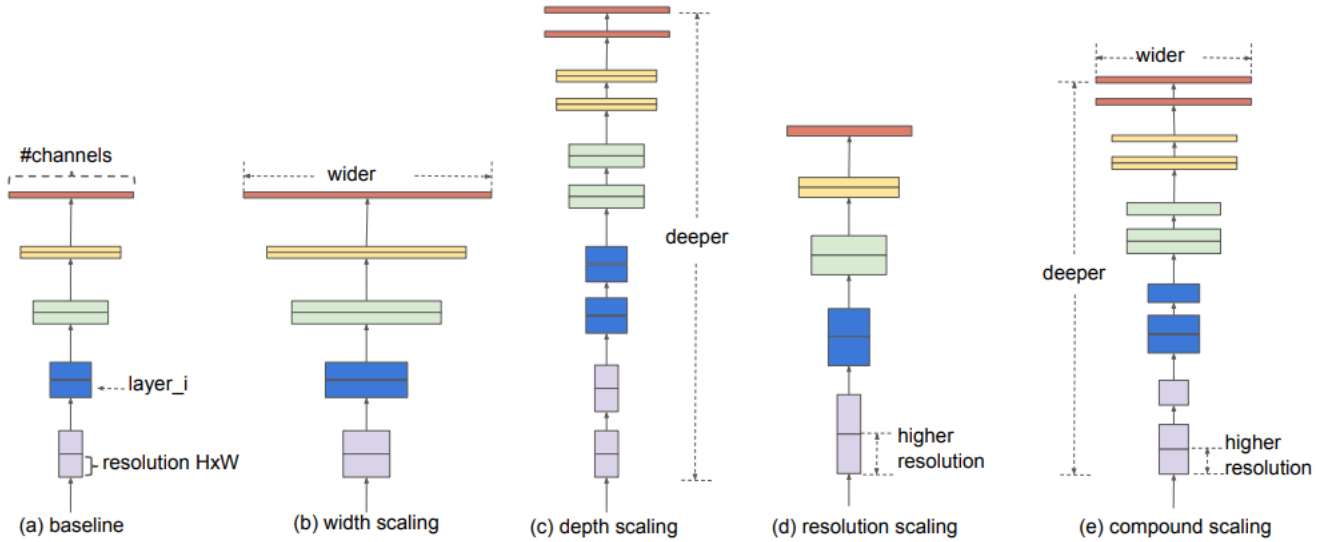


Figure 2.14: EfficientNet Architecture:(a) is a baseline network example; (b)-(d) are conventional scaling that only increases one dimension of network width, depth, or resolution. (e) Efficient net compound scaling method that uniformly scales all three dimensions with a fixed ratio. [139]

One can scale up/down existing ResNet architecture/some other architecture, however there is no clear understanding of suitability of network for a given input image of size w*h. This forces neural architecture to search for a new baseline network and grow it up to produce the EfficientNets family of models. EfficientNet-B7 achieves state-of-the-art 84.3% top-1 accuracy on ImageNet out of all EfficientNet-B0 to B7.

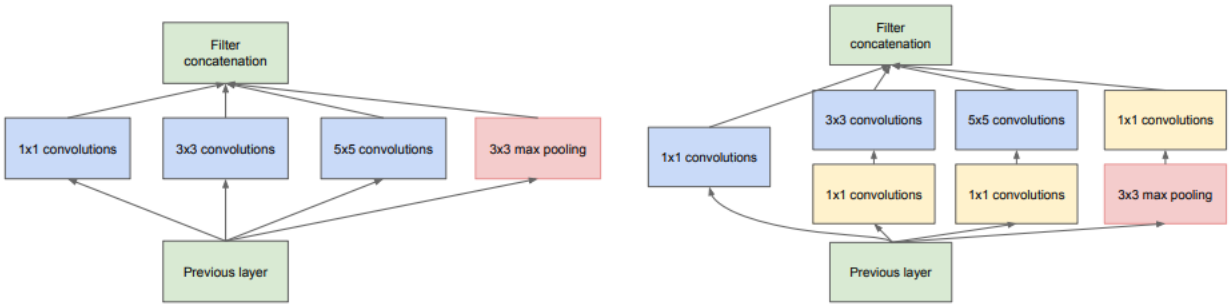
MobileNets [140] are class of EfficientNet models that are light weight in nature built by depthwise seperable convolutions. Depthwise seperabel convolutions are the core layers for MobileNets. This type of factorized convolution breaks down a regular convolution into two different types: a depthwise convolution and a pointwise convolution, which is a 1×1 convolution. The depthwise convolution for MobileNets gives each input channel a single filter. The outputs of the depthwise convolution are then combined by the pointwise convolution using a 1×1 convolution. In a single step, a conventional convolution filter and mixes inputs to create a new set of outputs. This is divided into two layers by the depthwise separable convolution: a filtering layer and a combining layer. The result of this factorization is a significant reduction in computation and model size. A representation of the model architecture is provided in Figure 2.15.

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
$5\times$ Conv dw / s1	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×1000	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$

Figure 2.15: Mobilenet Architecture [140]

GoogleNet is [30] an Inception network where Inception modules are stacked on top of each other. Inception module consist of 1*1,2*2,3*3,5*5 convolutional kernels connected depthwise between layers as shown in the Figure 2.16(a). This causes the increase in number of outputs between layers. It is reduced by a dimensionality reductions and projections. That is, 1×1 convolutions are used to compute reductions before the expensive 3×3 and 5×5 convolutions as shown in the Figure 2.16(b). Besides being used as reductions, they also include the use of rectified linear activation which makes them dual-purpose.

The entire architecture is shown in the Figure 2.17. Propagating gradients back through all the layers in an efficient manner was a challenge, considering the somewhat deep depth of the network. We would anticipate that by including auxiliary classifiers linked to these intermediate layers, we would give more regularization, boost the gradient signal that is transmitted back, and promote discrimination in the classifier's lower stages. They add their loss, which was weighted by 0.3 for the auxiliary classifier losses, to the network's overall loss during training. These backup networks are eliminated at inference time.



(a) Inception module, naïve version

(b) Inception module with dimension reductions

Figure 2.16: Inception architecture:Inception modules [30]

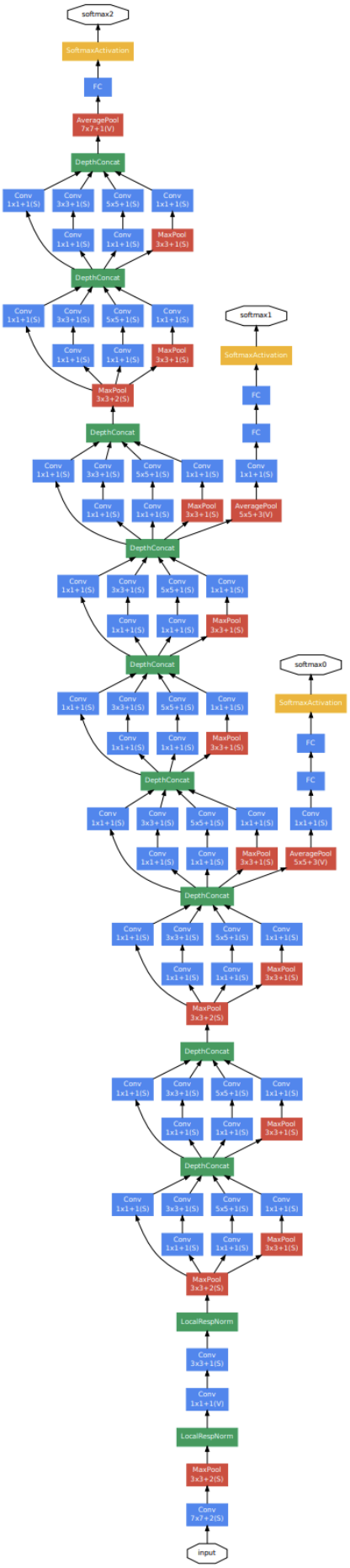


Figure 2.17: CNN architecture:Googlenet [30]

Some of the researchers also detected lung disease using object detection architectures DarkNet-19[61], DarkNet-53 [141] that are backbone for YOLO. These architectures are primarly designed for real time object detection. Like the VGG models, DarkNet-19 uses filters primarily and doubles the number of channels after each pooling phase. It employs filters to compress the feature representation between convolutions and global average pooling to produce predictions, in line with the work on Network in Network (NIN). The model batch is regularized, convergence is accelerated, and training stability are achieved using batch normalization. The CNN known as DarkNet-53 serves as the foundation for the YOLOv3 object identification methodology. The detailed architectures are shown in the Figure 2.18(a) and (b). One can observe the DarkNet-53 is more deeper and residual layers in between instead of maxpool layers.

Type	Filters	Size/Stride	Output
Convolutional	32	3×3	224×224
Maxpool		$2 \times 2/2$	112×112
Convolutional	64	3×3	112×112
Maxpool		$2 \times 2/2$	56×56
Convolutional	128	3×3	56×56
Convolutional	64	1×1	56×56
Convolutional	128	3×3	56×56
Maxpool		$2 \times 2/2$	28×28
Convolutional	256	3×3	28×28
Convolutional	128	1×1	28×28
Convolutional	256	3×3	28×28
Maxpool		$2 \times 2/2$	14×14
Convolutional	512	3×3	14×14
Convolutional	256	1×1	14×14
Convolutional	512	3×3	14×14
Convolutional	256	1×1	14×14
Convolutional	512	3×3	14×14
Maxpool		$2 \times 2/2$	7×7
Convolutional	1024	3×3	7×7
Convolutional	512	1×1	7×7
Convolutional	1024	3×3	7×7
Convolutional	512	1×1	7×7
Convolutional	1024	3×3	7×7
Convolutional	1000	1×1	7×7
Avgpool		Global	1000
Softmax			

(a) DarkNet-19

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

(b) DarkNet-53

Figure 2.18: DarkNet Architectures [61, 141]

An extreme version of Inception called XceptionNet [52] was created on the theory that there is complete decoupling between the mapping of cross-channel

correlations and spatial correlations in feature maps. As seen in Figure 2.19, the Xception architecture has 36 convolutional layers that serve as the network's foundation for feature extraction. With the exception of the first and last modules, each of the 14 modules made up of the 36 convolutional layers has a linear residual connection surrounding it. The input flow receives the data first, followed by the middle flow—which is repeated eight times—and the exit flow. Batch normalization and a depth multiplier of 1 (no depth expansion) are applied to all convolution and separable convolution layers.

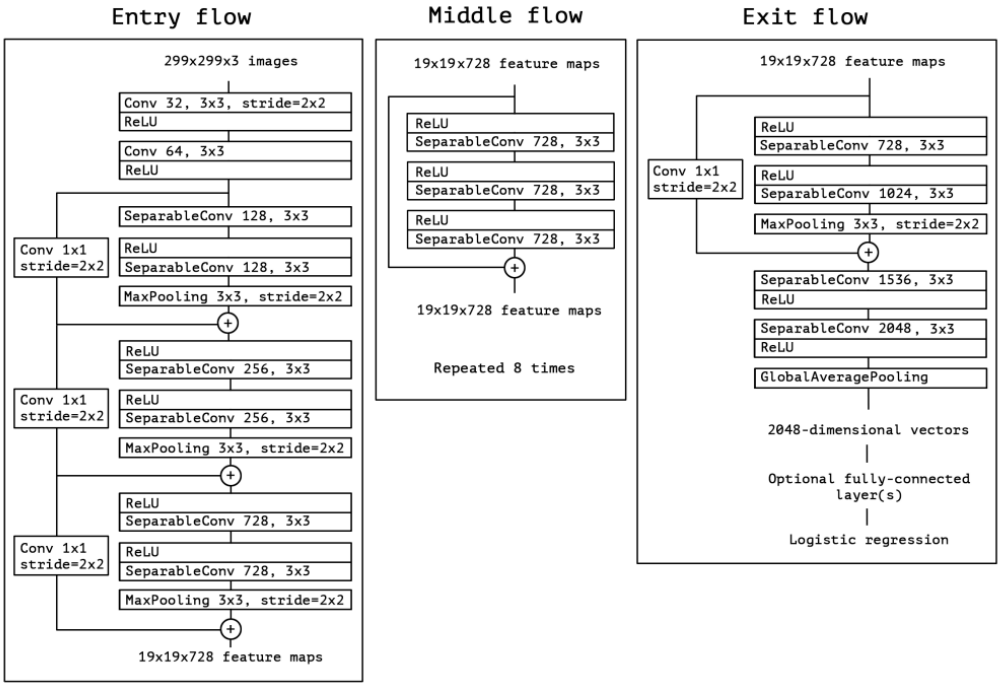


Figure 2.19: The Xception architecture [52]

2.5.3 VISION TRANSFORMERS

Vision Transformers (ViTs), an increasingly prominent neural network architecture, are utilized for computer vision tasks, particularly in the field of medical image processing. ViTs utilize the transformer architecture, initially designed for natural language processing, to handle image data. Here is an overview of how ViTs might be used to identify lung infections in CXR images, following the general principles

of ViT application in computer vision[142].

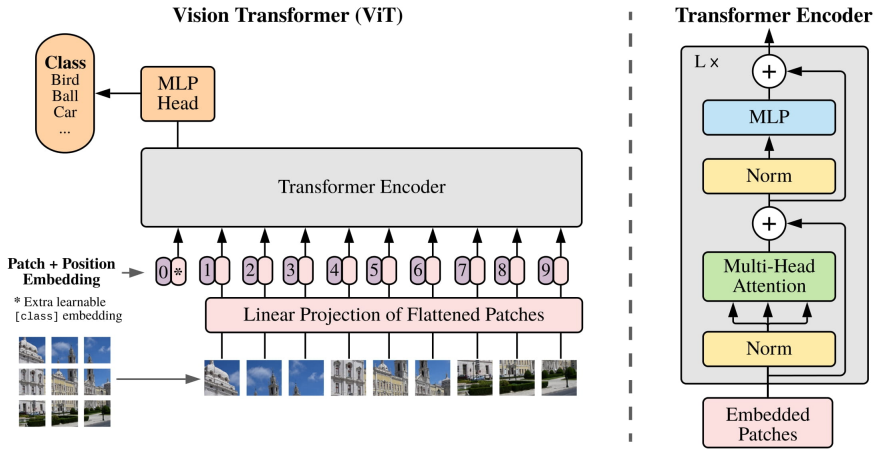


Figure 2.20: Vision Transformer architecture [143]

ViT, developed by [143] as shown in the Figure 2.20, utilizes a transformer to achieve image identification. The transformer requires one-dimensional sequences as input. To accommodate this, the image was initially separated into patches of equal size. These patches were then flattened to generate a sequence of two-dimensional patches. After applying linear projection, patch embeddings are obtained. These patch embeddings are then inputted into the transformer encoder, where position embeddings are added to them. The Multi-layer Perceptron (MLP) was included into the transformer encoder while the multi-head self-attention (MSA) component from the transformer model was kept. Ultimately, the picture classification was produced using the MLP head module. A module in which many attention modules learn distinct features of attention in separate sub-spaces is called the multi-headed attention module [144]. First, the vector matrix X was obtained and subsequently mapped to various sub-spaces. Using the learnable matrices W_q, W_k, W_v accordingly, to produce the matrices query (Q), key (K), and value (v).

Q, K, and V are all equal in the encoder. In this case, the decoder's first layer's output is Q, and its second layer's outputs are K and V. The final feature representation is obtained by projecting these via the remaining learnable weights. Sev-

eral subspaces can be created thanks to the multi-head attention process, where each head focuses on different informational components. Multi-head attention enhances the network's resilience and stability at the same time. The number of people working on a given task may change. It is important to note that although ViTs have shown potential in various computer vision tasks, their use in medical image analysis, specifically in detecting lung infections [145, 146, 147] may require adjustments to accommodate the distinct characteristics and challenges of medical images. Furthermore, in order to ensure the accuracy and applicability of ViT-based models in healthcare settings, it is crucial to utilize comprehensive and varied datasets, conduct meticulous model evaluation, and validate the model. It is also outperforming deep learning methods [148] while classifying lung diseases from CXRs.

2.5.4 EXPLAINABLE ARTIFICIAL INTELLIGENCE(XAI) IN LUNG DISEASE DETECTION

There are several Explainable Artifcial Interlligence(XAI) methods that are divided based on the way of applying. Few XAI methods are applied after training the model, few are during training the model it self and few are tested by computing predictions for a group of samples. The detailed division has been shown in the Figure 2.21.

Among these, few can be applied to images: Class Activation Mappings(CAM) is one among those. It makes localization possible for CNNs specialized in categorization. The discriminative regions that are used to identify that category are indicated by CAM. Although they are not necessary, explicit bounding box annotations alter the model's architecture. Just prior to the final output layer, global average pooling is applied on convolutional feature maps. After that, a fully linked layer receives these features and generates the required output.

For a given image, let $f_k(x, y)$ represent the activation of unit 'k' in the last convolutional layer at spatial location (x, y). Then, for unit 'k', the result of performing

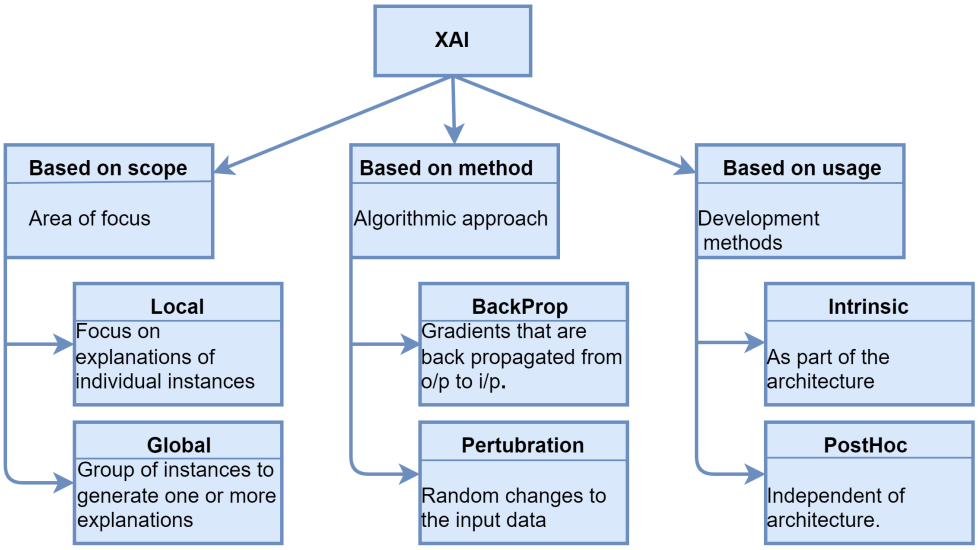


Figure 2.21: Categorization of XAI methods[149]

global average pooling,

$$F_k = \sum_{x,y} f_k(x, y)$$

Therefore, for a given class c, the weight corresponding to class 'c' for unit 'k' is the input to the softmax, $S_c = \sum_k w_k^c F_k$, w_k^c. In essence, it shows how significant F_k is for class c. Ultimately,

$$exp(S_c)/(\sum_c exp(S_c)).$$

yields the softmax output for class c, or P_c.

$$S_c = \sum_k w_k^c \sum_{x,y} f_k(x, y) => \sum_{x,y} \sum_k^c f_k(x, y),$$

can be obtained by plugging $F_k = \sum_{x,y} f(x, y)$ into the class score.

$$M_c = \sum_k w_k^c f_k(x, y)$$

where the class activation map for class c is denoted by M_c; hence, $S_c = \sum_{x,y} M_c(x, y)$.

This indicates the importance of the activation at spatial grid (x, y) leading to the classification of an image to class c. One drawback of CAM is that it needs feature maps to come before softmax layers. These designs might perform less accurately than general networks on other tasks. Hence a technique that doesn't require changing the current architecture is required. That is where Gradient-weighted Class Activation Mapping (Grad-CAM) has come.

Grad-CAM uses gradients of the target class flowing into. This class discriminative localization method doesn't require architectural modifications or retraining to operate on any CNN-based network. This class discriminative localization method doesn't require architectural modifications or retraining to operate on any CNN-based network. It has been tested on ResNet. It assess the impact of moving from deep to shallow layers. It has been applied to the best classification, VQA, and captioning models currently in use. Convolutional layers typically preserve spatial information, whereas completely connected layers lose this information. Grad-CAM determines each neuron's significance for a decision of interest by utilizing gradient information coming from the last layer. Grad-CAM works as follows:

- Compute gradient of 'y' for class 'c' with repect to feature map y_k: $\frac{\partial y^c}{\partial A_k}$.
- Then, Global average pool these gradients to obtain neuron importance weights. $\alpha_k^c = \frac{1}{z}\sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k}$.
- Perform weighted combination of forward activation maps and follow it by relu to obtain $L_{Grad-CAM}^c = ReLU(\sum_k \alpha_k^c A^k)$.

The concept originated from the fact that fully connected layers lose spatial information while convolutional layers retain it. When gradients go into the final convolutional layer, the information is used to assign importance to each neuron. It computes the gradient of the class score with respect to the feature map activations of the convolutional layer first, and then flows back with each convolutional layer to obtain the localization map. Every layer after that goes through the same procedure

again. Gradients are global averages pooled as an aggregation process as they flow back. These techniques are more trustworthy, and they were one of the successful techniques to describe visuals from CNN output.

Grad-CAM was used to put multiple pre-trained models to the test by generating overlay heatmaps across the region of interest. For example, [150] has observed heat maps superimposed on bilateral multifocal ground-glass opacities with patchy consolidations. In the case of bacterial pneumonia, localized heat maps show opacities with consolidation on the lower and upper lobes in some of the images.

Further, Brunese and his colleagues [151] have employed a three-fold method to classify COVID-19 from other pulmonary diseases. Grad-CAM visuals considered only lung regions while classifying disease. These activations matched with radiologist results. As shown in the Figure 2.22, one patient has bronchial wall thickening with tiny peripheral patchy infiltrates, while another has multi-focal patchy opacities while being diagnosed with COVID-19.

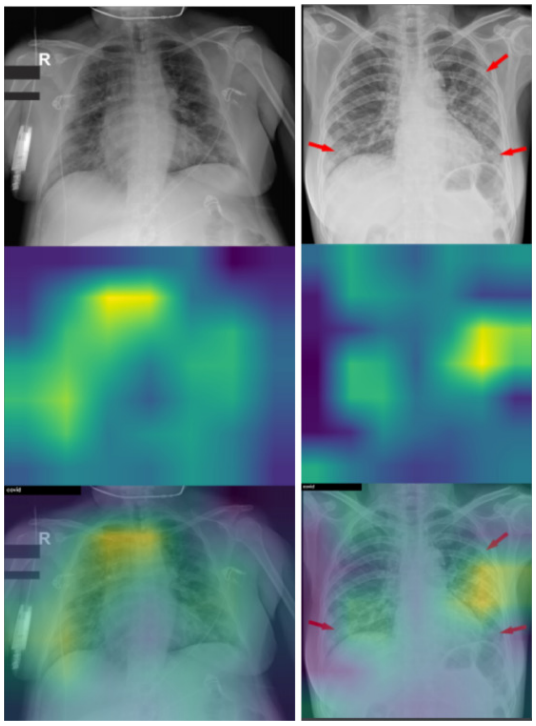


Figure 2.22: Examples: Grad-CAM activation maps [151]

Local Interpretable Model Agnostic Explanations(LIME) [152] is such another

model that works in a different way, it does modify the samples. LIME follows the following approach:

- LIME decomposes image 'I' into 'd' superpixels. These are small homogeneous image patches using quick shift algorithm; Quickshift is a mode-seeking technique that treats the pixels as samples over a 5-dimensional space, which consists of three color dimensions and two space dimensions. These little homogenous image patches are created using quickshift.
- generates several new images $x_1, ..., x_n$ by alternating these superpixels at random.
- finds the predictions by querying the model in the image $y_i = f(x_i)$.
- constructs a local weighted surrogate model 'm' that fits the y_i's to the existence or non-existence of superpixels.
- An original image 'I' superpixel is linked to every coefficient of'm'. It makes sense that a superpixel is more significant for the prediction at 'I' if it is more positive, according to LIME.
- Typically, the superpixels linked to the highest positive coefficients are highlighted when a user visualizes the 'm' hat.

Saliency map [153] is a ranking-based technique that ranks pixels based on their influence on the class score. CNNs are assessed [154] CAM by classifying normal and abnormal. Several cases were studied concerning correct and wrong predictions. The model was accurately impacted by the lower left cardiac area. The relatively high CNN score for this negative example (0.48) suggested that this study was rated as borderline abnormal by both the CNN and the physician. Local lesion attention guided network(LLAGnet) [155] is built to classify different thoracic diseases in CXR images. The net's objective is to find local lesions. Back-propagating gradients will be used by a weakly-supervised attention mechanism integrated into

the global branch to acquire visual regions of lesion locations. To give more granular features for visual categorization, the ideal attention region is amplified and applied to the local branch. A multi-attention CNN [156] is designed for automatic diagnosis. It learns discriminative features for each category. It has been done in two phases—one extracts convolutional feature representations with DenseNet121 as a backbone. Two is applying weekly supervised learning of deep convnets to extract local patches.

Adversarially Robust Optimization[157] is a method that improves learned feature representations that are robust against adversarial examples. Adversarial examples are inputs that are perturbed. It is a min-max problem. Objective one is to find perturbed inputs which maximize the loss. Objective two is to minimize the overall loss of the classifier while learning interpretable features. All the networks in the experiment are learned using SGD with a learning rate of $10^{-2}$and a momentum of 9*10^{-2} with binary cross entropy loss which is defined as:

$$-(1/N)\sum_{i}^{N} y_i * log(y_i^p) + (1 - y_i) * log(1 - y_i^p)$$

A segmentation-based deep fusion network(SDFN)[158] was trained on Thoracic disease classification. This network focuses on local regions rather than the whole image. Local regions are identified by the Local Region Generator(LRG). LRG uses Unit architecture to segment the lung regions. A couple of Densenets consisting of 121 layers were used as feature extractors and further fine-tuned to classify the diseases. CAMs were used to generate heatmaps to test the model. CNN with attention feedback(CONAF) [159] is designed to identify a chest radiograph that is likely to contain one or more lesions. Pediatric chest radiographs[160] could be explained using profound learning predictions. Radiographs of patients with pneumonia were used to visualize the model's predictions. The CAMs were displayed using LIME and Gard-CAM, respectively. Infections from lung inflammation typically result in puss or fluids in air sacs.

CheXNeXt [161] have used class activation mappings to create heatmaps that indicated regions in the CXR images. It is a 121-layer DenseNet architecture used to detect 14 different chest diseases, such as pneumonia, pneumothorax and etc. Multiple instance learning is another localization approach that divides the CXR image into overlapping patches. Probability scores for each patch will be generated by training patches through CNN. Patch-level probabilistic localization helps in critical findings.

Tang was [162] able to determine abnormalities from CXR images with proper explanations for each group of images with overlaid heat maps. The heatmap represents the forecast of anomalies with a high probability. The CNN model, for example, forecasts it as abnormal with 0.99 probability, and a heat map will be created, whereas four radiologists categorized it as abnormal.

Grad-CAM visualization method used to explain the COVID-19 disease [55, 163]. However, the explanations are only for samples of images, where local features were not given more importance. Grad-CAM was used on top of the existing classifier, as shown in the Figure 2.23. An improved version of Grad-CAM has used by authors of [164] to segment the medical images.

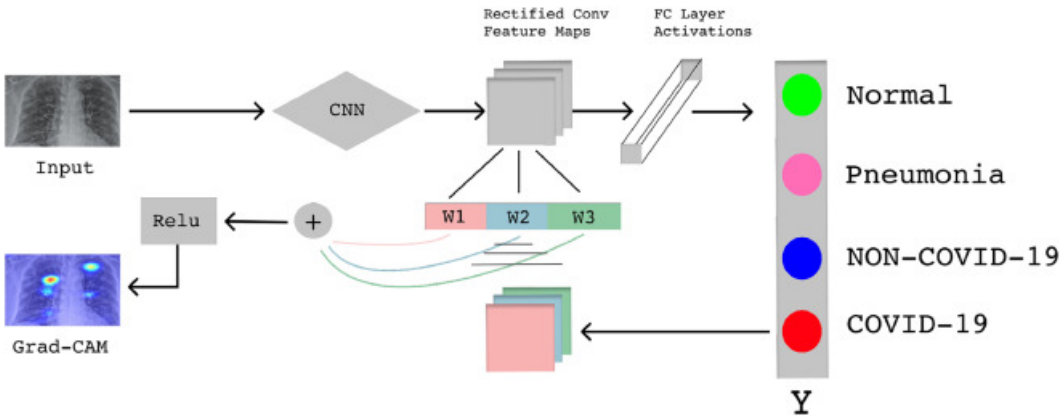


Figure 2.23: Usage of Grad CAM to explain infections[55]

Several Cancer diseases were explained through XAI methods [165]. For example, Lung cancer is explained through Supervised Iterative Descent, Grad-CAM, Deep Hierarchical Semantic CNN; Breast cancer with Deep-Miner, Grad-CAM,

Integrated Gradients; Brain cancer with Capsule Network, Pyramid Grad-CAM; Liver cancer with saliency maps, activaton maps. Deep network activation maps were studied in [166] with joint localization and severity grading aids in predicting disease. However, they remain inaccurate in localizing the actual infestation termed infection maps.

There are other XAI methods such as Partial Dependency Plots(PDP), Accumulated Local Effects (ALE), Individual conditional expectation (ICE), SHAP, ELI5, etc. PDP plots [167] illustrate the incremental impact of one or two features on the anticipated outcome of a machine-learning model. ALEs are more similar to PDP with a change in the way of computing the feature importance. It is based on differences in predictions rather than averages [168]. On the other hand, ICE plots [169] are used to assess the impact of a variable on a trained machine-learning model, assuming that all other variables remain constant. By estimating each feature's contribution to the prediction, SHAP seeks to clarify the forecast of an instance x. Shapley values are computed using coalitional game theory by the SHAP explanation technique [170]. The objective of SHAP is to elucidate the prediction of an instance x by calculating the contribution of each feature to the prediction. The feature values of a data instance function as participants in a coalition. Shapley values provide a method for equitably allocating the "payout" (i.e., the prediction) across the different attributes. It is inspired by a local surrogate model. ELI5 [171] is primarily designed for text data to explain the answers in a simple way. Among all these methods LIME, Grad-CAM, and Saliency mappings are visualization methods that explain images. Methods like PDP, ALE, and ICE work on lower dimensions and structured data for example predicting a customer churn from customer demographical, potential to buy, and job-related and etc..features. SHAP has not been applied to any medical images so far. So we employ LIME, Grad-CAM, and Saliency maps to get the region of interest features.

Chapter 3

PROBLEM FORMULATION

3.1 BACKGROUND

Automated medical image analysis began when the first medical image was digitized. It has combined low-level pixel processing (edge and line detector filters, extracting region) and computational analysis (fitting lines, circles, and ellipses) to develop compound rule-based systems that perform complex operations during the 1970s to 1990s [172].

CXR images can convey a great deal about a patient's condition; hence the standard chest radiograph should be reconsidered [81]. Early detection of lung disorders is crucial for effective treatment and may reduce stress in the healthcare system. CXR images and CT scans are the standard image diagnosis tests for lung diseases. Although CT scans are the gold standard, CXRs are still valuable because they are less expensive, faster, and widely available. A detailed literature review on the techniques to identify lung diseases is presented by [172, 173]. Even though the authors [174] discussed various deep learning methods published between 2015 and March 2023, **To the best knowledge of the authors, nothing substantial was found in the literature that could explain the localization of a disease with the region of interest features and how feature extraction is migrating from traditional approaches(geometrical-based methods) to deep learning methods.**

3.2 MOTIVATION

- Though much work has been done on the CXR images in medical imaging, it

is challenging to identify very tiny lesions similar to other tissues. **There is a need to develop a complete AI system to address the same.**

- The open-source and open-access study community's work inspires and motivates me as I try to figure out how well AI systems work.
- As learning is a continuous process, one should update their skill set on the current research andutting-edge techniques used in academia and business.

3.3 PROBLEM STATEMENT

Identifying lung diseases like COVID-19 from medical images is challenging due to the difficulty in accurately localizing affected regions. Current AI models lack precise localization capabilities, leading to reduced diagnostic accuracy. A new AI system is needed to support medical practitioners by accurately identifying lung diseases through the localization of key regions in medical images, improving both speed and precision in diagnosis. .

The modality used here is the chest X-ray imaging technique. Its usage is because these are available at less cost and accessible to the public. This work has taken portable network graphics(png) and joint pixel export group(jpeg) format images during the implementation. The proposed approach experimented with limited system capacity and it may take a while to generalize this method as most of the work needs to be certified by an expert radiologist. This work is only to be used for academic and research purposes and is not to be commercialized.

3.4 OBJECTIVE

To build an AI system that supports medical practitioners in identifying lung disease(for example COVID-19) through localization.

Sub Objectives:

1. **To develop local indicators (Region of Interest features) using an attention-guided mechanism and associate them with global features extracted for determining the chest infection.**
2. **To design a mechanism using a majority voting attention-based classifier for data imbalance problems in the architecture.**
3. **To analyze the correlation of multiple architectures on variable patterns**

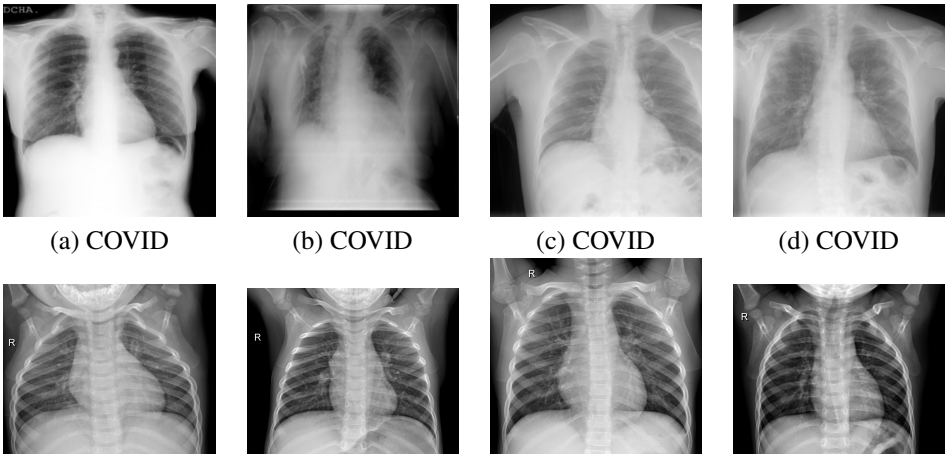
3.5 DATA

The proposed approach utilized COVID-19 and Normal CXR images during the experiments. A total of 2,896 COVID-19 images were sampled from Dr. J. P. Cohen's GitHub repository. Dr. J. P. Cohen, from the University of Montreal, collected data consisting of CXR and CT scan images [175]. To the best of the author's knowledge, this was the first COVID-19 dataset shared on GitHub. Additionally, 1,341 Normal CXR images were sampled from the RSNA Pneumonia Detection Challenge dataset [176]. These images were divided into training and testing datasets, as shown in Table 3.1. Few samples from COVID-19 and Normal images has been given in the Figure 3.1.

S.NO	Type	COVID-19	Normal	Total
1	Train data	2396	1041	3437
2	Test data	500	300	800

Table 3.1: Data distribution proposed approach (Dataset created by sampling COVID-19 images from [150,175], Normal images from [176]).

COVIDx dataset [82] comprising 13,975 CXR images across 13,870 patient cases. The authors integrated and modified five publicly available data repositories to create this dataset. i. COVID-19 image collection of Dr. J.P. Cohen, ii. COVID-19 chest X-ray Dataset Initiative [177], iii. Actual Med COVID-19 chest X-ray Dataset Initiative[178], iv. RSNA Pneumonia Detection Challenge dataset [176] v.



(a) COVID (b) COVID (c) COVID (d) COVID

(e) NORMAL (f) NORMAL (g) NORMAL (h) NORMAL

Figure 3.1: COVID-19 and Normal images from the dataset

COVID-19 radiography database [179]. The dataset can be reproduced through the codes from [180]. Data distribution is shown in the Table 3.2.

Type	**Normal**	**Pneumonia**	**COVID-19**	**Total**
Train	7966	5421	152	13539
Test	100	100	31	231

Table 3.2: Data distribution in COVIDx Dataset

Non-COVID and COVID instances of both CXR and CT images are included in the COVID-19 dataset. A total of 17,099 X-ray and CT images are generated from the dataset using various techniques [181]. COVID-R [51] dataset consists of 2843 COVID-19, 3,108 Normal, and 1,439 Pneumonia CXR images. Some of the publicly available datasets are shown in the Table 3.3.

Author	Data Distribution	Available@
[43]	183-COVID-19 5521-Pneumonia 8066-Normal	https: //github.com/linda wangg/COVID-Net
[181]	4044-COVID-19 5500-Normal	https://data.mendeley.com /datasets/8h65ywd2jr/3
[83]	3616-COVID-19 1345-Pneumonia 10192-Normal	https://www.kaggle.com/tawsifurrahman /covid19-radiography-database
[99]	11956-COVID-19 11,263-Pneumonia 10701-Normal	https://www.kaggle.com/anasmohammedtahir /covidqu/version/7
[106]	105-COVID-19 80-Normal	https://github.com/ieee8023/ covid-chestxray-dataset.
[182]	4272-Pneumonia 1583-Normal	https://www.kaggle.com /paultimothymooney/chest-xray-pneumonia
[48]	520 COVID-19 5000 Normal	https://github.com/shervinmin/DeepCovid/ tree/master/data
[118]	305-COVID-19 305-Pneumonia viral 305-Pneumonia bacterial 305-N	https://data.mendeley.com/datasets /rscbjbr9sj/2

Table 3.3: Data sets available for public-COVID-19,Pneumonia,Normal

The National Institute of Health released one of the largest publicly labeled datasets Chest X-ray8 [183] in this field, containing 1,08,948 images of 32,717 different patients, classified into eight different categories. The radiologists' annotations were labeled using natural language processing techniques. Chexpert [184] is another large dataset containing 2,24,316 CXRs from 65,240 people divided into 14 classes. Tuberculosis chest X-ray database [129] consists of 3,500 TB images and 3,500 Normal CXR images.Montgomery County X-ray Set [185] consists of 80 Normal, 58 TB from the Department of Health and Human Services of Montgomery County, MD, USA. Shenzhen Hospital X-ray 336 TB, 326 Normal CXR images. X-ray images in this data set have been collected by Shenzhen No.3 Hospital in Shenzhen, Guangdong Province, China. Two datasets named Dataset A(DA) and Dataset B(DB) [130]. The DA was composed of the test set (26 non-TB and 26 TB CXRs) and the training set (52 non-TB and 52 TB CXRs). In contrast, the test set consists of 25 non-TB and 25 TB CXRs, whereas the database has 50 TB and 50 non-TB CXRs. Two highly qualified chest radiologists from the National Institute of Tuberculosis and Respiratory Diseases in New Delhi, India, and Indira Gandhi Medical College in Shimla, India, independently assessed each CXR and determined which cases were TB and which were not. [186] Two radiologists collected chest X-rays at each center (BWC, with 19 years of experience in chest radiology, MJC, 18 years of experience, and six chest radiologists with more than ten years of experience) from four hospitals from 2015 to 2017. This dataset consists of 200 Abnormal and 800 Normal images. The US National Library of Medicine [128] has made two datasets with a significant focus on TB. One contains 80 normal cases and 58 TB; with 326 normal cases and 336 TB. [187] CXR images with lung nodules are 154 and 93 without a nodule were selected from 14 medical centers. Some of the datasets that are available for the public have shown in the Table 3.4.

Author	Data set Name and Distribution	Available@
[129]	Tuberculosis chest Xray database: 3500-TuberculosisB , 3500-Normal	https://ieee-dataport.org documents/ tuberculosis-tb-chest-x-ray -database#files
[185]	Montgomery County X-ray Set(MC):58-TB,80-Normal Shenzhen Hospital X-ray Set:336-Tuberculosis ,326-Normal	https://lhncbc.nlm.nih.gov/ LHC-publications/pubs/Tuberculosis ChestXrayImageDataSets.html
[130]	Dataset A:78-Tuberculosis, 78-Normal Dataset B: 75-TuberculosisB,75-Normal	https://sourceforge.net/ projects/tbxpredict/files/data/
[188]	The Japanese Society of Radiological Technology (JSRT): 154-TuberculosisB,78-Normal	http://db.jsrt.or.jp/eng.php

Table 3.4: Datasets available for public for the detection of Tuberculosis

Chapter 4

METHODOLOGY

To classify the lung images to COVID-19 and healthy images, we have proposed a system architecture that consists of generating local features from images with the XAI method and training the CNN network twice, one model on global features and another on local features and the final one with a fusion of local and global features. There are various options to extract features. Employ a custom CNN network with sigmoid as output activation, another is to employ a transfer learning network for feature extraction. Class imbalnce is one of the problem and it is addressed by designing a multi stage approach while handling through ensembles methods. Preprocessing methods such as image resize, contrast enhancement, and otsu's thresholding algorithm for segmentation are included as part of the system. The entire proposed system has been shown in the Figure 4.1

The rest of the sections will present overview of each sub objective in details.

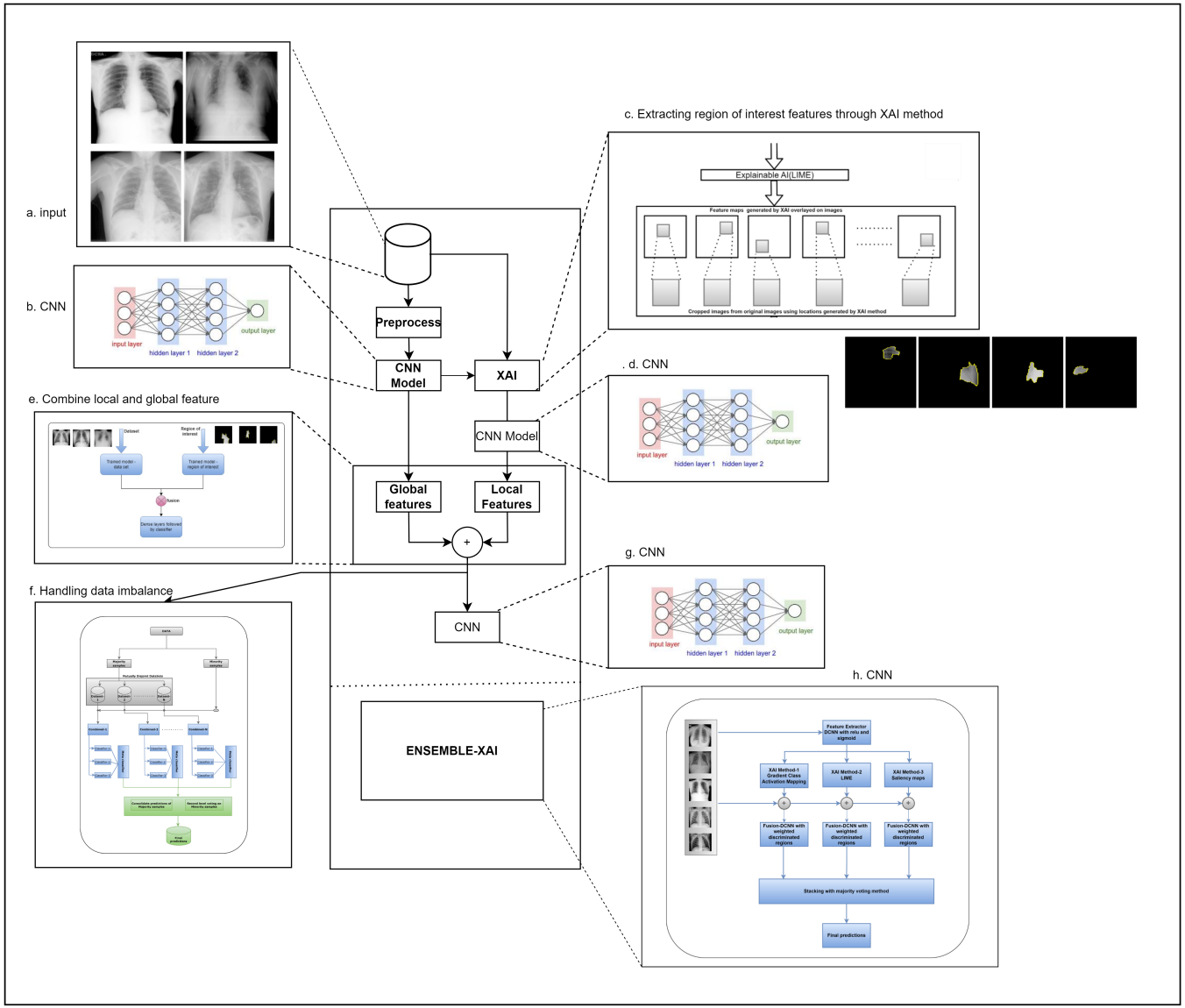


Figure 4.1: Overall system architecture for the proposed approach

4.1 SUB OBJECTIVE 1

To develop local indicators (Region of Interest features) using an attention-guided mechanism and associate them with global features extracted for determining the chest infection.

The suggestedd approach is to obtain local features through the attention-guided mechanism. This can be achieved through the Explainable AI method by applying it to a trained classifier. A trained classifier is a model that was designed and trained on a binary classification problem of lung diseases. Then, the XAI method was used to generate a Region of Interest(ROI) features that reveals why an image is classified into that class. It explains the hidden nature of a typical CNN-based model. Using the XAI method, annotate the ROI features on the trained data set. Here the goal is to super impose a binary mask on the region of interest. Train a model on the ROI of images. These are additional features that further explain and help the classifier. The challenge here is that while extracting ROI, there is a chance of missing the spatial information of features. This information is captured through a proposed quadrant approach, shown in the algorithm 1. The proposed method's architecture is shown in the Figure 4.2. It has multiple phases, i. Train a CNN model on the given set of images using the best architecture, ii. Generate local features from the set of images using the XAI method, e.g., LIME., iii. Train another CNN model using the local discriminant features, iv. Combine the generated local features with features generated in phase one, v. Perform a simple classification model on the fusion data set. For example, the ROI mask can be observed from COVID-19 class images in the Figure 4.4. These ROIs are the output images from phase two. Local and global features are given equal weight during fusion. The architecture used to train images is a light weight model with four CNN layers,max-pooling, and dropout layers.

In addition to local discriminant features, the location of features will be extracted as given in the Algorithm 1. CXR image is partitioned into four quadrants labeled Q1, Q2, Q3, and Q4. The centroid is computed from all non-zero pixel

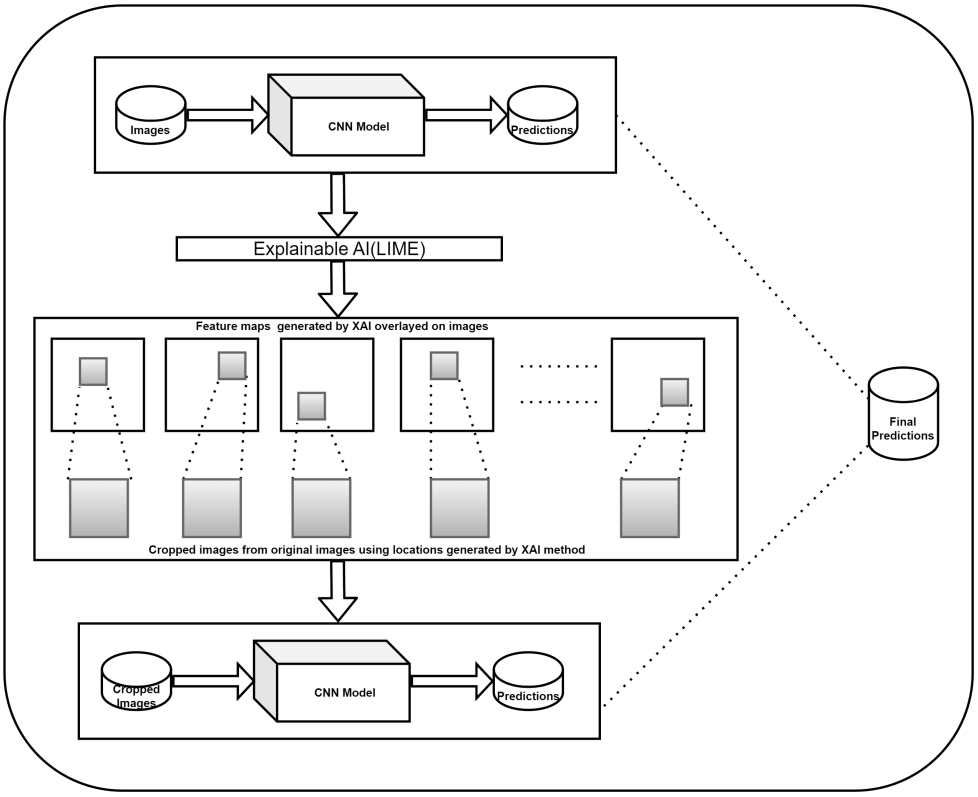


Figure 4.2: Proposed method to extract local indicators

locations. A pixel at location (i,j) is denoted as i'th row j'th column. Every image mask can be denoted with two pointers; one is on top of the left corner and the other from the bottom right corner. The top left corner pointer is the least of the coordinates in the given locations, and the other right corner is the higher coordinate. These two reference pointers will helps to get the feature's centroid which further aids in labeling.

The fusion of local and global features which helps to classify the disease as shown in the Figure 4.3. Here the fusion of features is carried out in the following way:

- CXR images sent to a trained CNN network and features gets extracted: say output of one image is a flattened vector of size 1*500 - global vector
- similarly local features extracted from other CNN network that is trained on local features: say output is a flattened vector of size 1*500 - local vector

Algorithm 1 Extracting spatial location of features with Quadrant approach

while $i \leq$len(images) **do**
  if $image_i$ is valid **then**
    $x_1 \leftarrow non - zero - locations(image_i)$
    $x_{count} \leftarrow size(x1)$
    $x_{11} \leftarrow min(x1, axis = 1)[0]$
    $x_{12} \leftarrow min(x1, axis = 1)[1]$
    $x_{21} \leftarrow max(x1, axis = 1)[0]$
    $x_{22} \leftarrow max(x1, axis = 1)[1]$
    $r_r \leftarrow \frac{x_{11}+x_{21}}{2}$
    $r_c \leftarrow \frac{x_{12}+x_{22}}{2}$
    $t \leftarrow \frac{sizeofimage}{2}$
  else if $r_r <= t$ and $r_c <= t$ **then**
    append(l1,Q1)
  else if $r_r >= t$ and $r_c <= t$ **then**
    append(l1,Q2)
  else if $r_r <= t$ and $r_c >= t$ **then**
    append(l1,Q3)
  else
    append(l1,Q4)
  end if
end while

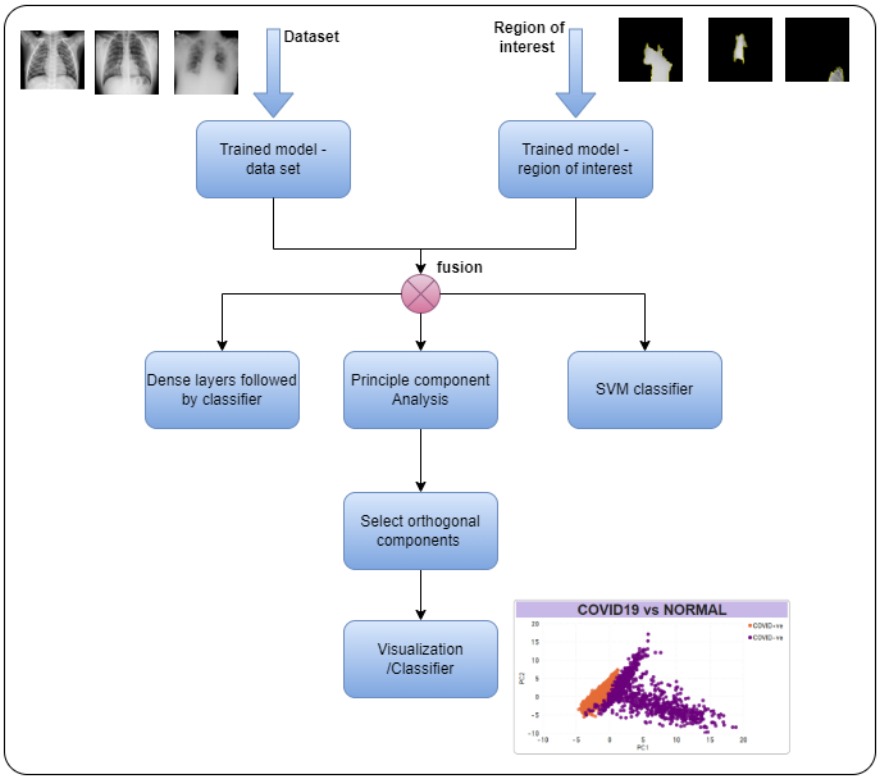


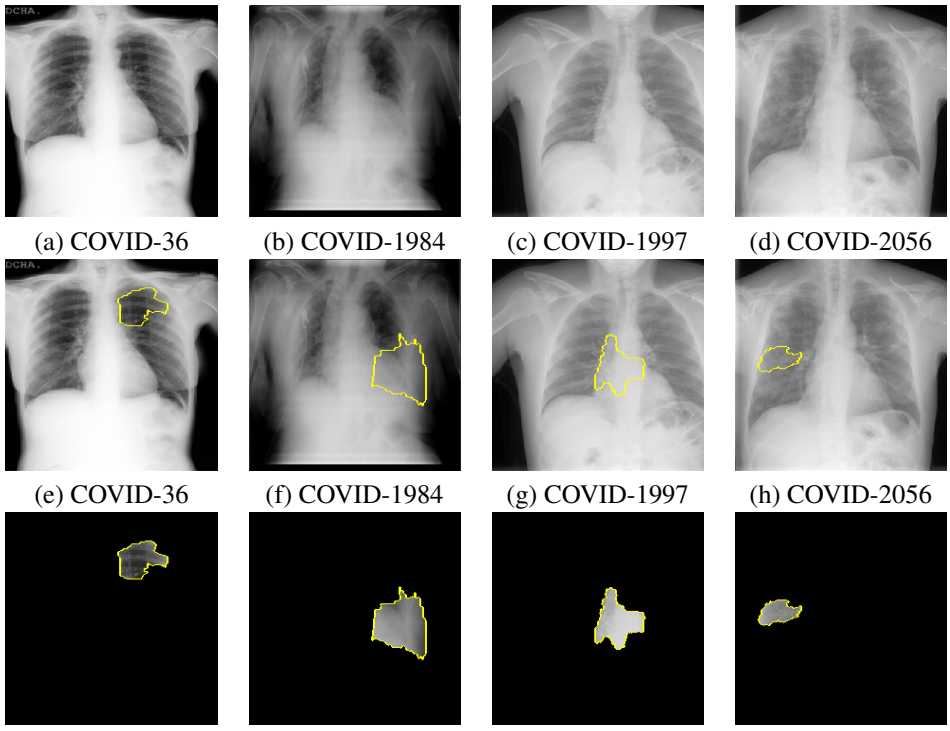
Figure 4.3: fusion of local and global features

- now, we have two flattened vectors of same size, perform element wise addition.

- element wise additon could be weighted sum. i.e w1* global vector + w2* local vector ; w1, w2 could be varied from 0.1 to 0.9 such that w1+w2=1.

One way to reduce the featues is through a traditional principal component analysis, however one looses the explainability,it could be considered for a quick understanding of features.

The proposed methods are evaluated based on computing ROI from multiple-resolution images. This is further discussed in detail in the next section.



(a) COVID-36 (b) COVID-1984 (c) COVID-1997 (d) COVID-2056

(e) COVID-36 (f) COVID-1984 (g) COVID-1997 (h) COVID-2056

(i) COVID-36 (j) COVID-1984 (k) COVID-1997 (l) COVID-2056

Figure 4.4: COVID-19 images and their corresponding localization maps along with masks

4.2 SUB OBJECTIVE 2

To design a mechanism using a majority voting attention-based classifier for data imbalance problems in the architecture.

An imbalanced learning problem is a challenge that has grabbed the interest of academics and industry. Class imbalance is a wide spread problem in ML due to the unavailability of data in a specific category. Several under-sampling and over-sampling methods handle class imbalance; however, finding the correct pattern heavily depends on the sub-sampling. This work proposes a two-stage approach with a data sampling method with Mutually Disjoint Data Sets(MDS) and ensemble models. Ensembles give more importance to the minority samples with a double voting method. The proposed method has an improvement of 3.78, and 13.78 percent of average recall on the train and test data respectively compared to best-base model results. The suggested method comprises two stages: stage one separates the data from the majority class into Mutually Disjoint Datasets (MDS). By merging MDS with minority samples, a subsample is created. Most samples will be used in this way for building the model. Each sub-sample will be used to train a set of classifiers. The second stage creates meta-classifiers using learned models. The outcomes of each model's predictions are combined. It can be done in the following way.

- divide the predictions-MDS predictions, minority predictions.
- combine all MDS prediction sets.
- second level voting on minority samples.

This approach's schematic is shown in the Figure 4.5. Further, the trained models can be stored and used for future data in the following way.

- send unseen data to all trained models
- get the prediction sets
- use majority voting method to get the final predictions.

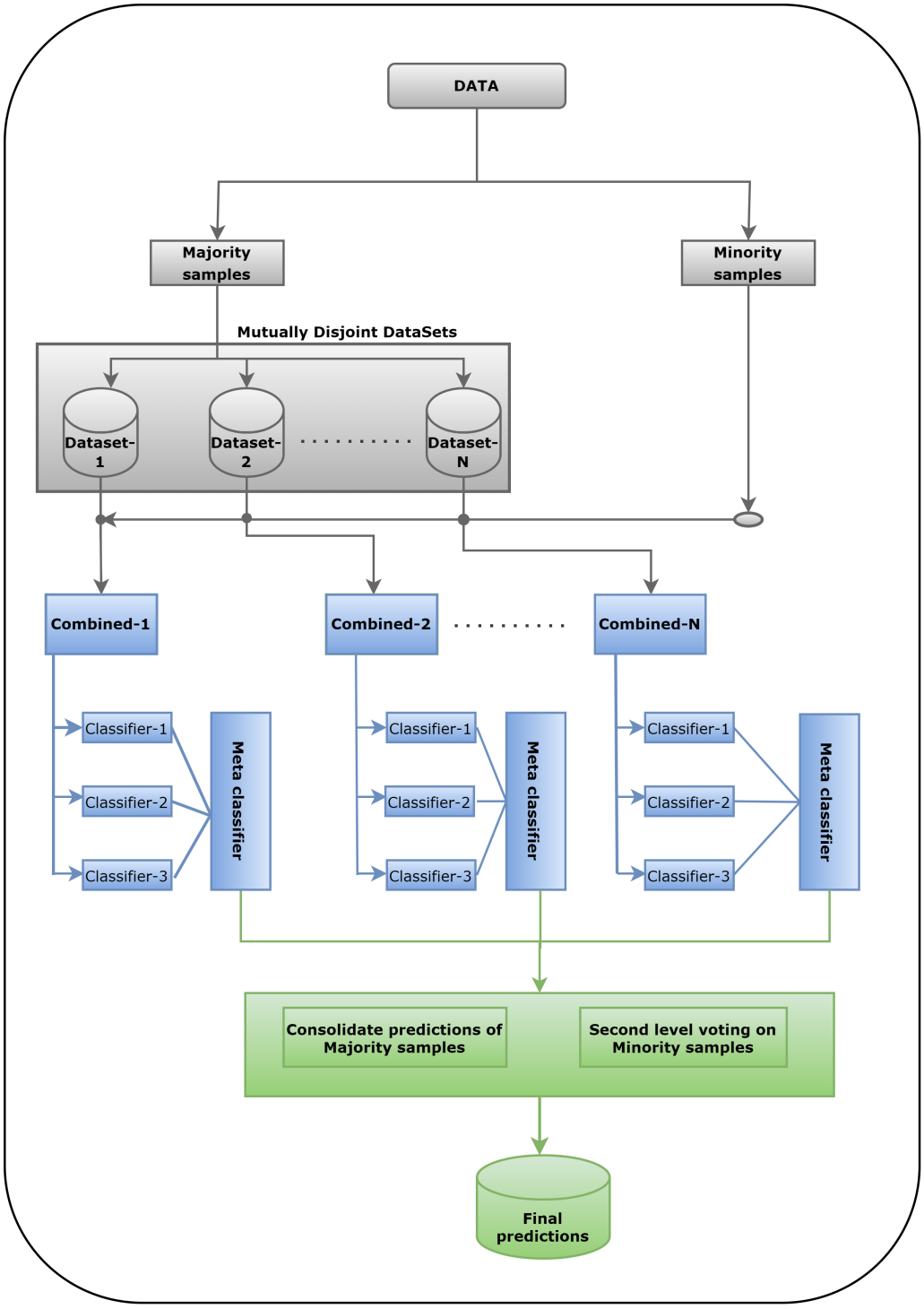


Figure 4.5: Proposed approach: A two stage approach to handle class imbalance.

4.3 SUB OBJECTIVE 3

To analyze the correlation of multiple architectures on variable patterns.

Gradient-weighted Using gradients of the target class flowing into the final convolutional layer, Class Activation Mapping generates a coarse localization map that highlights the critical places in the image for predicting the target class [189]. The concept originated from the fact that fully connected layers lose spatial information while convolutional layers retain it. When gradients go into the final convolutional layer, the information is used to assign importance to each neuron. It computes the gradient of the class score with respect to the feature map activations of the convolutional layer first, and then flows back with each convolutional layer to obtain the localization map. The same process is repeated at each other layer. Gradients are global averages pooled as an aggregation process as they flow back. These techniques are more trustworthy, and they were one of the successful techniques for describing visuals from CNN output.

Saliency map [153] is a ranking-based technique that ranks pixels based on their influence on the class score. Because the class score function is highly nonlinear, identifying the significance of a pixel is difficult. However, the importance of a pixel can be computed using a linear function utilizing the first-order Taylor expansion. $S_c(I) = \|w\|^T I + b$ where w is the derivative of S with respect to the image I at the point (image) I_0: $w = [\frac{dS_c}{dI}]_{I_0}$.

LIME [152] is another innovative explanation method that learns an interpretable model locally around the predictions in order to explain the predictions of any classifier in an understandable and accurate manner. Interpretable explanations must use a representation that is understandable to humans, regardless of the model's real attributes. The output, which allows the model to most accurately predict the class, is a binary vector that indicates whether a continuous patch of pixels is present or absent.

A collection of XAI models is created and trained to address the problem of

Algorithm 2 Ensemble of XAI models -Training

Read CXR images of 224*224 resolution.
Load customized trained model(base model) from h5 object.

Phase 1 - Generating ROIs-Local features from CXR images

while All images exhausted **do**
    Input model and image to the XAI model to generate ROI mask.
    $i \leftarrow XAImethod$
    if $i = GradCAM$ or $i = saliencymap$ **then**
        Send the image to the XAI method to get the ROI mask
        Apply Otsu's method on ROI to get the right segments
    else
        Send the image to the XAI method to get the ROI mask
    end if
    Write the roi generated image to disk.(named as local dataset(LD)
end while

Phase 2 - Training CNN model on local features- Local models

while On each dataset produced in the previous step **do**
    Train a CNN model.
    Store the model as an h5 object.
end while

Phase 3 - Fusion of original images with extracted local regions

while On each LD(s) **do**
    read images from LD and original dataset
    while All images exhausted **do**
        Point wise pixel addition local feature and original image
        prepare a numpy array
    end while
    Store numpy array for each - Fusion Dataset(FDs)
end while

Algorithm 2 Ensemble of XAI models -Training (Continued)

Phase 4 - Training a CNN model on combined dataset

while On each FD(s) **do**
    Train a CNN model
    Store it as h5 object (Fusion model-FM)
    Get the predictions
    Take the mode of predictions as the final output
end while

binary classification that detects the presence or absence of an individual who has COVID-19. GradCAM, LIME, and Saliency maps are some of the XAI approaches used in this work. Each XAI approach generates ROIs from CXR pictures. To produce ROIs, various XAI approaches are applied, and a CNN model is trained on each output before the final predictions are selected by a majority vote.

The process has been explained thorugh the algorithm 2 to train the model,

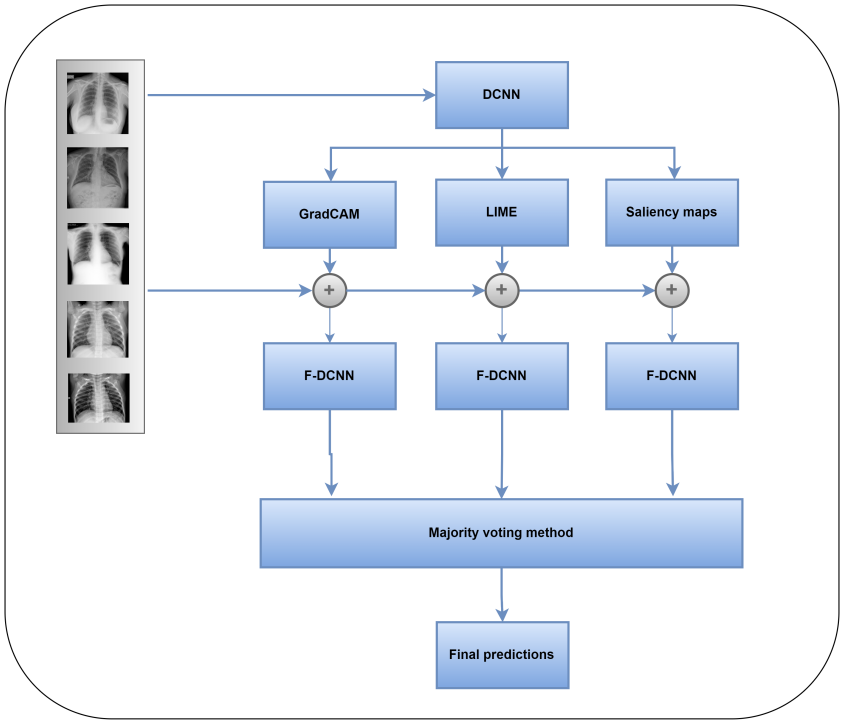


Figure 4.6: Block diagram for proposed method-Ensemble XAI; DCNN-Deep convolutional neural net;F-DCNN-Fusion DCNN

Algorithm 3 Ensemble of XAI models -Testing

```
Read test CXR images of 224*224 resolution.
Load customized trained model(base model) from h5 object.
Load three local models from the h5 object.
while All images exhausted do
    while each of three XAI method(s) do
        produces local features using the XAI model.
        Fusion with original data.
        Apply fusion model to get the prediction.
    end while
    Take the mode of all as the final prediction.
end while
```

testing is in Algorithm 3, and the schematic has been shown in the Figure 4.6. The following is the summary of the process i. train a CNN model on given input images, ii. sharing the stored model and images with XAI method to generate rois, iii. fusion of input images and generated rois and training another custom DCNN on data generated from phase ii, iv. get the predictions from the previous phase for all XAI techniques and finalize the prediction using majority voting.

The proposed methodologies offer several distinct advantages over conventional approaches for lung disease classification in CXR images. Unlike standard CNN models that primarily operate as "black boxes," our method leverages XAI techniques, such as Grad-CAM and LIME, to extract and visualize local features, thereby enhancing interpretability. This is particularly crucial in medical imaging, where understanding the rationale behind a model's predictions can significantly impact clinical decision-making. Additionally, the fusion of local and global features, derived from separate models, allows for a more comprehensive representation of the data, capturing subtle patterns that may be overlooked by single-model systems. This integrated approach not only improves classification accuracy but also provides deeper insights into the spatial distribution of disease-specific markers within the lungs. Furthermore, our multi-stage ensemble framework, which utilizes Mutually Disjoint Datasets (MDS) for addressing class imbalance, offers a robust alternative to traditional sampling methods, ensuring better performance on underrepresented classes. This comprehensive system architecture thus demonstrates superior efficacy and reliability compared to existing methodologies.

Chapter 5

EXPERIMENTS AND RESULTS

5.1 ARCHITECTURE DETAILS TO GENERATE LOCAL FEATURES

The problem formulation has been done in the following way: A disease classification problem with two classes COVID-19 and NORMAL. Data from [150] with a size of 3437 samples were used in the experiment. Images of different scales 120*120,150*150,180*180,224*224 were used for better generalization and to persist scale invariant transformations. Four global models were trained, local discriminant features were extracted, and four local models were trained using local features. Principle Component Analysis(PCA) is applied to local features while combining with other quadrant information. Further, rules are derived by applying a Decision Tree(DT). The experiments were conducted on different-resolution images, as shown in the Figure 5.1.

In phase one, we divided the data into train and validation sets while reading through the image data generator. Each image is cropped to a target resolution size, e.g.:224*224. A custom CNN network was taught for $5*10^1$ epochs with a learning rate of 10^{-3} and patience of five. A learning rate regularizer was used to avoid overfitting. T he a rchitecture c onsists o f C NN(2*64,2*128) a nd m ax p ooling, dropout layers between bound with a set of dense layers(64,1 neurons) towards the end. Most of the time, models try to learn patterns over the epochs, and very few errors shoot up towards the end, as shown in Figure 5.2.

In phase two, sharing the model and image with LIME [152] to work on the image; LIME is an agnostic model that tweaks the features in the dataset to see the change in the output. When we set the feature weight to a lesser value, we

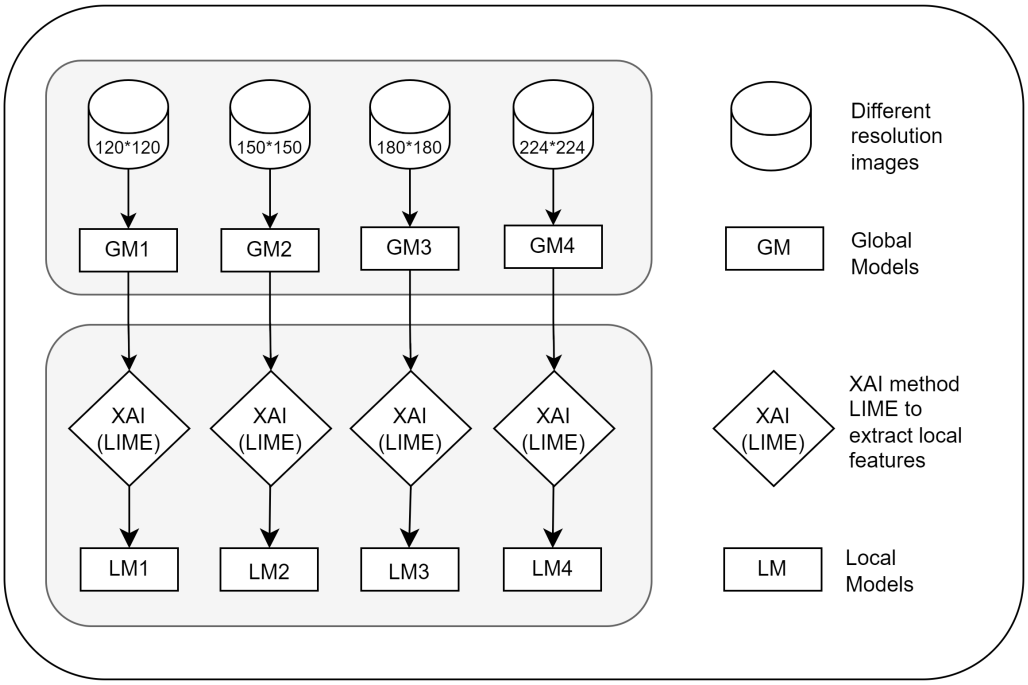


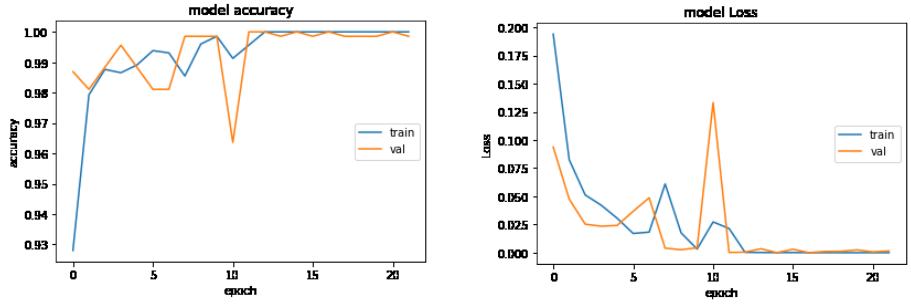
Figure 5.1: Experiment set up

are able to fetch the critical feature. These features are masked and produced as shown in the Figure 4.4. A sample of images from one class is shown along with the corresponding mask. For example, Figure 4.4 (a) COVID-36 corresponding annotation is Figure 4.4 (e) and its mask Figure 4.4 (i). All discriminant regional features were independently labeled and trained using the similar CNN architecture.

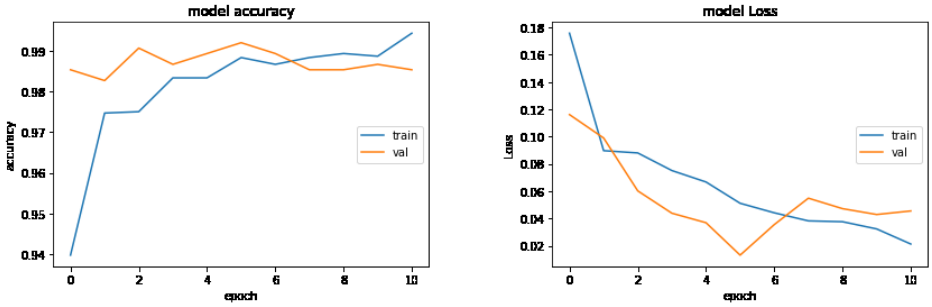
Phase three is extracting global and local features, as shown in the Figure 4.3 from the flattened layer separately; aggregated features are trained further on densely connected layers of 512, 256 neurons with an intermediate dropout rate of 0.2 and 50 epochs. This model has achieved mean accuracy of 99.29%, 97.81% across all resolutions on trian and test data respectively. Further, PCA was applied, and a DT was built to extract rules. All the experiments were conducted in python using google colab[1] by saving intermediate results between the phases.

We have conducted experiments with varying image resolutions from 120*120,150*150, 180*180, and 224*224. The predictions on global data were shown in confusion

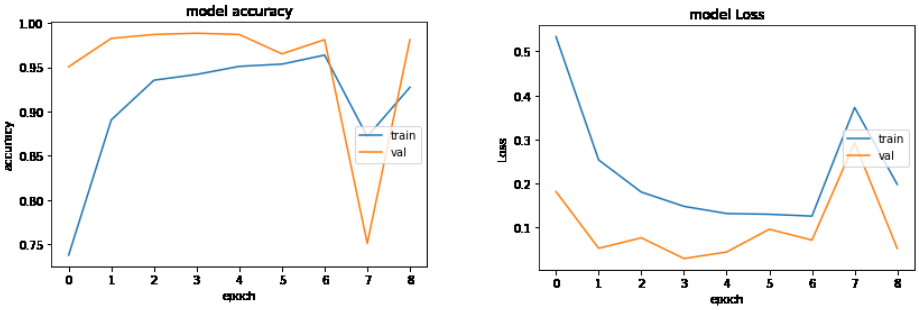
[1]https://colab.research.google.com/



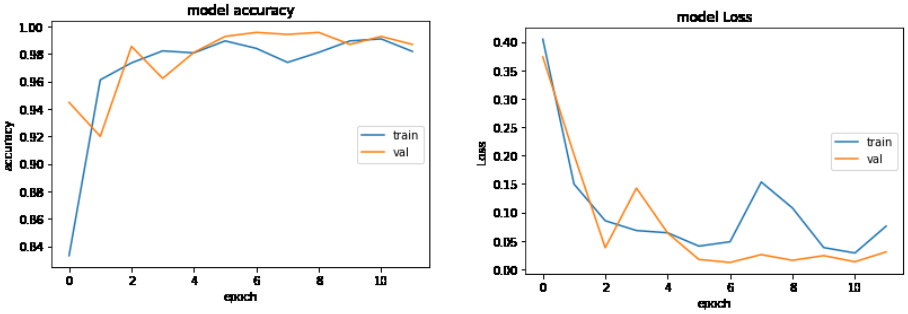
(a) image resolution (224,224)



(b) image resolution (180,180)

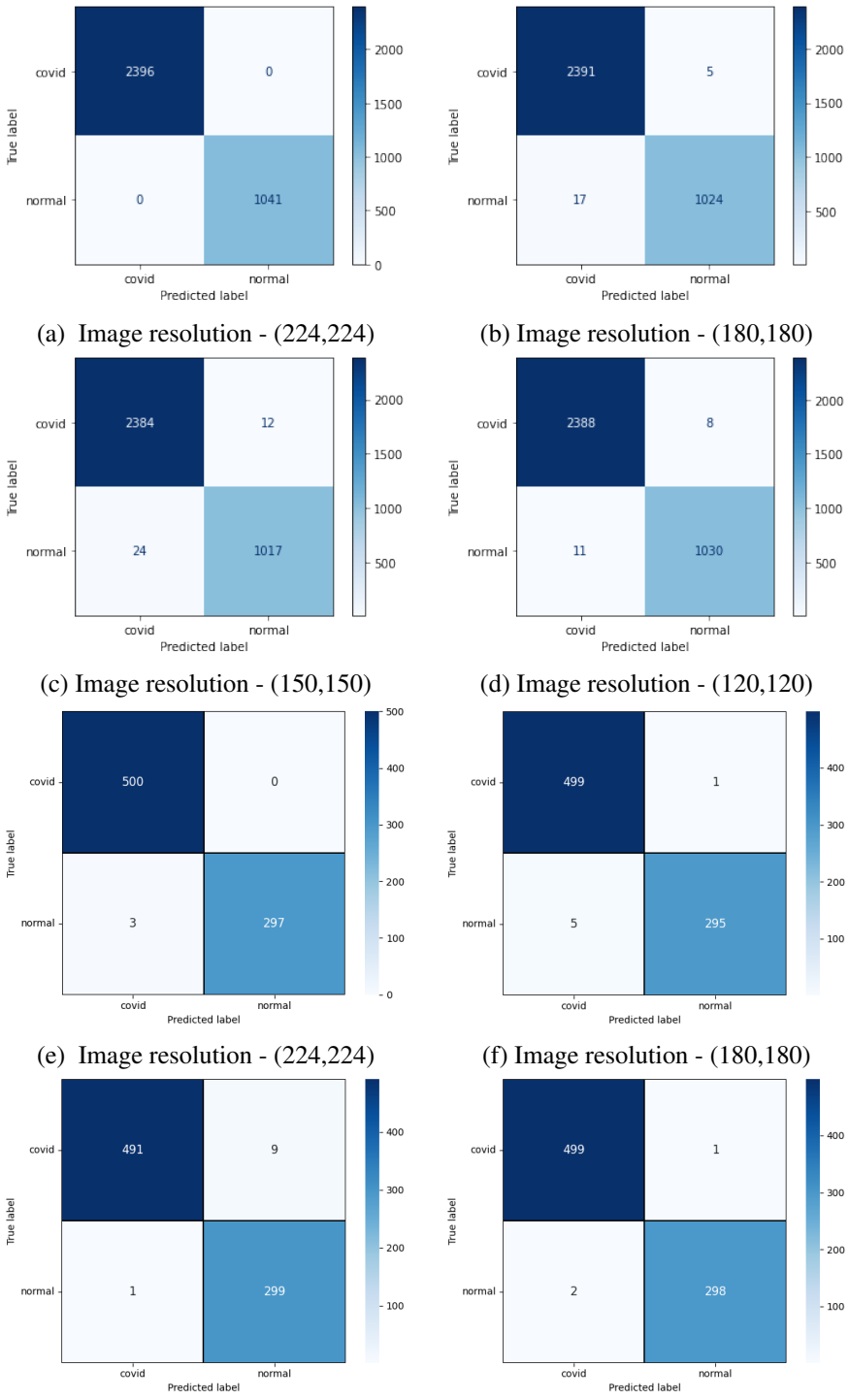


(c) image resolution (150,150)



(d) image resolution (120,120)

Figure 5.2: Accuracy and loss curves for various image resolutions trained on COVID-19 data set used from table 3.1.



(a) Image resolution - (224,224)

(b) Image resolution - (180,180)

(c) Image resolution - (150,150)

(d) Image resolution - (120,120)

(e) Image resolution - (224,224)

(f) Image resolution - (180,180)

(g) Image resolution - (150,150)

(h) Image resolution - (120,120)

Figure 5.3: Confusion matrices for various resolution images - global data ((a) to (d) are train, (e) to (h) are test), COVID-19 data set used from Table 3.1.

matrices in the Figure 5.3 and the the detailed results are shown for global and fusion features in the Table 5.1. When compared to state-of-the-art approaches as show in the Table 5.2, the suggested method yielded considerably improved results. All the authors listed in the Table 5.1 have used TL models except [105] Masud. This author has used a lightweight CNN model, which is an appropriate comparison to the proposed method.

S.No	Model	traindata	Resolution	Data size	TP	TN	FP	FN	Accuracy
1	Extract global features using CNN network, sigmoid at the end- Adam optimizer.	1	120*120	3437	2388	1030	8	11	99.45
2		0	120*120	800	499	298	2	1	99.63
3		1	150*150	3437	2384	1017	24	12	98.95
4		0	150*150	800	491	299	1	9	98.76
5		1	180*180	3437	2391	1024	5	17	99.36
6		0	180*180	800	499	295	5	1	99.25
7		1	224*224	3437	2396	1041	0	0	100.00
8		0	224*224	800	500	297	3	0	99.63
9	Local descriminant features+ Global features	1	120*120	3004	2161	765	78	0	97.40
10		0	120*120	800	499	255	45	1	94.25
11		1	150*150	2639	1891	748	0	0	100.00
12		0	150*150	800	497	292	3	8	98.63
13		1	180*180	3004	2175	823	3	3	99.80
14		0	180*180	800	493	297	3	7	98.75
15		1	224*224	3004	2177	826	1	0	99.97
16		0	224*224	800	500	297	3	0	99.63

Table 5.1: Predictions for various resolution images- TP-True Positive, TN-True Negative, FP-False Positive, FN-False Negative, Acc-Accuracy(%), in train data column: 1-train data,0-test data (Data used: COVID-19,Normal).

Author	No.of classes	Acc	Recall	Spec
[44]	2 (COVID-19,NORMAL)	98.08	95.13	95.30
[46]	2 (COVID-19,NORMAL)	90.00	100.00	83.00
[119]	2 (COVID-19,NORMAL)	96.83	96.26	95.54
[105]	2 (COVID-19,NORMAL)	98.78	-	-
[190]	2 (COVID-19,NORMAL)	80.28	-	-

Table 5.2: Metrics from state-of-the-art methods,Acc-Accuracy,Spec-Specificity

5.2 ENSEMBLE METHODS, HANDLING CLASS IMBALANCE

In the proposed method COVID-19 and normal CXR images of 224*224 size have been taken as input. In all the experiments DCNN architecture as shown in Figure 5.4 was used, where the input layer has a dimension of 224*224.

Once the DCNN model is trained on COVID-19 and normal classes, images and model were fed to the XAI technique for eg. GradCAM, through which annotated

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 222, 222, 32)      896

 conv2d_1 (Conv2D)           (None, 220, 220, 64)      18496

 max_pooling2d (MaxPooling2D  (None, 110, 110, 64)     0
 )

 dropout (Dropout)           (None, 110, 110, 64)      0

 conv2d_2 (Conv2D)           (None, 108, 108, 64)      36928

 max_pooling2d_1 (MaxPooling  (None, 54, 54, 64)       0
 2D)

 dropout_1 (Dropout)         (None, 54, 54, 64)        0

 conv2d_3 (Conv2D)           (None, 52, 52, 128)       73856

 max_pooling2d_2 (MaxPooling  (None, 26, 26, 128)      0
 2D)

 dropout_2 (Dropout)         (None, 26, 26, 128)       0

 conv2d_4 (Conv2D)           (None, 24, 24, 128)       147584

 max_pooling2d_3 (MaxPooling  (None, 12, 12, 128)      0
 2D)

 dropout_3 (Dropout)         (None, 12, 12, 128)       0

 flatten (Flatten)           (None, 18432)             0

 dense (Dense)               (None, 64)                1179712

 dropout_4 (Dropout)         (None, 64)                0

 dense_1 (Dense)             (None, 1)                 65

=================================================================
Total params: 1,457,537
Trainable params: 1,457,537
Non-trainable params: 0
_________________________________________________________________
```

Figure 5.4: Layer-wise detail of DCNN model

images were generated. The annotated images generated through GradCAM and saliency maps are preprocessed further with Otsu's thresholding method to get the segments. Otsu is a segmentation algorithm that generates lower and upper thresholds to segment the image based on the continuity of pixels [2]. Similarly, LIME is another XAI model applied to this problem [191]. When we use LIME the feature weights will vary between 0.0001 to 0.00000001

Fusion is a simple process of the addition of pixels, which intuitively gives more weight to the important regions though it does not show much interpretation when we plot the fusion image. Of course, it can be done through normalization but it does not convey anything to the human eye. Three XAI techniques GradCAM, LIME, and saliency maps were used during experimentation to generate critical regions. Each generates local discriminant regions which are nothing but masked images and were used further to train a model along with original data.

XAI method	train	Acc	Recall	Prc	F1
GradCam	1	99.44	99.45	99.80	1.00
	0	98.79	99.23	99.10	0.99
Saliency	1	99.24	99.47	99.1	0.99
	0	98.05	98.22	98.20	0.98
LIME	1	99.96	100.00	100.00	1.00
	0	99.62	100.00	99.40	1.00
Ensembles	0	**98.85**	**97.88**	**100.0**	0.99

Table 5.3: Metrics captured from the models built on XAI generated annotations when combined with original data(COVID-19); In train column, one indicates train data, zero indicates test data,Acc-Accuracy,Prc-Precision

The results reported in the Table 5.3 for an ensemble of XAI is predicting all positive cases correctly without missing anything although the proposed method is on par with individual models in other metrics. Since the final output is voting from three models, it is much more reliable.

Handling class imbalance: The images are sampled to create class imbalance data with 100 images from COVID-19 and 900 from normal class with the intention of creating a 90:10 ratio. Here the minority class samples are from the COVID-19 class. In general, positive samples are rare for any disease due to lack of data

[2]https://learnopencv.com/otsu-thresholding-with-opencv/

collection. The experiments are conducted in the following way.

- a customized CNN model has been trained on the sample data set.
- mutual disjoint datasets(MDS) were created with repeated sampling from the majority class while combining with minority class samples.
- a CNN model has been trained on each MDS and saved the predictions.
- predictions were consolidated from all MDS with two-stage voting for minority class samples.
- results were compared before and after applying the method.

The experiments have a choice of choosing models between customized models and TL methods. However, TL methods are a good choice as there are fewer images. The results were recorded during the training original dataset and MDS are given in Table 5.4. While creating MDS, COVID-19 class has been brought to 40% of the total with the undersampling method since these samples are less than 10% in the dataset. The experiments were conducted on the original dataset before creating MDS and can be compared with the results generated from MDS. We can observe that all methods gave results on par with the base metric. Among TL methods, VGG16 has received low scores. However, it has performed well on test data using the proposed method.

Correlation analysis with transfer learning techniques: In this work, we have used some of the TL methods by restricting to one version from every family of TL methods. The experiments were conducted on the FD. A TL method is trained on FD with binary cross-entropy loss, Adam optimizer for 50 epochs on every XAI-generated output. The results are shown in the Table 5.5.

TL methods have been applied to the original dataset and the metrics are reported in the Table 5.6. One can compare these results with the average metric computed across all the TL methods built on fusion datasets is slightly higher than the individual models on the original data. It will be efficient if we compare at the

S.NO	Model	Data	train	Accuracy	Recall	Precision	F1 score
1	Customized Model	DS	1	98.17	99.88	98.10	0.99
			0	99.04	98.91	100.00	0.99
		MDS1	1	97.86	94.44	100.00	0.97
			0	96.15	90.00	100.00	0.95
		MDS2	1	99.15	100.00	98.59	0.99
			0	96.15	100.00	94.44	0.97
		MDS3	1	99.15	100.00	98.62	0.99
			0	96.15	100.00	93.33	0.97
		Testdata - proposed method	0	97.61	98.13	98.50	0.98
		Test data - vanila model	0	98.14	98.88	98.51	0.99
2	DenseNet	DS	1	99.68	99.84	99.84	1.00
			0	99.71	99.67	100.00	1.00
		MDS1	1	99.43	98.99	100.00	0.99
			0	100.00	100.00	100.00	1.00
		MDS2	1	97.70	96.36	100.00	0.98
			0	97.67	96.43	100.00	0.98
		MDS3	1	99.43	99.12	100.00	1.00
			0	95.35	93.88	97.87	0.96
		Testdata - proposedmethod	1	96.28	89.83	98.15	0.94
		Testdata - vanila model	0	96.54	89.26	100.00	0.94
3	ResNet	DS	1	100.00	100.00	100.00	1.00
			0	99.71	99.67	100.00	1.00
		MDS1	1	100.00	100.00	100.00	1.00
			0	100.00	100.00	100.00	1.00
		MDS2	1	99.42	99.07	100.00	1.00
			0	100.00	100.00	100.00	1.00
		MDS3	1	99.43	99.05	100.00	1.00
			0	97.67	96.55	100.00	0.98
		Testdata - proposedmethod	0	97.91	93.86	99.07	0.96
		Testdata - vanila model	0	94.72	85.71	100.00	0.92
4	VGG16	DS	1	94.80	94.62	100.00	0.97
			0	92.96	92.59	100.00	0.96
		MDS1	1	85.63	81.62	100.00	0.90
			0	83.72	77.78	100.00	0.88
		MDS2	1	87.93	83.33	100.00	0.91
			0	86.05	82.09	100.00	0.90
		MDS3	1	84.48	80.58	100.00	0.89
			0	76.74	70.59	100.00	0.83
		Testdata - proposedmethod	0	73.94	52.43	100.00	0.69
		Testdata - vanila model	0	59.31	41.38	100.00	0.59

Table 5.4: Comparison of metrics - handling class imbalance through an ensemble method which includes customized model and TL methods;In train column one indicates train data, zero indicates test data,(COVID-19 data).

individual model level. However, the panel of the TL model's decisions is much more reliable than the base one. This analysis can be done by comparing with the results from the Table 5.3 and claim that the ensemble of customized models results are consistent.

Some assumptions were made while training all the networks for standardization purposes. Each network was trained for 50 epochs with a learning rate of 0.001, using the Adam optimizer and a patience of five epochs. No other parameter

S.NO	XAI method	Transfer learning method	train	Accuracy	Recall	Precision	F1 score
1	GradCAM	ResNet50v2	1	99.52	99.72	99.61	1.00
			0	97.74	98.23	98.67	0.98
		MobileNetv3small	1	99.20	99.56	99.34	0.99
			0	**99.44**	99.45	**99.78**	**1.00**
		Xception	1	99.48	99.56	99.72	1.00
			0	96.37	96.49	98.54	0.98
		ConvNextTiny	1	99.68	99.73	99.84	1.00
			0	97.66	98.73	97.95	0.98
		EfficientNetV2S	1	99.68	99.78	99.78	1.00
			0	**99.19**	99.33	**99.55**	**0.99**
		DeseNet121	1	99.08	98.96	99.78	0.99
			0	**99.68**	99.77	**99.77**	**1.00**
		VGG16	1	99.68	99.78	99.78	1.00
			0	98.87	99.31	99.08	0.99
2	LIME	ResNet50V2	1	99.76	99.89	99.78	1.00
			0	**99.60**	99.67	**99.78**	1.00
		MobileNetv3small	1	84.10	96.47	84.13	0.90
			0	82.10	96.22	81.65	0.88
		Xception	1	99.96	99.95	100.00	1.00
			0	**99.84**	99.77	**100.00**	**1.00**
		ConvNextTiny	1	94.46	94.82	97.48	0.96
			0	94.92	95.63	97.26	0.96
		EfficientNetV2S	1	96.66	97.03	98.33	0.98
			0	95.73	97.21	96.89	0.97
		DenseNet121	1	99.88	99.89	99.94	1.00
			0	**99.92**	100.00	**99.89**	**1.00**
		VGG16	1	98.77	99.28	99.01	0.99
			0	**99.59**	99.89	**99.55**	1.00
3	Saliency	ResNet50v2	1	92.35	99.34	87.94	0.93
			0	91.77	99.23	87.39	0.93
		MobileNetv3small	1	97.66	97.44	98.19	0.98
			0	96.93	96.79	97.31	0.97
		Xception	1	98.28	97.86	98.78	0.98
			0	95.40	93.63	98.20	0.96
		ConvNextTiny	1	98.42	97.53	99.47	0.98
			0	99.16	98.96	99.48	0.99
		EfficientNetV2S	1	93.04	96.80	93.90	0.95
			0	93.23	97.47	93.19	0.95
		DenseNet121	1	95.81	92.15	99.86	0.96
			0	95.96	92.56	100.00	0.96
		VGG16	1	98.77	99.28	99.01	0.99
			0	**99.59**	99.89	**99.55**	**1.00**

]

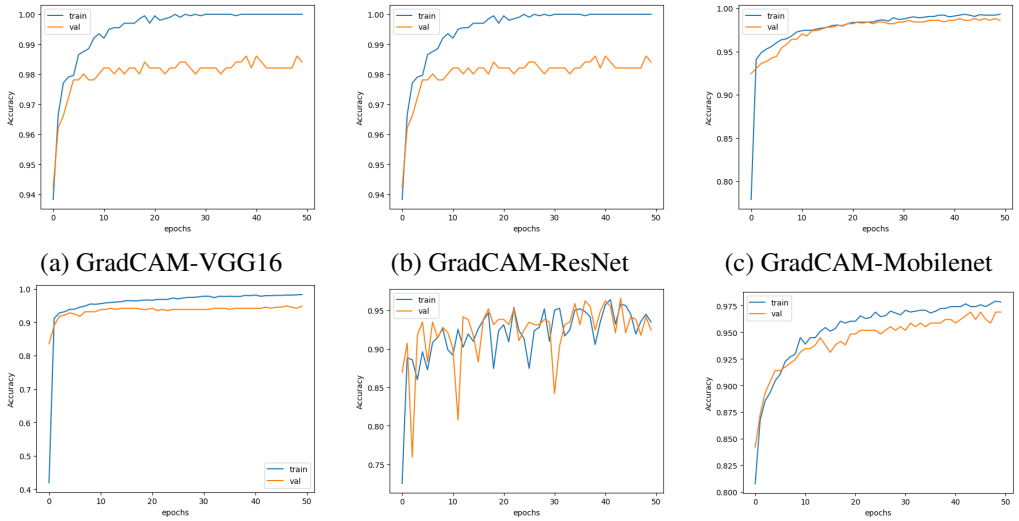
Table 5.5: Metrics recorded after applying various TL methods on different XAI generated fusion datasets; In train column one indicates train data, zero indicates test data - An ensemble model(COVID19 data).

tuning was done to improve the results in order to avoid overfitting. With the approach adopted, we claim that some of the networks are weak classifiers that may perform well on a few data points compared to others, which is actually the concept of ensembles. Some of the error curves are shown in Figure 5.5.

Further the experimentation was done on identifying Lung nodules from CXR

S.NO	Transfer learning method	train	Accuracy	Recall	Precision	F1 score
1	ResNet50v2	1	99.96	100.00	99.94	1.00
		0	99.68	99.67	99.89	1.00
2	MobileNetv3small	1	86.88	97.48	86.25	0.92
		0	87.02	96.84	86.59	0.91
3	Xception	1	99.92	99.89	100.00	1.00
		0	99.44	99.55	99.66	1.00
4	ConvNextTiny	1	93.04	96.80	93.90	0.95
		0	93.23	97.47	93.19	0.95
5	EfficientNetV2S	1	98.21	98.83	98.67	0.99
		0	98.15	99.01	98.47	0.99
6	DenseNet121	1	99.92	99.95	99.95	1.00
		0	99.84	99.78	100.00	1.00
7	VGG16	1	99.01	99.12	99.50	0.99
		0	98.47	98.55	99.32	0.99

Table 5.6: Metrics generated from TL methods- on original images of 224*224 resolution; In the train column, one indicates train data, and zero indicates test data(COVID-19 data.



(a) GradCAM-VGG16 (b) GradCAM-ResNet (c) GradCAM-Mobilenet

(d) Saliency-VGG16 (e) Saliency-ResNet (f) Saliency-Mobilenet

Figure 5.5: Error curves-a,b,c are generated while training on the GradCAM fusion dataset; d,e,f are generated while training on the Saliency fusion dataset.

images by sampling from JSRT Dataset [3], NIH Chest X-ray 14 dataset [4]. The dataset used in the experiments consists of 318 nodule images sampled from the NIH Chest X-ray dataset and 1041 normal images sampled from the Kaggle pneumonia dataset.

[3]https://www.kaggle.com/datasets/raddar/nodules-in-chest-xrays-jsrt

[4]https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia

S.No	Fusion dataset : composition of features		train-1 test -0	Accuracy	Recall	Precision	F1score
	weight of global features	weight of local features					
1	0.5	0.5	1 0	99.63 99.25	98.41 97.26	100.00 100.00	99.20 98.61
2	0.6	0.4	1 0	99.91 99.63	99.59 100.00	100.00 98.67	99.79 99.33
3	0.7	0.3	1 0	99.91 99.63	99.61 100.0	100.00 98.41	99.80 99.20
4	0.8	0.2	1 0	98.43 99.26	97.29 98.33	95.80 98.33	96.54 98.33
5	0.9	0.1	1 0	99.17 98.16	96.60 95.24	100.00 96.77	98.27 96.00
6	0.1	0.9	1 0	98.71 98.16	96.03 94.67	98.37 98.61	97.19 96.60
7	0.2	0.8	1 0	99.36 97.06	97.77 91.07	99.62 94.44	98.69 92.73
8	0.3	0.7	1 0	92.53 98.90	50.00 100.00	52.38 94.92	51.16 97.39
9	0.4	0.6	1 0	100.00 99.63	100.00 98.57	100.00 100.00	100.00 99.28

Table 5.7: Metrics recorded on fusion dataset with various weight compositions of local and global features trained with deep learning model to predict lung nodules.

Similar CNN model trained for 50 epochs by setting parameter patience of ten with early stopping criteria. The weights vary from 0.1 to 0.9 for local and global features. The corresponding results are shown in the Table 5.7. The second and third columns show the feature's contribution to the final data. The local features are observed in small sizes in many images however in a few cases they can be seen on the sides of the lungs as shown in the Figure 5.6. Some of the error curves captured for the different models are shown in the 5.7 all the experiments were conducted in Google Collab[5] running in a local machine with 8 GB RAM.

Inaddition, Vision Transformers are used to extract features from the local features and classified COVID-19 from normal images. Images are trained with custom CNN with Adam optimizer and a learning rate of 0.001. The network is trained

[5]https://colab.research.google.com/

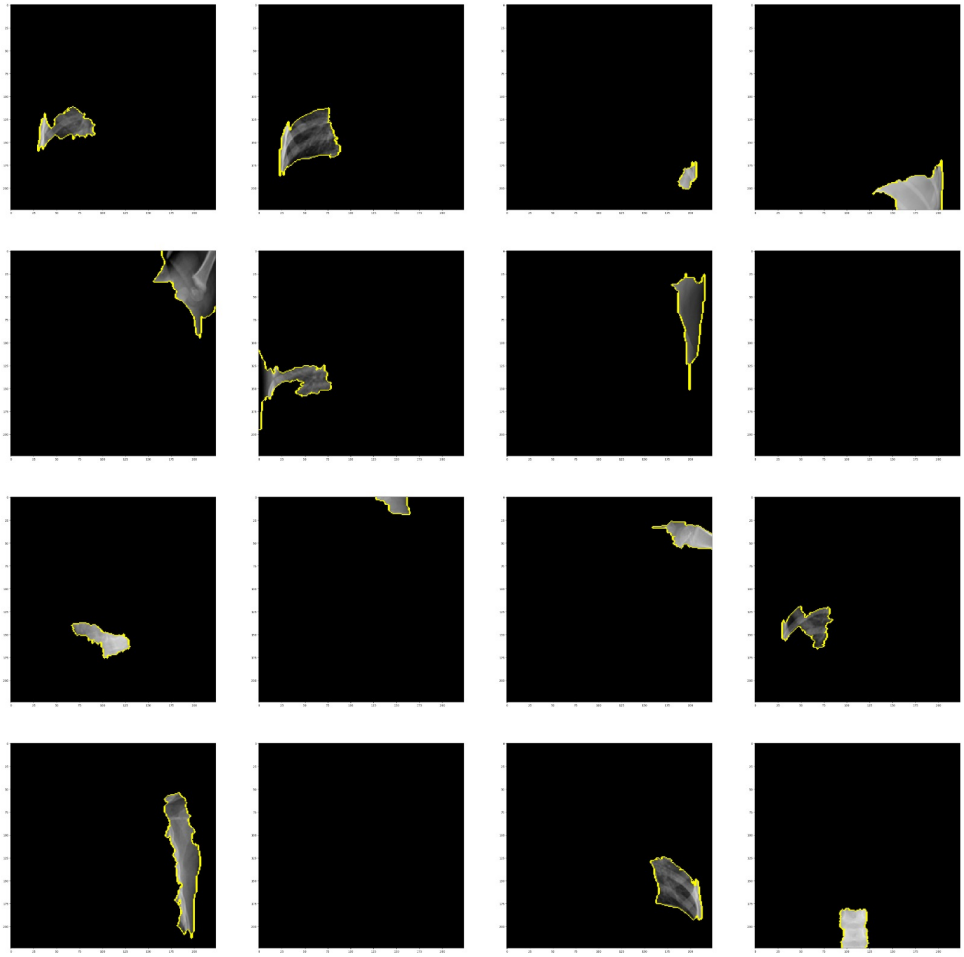


Figure 5.6: Local features observed from CXR images of Nodule class.

for 50 epochs with a patience parameter set to 10. The deep learning model has been saved in the h5 object. This makes the user load the model and image to get predictions. This model and each image have been passed to LIME to get the region of interest features. The region of interest features are combined with original images. Now each image has been divided into 64 patches with a patch of size 8*8. Each patch gets converted into a vector. Each vector will be passed to the transformer for further training. The model was trained for 100 epochs and the correpsonding error curves has been show in the Figure 5.8. Adam optimizer was used to optimize the weights. The experiment has shown an accuracy of 98.08% on test data when compared to the existing CNN performance [46, 192]. Corresponding confusion matrix on test data is shown in the Figure 5.9. We have used Google Colab free

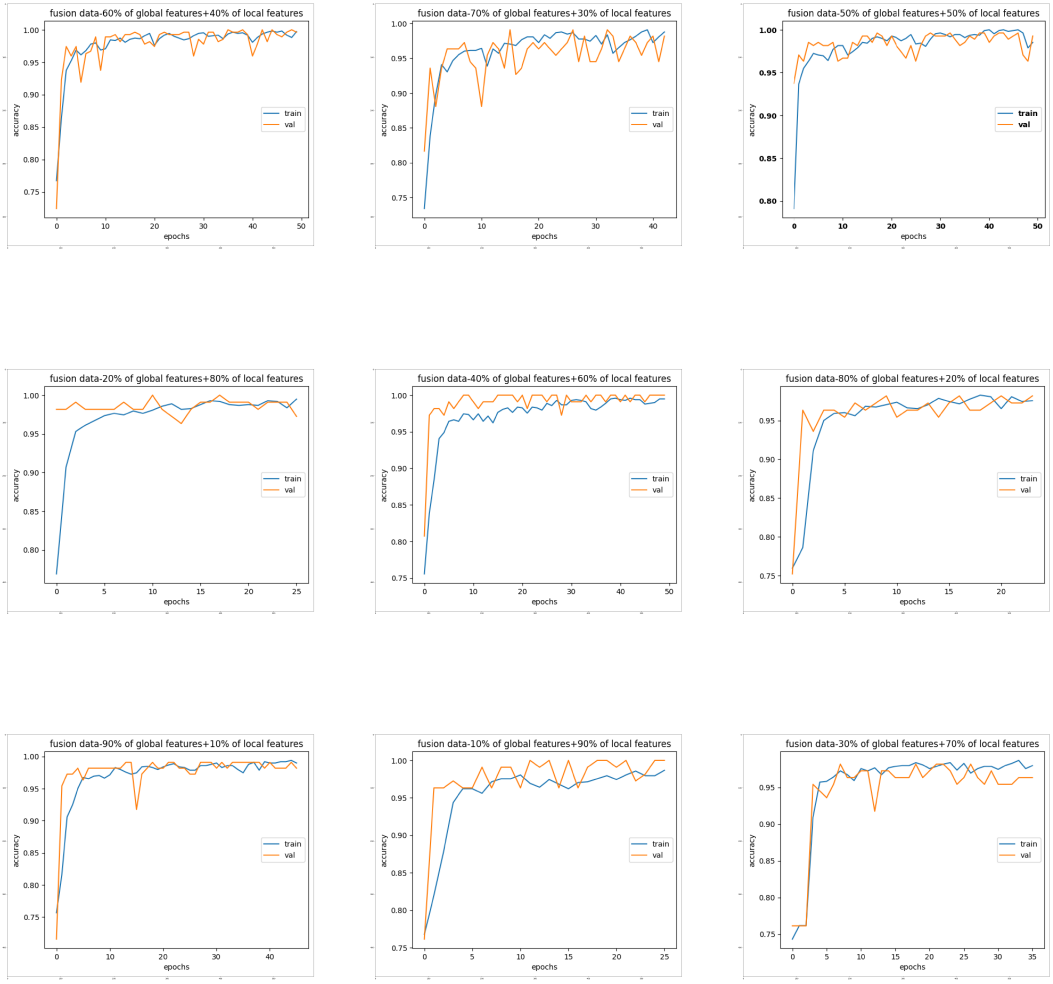


Figure 5.7: Error curves observed during training lung nodule detection.

GPU access to train transformers.

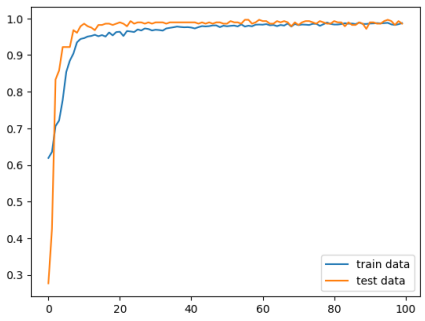


Figure 5.8: Error curve obtained during training-VIT

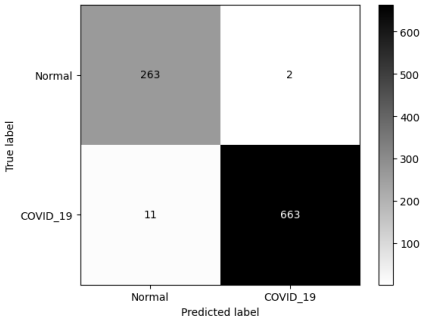


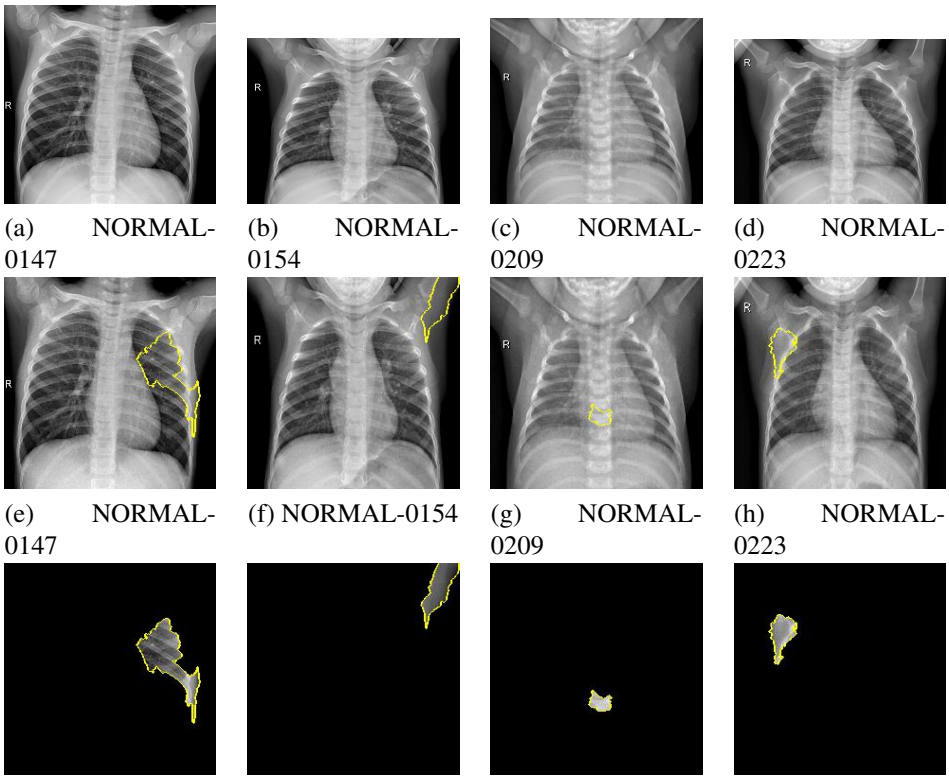
Figure 5.9: Confusion matrix recorded on test data-VIT

Chapter 6

DISCUSSION

6.1 INTRODUCTION

In this work, CXR images of different resolutions have been taken to produce the discriminative features through the XAI method in order to generalize the solution. XAI method LIME focuses on a part of the image(local features) to produce discriminative regions. Disease classification through local features has achieved convincing results. Several investigations [193, 194] have revealed that CXR images contain abnormalities that are consolidations that largely affect the perimeter of the contralateral lower lobes. The most common extrapulmonary findings were increased resistance to flow in the kidneys, thickness of arterial walls, fatty liver, pancreas, and heart inflammation. Localization results emphasized the region of the lower lobes; nevertheless, some of the upper lobes still displayed anomalies. Some of them are denoted at airways near the throat area in bubble size, which determines the starting stage of the disease. However, these are to be confirmed by an expert radiologist. Some standard images did not find discriminant features as part of the lungs—for example, the arm area and outside of the lobes. The proposed method has shown significant improvement over the base and state-of-the-art models. **Images misclassified with the base model are correctly classified with local discriminative feature training.** It is because of the following reasons. The local discriminant features learned through the XAI method on the base model impact other misclassified samples while training. It could happen by identifying the correct patterns, sometimes when a learned pattern is not similar to one class but is classified to another. When we did the sample-wise analysis, many images belong-



(a) NORMAL-0147 (b) NORMAL-0154 (c) NORMAL-0209 (d) NORMAL-0223

(e) NORMAL-0147 (f) NORMAL-0154 (g) NORMAL-0209 (h) NORMAL-0223

(i) NORMAL-0147 (j) NORMAL-0154 (k) NORMAL-0209 (l) NORMAL-0223

Figure 6.1: NORMAL images and their corresponding localization maps along with masks

ing to the normal class showed features outside of the lung regions, as shown in the Figure 6.1. A DT was built on the quadrant information to draw rules to classify images; and when we applied PCA to resultant features, it showed promising results. A sample tree is shown in the Figure 6.2 on the resolution of 150,150 images.

Since images are trained with varying resolutions, we are in a position to compare the local discriminant features. It helps to show consistency in the results when we see the exact location the majority time. However, there are cases where spatial information has changed from one to another. Sometimes, no local features were learned, leading to dropping of those images during training; however, the number of such images is less. **To overcome this problem, we have done a fusion of features. Fusion is to combine local and global features and trains a classifier. Local discriminant features are generated from various other XAI methods such as GradCAM, LIME, and saliency maps, and comparisons are**

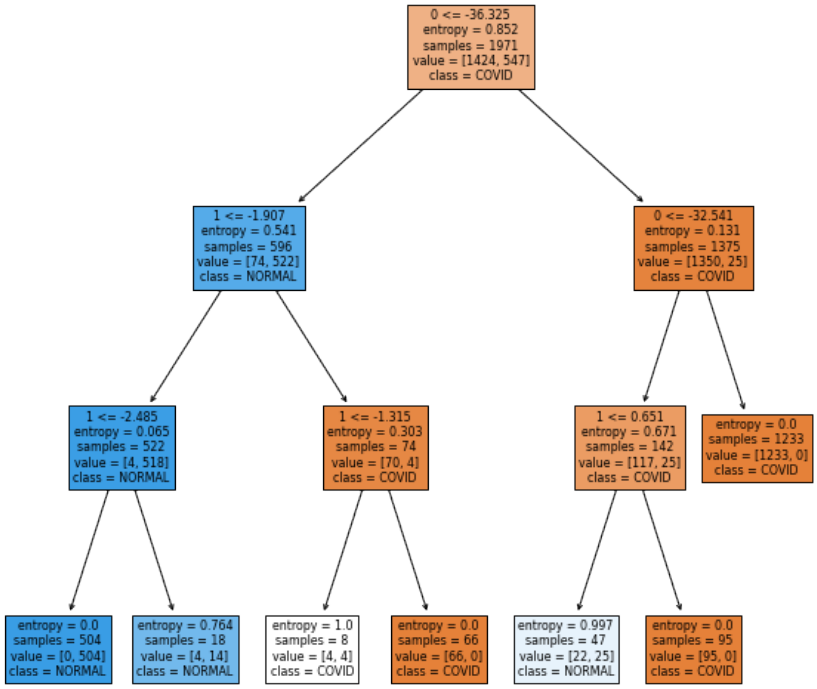


Figure 6.2: Decision tree built on top of three PCA components with a depth of three

made among these. In addition, an ensemble of these models was built.

The annotations produced by XAI methods like GradCAM, LIME, and saliency maps are subject to the base model built on the data, so it's critical to produce trustworthy interpretations based on a fixed model. Especially in contrast to individual models, Ensemble XAI has the benefit of stable interpretation because it simply applies weights to the appropriate pixel attributes by studying a limited number of annotations. By extracting the highly contributed pixel features through LIME and GradCAM, Ensemble XAI produces stable interpretation when combined with original data. Additionally, due to the presence of text, catheters, or lines in the X-ray image, the base heat maps produced by GradCAM and LIME rarely highlight the regions outside the lungs. Even though this distinct and interrupting area may be a sign of a serious lung condition, it is not useful for making decisions. In such images, the ensemble XAI performs better than the individual XAI by assigning less

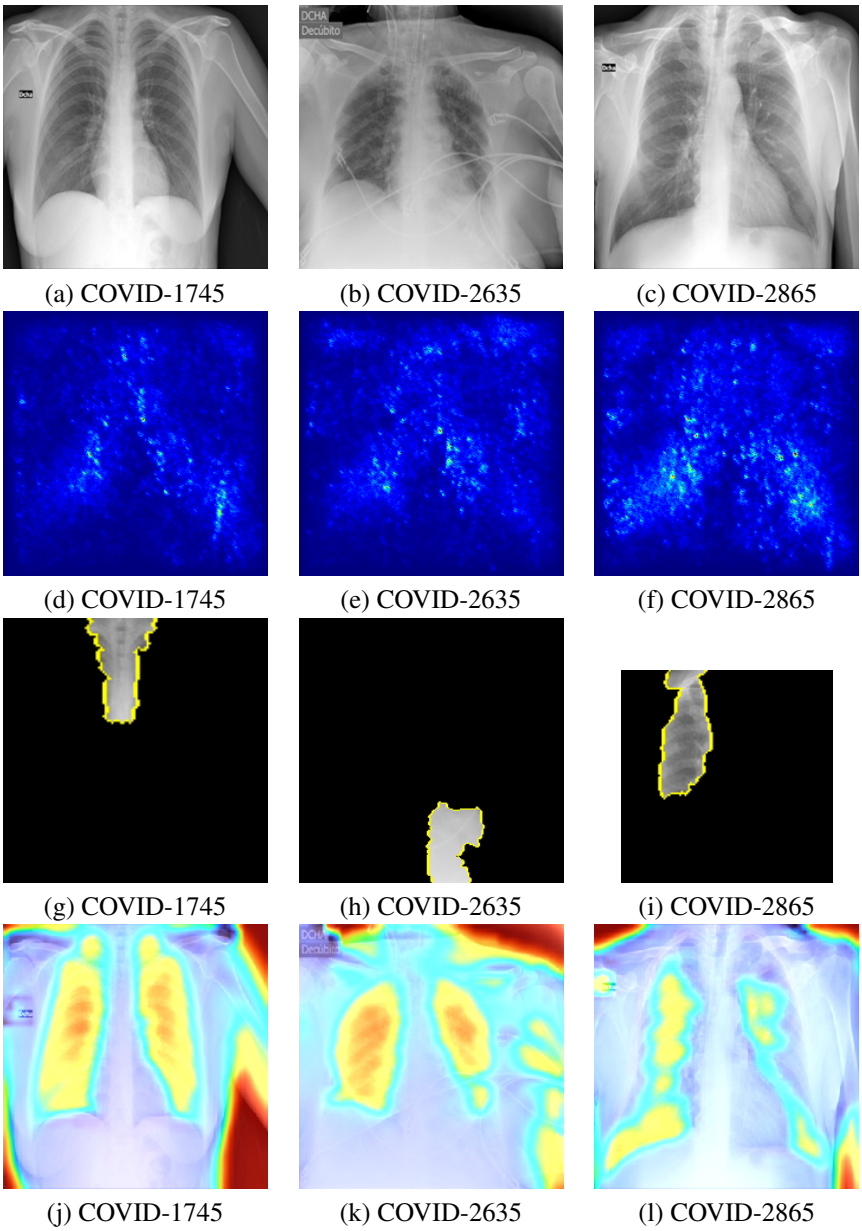
importance. Both, the original data and the annotated data, produced better outcomes when TL techniques were used. The topic of whether it is wise to rely on the transfer learning approach or the customized method has been raised, and the solution may lie in a combination of the two. It requires a great deal of experimentation and takes quite a while making predictions based on actual facts. However, if one has sufficient computing power can overcome the impediment. One can observe the variation between different XAI methods as show in the Figure 6.3.

The findings demonstrate that integrating local and global features through XAI methods, such as LIME and GradCAM, improves disease classification accuracy in CXR images and provides more interpretable model outputs. This approach enhances our understanding of AI interpretability by showing how localized feature analysis can identify clinically relevant regions, facilitating better transparency and trust in AI-assisted diagnostics. The ensemble XAI methods used offer stable, consistent interpretations, mitigating issues caused by artifacts and external factors. These insights suggest that XAI not only aids in explaining model decisions but also helps bridge the gap between AI and clinical expertise, promoting broader acceptance of AI tools in medical practice. This work is intended solely for academic and research purposes. It is generalized and can be applied across various levels by training the models on larger datasets. This approach could be further extended by collaborating with expert radiologists for real-time usage.

6.2 TIME COMPLEXITY ANALYSIS

The time complexity for the proposed approach is estimated in a model-specific way [195] as per the architecture in a total of three parts.

- one is training the base model before applying XAI methods.
- second, time taken to generate rois from model and XAI methods,say 'x' is the number of XAI methods.
- third, training fusion datasets.



(a) COVID-1745 (b) COVID-2635 (c) COVID-2865

(d) COVID-1745 (e) COVID-2635 (f) COVID-2865

(g) COVID-1745 (h) COVID-2635 (i) COVID-2865

(j) COVID-1745 (k) COVID-2635 (l) COVID-2865

Figure 6.3: XAI generated images - a,b,c are original images; d,e,f are saliency maps g,h,i are LIME generated masks;j,k,l are GradCAM output

The first and third uses the same architecture with the goal of capturing the features that are not captured before fusion. It has been defined [196] in the following way: $\sum_{i=1}^{d} f(m_i, n_i, k_i, s_i)$. Here, 'm' and 'n' denote the size of features, 'k' is the number of kernels, 's' is the size of the kernel, and 'd' denotes the depth of the network. The second part again consists of three parts since we have different XAI

methods. For simplification t_i is the time taken by an XAI method. It defined as: $t = f(XAImethod, Model)$, here the 'Model' is one of the components, because the image and model will be passed to the XAI method. Then the total time calculated as: $\sum_{i=1}^{x} f(t_i)$. XAI method time depends on the architecture of the model as well. Finally the total time is 2* $\sum_{i=1}^{d} f(m_i, n_i, k_i, s_i)$ + $\sum_{i=1}^{x} f(t_i)$. We have also measured the time for an epoch in each of these XAI methods. It excludes read/write intermediate results.

6.3 LIMITATIONS OF THE STUDY

The study faces several limitations, including its restricted focus on specific diseases, image quality variability, potential data biases, and dependency on particular model architectures, which can hinder its generalizability. The high computational complexity of the proposed methods also poses challenges for real-time clinical use. Moreover, interpreting XAI outputs without expert medical input can be problematic, as radiologists' insights are crucial for accurately correlating AI-generated visual explanations with clinical findings. Collaboration with radiologists could significantly improve the reliability of these interpretations, helping to bridge the gap between AI and clinical practice. While XAI-generated annotations offer a level of automation, they still require refinement to ensure clinical validity. Future research should focus on expanding disease coverage, standardizing datasets, optimizing computational efficiency, and integrating multi-modal data sources to enhance the model's clinical applicability, interpretability, and impact on healthcare workflows.

Chapter 7

CONCLUSION AND FUTURE SCOPE

7.1 CONCLUSION

This work investigates the finding of Lung disease that is COVID-19 with the help of localization approach, which is similar to radiologist way of looking at the image. It has been achieved through Ensemble of LIME, GradCAM and Saliency maps. Original image features and XAI generated features are combined with the fusion of local and global models. Specifically LIME, for the purpose of extracting local features from images, with a specific focus on C OVID-19. The LIME algorithm generates features in a random manner, occasionally obscuring the identity of the rightsholder. Consequently, further evidence is necessary to identify these features accurately. The findings exhibit promise; nonetheless, additional research on more extensive datasets is necessary. As the field of XAI continues to progress, the LIME technique will be subject to comparative analysis with more sophisticated methodologies in order to validate its outcomes. These attributes can be certified by radiologists based on their expertise and professional experience. On the other hand, GradCAM and saliency maps are also used to generate region of interest features which helped to assess LIME method annotations. In addition class imbalance is handled through a multi stage approach of undersampling.

This method is an versatile approach to computing local features through XAI methods and shown better accuracy of 98.85% and 99.62% for ensemble XAI and LIME respectively. With the transfer learning approaches XAI-Xception net gave the best test result of 99.84% and while handling class imbalance custom network gave 97.61% on test data. As the number of options are increasing as technology

progresses, one can choose heterogeneous approaches to build systems to predict lung disease in the near future. As the research grows, these XAI method-generated images to be inspected by radiologists and come up with the right annotations may increase the trust in the patients and healthcare community to make use of the automation process. The proposed approach may be extended to other disease images and in other modalities like CT scans, magnetic resonance imaging(MRI), etc. There is a possibility of building a segmentation model on each XAI method-generated output. Further ensemble of segmentation may provide the right annotations instead of depending on XAI methods once there is a well-trained segmentation model similar to [100]. Unet architecture is a network and a training strategy that has a contracting path to capture context and a symmetric expanding path that enables precise localization. Training such a network requires image masks along with images that need to be verified by experts. Here an attempt is made to build Ensemble XAI methods to predict lung disease (COVID-19 vs Normal) and it has produced sustainable results.

The dataset presented here is the initial outcome of feature extraction using a binary masking approach applied to the COVID illness. So far, it is the first data set produced by extracting features through a binary masking mechanism on COVID-19 disease as per the author's knowledge. The dataset(for more sample images, refer the Figure 7.1) will be shared on a special request. The dataset will be disseminated to other individuals and contributed to the research community for subsequent investigation.

7.2 FUTURE SCOPE

The authors have future plans to expand the application of the pipeline to encompass multiple lung illnesses by employing various XAI techniques. The proposed strategy has not yet been subjected to experimentation using transfer learning methods. The utilization of binary masks generated by the proposed methodology contributes to the development of an exemplary classifier. However, it is important to note that the determination of the disease's severity ultimately relies on the exper-

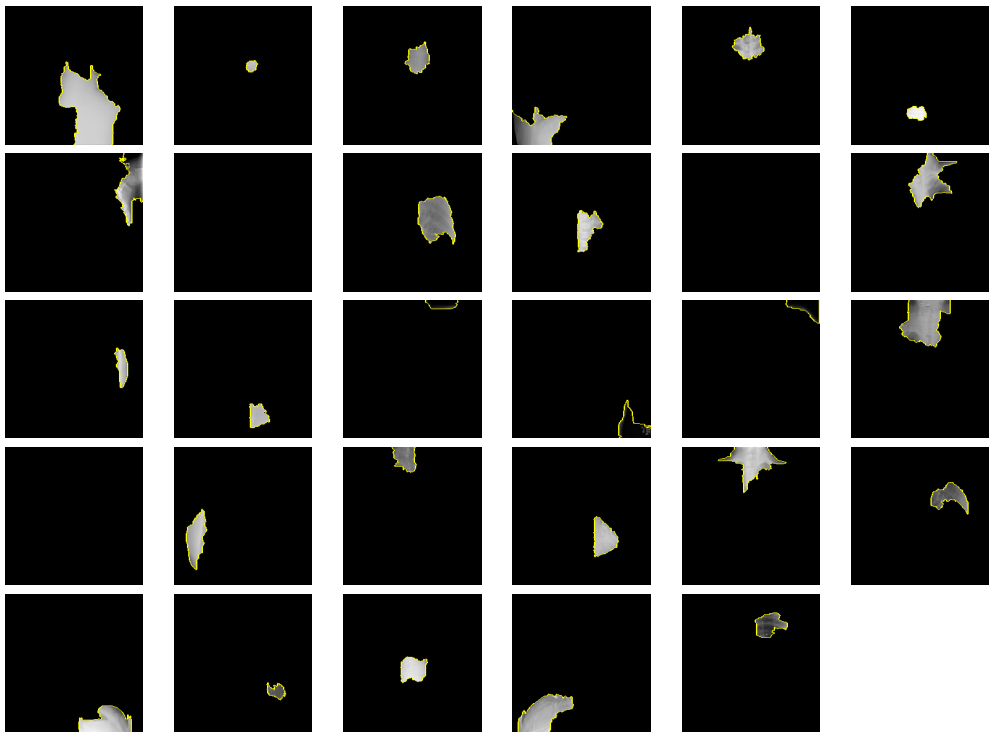


Figure 7.1: COVID images and their corresponding localization maps along with masks

tise and judgment of a skilled radiologist. The dataset pertaining to illness masks can effectively train segmentation networks and accurately detect local segments. As the field of study expands, the utilization of XAI holds potential for radiologists to examine and annotate them, thereby fostering enhanced confidence among patients and healthcare communities. The proposed methodology has the potential to be applied to a broader range of medical conditions and imaging techniques, such as CT scans and MRI. A segmentation model that has undergone extensive training, such as the Unet architecture, has the potential to accurately generate annotations without the need of XAI techniques.

REFERENCES

[1] Ksiazek, T. G., Erdman, D., Goldsmith, C. S., Zaki, S. R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J. A., Lim, W., Rollin, P. E., Dowell, S. F., Ling, A.-E., Humphrey, C. D., Shieh, W.-J., Guarner, J., Paddock, C. D., Rota, P., Fields, B., & DeRisi, J. (2003). A novel coronavirus associated with severe acute respiratory syndrome. *The New England Journal of Medicine*, *348*(20), 1953–1966. https://doi.org/10.1056/NEJMoa030781

[2] Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., & Zhang, Y.-Z. (2020). A new coronavirus associated with human respiratory disease in China. *Nature*, *579*(7798), 265–269. https://doi.org/10.1038/s41586-020-2008-3

[3] Miranda, E., Aryuni, M., & Irwansyah, E. (2016). A survey of medical image classification techniques. *2016 International Conference on Information Management and Technology (ICIMTech)*. https://doi.org/10.1109/icimtech.2016.7930302

[4] Cheng, C. P. (2019). Medical Imaging Modalities and Protocols. *Handbook of Vascular Motion*, 23–43. https://doi.org/10.1016/b978-0-12-815713-8.00003-6

[5] Lodwick, G. S. (1966). Computer-aided Diagnosis in Radiology. *Investigative Radiology*, *1*(1), 72–80. https://doi.org/10.1097/00004424-196601000-00032

[6] Lodwick, G. S., Keats, T. E., & Dorst, J. P. (1963). The Coding of Roentgen Images for Computer Analysis as Applied to Lung Cancer. *Radiology*, *81*(2), 185–200. https://doi.org/10.1148/81.2.185

[7] Toriwaki, J.-I., Suenaga, Y., Negoro, T., & Fukumura, T. (1973). Pattern recognition of chest X-ray images. *Computer Graphics and Image Processing*, *2*(3-4), 252–271. https://doi.org/10.1016/0146-664x(73)90005-1

[8] Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791

[9] None Ekata, Praveen Kumar Tyagi, Neeraj Kumar Gupta, & Gupta, S. (2016). *Diagnosis of Pulmonary Tuberculosis using fuzzy Inference System*. https://doi.org/10.1109/cipech.2016.7918726

[10] van Ginneken, B. (2017). Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. *Radiological Physics and Technology*, *10*(1), 23–32. https://doi.org/10.1007/s12194-017-0394-5

[11] Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C., Shpanskaya, K., Lungren, M. P., & Ng, A. Y. (2017). *CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning*. ArXiv.org. https://arxiv.org/abs/1711.05225

[12] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386

[13] Ashizawa, K., Ishida, T., MacMahon, H., Vyborny, C. J., Katsuragawa, S., & Doi, K. (1999). Artificial neural networks in chest radiography: Application to the differential diagnosis of interstitial lung disease. *Academic Radiology*, *6*(1), 2–9. https://doi.org/10.1016/s1076-6332(99)80055-5

[14] Heckerling, P. S., Gerber, B. S., Tape, T. G., & Wigton, R. S. (2003). Prediction of Community-Acquired Pneumonia Using Artificial Neural Networks. *Medical Decision Making*, *23*(2), 112–121. https://doi.org/10.1177/0272989x03251247

[15] Grafakou, O., Moustaki, M., Tsolia, M., Kavazarakis, E., Mathioudakis, J., Fretzayas, A., Nicolaidou, P., & Karpathios, T. (2004). Can chest X-ray predict pneumonia severity? *Pediatric Pulmonology*, *38*(6), 465–469. https://doi.org/10.1002/ppul.20112

[16] Franquet, T. (2001). Imaging of pneumonia: trends and algorithms. *European Respiratory Journal*, *18*(1), 196–208. https://doi.org/10.1183/09031936.01.00213501

[17] Oliveira, L. S., Almeida, S., Feitosa, M., Aparecido, R., Clarimar José Coelho, & Andrade, A. (2008). Computer-aided diagnosis in chest radiography for detection of childhood pneumonia. *International Journal of Medical Informatics*, *77*(8), 555–564.https://doi.org/10.1016/j.ijmedinf.2007.10.010

[18] Er, O., Yumusak, N., & Temurtas, F. (2010). Chest diseases diagnosis using artificial neural networks. *Expert Systems with Applications*, *37*(12), 7648–7655. https://doi.org/10.1016/j.eswa.2010.04.078

[19] Noor, N. Mohd., Rijal, O. Mohd., Yunus, A., & Abu-Bakar, S. A. R. (2010). A discrimination method for the detection of pneumonia using chest radiograph. *Computerized Medical Imaging and Graphics*, *34*(2), 160–166. https://doi.org/10.1016/j.compmedimag.2009.08.005

[20] Karargyris, A., Siegelman, J., Tzortzis, D., Jaeger, S., Candemir, S., Xue, Z., Santosh, K. C., Vajda, S., Antani, S., Folio, L., & Thoma, G. R. (2015). Combination of texture and shape features to detect pulmonary abnormalities in digital chest X-rays. *International Journal of Computer Assisted Radiology and Surgery*, *11*(1), 99–106. https://doi.org/10.1007/s11548-015-1242-x

[21] Melendez, J., Sánchez, C. I., Philipsen, R. H. H. M., Maduskar, P., Dawson, R., Theron, G., Dheda, K., & van Ginneken, B. (2016). An automated tuberculosis screening strategy combining X-ray-based computer-aided detection and clinical information. *Scientific Reports*, *6*(1). https://doi.org/10.1038/srep25265

[22] Khobragade, R. N., Kelkar, R. U., M, S., B, C., Murthy, N., Surendran, D., PS, R., & Balakrishnan, S. (2021). Health System Resilience: Ensuring TB services during COVID-19 pandemic in Kerala, India. *Indian Journal of Tuberculosis*. https://doi.org/10.1016/j.ijtb.2021.10.004

[23] Li, Q., Cai, W., Wang, X., Zhou, Y., Feng, D. D., & Chen, M. (2014). Medical image classification with convolutional neural network. *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*. https://doi.org/10.1109/icarcv.2014.7064414

[24] Lakhani, P., & Sundaram, B. (2017). Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. *Radiology*, *284*(2), 574–582. https://doi.org/10.1148/radiol.2017162326

[25] Lopes, U. K., & Valiati, J. F. (2017). Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. *Computers in Biology and Medicine*, *89*, 135–143. https://doi.org/10.1016/j.compbiomed.2017.08.001

[26] Liu, C., Cao, Y., Alcantara, M., Liu, B., Brunette, M., Peinado, J., & Curioso, W. (2017, September 1). *TX-CNN: Detecting tuberculosis in chest X-ray images using convolutional neural network*. IEEE Xplore. https://doi.org/10.1109/ICIP.2017.8296695

[27] Xu, X., Guo, Q., Guo, J., & Yi, Z. (2018). DeepCXray: Automatically Diagnosing Diseases on Chest X-Rays Using Deep Neural Networks. *IEEE Access*, *6*, 66972–66983. https://doi.org/10.1109/access.2018.2875406

[28] Heo, S.-J., Kim, Y., Yun, S., Lim, S.-S., Kim, J., Nam, C.-M., Park, E.-C., Jung, I., & Yoon, J.-H. (2019). Deep Learning Algorithms with Demographic Information Help to Detect Tuberculosis in Chest Radiographs in Annual Workers' Health Examination Data. *International Journal of Environmental Research and Public Health*, *16*(2), 250. https://doi.org/10.3390/ijerph16020250

[29] Nguyen, Q. H., Nguyen, B. P., Dao, S. D., Balagopal Unnikrishnan, Dhingra, R., Ravichandran, S., Satpathy, S. C., Palaparthi Nirmal Raja, & Chin, M. (2019). *Deep Learning Models for Tuberculosis Detection from Chest X-ray Images*. https://doi.org/10.1109/ict.2019.8798798

[30] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Dumitru Erhan, Vanhoucke, V., & Rabinovich, A. (2014). Going Deeper with Convolutions. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1409.4842

[31] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the Inception Architecture for Computer Vision. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2818–2826. https://doi.org/10.1109/cvpr.2016.308

[32] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/cvpr.2016.90

[33] Simonyan, K., & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Science*. https://doi.org/10.48550/arXiv.1409.1556

[34] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2261–2269. https://doi.org/10.1109/cvpr.2017.243

[35] Rajaraman, S., Candemir, S., Kim, I., Thoma, G., & Antani, S. (2018). Visualization and Interpretation of Convolutional Neural Network Predictions in Detecting Pneumonia in Pediatric Chest Radiographs. *Applied Sciences*, *8*(10), 1715. https://doi.org/10.3390/app8101715

[36] Stephen, O., Sain, M., Maduh, U. J., & Jeong, D.-U. (2019). An Efficient Deep Learning Approach to Pneumonia Classification in Healthcare. *Journal of Healthcare Engineering*, *2019*, 1–7. https://doi.org/10.1155/2019/4180949

[37] Zhang, R., Duan, H., Cheng, J., & Zheng, Y. (2020). A Study on Tuberculosis Classification in Chest X-ray Using Deep Residual Attention Networks. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. https://doi.org/10.1109/embc44109.2020.9175919

[38] Singhal, T. (2020). A review of coronavirus disease-2019 (COVID-19). *The Indian Journal of Pediatrics*, *87*(4), 281–286. https://doi.org/10.1007/s12098-020-03263-6

[39] Waleed Salehi, A., Baglat, P., & Gupta, G. (2020). Review on machine and deep learning models for the detection and prediction of Coronavirus. Materials Today. Proceedings, 33, 3896–3901. https://doi.org/10.1016/j.matpr.2020.06.245

[40] Alzubaidi, M., Zubaydi, H. D., Bin-Salem, A. A., Abd-Alrazaq, A. A., Ahmed, A., & Househ, M. (2021). Role of deep learning in early detection of COVID-19: Scoping review. *Computer Methods and Programs in Biomedicine Update*, *1*, 100025. https://doi.org/10.1016/j.cmpbup.2021.100025

[41] Adamidi, E. S., Mitsis, K., & Nikita, K. S. (2021). Artificial intelligence in clinical care amidst COVID-19 pandemic: A systematic review. *Computational and Structural Biotechnology Journal*, *19*, 2833–2850. https://doi.org/10.1016/j.csbj.2021.05.010

[42] Kora, P., Ooi, C. P., Faust, O., Raghavendra, U., Gudigar, A., Chan, W. Y., Meenakshi, K., Swaraja, K., Plawiak, P., & Rajendra Acharya, U. (2022). Transfer learning techniques for medical image analysis: A review. *Biocybernetics and Biomedical Engineering*, *42*(1), 79–107. https://doi.org/10.1016/j.bbe.2021.11.004

[43] Wang, H., Gu, H., Qin, P., & Wang, J. (2020). CheXLocNet: Automatic localization of pneumothorax in chest radiographs using deep convolutional neural networks. *PLOS ONE*, *15*(11), e0242013. https://doi.org/10.1371/journal.pone.0242013

[44] Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Rajendra Acharya, U. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine*. https://doi.org/10.1016/j.compbiomed.2020.103792

[45] Tan, M., & Le, Q. V. (2020, September 11). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. ArXiv.org. https://doi.org/10.48550/arXiv.1905.11946

[46] Hemdan, E. E.-D., Shouman, M. A., & Karar, M. E. (2020). COVIDX-Net: A Framework of Deep Learning Classifiers to Diagnose COVID-19 in X-Ray Images. *ArXiv:2003.11055 [Cs, Eess]*. https://arxiv.org/abs/2003.11055

[47] Farooq, M., & Hafeez, A. (2020, March 1). *COVID-ResNet: A Deep Learning Framework for Screening of COVID19 from Radiographs*. NASA ADS. https://doi.org/10.48550/arXiv.2003.14395

[48] Minaee, S., Kafieh, R., Sonka, M., Yazdani, S., & Jamalipour Soufi, G. (2020). Deep-COVID: Predicting COVID-19 From Chest X-Ray Images Using Deep Transfer Learning. *Medical Image Analysis*, 101794. https://doi.org/10.1016/j.media.2020.101794

[49] Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. (2016). SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1602.07360

[50] Kedia, P., Anjum, & Katarya, R. (2021). CoVNet-19: A Deep Learning model for the detection and analysis of COVID-19 patients. *Applied Soft Computing*, *104*, 107184. https://doi.org/10.1016/j.asoc.2021.107184

[51] Hussain, E., Hasan, M., Rahman, M. A., Lee, I., Tamanna, T., & Parvez, M. Z. (2020). CoroDet: A deep learning based classification for COVID-19 detection using chest X-ray images. *Chaos, Solitons & Fractals*, 110495. https://doi.org/10.1016/j.chaos.2020.110495

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

[53] Chollet, F. (2017). Xception: Deep Learning with Depthwise Separable Convolutions. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2017.195

[54] Luz, E., Silva, P., Silva, R., Silva, L., Guimarães, J., Miozzo, G., Moreira, G., & Menotti, D. (2021). Towards an effective and efficient deep learning model for COVID-19 patterns detection in X-ray images. *Research on Biomedical Engineering*. https://doi.org/10.1007/s42600-021-00151-6

[55] Panwar, H., Gupta, P. K., Siddiqui, M. K., Morales-Menendez, R., Bhardwaj, P., & Singh, V. (2020). A Deep Learning and Grad-CAM based Color Visualization Approach for Fast Detection of COVID-19 Cases using Chest X-ray and CT-Scan Images. *Chaos, Solitons & Fractals*, 110190. https://doi.org/10.1016/j.chaos.2020.110190

[56] Oh, Y., Park, S., & Ye, J. C. (2020). Deep Learning COVID-19 Features on CXR using Limited Training Data Sets. *IEEE Transactions on Medical Imaging*, 1–1. https://doi.org/10.1109/TMI.2020.2993291

[57] Núñez-Regueiro, M. (2020). Yaşlı Kadınlarda Üreme Sağlığı. *DergiPark (Istanbul University)*, *1*(1). https://doi.org/10.1016/j

[58] Shaban, W. M., Rabie, A. H., Saleh, A. I., & Abo-Elsoud, M. A. (2021). Accurate detection of COVID-19 patients based on distance biased Naïve Bayes (DBNB) classification strategy. *Pattern Recognition*, *119*, 108110. https://doi.org/10.1016/j.patcog.2021.108110

[59] Dey, S., Bhattacharya, R., Malakar, S., Mirjalili, S., & Sarkar, R. (2021). Choquet fuzzy integral-based classifier ensemble technique for COVID-19 detection. *Computers in Biology and Medicine*, *135*, 104585. https://doi.org/10.1016/j.compbiomed.2021.104585

[60] Apostolopoulos, I. D., & Mpesiana, T. A. (2020). Covid-19: automatic detection from X-ray images utilizing transfer learning with convolutional neural networks. *Physical and Engineering Sciences in Medicine*. https://doi.org/10.1007/s13246-020-00865-4

[61] Redmon, J., & Farhadi, A. (2017). YOLO9000: Better, Faster, Stronger. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/cvpr.2017.690

[62] Toğaçar, M., Ergen, B., & Cömert, Z. (2020). COVID-19 detection using deep learning models to exploit Social Mimic Optimization and structured chest X-ray images using fuzzy color and stacking approaches. *Computers in Biology and Medicine*, *121*, 103805. https://doi.org/10.1016/j.compbiomed.2020.103805

[63] Gaál, G., Maga, B., & Lukács, A. (2020, March 23). *Attention U-Net Based Adversarial Architectures for Chest X-ray Lung Segmentation*. ArXiv.org. https://doi.org/10.48550/arXiv.2003.10304

[64] Rajaraman, S., Siegelman, J., Alderson, P. O., Folio, L. S., Folio, L. R., & Antani, S. K. (2020). Iteratively Pruned Deep Learning Ensembles for COVID-19 Detection in Chest X-rays. *IEEE Access*, 1–1. https://doi.org/10.1109/access.2020.3003810

[65] Hassantabar, S., Ahmadi, M., & Sharifi, A. (2020). Diagnosis and detection of infected tissue of COVID-19 patients based on lung x-ray image using convolutional neural

network approaches. *Chaos, Solitons & Fractals*, *140*, 110170. https://doi.org/10.1016/j.chaos.2020.110170

[66] Giger, M. L., Doi, K., MacMahon, H., Metz, C. E., & Yin, F. F. (1990). Pulmonary nodules: computer-aided detection in digital chest images. *RadioGraphics*, *10*(1), 41–51. https://doi.org/10.1148/radiographics.10.1.2296696

[67] Suzuki, H., Inaoka, N., Hirotsugu Takabatake, Mori, M., Hiroshi Natori, & Suzuki, A. (1991). Experimental system for detecting lung nodules by chest x-ray image processing. *Proceedings of SPIE, the International Society for Optical Engineering/Proceedings of SPIE*. https://doi.org/10.1117/12.44289

[68] Sanada, S., Doi, K., & MacMahon, H. (1992). Image feature analysis and computer-aided diagnosis in digital radiography: Automated detection of pneumothorax in chest images. *Medical Physics*, *19*(5), 1153–1160. https://doi.org/10.1118/1.596790

[69] Lo, S.-C. B., Lin, J.-S., Freedman, M. T., & Mun, S. K. (1993). Computer-assisted diagnosis of lung nodule detection using artificial convoultion neural network. *NASA ADS*, *1898*, 859–869. https://doi.org/10.1117/12.154572

[70] Lo, S.-C. .B., Lou, S.-L. .A., Jyh-Shyan Lin, Freedman, M. T., Chien, M. V., & Mun, S. K. (1995). Artificial convolution neural network techniques and applications for lung nodule detection. *IEEE Transactions on Medical Imaging*, *14*(4), 711–718. https://doi.org/10.1109/42.476112

[71] Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, *36*(4), 193–202. https://doi.org/10.1007/bf00344251

[72] Fukushima, K. (1975). Cognitron: A self-organizing multilayered neural network. *Biological Cybernetics*, *20*(3-4), 121–136. https://doi.org/10.1007/bf00342633

[73] Stoffey, R. D. (2013). Primer of Diagnostic Imaging, 5th ed.Primer of Diagnostic Imaging, 5th ed. By Ralph Weissleder, Jack Wittenberg, Mukesh Harisinghani, and John W. Chen. St. Louis, MO: Mosby Elsevier, 816 pp., 2011. $127.95 softcover

(ISBN: 978-0323065382). *American Journal of Roentgenology*, *200*(6), W690–W690. https://doi.org/10.2214/ajr.12.10515

[74] Haghanifar, A., Majdabadi, M. M., Choi, Y., Deivalakshmi, S., & Ko, S. (2022). COVID-CXNet: Detecting COVID-19 in frontal chest X-ray images using deep learning. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-022-12156-z

[75] Saha, P., & Neogy, S. (2022). Concat_CNN: A Model to Detect COVID-19 from Chest X-ray Images with Deep Learning. *SN Computer Science*, *3*(4). https://doi.org/10.1007/s42979-022-01182-1

[76] Geetha, R., Balasubramanian, M., & Devi, K. R. (2022). COVIDetection: deep convolutional neural networks-based automatic detection of COVID-19 with chest x-ray images. *Research on Biomedical Engineering*. https://doi.org/10.1007/s42600-022-00230-2

[77] Agrawal, S., Venkatesh Honnakasturi, Nara, M., & Patil, N. (2023). Utilizing Deep Learning Models and Transfer Learning for COVID-19 Detection from X-Ray Images. *SN Computer Science*, *4*(4). https://doi.org/10.1007/s42979-022-01655-3

[78] Otsu, N. (1979). A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. https://doi.org/10.1109/tsmc.1979.4310076

[79] Russ, J. C. (2017). Introduction to Image Processing and Analysis. In *CRC Press eBooks*. Informa. https://doi.org/10.1201/9781315221939

[80] Joshi, D., & Singh, T. P. (2020). A survey of fracture detection techniques in bone X-ray images. *Artificial Intelligence Review*. https://doi.org/10.1007/s10462-019-09799-0

[81] Van Ginneken, B., Ter Haar Romeny, B. M., & Viergever, M. A. (2001). Computer-aided diagnosis in chest radiography: a survey. *IEEE Transactions on Medical Imaging*, *20*(12), 1228–1241. https://doi.org/10.1109/42.974918

[82] Wang, L., Lin, Z. Q., & Wong, A. (2020). COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-76550-z

[83] Chowdhury, M. E. H., Rahman, T., Khandakar, A., Mazhar, R., Kadir, M. A., Mahbub, Z. B., Islam, K. R., Khan, M. S., Iqbal, A., Emadi, N. A., Reaz, M. B. I., & Islam, M. T. (2020). Can AI Help in Screening Viral and COVID-19 Pneumonia? *IEEE Access*, *8*, 132665–132676. https://doi.org/10.1109/access.2020.3010287

[84] Howard, J., & Ruder, S. (2018). Universal Language Model Fine-tuning for Text Classification. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1801.06146

[85] Sethy, P. K., Behera, S. K., Ratha, P. K., & Biswas, P. (2020). Detection of coronavirus Disease (COVID-19) based on Deep Features and Support Vector Machine. *International Journal of Mathematical, Engineering and Management Sciences*, *5*(4), 643–651. https://doi.org/10.33889/ijmems.2020.5.4.052

[86] Patrascu, V., & Buzuloiu, V. (2005). *Color Image Enhancement Using the Support Fuzzification in the Framework of the Logarithmic Model*. Semantic Scholar. https://api.semanticscholar.org/CorpusID:125890228

[87] Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for Activation Functions. *Arxiv.org*. https://doi.org/10.48550/arXiv.1710.05941

[88] Kingma, D., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *Computer Science*. https://doi.org/10.48550/arXiv.1412.6980

[89] Recognition of COVID-19 disease from X-ray images by hybrid model consisting of 2D curvelet transform, chaotic salp swarm algorithm and deep learning technique. (2020). *Chaos, Solitons & Fractals*, *140*, 110071. https://doi.org/10.1016/j.chaos.2020.110071

[90] Waheed, A., Goyal, M., Gupta, D., Khanna, A., Al-Turjman, F., & Pinheiro, P. R. (2020). CovidGAN: Data Augmentation using Auxiliary Classifier GAN for Improved Covid-19 Detection. *IEEE Access*, 1–1. https://doi.org/10.1109/ACCESS.2020.2994762

[91] Chauhan, T., Palivela, H., & Tiwari, S. (2021). Optimization and fine-tuning of DenseNet model for classification of COVID-19 cases in medical imaging. *International Journal of Information Management Data Insights*, *1*(2), 100020. https://doi.org/10.1016/j.jjimei.2021.100020

[92] Huy Quan Vu, Gleb Beliakov, & Li, G. (2014). A Choquet Integral Toolbox and Its Application in Customer Preference Analysis. *Elsevier EBooks*, 247–272. https://doi.org/10.1016/b978-0-12-411511-8.00009-8

[93] Fan, Y., Liu, J., Yao, R., & Yuan, X. (2021). COVID-19 Detection from X-ray Images using Multi-Kernel-Size Spatial-Channel Attention Network. *Pattern Recognition*, *119*, 108055. https://doi.org/10.1016/j.patcog.2021.108055

[94] Saddam Hussain Khan, Sohail, A., Khan, A., Hassan, M., Yeon Joo Lee, Alam, J., Basit, A., & Zubair, S. (2021). *COVID-19 detection in chest X-ray images using deep boosted hybrid learning*. *137*, 104816–104816. https://doi.org/10.1016/j.compbiomed.2021.104816

[95] Kumar, M., Shakya, D., Kurup, V., & Suksatan, W. (2021). COVID-19 prediction through X-ray images using transfer learning-based hybrid deep learning approach. *Materials Today: Proceedings*. https://doi.org/10.1016/j.matpr.2021.12.123

[96] Hu, T., Khishe, M., Mohammadi, M., Parvizi, G.-R., Taher Karim, S. H., & Rashid, T. A. (2021). Real-time COVID-19 diagnosis from X-Ray images using deep CNN and extreme learning machines stabilized by chimp optimization algorithm. *Biomedical Signal Processing and Control*, *68*, 102764. https://doi.org/10.1016/j.bspc.2021.102764

[97] Huang, G.-B., Wang, D. H., & Lan, Y. (2011). Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics*, *2*(2), 107–122. https://doi.org/10.1007/s13042-011-0019-y

[98] Khishe, M., & Mosavi, M. R. (2020). Chimp optimization algorithm. *Expert Systems with Applications*, *149*, 113338. https://doi.org/10.1016/j.eswa.2020.113338

[99] Tahir, A. M., Chowdhury, M. E. H., Khandakar, A., Rahman, T., Qiblawey, Y., Khurshid, U., Kiranyaz, S., Ibtehaz, N., Rahman, M. S., Al-Maadeed, S., Mahmud, S.,

Ezeddin, M., Hameed, K., & Hamid, T. (2021). COVID-19 infection localization and severity grading from chest X-ray images. *Computers in Biology and Medicine*, *139*, 105002. https://doi.org/10.1016/j.compbiomed.2021.105002

[100] Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. (2015). In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Lecture Notes in Computer Science*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4

[101] Balaha, H. M., Balaha, M. H., & Ali, H. A. (2021). Hybrid COVID-19 segmentation and recognition framework (HMB-HCF) using deep learning and genetic algorithms. *Artificial Intelligence in Medicine*, 102156. https://doi.org/10.1016/j.artmed.2021.102156

[102] Vieira, P., Sousa, O., Magalhães, D., Rabêlo, R., & Silva, R. (2021). Detecting pulmonary diseases using deep features in X-ray images. *Pattern Recognition*, *119*, 108081. https://doi.org/10.1016/j.patcog.2021.108081

[103] Zhong, A., Li, X., Wu, D., Ren, H., Kim, K., Kim, Y., Buch, V., Neumark, N., Bizzo, B., Tak, W. Y., Park, S. Y., Lee, Y. R., Kang, M. K., Park, J. G., Kim, B. S., Chung, W. J., Guo, N., Dayan, I., Kalra, M. K., & Li, Q. (2021). Deep metric learning-based image retrieval system for chest radiograph and its clinical applications in COVID-19. *Medical Image Analysis*, *70*, 101993. https://doi.org/10.1016/j.media.2021.101993

[104] Wang, C., Elazab, A., Wu, J., & Hu, Q. (2017). Lung nodule classification using deep feature fusion in chest radiography. *Computerized Medical Imaging and Graphics*, *57*, 10–18. https://doi.org/10.1016/j.compmedimag.2016.11.004

[105] Masud, M. (2022). A light-weight convolutional Neural Network Architecture for classification of COVID-19 chest X-Ray images. *Multimedia Systems*. https://doi.org/10.1007/s00530-021-00857-8

[106] Maguolo, G., & Nanni, L. (2021). A critic evaluation of methods for COVID-19 automatic detection from X-ray images. *Information Fusion*, *76*, 1–7. https://doi.org/10.1016/j.inffus.2021.04.008

[107] Redie, D. K., Sirko, A. E., Demissie, T. M., Teferi, S. S., Shrivastava, V. K., Verma, O. P., & Sharma, T. K. (2022). Diagnosis of COVID-19 using chest X-ray images based on modified DarkCovidNet model. *Evolutionary Intelligence*. https://doi.org/10.1007/s12065-021-00679-7

[108] Gupta, V., Jain, N., Sachdeva, J., Gupta, M., Mohan, S., Mohd Yazid Bajuri, & Ahmadian, A. (2022). Improved COVID-19 detection with chest x-ray images using deep learning. *Multimedia Tools and Applications*, *81*(26), 37657–37680. https://doi.org/10.1007/s11042-022-13509-4

[109] Agrawal, T., & Choudhary, P. (2021). FocusCovid: automated COVID-19 detection using deep learning with chest X-ray images. *Evolving Systems*. https://doi.org/10.1007/s12530-021-09385-2

[110] Asif, S., Zhao, M., Tang, F., & Zhu, Y. (2022). A deep learning-based framework for detecting COVID-19 patients using chest X-rays. *Multimedia Systems*. https://doi.org/10.1007/s00530-022-00917-7

[111] K. Kiran Kumar, & K.V. Kanimozhi. (2022). Detection of Lung Disease using Novel Genetic Algorithm in comparison with Particle Swarm Optimization to improve Accuracy. *2022 International Conference on Business Analytics for Technology and Security (ICBATS)*. https://doi.org/10.1109/icbats54253.2022.9759044

[112] Chow, L. S., Tang, G. S., Solihin, M. I., Gowdh, N. M., Ramli, N., & Rahmat, K. (2023). Quantitative and Qualitative Analysis of 18 Deep Convolutional Neural Network (CNN) Models with Transfer Learning to Diagnose COVID-19 on Chest X-Ray (CXR) Images. *SN Computer Science*, *4*(2). https://doi.org/10.1007/s42979-022-01545-8

[113] Pradhan, A. K., Mishra, D., Das, K., Obaidat, M. S., & Kumar, M. (2022). A COVID-19 X-ray image classification model based on an enhanced convolutional neural network and hill climbing algorithms. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-022-13826-8

[114] Qi, A., Zhao, D., Yu, F., Ali Asghar Heidari, Cui, Z., Cai, Z., Fayadh Alenezi, Mansour, R. F., Chen, H., & Chen, M. (2022). Directional mutation and crossover boosted ant colony optimization with application to COVID-19 X-ray image

segmentation. *Computers in Biology and Medicine*, *148*, 105810–105810. https://doi.org/10.1016/j.compbiomed.2022.105810

[115] Goel, T., Murugan, R., Mirjalili, S., & Chakrabartty, D. K. (2021). Multi-COVID-Net: Multi-objective optimized network for COVID-19 diagnosis from chest X-ray images. *Applied Soft Computing*, 108250. https://doi.org/10.1016/j.asoc.2021.108250

[116] Manickam, A., Jiang, J., Zhou, Y., Sagar, A., Soundrapandiyan, R., & Dinesh Jackson Samuel, R. (2021). Automated pneumonia detection on chest X-ray images: A deep learning approach with different optimizers and transfer learning architectures. *Measurement*, *184*, 109953. https://doi.org/10.1016/j.measurement.2021.109953

[117] Parveen, N. R. S., & Sathik, M. M. (2011). Detection of Pneumonia in chest X-ray images. *Journal of X-Ray Science and Technology*, *19*(4), 423–428. https://doi.org/10.3233/xst-2011-0304

[118] Parveen, N. R. S., & Sathik, M. M. (2011). Detection of Pneumonia in chest X-ray images. *Journal of X-Ray Science and Technology*, *19*(4), 423–428. https://doi.org/10.3233/xst-2011-0304

[119] Rajaraman, S., & Antani, S. K. (2020). Modality-Specific Deep Learning Model Ensembles Toward Improving TB Detection in Chest Radiographs. *IEEE Access*, *8*, 27318–27326. https://doi.org/10.1109/access.2020.2971257

[120] Mahmud, T., Rahman, M. A., & Fattah, S. A. (2020). CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. *Computers in Biology and Medicine*, *122*, 103869. https://doi.org/10.1016/j.compbiomed.2020.103869

[124] Adedigba, A. P., Adeshina, S. A., Aina, O. E., & Aibinu, A. M. (2021). Optimal hyperparameter selection of deep learning models for COVID-19 chest X-ray classification. *Intelligence-Based Medicine*, *5*, 100034. https://doi.org/10.1016/j.ibmed.2021.100034

[125] Ratnasari Nur Rohmah, Susanto, A., & Indah Soesanti. (2013). *Lung tuberculosis identification based on statistical feature of thoracic X-ray*. https://doi.org/10.1109/qir.2013.6632528

[126] Guan, Q., Huang, Y., Zhong, Z., Zheng, Z., Zheng, L., & Yang, Y. (2020). Thorax disease classification with attention guided convolutional neural network. *Pattern Recognition Letters*, *131*, 38–45. https://doi.org/10.1016/j.patrec.2019.11.040

[127] He, K., Gkioxari, G., Dollar, P., & Girshick, R. (2017). Mask R-CNN. *2017 IEEE International Conference on Computer Vision (ICCV)*. https://doi.org/10.1109/iccv.2017.322

[128] Rahman, M., Cao, Y., Sun, X., Li, B., & Hao, Y. (2021). Deep pre-trained networks as a feature extractor with XGBoost to detect tuberculosis from chest X-ray. *Computers & Electrical Engineering*, *93*, 107252–107252. https://doi.org/10.1016/j.compeleceng.2021.107252

[129] Verma, D., Bose, C., Tufchi, N., Pant, K., Tripathi, V., & Thapliyal, A. (2020). An efficient framework for identification of Tuberculosis and Pneumonia in chest X-ray images using Neural Network. *Procedia Computer Science*, *171*, 217–224. https://doi.org/10.1016/j.procs.2020.04.023

[130] Antani, S., Candemir, S., Jaeger, P.F.S., Karargyris, R.F.A., McDonald, C., Kohli, M., Siegelman, J.(2015). Auto- mated detection of lung diseases in chest x-rays a report to the board of scientific counselors. Technical Report to the LHNCBC Board of Scientific Counselors.

[131] Jaeger, S., Karargyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., Zhiyun Xue, Palaniappan, K., Singh, R. K., Antani, S., Thoma, G., Yi-Xiang Wang, Pu-Xuan Lu, & McDonald, C. J. (2014). Automatic Tuberculosis Screening Using Chest Radiographs. *IEEE Transactions on Medical Imaging*, *33*(2), 233–245. https://doi.org/10.1109/tmi.2013.2284099

[132] Hariharan, S., Ray, A. K., & Ghosh, M. K. (2005). *An algorithm for the enhancement of chest X-ray images of tuberculosis patients*. https://doi.org/10.1109/icit.2000.854108

[133] Hashmi, M. F., Katiyar, S., Keskar, A. G., Bokde, N. D., & Geem, Z. W. (2020). Efficient Pneumonia Detection in Chest Xray Images Using Deep Transfer Learning. *Diagnostics*, *10*(6), 417. https://doi.org/10.3390/diagnostics10060417

[134] Carreira, M. J., Cabello, D., Mosquera, A., & Penedo, M. G. (1992). Medical images segmentation using region and edges information. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. https://doi.org/10.1109/iembs.1992.5762095

[135] van Ginneken, B., & ter Haar Romeny, B. M. (2000). Automatic segmentation of lung fields in chest radiographs. *Medical Physics*, *27*(10), 2445–2455. https://doi.org/10.1118/1.1312192

[136] Zhou, Z., Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, & Liang, J. (2018). *UNet++: A Nested U-Net Architecture for Medical Image Segmentation*. https://doi.org/10.48550/arxiv.1807.10165

[137] Souza, J. C., Bandeira Diniz, J. O., Ferreira, J. L., França da Silva, G. L., Corrêa Silva, A., & de Paiva, A. C. (2019). An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks. *Computer Methods and Programs in Biomedicine*, *177*, 285–296. https://doi.org/10.1016/j.cmpb.2019.06.005

[138] Tan, M., & Le, Q. V. (2020, September 11). *EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks*. ArXiv.org. https://doi.org/10.48550/arXiv.1905.11946

[139] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., M. Andreetto, & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1704.04861

[140] Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement. *Arxiv.org*. https://doi.org/10.48550/arXiv.1804.02767

[141] Xiao, H., Li, L., Liu, Q., Zhu, X., & Zhang, Q. (2023). Transformers in medical image segmentation: A review. *Biomedical Signal Processing and Control*, *84*, 104791. https://doi.org/10.1016/j.bspc.2023.104791

[142] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021, June 3). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. ArXiv.org. https://doi.org/10.48550/arXiv.2010.11929

[143] Azad, R., Kazerouni, A., Heidari, M., Aghdam, E. K., Molaei, A., Jia, Y., Jose, A., Roy, R., & Merhof, D. (2024). Advances in medical image analysis with vision Transformers: A comprehensive review. *Medical Image Analysis*, *91*, 103000. https://doi.org/10.1016/j.media.2023.103000

[144] Krishnan, K. S., & Krishnan, K. S. (2021). Vision Transformer based COVID-19 Detection using Chest X-rays. *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)*, 644–648. https://doi.org/10.1109/ISPCC53510.2021.9609375

[145] Tyagi, K., Pathak, G., Nijhawan, R., & Mittal, A. (2021, December 1). *Detecting Pneumonia using Vision Transformer and comparing with other techniques*. IEEE Xplore. https://doi.org/10.1109/ICECA52323.2021.9676146

[146] Duong, L. T., Le, N. H., Tran, T. B., Ngo, V. M., & Nguyen, P. T. (2021). Detection of tuberculosis from chest X-ray images: Boosting the performance with vision transformer and transfer learning. *Expert Systems with Applications*, *184*, 115519. https://doi.org/10.1016/j.eswa.2021.115519

[147] Uparkar, O., Bharti, J., Pateriya, R. K., Gupta, R. K., & Sharma, A. (2023). Vision Transformer Outperforms Deep Convolutional Neural Network-based Model in Classifying X-ray Images. *Procedia Computer Science*, *218*, 2338–2349. https://doi.org/10.1016/j.procs.2023.01.209

[148] Das, A., & Rad, P. (2020, June 22). *Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey*. ArXiv.org. https://doi.org/10.48550/arXiv.2006.11371

[149] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based

Localization. *International Journal of Computer Vision*, *128*(2), 336–359. https://doi.org/10.1007/s11263-019-01228-7

[150] Mahmud, T., Rahman, M. A., & Fattah, S. A. (2020). CovXNet: A multi-dilation convolutional neural network for automatic COVID-19 and other pneumonia detection from chest X-ray images with transferable multi-receptive feature optimization. *Computers in Biology and Medicine*, *122*, 103869. https://doi.org/10.1016/j.compbiomed.2020.103869

[151] Brunese, L., Mercaldo, F., Reginelli, A., & Santone, A. (2020). Explainable Deep Learning for Pulmonary Disease and Coronavirus COVID-19 Detection from X-rays. *Computer Methods and Programs in Biomedicine*, *196*, 105608. https://doi.org/10.1016/j.cmpb.2020.105608

[152] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). *"Why Should I Trust You?": Explaining the Predictions of Any Classifier*. https://doi.org/10.48550/arxiv.1602.04938

[153] Simonyan, K., Vedaldi, A., & Zisserman, A. (2014, April 19). *Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps*. ArXiv.org. https://doi.org/10.48550/arXiv.1312.6034

[154] Dunnmon, J. A., Yi, D., Langlotz, C. P., Ré, C., Rubin, D. L., & Lungren, M. P. (2019). Assessment of Convolutional Neural Networks for Automated Classification of Chest Radiographs. *Radiology*, *290*(2), 537–544. https://doi.org/10.1148/radiol.2018181422

[155] Chen, B., Li, J., Lu, G., & Zhang, D. (2020). Lesion Location Attention Guided Network for Multi-Label Thoracic Disease Classification in Chest X-Rays. *IEEE Journal of Biomedical and Health Informatics*, *24*(7), 2016–2027. https://doi.org/10.1109/jbhi.2019.2952597

[156] Huang, Z., & Fu, D. (2019). *Diagnose Chest Pathology in X-ray Images by Learning Multi-Attention Convolutional Neural Network*. https://doi.org/10.1109/itaic.2019.8785431

[157] Khakzar, A., Albarqouni, S., & Navab, N. (2019). Learning Interpretable Features via Adversarially Robust Optimization. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1905.03767

[158] Liu, R., Wang, L., Nan, Y., Jin, F., Wang, Q., & Pu, J. (2019). *SDFN: Segmentation-based deep fusion network for thoracic disease classification in chest X-ray images*. *75*, 66–73. https://doi.org/10.1016/j.compmedimag.2019.05.005

[159] Pesce, E., Joseph Withey, S., Ypsilantis, P.-P., Bakewell, R., Goh, V., & Montana, G. (2019). Learning to detect chest radiographs containing pulmonary lesions using visual attention networks. *Medical Image Analysis*, *53*, 26–38. https://doi.org/10.1016/j.media.2018.12.007

[160] Rajaraman, S., Thoma, G., Antani, S. ., & Candemir, S. (2019). Visualizing and explaining deep learning predictions for pneumonia detection in pediatric chest radiographs. *Medical Imaging 2019: Computer-Aided Diagnosis*. https://doi.org/10.1117/12.2512752

[161] Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Langlotz, C. P., Patel, B. N., Yeom, K. W., Shpanskaya, K., Blankenberg, F. G., Seekins, J., Amrhein, T. J., Mong, D. A., Halabi, S. S., Zucker, E. J., & Ng, A. Y. (2018). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLOS Medicine*, *15*(11), e1002686. https://doi.org/10.1371/journal.pmed.1002686

[162] Tang, Y.-X., Tang, Y.-B., Peng, Y., Yan, K., Bagheri, M., Redd, B. A., Brandon, C. J., Lu, Z., Han, M., Xiao, J., & Summers, R. M. (2020). Automated abnormality classification of chest radiographs using deep convolutional neural networks. *Npj Digital Medicine*, *3*(1). https://doi.org/10.1038/s41746-020-0273-z

[163] Morris, B. (2003). The components of the Wired Spanning Forest are recurrent. *Probability Theory and Related Fields*, *125*(2), 259–265. https://doi.org/10.1007/s00440-002-0236-0

[164] Xiao, M., Zhang, L., Shi, W., Liu, J., He, W., & Jiang, Z. (2021). A visualization method based on the Grad-CAM for medical image segmentation model. *2021*

International Conference on Electronic Information Engineering and Computer Science (EIECS). https://doi.org/10.1109/eiecs53707.2021.9587953

[165] Gulum, M. A., Trombley, C. M., & Kantardzic, M. (2021). A Review of Explainable Deep Learning Cancer Detection Models in Medical Imaging. *Applied Sciences*, *11*(10), 4573. https://doi.org/10.3390/app11104573

[166] Degerli, A., Ahishali, M., Yamac, M., Kiranyaz, S., Chowdhury, M. E. H., Hameed, K., Hamid, T., Mazhar, R., & Gabbouj, M. (2021). COVID-19 infection map generation and detection from chest X-ray images. *Health Information Science and Systems*, *9*(1). https://doi.org/10.1007/s13755-021-00146-8

[167] Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

[168] Apley, D. W., & Zhu, J. (2020). Visualizing the effects of predictor variables in black box supervised learning models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *82*(4), 1059–1086. https://doi.org/10.1111/rssb.12377

[169] Goldstein, A., Kapelner, A., Bleich, J., & Pitkin, E. (2015). Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional Expectation. *Journal of Computational and Graphical Statistics*, *24*(1), 44–65. https://doi.org/10.1080/10618600.2014.907095

[170] Lundberg, S., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. *ArXiv:1705.07874 [Cs, Stat]*. https://arxiv.org/abs/1705.07874

[171] Fan, A., Jernite, Y., Perez, E., Grangier, D., Weston, J., Auli, M., & Research, F. (n.d.). *ELI5: Long Form Question Answering*. Retrieved May 18, 2024, from https://arxiv.org/pdf/1907.09190

[172] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, *42*, 60–88. https://doi.org/10.1016/j.media.2017.07.005

[173] Shen, D., Wu, G., & Suk, H.-I. (2017). Deep Learning in Medical Image Analysis. *Annual Review of Biomedical Engineering*, *19*(1), 221–248. https://doi.org/10.1146/annurev-bioeng-071516-044442

[174] Çallı, E., Sogancioglu, E., van Ginneken, B., van Leeuwen, K. G., & Murphy, K. (2021). Deep learning for chest X-ray analysis: A survey. *Medical Image Analysis*, *72*, 102125. https://doi.org/10.1016/j.media.2021.102125

[175] Cohen, J. P. (2020, June 10). *ieee8023/covid-chestxray-dataset*. GitHub. https://github.com/ieee8023/covid-chestxray-dataset

[176] *agchung - Overview*. (n.d.). GitHub. Retrieved May 23, 2024, from https://github.com/agchung/

[177] agchung. (2021, July 30). *Actualmed COVID-19 Chest X-ray Dataset Initiative*. GitHub. https://github.com/agchung/Actualmed-COVID-chestxray-dataset

[178] *RSNA Pneumonia Detection Challenge*. (n.d.). Kaggle.com. Retrieved May 23, 2024, from https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/

[179] *COVID-19 Radiography Database*. (n.d.). Www.kaggle.com. https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database

[180] *COVID-19 Radiography Database*. (n.d.). Www.kaggle.com. https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database

[181] El-Shafai, W., & E. Abd El-Samie, F. (2020). Extensive COVID-19 X-Ray and CT Chest Images Dataset. *Data.mendeley.com*, *3*. https://doi.org/10.17632/8h65ywd2jr.3

[182] Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C. S., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F., Dong, J., Prasadha, M. K., Pei, J., Ting, M. Y. L., Zhu, J., Li, C., Hewett, S., Dong, J., Ziyar, I., & Shi, A. (2018). Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell*, *172*(5), 1122-1131.e9. https://doi.org/10.1016/j.cell.2018.02.010

[183] Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2017). ChestX-Ray8: Hospital-Scale Chest X-Ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases. *2017 IEEE Conference*

on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr.2017.369

[184] Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., Seekins, J., Mong, D. A., Halabi, S. S., Sandberg, J. K., Jones, R., Larson, D. B., Langlotz, C. P., Patel, B. N., Lungren, M. P., & Ng, A. Y. (2019). CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. *ArXiv:1901.07031 [Cs, Eess]*. https://arxiv.org/abs/1901.07031

[185] Jaeger, S., Candemir, S., Antani, S., Wáng, Y.-X. J., Lu, P.-X., & Thoma, G. (2014). Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative Imaging in Medicine and Surgery*, *4*(6), 475–477. https://doi.org/10.3978/j.issn.2223-4292.2014.11.20

[186] Sim, Y., Chung, M. J., Kotter, E., Yune, S., Kim, M., Do, S., Han, K., Kim, H., Yang, S., Lee, D.-J., & Choi, B. W. (2020). Deep Convolutional Neural Network–based Software Improves Radiologist Detection of Malignant Lung Nodules on Chest Radiographs. *Radiology*, *294*(1), 199–209. https://doi.org/10.1148/radiol.2019182465

[187] Abe, H., MacMahon, H., Engelmann, R., Li, Q., Shiraishi, J., Katsuragawa, S., Aoyama, M., Ishida, T., Ashizawa, K., Metz, C. E., & Doi, K. (2003). Computer-aided Diagnosis in Chest Radiography: Results of Large-Scale Observer Tests at the 1996–2001 RSNA Scientific Assemblies. *RadioGraphics*, *23*(1), 255–265. https://doi.org/10.1148/rg.231025129

[188] Shiraishi, J., Katsuragawa, S., Ikezoe, J., Matsumoto, T., Kobayashi, T., Komatsu, K., Matsui, M., Fujita, H., Kodera, Y., & Doi, K. (2000). Development of a Digital Image Database for Chest Radiographs With and Without a Lung Nodule. *American Journal of Roentgenology*, *174*(1), 71–74. https://doi.org/10.2214/ajr.174.1.1740071

[189] De Falco, I., De Pietro, G., & Sannino, G. (2022). Classification of Covid-19 chest X-ray images by means of an interpretable evolutionary rule-based approach. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-021-06806-w

[190] Koyyada, S. prasad, & Singh, T. P. (2023). An explainable artificial intelligence model for identifying local indicators and detecting lung disease from chest X-ray images. *Healthcare Analytics*, *4*, 100206. https://doi.org/10.1016/j.health.2023.100206

[191] Koyyada, S. prasad, & Singh, T. P. (2023). An explainable artificial intelligence model for identifying local indicators and detecting lung disease from chest X-ray images. *Healthcare Analytics*, *4*, 100206. https://doi.org/10.1016/j.health.2023.100206

[192] KARACAN, A., AKSOY, Y. E., & ÖZTÜRK, M. H. (2021). The radiological findings of COVID-19. *TURKISH JOURNAL of MEDICAL SCIENCES*, *51*(SI-1), 3328–3339. https://doi.org/10.3906/sag-2106-203

[193] Durrani, M., Haq, I. ul, Kalsoom, U., & Yousaf, A. (2020). Chest X-rays findings in COVID 19 patients at a University Teaching Hospital - A descriptive study. *Pakistan Journal of Medical Sciences*, *36*(COVID19-S4), S22–S26. https://doi.org/10.12669/pjms.36.COVID19-S4.2778

[194] Hu, X., Chu, L., Pei, J., Liu, W., & Bian, J. (2021). Model Complexity of Deep Learning: A Survey. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2103.05127

[195] Shah, B., & Bhavsar, H. (2022). Time Complexity in Deep Learning Models. *Procedia Computer Science*, *215*, 202–210. https://doi.org/10.1016/j.procs.2022.12.023

CURRICULUM VITAE WITH LIST OF PUBLICATIONS

PROFILE:

Shiva Prasad Koyyada is a Data Science Professional with over 16 years of experience; having 8+ in **AI and Machine Learning** consulting and **corporate training** and seven years of experience in Training and teaching in Computer Applications and Programming. A quick learner who understands and provides solutions to problems and easy to adapt to new learning environments. He believes in is keep it simple. Work experience and other details could be accessed @www.linkedin.com/in/shiva-prasad-6b478b33

LIST OF PUBLICATIONS

1. **Koyyada, S. prasad**, & **Singh, T. P**. (2023). An explainable artificial intelligence model for identifying local indicators and detecting lung disease from chest X-ray images. *Healthcare Analytics*, *4*, 100206. https://doi.org/10.1016/j.health.2023.100206
2. **Shiva Prasad Koyyada**, & **Singh, T. P**. (2023). Ensemble of explainable artificial intelligence predictions through discriminate regions: A model to identify COVID-19 from chest X-ray images. *Journal of Intelligent Systems*, *32*(1). https://doi.org/10.1515/jisys-2023-0163
3. **Shiva Prasad Koyyada**, & **Singh, T. P.** (2024). A Systematic Survey of Automatic Detection of Lung Diseases from Chest X-Ray Images: COVID-19, Pneumonia, and Tuberculosis. *SN Computer Science/SN Computer Science*, *5*(2). https://doi.org/10.1007/s42979-023-02573-8

4. **Shiva Prasad Koyyada, & Singh, T. P.** (2023). *An AI Decision System to Predict Lung Nodules through Localization from Chest X-ray Images*. https://doi.org/10.1109/icsc60394.2023.10441301
5. **Koyyada, S. prasad, & Singh, T. P.** (2023). A multi stage approach to handle class imbalance:An ensemble method. *Procedia Computer Science*, *218*,2666–2674. https://doi.org/10.1016/j.procs.2023.01.239
6. **Shiva Prasad Koyyada,** Rawat, A., & Singh, T. P. (2022). *Lung Infection Detection using Contemporary Techniques of Artificial Intellligence*. *2*(2),14–22. https://doi.org/10.17492/computology.v2i2.2202
7. **Shiva Prasad Koyyada**, T. P. Singh, *Lung Infection Detection through Localization: A Vision Transformer's Approach* 4th International Conference on Computational Intelligence & Internet of Things (ICCIIoT), 2023. [Accepted, Presented in December 2023]
8. **Shiva Prasad Koyyada**, **T. P. Singh, Hitesh Kumar Sharma** *An improved Lung diseases prediction system based on Ensemble of XAI methods through Localization while handling Data imbalance* International Journal of Imaging Systems and Technology. [communicated on 2nd April 2024].

UPES
UNIVERSITY OF TOMORROW

PLAGIARISM CERTIFICATE

1. We Dr. Hitesh Kumar Sharma (Internal Guide), Dr. T P Singh (~~Co-Guide~~/ External Guide) certify that the Thesis titled

LUNG DISEASES CLASSIFICATION BASED ON MICRO PATTERN IDENTIFICATION THROUGH ATTENTION GUIDED MECHANISM FROM X-RAY IMAGES

submitted by Scholar Mr/ Ms Shiva Prasad Koyyada having SAP ID 500073479 has been run through a Plagiarism Check Software and the Plagiarism Percentage is reported to be 10 %.

2. Plagiarism Report generated by the Plagiarism Software is attached.

23/9/24

Signature of the Internal Guide

Tripendra.

Signature of External Guide/Co Guide

K Shiva Prasad.

Signature of the Scholar

LUNG DISEASES CLASSIFICATION BASED ON MICRO PATTERNS IDENTIFICATION THROUGH ATTENTION GUIDED MECHANISM FROM X-RAY IMAGES

ORIGINALITY REPORT

10%	5%	8%	4%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to University of Hertfordshire Student Paper	<1%
2	dokumen.pub Internet Source	<1%
3	Submitted to Manchester Metropolitan University Student Paper	<1%
4	Anusha Leela Somashekharaiah, Abhay Deshpande. "Preprocessing techniques for recognition of ancient Kannada epigraphs", International Journal of Electrical and Computer Engineering (IJECE), 2024 Publication	<1%
5	pages.cs.wisc.edu Internet Source	<1%
6	vincmazet.github.io Internet Source	<1%