People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
University Echahid Cheikh Larbi Tebessi- Tebessa

Faculty of Exact Sciences and Natural and Life Sciences
Department of computer science
Domiciliation laboratory: laboratory of mathematics, informatics and systems (LAMIS)

Thesis

Presented and Publicly Defended for Obtaining the Diploma of Doctorate in Third Cycle

By:
Yousra HEDHOUD

Domain: Computer science **Field:** Computer science
Speciality: Artificial intelligence and its application

Title

X-ray Imaging System for diseases diagnostics

Defended on : 05/06/2026 Before the jury composed of:

Full name	Rank	University	
Mr Abdeljalil GATTAL	**Prof.**	Univ. of Cheikh Larbi Tebessi, Tebessa	President
Mr Tahar MEKHAZNIA	**Prof.**	Univ. of Cheikh Larbi Tebessi, Tebessa	Supervisor
Mr Mohamed AMROUNE	**Prof.**	Univ. of Cheikh Larbi Tebessi, Tebessa	Co- supervisor
Mme Nassira GHOUALMI ZINE	**Prof.**	Univ. of Baji Mokhtar, Annaba	Examiner
Mr Messaoud ABBAS	**Prof.**	Univ. of Martyr Hamma Lakhdar, El-Oued	Examiner

Academic year: 2024-2025

Acknowledgments

In the name of Allah, the most gracious and the most merciful. Alhamdulillah, all praises to Allah for giving me the courage and the patience to finish this thesis.

I would like to express my deepest gratitude to my supervisor, **Pr. Tahar MEKHAZNIA**, for his invaluable guidance, constant support, and insightful feedback throughout the course of this research. His expertise and encouragement have been instrumental in the successful completion of this thesis.

I would also like to extend my sincere thanks to my co-supervisor, **Pr. Mohamed AMROUNE**, whose constructive suggestions and thoughtful discussions, enriched the quality of this work.

I am equally grateful to the **Jury members**, whose careful reading, pertinent questions, and valuable recommendations contributed meaningfully to the final refinement of this work. I am honored by their time, consideration, and scientific rigor.

My sincere appreciation also goes to the **LAMIS laboratory team**, whose collaboration, technical assistance, and academic environment provided a strong foundation for this research. Working with this team has been a rewarding and intellectually stimulating experience.

I am deeply thankful to my **family** for their unconditional love, endless patience, and unwavering support. Their belief in me has been my greatest motivation through every stage of this journey.

Dedication

This dissertation is dedicated to the soul of my dear mother, which was the first supporter and motivator during my research journey.

Yousra 

Table of contents

General introduction 1
1.1 Background and problem statement 1
1.2 Objective and motivation 1
1.3 Contributions 2
1.4 Thesis organization 3

Chapter 1: Chest diseases overview

1. Introduction 4
2. Chest diseases 4
2.1 Pneumonia 5
2.2 Tuberculosis 6
2.3 Radiological mimicry between Pneumonia and Tuberculosis 8
3. Traditional methods for chest disease diagnosis 9
3.1 Clinical evaluation 9
3.2 Laboratory tests 9
3.2.1 Blood tests 9
3.2.2 Pulmonary function tests (PFTs) 9
3.2.3 Arterial blood gas (ABG) analysis 10
3.3 Limitation of traditional techniques 10
4. Medical technique imaging 11
4.1 Chest X-rays 11
4.2 Chest Computer Tomography (CT) scans 12
4.3 Chest Magnetic Resonance Imaging (MRI) 13
5. Type of Chest X-rays projections and Views 15
5.1 Posteroanterior (PA) 15
5.2 Anteroposterior (AP) 15
5.3 Lateral view 16
5.4 Lateral Decubitus view 16
6. Computer Aided diagnosis systems for Chest disease interpretation 17
6.1 Early CAD systems 18
6.2 Emergence of deep learning in medical imaging 18
7. Existing Chest disease datasets 19

8. Conclusion 21

Chapter 2: Deep learning fundamentals

1. Introduction 23
2. Deep learning in medical image analysis 23
3. Neural network fundamentals 24
3.1 Architecture of neural network 25
3.2 Training and optimization of deep learning models 26
3.2.1 Forward propagation 27
3.2.2 Loss function 29
3.2.3 Backward propagation 30
3.2.4 Optimization algorithms 30
3.3 Evaluation of deep learning models 32
3.3.1 Quantitative evaluation metrics 32
3.3.2 Graphic evaluation metrics 34
3.4 Types of Artificial Neural Networks 36
3.4.1 Convolutional Neural Networks 36
3.4.2 Vision Transformers 39
3.4.3 Vision Mamba 41
4. Transfer learning 43
5. Ensemble learning 45
5.1 Bootstrap aggregating (Bagging) 45
5.2 Boosting 46
5.3 Stacking 46
6. Deep learning models challenges 47
6.1 Data-related challenges 47
6.1.1 Data quality and quantity 47
6.1.2 Data imbalance 49
6.2 Model-related challenges 51
6.2.1 Overfitting and underfitting 52
6.2.2 Vanishing gradient problem 53
6.3 Computational challenges 53
7. Explainability and interpretability of deep learning model 53
7.1 Gradient-based methods 54
7.2 Perturbation-based methods 54
8. Conclusion 55

Chapter 3: The application of deep learning in chest disease diagnosis

1. Introduction 56
2. Methodology of the systematic review 56
   2.1 Objective 56
   2.2 Search strategy 56
   2.3 Study selection process 57
3. Chest disease classification 59
   3.1 Chest disease classification with CNNs trained from scratch 59
   3.2 Chest disease classification with Transfer learning 60
   3.3 Chest disease classification with Vision transformers 61
   3.4 Chest disease classification using hybrid models 62
4. Anomalies localization 68
5. Explainable Artificial Intelligence in Chest disease interpretation 70
6. Analysis and discussion 71
7. Conclusion 74

Chapter 4: Contributions

1. Introduction 76
2. Datasets and training objective 76
   2.1 Criteria for data selection 77
   2.2 The target objective of deep learning model's training 79
3. Experiment 1: An Improved CNN-XGboost Model for Pneumonia Classification 79
   3.1 Preprocessing and data preparation 79
   3.2 Model architecture 81
   3.3 Model training 83
   3.4 CNN-XGboost pseudo-code 84
   3.5 Results and discussion 85
4. Experiment 2: A Hybrid CNN-ViT Model for Chest Disease Classification 89
   4.1 Preprocessing and data preparation 89
   4.2 Model architecture 90
   4.3 Model training 91
   4.4 Pseudo-code of ResNet50-ViT16 model 92
   4.5 Results and discussion 93
5. Experiment 3: Application of Vision Mamba for Tuberculosis Detection 96
   5.1 Preprocessing and data preparation 97
   5.2 Model architecture 97

5.3 Model training 98
5.4 Pseudo-code of Vision Mamba for TB detection 99
5.5 Results and discussion 100
6. General discussion 102
7. Limitations and gaps 105
8. Conclusion 107
Conclusion and perspectives 107
Scientific contributions
References
Glossary

List of figures

Figure 1.1 Deaths from pneumonia, by age, world, 1980 to 2021 [5] 6
Figure 1.2 Estimated number of incident TB cases in 2023, for countries with at least 100,000 incident cases 7
Figure 1.3 Distribution of pathogens in Pneumonia cases: USA and Europe [7] 7
Figure 1.4 Sample of chest X-ray images (a) Bacterial Pneumonia, (b) Tuberculosis, (c) Covid-19 12
Figure 1.5 Example of Chest CT scan image 13
Figure 1.6 Example of chest MRI images 13
Figure 1.7 Common Chest X-ray positions [12] 17
Figure 2.1 Illustration of biological and artificial neuron 26
Figure 2.2 Basic Artificial neural network architecture 27
Figure 2.3 Representation of a confusion matrix for a binary classification task 35
Figure 2.4 K-fold cross validation process 36
Figure 2.5 Max pooling Vs average pooling techniques 38
Figure 2.6 Feature extraction Vs fine tuning concept 44
Figure 2.7 Illustration of ensemble learning techniques mechanism [16] 46
Figure 3.1 PRISMA flow diagram illustrating the selection process of studies in the systematic literature review 58
Figure 3.2 CAMs visualization of different chest diseases: (a) large pleural effusion, (b) Congestive heart failure and cardiomegaly detected, (c) Primary lung malignancy with two masses detected 71
Figure 4.1 The architecture of the proposed shallow CNN for Pneumonia detection 82
Figure 4.2 Block diagram of the hybrid CNN-XGboost model for Pneumonia detection 83
Figure 4.3 Comparison of model's performance in term of accuracy and loss for 3-class classification 87
Figure 4.4 Comparison of CNN-XGboost performance with and without data augmentation 87
Figure 4.5 Comparison of model's performance in term of accuracy and loss for 2-class classification 88
Figure 4.6 Example of Tuberculosis. (a) The original image, (b) Image after CLAHE application 90
Figure 4.7 The overall ensemble ResNet50-ViTb16 architecture 91
Figure 4.8 Performance of the ensemble ResNet50-ViTb16 for binary classification in term of accuracy and loss 94
Figure 4.9 Performance of the ensemble ResNet50-ViTb16 for binary classification in term of precision and recall 95
Figure 4.10 Performance of the ensemble ResNet50-ViTb16 for multi-class classification in term of accuracy and loss 95
Figure 4.11 Performance of the ensemble ResNet50-ViTb16 for multi-class classification in term of precision and recall 96
Figure 4.12 The architecture of Vision Mamba modem for Tuberculosis detection 97
Figure 4.13 The architecture of Vision Mamba modem for Tuberculosis detection 98
Figure 4.14 The performance of the proposed Vision Mamba for Tuberculosis detection in term of accuracy and loss 101
Figure 4.15 (a) ROC curve for Tuberculosis detection with the proposed Vision Mamba, (b) Confusion matrix of the proposed Vision Mamba model 102
Figure 4.16 Synthetic blury Pneumonia CXR images generated by DGAN……..……107

List of tables

Table 1.1 Radiological mimicry between Pneumonia and Tuberculosis 8
Table 1.2 Limitations of traditional chest disease diagnosis tools 11
Table 1.3 Detailed comparison between chest imaging techniques 14
Table 1.4 Technical and clinical characteristics of CXR views 16
Table 1.5 An overview of the public Chest X-ray datasets for deep learning applications 20
Table 2.1 Comparison between transfer learning technique 45
Table 3.1 Performance comparison of the state-of-the-art methods for Uni-chest disease detection 64
Table 3.2 Performance comparison of the state-of-the-art methods for multi-chest disease detection 66
Table 3.3 An evaluation of existing work for chest disease detection-based Transfer learning 72
Table 4.1 The applied data augmentation for CNN-XGboost training 80
Table 4.2 Comparative analysis of the proposed model with various tested model for Pneumonia classification 87
Table 4.3 Comparative Performance of some pretrained models and the Proposed Ensemble Model for Binary Tuberculosis detection 94
Table 4.4 Comparative Performance of some pretrained models and the Proposed Ensemble Model for multi-class classification 96
Table 4.5 Classification performance of the tested DL models for Tuberculosis detection 101
Table 4.6 Comparison of our proposed model with the state-of-the-art models 105

List of acronyms

A

Accuracy
acc32
acute respiratory distress syndrome
ARDS 5, 79
Adaptive Histogram equalization
AHE 49
Adaptive Moment Estimation
ADAM 32
Area Under Curve
AUC 59
Area under the receiver operating characteristic curve
AUC-ROC 34
Arterial blood gas
ABG 10
Artificial Intelligence
AI 4
Artificial neural networks
ANNs 24

B

Binary Cross Entropy
BCE 29

C

Categorical Cross Entropy
CCE 30
Chest Magnetic Resonance Imaging
MRI 13
chest x-ray
CXR 2
chronic obstructive pulmonary disease
(COPD) 4
Computer Aided Diagnosis
CAD 17
Computer Tomography
CT12
Contrast-Limited Adaptive Histogram Equalization
CLAHE 49
Convolutional neural network
CNN 2
C-reactive Protein
CRP 9
Cross Entropy
CE64

D

Deep learning
DL2
Deep Neural Networks
DNNs 53
Digital Imaging and Communications in Medicine
DICOM 20

E

Ensemble learning
EL45
Equation
eq 27
Example
e.g. 5
Explainable Artificial Intelligence
XAI 54

F

False Negative
FN35
False Positive
FP 35
Focal Loss
FL 64

G

Generative Adversarial Network
GAN 48
Gigabyte
GB 15
Gradient Descent
GE30
Gradient-weighted Class Activation Mapping
Grad-CAM 54

I

Intersection Of Union 64

L

Learning rate
Lr 64

Local Interpretable Model-agnostic Explanations
LIME 55

M

machine learning
ML 17
mean Average Precision
mAP 69
Megabyte
MB 15
millisievert
mSv 13
Montgomery County dataset
MC 77
Multilayer perceptron
MLP 41

N

Natural Language Processing
NLP 20

P

Polymerase Chain Reaction
PCR 11
Precision
prec 33
Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA 56
Procalcitonin Tests
PCT 9
Pulmonary function tests
PFTs 9

R

Radiological Society of North America
RSNA 1
Random Oversampling
ROS 50
Random Undersampling
RUS 51
Receiver Operating Characteristic curve
ROC 34
Residual Network
ResNet 39

S

Sensitivity
sens 64
Shapley Additive explanations
SHAP 55
Single Shot Detector
SSD 69
Specificity
spec 64
State space models
SSMs 41
Stochastic Gradient Descent
SGD 30
Synthetic Minority Over-sampling Technique
SMOTE 50

T

Transfer learning
TL 43
True Negative
TN 34
True Positive
TP 34

V

Vision transformer
ViT 2

W

World Health Organization
WHO 1

Y

You Only Look Once
YOLO 69

Abstract

Chest diseases, particularly Pneumonia and Tuberculosis remain among world's leading causes of morbidity and mortality, posing ongoing healthcare challenge, especially in resource-limited sources. Therefore, early detection of these diseases is crucial to save human lives. Chest X-rays (CXR) images represent the most common tool used for chest disease diagnosis due to its painless, fast acquisition, and widespread availability. However, the interpretation of these images is often hindered by weak image resolution, overlapping features, and shortage of experienced radiologists. These limitations emphasis the need of automated diagnostic tools to support clinicians' decision making.

The primary objective of this thesis is to develop deep learning-based systems for accurate detection of Pneumonia and Tuberculosis using CXR images. The research is structured around three contributions designed in a hierarchical manner. Each successive approach addresses specific limitations encountered in the preceding one. First, a hybrid CNN-XGboost model is introduced to detect Pneumonia and distinguish between viral and bacterial Pneumonia. The model showed promising results in binary classification and reduced performance in multi-class classification due to its inability to capture long-range dependencies and complex patterns.

To address this limitation, an ensemble model combining ResNet-50 and ViT-b16 (a Vision Transformer–based model) was developed—first for Tuberculosis detection, and then for multi-class classification of normal, Tuberculosis, and Pneumonia CXR images. The ensemble model leverages the strength of Convolutional Neural Network and Vision transformer, showing high performance in both binary classification and multi-class classification. Despite the strong performance of Vision Transformers in analyzing CXR images, the high memory consumption caused by their quadratic complexity, hinders the training process. Vision Mamba, a new deep learning architecture, was recently developed to deal with this issue with their ability to reduce computational overhead, while maintaining high accuracy. Based on this concept, a fine-tuned Vision Mamba model was designed for efficient Tuberculosis detection using CXR images. The obtained results demonstrate that the Vision Mamba-based model significantly reduced memory consumption, while achieving high accuracy.

Keywords: Chest diseases, X-ray images, classification, deep learning, Convolutional Neural Networks, Vision transformers, Vision Mamba

Résumé

Les maladies thoraciques, spécifiquement la Pneumonie et la Tuberculose, demeurent parmi les principales causes de morbidité et de mortalité dans le monde, représentant un défi constant, notamment dans les régions à ressources limitées. La détection précoce de ces maladies est cruciale pour sauver des vies humaines. Les images de radiographie thoracique constituent l'outil le plus utilisé pour le diagnostic des maladies thoraciques grâce à leur acquisition rapide, indolore et aussi grâce à leur large disponibilité dans la plupart des services de santé. Néanmoins, l'interprétation de ces radiographies est souvent entravée par la faible résolution, la similarité entre les symptômes radiologiques, et le manque de radiologues qualifiés. Ces limitations soulignent la nécessité de développer des outils de diagnostic automatisés pour soutenir la prise de décision clinique.

L'objectif principal de cette thèse est de développer des systèmes basés sur l'apprentissage profond pour une détection précise de la Pneumonie et de la Tuberculose à partir d'images X-ray. La recherche s'articule autour de trois contributions conçues de manière hiérarchique. Chaque approche vise à surmonter des limitations identifiées dans l'approche précédente.

Premièrement, un modèle hybride CNN-XGboost est proposé pour détecter la pneumonie et différencier ses formes virale et bactérienne. Ce modèle a démontré des performances prometteuses dans le cadre de la classification binaire, en fournissant une précision élevée. Toutefois, ses performances se sont révélées limitées dans des contextes de classification multi-classes, en raison notamment de son incapacité de gérer efficacement les dépendances à longue portée et à extraire des caractéristiques complexes présents dans les images radiographiques thoraciques. Pour surmonter cette limitation, un modèle ensembliste combinant ResNet-50 et ViT-b16 a été développé, d'abord pour la détection de la Tuberculose, puis pour la multi-classification des images X-ray en : normal, Tuberculose, Pneumonie virale et Pneumonie bactérienne. Ce modèle exploite les avantages du réseau neuronal convolutif et du Vision Transformer, et a montré de hautes performances en classification binaire comme en classification multi-classes.

Malgré les excellentes performances des Vision Transformers dans l'analyse des images X-ray, leur consommation mémoire élevée, due à leur complexité quadratique, constitue un obstacle à l'entraînement. Vision Mamba, une architecture d'apprentissage profond récemment développée, a été conçue pour surmonter ce problème en réduisant la charge computationnelle tout en maintenant une grande précision. Sur la base de ce concept, un modèle Vision Mamba ajusté a été conçu pour une détection efficace de la Tuberculose. Les résultats obtenus démontrent que le modèle basé sur Vision Mamba réduit significativement la consommation de mémoire tout en atteignant une précision élevée.

Mots clé : Maladies thoraciques, X-rays, Classification, Réseaux de neurones profonds, Réseaux de neurones convolutifs Vision Transformers, Vision Mamba,

ملخص

تُعدّ الأمراض الصدرية، وعلى وجه الخصوص مرضي الالتهاب الرئوي والسلّ، من أبرز الأسباب الرئيسية للوفيات على مستوى العالم، إذ تُشكّل تحديًا صحيًا مستمرًا، لا سيما في المناطق ذات الموارد الطبية المحدودة. ومن ثم، فإن الكشف المبكر عن هذه الأمراض يُعدّ أمرًا بالغ الأهمية لإنقاذ الأرواح البشرية. وتُعدّ صور الأشعة السينية للصدر (CXR) الأداة الأكثر شيوعًا في تشخيص أمراض الصدر، نظرًا لسهولة الحصول عليها، وخلوّها من الألم، وتوفرها الواسع. ومع ذلك، فإن تفسير هذه الصور غالبًا ما يواجه صعوبات تعود إلى ضعف دقتها، وتداخل السمات المرضية فيها، والنقص في عدد أخصائيي الأشعة ذوي الخبرة. حيث تعزز هذه القيود الحاجة المُلحّة إلى تطوير أدوات تشخيصية أوتوماتيكية تُسهم في دعم قرارات الأطباء السريرية.

يتمثل الهدف الرئيسي من هذه الأطروحة في تطوير أنظمة تشخيص تعتمد على تقنيات التعلم العميق، بهدف الكشف الدقيق عن مرضي الالتهاب الرئوي والسلّ باستخدام صور الأشعة السينية للصدر. حيث تم بناء هذا البحث اعتمادا على ثلاث مساهمات علمية مترابطة تم تنظيمها ضمن إطار هرمي، بحيث يعالج كل نموذج مقترح جوانب القصور التي ظهرت في النموذج السابق له.

في البداية، تم اقتراح نموذج هجين يجمع بين الشبكات العصبية الالتفافية (CNN) وخوارزمية XGBoost لتشخيص الالتهاب الرئوي والتمييز بين شكليه الفيروسي والبكتيري. وقد أظهر هذا النموذج نتائج واعدة في التصنيف الثنائي، إلا أن أداءه تراجع نوعا ما في التصنيف متعدد الفئات، ويرجع ذلك إلى محدوديته في تمثيل الاعتمادات بعيدة المدى و استنباط والخصائص المعقدة في الصور.

ولتجاوز هذه القيود، تم تطوير نموذج تجميعي يدمج بين شبكة ResNet-50 ونموذج ViT-b16 القائم على المحولات البصرية Vision Transformer، حيث طُبّق أولًا في الكشف عن حالات السل، ثم في تصنيف صور الأشعة السينية إلى ثلاث فئات: طبيعي، سل، والتهاب رئوي. واستفاد هذا النموذج من مزايا كل من الشبكات الالتفافية والمحولات البصرية، محققًا أداءً عاليًا في كل من التصنيف الثنائي ومتعدد الفئات.

ورغم الكفاءة العالية التي أظهرها المحولات البصرية في تحليل صور الأشعة السينية، فإن الاستهلاك العالي للذاكرة الناتج عن تعقيدها الحسابي الرباعي يشكّل عائقًا في عملية التدريب. ومن هذا المنطلق، تم تطوير بنية جديدة تُعرف بـVision Mamba ، وهي مصممة خصيصًا للتقليل من العبء الحسابي، مع الحفاظ على دقة تصنيف عالية. استنادًا إلى هذا المفهوم، تم تصميم نموذج دقيق من Vision Mambaليُستخدم بفعالية في الكشف عن حالات السل. وقد أظهرت النتائج المحصلة أن هذا النموذج يُقلّل بشكل كبير من استهلاك الذاكرة، مع الحفاظ على دقة تصنيف مرتفعة.

الكلمات المفتاحية: الأمراض الصدر، صور الأشعة السينية، التصنيف، التعلم العميق، الشبكات العصبية الالتفافية، المحولات البصرية.

General introduction

1.1 Background and problem statement

Chest diseases represent one of the major global health challenges, accounting for millions of deaths annually and causing an immense burden on healthcare systems around the world, especially in developing countries. Among these diseases, Pneumonia and Tuberculosis are particularly devastating conditions and together contribute a massive share for worldwide deaths. Chest disease diagnosis is a complex clinical process that involves multiple sources of patient information, including medical history, physical examination, laboratory tests, and medical imaging. Chest X-ray is the most common and accessible first-line imaging modality for diagnosis of chest disease due to its speed, low cost, and widespread availability. Nevertheless, accurate and timely diagnosis of these diseases remains a significant clinical challenge. This difficulty stems primarily from the remarkable overlapping radiological features between several chest diseases, causing complex interpretation landscape. This limitation is compounded by the global lack of experienced radiologists, especially in resource-limited settings, resulting in delayed diagnoses and potential misinterpretations. The World Health Organization (WHO) and Radiological Society of North America (RSNA) have acknowledged the radiology workforce gap as a serious health system bottleneck [1], [2]. Additionally, variability and subjectivity in human interpretation potentially leads to inconsistent clinical decisions.

Despite advancement of diagnostic imaging technologies, their potential is inherently limited by low spatial resolution and two-dimensional projection, which can obscure or confuse pathological findings. These limitations highlight the urgent need for intelligent, automated decision-support systems to improve diagnostic accuracy and ensure consistent healthcare delivery.

1.2 Objective and motivation

Medical imaging tools have significantly improved the understanding of human anatomy, physiology and pathology patterns. The evaluation of these images offers clinicians with an objective basis for disease diagnosis. However, the accurate interpretation of medical imaging, particularly chest x-ray images require extensive medical training and experience.

The motivation behind this experimental research is inspired from the potential of artificial intelligence (AI), and particularly deep learning to address the aforementioned limitations and

challenges in chest X-ray images interpretation. Deep learning models with their ability to automatically learn complex and hierarchical features from raw data offer significant advantages in chest x-ray interpretation, eliminating the need of capturing subtle patterns that can be invisible by human eye, and manual feature engineering. Additionally, the increasing availability of public chest x-ray datasets with the development of advanced deep learning architectures provides a great opportunity to build robust and clinically efficient deep learning-based diagnosis systems.

To achieve the objective of this thesis, the research is guided by the following questions:

1. Which deep learning architecture can be designed to effectively learn complex patterns from x-ray images?
2. How to treat problems like small datasets, data imbalance, which represent a big challenge in deep learning model training.
3. How to optimize deep learning architecture to achieve better results classification tasks.
4. How to evaluate these models to ensure that the obtained results are reliable, clinically relevant, and generalizable?
5. How can model interpretability be enhanced to ensure clinical transparency and trust?

1.3 Contributions

This thesis aims to address the real-world challenges related to chest disease diagnosis and to develop automated diagnostic support systems-based learning to provide consistent and accurate interpretation of Pneumonia and Tuberculosis using X-ray images. This is done through the design of three innovative approaches:

1. A hybrid CNN-XGboost model designed to detect Pneumonia and differentiate between viral and bacterial types, which is crucial for accurate treatment planning.
2. An ensemble model based on Convolutional neural network (ResNet-50) and Vision transformer (ViT-b16) for Tuberculosis detection, and to differentiate between Tuberculosis and Pneumonia, which share together several overlapping patterns.
3. A new fine-tuned Vision Mamba model for Tuberculosis classification. The model aims to address the challenges related to computational complexity and memory consumption, while maintaining high performance.

The proposed approaches are evaluated on multiple large-scale datasets, benchmarking the obtained results with existing state-of-the-art researches.

1.4 Thesis organization

This thesis is organized to provide a comprehensive exploration of the application of deep learning for chest disease diagnosis using x-ray images. The research encompasses both theoretical innovations and practical validation, starting from fundamental concepts through experimental implementation and evaluation.

Chapter 1 provides insights into chest diseases, their clinical relevance, diagnostic techniques, and the different imaging modalities with their characteristics, focusing on x-ray imaging. The chapter also introduces the motivation behind going through AI-assisted diagnosis.

Chapter 2 highlights the theoretical concepts of deep learning including model architectures, training and optimization, and the challenges related to each architecture. Detailing these fundamental concepts is crucial to understand the reasoning behind architecture selection, performance behavior, and the methodological choices made throughout this thesis.

Chapter 3 provides a systematic review of the existing literature on deep learning-based approaches for chest disease diagnosis using x-ray modality, identifying the strengths and gaps of current methodologies, and highlighting opportunities for improvement. This comprehensive analysis serves as a foundation for positioning the contributions of this thesis.

Chapter 4 presents the experimental contribution of this thesis, detailing the proposed approaches, the obtained results, the encountered limitations, and a comprehensive discussion of their performance in comparison to existing methods.

General conclusion concludes the thesis, summarizes the main findings, the encountered limitations and outlines future directions.

Chapter 1

Chest diseases overview

1. Introduction

Chest diseases remain one of the leading causes of morbidity and mortality worldwide, with a high burden on healthcare systems, especially in third-world countries. Despite acknowledging clinical progress and diagnostic technologies, early and accurate detection of chest diseases remains a major challenge. Particularly, Pneumonia and Tuberculosis —The main priority areas in this thesis— cause a high rate of mortality, especially in vulnerable population like elderly, children under the age of five years, and immunocompromised individuals. In this chapter, a detailed explanation of chest diseases is given with particular focus on Pneumonia and Tuberculosis. It introduces the epidemiological significance, clinical presentation, and radiological features of such diseases, followed by a discussion of the diagnostic imaging modalities, namely chest X-rays. Furthermore, it mentions the challenges of manual image analysis and the motivation behind the integration of AI-based methods into clinical workflows. Finally, it gives an overview of the major publicly available chest X-ray datasets enabling research in automated diagnosis.

2. Chest diseases

Chest diseases, also known as thoracic diseases, are a board group of anomalies affecting the lungs, pleura, bronchi, mediastinum, and other thoracic organs. These diseases vary widely in causes, progression, and prognosis. Common chest conditions include Pneumonia, Asthma, pleural effusion, lung cancer, pulmonary fibrosis, Tuberculosis and chronic obstructive pulmonary disease (COPD). Each of these diseases differs in origin, seriousness, and progression. Together, they represent a significant portion of global morbidity and are major contributors to hospital admissions and respiratory-related deaths.

Clinical symptoms of chest diseases usually overlap, with the most frequent manifestations being chronic cough, chest pain, shortness of breath (dyspnea), fever, fatigue, and in advanced cases, respiratory failures, and cyanosis (see glossary). Although they point towards chest disease, these symptoms are not specific and are usually confirmed by imaging and laboratory tests to make a definitive medical assessment.

The etiologies of chest diseases are complex. Infectious agents such as bacteria, viruses, and fungi are the underlying causes in diseases like Pneumonia and Tuberculosis. The non-infectious

etiologies include exposures to the environment (e.g., occupational exposure and air pollution), autoimmune diseases, genetic predispositions, and cancers (Lung cancer). The chronic nature of some diseases, e.g., Tuberculosis or COPD, have a tendency to result in long-term lung damage and systemic complications if left untreated.

In the wide spectrum of chest diseases, Pneumonia and Tuberculosis remain two of the most prevalent and clinically significant diseases worldwide, both from a public health and radiological perspective. These are the diseases that form the main topic of this research due to their diagnostic difficulty and the global pressing need for scalable and automated screening tools.

2.1 Pneumonia

Pneumonia is a serious inflammatory condition of the lungs, it affects the alveoli which can become filled with fluid or pus, or cellular debris, impairing gas exchange and causing symptoms like productive cough with phlegm, fever, trouble breathing and impairing oxygen transfer to the bloodstream. In severe cases, complications such as sepsis, pleural effusion, and acute respiratory distress syndrome (ARDS) may develop.

Pneumonia can develop from a range of different infection caused by different pathogens like bacteria, viruses or fungi. Bacterial pneumonia is the most common and severe, *Streptococcus* p*neumoniae* being the primary bacterial cause. Influenza, Corona viruses such as SARS-Cov2 may cause viral pneumonia. However, fungal pneumonia is less common and typical affects peoples with weakened immune systems, such as those getting chemotherapy or those who have HIV/AIDS.

The chart in *figure 1.1* shows the global number of deaths caused by pneumonia by age group from 1980 and 2021 [3], in 2023, 450 million cases of pneumonia were detected, it caused 2,5 million deaths including 672,000 children under 5 [4].

Radiology, Pneumonia appears as greater opacity on chest X-ray images, which in most cases represent consolidation or infiltrates. The patterns can differ depending on the pathogen and the host's immune response, for example, lobar consolidation frequently occurs in bacterial pneumonia, whereas viral infections often show diffuse interstitial patterns. These patterns are crucial and critical for diagnosis, but may sometimes overlap with other thoracic pathology, posing a challenge to non-expert radiologists.

2.2 Tuberculosis

Tuberculosis (TB) is a chronic, airborne infectious disease caused by *Mycobacterium tuberculosis*. It mainly affects the lungs (pulmonary TB), but is also able to spread to other organs (extrapulmonary TB) such as pleura, lymph nodes, or central nervous system. Tuberculosis is transmitted through the air when infected individuals cough, sneeze, or spit. It progresses more slowly and requires prolonged treatment, which is usually with antibiotics. People infected with TB don't feel sick and aren't contagious, symptoms may occur for many months as chest pain, prolonged cough, fever, weakness and fatigue.

TB pathogenesis remains a significant public health concern; it may remain latent for months or years before developing into active disease. Asymptomatic individuals with latent TB infection are not contagious but have the risk of reactivation, particularly if immunocompromised.

Despite substantial global efforts to control Tuberculosis, an estimated 10,8 million people fell with active TB in 2023, and around 1,25 million deaths were attributed to this disease. According to the World Health Organization (WHO), about 25% of the world's population is thought to have been exposed to TB bacteria. *Figure 1.2* highlights the estimated number of incident TB cases in 2023.

Radiologically, primary TB appears as patchy or segmental consolidation accompanied with pleural effusion. However post-primary TB, which is more common in adults and manifests as nodular opacities and cavitary lesions.

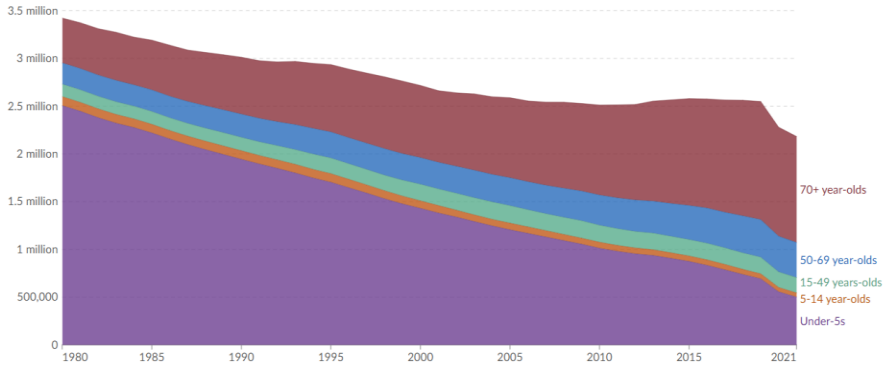


***Figure 1.1 Deaths from pneumonia, by age, world, 1980 to 2021** [5]*

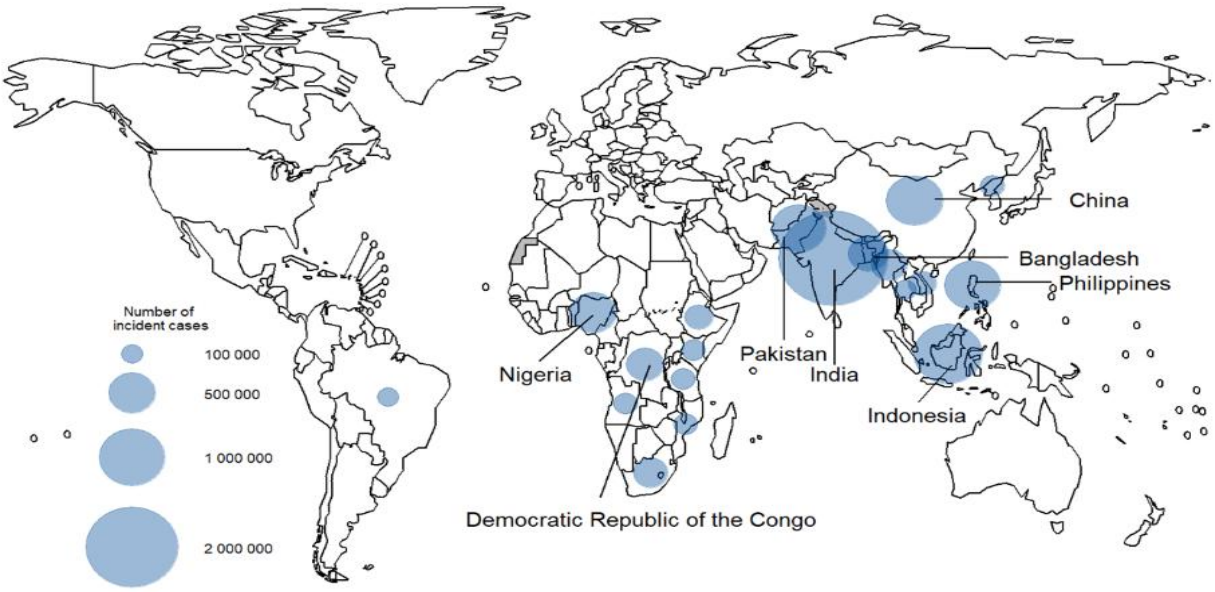


Figure 1.2 Estimated number of incident TB cases in 2023, for countries with at least 100,000 incident cases *[6]*

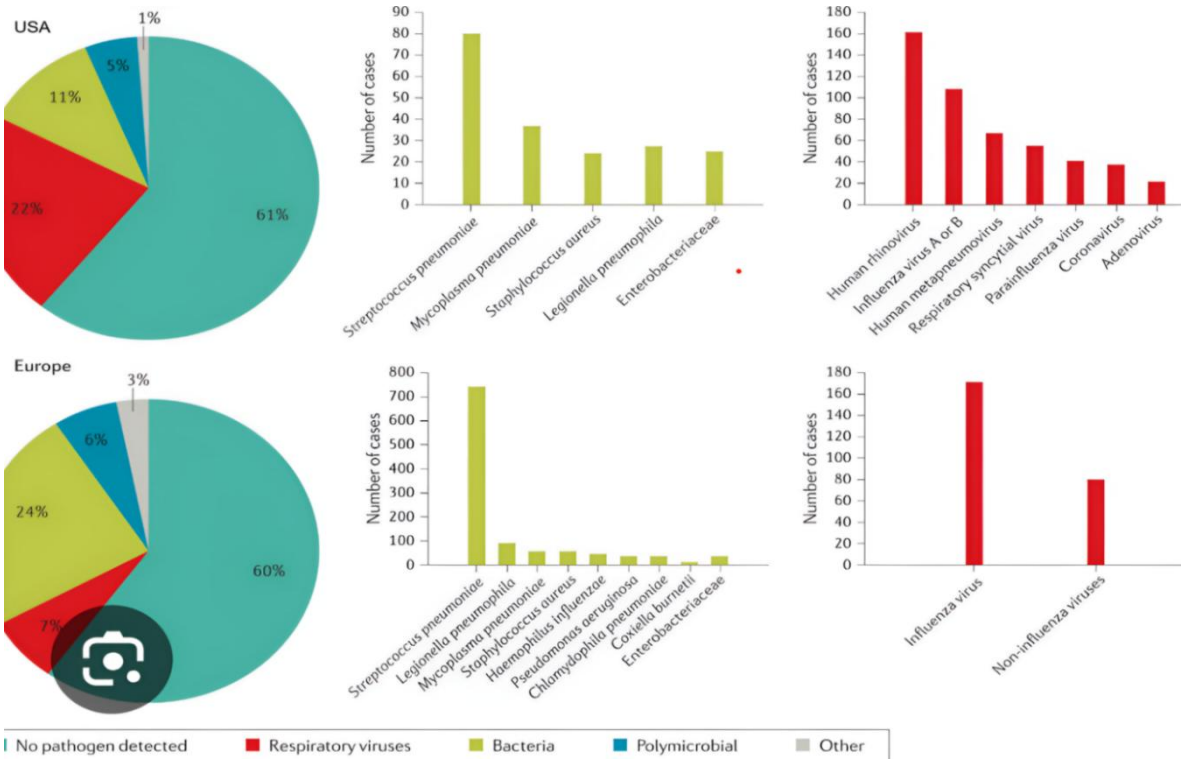


Figure 1.3 Distribution of pathogens in Pneumonia cases: USA and Europe *[7]*

2.3 Radiological mimicry between Pneumonia and Tuberculosis

Despite different underlying of Tuberculosis and Pneumonia, both diseases may produce similar radiographic features on chest x-ray images through consolidation, ground-glass opacities, inflammatory thickening of bronchial walls, or secondary pleural involvement [8]. Misdiagnosis rates between TB and pneumonia range from 20-40% in some settings, in some cases of bacterial Pneumonia may closely mimic TB, especially when cavitation is present [9]. This is commonly observed in infections caused by germs like *Klebsiella pneumoniae* or *Staphylococcus aureus*, creating lung damage that looks like TB. Such radiographic overlap leads to diagnostic confusion, delayed or inappropriate treatment, increasing morbidity and mortality for both conditions. For instance, inappropriate antibiotic use for misdiagnosed TB contributes to antimicrobial resistance. Table 1.1 summarizes radiological mimicry between Tuberculosis and Pneumonia in detail.

Table 1.1 Radiological mimicry between Pneumonia and Tuberculosis

Radiological patterns	Similarities in appearance	Common mimicry scenarios
Consolidation	Both present with airspace opacification	Primary TB can mimic lobar Pneumonia; both can show air bronchograms
Cavitation	Both can develop cavitation lesions	Necrotizing pneumonia can mimic cavitary TB
Pleural effusion	The two diseases manifest with pleural effusion	Generally, unilateral effusions appear in both conditions
Distribution	Both conditions affect lower and middle lobes	Primary TB often involves lobes similar to bacterial pneumonia
Bilateral Involvement	Both can manifest with bilateral lung involvement	Disseminated TB can appear similar to lobar pneumonia
Bronchiectasis	The two diseases can be associated with bronchiectasis changes	Chronic forms of both conditions may show similar airway involvement

3. Traditional methods for chest disease diagnosis

In order to diagnose chest diseases, different examinations are carried out such as clinical evaluations, laboratory tests, and imaging techniques. These techniques are often combined to help doctors confirm the presence of disease and identify the specific causes to assess the severity.

3.1 Clinical evaluation

Clinical evaluation is the initial phase of diagnosis, during which doctors assess the patient's medical history, the symptoms (fever, cough, chest pain, etc.) and physical examination findings based on palpation, auscultation or percussion.

3.2 Laboratory tests

The clinical evaluation is generally accompanied with laboratory tests to complement the diagnosis. These tests include:

3.2.1 Blood tests

Blood tests play an important role in pathogen detection and confirmation of diagnosis. For instance, C-reactive Protein (CRP) and Erythrocyte Sedimentation Rate (ESR) are inflammatory markers, which is useful for conditions like Pneumonia or Tuberculosis. Procalcitonin Tests (PCT) help to differentiate between viral and bacterial Pneumonia, guiding antibiotic treatments. However, an early stage of bacterial Pneumonia may not cause a rise in PCT level, CRP and white blood cell (WBC) detect inflammation but do not specify the cause (viral, bacterial, cancerous), which makes blood tests less accurate for the diagnosis process. On the other hand, some blood tests can be time-consuming, often taking days to confirm the pathology, which risks to false-negative results due to sample contamination or improper specimen collection.

3.2.2 Pulmonary function tests (PFTs)

This kind of test enables to show how well the lungs working. It assesses the lung volume, the airflow obstruction, and gas exchange in conditions like COPD or Asthma. PFTs are non-invasive, safe in general and quick to identify diseases at an early stage. However, there is a risk in some cases such as recent belly or chest surgeries, recent eye surgery, case of recent heart attack, chest pain or unstable heart conditions. In some cases, PFTs are less

accurate including weak patient effort during the examination, some patients may struggle to follow instructions, especially children and elderly individuals, inadequate breathing maneuvers can lead to false readings. Additionally, some environmental factors such as temperature, humidity and altitude may affect lung function measurements.

3.2.3 Arterial blood gas (ABG) analysis

This analysis plays a vital role in detecting some chest diseases. It aims to measure oxygen, carbon dioxide and pH levels in blood obtained from an artery to check breathing or metabolic problems. Despite this technique helping to detect several diseases such as asthma, pneumonia, COPD and even Covid-19, it is painful since it's taken from an artery, it only provides a momentary view of blood gazes at that specific time.

3.3 Limitation of traditional techniques

While traditional techniques of chest disease diagnosis are widely used, they are frequently insufficient when used alone, particularly if early or subtle disease manifestations are involved. However, timely and accurate diagnosis is crucial to avoid complications and prevent disease transmission. These limitations are related to several factors impacting the diagnosis process.

- **Overlapping symptoms:** Many chest diseases manifest with similar symptoms, making it difficult to distinguish between them.
- **Subjectivity and variability in interpretation**: The doctor's experience play a crucial role in clinical evaluation trust. Missed signs can introduce confirmation biases, which impact the accuracy judgments. Additionally, the interpretation may vary from one clinician to another, leading to inconsistent treatment plan especially in complex and ambiguous cases.
- **Lack of anatomical visualization:** Clinical evaluations and laboratory tests don't visualize what the lungs actually look like inside, making it challenging to measure the existence of anomaly or localize it.
- **Delayed or inconclusive results:** Laboratory tests especially PCR or those destinated for bacterial conditions such as Tuberculosis can yield several days to achieve results, which is not feasible for serious and urgent cases.

Table 1.2 illustrates the limitation related to each traditional diagnosis technique. These challenges underscore the essential need to incorporate direct imaging of the chest to ensure an accurate and fast assessment of thoracic conditions.

Table 1.2 Limitations of traditional chest disease diagnosis tools

Diagnosis tool	Limitations
Clinical evaluations	- Overlapping symptoms between diseases - Difficulty of some physical examinations - Challenges of accurate history-taking - Weak experience of some physicians
Laboratory tests	- Time consuming to obtain results - Some markers detect only infections and inflammation but not detect the exact cause (viral, bacterial, …) - Difficulty of tests that require sputum especially for very ill patients - PCR-based tests are unavailable in many resource-constrained settings
Pulmonary Function Testing	- Not suitable for infectious cases - Patients should follow hard instruction accurately to perform breathing maneuvers correctly. - External factors such as temperature, humidity, or altitude can affect results.

4. Medical technique imaging

Medical imaging including X-rays, CT scans and MRI provide direct visualization to the chest area, allowing physicians to spot abnormalities that might be missed through clinical assessment alone.

4.1 Chest X-rays

A chest X-ray (CXR), also known as a chest radiograph produces a 2D image (*Figure 1.4*) of the chest area (lung, heart, airways, bone of the chest and spine) using very small dose of ionizing radiation to capture the image. When the x-ray beams pass through the body's tissues, each structure absorbs different amount of this radiation. Dense structures such as bones,

calcifications and foreign objects don't allow much radiation to pass through, and therefore appear white, soft tissues and organs like heart and blood vessels absorb a portion of radiation, resulting in various shade of gray. However, air filled structures such lungs appear dark on the X-ray image.

X-rays are cost-effective, non-invasive and quick, usually takes a few minutes to complete, it is available in most healthcare facilities even small clinics. In addition, the low dose of ionizing radiation makes it relatively safe, especially for routine examinations. However, any movement, even breathing can blur the image.

(a) *(b)* *(c)*

Figure 1.4 Sample of chest X-ray images (a) Bacterial Pneumonia, (b) Tuberculosis, (c) Covid-19

4.2 Chest Computer Tomography (CT) scans

CT scans are an advanced medical imaging technique, useful for detecting small abnormalities. It takes several slices (detailed pictures -*Figure 1.5*-) of the lungs and the inside of the chest, providing high-resolution images with a detailed view. The concept of a chest CT scan is to take multiple X-ray images from different angles around the chest, then a computer combines these images and creates high-resolution slices of the chest.

Despite the chest CT scans provide detailed images enabling the detection of small lesions that might be missed on standard CXR, it uses high doses of radiation, making it not suitable for

pregnant women. Additionally, it is more expansive and, in some cases, the contrast used to enhance the clarity of the image may cause complications and allergic reaction.

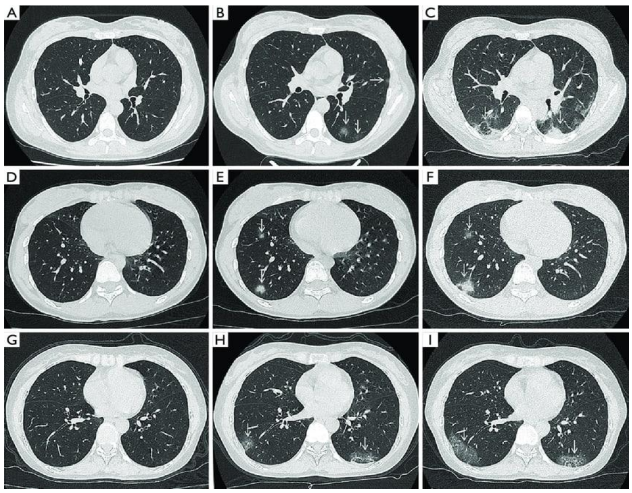


Figure 1.5 Example of Chest CT scan image

4.3 Chest Magnetic Resonance Imaging (MRI)

This technique of images (Figure 1.6) uses a powerful magnetic field, wave frequencies, and a computer to create detailed images of the components found within the chest. Generally used to confirm abnormalities noticed in CXR or CT scans. A contrast is given before the test within the vein to help radiologists see some areas more clearly.

MRI does not use radiation, making it safer for people requiring frequent imaging and pregnant women. However, the test takes 30 to 60 minutes and may take longer, which makes it disturbing and upsetting for patients. Moreover, it is not as effective in detecting lung diseases when compared to CT scans, since the presence of air in the lungs restricts MRI's capability to capture fine details.

Figure 1.6 Example of chest MRI images

Among all diagnostic techniques, CXR is the most frequently used tool for detecting chest diseases due to a combination of practical, clinical and economic factors that make it useful and valuable. It provides direct visualization of abnormalities instead of relying on indirect indicators. For instance, clinical evaluations are subjective and dependent on physician expertise, they can miss asymptomatic or early-stage diseases. Laboratory tests take time for the results and cannot localize the disease. However, the accuracy of PFTs depends on patient effort, they are less effective for infectious diseases and do not provide structural information. On the other hand, chest X-ray machines are commonly available in hospitals, clinics and certain primary care settings. This availability makes them an easily accessible diagnostic option for a broad range of patients. CXR images are less expansive compared to CT scans and MRI, which makes them the first-line diagnostic tool. Additionally, the test using CXR can be completed in a few minutes, which is critical in emergencies such as acute infection, pneumothorax, etc. Moreover, unlike CT scans and MRI, CXR test don't requires contrast agents or injections, it is non-invasive, painless and uses a low dose of radiation. Table 1.3 offer a detailed comparison between the different modalities.

Table 1.3 Detailed comparison between chest imaging techniques

Features	Chest X-ray	Chest CT scan	Chest MRI
Imaging principal	Uses small doses of ionizing radiation to create a 2D images.	Uses multiple X-ray beams from different angles to create detailed cross-sectional images	Uses strong magnetic fields and radio waves to produce detailed images without ionizing radiation
Image detail	Basic visualization of lungs bones, and heart	High-resolution images, good for detecting small abnormalities	Excellent for soft tissue visualization but not ideal for air-filled lungs
Radiation	Low (0,1 mSv)	High (5-7 mSv)	None
Acquisition time	Very fast (some seconds)	Fast (some minutes)	Slow (30-60 minutes) and even more
Cost	Low	Moderate to high	High
Portability	Highly portable (mobile X-ray units exist)	Not portable	Not portable

Availability	Widely available in all hospitals and clinics	Available in hospitals with advanced imaging facilities	Available in specialized centers of radiology
Contrast agent	Not required	Often requires contrast (iodine-based) for better clarity	Sometimes requires contrast (gadolinium-based)
Limitations	- Low resolution - Structures overlap - Small lesion may be missed. - Detect only nodules larger than 8 mm	- High radiation exposure - Motion artifacts - limited soft tissue contrast compared to MRI	- Very expansive - Long acquisition time - Contraindicated with certain metallic implants. - Not suitable for lung-specific diseases due to air interference
Data size	Small (5-15 MB per study)	Large (100-500 MB per study)	Very large (Several GB per study)

5. Type of Chest X-rays projections and Views

To perform CXR images, different views can be used to visualize the anatomical organs and pathologies in the chest area, each designed to highlight specific structure. These techniques differ depending on the projections and the digital image types. Understanding these views is crucial for radiologists and even automated systems, as they help to minimize misinterpretations caused by patient positioning or image distortion.

5.1 Posteroanterior (PA)

This is the standard frontal chest projection; the x-ray beam passes from back (posterior) to front (anterior). The patient stands upright in front of the detector, and the x-ray tube is positioned behind them, making patient's chest in contact with the detector. PA is the most used technique because it provides the most accurate representation of lungs and heart.

5.2 Anteroposterior (AP)

This is an alternative frontal projection, where the beams travers the patient from anterior (front) to posterior (back). The detector is placed behind the patient, and the x-ray tube is positioned in front of them. In this case the patient can be in supine, semi-recumbent or upright

position. This type of projection is generally used in case of ill patient who cannot stand, such as those in the intensive care unit [10]. Due to divergence of X-ray beam, the heart appears larger, which can mimic cardiomegaly.

5.3 Lateral view

This view is taken from the side (Generally left lateral), it is used to localize lesions such as tumors, effusions which are obscure in frontal views. The patient stands with left against the detector, the arms lifted above the head to clear the hand area, then the x-ray travels horizontally from right to left side.

5.4 Lateral Decubitus view

This view commonly used to identify minor amounts of fluid in the pleural cavity, which means the area located between the lungs and the chest wall [11]. In this case, the patient lies on their side to take the image.

Table 1.4 recaps the clinical characteristics of CXR views and provides a deep comparison between them.

Table 1.4 Technical and clinical characteristics of CXR views

AP view	PA view	Lateral view
X-ray beam enters through back	X-ray beam enters through front	X-ray beam enters from one side of the chest and exits through the other side
Patient's chest is close to detector	Patient's chest is close to detector	The patient is positioned sideways, with one side of the torso placed against the detector.
Heart size magnified	Normal appearance of heart size	The heart features are visible but might not be the main emphasis.
Hight radiation dose	Lower radiation dose	Similar to PA but with slightly higher dose
Low image quality due to increased scatter radiation	Image with high quality	Good for visualizing areas behind the heart and sternum
Convenient for immobile patients and emergency cases	Better for lung assessment	Useful for detecting air/fluid levels and lung consolidation

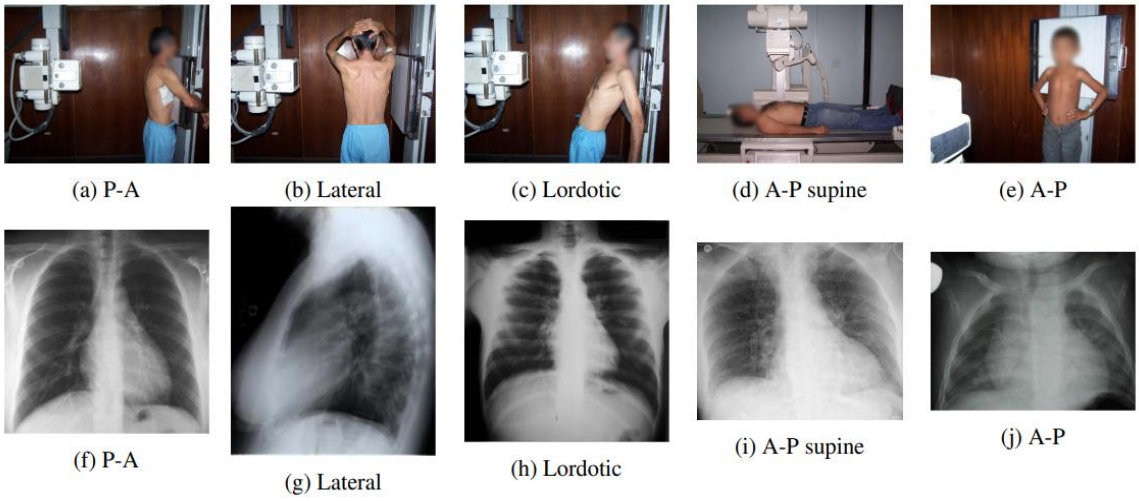


Figure 1.7 Common Chest X-ray positions *[12]*

6. Computer Aided diagnosis systems for Chest disease interpretation

Manual interpretation of CXR images remains a challenging task due to the complex anatomy of the human chest. Since CXR produces a 2D image, the anatomical structures such as lungs, heart and ribs overlap, making abnormalities difficult to detect. The interpretation of CXR images also requires expert radiologists, gaining expertise demand years of training and practical experience. Moreover, there is a great variability in observations among both between and within radiologists, resulting in inconsistent diagnoses. Examining numerous images daily creates bottlenecks, causing delays in the analysis or reducing diagnostic accuracy. These challenges highlight the need of sophisticated automatic diagnosis systems to ensure early disease detection and reduce human error.

Computer Aided Diagnosis (CAD) systems have emerged as a powerful solution to support physicians and radiologists in analyzing and interpreting medical images more efficiently. These systems leverage machine learning (ML) and deep learning (DL) algorithms to detect subtle features that may be missed by human observes. They are able to reduce the workload of radiologists and precisely accelerate the decision-making, improving the quality of patient care. The ongoing advancements of these algorithms, increase the capacity of CAD systems, offering automated anomaly detection and localization, real-time analysis and explainability features that give more information about the model's decision process, offering a human-interpretable models [13].

6.1 Early CAD systems

The development of CAD systems for medical imaging analysis has emerged with the high need of enhancing human interpretation capabilities and reducing the subjectivity and variability attendant in traditional diagnosis tools.

Early CAD systems are primarily based on handcrafted features and rule-based algorithms. These systems use predefined image characteristics explicitly engineered by domain experts to identify abnormalities by analyzing edges, textures, shapes, and intensity features. For instance, for a chest X-ray image, Sobel, Canny or other edge detectors can be used to identify the lung boundaries, which help to identify opacities. Histogram algorithms are applied to analyze the density distribution to characterize different opacity patterns. However, Gray Level Size Zone Matrix (GLSZM) extract features from grayscale images, enabling the capture of consolidation heterogeneity. Among these features expert choose the most relevant and discriminative one for the diagnostic task.

The performance of handcrafted-based systems is often limited. Designing informative features needs considerable domain expertise and deep understanding of the imaging modality and disease characteristics. This process is generally iterative and requires a substantial effort. Additionally, it cannot be generalizable for a variation of images acquisition, patient populations, or nuanced disease manifestations that were not specifically taken into account during their development. Even slight alterations in image quality or appearance can significantly impair performance.

6.2 Emergence of deep learning in medical imaging

With the advanced development of deep learning — A subfield of Artificial intelligence— CAD systems marked a revolutionary turning point for medical image analysis. Deep learning fundamentally changed the paradigm of CAD systems by passing from handcrafted features to automatic learning and hierarchical features representations directly from raw image data. This allows researchers and clinicians to avoid the tedious and frequently subjective task of creating explicit feature extractors.

Deep learning excels at complex and high dimensional data, which is a significant advantage over traditional methods that require data to be transformed into structured and numerical attributes. Additionally, deep learning tends to be more robust to variations in image

acquisition protocols, subtle differences in disease manifestation and patient demographics. The power of deep learning to learn highly complex patterns allows them to detect early abnormalities presentation that might be easily missed by the human eye, especially in the context of fatigue or large amount of data. This early detection is crucial, where early treatment can dramatically improve prognosis.

7. Existing Chest disease datasets

The advancement of automated detection of chest disease is based on the availability of large-scale and well-annotated datasets. Several chest X-ray datasets from various geographic regions, demographic groups, and clinical settings are publicly available. The datasets vary considerably in size, viewing position, annotation methodology, and pathologies type. This rich variety has become instrumental in developing and validating CAD systems for accurate chest disease detection. Among these datasets:

- **RSNA Pneumonia Detection Challenge Dataset** [14]**:** It is a large dataset that contains around 30,000 CXR images with associated labels extracted from radiological reports, indicating the presence or absence of Pneumonia. The images are provided by the Radiological Society of North America (RSNA) and made available for a challenge on Kaggle by the National Institute of Health Clinical Center of US [15].
- **Pediatric CXR dataset (Pneumonia)** [16]: The dataset contains 5,232 images of Normal and Pneumonia cases obtained from the Guangzhou Women's and Children's Medical Center. The Pneumonia cases include both viral (1,345 images) and bacterial (2,538 images) types.
- **Indiana dataset** [17]**:** This dataset consists of 7,470 images including frontal and lateral views. The images were collected from different hospitals affiliated with the Indiana University School of Medicine, and are accompanied with radiologist reports. It contains different diseases such as Pneumonia, Edema, Tuberculosis, Cardiac hypertrophy and others.
- **TBX11K dataset** [18]: The dataset contains 11,200 CXR images of healthy, sick but non-Tuberculosis, active Tuberculosis, latent Tuberculosis and uncertain Tuberculosis cases. The images are with a resolution of 512×512.
- **Montgomery dataset** [19]: It comprises 138 frontal CXR images of Tuberculosis manifestation (80 normal cases and 58 Tuberculosis manifestation), obtained from the

Tuberculosis screening program in Montgomery County, Maryland, USA. The images are portable in PNG format with a size of 4020×4890 pixels.

- **Shenzhen dataset** [20]**:** It is a collection of 662 frontal CXR images representing normal (326 images) and Tuberculosis cases (336 images), obtained from Guangdong Medical College, Shenzhen, China. The images provided in PNG format with a size of 3000×3000 pixels.
- **COVIDx CXR-4 Dataset** [21]: It is an expanded version of the COVIDx CXR-3 dataset [22]. It comprises 84,818 images obtained from 45,342 patients across different medical centers to enhance research in AI-driven Covid-19 diagnostics, by providing a large and diverse dataset.
- **ChexPert dataset** [23] **:** It consists of 224,316 images of 65,240 patients from Standford university hospital. It includes both frontal and lateral views of 14 diseases labeled by a Natural Language Processing (NLP) algorithm using radiology reports.
- **PadChest dataset** [12]**:** It is a large dataset that contains more than 160,000 high resolution images from 67,000 patients interpreted at San Juan Hospital in Spain. The images cover six position views and labeled with 174 different radiographic findings.
- **Chest X-ray 14** [24]: Comprised of 112,120 of frontal CXR images from 30,805 patients with 14 diseases. The images obtained from the National Institutes of Health (NIH), and labeled using NLP based on the associated radiological reports, with an expected accuracy of over 90%**.**
- **Vin-Dr CXR** [25]: It comprises 18,000 Posterior-anterior (PA) view CXR scans created by the Vingroup Big Data Institute and collected from Hanoi Medical University Hospital and Hospital H100. The images are publicly available in DICOM format. They were annotated by a group of 17 radiologists with at least 8 years of experience. 22 different finding were detected, and each finding in the images is localized with a bounding box.
- **PediCXR dataset** [26]: It's a newly available, freely accessible pediatric chest X-ray dataset. It comprises 9,125 images collected from a major pediatric hospital in Vietnam between 2020 and 2021, stored in DICOM format with a resolution of 1643×1349. The images are labeled by a team of three radiologists with an experience over 10 years of experience. 15 diseases and 36 critical findings are present in the images, and each abnormal finding was identified with a rectangle bounding box.

Table 1.5 An overview of the public Chest X-ray datasets for deep learning applications

Dataset name	Number of images	Image resolution	Number of patients	Number of classes	View type	Annotation method
RSNA Dataset [14]	~30,000	1024×1024	Not mentioned	2 (Presence and absence of Pneumonia)	Frontal	Expert validation of reports
Pediatric CXR dataset [16]	5,232	1024×1024	Not mentioned	2 (Normal – Pneumonia: Viral and bacterial)	Frontal	Expert annotations
Indiana dataset [17]	7,470	Range of 512×512 to 1024×1024 pixels	Not mentioned	12 diseases labels	Frontal, Lateral	Expert validation of reports
TBX11K dataset [18]	11,200	512×512	Not mentioned	5 (Healthy, uncertain, latent-TB, active TB, non-TB)	Frontal (PA, AP)	Expert annotations
Montgomery dataset [19]	138	4020×4890	138	2 (Normal, Tuberculosis)	Frontal	Expert validation of reports
Shenzhen dataset [20]	662	3000×3000	Not mentioned	2 (Normal, Tuberculosis)	Frontal (PA, AP)	Expert validation of reports
COVIDx CXR-4 Dataset [21]	84,818	Varied	45,342	Multiple (Covid-19 focused)	Frontal (PA, AP, AP supine)	Expert validation of reports
ChexPert dataset [23]	224,316	Varied	65,240	14 diseases labels	Frontal, Lateral	NLP+ Expert validation
PadChest dataset [12]	>160,000	High resolution	67,000	174 findings	7 views (PA, AP, AP horizontal, AP supine, Lateral, Lordotic, Decubitus)	Expert annotations+ NLP
Chest X-ray 14 [24]	112,120	Varied	30,805	14 diseases labels	Frontal	NLP
Vin-Dr CXR [25]	18,000	DICOM standard	Not mentioned	22 diseases labels	PA views	17 expert radiologists
PediCXR dataset [26]	9,125	1643×49	Not mentioned	15 Diseases, 36 findings	Frontal (PA)	3 expert radiologists

8. Conclusion

Chest diseases remain significant global health burden, especially in middle- and low-income nations where early and accurate diagnosis is often limited by resource availability,

specialized expertise, and infrastructure. This chapter highlighted the critical medical context necessary for understanding chest diseases, their symptoms, the mimic between them, and the different diagnosis tools with limitations related to each one.

The analysis presented in this chapter uncovers numerous significant deficits in the current systems of diagnosis. Although conventional diagnostic methods remain essential in clinical practice, they exhibit inherent limitations that complicate the diagnostic process. These complications not only suggest a technical issue, but also a significant public health concern with life-threatening implications. The fact that numerous thoracic disorders share similar patterns generates challenging diagnosis obstacle that even expert radiologists find difficult to unravel. For instance, the radiological overlap between Tuberculosis and Pneumonia continues to pose significant challenges to early and specific diagnosis. The visual pattern overlap on CXR images is to blame for excessive misdiagnosis rates, particularly in resource-poor settings where access to expert radiologists is inconsistent.

Computer-Aided Diagnosis (CAD) systems have emerged as a promising augmentation to address these gaps. While early CAD models were hindered by the rigidity of handcrafted features and lack of generalization, the recent integration of deep learning has revolutionized this domain by enabling automatic, scalable, and often more accurate feature extraction and classification directly from raw images.

Chapter 2

Deep learning fundamentals

1. Introduction

Deep learning has revolutionized the landscape of artificial intelligence and machine learning. This paradigm shift is notably worthwhile for medical image analysis task, where the convergence of computational capacities, algorithmic complexities, and clinical demand has provided outstanding potential for advancing diagnostic skills. As detailed in *chapter 1*, traditional machine learning methods often face limitations during feature engineering, this chapter delves into the core concepts related to deep learning, with a particular focus on medical image analysis. The chapter explored different architectures and algorithm, training strategies, and evaluation metrics, while highlighting the different challenges encountered during the training deep learning models.

2. Deep learning in medical image analysis

Deep learning (DL), as a subset of machine learning (ML), has emerged as a powerful tool for medical image analysis and treatment planning, offering efficient and more accurate diagnosis and disease's interpretation. Unlike traditional ML methods that rely on handcrafted features, which struggle in capturing complex patterns and variability present in medical images, DL-based approaches draw its inspiration from the hierarchical structure and information processing mechanisms of the human brain, enabling advanced features extraction, and reducing human bias.

DL is a kind of representation learning technique where an intricate multi-layer neural network model automatically derives representations from data by converting the input information into various levels of abstraction. It provides objective and standardized analysis of input data, ensuring improved features extraction and enhanced patterns recognition, which increase generalization capabilities across diverse applications such as anomalies detection [27], diseases classification [28], organ segmentation [29], etc.

In the field of radiology, medical images interpretation-based DL achieved superior accuracy in various diseases diagnosis through different modalities, including MRI [30], CT scans [31], and X-rays [32]. Studies demonstrated that DL models performance often outperformed expert radiologists [33]. Moreover, several DL architectures have been developed to address the challenges of medical images analysis. Multi-scale networks are designed to

capture both fine-grained details and larger contextual patterns, which is particularly valuable for detecting heterogeneous pathologies. 3D-architectures such as V-Net [34] exploits the volumetric nature of medical images like MRI images. Additionally, the incorporation of attention mechanism in some DL architectures helps in focusing on most relevant regions within the image, improving diagnosis accuracy.

Literature search for publication in google scholar from 2015 to 2025 using keywords: (‘medical image analysis’ AND ‘deep learning’) or (‘medical image classification’ AND ‘deep learning’) highlights the increasing application of deep learning in the medical field. This trend indicates a high shift from standard diagnosis techniques towards more data-driven and end-to-end models able to learn complex patterns directly from raw medical images.

The practical implementation of DL in medical image analysis tasks involves many important aspects that should be carefully handled and optimized, the choice of model’s architecture significantly impacts its ability to capture relevant features in medical images. The training process of the chosen model requires careful attention to the optimization algorithms used, the choice of loss function, and the use of the regularization techniques to obtain robust and generalizable performance. Further evaluation of DL models in medical imaging necessitates rigorous testing with appropriate metrics reflecting clinical utility and significance.

On the other hand, data quality and diversity used for model training is also a critical aspect which must be taken into consideration. Models trained on limited, imbalanced, or poorly labelled datasets may struggle to generalize across populations or clinical settings. Ensuring high-quality labeling and representative training samples is critical for attaining clinically valid results. Another significant factor is model interpretability. Practitioners, and clinician experts must be able to comprehend and trust system’s prediction. Techniques such as explainable AI are being applied provide insight into model decisions.

3. Neural network fundamentals

Artificial neural networks (ANNs), which form the foundational building blocks of deep learning, are designed to identify patterns and complex relationships within a given training data. ANNs take inspiration from the biological neural networks found in human brains. They consist of layers of interconnected processing units called neurons or nodes. Information passes through

the network, where each neuron interprets incoming signals and generates an output signal that affects other neurons in the system.

The learning process is called deep due to its architectural configuration, which integrates multiple layers: input and output layers, as well as several hidden layers in between.

3.1 Architecture of neural network

A basic neural network is composed of three main types of layers, each of them provides different but related tasks in the information processing pipeline.

- **Input layer:** It is the first layer of the neural network that receives the raw information Each neuron in this layer corresponds to a feature in the input data such as pixel's intensity. The input layer doesn't perform any computation; it simply passes the information to the next layer (The first hidden layer). However, it is intimately related to preprocessing operations, which prepare raw data for network consumption.
- **Hidden layers:** Hidden layers are intermediate units placed between the input layer and the output layer. They play a key role in mapping the input data through computational transformations to produce internal representations of the data. The number of hidden layers is typically determined by the size of the dataset and the complexity of the problem being addressed: Each hidden layer receives internal representations of the data via weights, biases, and non-linear activation functions that primarily identify basic features like edges or colors in the first few hidden layers, and increasingly identify more abstract and complex patterns such as shapes, textures, or components of objects in deeper layers.
- **Output layer:** It is the final layer in an ANN and is responsible for producing the final prediction. The number of neurons in the output layer depends on the nature of task. For example, in classification tasks the number of neurons is the number of classes, however in regression tasks, the output layer typically consists of one neuron that employs a linear activation function.

ANNs process the data at each layer to produce the predicted output value following this equation:

$$\hat{y} = \sigma\left(W^T x + b\right) \qquad (eq1)$$

Where $\hat{y}$ is the predicted output, x is the input data, W^T represents the transpose of the weight matrix corresponding to the inputs, b is the bias, and σ denotes the activation function, which brings non-linearity to the model.

This calculation is performed at each neuron in the network, enabling the ANN to learn and approximate complicated functions by adjusting weights and biases during the training.

Weights and biases are fundamental elements in ANNs, as they control how input data is processed and propagated across the network. Weights determine the direction of each input depending on its importance. However, biases allow for adjustments in activation thresholds throughout the network, resulting in a more flexible and accurate representation of complex, non-linear interactions between variables. Weights and biases together apply continuing adjustments during network training to get outputs and target values closest together, eventually allowing the network to detect meaningful features more accurately.

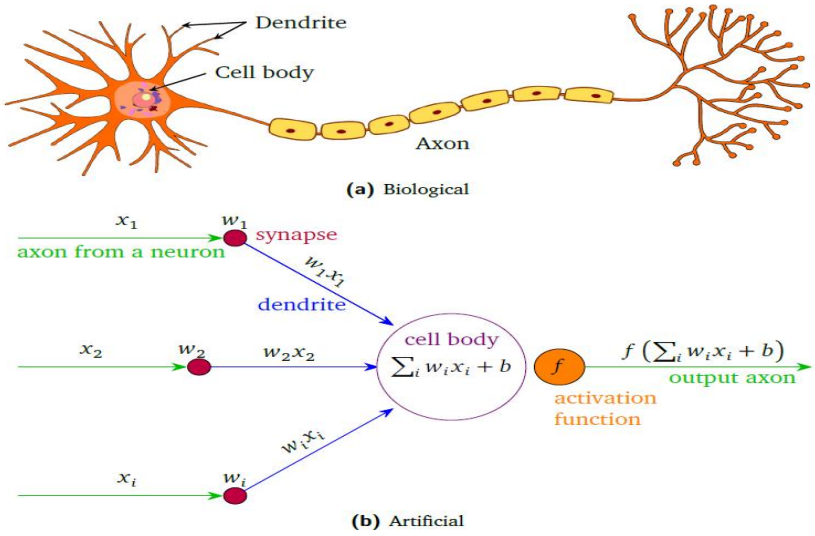


Figure 2.1 Illustration of biological and artificial neuron

3.2 Training and optimization of deep learning models

Training and optimization represent the core of how a DL model learn to perform intricate tasks. After the model is constructed, its ability to deliver prediction depends on a carful process of iteratively adjusting its internal parameters (Weights and biases). The model gradually minimizes the difference between the target and the predicted values by alternating between the forward propagation and backpropagation of parameters, enhancing its predictive ability and enabling it to generalize well on new and unseen data.

During training, the input is loaded into the ANN in a process called forward propagation. After the model generates the output, the predicted result is evaluated in relation to the given target output in a process called backpropagation.

3.2.1 Forward propagation

Forward propagation is the process by which the input data is passed through the network to generate an output. It represents the initial phase of the learning cycle. This process converts the data via a sequence of mathematical operations to construct hierarchical representations resulting in task-specific outputs.

When the input data travers the layers of the neural network, each neuron applies weighted sums, adds biases (eq1), and then passes the output through an activation function to produce the final prediction. During forward propagation, the weights and biases remain fixed; their values are only updated during the subsequent backward propagation step.

3.2.1.1 Activation function

The activation function is a crucial mathematical aspect of DL models. It calculates each neuron's output from its input and its corresponding weights by adding a non-linear transformation to the network, thereby enabling the network to learn complex patterns. The activation function determines whether a neuron is appropriate for the information being processed to be activated or not. It takes the input to the neuron and transforms the signal into an output, which is propagated to the next layer in the neural network.

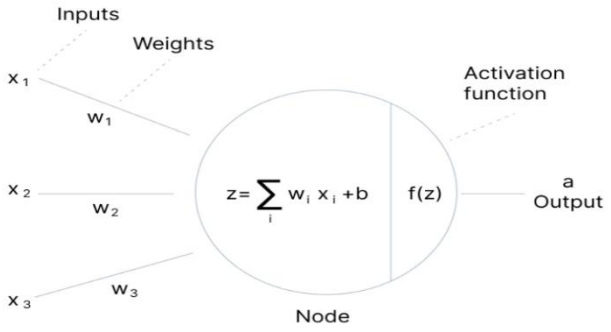


Figure 2.2 Basic Artificial neural network architecture

• **Sigmoid:** It is a logistic activation function, mostly used in binary classification tasks. The output range between 0 and 1, making it useful for probability estimation. This function is defined as:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \qquad (eq2)$$

Where x is the input value, e is the base of the natural logarithm ($e \approx 2.71828$).

Sigmoid produce a smooth gradient, enabling stable learning during backpropagation. However, for output with large magnitude either negative or positive, the gradient becomes very small causing vanishing gradient problem. This limitation making it not recommended for deep ANNs.

- **Hyperbolic tangent (*Tanh*):** This activation function is particularly used in hidden layers of the neural network. It is defined as:

$$Tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \qquad (eq3)$$

Where x is the input of the activation function, and e is defined above.

Unlike the sigmoid which outputs between (0, 1), *Tanh* function outputs between (-1, 1) which enable the gradient to update both positive and negative values, allowing more balanced updates.

Vanishing gradient problem can also be present using *Tanh* function in case of very large or very small values of x. Moreover, the exponential terms increase the computational complexity.

- **Relu:** The Rectified Linear Unit (Relu) activation function is generally used in convolutional and fully connected neural networks as default choice due to its simplicity and efficiency. It introduces non-linearity while being computationally efficient.

The relu is defined as:

$$f(x) = \max\ (0, x) \quad (eq4)$$

Where x is the input of the activation function; if $x > 0$ Relu outputs x and if $x \leq 0$ it outputs 0.

Unlike sigmoid and tanh, Relu does not saturate for positive inputs, allowing gradients to flow better during backpropagation.

- **Softmax:** This function is widely used for multi-class classification problems, it assigns a probability between 0 and 1 for each class, by transforming the vector of logits into a probability distribution. It emphasizes larger values while suppressing smaller ones, leading to a more confident classification.

  The softmax is defined as:

$$softmax\ (x_i) = \frac{e^{x_i}}{\sum_{j=1}^{n} e^{x_j}} \qquad (eq5)$$

  Where x_i is the input value (logit) for class *i,* and *n* the number of classes. The determinator sums across all classes to normalize the output, ensuring that the sum of all probabilities equals 1.

3.2.2 Loss function

Loss functions are an important aspect of neural networks, they measure how well the neural network models the training data, by comparing the target and the predicted outputs. The goal is to minimize the loss between these values by adjusting the weights, the biases and even the hyperparameters over time to improve the model's prediction.

For classification tasks, where the goal is to assign the input data to one or more categories, the loss (error) is typically calculated using cross entropy.

3.2.2.1 Loss function for binary classification

The most common loss function for binary classification tasks is *binary cross entropy (BCE)*. It quantifies the classification performance for models whose output is probability value between 0 and 1.

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i) \log\ (1 - p_i)] \quad (eq6)$$

Where N is the number of observations, y_i is the target value of i^{th} observation, and p_i is predicted value of the i^{th} observation being in class 1. A *BCE* with low value indicates a good prediction.

3.2.2.2 Loss function for multi-class classification

In the case of multi-class classification tasks, *Categorical Cross Entropy* (CCE) is the most widely used function to calculate the loss. It measures the dissimilarity between the predicted and the correct probability distribution of class labels. CCE is calculated as:

$$CCE = -\sum(y_{true} \times \log(y_{pred})) \quad (eq7)$$

Where y_{true} represents the true label, and y_{pred} is the predicted probability distribution.

3.2.3 Backward propagation

After the forward propagation generates a prediction, back ward propagation or backpropagation is used to adjust weights and biases and reduce the difference between target and predicted values. Once the error is calculated, it is propagated backward through the network.

The key idea of backpropagation is to calculate the gradient of the loss function with respect to each trainable parameter using a chain rule of computations, allowing the model to determine how much each weight contributed to the final error. This is done using an optimization algorithm.

3.2.4 Optimization algorithms

The optimization is the process of minimizing the error (loss function) when mapping inputs to outputs to enhance the DL model performance. They consist of mathematical functions that iteratively adjust the model's parameters (such as weights and biases) in order to reduce the difference between predicted outputs and actual targets. Common optimization techniques include Gradient Descent and its variants such as Stochastic Gradient Descent (SGD), RMSprop, Adam, and others. The choice of optimization algorithm significantly impacts the speed of convergence and the final performance of the model.

3.2.4.1 Gradient Descent (DC)

Is a popular optimization algorithm that aims to find the local minima of the loss function with respect to the given training data. At each point in time, it seeks to determine the optimal direction to move, taking small steps to ensure that the overall

loss value decreases. To identify the most favorable direction at a given point, it calculates the gradient (partial derivatives) of the loss function at this point. The update rule of the gradient descent is:

$$\theta = \theta - \alpha \nabla J(\theta) \quad (eq8)$$

Where θ represents the model parameters (Weights and biases), α is the learning rate and $\nabla J(\theta)$ is the derivative of the loss function.

The learning rate controls the size of the steps taken in the direction of the negative gradient when updating model parameters. Specifically, deep learning models learn by adjusting their internal parameters weights and biases in order to reduce the difference between their predicted outcomes and the actual target values. They use backpropagation to compute the gradient of the loss function with respect to each weight. Since deep learning models consist of multiple layers, adjusting the weights necessitates calculating gradients for each layer by applying the chain rule:

$$\frac{\partial J}{\partial W} = \frac{\partial J}{\partial A} \cdot \frac{\partial A}{\partial W} \qquad (eq9)$$

Where $\frac{\partial J}{\partial A}$ is the gradient of the loss with respect to the activations, and $\frac{\partial A}{\partial W}$ is the gradient of the activation with respect to the weights. Backpropagation propagates these gradients from the last layer to the first, adjusting weights accordingly.

3.2.4.2 Stochastic Gradient Descent (SGD)

It is a variant of the gradient descent that updates the model's parameters following the processing of each single training example rather than using the entire dataset as in the case of GD, which can be slow for large datasets. The update rule of the SGD is defined us:

$$\theta = \theta - \alpha \nabla \mathrm{J}(\theta ; x_i ; y_i) \quad (eq10)$$

In this case (x_i , y_i) represents a single training sample, and $\nabla \mathrm{J}(\theta ; x_i ; y_i)$ is the gradient of the loss function computed for this one sample.

This technique speeds up the training process, which make it suitable for big data and online training.

3.2.4.3 Adaptive Moment Estimation (Adam)

Adam is a commonly used optimization method in deep learning [35]. It combines the advantages of RMSprop [36] and AdaGrad [37] to individually adjust the learning rate of each parameter, resulting in faster and more efficient convergence, especially for complex models and large datasets. Adam keeps two moving averages for each parameter: The first moment (mean of the gradients) and second moment (uncentered variance of the gradients) for each parameter, computed with exponential decay rates that enable the optimizer to adapt to the geometry of the loss surface. Adam also incorporates biases correction to deal with the moment estimates being initialized at zero.

3.3 Evaluation of deep learning models

Evaluation metrics are indispensable for quantifying DL models performance. They serve as a critical bridge between theoretical and mathematical concepts and their practical application in DL models. In classification scenarios, where models need to assign distinct categories to input data, evaluation metrics offer the quantitative structure required to measure predictive performance, analyze error trends, and assess different architectures. There is a variety of evaluation metrics such as accuracy, specificity, sensitivity, etc, used to provide a more comprehensive understanding of model effectiveness.

Understanding the mathematical foundation of evaluation metrics is essential to develop robust DL models and ensure their deployment for real-world applications. The evaluation metrics intended for classification can be classified into two categories:

3.3.1 Quantitative evaluation metrics

These are numerical values that provide a concrete measure of a model's performance. Among these metrics we find:

3.3.1.1 Accuracy

The accuracy is defined as the proportion of correctly classified instances out of the total instances. This metric is suitable for balanced datasets, and useful for quickly comparing the performance of different models during the initial stages of development. The accuracy can be calculated as:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (eq\ 11)$$

3.3.1.2 Precision

This metric measures the proportion of positive identifications made by the model. It is particularly important in applications where false positives are notably costly, leading to invasive, expensive, or risky follow-up procedures.

$$precision = \frac{TP}{TP + FP} \quad (eq12)$$

3.3.1.3 Recall

The recall metric is also known as *sensitivity* or *true positive.* It measures the proportion of correctly predicted positive instances among all actual positives. A model with high recall means that is successfully identifies most of the actual positive cases, which is crucial in medical and safety-critical applications when missing a positive case (false negative) can have serious consequences. It is often to use recall in combination with the precision to provide a more balanced evaluation. The recall is defined as:

$$recall = \frac{TP}{TP + FN} \quad (eq13)$$

3.3.1.4 F1-score

The role of the F1-score metric is to balance both precision and recall. It is particularly beneficial in situations of imbalanced data, and where there are concerns regarding both false positives and false negatives. The F1-score is calculated following this equation:

$$F1 - score = 2 \times \frac{precision \times recall}{precision + recall} \quad (eq14)$$

3.3.1.5 Jaccard-score

Jaccard-score, also known as Jaccard similarity coefficient, is a commonly used evaluation metric in multilabel classification tasks. It quantifies the similarity between the predicted set and the ground truth set, by calculating the ratio of the size of intersection and the size of the union of the training sets.

Jaccard-score is useful in case of large and imbalanced datasets, it penalizes both positives and false negatives, giving it a more strict and reliable measure of prediction quality in such situations. It is defined as:

$$Jaccard - score = \frac{TP}{TP + FP + FN} \ (eq15)$$

3.3.2 Graphic evaluation metrics

Graphical evaluation metrics enable the analyze of the trade-off between different performance aspects. They offer a visual representation of the classification model's performance, allowing for better interpretation and comparison of different models.

3.3.2.1 Receiver Operating Characteristic curve (ROC)

The ROC curve is a graphical tool that represents the classification performance of the constructed model at multiple threshold levels. Its graphical representation best captures the balance between the True Positive Rate (1 - Specificity) at different decision thresholds, giving a valuable information regarding the ability of the model to distinguish between different classes.

One of the major advantages of ROC curves is that they are not heavily sensitive to class imbalance, assessing the model's ability to predict rankings instead of focusing primarily on overall classification ability. Therefore, this makes them particularly valuable when applied to medical diagnosis, where data sets tend to have an imbalanced nature.

Area under the receiver operating characteristic curve (AUC-ROC) is a broad measurement of the overall performance of a predictive model. A value of AUC that is approaching 1 implies a high ability to classify, while a value of 0.5 implies a classification performance that is no better than random chance.

3.3.2.2 Confusion matrix

The confusion matrix is a performance evaluation tool in DL used to represent the accuracy of a classification model. It provides a clear visualization to facilitate the comparison between the predicted values against the actual values for a dataset. This

visualization is represented by a table that provides the number of predicted values compared to the actual class labels, helping to control the model's performance in term of correct and incorrect classification.

For a binary classification task, the confusion matrix is a 2×2 table of the four key elements TP, TN, FP and FN (Figure 2.3). However, in multi-classification tasks, it tends to an N×N table, where N is the number of classes.

Unlike single value metrics such as accuracy, precision, etc. Confusion matrix provides a much deeper understanding of the model's behavior; it highlights where the model is making mistakes and offer a clear representation of model's performance across all classes. However, relying only on the confusion matrix and overlooking metrics such as precision and recall may lead to misleading conclusions.

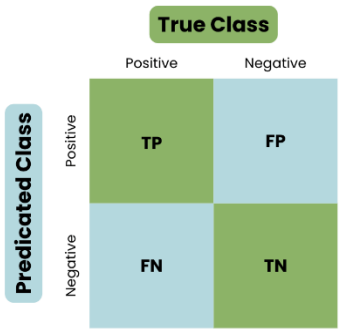


Figure 2.3 Representation of a confusion matrix for a binary classification task

3.3.2.3 Cross validation

Cross validation is a statistical technique to evaluate the model's performance more reliably. It involves dividing the datasets into k equal sized folds or subsets, then training the model on k-1 folds and test it on the remaining fold. This process is carried out K times, with a different fold used for testing during each iteration. Finally, the model's performance is the average of all the k-iterations. Using this technique helps to ensure that the model generalize well and isn't overfitting by ensuring that is evaluated on unseen data. Figure 2.4 illustrates an example of k-fold partition of a full training data.

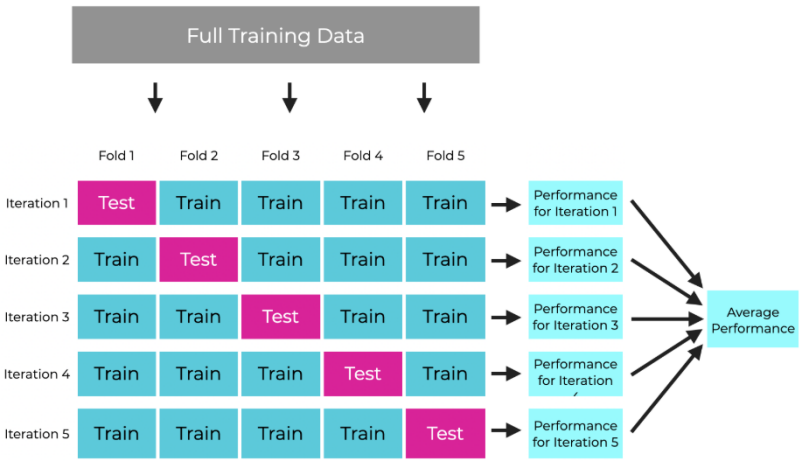


Figure 2.4 K-fold cross validation process

3.4 Types of Artificial Neural Networks

Different neural network architectures have been developed to address variable tasks through variable types of data. Each type is designed with a specific structure and functionality that makes it suitable for certain applications.

3.4.1 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) is an extended version of ANNs, firstly introduced by LeCun et al. [11] . They are being a widespread tool for image analysis in general and medical images specifically due to their capacity to automatically learn and extract features, eliminating the need for manual features engineering. CNNs excel at capturing intricate patterns and spatial relationships within the image through successive layers. They progressively build hierarchical representations of visual data, from low-level features (such as edges, textures, and gradients) in the early layers to increasingly abstract, high-level semantic concepts (such as shapes, objects, and anatomical structures) in the deeper layers. They can often discover subtle patterns that might escape human observation.

The core component of CNNs is the convolutional layers. These layers process the data with filters that moves across the image to extract local spatial patterns. They learn features from low level to high level, building a hierarchy of increasing complexity through multiple convolutional layers. The convolution layers are followed by pooling layers that reduce the spatial dimension of the feature map to down sample the information while

retaining important features making the network more efficient to achieve spatial invariance. At the end of the network, the fully connected layers are placed. These layers integrate the features learned by the convolutional layers for the classification. More precisely:

- **Convolutional layer:** A convolution layer plays a key role in CNNs, it consists of a set of learnable kernels that perform features extraction. The kernels are small 2D matrices containing weights. They slide across the input image, calculating at each position the dot product between the kernel value and the relevant pixel value. This operation is expressed as:

$$F(i,j) = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} I(i+m, j+n) \, . \, K(m,n) \qquad (eq16)$$

Where: I is the input image, K is the convolution kernel, the dot represents an element-wise multiplication, while $F\ (i,\ j)$ is the output features map at position (i,j).

Generally, the initial convolution layers are used to extract low-level features such as edges, textures and colors. The addition of more successive convolution layers increases the complexity of the extracted information, enabling the capture of higher-level features and resulting in a more thorough and comprehensive representation of the image. Two essential hyperparameters represent the convolution layers which are “the padding” and “the stride”.

The padding refers to adding specific values (usually zeros) around the input image to preserve information at the edges of the input, which could be otherwise discarded during convolution and to control the spatial size of the output of the feature map. Without padding, the output of the feature map decrease as the convolution kernel shifts over the input, particularly at the borders. Padding is useful for preserving the original size or managing the dimensions of the output.

The stride refers to how much the kernel shifts across the input image during each convolution operation. By default, the stride is (1, 1), which indicates the kernel shifts one pixel at a time. It is used to reduce the spatial size of the feature maps.

- **Pooling layer:** The pooling layers are placed between convolutional layers. The main goal of these layers is to reduce the spatial dimension of the feature maps while retaining important information. It helps to control overfitting and reduce the computational complexity by reducing the number of parameters and the amount of computation needed in the following layers.

  The concept of pooling layer is to divide the input into a set of non-overlapping regions, and replacing each region by a pooling operation such as max pooling and average pooling.

  - ✓ **Max pooling:** It is the most popular pooling technique, it takes the maximum value from each kernel of the feature map, preserving the strongest feature.
  - ✓ **Average pooling:** It calculates the average of the kernel's elements as it slides over the input feature map. For each region covered by the kernel, it computes the mean value of all the pixels and assigns this average to the corresponding position in the output feature map. This operation reduces the spatial dimensions while preserving the overall structure and smoothing the representation of features.

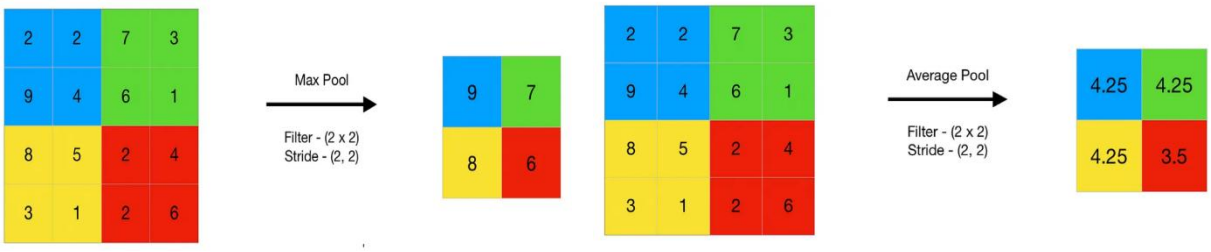


Figure 2.5 Max pooling Vs average pooling techniques

- **Fully connected layers:** Also known as dense layers, are a critical component of CNNs, they follow the convolutional and the pooling layers. Each neuron is fully connected to all the neurons in the previous layers. The role of fully connected layers is to integrate the features extracted by the previous layers and learn complex, non-linear combinations of these features to make final predictions, and then apply an activation function to classify the input image. The result is transmitted to the output layer as a vector whose size is equal to the number of target classes.

Advanced CNN architectures incorporate novel mechanisms such as residual connection (ResNet), dense connectivity (DenseNet), attention module (Squeeze-and-excitation) or depthwise separable convolutions (MobileNet). These advancements are designed to enhance models' efficiency, feature representations, and learning depth, which make them particularly valuable when analyzing medical images with multiscale pathological features, each one address specific challenge such as feature representation, model size, feature reuse, etc.

Architectures like ResNet apply residual connections in deep networks, allowing for more stable and deeper learning. DenseNet promotes feature reuse and gradient flow through dense connectivity between layers, which is particularly useful for capturing subtle variations in chest X-ray images. MobileNet employs depthwise separable convolutions to significantly reduce model complexity and computational load, making it suitable for real-time or edge-based diagnostic systems. Additionally, attention-based models such as Squeeze-and-Excitation (SE) blocks allow the network to dynamically recalibrate feature maps, improving the focus on diagnostically relevant regions. These architectural advancements are particularly valuable when analyzing medical images that contain multiscale and low-contrast pathological features, such as overlapping opacities or diffuse patterns (e.g. Pneumonia and Tuberculosis). Each mechanism addresses a specific challenge, and thereby improving diagnostic performance while maintaining scalability and interpretability.

3.4.2 Vision Transformers

Vision transformers (ViTs) have emerged as a powerful alternative to CNNs, particularly in capturing long-range dependencies and contextual relationships within an image. Originally designed for Natural Language Processing (NLP) tasks, transformers were adapted to computer vision by Disovitskiy et al. in their paper '**An image is worth 16×16 words: Transformers for image recognition at scale**' [39]. This approach fundamentally changes the image analysis process. Instead of analyzing images as a grid of pixels like traditional CNNs, the image is divided into small patches.

ViT treats images as a sequence of fixed-size patches, linearly embed each patch, and add positional encodings before processing the resulting sequence through multiple transformer encoder layers.

3.4.2.1 Vision transformers mechanism

Given an input image $x \in \mathbb{R}^{H\times W\times C}$, where H and W are the height and width, and C is the number of channels, the image is divided into non-overlapping 2D patches of size $P \times P$, resulting in:

$$N = \frac{HW}{P^2} \qquad (eq17)$$

Each patch is then flattened and projected into a fixed-dimensional embedding space:

$$z_i = W_p.p_i + b_i \qquad (eq18)$$

Where W_p represents the trainable projection matrix and b represents the bias. Subsequently, a trainable position embedding is incorporated to retain spatial information.

$$z_i' = z_i + Position\ embedding \quad (eq19)$$

These embedded patches are fed to the ViT encoder bloc. The core operation within each block is the self-attention mechanism, which allows the model to weigh the importance of different patches when processing each patch. It is defined as:

$$Attention\ (Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (eq20)$$

Where Q, K, and V represent query, key, and value matrices, respectively. These are obtained from performing linear transformations on the input embeddings, they are used in the computation of the attention weights, which specify how each of the patches attends to all the other patches.

In order to obtain wider and more diverse dependencies, multi-head self-attention (MHSA) is used. This mechanism involves running several self-attention operations in parallel, where each operation uses different sets of the parameters Q, K, and V, as well as their integrated outputs, enable the model to concentrate on different spaces of representation concurrently.

Following the attention layer, every encoder block contains a feed-forward network (FFN) which is applied to each token individually. Furthermore, the residual connections and layer normalization are used to enhance gradient flow and stability during training.

Finally, the output corresponding to the unique class token, which is embedded in the beginning of the sequence, is used in the classification process via a multilayer perceptron (MLP) head.

3.4.2.2 Limitation of Vision transformers

One of the key limitations of Vision Transformers (ViTs) is their quadratic computational and memory complexity with respect to the number of input tokens (image patches), which arises from the self-attention mechanism.

Given an input image of size $H \times W$ with N obtained patches *(eq17)*. In each Transformer encoder layer, self-attention computes pairwise interactions between all tokens. For an input sequence of N tokens, the attention operation involves computing a similarity matrix of size $N \times N$. where each element represents the attention score between two tokens. The matrix multiplication QK^T detailed in *(eq18)* requires $O(N^2.d_k)$ operations, which dominates the overall complexity.

3.4.3 Vision Mamba

Vision Mamba [40] is a recent deep learning architecture that utilize a different sequence modeling methodology based on state space models (SSMs) rather than the standard self-attention that defines Transformers. With the focus on increased efficiency and better expressiveness, Mamba is also a potential rival to Vision Transformers (ViTs) by capturing long-range dependencies with linear time and memory complexity.

3.4.3.1 Vision Mamba mechanism

In contrast to ViTs that rely on explicit pairwise interactions among tokens based on self-attention mechanism with its quadratic complexity. Vision Mamba uses a learned dynamic filtering mechanism through a selective state-space sequence model (SSM).

The selected SSM is implemented using a combination of convolution, gating, and parametrized recurrence. The Mamba block [40] can be defined as a sequence-to-sequence mapping that applies a learned state evolution equation to the input features.

Similar to ViTs, given an input image $x \in \mathbb{R}^{N \times d}$, where *N* is the number of patches, and *d* represents the feature dimension. The obtained sequence of embedded patches is processed using bidirectional filtering, with a forward and backward pass through parameterized SSMs.

An internal representation is used in SSM called hidden state, which summarizes the input past history. At each time step *t*, a latent space *h(t)* that carries the necessary context forward is maintained instead of storing the entire sequence history as in the case of ViTs. This process is represented by the subsequent continuous-time equations:

$$h'(t) = \boldsymbol{A}h(t) + \boldsymbol{B}x(t) \qquad (eq21)$$
$$y(t) = \boldsymbol{C}h(t) + \boldsymbol{D}x(t) \qquad (eq22)$$

$x(t)$ represents the input patch at time *t*, $y(t)$ is the output at time *t*; however, *A, B, C*, and *D* represents the learnable parameters that control how the input affects the hidden state and output. More precisely:

- The matrix *A* controls how the hidden state updates over time.
- The matrix *B* introduces the new input in the state.
- The matrix *C* maps the current hidden state to an output.
- The matrix *D* represents a skip connection from input to output, and it can be ignored.

Vision Mamba discretizes the continuous-time equations (*eq 21*, *eq22*) using zero-order hold (ZOH) to make the model trainable on digital hardware, resulting the following equations:

$$h(t) = \overline{\boldsymbol{A}}\, h_{t-1} + \overline{\mathbf{B}}\, X_t$$
$$y(t) = \boldsymbol{C}h(t) \qquad (eq23)$$

At this step, the model updates the hidden state and generates the output evolving from step *t-1*. This formulation enables the model to retrain itself across sequence elements, mimicking the temporal behavior of attention mechanisms without the need for pairwise comparisons. Then to ensure parallel training and fast computation, the aforementioned recurrence *(eq23)* is reformulated as a convolutional operation over the input sequence:

$$\begin{cases} y = x * \overline{K} \\ \overline{K} = (C\bar{B}, C\bar{A}\bar{B}, \dots, C\bar{A}^{N-1}\bar{B}) \end{cases} \quad (eq24)$$

Where, $\overline{K}$ represents the convolutional kernel and N denotes the sequence length. This enables the model to process sequences with linear time and memory complexity $O(N.d)$. Additionally, Vision Mamba introduces selective parameters, enabling the matrices detailed above to vary dynamically based on the input at each time step. This selectivity enhances the model's ability to adapt to varying patterns and suppress irrelevant features.

To improve contextual understanding, Vision Mamba applies bidirectional filtering to analyze the sequence in both forward and backward directions. The results from this processing technique are combined with gated integration to ensure that both prior and future context are involved to contribute to each token's representation.

Finally, the processed sequence is passed through a gated feedforward layer to enable prediction. This architecture excels at modeling global dependencies effectively and is especially beneficial for high-resolution takes like chest X-ray classification, showing remarkable balance between accuracy, speed, and scalability.

4. Transfer learning

Transfer learning (TL) is a widely used technique in deep learning that leverage knowledge acquired from a pre-trained model on a source task to improve performance on a different but related target task. This approach captures generalizable features learned from large scale datasets to accelerate training and improve model accuracy, especially in case of tasks with limited data availability. For instance, in classification tasks, ImageNet [14] serves as an exemplary foundational dataset, comprising over 14 million labeled images covering more than 20,000 classes. This dataset provides a rich foundation for pre-trained model to acquire hierarchical feature representations.

The effectiveness of TL is driven by the insight that initial layers in neural networks capture generic features (edges, texture, shapes), which are fundamental and applicable across various visual tasks, while deeper layers capture more task-specific features. Transfer learning exploits this hierarchical feature learning by employing several strategies:

- **Features extraction:** A pre-trained model is used to extract meaningful features from the input data. The parameters of the pre-trained model kept frozen, only the final layer which represents the classification layer is retrained.
- **Fine tuning:** In this case, instead of freezing all the layers, some top-level layers are retrained to learn new features related to the new task. The number of unfreeze layers is related to the complexity of the task; additional layers can be added and jointly training both the newly added layers and the last layers of the base model.
- **Domain adaptation:** This technique addresses problems of domain shift, where there is a difference in the data distribution between the source domain and the target domain. It aims to minimize this domain shift by learning invariant representations and aligning feature distributions across domains.

Table 2.1 highlights more details about TL strategies. The choice between these strategies depends on various factors such as similarity between the original task and the target task, computational resources and data availability.

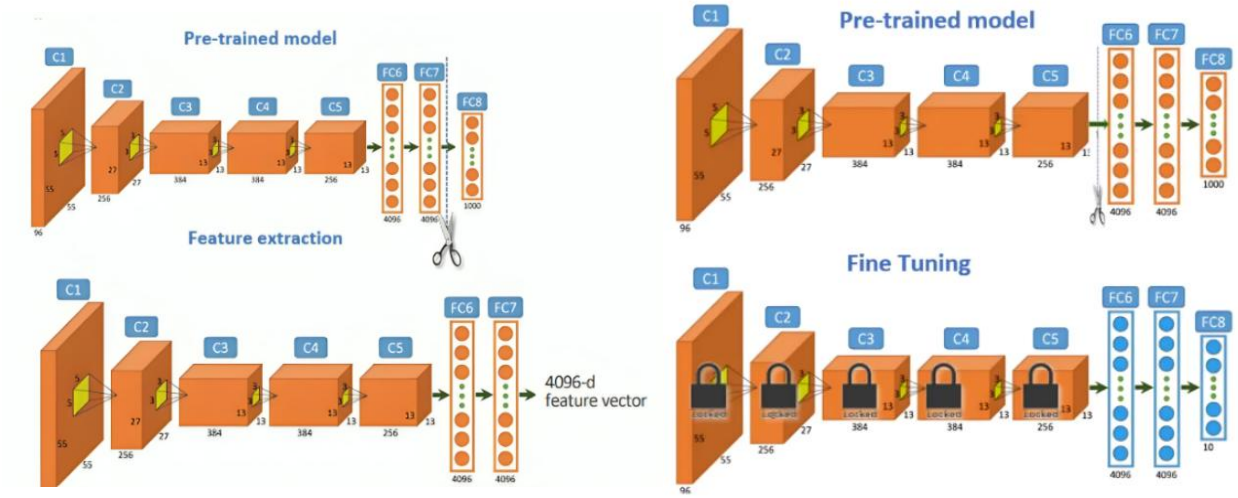


Figure 2.6 Feature extraction Vs fine tuning concept

Table 2.1 Comparison between transfer learning technique

Aspect	Feature extraction	Fine tuning	Domain adaptation
Training process	Only the final classification layers are trained; the preceding layers stay frozen.	Adapts a pre-trained model by unfreezing some or all layers and training them on the target dataset.	Adapting a model that has been trained in one domain to achieve good performance in another domain.
Computational cost	Minimal training with low cost	Depends on the number of layers being retrained	High cost; it requires additional training techniques
Use cases	Works well with small datasets	Useful in cases where there is some difference between the source and target tasks	Useful in cases of different between training and testing
Advantages	✓ Preserves original features ✓ Prevents overfitting with small datasets.	✓ Flexible according to the new task.	✓ Improves generalization to new models. ✓ Adress covariate shift
Limitations	- Less adaptable to new task.	✓ Fine tuning a model with small dataset may cause an overfitting ✓ Requires careful regularization.	- Complex to implement - May require specialized architectures.

5. Ensemble learning

Ensemble learning is a powerful paradigm in machine learning and deep learning that consists of combining multiple individual models often called ‘base models’ or ‘weak learners’ to create a more accurate and robust prediction system. The key idea of ensemble learning is to compensate for the weaknesses of individual models by leveraging their strengths, thereby reducing bias (systematic errors due to overly simplistic assumptions), variance (errors stemming from excessive sensitivity to training data noise), and overall errors [15]. This approach draws inspiration from the "wisdom of the crowd" phenomenon, where aggregated decisions from a diverse group often outperform those of individual ones, even experts. There are three main techniques of ensemble learning which are:

5.1 Bootstrap aggregating (Bagging)

Bagging involves training multiple neural networks with different architectures or random initializations and averaging their predictions. Each model is trained on a different subset of the training data, created through bootstrap sampling (random sampling with replacement). For classification tasks, predictions are typically aggregated using majority

voting. In this way, bagging lowers variance, improves generalization, and decreases the likelihood of overfitting.

5.2 Boosting

Boosting is an iterative ensemble learning technique. The neural networks are trained sequentially, with each new model addressing the errors of the prior one. Unlike bagging, which trains models separately, boosting dynamically adjusts weights to focus more on instances that are challenging to classify.

5.3 Stacking

Also known as stacked generalization, it is an advanced technique, where multiple base models are trained, generating predictions on the validation set. Then these predictions are used as input for another high-level model often known as meta-learner to learn how to optimally combine the predictions from the base models. This technique leverages the power of each model and mitigates their limitations, resulting in a highly adaptable solution for enhancing performance across different datasets.

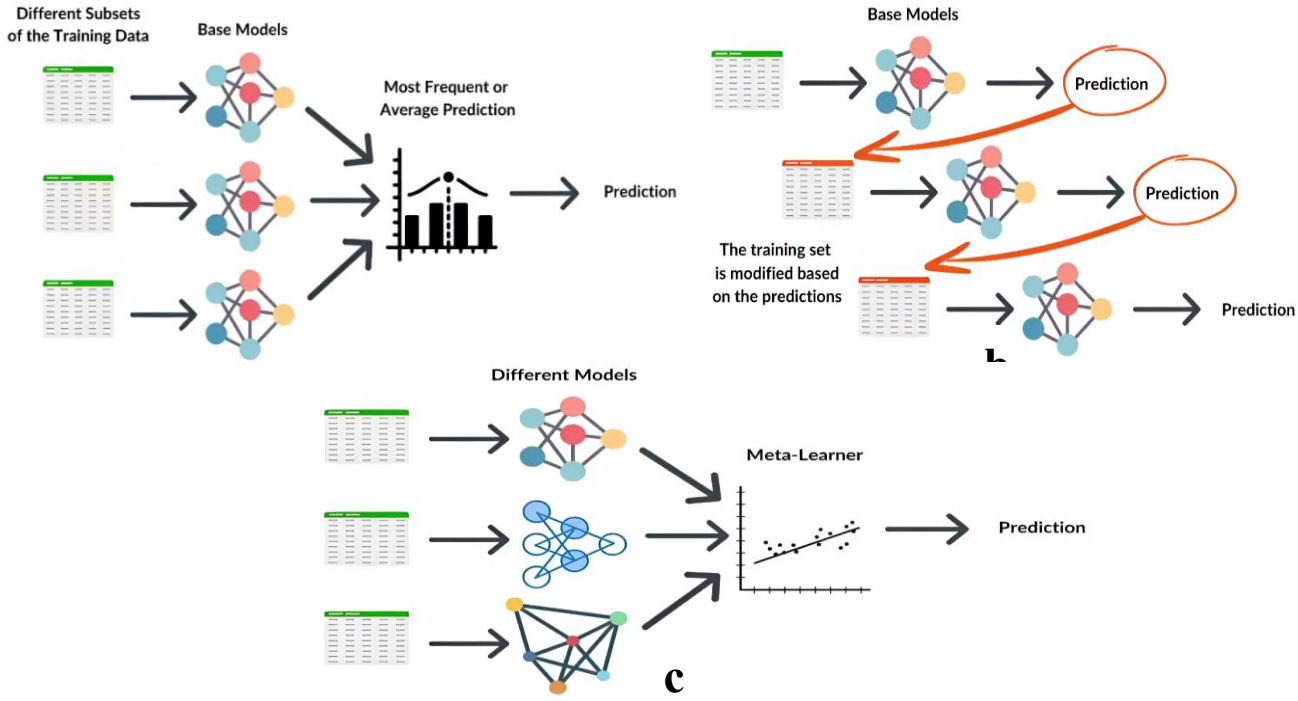


Figure 2.7 Illustration of ensemble learning techniques mechanism [16]

6. Deep learning models challenges

Despite the success of deep learning models in different fields such as computer vision, natural language processing, and medical diagnostics, behind these impressive achievements lies a complex landscape of challenges that researchers and practitioners continue to navigate. Neural network architectures still growing rapidly in depth and complexity, which introduce challenges in data quality, computational efficiency, generalization and model interpretability. These challenges can be summarized in:

6.1 Data-related challenges

The effectiveness of deep learning models largely depends on the quality, diversity, and availability of data used during training and evaluation. Several critical data-related challenges persist, which can significantly affect model performance, generalization, and clinical applicability including:

6.1.1 Data quality and quantity

Deep learning models often requires large labeled datasets for training to achieve optimal performance. However, the collection of this can be expensive and time-consuming, especially in medical data, which is subject to strict privacy regularization. Additionally, noisy data can also impact the model's learning capability and generalization, leading to confusion in understanding why the model makes certain predictions. Addressing these challenges requires advanced techniques such as data augmentation or data preprocessing.

6.1.1.1 Data augmentation

This technique aims to artificially increases the size of the dataset by creating synthetic images. Different techniques can be used to augment data, which can be summarized in:

- **Geometric transformations:** These involve applying transformations such as rotations, flipping, cropping, scaling, translation, and shearing to create variations of the original images while preserving their key features. Geometric transformations are easy to apply and help models become more invariant to positional and orientation changes. However, they cannot generate entirely new

images with different patterns, as they only modify spatial properties. This makes them unable to capture unseen variations in the data.

- **Intensity and color transformations:** Intensity and color transformations are additional methods for data augmentation; they involve changing the color distribution of images to enhance data diversity. This is especially crucial in medical imaging where tissue contrast and intensity variations can significantly impact diagnostic accuracy. Techniques include grayscale conversion, channel shuffling, color jittering or gamma correction are widely used.

  These transformations help models become robust to variations in lighting conditions, and imaging parameters that commonly occur across different imaging devices and protocols. However, careful application is necessary in medical contexts to avoid introducing artifacts that could alter clinically relevant features.

- **Generative methods:** Generative Adversarial Networks (GANs) [17] have become a powerful method for tackling the challenge of insufficient datasets by producing new artificial images. GAN models consist of two competing neural networks: A generator network which is trained to creates synthetic images closely resemble to the real ones, and a discriminator which differentiate between reel and fake images. These models have been widely applied to enhance dataset diversity and supplementing underrepresented classes.

6.1.1.2 Data preprocessing

To ensure the efficiency of DL models in classification tasks, preprocessing is an important step after choosing the adequate dataset. It helps in improving interpretability and image quality, which enhance the features extraction process by focusing on the most relevant information. Different preprocessing techniques can be applied to deal with image quality challenges including:

- **Image denoising:** It aims to remove noise from images while preserving important details and features. Common denoising techniques include Gaussian filtering, median filtering, and denoising using wavelet transforms. Additionally, more sophisticated approaches based on deep learning, such as autoencoders [18] and deep image prior [19], have also been used to enhance image clarity, especially in medical imaging applications.

- **Contrast enhancement:** This process aims to adjust the intensity distribution to improve the visibility of structures within an image. Techniques such as Histogram Equalization, Adaptive Histogram equalization (AHE), Contrast-Limited Adaptive Histogram Equalization (CLAHE) are widely used to enhance images quality. In the context of medical images, contrast enhancement helps highlights abnormalities such as nodules, infections or opacity, enabling the model to effectively differentiate between features related to the diseases.

6.1.1.3 Data normalization

Data normalization is a crucial preprocessing technique, especially for classification tasks, it consists of changing the range of pixel intensity values into a common scale, generally a range of [0, 1]. The goal of normalization is to prevent features with large ranges from dominating those with small ranges, accelerating the model's convergence. Normalization can improve the accuracy of classification models by ensuring that all features contribute equally to the learning process.

6.1.2 Data imbalance

Working with imbalanced datasets is a prevalent issue in deep learning; particularly in classification tasks. It refers to a situation where the distribution of classes in a dataset is heavily skewed. Specifically, the minority class denotes the class with the fewest data point, while the majority class is the class with the most.

Training a model with an imbalanced dataset often leads to a model that is biased to the majority class. This bias results in poor generalization and misclassification of the minority classes, which is often the most important class to predict accurately. For instance, in the context of pneumonia detection, misclassifying viral or bacterial pneumonia (which may be less frequent than other conditions) can lead to significant repercussions for how patients are treated.

The inclination to balance datasets arises from the goal of developing models that demonstrate consistent performance across all classes. Several techniques are used to deal with data imbalance challenge including resampling and class weighting. By adjusting the data distribution, the model will be able to generalize better to unseen data by reducing bias towards dominant patterns in the training data.

6.1.2.1 Resampling techniques

One way to handle an imbalanced dataset is to neutralize the effect of the imbalance by modifying the distribution of classes. The goal of resampling techniques is to balance between majority and minority classes to ensure that they contribute equally to the model learning process, reducing bias and improving generalization.

A. Oversampling

Oversampling involves increasing the number of samples in the minority class to balance the dataset. This is done by generating new synthetic images using techniques or by duplicating existing samples.

- **Random Oversampling (ROS):** It randomly duplicates existing minority class samples to balance the dataset. This method increases the risk of overfitting, as the model can memorize the repeated features instead of focusing on the most relevant patterns.
- **Synthetic Minority Over-sampling Technique (SMOTE):** It generates new illustrations by randomly selecting an instance in the minority class, then identifying its k-nearest neighbors (with k typically set to 5). From these neighbors, one is chosen at random to serve as the basis for the new synthetic data point. The vector representing the distance between the initial data point and its chosen neighbor is subsequently computed. Finally, a random value between 0 and 1 is multiplied with this distance to create a new synthetic data point that enhances the minority class by introducing more diverse examples. SMOTE reduces the risk of overfitting by creating more diverse data. But if the generated samples do not represent real patterns, noise may be added, which makes the decision boundary blurry.
- **Adaptive Synthetic Sampling (ADASYN):** It is a variation of the SMOTE technique that generates synthetic data for 'the harder to learn' samples. It introduces an adaptive mechanism to identify complex samples by calculating the density distribution of the minority class. A higher weight is assigned to these samples, and then, adjusting the number of synthetic samples generated based on the difficulty of learning each sample.

B. Undersampling

Undersampling consists of removing samples from the majority class. The common methods of undersampling include:

- **Random Undersampling (RUS):** It randomly removes samples from the majority class to equal the size of the minority class. This technique is generally effective for small datasets. However, a risk of losing valuable information is present.
- **NearMiss Undersampling:** It is a more advanced undersampling technique that reduces the number of majority class samples by selecting the closest ones to the minority class using Euclidian distance. It keeps only the most informative samples.

In general, oversampling is preferred more than undersampling. Removing samples is not optimal as they may include important information. On the other hand, undersampling is useful when the majority class has a much larger number of instances than the minority class.

6.1.2.2 Class weighting

Class weighting is a cost-sensitive learning technique used to balance the data by assigning a higher weight to the samples of the minority class and lower weights to the majority class during the training process. It encourages the model to focus more on rare events, even if they are not well-represented in the dataset.

During model training, the loss function treats all the samples equally. In the case of imbalanced data, the class weights modify this function to give more attention to the minority class. Mathematically, class weight is defined as:

$$w_c = \frac{N}{|C| \times N_c} \qquad (eq17)$$

Where w_c is the weight assigned to a class *C, N* is the total number of samples, N_c the number of samples in class *C*, and $|C|$ is the total number of classes.

Compared to oversampling and undersampling techniques, class weighting is computationally effective with no data loss. However, if the affected weight is too high, it may cause an overfitting to the minority class.

6.2 Model-related challenges

While deep learning has achieved remarkable success, particularly in medical image analysis, the training and optimization of these models remains technically demanding. Several inherent

challenges can affect the learning process and model performance, especially when dealing with complex or imbalanced medical datasets. Issues such as overfitting, underfitting, vanishing gradients, and optimization instability must be carefully addressed to ensure reliable and generalizable diagnostic outcomes.

6.2.1 Overfitting and underfitting

The overfitting is a common issue in deep learning; it occurs when a model starts to memorize and performs exceptionally with the training data but fails to accurately predict outcomes on unseen data. In this case the model entirely fit the patterns of training data including noise, irrelevant details and random fluctuations, resulting a high accuracy on the training set (e.g. 99%) and fails catastrophically with validation or test sets. Several factors allow the model to overfit. One major reason is the scarcity of training data, when the model is trained on a small or imbalanced dataset, focusing on specific details that do not accurately reflect real-world scenarios. Similarly, if the model is too complex with multiple layers and high number of parameters. This complexity makes the model overly sensitive to the specificities of the training data rather than general patterns. Another major factor of overfitting is the prolonged training, allowing the model to grind the training data into its weights, including noise and artifacts.

An overfitting is recognized when there is a large gap between training and validation accuracies, or when the validation loss starts rising, while the training loss keeps decreasing.

Underfitting is another pervasive challenge in deep learning. It occurs when a model is too simplistic to capture valuable patterns within the data, leading to inaccurate predictions on both training and unseen data. The model underfits in case of insufficient training data, either due to an overly simple architectures (e.g. shallow CNN), making it incapable to represent the complexities in the data. The model can also underfits if the input features used for training do not sufficiently represent the underlying factors influencing the target variable. If essential patterns are missing or poorly encoded, the model will fail to learn useful relationships.

Addressing overfitting and underfitting problems requires the application of optimization and techniques including, data augmentation, data balancing, cross validation, etc, in combination with regularization techniques such as dropout, early stopping, and (L1, L2) regularization.

6.2.2 Vanishing gradient problem

Vanishing gradient is a major roadblock in training DL models -Particularly models with many layers-, it occurs when gradients, which are the values used to update the weights of the network during backpropagation become extremely small and vanish. This problem leads to a very slow training and poor performance; it hinders the ability to learn complex features and capture long-term dependencies.

Sigmoid and *tanh* activation functions are a common cause of vanishing gradient due to their small derivatives for large input values, causing a decreasing in the gradients as they are propagated back. To mitigate this issue, it is preferred to use *relu* activation function to introduce non-linearity and avoid saturating derivatives. Additionally, adding batch normalization layers can stabilize gradients by scaling the activations to have unit variances and making them less sensitive to small variations of the model's parameters. Moreover, using architectures with skip connections and residual blocks by creating shortcut pathways, allowing the gradient to pass directly through the network during backpropagation.

6.3 Computational challenges

Training DL models requires enormous computational power, given their complexity and large amounts of training data. Some DNNs require powerful hardware, including GPUs and TPUs, and important energy resources. These requirements are posing increasing challenges, particularly regarding sustainability and cost. For instance, DL models in healthcare contexts need to be inferenced with quick and accurate results, which is likely to be a challenge in real-world hospital environments where available resources are limited.

Leveraging cloud computing platforms like Azure, google cloud or AWS can deal with issue, offering robust and scalable computational resources without requiring a large initial investment.

7. Explainability and interpretability of deep learning model

DL models generate outputs that are somewhat of a "black box," without providing much information about the reasoning or processes underlying these predictions. As the complexity of DL architectures increases, understanding their decision processes becomes more difficult to discern; however, this understanding is critical in consequential applications such as healthcare. This complexity poses major challenges for clinical acceptance and ethical use. In medical

imaging, where decisions made can strongly impact patient therapy, it is crucial that clinicians are equipped with a means of understanding and verifying model predictions before their implementation in clinical practice.

Explainable Artificial Intelligence (XAI) have emerged to deal with this challenge by providing different methodologies and strategies to improve the understandability, transparency, and interpretability of AI models for human users. XAI methods can broadly be categorized into:

7.1 Gradient-based methods

Gradient-based methods aim to interpret the decision-making process of DL models by analyzing how small changes in the input affect the model's output. These methods include Grad-CAM (Gradient-weighted Class Activation Mapping) [47], Score-CAM [48], and SmoothGrad [49], which highlight salient regions in the input image that are important for the model's decision by analyzing gradients flowing back to the input layer or intermediate feature maps. These are particularly useful for CNNs.

7.2 Perturbation-based methods

Perturbation-based methods are a class of XAI methods that interpret a model's decision-making process through systematically perturbing or masking portions of the input and examining the resulting differences in the model's output. A major purpose of these methods is to determine the areas or features of the input that most strongly impact a particular prediction.

In the context of medical image analysis, and specifically in chest X-ray classification, perturbation-based methods involve modifying or occluding small parts of the image (e.g., by blurring, masking, or adding noise) in an attempt to estimate how much models' predictions are impacted. A significant decrease in prediction confidence due to the occlusion of a specific area suggests that this area is an essential part in model decision-making processes.

LIME (Local Interpretable Model-agnostic Explanations) and SHAP (Shapley Additive explanations) are the commonly Perturbation-based methods used in the literature. They analyze how small changes to the input affect the output, providing local fidelity to the model's behavior. These are model-agnostic and can be applied to any deep learning architecture.

8. Conclusion

This chapter provided an in-depth exploration of the theoretical and technical foundations of deep learning (DL), with particular emphasis on its relevance to medical image analysis. It detailed the fundamental architectures such as CNNs, Vision transformers, and training strategies, including optimization and regularization techniques. The chapter also addressed the practical challenges associated with model development, in addition to explainability and interpretability that represent a key factor that significantly influence real-world deployment, especially in the medical domain. Establishing a strong understanding of DL theory is crucial to provide the conceptual framework needed to select the optimal architecture for automated chest disease diagnosis, to enhance the model's training, to ensure accurate performance, to critically evaluate the provided results, and to value their efficiency for real-world application. The next chapter will detail the literature related to the application of deep learning algorithms for chest disease diagnosis, and systematically review the state-of-the-art studies.

Chapter 3

The application of deep learning in chest diseases diagnosis

1. Introduction

Chest disease diagnosis using deep learning algorithms has gained significant attention in recent years due to the increasing availability of medical imaging and the advances of these algorithms' abilities to improve diagnostic accuracy, reduce radiologists' workload, and support clinical decision-making. This chapter introduces an in-depth systematic review of the literature related to chest disease diagnosis-based DL including classification, anomalies localization, and the integration of explainable AI (XAI) techniques. By comprehensively examining the current state-of-the art researches, this chapter provides a solid foundation for understanding existing methodologies, identifying research gaps, and exploring innovative approaches to improve patient care through deep learning.

2. Methodology of the systematic review

In this thesis, we conducted a systematic review using PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) methodology to highlight the application of DL algorithms for chest disease diagnosis using CXR images.

2.1 Objective

The main goal of this systematic review is to capture and analyze the most recent and relevant research on the application of DL for chest disease diagnosis. The review focuses on disease detection and classification, anomaly localization, and the integration of explainable artificial intelligence (XAI) in DL-based diagnostic systems. It also highlights the different challenges and uncovers the gaps in literature that can be used to direct future studies and clinical practice.

2.2 Search strategy

The search strategy was carefully designed and iteratively refined to guarantee that the review included the most recent advancements in the field. Multiple electronic databases are used to cover a broad spectrum of published works including google scholar, IEEE Xplore, Science direct, PubMed and ArXiv. Keywords including “chest disease diagnosis”, “chest disease classification”, “deep learning”, “chest x-ray” combined with Boolean operators (AND, OR) are used to find the adequate research papers. As example of this queries:

- (“Chest disease diagnosis” OR “Chest disease classification” AND “Deep learning”).
- (“Chest disease classification” AND “X-ray” AND “CNN” or “Vision transformers”).
- (“Chest disease classification” AND “X-ray” AND “Explainable Artificial Intelligence”).
- (“Chest disease” AND “localization”).

In PubMed, supplementary Medical Subject Headings (MeSH) such as “Chest Disease” and “Radiography” were integrated to encompass indexed items that may not have been identified by keyword searches alone. In IEEE Xplore and Scopus, the queries were refined to focus on both titles and abstract fields to increase the relevance of the retrieved results. However, in ArXiv, more broad terminology was used to account for the preprint nature of the repository, ensuring a corporation of emerging trends.

2.3 Study selection process

Searches were restricted to recent English-language articles published from 2020 to 2025. Then a series of inclusion and exclusion criteria were conducted to ensure the selection of high-quality and relevant studies that directly address the application of DL algorithms to chest disease diagnosis using CXR images.

- **Inclusion criteria**

The selection of papers is based on some criteria including:

- ✓ English language publication only.
- ✓ Peer-reviewed journal articles, conference proceedings, or high-quality preprints.
- ✓ Articles providing detailed methodology and clear explanation of the developed models.
- ✓ Papers focused on the classification of Pneumonia, Tuberculosis, and Covid-19.
- ✓ Only studies that used X-ray modality are selected.
- ✓ Only methods that underwent detailed evaluation with quantitative performance metrics were included.

- **Exclusion criteria**

  ✓ Papers representing a weak approach and overlapping datasets.

  ✓ Scientific reports are not included in this review.

  ✓ Studies not focused on DL-based models (e.g. Traditional ML methods only).

All records are integrated into Mendeley reference management tool, where duplicate papers are automatically removed. Then, based on the abstract of papers, those with clearly irrelevant with the research scope are also excluded.

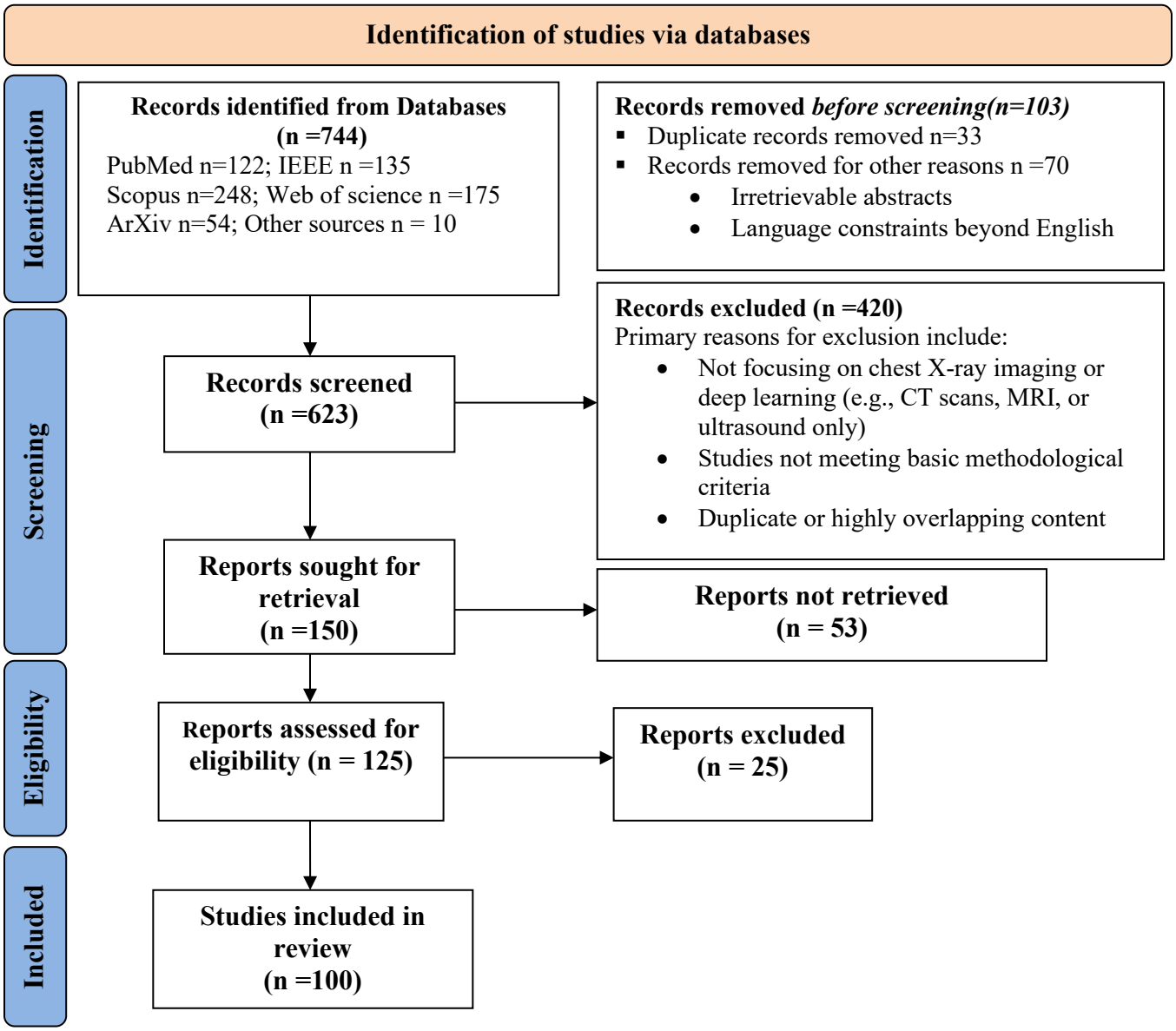


Figure 3.1 PRISMA flow diagram illustrating the selection process of studies in the systematic literature review

3. Chest disease classification

Chest disease classification is a fundamental task that aims to distinguish between healthy and pathological cases based on CXR images. The application of DL techniques to chest disease classification represents one of the most active areas of research in medical imaging analysis. This subfield has evolved rapidly encompassing a wide range of techniques, with differences in neural network architectures, training methods, data types, and clinical goals.

3.1 Chest disease classification with CNNs trained from scratch

Shallow CNNs trained from scratch have been widely explored for chest disease classification tasks using CXR images. These models with their few layers and parameters provide fast training and good performance especially in cases with limited datasets. Shallow CNNs excel in capturing fundamental features, which are crucial in localized abnormalities like pleura, airways or any changes in the lung's region.

M. Nahiduzzaman et al. [50] designed a lightweights CNN model, named ChestX-ray6, consisting of six convolutional layers and two fully connected layers for the detection of six chest diseases. The model was trained using combined datasets and achieved an accuracy of 80%. It was tested for binary classification of Pneumonia disease; were it reached an accuracy and recall of 97.94%. Another lightweights CNN with three convolutional layers and three fully connected layers was proposed by T. Sanida et al. [51] for seven-category identification of chest diseases using four merged datasets. The model was evaluated using 5-fold cross validation, achieving an accuracy of approximately 98.56%. Additionally, B. Maheswar [52] used a lightweight CNN with four convolution-max-pooling layers for screening of Tuberculosis conditions from CXR images. The created model achieved an accuracy, F1-score and sensitivity of 95% with a peak Area Under Curve value of 0.976. Authors in [53] explored the use of shallow CNN to classify CXR images for Pneumonia detection, achieving an accuracy over than 90%. Moreover, N. Alrefai et al. [54] proposed CNN from scratch for Covid-19 detection, achieving an accuracy of 96.66%. Another custom CNN architecture, consisting of four convolutional layers, was developed to classify chest X-ray images into four categories: Normal, Pneumonia, COVID-19, and Tuberculosis. The model achieved an accuracy of 93% [55]. Q. Li [56] proposed a self-designed CNN model with two convolution layers for Pneumonia detection, achieving an accuracy of 83.33% and a precision of

85.57%. For the same task, I. Naskinova [57] applied a fine-tuned CNN with two convolution layers, achieving an accuracy higher than 90%.

3.2 Chest disease classification with Transfer learning

While CNNs trained from scratch have shown promising results in chest disease classification, especially with limited datasets, they often suffer from generalization issues. To deal with this limitation, Transfer learning allows the reuse of pretrained models on large-scale datasets for specific medical imaging tasks.

Many researchers relied on the application of transfer learning techniques for chest disease classification by fine-tuning models such as ResNet, DenseNet, Inception, etc. S. Guefrechi et al. [58] fine-tuned ResNet-50, VGG-16 and Inception-V3 for Covid-19 detection using a combination of CXR images collected from different datasets. A best accuracy of 98.30% was achieved by the VGG-16 model, and an accuracy of 98.10% and 97.20% was achieved by Inception-V3 and ResNet-50 respectively. K. Kansal et al. [59] conducted a performance comparison between ResNet-50 and EfficientNet-B0 for distinguishing between COVID-19 and Pneumonia. The EfficientNet-B0 model outperformed ResNet-50, achieving an accuracy of 98.74%, which is attributed to its compound scaling strategy. Authors in [60] trained the AlexNet model to classify Covid-19, non-Covid-19, viral Pneumonia, Bacterial Pneumonia and Normal CXR images collected from different datasets, achieving an accuracy of 93.42%. K. Kalaiselvi et al. [61] optimized a VGG-19 model for Pneumonia classification, achieving an accuracy of 99.8%. However, a fine-tuned VGG-19 was used in [62] for multi-class classification and reached an accuracy of 93.75%. Authors in [63] fine-tuned MobileNetV2 to classify diseases using Chest-Xray14 dataset [24] by adding multiple fully convolutional layers and residual bottleneck layers. The resulting MobileLungV2 achieved an accuracy of 96.97%, outperforming five pretrained CNNs (InceptionV3, AlexNet, DenseNet121, VGG19 and MobileNetV2).

V. Ravi et al. [64] applied multiple EfficientNet models using a stacking ensemble approach to detect lung diseases such as Pneumonia, Tuberculosis, and COVID-19, achieving an accuracy of 98%. To further enhance the detection of COVID-19 from chest X-ray images, S. Asif et al. [65] introduced a lightweight stacked ensemble model called LWSE, which integrates MobileNet with a lightweight CNN. This combination not only improved detection performance but also reduced computational complexity, resulting in an accuracy of 96.40%. In a similar task, Q. Hamad et al. [66] employed a pre-trained VGG-19 model alongside the Q-Learning Embedded Sine Cosine

Algorithm (QLESCA) for feature selection, with the selected features subsequently classified using a Support Vector Machine (SVM) classifier. I. Sirazitdinov et al. [67] proposed an ensemble model composed of RetinaNet [68] and Mask R-CNN [69] to detect Pneumonia using RSNA dataset [14]. It achieved an F1-score of 77.5%. A. Sahlol et al. [70] proposed a novel hybrid model for Tuberculosis detection. The model combines MobileNet, which is used as features extractor, and an Artificial Ecosystem-Based Optimization (AEO) to select the most relevant features. The model was trained on two different datasets [20], [71] achieving an accuracy of 90.2% and 94.1% respectively.

Hashmi et al. [72] developed a deep transfer learning-based method for pneumonia detection. They fine-tuned five pretrained CNN model (ResNet18, DenseNet121, InceptionV3, Xception, and MobileNetV2), and introduced novel weighted classifier that linearly combines predictions from all five models based on their individual performance. The designed ensemble model achieved test accuracy of 98.43% and an AUC of 99.76. Chowdhury et al. [73] fine-tuned eight pretrained CNNs including DenseNet201, ResNet18/101, VGG19, InceptionV3, MobileNetV2, and SqueezeNet for detecting Covid-19 and viral Pneumonia using a combination of multiple datasets. The models are evaluated under two classification tasks (Normal-Covid19) and (Normal-Covid19-Viral Pneumonia). DenseNet201 achieved the better result reaching an accuracy of 97.9% in the three-class classification task.

3.3 Chest disease classification with Vision transformers

As transformer models have gained popularity in computer vision and natural language processing, ViTs have increasingly been used for medical image analysis tasks. They have several distinct advantages compared to CNNs, including their capability to model long-range dependencies and global image context.

ViTs are applied in various chest disease detection tasks using CXR images, offering distinctive advantages over traditional convolutional neural networks for analyzing CXR images. T. Chen et al. [74], fine-tuned a ViT for Covid-19 detection. The model exceeded EfficientNet [75], multi-scale Vision Transformer (MViT) [75] and EfficientVit [76], achieving an accuracy of 95.79% in four-class classification (Covid-19, Lung opacity, Viral Pneumonia and Normal) and 99.57% in three-class classification. O. Uparkar et al. [77] trained a huge ViT model on the full Chest-Xray14 dataset [24] , and demonstrated that this model outperforms VDSNet model [78], which integrates classic CNN layers with a spatial transformer network. The huge ViT achieved an accuracy of

70.24% surpassing the VDSNET which achieved 69.86%. Building on the same dataset, authors in [79] developed a model named SwinCheX by fine-tuning the Swin Transformer model [80]. The Swin Transformer concept replaces the multi-head self-attention module in traditional ViTs with shifted windows, enabling the model to capture both local and global context with a linear complexity instead of quadratic complexity. The SwinCheX model achieved an AUC of 0.81.

M. Chetoui et al. [81] tested different pretrained ViT models to detect Covid-19, Pneumonia and Normal cases. The Covid-19 images are obtained from SIIM-FISABIO-RSNA Covid-19 dataset [82], The Pneumonia and Normal cases images are obtained from the RSNA dataset [14]. A best performance was achieved by ViT-b32 with an accuracy of 96% and AUC of 99%. For the same task, K. Krishnan et al. [83] fine-tuned a ViT-b32 achieving an accuracy of 97.61% and f1-score of 94.58%.

C. Liu et al. [84] proposed a fine-tuned model called Vision Outlooker (VOLO) [85] for COVID-19 detection. The VOLO architecture introduces a novel approach to token generation by using Outlook attention, which incorporates fine-grained features and local contextual information into tokens. These tokens are then processed through a series of stacked transformer blocks to capture global representations. The model demonstrated excellent performance, achieving an accuracy of 99.7%. To improve thoracic disease classification, X. Jiang et al. [86] proposed TransDD, a novel Transformer model based-dual-path-decoder. This model introduces a learnable label embedding that links class-related features directly to corresponding diseases labels. It achieved an AUC of 83.1% across all classes of chest-14 dataset, and an accuracy of 94.31% in three-class classification. J. Ko et al. [87] compared different optimization techniques to enhance the performance of ViT models (ViT, FastViT [88], CrossViT [89]) in classifying chest diseases. In the case of an imbalanced data with seven class diseases, FastViT with Adam optimizer achieved the best performance, with an accuracy of 97.93%. However, for a balanced dataset with four class diseases, ViT with the *Rectified Adam* optimizer achieved the highest accuracy of 95.87%.

3.4 Chest disease classification using hybrid models

Hybrid architectures, which combine the strengths of different DL architectures, are gaining traction in chest disease classification. Such approaches typically combine CNN vision transformers, attention mechanisms, machine learning techniques, or advanced feature fusion techniques to enhance the model's generalizability and classification performance.

Researchers have increasingly explored the development of hybrid architectures to leverage the complementary strength of different models, leading to more precise and reliable results.

C. Ejiyi et al. [90] proposed a novel model named ResfEANet to improve the classification of Tuberculosis using CXR images. They combined a shallow ResNet with a reduced number of residual blocks with an external attention mechanism without using the pretrained route. The ResfEANet model achieved an accuracy of 97.59% and a sensitivity of 100%. For the same task, S. Rajaraman et al. [91] designed an ensemble model by combining several pretrained CNN and ViT models, using lateral CXR views instead of traditional frontal views. Different ensemble strategies are used, where the SLSQP-based weighted averaging ensemble technique was the most effective. The model achieved an accuracy of 90.57%. Authors in [92] suggested a two-step approach to support Tuberculosis identification. They presented a new TB-UNet model for segmenting the lung region and a TB-DenseNet model for accurately classifying Tuberculosis images, reaching an accuracy of 95.10%.

C. Ukwuoma et al. [93] developed and ensemble model for Pneumonia detection. They used multiple pretrained CNNs for features extraction, then the extracted features are fed into a fine-tuned ViT for the identification and classification process, achieving an accuracy of 98.01%. A new model called PneuNet was developed by T. Wang et al. [94] for Pneumonia and Covid-19 classification. The model extracts feature from CXR images using ResNet18, and then applies a channel-based Vision Transformer with multi-head attention to perform the final classification. This approach achieved an accuracy of 99.32%. Another hybrid model proposed by Monday et al. [95] using capsule network and Multi-resolution Discrete Wavelet Transform (NW-CapsNet) to classify Pneumonia and Covid-19 using CXR images. The model addresses the challenges of spatial detail loss and noise in CXR images, achieving an accuracy of 99.6% and a sensitivity of 99.2%

On the other hand, S. Öztürk et al. [96] designed a hybrid convolutional-transformer model named HydraViT for multi-label thoracic disease classification using Chest X-ray14 dataset [24]. The model applies adaptive weighting to tackle the issue of co-occurring diseases such as pneumonia and effusion. These weights assist in normalizing the model's predictions, resulting in an AUC of 83.8 for all conditions.

Table 3.1 provides a detailed summary of the various researchs carried out binary classification of chest disease. However, table 3.2 focuses on multi-class classification, providing an overview of

the methods, datasets, and performance metrics used in studies that aim to distinguish among multiple chest conditions simultaneously.

Table 3.1 Performance comparison of the state-of-the-art methods for Uni-chest disease detection

Study	Pathology	Approach	Preprocessing technique	Hyperparameters	Dataset	Obtained result	Train and test time
[50]	Pneumonia detection	Lightweight CNN of 6 conv layers	✓ Image resizing (150×150) ✓ HE ✓ Image normalization ✓ DA with Geometric transformations	✓ Optimizer ✓ Lr = 0.001 ✓ Dropout = 0.5 ✓ Batch size= 128 ✓ Epochs = 100	Pediatric CXR dataset (Pneumonia) [16]	acc= 97.94% Recall= 98% AUC= 99.76%	Not mentioned
[53]		CNN with 3 blocks of Conv layers and 2 dense layers	✓ Image resizing (150×150)	Optimizer: Adam, SGD Lr = 0.001 Epochs = 30	[16]	acc= 92.08% Recall = 94.31% prec = 94.98% AUC =96.31 Better results with Adam optimizer	Not mentioned
[61]		Optimized VGG-19	Not detailed	Optimizer: Adam Lr = 0.0001 Batch size = 128 Epochs = 10	[16]	acc = 99.8% F1-score = 99.9%	Train = 17 min 59 s
[67]		Ensemble model (RetinaNet + Mask R-CNN)	✓ Image resizing (512×512) ✓ Image normalization ✓ DA with geometric transformations ✓ GC and random brightness adjustment	Optimizer: Adam Lr = 0.0001 and 0.001 for each model respectively Batch size= 8 and 6 Loss function = FL and bounding box regression loss	[97]	Recall = 79.3% prec = 75.8% F1-score = 77.5%	Train = 8 h (RetinaNet) and 6h (Mask R-CNN)
[93]		Ensemble of pretrained CNNS for features extraction + Fine-tuned transformer encoder for Pneumonia detection	✓ Image resizing (224×224) ✓ Image normalization ✓ DA with geometric transformations ✓ Data splitting	Optimizer: Adam Lr = 0.0001 Early stopping Transformer head = 8 Epochs = 100	[71], [98]	acc = 99.21% F1-score = 99.21%	Not mentioned
[72]		Ensemble model with	✓ Image resizing (224×224) and (229×229)	Optimizer: SGD Lr = 0,001	[16]	acc = 98,43% AUC = 99.76%	Inference time = 0,045s

		majority voting (ResNet18, DenseNet121, InceptionV3, Xception, MobileNetV2)	✓ Image normalization ✓ DA with geometric transformations	Weight decay = 0,0001 Epochs = 25			
[54]	**Covid-19 detection**	Lightweight CNN with 4 conv layers	✓ Image resizing (224×224)	Batch size = 16 Loss function: CE Lr = 3e-4 Dropout= 0,25 and 0,5 for the last layers Epochs = 30	[99]	acc= 96.66% F1-score = 96.77%	Train = 15,20 min
[58]		Fine- tuned ResNet-50 VGG-16 and Inception-V3	✓ Image resizing (224×224) ✓DA with geometric transformations	Optimizer: Adam Batch size = 32 Loss function: CE Lr = 0,0001 Epochs = 25	[100] [16] [101]	ResNet-50 acc = 97,2% prec = 97% Recall 96% VGG-16 acc = **98.3%** prec = **98.33%** Recall = 98% Inception-V3 acc = 98.10% prec = 98% Recall = 98%	Not mentioned
[83]		Fine-tuned ViT-b32	✓Image resizing (224×224) ✓DA with geometric transformation ✓Image enhancement with CLAHE	Optimizer: Rectified Adam with ReduceOnPlateau and early stopping Lr = 1e-4	[101] [102]	acc = 97.6% prec = 95.3% Recall = 93.87% F1-score = 94.6%	Not mentioned
[95]		CapsuleNet+ Wavelet multi resolution-CNN	✓Image resizing (224×224) ✓Image denoising with Wavelet Transform	Optimizer: AdamW Lr = 1 e-4 Batch size = 16 Dropout = 0.5 Epochs = 30	[24] [97] [102]	acc = 99.6% prec = 99.7% Recall = 99.2% AUC= 99.96%	Train = 23,2 min
[84]		VOLO	✓Image resizing (224×224) Random shuffle of normal images	Optimizer: AdamW Lr= 1,6 e-4 Decay rate = 0.1 Epochs = 30	[103] For training [99] For testing	acc = 99.7%	Not mentioned
[52]	**Tuberculosis (TB) detection**	Shallow CNN with 4 conv+ 2 dense layers	✓Image resizing (224×224) ✓Data splitting	Kernel size: 1 to 11 Kernel stride: 1 to 5 Optimizer: Adam Learning rate = 0.001	[104]	Accuracy = 95% F1-score = 95% AUC = 97.6%	Not mentioned
[90]		Shallow ResNet + External attention mechanism	✓Image resizing (224×224) ✓Data normalization ✓DA with geometric transformation	Optimizer: Adam Learning rate = 0,0001 Batch size = 32 Loss function: BCE Epochs = 300	[19]	acc = 97.59% prec = 95% spec = 95.56% sens = 100% F1-score = 97.44 AUC = 97.44%	Not mentioned

[91]		EM with SLSQP-Weighted Averaging (DenseNet-121 + ViT-b32	✓Image resizing (224×224) ✓Data normalization ✓DA with geometric transformation	Optimizer: SGD Lr = 1e-4 Loss function: CCE Epochs = Adaptive (Early stopping used)	[12] [23]	acc = 90.57% F1-score = 90.20% AUROC = 94.09%		Train = 110s
[92]		Modified U-Net (with attention block) for lung region segmentation Fusion of DenseNet-169 with dual conv blocks for TB classification	✓Image resizing (224×224×1) ✓Image normalization ✓Lung region segmentation ✓Data splitting	Lr = 0.001 Weight decay = 0.005 Dropout = 0.2 Epochs = 100	[19] [20] [105]	TB-UNet	acc= 95.74% F1-score = 89.88% IoU = 81.86	Train = 4,30 h Test = 2,691 s
						TB-DenseNet	acc = 98,98% F1-score = 99.01%	

Table 3.2 Performance comparison of the state-of-the-art methods for multi-chest disease detection

Study	Approach	Preprocessing technique	Hyperparameters	Dataset	Obtained result		Train and test time
[50]	Lightweight CNN with 6 conv layers	Mentioned in table 3.1	Mentioned in table 3.1 Loss function: Sparse CE	Combination of datasets [16] [106] [107]	acc=80%		Train = 6152,62 s Test= 2,72 s
[51]	Lightweight CNN with 6 conv layers	✓ Image resizing (224×224) ✓ Image normalization	Optimizer: Adam Batch size = 16 Loss function: CE, FL Lr= 0,0001	Combination of datasets [101] [104] [98] [16] [102]	acc = 98.56% Recall= 98.91% prec = 99.10% F1-score = 99.01% AUC = 99.30%		Train = 55 min Test = 18,42 s
[55]	Lightweight CNN with 4 conv layers each paired with BN layer and one dense layer	✓ Image resizing (224×224) ✓ Data balancing with under sampling technique	Epochs = 15 Early stopping to prevent overfitting	[99] [108] [109]	acc = 93%		Not mentioned
[60]	AlexNet	Data splitting: 70% training 30% testing	Optimizer: SGD Learning rate = 0.0001 Epochs = 20	Combination of Covid-19 CXR images [71] [100] [110]	acc = 93.42% sens = 89.18% spec = 98.92%		Not mentioned
[64]	Ensemble stacking approach of	✓Image resizing (224×224)	Optimizer: Adam Batch size = 64 Learning rate = 0.0001	[16] [111] [112]	Covid-19	acc = 98%	Not mentioned
					TB	acc= **99%**	

	EfficientNet models		Epochs = 15		Pneumonia	acc= 98%	
[63]	Fine-tuned MobileNet	✓Image resizing (224×224) ✓ Image denoising with gaussian filter ✓ Contrast enhancement with CLAHE ✓ DA with geometric transformation. ✓ Oversampling and undersampling for data balancing	Optimizer: RMSProp Learning rate = 0.001	[24]	acc= 96.97% prec = 96.71% Recall = 96.83% sens = 99.78%		Not mentioned
[81]	Fine-tuned ViT-b16 ViT-b32 ViT-L32	✓Image resizing (512×512) ✓Data augmentation with geometric transformation	Optimizer: Rectified Adam Lr = 1×10^{-4} Batch size = 16 (ViT-b16 and ViT-b32) Batch size = 4 (ViT-L32) Epochs = 200	[82] [113]	ViT-b16	acc = 87% AUC = 96%	Not mentioned
					ViT-b32	acc = **96%** AUC = **99%**	
					ViT-L32	acc = 53% AUC = 79%	
[94]	EM: ResNet-18 with multi-head attention	✓Image resizing (224×224) Data augmentation with geometric transformation	Optimizer: AdamW Lr = 0,001 Weight decay = 0,00001 Batch size = 16 Dropout rate = 0,2 Epochs = 200	[16] [114] [115] [97] [12]	Three-category classification	acc = 95.16% prec = 97.11% Recall = 97.33% F1-score= 97.26%	Not mentioned
					Four-category classification	acc = 90.03% prec = 89.58% Recall = 89.62% F1-score= 89.59%	
[77]	ViT-huge	✓Image resizing (224×224) ✓Image normalization	Optimizer: AdamW Lr = 0.0001 Weight decay = 0.000001 Loss function = Cross entropy Dropout rate = 0,2 Epochs = 20	[24]	Accuracy = 70.24% F1-score = 65%		Not mentioned

[79]	Pretrained Swin Transformer + MLP	✓Image resizing (224×224)	Optimizer: Adam Lr = 3×10^{-5} Weight decay = 0,000001 Loss function = BCR Epochs = 20	[24]	AUC = 81%	Not mentioned
[73]	Fine-tuned DenseNet201	✓Image resizing (224×224) ✓Image normalization Data augmentation with geometric transformation	Optimizer: SGD Lr= 0.001 Batch size=16 Epochs = 80	[101], [116]	Accuracy = 97.9% Sensitivity = 97.95% Specificity= 98.8%%	Train = 264,79 s for each epoch
[96]	Local CNN features with global ViT attention	✓Image resizing (224×224) ✓Data splitting	Optimizer: Adam Learning rate = 1e - 4 Batch size = 35 Attention heads = 20 Patch size = 4×4 Epochs = 120	[24]	AUC = 83.8%	Not mentioned

4. Anomalies localization

Surpassing the task of just classifying CXR images into “diseased” or “healthy” and the distinction between different diseases, anomalies localization plays a crucial role in medical image analysis. Effectively identifying the precise region of pathology helps to enhance the diagnosis and improve the treatment planning.

The utility of applying DL algorithms for chest disease diagnosis enhanced and becomes more trustworthy when the classified disease precisely resides in the CXR image. Localization is important for guiding clinicians to the relevant regions, enabling differential diagnosis, tracking disease progression, and applying appropriate interventions. Anomaly localization methods in deep learning generally include object detection (detecting the presence of a bounding box for the detected anomalies), and semantic segmentation by producing a pixel-level image of the abnormal area.

Anomalies localization is not merely a separate task and must be appreciated as being intrinsically tied to the process of classification. Accurate localization can provide visual confirmation of the outcomes of classification, while robust classification can, in turn, help refine and validate localized findings. The two-way reinforcement present from the interaction provides a further

enriched and pragmatic diagnostic result, effectively transcending the gap between generic disease prediction and anatomical particulars.

This section reviews deep learning architectures developed for chest-related anomalies in CXR images including Faster R-CNN, YOLO (You Only Look Once), RetinaNet, and SSD (Single Shot Detector). These models require pixel-level or bounding box annotations, which might be costly but produce highly exact localization results.

For instance, Hao et al. [117] applied a modified YOLOv8s architecture to improve the accuracy of multi-disease detection and small-lesion localization in CXR images. They introduced a novel attention mechanism to boost the sensitivity to spatial location information and captures long-range dependencies. The model achieved an mAP@0.5 of 0.338, which means it detected lesions fairly accurately. Authors in [118] used YOLOv5s and Faster R-CNN to detect and localize five different chest diseases through CXR images. YOLOv5s was used to extract lung regions, then Faster R-CNN was applied to localize abnormalities using bounding boxes, achieving an AP@0.5 of 0.58 and AP@0.5:0.95 of 0.15 across five types of chest anomalies. Jain et al. [119] worked on detecting and localizing Pneumonia and lung opacities using Faster R-CNN with a ResNet backbone. Authors confirm improved localization and detection achieving an accuracy of 95.28%.

On the other hand, Fan et al. [120] used YOLOX (You Only Look Once - eXtreme) model to detect and localize 14 chest diseases using a large dataset. The model achieved an mAP@0.5 of 0.629 outperforming Faster R-CNN, RetinaNet, and even radiologists' localization. However, in another study, authors explored multiple cutting-edge object detection models. YOLOX, Faster R-CNN-based few-shot models, and binary classification were combined to detect and localize 12 chest diseases in X-rays. FSCE 10-shot achieved better performance with an mAP@0.5 of 0.3343. V. Tiwari et al. [121] applied YOLO and RetinaNet to accurately localize regions associated with Covid-19 the ensemble model achieve a mAP of 0.552, representing an improvement over their individual predictions. However, L. Mao et al. [122] developed an ensemble of improved RetinaNet and Mask R-CNN to detect and localize Pneumonia in CXR images, achieving a recall of 0.813 and a mAP of 0.2283.

5. Explainable Artificial Intelligence in Chest disease interpretation

XAI has gained more importance in the field of chest disease classification, noting the vital need for transparency in clinical decision support systems. Knowing the reason for diagnostic recommendations is vital for physician trust and patient safety. When classifying chest disease based on CXR images, the use of XAI renders the model transparent and interpretable, pointing out regions of the image that affected the decision of the model.

Lee et al. [123] demonstrated the application of visualization-based segmentation for cardiomegaly detection using CXR images. This approach clearly visualizes the lung and heart regions to effectively detect cardiomegaly. Authors in [124] explored visualization and local feature extraction techniques for model interpretability. They applied the Local Interpretable Model-Agnostic Explanations (LIME) approach, which explains predictions by perturbing input data and analyzing its effect on model outputs. LIME was also applied in [125] to evaluate explanation quality in chest disease classification using CXR images. Similarly, B. Maheswari et al. [52] employed both CAM and LIME methods to analyze the decision made by the shallow CNN in detecting Tuberculosis. They demonstrated that their model delivers significant and interpretable predictions by concentrating on clinically important regions of the lungs.

Additionally, to improve chest disease diagnosis using CXR images, more XAI advancements have been made. A. Aasem et al. introduced Ensemble-CAM, a novel XAI framework to enhance the localization of chest disease by combining multiple CAM techniques including Grad-CAM and Grad-CAM++. Ensemble-CAM generates bounding boxes and interpretable heatmaps without the need for strong supervision during training. J. Zhao et al. [126] enhanced the accuracy of generating bounding boxes for chest X-ray image diagnosis. They integrated two XAI techniques, Guided Backpropagation and Grad-CAM++. The Guided Backpropagation highlights the most influential pixels in the model's prediction, producing a high-resolution gradient map. However, it is not class-discriminative, highlighting regions that may overlap unrelated classes. The Grad-CAM++ produces class-discriminative heatmaps but with low resolution. Combining the two approaches with a weighted average ensures that the final bounding box is more focused on the disease region.

(a) **(b)** **(c)**

Figure 3.2 CAMs visualization of different chest diseases: (a) large pleural effusion, (b) Congestive heart failure and cardiomegaly detected, (c) Primary lung malignancy with two masses detected

6. Analysis and discussion

The choice of the appropriate deep learning model for an accurate chest disease detection and classification is a critical task. CNNs and ViTs are the two leading architectures used for medical image analysis. Several studies adopted CNNs trained from scratch, especially lightweight models, due to their computational simplicity and suitability for limited datasets such as [109]. These models generally achieved better performance in binary classification [52] [54] compared to multi-class classification [50]. The application of cross-validation enhances the model performance in detecting seven diseases [51]. For instance, using more than four convolutional layers helps in achieving better results than using two convolytional layers as in the case of [56] [57]. The high number of convolution layers enables the model to learn high-level features, which leads to better performance.

With the growth of large publicly available CXR datasets such as ChestX-ray14 and COVIDx, researchers began to use pretrained models such as VGG-16, ResNet-50, InceptionV3, and EfficientNet, etc. Several studies demonstrated that ResNet and DenseNet show excellent performances in chest disease detection [127], [128], [129], [130], and even for multi-disease classification [131], [132]. The simplified architecture of MobileNet make it more efficient for binary classification [63],[133] more than complex multi-class classification involving mimicking diseases [134].

Moreover, ensemble techniques further enhanced classification performance. Combining multiple models, either through stacking, majority voting, or weighted averaging. These models consistently yielded superior results in robustness and prediction stability. Notable examples include the ensemble approaches in studies [94], [64]. Table 3.2, and table 3.3 provide more details about these studies.

On the other hand, Vision transformer-based architectures have introduced a paradigm shift. Multiple studies, including [74] [81] demonstrated that ViTs achieve high performance in multi-class classification, outperforming CNNs in certain settings. The majority of research relies on based vision transformers (e.g. ViTb-16, ViTb-32) rather than huge architecture (e.g. VitL-16, VitL-32). This is primarily due to high computational cost and memory requirements associated with training and deploying large-scale transformers, which can be prohibitively expensive, particularly in medical context where access to high-performance computing resources is limited. Additionally, despite the deeper architecture of large-scale transformer models, studies [81] and [77] reported that they achieved lower performance compared to their base counterparts. This may be attributed to their higher sensitivity to data quality, the need for extensive training data, and increased risk of overfitting in medical imaging tasks with limited datasets.

In addition to classification, anomaly localization was identified as a critical task to enhance the clinical utility and trustworthiness of DL models. Techniques such as YOLO, Faster R-CNN, and RetinaNet have been effectively applied to localize disease-specific regions in CXR images. Studies [117] and [122] highlighted improvements in detection precision and visual explanation, with some models outperforming human-level annotation under specific settings.

Furthermore, the integration of XAI techniques such as CAM, Grad-CAM++, LIME, and Ensemble-CAM, has significantly enhanced model interpretability. These tools allow clinicians to visualize which parts of the image influenced the decision, improving clinical trust and decision support.

Table 3.3 An evaluation of existing work for chest disease detection-based Transfer learning

Model architecture	Key architectural features	Advantages	Limitations	Implementation studies		
				Ref	Pathology	Performance
VGG-16 [135]	• A sequential model with13 Convolutional layers, 5 max pooling	✓ Good feature extracted for classification tasks.	- High computational cost and memory usage.	[136]	Pneumonia	Accuracy = 92,15% and 95,40% using two different datasets

	layers and3 dense layers (4096, 4096, 1000 channel respectively). • For the VGG-19: There is 16 Convolutional layers and 3 dense layers	✓Power at capturing special hierarchies	- Less effective for small datasets and prone to overfitting.	[137]	Covid-19	Accuracy = 97,3%
				[138]	Pneumonia	Accuracy ≈87%
				[139]	Tuberculosis	Accuracy ≈ 90%
				[140]	Multi-chest disease	Accuracy = 98.05%
ResNet [141]	• Verry deep network with shortcut connections (ResNet-50, ResNet-101) • Incorporates residual blocks with skip connections that directly connect the input block to its output. • Include batch normalization	✓ Reduce training time with skip connections. ✓ Mitigates vanishing gradient problem ✓ Strong generalization capacities and powerful for complex datasets.	- Computationally expensive for deeper version. - Training on imbalanced data can lead to suboptimal performance.	[127]	Pneumonia	Accuracy = 94,01%
				[142]	Covid-19 Pneumonia	Accuracy = 92,63%
				[131]	Detection of conventional Pneumonia from Covid-19 cases	Accuracy = 99,51%
				[143]	Cardiomegaly	Accuracy ≈ 80%
				[128]	Covid-19	Accuracy = 98,20%
DenseNet [144]	• Densely connected network: each layer receive input from all the previous layers. • Employs large number of filters per layer to control complexity.	✓ Encourages feature reuse, reducing redundancy and improving efficiency	- Bottleneck issue may occur in extremely deep network. - The dense connection increases the computational overhead during training.	[129]	Cardiomegaly	Accuracy = 95%
				[145]	Tuberculosis	Accuracy = 98,80%
				[132]	Multi-chest disease	90,4<Accuracy<98,5%
				[130]	COPD	Accuracy = 96.8%
				[146]	Multi-chest disease	0,68<AUC<0,93
EfficientNet [75]	• Lightweight architecture • Uses compound scaling to scale width, depth and resolution. • Uses depthwise separable convolutions and inverted residual structure to reduce the number of parameters.	✓ Efficient scaling across different resource constraints	-Requires large amounts of high-quality data for optimal performance. - Complicated to design and require specialized scaling rules.	[147]	Pulmonary Edema and Pleural effusion	Accuracy = 98,3%
				[148]	Multi-chest disease	AUC = 0,9080
				[149]	Pneumonia and Pneumothorax	Accuracy = 82,20%
				[150]	Tuberculosis	Accuracy = 99%
MobileNet [151]	• Uses depth-wise separable convolutions to reduce computational complexity.	✓Ideal for real-time applications. ✓Low memory and	- Lower performance compared to complex and large model.	[133]	Pneumonia	Accuracy = 94,23% and 93,75%
				[134]	Multi-chest disease	Accuracy > 90%
				[152]	Multi-chest disease	Accuracy = 93,30%

	• Expands the number of channels prior to implementing depthwise convolutions, which enhances the efficiency of feature extraction.	computational needs.	- Limited capacity to learn complex features as a result of the reduced parameters.	[153]	Tuberculosis	Accuracy = 98,66%
Inception V3 [154]	• Updated version of GoogleNet network. • Parallel convolutions to extract features at different scales, by applying a combination of convolutions (1×1, 3×3, 5×5) • Combines features from different paths for a more robust representation.	✓ Extracting features at multiple scales reduces overfitting. ✓ Efficient computation, reducing cost by using 1×1 convolution	-Complex architecture and difficult to fine-tune.	[155]	Pneumonia	Accuracy = 99,29%
				[156]	Tuberculosis	Accuracy = 99%
				[157]	Cardiomegaly	Accuracy ≈ 99%

7. Conclusion

This systematic review provides a comprehensive analysis of recent deep learning methods applied to chest disease diagnosis using chest X-ray images. The reviewed studies highlight significant advancements in classification performance, localization precision, and model interpretability. A clear evolution is visible, from shallow CNNs trained on small datasets to ensemble and transformer-based models leveraged on large-scale public datasets.

The key findings indicate that while fine-tuned pretrained CNNs like ResNet, DenseNet, and EfficientNet remain robust and widely used baselines, the emergence of Vision Transformers and hybrid architectures marks a significant step forward, particularly in handling complex multi-class problems with overlapping radiological features. The consistent outperformance of ensemble models underscores the benefit of combining diverse architectures to enhance predictive power and reliability.

However, several gaps and challenges persist. First, while many studies report high accuracy, the direct comparison remains difficult due to the high variability in datasets, preprocessing pipelines, and evaluation metrics used.

Standardization in benchmarking is needed for more objective performance assessment. Second, the issue of data imbalance continues to be a significant challenge; although methods such as SMOTE and class weighting are used, their use is not fully generalizable, which may result in models being biased towards majority classes.

Finally, despite the progress in XAI and localization, most studies are still in a validation phase. The translation of these models into routine clinical practice is hindered by the need for more extensive clinical trials, regulatory approval, and seamless integration with existing hospital information systems.

The insights gathered from this review directly inform the direction of this thesis. The identified limitations in existing works, such as the struggle of CNNs with long-range dependencies, the computational cost of ViTs, and the need for powerful models capable of distinguishing between mimicking pathologies (e.g., Tuberculosis and Pneumonia), serve as an impetus for the novel approaches introduced in the chapter to come. This study will try to transcend these limitations by suggesting novel hybrid and efficient models to make precise and trustworthy diagnoses of chest disease.

Chapter 4

Contributions

1. Introduction

Chest disease detection and classification using X-ray images pose considerable challenges in contemporary medical practice. These challenges vary from overlapping and mimic symptoms of diseases, complexity of manual interpretation of chest X-ray images, the lack of expert radiologists especially in the third world countries, to inconsistencies in interpretation between observers. Even experienced radiologists face significant difficulties in differentiating between closely related diseases, like early-stage Tuberculosis, subtle manifestations of Covid-19, viral versus bacterial Pneumonia, etc. Such diagnostic uncertainty leads to delay in treatment, inappropriate therapeutic interventions, and ultimately patient outcome. The development of artificial intelligence tools, especially deep learning algorithms had the potential to reduce diagnosis time drastically, while attaining or even exceeding human interpretive accuracy. Convolutional Neural Networks (CNNs) and Transformer-based architectures, have emerged as useful tools for medical image analysis. CNNs are effective in capturing local and hierarchical patterns within chest X-ray images, while Vision Transformers (ViTs) and more recently Vision Mamba models have shown such potential in global contextual relationships and long-range dependencies encoding.

Through the integration and comparative evaluation of these architectures, this chapter presents the main scientific contribution of this thesis, which is a series of studies that cumulatively address the challenging issues of automatic chest disease detection using deep learning algorithms. The studies cover the designing of models, optimization of their performance, and comparison, to ultimately enhance the diagnostic accuracy and robustness in detecting chest disease from X-ray images.

2. Datasets and training objective

Among the chest diseases prevalent around the world, we chose to focus on Pneumonia and Tuberculosis. These diseases together account for a massive share of global respiratory mortality. Interpretation of Pneumonia and Tuberculosis on CXR images suffer from inter-reader variability. The overlapping clinical patterns (e.g. consolidations, infiltrates, opacities) often make differential diagnosis difficult, even for expert radiologists.

Furthermore, the choice of X-ray images over other modalities such as CT scans or MRI is strategically supported for many reasons:

- The widespread availability of CXR images, they are available in nearly all hospitals and many clinics worldwide.
- The low cost of CXR images compared to CT scans and MRI, making them the primary screening tool in many diagnoses process.
- The rapid acquisition of CXR images makes them crucial for efficient patient management in endemic areas.
- CXR images are the primary imaging modality for suspected pneumonia and tuberculosis, as per WHO and the majority of national guidelines.

2.1 Criteria for data selection

For Pneumonia, we used the dataset proposed by Kermany et al. [71]. This dataset contains 5863 CXR images of normal, viral and bacterial pneumonia cases, obtained from Guangzhou Women's and Children's Medical Centre.

For Tuberculosis, we used three different datasets which are:

- The Montgomery County (MC) CXR dataset [19], which contains 58 CXR images of Tuberculosis manifestation and 80 normal cases collected from Montgomery County (MC), Maryland, USA's Department of Health and Human Services (HHS) tuberculosis control program.
- The dataset from the National Institute of Tuberculosis and Respiratory Diseases [158] comprises a total of 278 CXR images, of which 125 represent Tuberculosis and 153 are Tuberculosis-negative.
- The Tuberculosis dataset of Kaggle, collected by a team of researchers from universities in Qatar and Dhaka (Bangladesh) in collaboration with Malaysian partners [109]. It contains 4200 CXR images in total, with 3500 images of normal cases and 700 Tuberculosis cases.

These datasets are described in greater detail in Chapter 1, along with the reasons for selecting them can be summarized in:

- All the mentioned datasets are publicly available.

- The Pneumonia dataset of Kermany contains CXR images from children aged between one and five years, the highly demographic infected with this disease. The World Health Organization (WHO) announced that Pneumonia caused more than 800,000 child fatalities in 2017, and in 2018, a child died from the illness every 39 seconds. It is estimated that 11 million children will die from pneumonia by 2030.
- Unlike many other datasets, the children Pneumonia dataset of Kermany is the only one that contains the details about viral and bacterial cases. This separation directly addresses the clinical decision supports to determine if antibiotics should be given, as they work for bacterial pneumonia but are ineffective against viral cases.
- All the images of Pneumonia datasets are reviewed by two experts' radiologists ensuring truth labels for model's training and evaluation. Additionally, this dataset is widely used in published researches, enabling efficient and objective comparison.
- For Tuberculosis detection process, the MC dataset contains high quality images (12-bit grayscale) with radiologist annotations and segmentation masks of lung regions.
- The MC dataset was with appropriate consent, tackling standard ethical issues related to data.
- The MC dataset incorporates an acceptable balance between normal and abnormal cases, which is advantageous for DL models training.
- The availability of the ground truth lung mask in the CXR images of the MC dataset facilitates the integration of XAI to increase the trust of model.
- The Tuberculosis dataset of Kaggle contains subtle and difficult-to-detect cases that can push the boundaries of DL learning models.
- The dataset from the National Institute of Tuberculosis and Respiratory Diseases was selected due to its clinical relevance, high-quality annotations, and balanced representation of both diseased and healthy cases. This makes it a valuable resource for training and evaluating models aimed at accurate Tuberculosis detection.

2.2 The target objective of deep learning model's training

The main objective of training deep learning models in this overall research is to develop a robust and accurate CAD system for the automated detection and classification of chest disease, focusing on Pneumonia and Tuberculosis using CXR images.

Misinterpretation of chest diseases is common due to low contrast of images, human subjectivity, or insufficient radiological expertise. Training DL models efficiently helps to improve accuracy and reproducibility of the diagnosis process. To address this purpose, we proposed:

- A hybrid CNN-XGboost for Pneumonia detection and the distinction between viral Pneumonia and bacterial Pneumonia.
- An ensemble ResNet50-ViTb16 model for Tuberculosis and Pneumonia classification.
- An efficient fine-tuned Vision Mamba model for Tuberculosis detection.

3. Experiment 1: An Improved CNN-XGboost Model for Pneumonia Classification

To tackle the challenges related to chest disease detection and classification through deep learning algorithms, we conducted a series of structured experiments designed to progressively enhance the diagnosis process. As a first experimentation step, a hybrid model combining CNN and Extreme Gradient Boosting (XGboost) was developed to accurately distinguish between Normal cases, Viral Pneumonia and Bacterial Pneumonia [159].

The motivation behind focusing on Pneumonia lies particularly in the importance of distinguishing between viral and bacterial types. This distinction directly impacts the diagnosis, treatment and the patient outcomes. Viral pneumonia requires antiviral medications or just a supportive care, whereas bacterial pneumonia is more aggressive and treated with antibiotics. If not treated promptly, bacterial pneumonia can cause severe complications such as pleural effusion, acute respiratory distress syndrome (ARDS) [160] or lung abscess [161].

3.1 Preprocessing and data preparation

In this study we used the children Pneumonia CXR dataset [71]. To increase the volume of data and address the challenge of data imbalance, Albumentations, a widely

used data augmentation tool, was employed to enhance the variability of the training data [162].

Albumentations is a python library, particularly designed for data augmentation in machine learning tasks. It provides an extensive range of effective augmentation methods to artificially increase the size and diversity of training datasets, aiding models in improving their generalization capabilities. This tool is easy to integrate into DL pipelines and contains advanced transformations like random brightness or contrast, histogram equalization, grid distortions and elastic deformations, etc.

The transformation pipeline used in this experiment is detailed in table 4.1. The horizontal flip helps the model become robust to right and left variations, applying random rotation helps the model become more tolerant to slight angle variations that often occur in real-world clinical imaging, such as when patients move slightly or are positioned inconsistently. This makes the model more robust and improves its classification accuracy. Furthermore, the integration of random Gaussian noise helps the model to focus on meaningful features instead of noisy patterns, improving the model's generalization.

The specific values of these augmentation operations were determined after multiple phases of testing and experiments; different combinations were evaluated to determine the optimal one between data variety and model stability.

These augmentations were applied after resizing the input images to 64×64 pixels, and converting them from BGR to RGB format to maintains consistent channel ordering, ensuring compatibility with the input expectations of most deep learning libraries. Finally, all the images are normalized to the interval [0, 1] to enhance the training process and improve model's convergence.

Table 4.1 The applied data augmentation for CNN-XGboost training

Augmentation pipeline	Parameters
Horizontal flip	P = 0,9
Random degree rotation	Limit = 10
Gaussian noise	P = 0,9

3.2 Model architecture

The use of CNNs in Pneumonia detection has been the subject of several studies in the last years. These models are specially designed for visual data analysis. The hierarchical structures of CNNs enables them to detect the progression of lungs infection, eliminating the need for manual features engineering. Additionally, the pooling mechanism of these models manages the complexity of high dimensionality of CXR images, enabling the learning of meaningful representations without needing an unwieldy number of parameters.

On the other hand, XGboost, being a tree-based ensemble method, excels at capturing non-linear relationships across high-dimensional feature spaces by building decision trees one after the other to correct the errors of previous ones. Its gradient boosting framework optimizes a differentiable loss function which makes it effective in classification tasks with imbalanced data distribution which is common in medical imaging datasets. XGboost provides valuable interpretability by quantifying how much each feature contributes to model decision. Moreover, the regularization mechanisms of XGboost helps to prevent overfitting which is optimal with moderate sized datasets.

The combination of CNN with XGboost model enables the extraction of high-level features from CXR images through the convolutional layers, followed by robust classification using XGboost. This combination is designed to benefit from the effectiveness of CNNs at learning spatial patterns in medical images and the advanced regularization techniques of XGboost to reduce overfitting. Additionally, XGboost supports efficient parallelization, it calculates the gradient statistics in parallel. During the construction of each decision tree, it performs parallel computation to determine the optimal split for every feature across all data points. These operations can be executed simultaneously on multiple CPU cores because the calculation of each feature is independent of the others.

The suggested architecture is built upon a shallow CNN composed of five convolutional layers with increasing numbers of filters (16, 32, 64, 128, 256) respectively, followed by a fully connected layer of 128 neurons (*Figure 4.1*). After each convolutional block, we added a batch normalization layer to improve the gradient flow and reduce the problem of internal covariate shift by stabilizing the input distributions across training.

We also incorporated the fully connected layer with L2 regularization to penalize large weights and prevent the model from learning only complex patterns. Additionally, a dropout layer with a rate of 0.7 was added after this fully connected layer to prevent overfitting. Finally, we integrated the XGboost as a final classifier to process the features extracted by the CNN to make the final classification decision.

The implemented XGboost consists of 200 sequential decision trees, each one models a portion of the classification task. The internal nodes perform binary splits based on feature tests on the 52,163-dimensional CNN-extracted feature vector. However, the leaf nodes contain the predictions score which are aggregated across all trees through weighted summation to produce the final classification. *Figure 4.2* shows the entire architecture of our hybrid CNN-XGboost model.

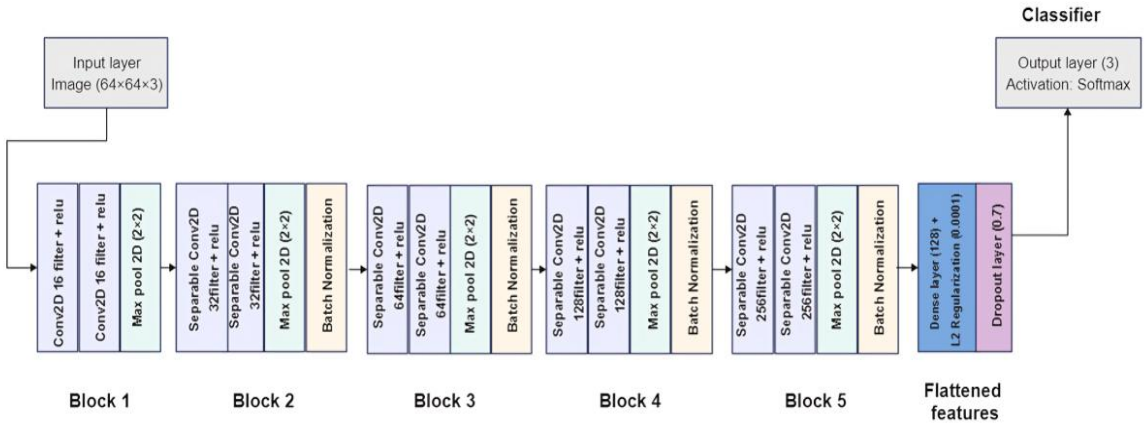


Figure 4.1 The architecture of the proposed shallow CNN for Pneumonia detection

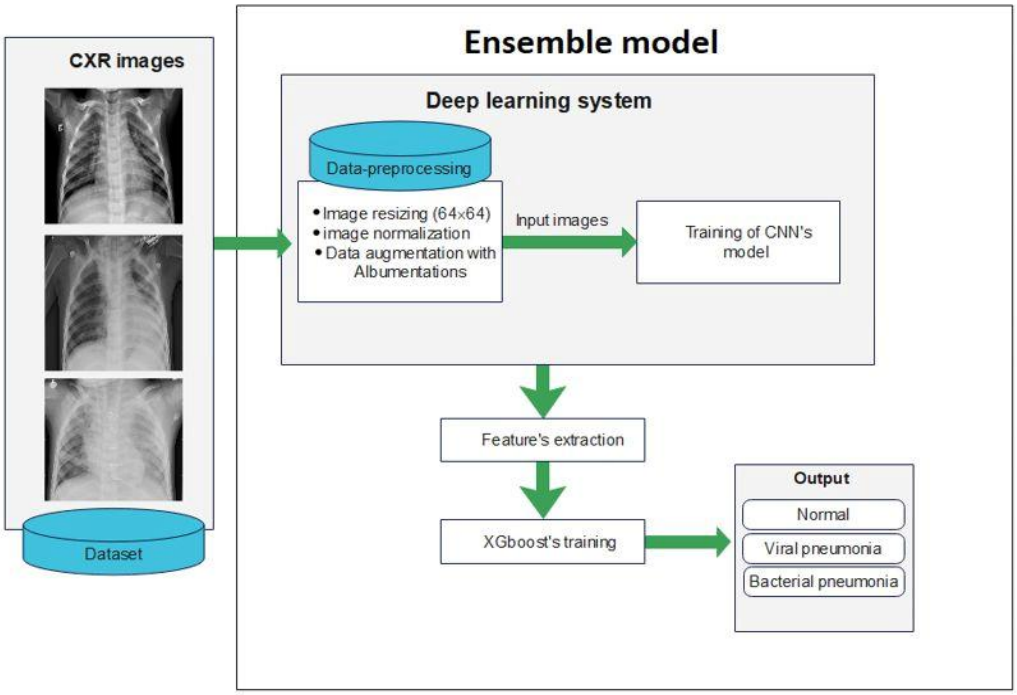


Figure 4.2 Block diagram of the hybrid CNN-XGboost model for Pneumonia detection

3.3 Model training

The primary objective of training our hybrid model is to learn discriminative patterns from CXR images to effectively distinguish between normal cases, viral Pneumonia and bacterial Pneumonia. This training process was carried out in two main phases: Training of CNN for features extraction, followed by the training of the XGboost classifier.

A. Shallow CNN training

The proposed shallow CNN model was trained for *40 epochs* accompanied with early stopping with a *patience* of *5* and a *min_delta* of *0,001*. After *40* epochs the model's performance did not show meaningful improvement. *Adam* optimizer with a *learning_rate* of *0,0001* was conducted to adjust the weights. Different regularization techniques were experimented with, and *L2 regularization* with $\lambda = 0.0001$ proved most effective in enhancing model's stability and dealing with the overfitting problem encountered during the training. The *batch size* was set to *128* based on memory constraints and convergence stability. To monitor the model's generalization, the dataset was split into 80% for training and 20% validation and test. Additionally, the

input image resolution of 64×64 helped us to reduce the computational requirements while preserving the pathological details.

These parameters and values were selected based on extensive evaluation, where different configurations were tested to fix the optimal setup that provides a stable balance model accuracy and generalization performance.

B. XGboost classifier training

Following the CNN training, the XGboost classifier was trained on the extracted features. A series of thorough experiments was carried out with learning rates varying from 0,00001 to 0,1. The extracted feature vector of length 52,163 from the CNN's penultimate layer served as input to the XGboost classifier, forming a critical bridge between deep feature extraction and gradient boosting classification.

The number of the trees (estimators) in the XGboost model was determined empirically through hyperparameter tuning to preserve the balance between computational complexity and the classification performance.

Our extensive grid search revealed that 200 estimators worked best, providing optimal performance. Further increment in this value introduced decreasing returns in *accuracy* with a significant increase in computational overhead. The *learning rate* was finally set to 0,001 after careful evaluation, since this rate provided sufficient gradient descent steps without overshooting optimal weight values.

3.4 CNN-XGboost pseudo-code

To provide a clear and systematic understanding of the proposed hybrid CNN-XGboost methodology, this section provides the detailed algorithmic framework that governs the implementation of the Pneumonia classification system. The pseudo code delineates the complete workflow from initial data preprocessing through feature extraction to final classification, highlighting the synergistic combination of CNN's spatial feature learning capabilities with XGboost classification framework.

ALGORITHM 1: CNN-XGBoost Pneumonia Classification

INPUT: Chest X-ray images dataset (Normal, Viral Pneumonia, Bacterial Pneumonia)

#PHASE 1: Data Preprocessing and Augmentation

FUNCTION preprocessing_data(dataset):
FOR each image in dataset:
Resize image to 64×64 pixels
Convert from BGR to RGB format
Apply augmentation pipeline
Normalize pixel values to [0,1]
RETURN preprocessed_dataset

#PHASE 2: CNN Feature Extraction Training

FUNCTION train_cnn_feature_extractor(train_data, val_data):
Initialize CNN architecture:
- Conv2D layers: [16, 32, 64, 128, 256] filters
- Batch Normalization after each conv layer
- Fully connected layer: 128 neurons with L2 regularization (λ=0.0001)
- Dropout layer: rate=0.7
SET optimizer = Adam(learning_rate=0.0001)
SET batch_size = 128, epochs = 40
SET early_stopping(patience=5, min_delta=0.001)
FOR epoch = 1 to epochs:
TRAIN CNN on train_data
VALIDATE on val_data
IF early_stopping_criteria_met:
BREAK
RETURN trained_cnn_model

#PHASE 3: Feature Extraction

FUNCTION extract_features(cnn_model, data):
features = []
FOR each image in data:
feature_vector = cnn_model.penultimate_layer(image) // 52,163 dimensions
features.append(feature_vector)
RETURN features

#PHASE 4: XGBoost Training

FUNCTION train_xgboost_classifier(features, labels):
Initialize XGBoost parameters: n_estimators = 200; learning_rate = 0.001
xgb_model = XGBoost(parameters)
xgb_model.fit(features, labels)
RETURN xgb_model

#MAIN EXECUTION:

train_data, val_data, test_data = split_dataset(ratio=[0.8, 0.1, 0.1])
preprocessed_train = preprocess_data(train_data)
cnn_model = train_cnn_feature_extractor(preprocessed_train, preprocessed_val)
train_features = extract_features(cnn_model, preprocessed_train)
test_features = extract_features(cnn_model, preprocessed_test)
xgb_classifier = train_xgboost_classifier(train_features, train_labels)
predictions = xgb_classifier.predict(test_features)
RETURN predictions, performance_metrics

3.5 Results and discussion

The proposed CNN-XGboost exhibited encouraging performance in Pneumonia classification (Table 4.2), particularly in viral vs. bacterial Pneumonia discrimination — a challenging endeavor considering overlapping clinical patterns. The model achieved an overall *accuracy* of 87% beating the baseline CNN which achieved an accuracy of 85%. The *accuracy* improvement from CNN (86%) to CNN-XGboost (89%) illustrates the effectiveness of our hybrid method in removing false positives by CNN boundary refinement through feature extraction. However, the overall *recall* of 85% illustrates a satisfactory capability in recalling most of the true viral and bacterial Pneumonia cases but indicates that some infected samples with noisy or confusing features are missed. Such *sensitivity* is a decision-support tool for

radiologists during diagnosis; however, expert confirmation would still be recommended in order to deal with the missed cases. In addition, the 87% *F1-score* demonstrated the balanced performance between *precision* and *recall*, which indicated that both false positives and false negatives are areas of further improvement, especially in a clinical environment where diagnostic accuracy is the priority.

To ensure the capacity of XGboost in the enhancement of Pneumonia classification, we also tested our model with Catboost and LightGBM, two alternative gradient boosting algorithms, using the same extracted CNN features.

The CNN-XGboost outperforms both alternatives in terms of accuracy, *precision, F1*-score, highlighting the effectiveness of XGboost in refining the feature representation learned by CNN to better capture Pneumonia features in CXR images.

Despite Catboost is designed to handle categorical data, it showed limited performance in this context with an *accuracy* of 81%, *precision* of 83%, and significantly lower *recall* of 75%, suggesting difficulties in identifying true positive Pneumonia cases. CNN-Catboost was more sensitive to the imbalanced data, it fails precisely in distinguishing between the underrepresented viral pneumonia and bacterial pneumonia cases. Additionally, overfitting was largely noticed during the models training.

On the other hand, LightGBM performed even poorer with the poorest *accuracy* (79%), precision (80%), and significantly lower recall (68%), which indicates an extreme deficiency in correctly classifying pneumonia. The failure of LightGBM is noteworthy considering its architecture for efficiency and speed. The comparative analysis reveals that these alternative gradient boosting classifiers struggle with the high-dimensional feature vectors (52,163 features) extracted from our CNN model when applied to the complex task of distinguishing between normal, viral Pneumonia, and bacterial Pneumonia patterns in CXR images.

The use of Albumentations helped the model to generalize better and likely prevented further *recall* degradation. As shown in *Figure 4.4*, models trained with data augmentation performed better than those without.

In the field of medical diagnosis, training time is an important factor. For real-world application, models that can be rapidly trained or fine-tuned on new data enable more frequent updates with emerging cases. Moreover, faster training time allows for more extensive

hyperparameter tuning and architecture exploration, potentially leading to superior final models.

The XGboost inherent parallel processing capability significantly contributed to the classifier's exceptional speed (7 seconds for training) compared to Catboost (2 min 13s) and LightGBM (57s) as illustrated in table 4.2. This efficiency stems from its column block structure for sparse data and the cache-aware prefetching algorithm that optimizes the memory access pattern, making it extremely efficient for real-time clinical practice where diagnosis in a timely manner is necessary.

Table 4.2 Comparative analysis of the proposed model with various tested model for Pneumonia classification

Trained models	Accuracy	Precision	Recall	F1-score	Training time
CNN	0,85	0,86	0,83	0,85	14 min
CNN-XGboost	0,87	0,89	0,85	0,85	7s
CNN-Catboost	0,81	0,83	0,75	0,75	2 min 13s
CNN-Light GBM	0,79	0,80	0,68	0,68	57 s

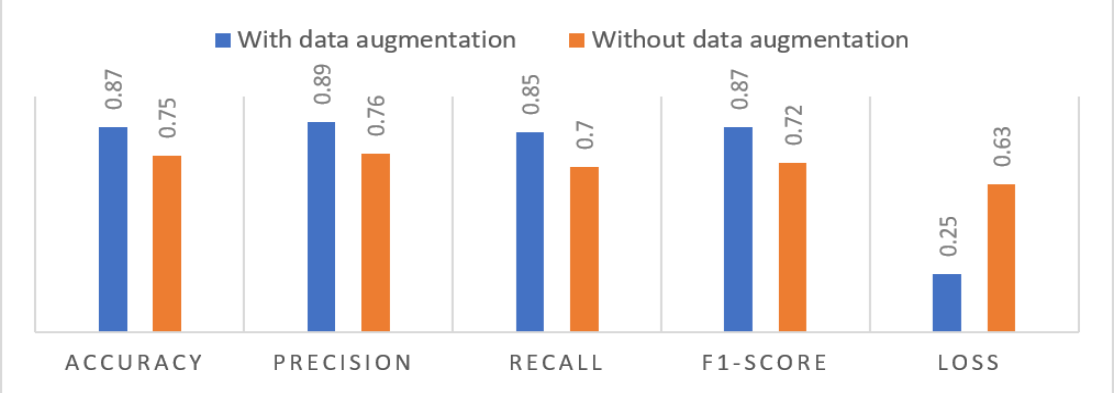


Figure 4.3 Comparison of model's performance in term of accuracy and loss for 3-class classification

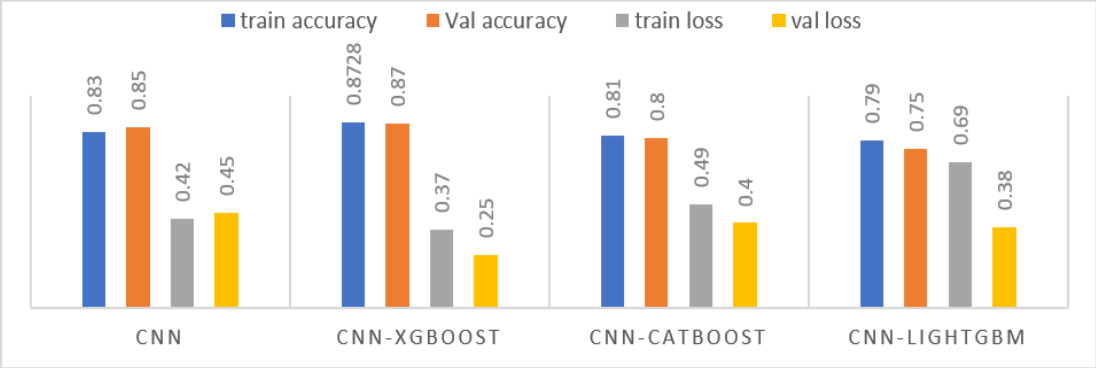


Figure 4.4 Comparison of CNN-XGboost performance with and without data augmentation

The proposed model was also tested for binary classification to only distinguish between normal CXR images and Pneumonia in general, achieving an accuracy of approximately 96% and a balanced F1-score of 95,22%. This result is crucial for clinical perspective, as minimizing false negatives (missed pneumonia cases) is critical for early intervention. The model showed high sensitivity (recall) of 95%, indicating its ability to correctly identify positive pneumonia cases that require medical treatment. Such high sensitivity is particularly beneficial in clinical screening scenarios, where false negatives for pneumonia would delay treatment and potentially have severe health consequences for the patients.

The performance difference between binary (96% accuracy) and multi-class classification (87% accuracy) reflects the challenging distinction between bacterial and viral pneumonia subtypes over predicting pneumonia presence. This aligns with clinical experience, in which even experienced radiologists struggle to distinguish etiologies of pneumonia based on CXR images alone. The significant performance improvement in binary classification indicates that our CNN-XGboost hybrid can be employed as an efficient first-line screen in low-resource settings to recognize probable pneumonia cases for subsequent investigation at the clinic.

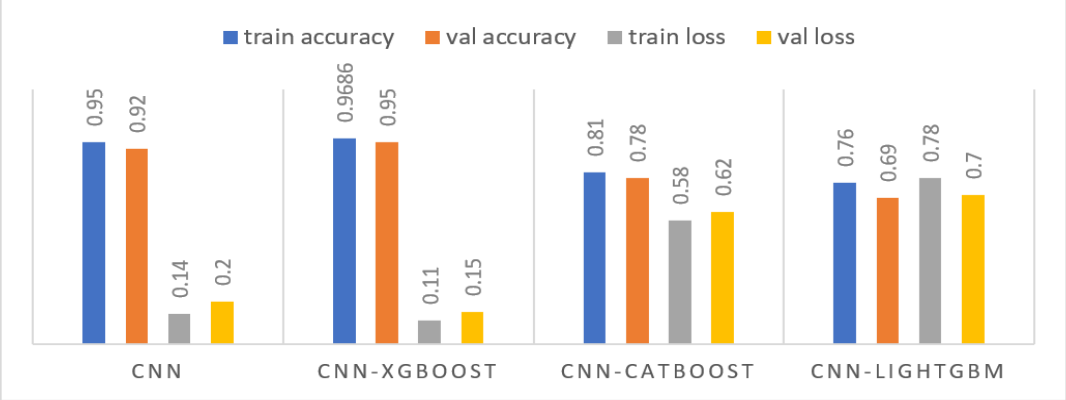


Figure 4.5 Comparison of model's performance in term of accuracy and loss for 2-class classification

4. Experiment 2: A Hybrid CNN-ViT Model for Chest Disease Classification

Continuing within the context of applying deep learning algorithms for chest disease detection, we conducted a second experiment aimed at developing a deep learning framework for efficient Tuberculosis and Pneumonia detection using CXR images.

CNNs are great at capturing local features such as edges, textures, and small patterns, using small receptive fields that slide over the image. However, one of their major limitations is that they struggle with long-range dependencies. This limitation is a critical issue in Tuberculosis diagnosis, where abnormalities like cavitation or diffuse nodules may appear across distant regions of the lungs, requiring a broader understanding. Vision Transformers (ViTs) with their self-attention mechanism deal with this limitation, they are able to capture long-range dependencies ensuring a comprehensive understanding of global contextual information within medical images.

We aimed to benefit from the strength of both models CNN and ViT, by developing an ensemble model that combines both of them for Tuberculosis detection.

4.1 Preprocessing and data preparation

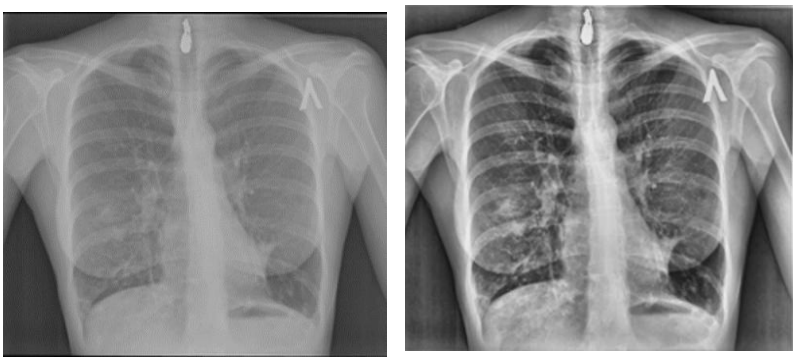
Our task of chest disease classification based- DL using CXR images, begins with a preprocessing pipeline that includes image enhancement, image normalization and data augmentation. In this experiment, we used the Tuberculosis dataset of Kaggle [109] and the children's Pneumonia dataset [71].

The CXR images were firstly resized to 224×224×3, followed by pixels normalization to standardize the input data. To enhance the visualization of fine details in CXR images, we applied Contrast Limited Adaptive Histogram Equalization (CLAHE), using a *clip_limit* of 3,5 and a *tile grid size* of (8×8).

- The *clip_limit* helps to avoid noise over-amplification in relatively homogeneous regions within the CXR image by limiting the contrast enhancement.
- The *tile grid size* divides the image into (8x8) blocks for local histogram equalization.

After image enhancement, we processed with data augmentation to increase the diversity of data by applying a *rotation* of 15°, *width* and *height shift* of 20%, a *zoom range* of 15° and *nearest fill mode* to address the gaps formed at the edges of the image by filling any

newly generated empty pixels with the value of the closest valid pixel. Then, we employed the oversampling technique targeted to the minority class by adding more aggressive augmentation including *rotation* up to 40°, vertical and horizontal flips, vertical and horizontal translation up to 30% and a zoom range of 30%. The application of oversampling helps the model to learn more balanced decision boundaries, and improves its ability to deal with unseen data. Finally, we randomly split the training dataset into: 80% for training and 20% for validation.



(a) **(b)**

Figure 4.6 Example of Tuberculosis. (a) The original image, (b) Image after CLAHE application

4.2 Model architecture

We designed an ensemble model that leverage CNN and ViT for Tuberculosis and Pneumonia detection. The input image of size (224×224×3) was fed into two parallel branches (*Figure 4.7*):

- A fine-tuned ResNet-50 for local features extraction.
- A fine-tuned ViT-b16 for global features extraction.

Both models are enhanced with self-attention mechanisms to ensure an efficient global features recognition while learning from the structured local details. Then, a feature-level-fusion was used to combine the extracted features from the two branches, followed by two dense layers of 510 and 256 neurons respectively. Batch normalization and dropout layers were added to improve the stability and the generalization of the model. Finally, an output layer with sigmoid activation function was added to produce the final binary classification decision (Tuberculosis, normal).

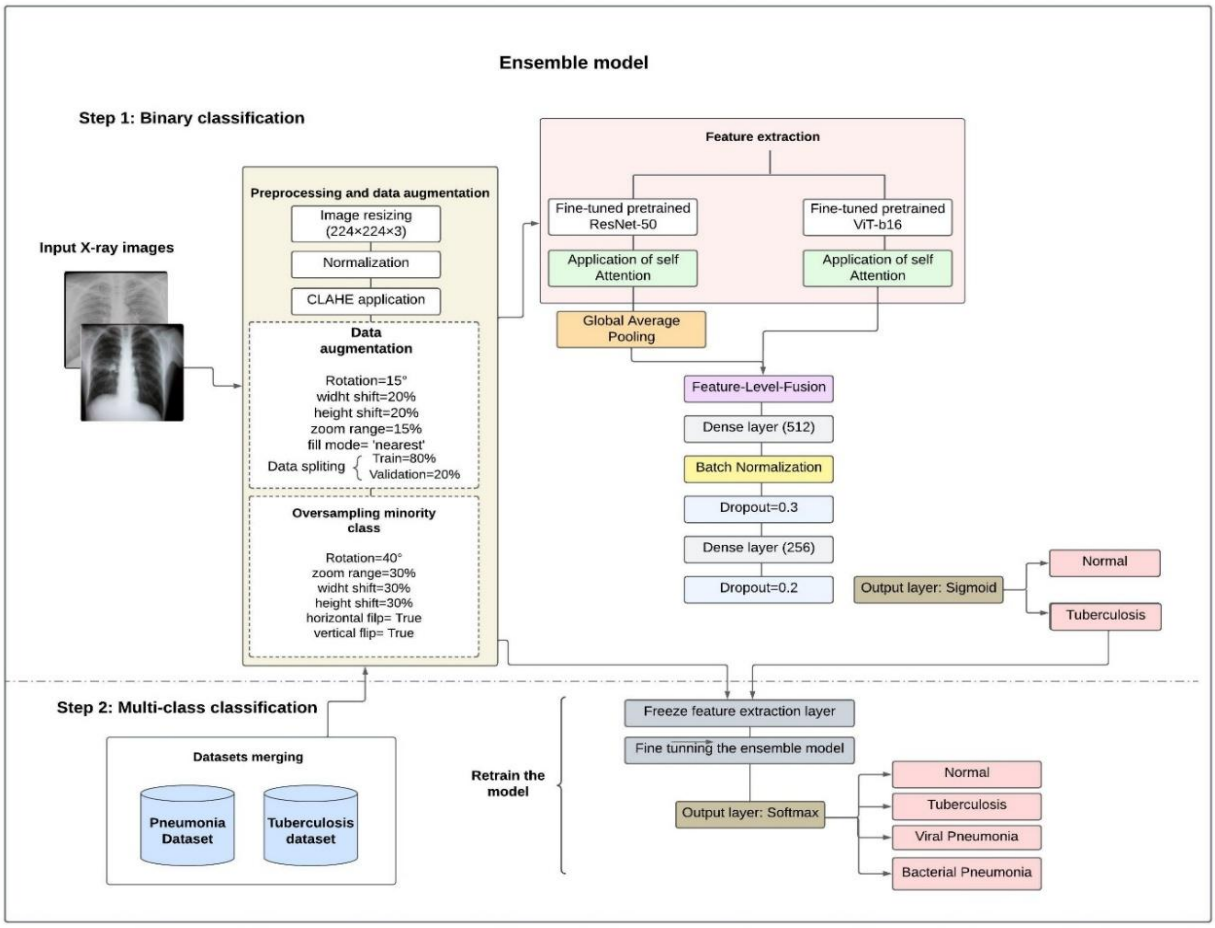


Figure 4.7 The overall ensemble ResNet50-ViTb16 architecture

4.3 Model training

The ensemble ResNet50-ViTb16 model performed classification in two steps: The first one involves binary classification for Tuberculosis detection, while the second step aims to classify CXR images into Normal, Tuberculosis, viral Pneumonia, and bacterial Pneumonia.

A. Binary classification step

For Tuberculosis detection, the model was trained during *30 epochs* with a small *batch size* of 16 to ensure compatibility with available GPU memory. We tested different optimizers such as *Stochastic gradient descent* (SGD), Adaboost, Adam, etc. The use of *Rectified Adam Optimizer* with a learning-rate of 1×10^{-4} helped us to achieve reliable convergence. We also used *binary-cross-entropy* as a loss function and *early stopping* based on validation F1-score to prevent overfitting. For the ViT branch, we trained the model with

different patch sizes such as (8×8), (16×16), and (32×32). The size of (8×8) caused a very slow training and high memory usage, however the size of (32×32) with a low image resolution of 224×224 composes only 49 patches. This small number of patches caused lower performance, which was interpreted as the possibility of losing fine-grained spatial details.

B. Multi-class classification step

For multi-class classification step, we merged the datasets of Tuberculosis and Pneumonia. A *class-weight* was implemented to address the class imbalance issue more effectively. The application of *class-weight* ensures that the minority class receives high penalty in the loss function, helping the model to pay more attention to it.

At this level, we extended the binary classification model to perform multi-class classification across four categories as it is mentioned above.

The same ensemble model, consisting of parallel ResNet-50 and ViT-b16 was fine-tuned, leveraging Transfer learning, with their fused extracted features processed through a dense layer of four neurons for the final classification. The *sigmoid* activation function was replaced by *softmax*, and we used *categorical-cross-entropy* as loss function. We kept the *Rectified Adam Optimizer* with the same learning rate, and increased the number of epochs to 40.

All these experiments detailed above were implemented using *Google Colab pro* with a premium *A100 NVIDIA GPU* and 27.3 GB of memory. The software setup comprised Python environment, including *TensorFlow* and *Keras* libraries.

4.4 Pseudo-code of ResNet50-ViT16 model

This section presents the comprehensive pseudo-code to elucidate the fusion of convolutional and transformer-based approaches. The pseudo code meticulously details the two-stage training protocol, beginning with binary Tuberculosis detection and progressing to multi-class classification across four disease categories.

ALGORITHM 2: Ensemble ResNet50-ViT Model for TB and Pneumonia Detection

INPUT: Chest X-ray images (Normal, TB, Viral Pneumonia, Bacterial Pneumonia)

PHASE 1: Advanced Preprocessing

1. FUNCTION advanced_preprocess(dataset):
2. FOR each image in dataset:
3. Resize to 224×224×3
4. Apply CLAHE enhancement
5. Normalize pixel values
6. Apply augmentation pipeline
7. Apply oversampling for minority classes
8. RETURN enhanced_dataset

PHASE 2: Ensemble Architecture Setup

9. FUNCTION create_ensemble_model():
10. resnet_branch = Fine-tuning ResNet50
11. vit_branch = Fine-tuning ViT_b16
12. fused_features = concatenate([resnet_features, vit_features])
13. dense1 = Dense (512, activation='relu')(fused_features)
14. batch_norm1 = BatchNormalization()(dense1)
15. dropout1 = Dropout (0.3)(batch_norm1)
16. dense2 = Dense (256, activation='relu')(dropout1)
17. batch_norm2 = BatchNormalization ()(dense2)
18. dropout2 = Dropout (0.3)(batch_norm2)
19. RETURN ensemble_model

PHASE 3: Binary Classification Training (TB Detection)

20. FUNCTION train_binary_classification(model, data):
21. output_layer = Dense (1, activation='sigmoid')(model.layers[-1])
22. model.compile(optimizer =RectifiedAdam(lr=1e-4, loss='binary_crossentropy'))
23. SET batch_size = 16, epochs = 30
24. SET early_stopping(monitor='val_f1_score', patience=5)
25. history = model.fit(data, validation_split=0.2, callbacks=[early_stopping])
26. RETURN trained_model, history

PHASE 4: Multi-class Classification Training

27. FUNCTION train_multiclass_classification(model, merged_data):
28. // Calculate class weights for imbalanced data
29. class_weights = compute_class_weight('balanced', classes, labels)
30. output_layer = Dense (4, activation='softmax')(model.layers[-2])
31. model.compile(optimizer = RectifiedAdam(lr=1e-4), loss='categorical_crossentropy')
32. SET batch_size = 16, epochs = 40
33. history = model.fit(merged_data, class_weight=class_weights, validation_split=0.2)
34. RETURN trained_model, history

MAIN EXECUTION:

35. // Step 1: Binary TB Detection
36. tb_data = load_tuberculosis_dataset()
37. processed_tb_data = advanced_preprocess(tb_data)
38. ensemble_model = create_ensemble_model()
39. binary_model = train_binary_classification(ensemble_model, processed_tb_data)
40. // Step 2: Multi-class Classification
41. pneumonia_data = load_pneumonia_dataset()
42. merged_data = merge_datasets(tb_data, pneumonia_data)
43. processed_merged_data = advanced_preprocess(merged_data)
44. multiclass_model = train_multiclass_classification(ensemble_model, processed_merged_data)
45. binary_predictions = binary_model.predict(test_data)
46. multiclass_predictions = multiclass_model.predict(test_data)
47. RETURN binary_predictions, multiclass_predictions, performance evaluation

4.5 Results and discussion

The ensemble model was evaluated using *accuracy, precision, recall, F-score, and Jaccard-score* metrics. For Tuberculosis detection, empirical results show that the model achieved an *accuracy* of 98.97%. The high *recall* value of 99,91% demonstrates the model's generalization ability in detecting the presence of Tuberculosis. Most true positive cases are correctly identified, and very few false negative cases are missed, which is a critical achievement in the medical context. In the other hand, the F1-score of 99,08% indicates the perfect balance between specificity (precision) and sensitivity

(recall). The model not only detects almost Tuberculosis cases, but also produces very few false positive results (High precision of 99.87%).

We also tested the model's performance in term of *Jaccard-score*. The *Jaccard-score* handles the class-imbalance effectively, it helped us to measure the overlap between predicted and actual positive regions, it is stricter than F1-score, as it penalizes both false positive and false negative harshly. Our ensemble model achieved a *Jaccard-score* of 98,38% indicates that the model can be trusted for making accurate decision.

To confirm the robustness of the proposed architecture and the advantage of combining the power of vision transformers with CNNs, we tested different pretrained models such as VGG-16, DenseNet-121, ViT-b16, etc. Table 4.3 summarizes the obtained results for each model, were our ensemble ResNet50-ViTb16 achieved best results.

Table 4.3 Comparative Performance of some pretrained models and the Proposed Ensemble Model for Binary Tuberculosis detection

Pretrained models	Accuracy	Precision	Recall	F1-score	Jaccard score
VGG-16	0,7756	0,7698	0,3309	0,4628	0,3713
DenseNet-121	0,8299	0,8098	0,8355	0,8224	0,7198
Xception	0,8754	0,8960	0,8934	0,8947	0,8063
ResNet-50	0,9115	0,9210	0,9077	0,9114	0,8464
ViT-b16	0,9418	0,9367	0,9549	0,9457	0,8964
ViT-b32	0,8909	0,9023	0,8809	0,8915	0,7890
Ensemble-model	0,9897	0,9987	0,9991	0,9908	0,9838

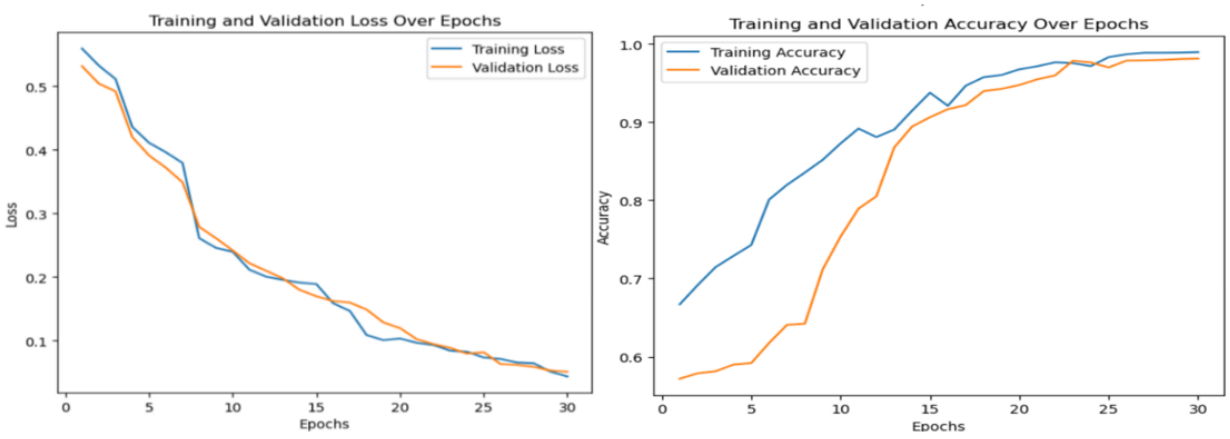


Figure 4.8 Performance of the ensemble ResNet50-ViTb16 for binary classification in term of accuracy and loss

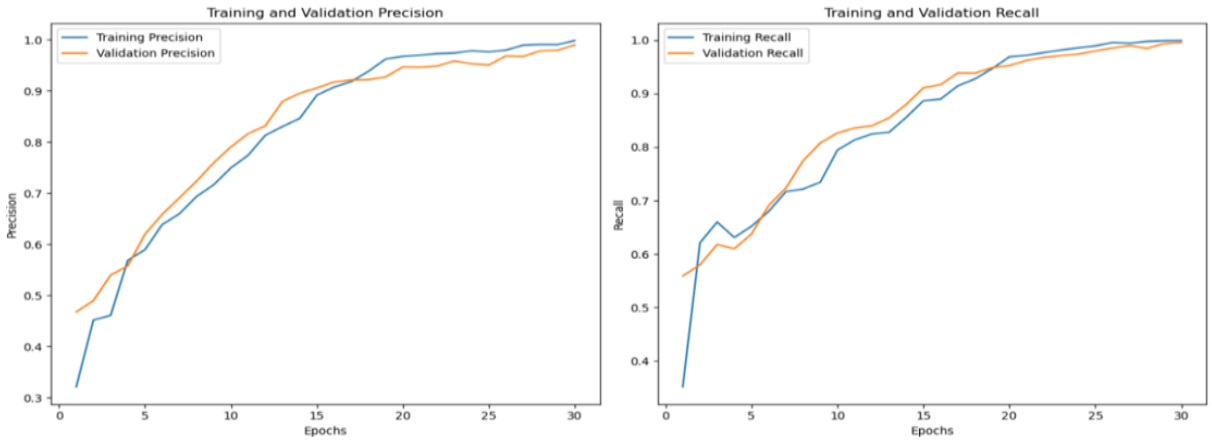


Figure 4.9 Performance of the ensemble ResNet50-ViTb16 for binary classification in term of precision and recall

Passing to the multi-class classification, as it is highlighted in table 4.4 we notice the model's performance have declined somewhat compared to the binary classification. This expected decrease is related to the increase of number of classes, each having overlapping radiological features. The close visual similarity between Tuberculosis and pneumonia augments the complexity of the task.

Despite these challenges, the model still robust with an overall *accuracy* of 96,18% and a weighted *F1-score* of 96,40%. Moreover, a *Jaccard-score* of 90,32% still reflects a strong performance in challenging multi-class classification tasks.

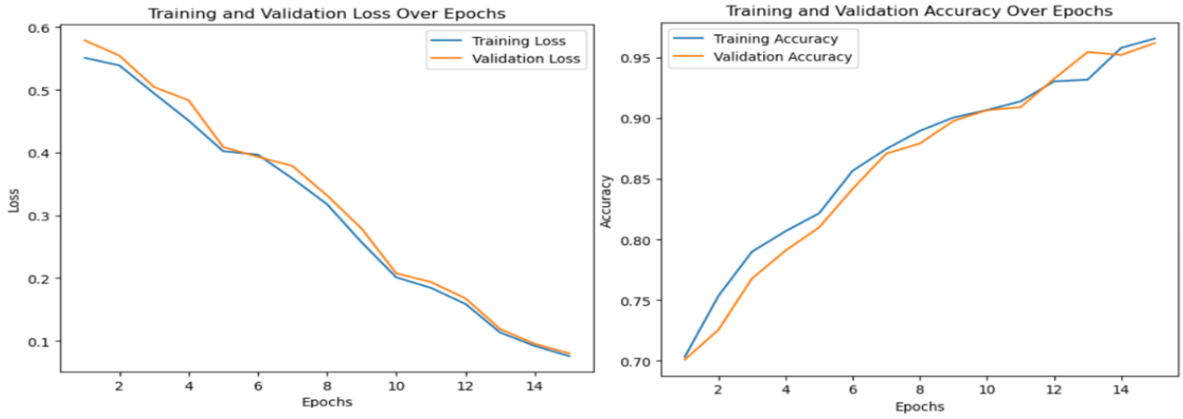


Figure 4.10 Performance of the ensemble ResNet50-ViTb16 for multi-class classification in term of accuracy and loss

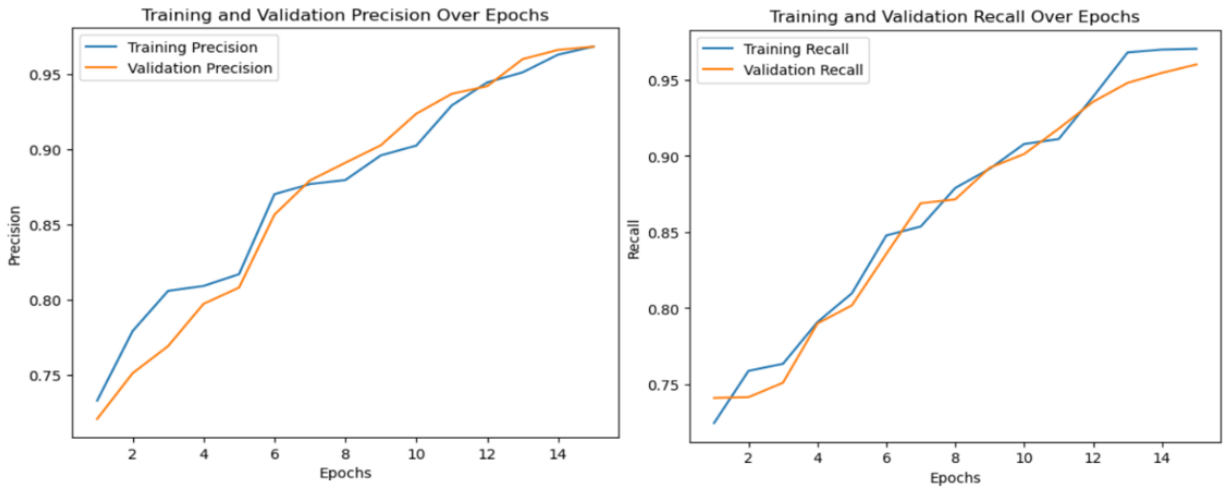


Figure 4.11 Performance of the ensemble ResNet50-ViTb16 for multi-class classification in term of precision and recall

Table 4.4 Comparative Performance of some pretrained models and the Proposed Ensemble Model for multi-class classification

Pretrained models	Accuracy	Precision	Recall	F1-score	Jaccard score
VGG-16	0,5518	0,5657	0,5157	0,5412	0,9756
DenseNet-121	0,9043	0,8945	0,8945	0,8921	0,8064
Xception	0,8514	0,8438	0,8499	0,8468	0,7543
ResNet-50	0,9365	0,9411	0,9387	0,9384	0,8839
ViT-b16	0,9725	0,9710	0,9766	0,9738	0,9245
ViT-b32	0,9156	0,8965	0,9076	0,9020	0,8623
Ensemble-model	0,9618	0,9680	0,9600	0,9640	0,9032

5. Experiment 3: Application of Vision Mamba for Tuberculosis Detection

To address the limitation of traditional CNNs in capturing long-range dependencies, and the quadratic complexity of ViTs. This research aims to deal with these limitations by investigating a new deep learning backbone -Vision Mamba- for efficient Tuberculosis detection using CXR images. Vision Mamba architecture built upon State Space Models (SSMs) introduces a powerful alternative for medical images analysis by combining the ability of global context modeling and the linear computational and memory complexity.

5.1 Preprocessing and data preparation

To ensure the generalizability and robustness of the developed model, we merged the three publicly Tuberculosis datasets mentioned above [19],[158], [109]. This integration offered more diverse samples of normal and Tuberculosis (Figure 4.12), taking into consideration differences in image resolution, acquisition settings, and patient demographics.

Prior to training, all the images are resized to 224×224 pixels and normalized to a standard intensity range. To handle the class imbalance, we performed data augmentation using Random cropping with ratio (3/4, 4/3), Vertical and horizontal flips of 30% to introduce positional diversity in lung field, rotation of 15°, dealing for movements in patient posture during X-ray capture. We also applied shear transformations to simulate the anatomical distortions and anomalies frequently observed in real word imaging scenarios.

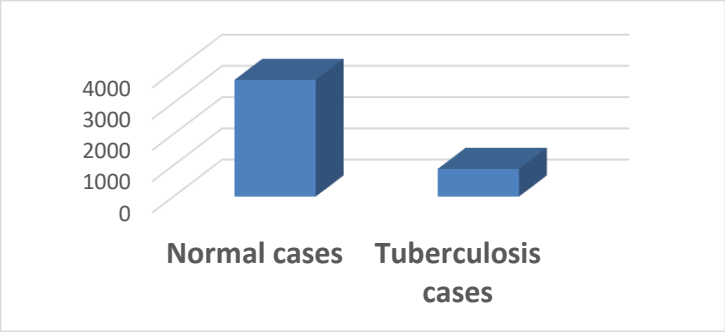


Figure 4.12 The architecture of Vision Mamba modem for Tuberculosis detection

5.2 Model architecture

The fine-tuned Vision Mamba model processed 224×224×3 input images and divided them into non-overlapping 32×32 patches. These patches are normalized, flattened and fed to the Bidirectional Mamba block (Figure 4.13) to implement both forward and backward processing, allowing a richer understanding of the image context. This processing takes information from past and future contexts simultaneously, enabling the capture each of local and global features while maintaining the linear complexity that makes Mamba suitable for resource-constrained environments.

The core component of each of each processing stream (Forward and backward) comprises:

- **Dilated convolutional layer:** Applied to capture local features, while maintaining an expanded receptive field, without augmenting the number of parameters or compromising resolution.
- **Depthwise convolution:** This block reduces the number of learnable parameters and minimizes redundancy and computational load.
- **Relu activation:** Accelerates convergence during training by introducing non-linearity and promoting sparsity in features maps.
- **State Space Model (SSM)-Based Blocks:** These two blocs incorporated by the forward and the backward streams dynamically compute the internal parameters B, C and Δ (detailed in chapter 2) based on the input sequence. This process enables the model to capture long-range features while maintaining linear computational complexity, rendering it suitable for high-resolution CXR analysis.

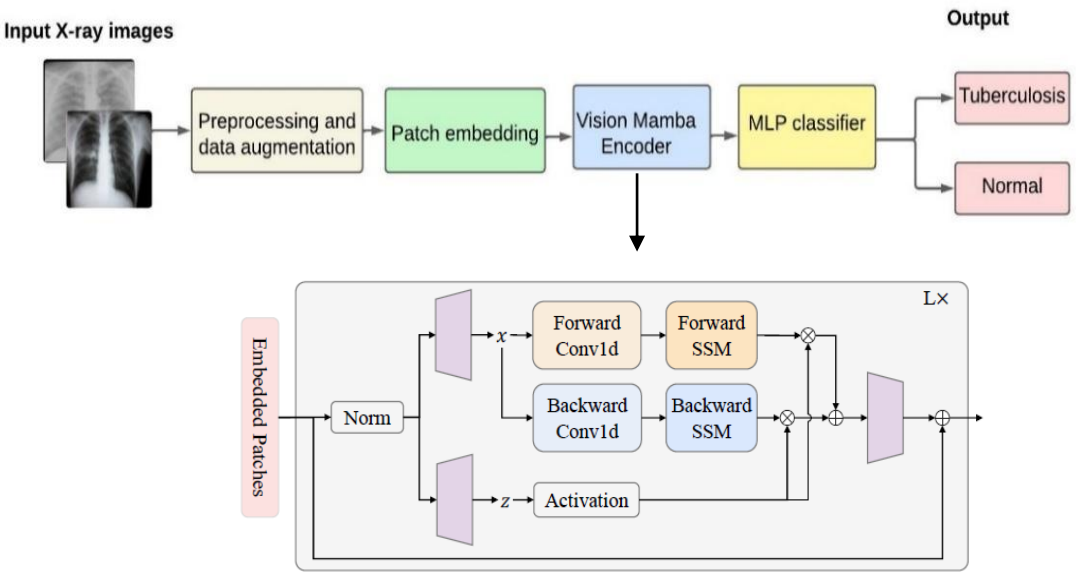


Figure 4.13 The architecture of Vision Mamba modem for Tuberculosis detection

5.3 Model training

To enhance the learning process of the developed model, we carried out a methodical investigation of hyperparameters to efficiently fine-tune the original Vision Mamba model (Vim) [40] for Tuberculosis detection purpose. We used *Rectified Adam* optimizer with a

learning rate of *1e-5*, which help us to stabilize the model's convergence better than the original *Adam*. We applied *ReduceLROnPlateau* to control and adjust the learning rate, and *Binary cross entropy* as loss function.

On the other hand, we increased the patch size from (16×16) to (32×32) due to computational resources reasons. Before moving to the SSM block, we applied residual multi-scale feature fusion by adding parallel dilated convolutions with varying rates and fuse them to improve sensitivity to subtle abnormalities.

To address the significant data imbalance as illustrated in figure 4.12, which can negatively affect the model's performance, we applied class weighting during the training. We assigned a high weight of 4,73 to the minority Tuberculosis class. This calibrated weighting mechanism encourages the learning of discriminative features for both classes, and penalize the misclassifications of the clinically significant Tuberculosis class more severely.

After fine-tuning, the model was trained during *30 epochs* with a batch size of 16. Implementation was conducted using *Pytorch 2.4.1* and *Cuda 12.1* via *Google Colab* Pro along with V100 and A100 GPUs and *83 GB* of RAM.

5.4 Pseudo-code of Vision Mamba for TB detection

The Vision Mamba architecture presents a paradigm shift for medical image analysis by utilizing State Space Models to attain linear computational complexity without compromising on the enhanced long-range dependency modeling capacity. This section describes the algorithmic design underlying the implementation of the optimized Vision Mamba model. The algorithm carefully details the novel bidirectional Mamba processing method, where forward and backward streams concurrently receive context information from the dynamic calculation of parameters associated with matrices B, C, and Δ, thereby facilitating a large spatial comprehension without a loss of computational efficiency.

ALGORITHM 3: Vision Mamba for Efficient TB Detection
INPUT:
Merged TB datasets

#PHASE 1: Multi-dataset Integration and Preprocessing
FUNCTION advanced_preprocessing(data):
1. FOR each image in data:
2. Resize to 224×224×3
3. Normalize to standard intensity range
4. Apply augmentation for class balance:
RETURN preprocessed_data

#PHASE 2: Vision Mamba Architecture Construction
FUNCTION create_vision_mamba_model():
5. patches = create_patches(input_image, patch_size=(32,32))
6. patch_embeddings = normalize_and_flatten(patches)
FUNCTION bidirectional_mamba_block(embeddings):
8. forward_stream:
9. dilated_conv = DilatedConv2D(expanded_receptive_field)
10. depthwise_conv = DepthwiseConv2D(reduce_parameters)
11. relu_activation = ReLU()
12. ssm_forward = StateSpaceModel_Forward()
13. backward_stream:
14. dilated_conv = DilatedConv2D(expanded_receptive_field)
15. depthwise_conv = DepthwiseConv2D(reduce_parameters)
16. relu_activation = ReLU()
17. ssm_backward = StateSpaceModel_Backward()
18. // Dynamic parameter computation
19. B, C, Δ = compute_dynamic_parameters(input_sequence)
20. forward_output = ssm_forward.process(embeddings, B, C, Δ)
21. backward_output = ssm_backward.process(embeddings, B, C, Δ)
22. // Feature fusion
23. fused_features = concatenate([forward_output, backward_output])
24. RETURN fused_features
25. // Multi-scale feature fusion
26. parallel_dilated_convs = [DilatedConv(rate=r) for r in [1,2,4,8]]
27. multi_scale_features = [conv(patch_embeddings) for conv in parallel_dilated_convs]
28. residual_features = add(multi_scale_features)
29. // Process through bidirectional Mamba
30. mamba_output = bidirectional_mamba_block(residual_features)
31. final_output = Dense(1, activation='sigmoid')(mamba_output)
32. RETURN vision_mamba_model

#PHASE 3: Training with Class Weighting
FUNCTION train_vision_mamba(model, data, labels):
33. // Handle class imbalance
34. class_weights = {0: 1.0, 1: 4.73} // Higher weight for TB class
35. model.compile(optimizer=RectifiedAdam(lr=1e-5), loss='binary_crossentropy')

5.5 Results and discussion

The proposed vision Mamba for Tuberculosis detection achieved a notable *accuracy* of 94,17% with a low loss of 0.0325% (*Figure 4.14*) and demonstrated a GPU memory saving of approximately 80%. A key advantage of Vision Mamba-based models is its hardware efficiency, under identical conditions, it consumed only 9.8 GB of GPU memory, significantly better than ViTb-16, which required 49.2 GB and suffered from the bound memory. To validate this ability of Vision Mamba, we gradually increased the images resolution from (224×224×3) to (250×250×3) and (512×512×3). Despite the high computational resources required by larger input size, the model performed consistently, with an overall *accuracy* decreasing by just 1.53%.

In terms of clinical relevance, the model achieved a high *recall* of 95,12%, highlighting its effectiveness in detecting positive Tuberculosis cases. This is crucial in case of contagious

disease, where misinterpretation might have serious health complications. However, the degradation in *precision* (79,75%) compared to *recall*, as well as the noticed imbalance in the confusion matrix (*Figure 4.15*), which shows 209 false positives but only 60 false negatives may be attributed to overlapping radiographic features between Tuberculosis and other non-TB conditions, and the artifacts presents in training data. The resulting *F1-score* of 85,95% is an imbalanced but still improvable performance. Such a score may be acceptable in limited-resource settings where radiologists are scarce. However, enhancing precision further to reduce unnecessary follow-ups increase clinical workflows remains necessary.

The improvement of Vision Mamba compered to CNN models such as VGG-16, DenseNet-121 or ResNet-50 demonstrates its ability to capture long-range dependencies. Moreover, in term of *accuracy*, Vision Mamba achieved comparable results to ViT-b16 with less computational memory and faster inference time. Table 4.5 provides a performance comparison of the different tested DL models.

Table 4.5 Classification performance of the tested DL models for Tuberculosis detection

Pretrained models	*Accuracy*	*Precision*	*Recall*	*F1-score*
VGG-16	0,7756	0,7698	0,3309	0,4628
DenseNet-121	0,8299	0,8098	0,8355	0,8224
Xception	0,8754	0,8960	0,8934	0,8947
ResNet-50	0,9115	0,9210	0,9077	0,9114
ViT-b16	0,9418	0,9367	0,9549	0,9457
ViT-b32	0,8909	0,9023	0,8809	0,8915
MedMamba Pneumonia [163]	0.8990	0.9210	0.8700	0.8860
Vision Mamba	0.9417	0.7975	0.9321	0.8595

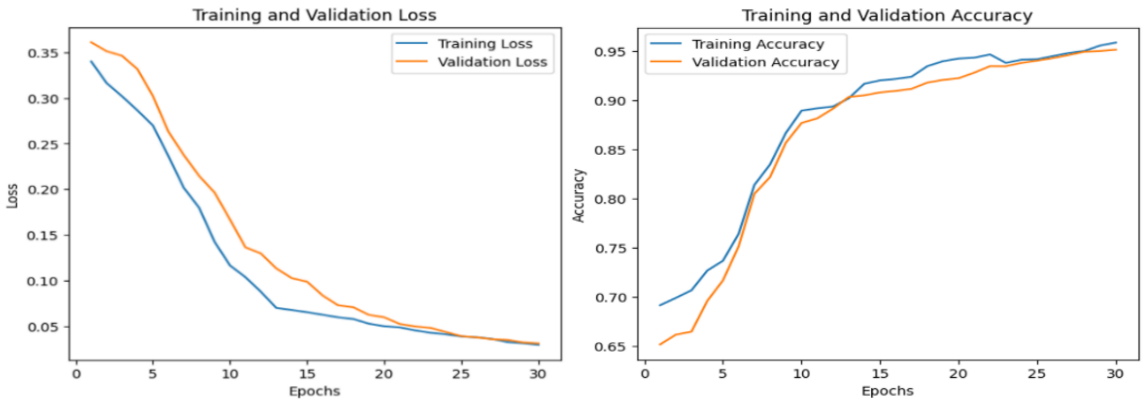
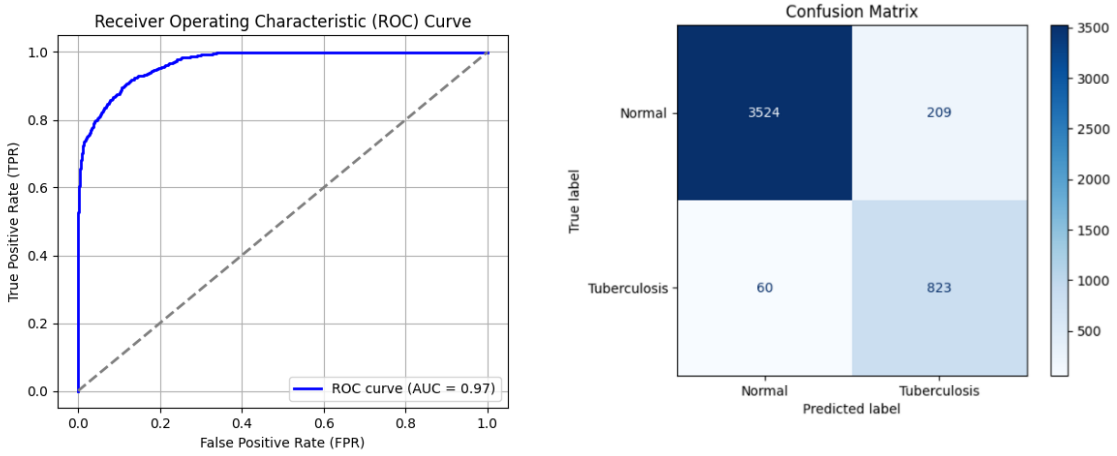


Figure 4.14 The performance of the proposed Vision Mamba for Tuberculosis detection in term of accuracy and loss



(a) **(b)**

Figure 4.15 (a) ROC curve for Tuberculosis detection with the proposed Vision Mamba, (b) Confusion matrix of the proposed Vision Mamba model

Overall, the obtained results demonstrate the feasibility of Vision Mamba as a computationally efficient, clinically, and diagnostically accurate CAD system. Its high rate of recall, good scalability to high resolution images, and hardware-aware architecture make it an excellent candidate for use in real healthcare environments. Nonetheless, the improvement of *precision* and *F1-score* is essential to ensure its adoption and trust in clinical workflows.

6. General discussion

This chapter presented the application of deep learning methodologies for automated diagnosis of chest disease using chest X-ray (CXR) images. Three experimental studies were conducted to demonstrate the effectiveness of these algorithms in accurately interpreting CXR images, with a focus on Pneumonia and Tuberculosis diseases. Each experiment addresses specific limitations identified in existing literature while contributing novel solutions to the challenging problem of automated chest disease diagnosis.

The CNN-XGboost model demonstrated its robustness in Pneumonia classification. The model achieved an accuracy of 87% in the distinction between normal, viral and bacterial Pneumonia, and an accuracy of 96.86%. Reviewing the existing literature highlight that most of studies focused on binary classification (normal Vs Pneumonia), where our model compares favorably in this domain (Table 4.6). For instance, Kermany et al. [71] reported accuracy rates of 92.8% using

traditional CNN architectures. Jiang et al. [164] achieved an accuracy of 94.2% using a variant of fine-tuned pretrained CNNs. However, Rajpurkar et al. [33] in their ChexNet study achieved an F1-score of 76.80 in detecting Pneumonia.

In contrast, relatively few studies have attempted multi-class classification to distinguish between normal, viral, and bacterial Pneumonia. This classification is clinically more meaningful, as it guides treatment decisions, in this case, antibiotics re effective for bacterial but not viral Pneumonia. However, it remains a challenging task, often resulting in lower performance due to inter-class ambiguity and limited annotated data.

For this task, Gu el. [165] achieved a best accuracy of 80.48% using an ensemble model-based CNN and SVM that consists of two parts: lung region segmentation and Pneumonia category classification. Nillmani et al. [166] achieved an accuracy of 92.67% in the classification of normal viral, bacterial Pneumonia using DenseNet201 and VGG16. However, Than et al. [167] reached an accuracy of approximately 86%.

The ensemble CNN-ViT framework, as detailed in the second experimentation, achieved outstanding results, reaching an accuracy of 98.97% for Tuberculosis detection and 96.18% for multi-class chest disease classification. This impressive performance stems from the combination of the two architectures: CNNs effectively extract local spatial features, while ViTs excel at modeling global contextual dependencies. The integration of ViT-b16 with ResNet50 resulting in achieving high performance even in the more challenging multi-class classification task, in contrast to the CNN-XGboost model, whose accuracy degraded significantly when transitioning from binary to multi-class settings. This highlights the superior generalization capability and robustness of the hybrid CNN-ViT architecture in handling complex diagnostic scenarios with overlapping radiological features, such as distinguishing between normal, viral, and bacterial pneumonia cases.

Compared with the state-of-the art, our proposed ResNet50-ViTb16 outperforms different studies that combined the two architectures of CNN and ViT. Xu et al. [168] used a hybrid VGG-16 and coordinate attention for binary classification of Tuberculosis, achieving an accuracy of 92.73%. For the same task, Cohen et al. [99] used ResNet with ViT, reaching an accuracy of 97%. However, Okolo et al. [169] developed a parallel CNN and ViT for multi-class classification of Tuberculosis, Pneumonia, and Covid-19, achieving a recall of around 93%.

During the training of the ensemble ResNet50-ViTb16 models, we encountered significant memory consummation. Despite the use of advanced GPU offered by Google Collab (A100 and V100), we have been extensively limited by the *out-of-memory* (OOM) errors due to the quadratic complexity of ViTs. This limitation motivates us to explore Vision Mamba the new deep learning framework, aiming to overcome the limitations of CNNs in capturing long-range dependencies and the quadratic complexity of ViTs.

The fine-tuned Vision-Mamba model for Tuberculosis detection achieved a notable accuracy of 94.17% with significant GPU memory savings, which is approximately 80% compared to ViTb-16. The model also demonstrated consistent performance with increasing image resolutions and a high recall of 95.12%. The obtained results make Vision Mamba a highly promising candidate for practical deployment in real healthcare environments. While its precision (79.75%) and F1-score (85.95%) were somewhat lower than its recall, potentially due to overlapping radiographic features and artifacts in the training data, these scores are still acceptable with future development and enhancement.

Most of studies that explored Vision Mamba for medical image analysis domain, focused on segmentation tasks. Few studies explored Vision Mamba for classification purposes. Yang et al. [170] proposed BI-Mamba model to detect cardiovascular diseases from CXR images, achieving promising performance compared to ResNet-50 and ViTs. Authors in this paper does not mention the exact obtained results. Authors in [171] introduced Vision Mamba with federated learning to detect Pneumonia and Covid-19 from CXR images. However, Yue et al.[163] developed Mamba-based model named MedMamba for medical image classification. The MedMamba achieved an accuracy of 89.90% for Pneumonia detection. These results demonstrated the robustness of our proposed Vision Mamba model.

Across all the experiments, deep learning models consistently achieved higher performance in binary classification tasks compared to multi-class classification, where the possibility of overlapping features increases (e.g. viral Pneumonia, bacterial Pneumonia, and Tuberculosis) making the task more complex.

On the other hand, all experiments benefited from the application of data augmentation strategies and class reweighting, which mitigated issues related to imbalanced datasets. Particularly, Albumentations and oversampling helped boost generalization performance. However, as

observed with Catboost and LightGBM, some models remained sensitive to feature sparsity and label imbalance, highlighting the need for adaptive preprocessing pipelines.

Table 4.6 Comparison of our proposed model with the state-of-the-art models

Study	Proposed model	Pathology	Classification method	Performance
[71]	CNN-based architecture	Pneumonia	Binary classification	Accuracy=92.8%
[164]				Accuracy =94.2%
[33]				F1-score=76.8%
CNN-XGboost			Binary classification	**Accuracy= 96.86% F1-score= 95.10%**
			Multi-class classification	Accuracy = 87% F1-score=85%
[165]			Multi-class classification	Accuracy=80.48%
[166]				**Accuracy=92.67%**
[167]				Accuracy=86%
[99]	Vision transformer-based architecture	Tuberculosis	Binary classification	Accuracy=97.92%
[169]		Tuberculosis Pneumonia Covid-19	Multi-class classification	Recall=93.5%
[168]		Tuberculosis	Binary classification	Accuracy = 92.73%
[90]		Tuberculosis	Binary classification	Accuracy= 97.95%
Our ResNet50-ViTb16		Tuberculosis	Binary classification	**Accuracy=98.97%**
		Tuberculosis Viral, bacterial Pneumonia	Multi-class classification	**Accuracy=96.18% Recall = 96%**
[163]	Vision Mamba-based architecture	Pneumonia	Binary classification	Accuracy =89.9%
Our fine-tuned Vision Mamba		Tuberculosis		**Accuracy= 94.17%**

7. Limitations and gaps

Despite the advanced outcomes of the proposed deep learning systems for Pneumonia and Tuberculosis, this research encounters different limitations that creates opportunities for future developments.

A. Dataset-related limitations

A primary limitation of this reseach lies in its dependance on publicly available datasets. For instance, the distinction between viral and bacterial Pneumonia was explored using a single pediatric dataset (The only public available one), which may not fully represent the demographic and pathological diversity found in real-world clinical environments, potentially affecting the models' generalizability.

B. Vision Mamba complexity

As detailed in the discussion of each experiment, each proposed model aimed to solve the limitations of its predecessor, be its classification accuracy or resource consumption. Despite the ability of the Vision Mamba in reducing memory consumption while maintaining high accuracy, the complication of its architecture introduced new practical challenge to solve the precision-recall imbalance observed in the final model, or to apply the model for more complex multi-class classification task.

C. Failed GAN-based data augmentation

To address the challenge of limited and imbalanced data, an experiment was conducted using Generative Adversarial Networks (GANs) for data augmentation. We trained a GAN model to synthetize artificial CXR images to expand the underrepresented disease classes. A deep Convolutional GAN (DCGAN) was implemented, consisting of:

- **Generator network:** It imploys a fully convolutional architecture, with an input layer of 100-dimensional random noise vector sampled from a standard normal distribution. The generator projects this vector and then reshapes to a 4×4×1024 tensor. This tensor is upsampled progressively into five transposed convolutional layers with kernel size 4 and stride 2. Each layer is followed by a batch normalization layer and *relu* activation. In the final layer, *tanh* activation is used to generate 128×128 grayscale synthetic CXR images.
- **Discriminator network:** It receives 128×128 grayscale images, and processes them through five convolutional layers to classify the real and generated images. *LeakyReLU* activation with a negative slope of 0.2 was used for all layers. Batch normalization was applied only after the first convolutional layer.

This architecture was fine-tuned iteratively according to the obtained results. The generated samples were not sufficiently realistic to benefit the downstream deep learning models. Since this study focuses on infectious diseases Tuberculosis and Pneumonia, where both of which manifest as white patches, radiographic opacities, and consolidations, preserving pathological features is critical for training data. However, the generated images by the proposed DCGAN were frequently blurry and lacked the structural clarity necessary to

represent these critical features. As a result, training the proposed models with these images introduced confusion, ultimately degrading the models performance.

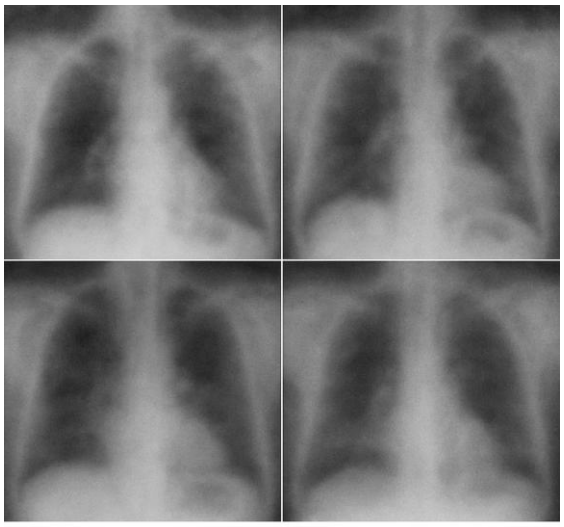


Figure 4.16 Synthetic blury Pneumonia CXR images generated by DGAN

8. Conclusion

This chapter has presented the core scientific contributions of this thesis, showcasing a comprehensive investigation of deep learning approaches for automated chest disease detection, specifically targeting Pneumonia and Tuberculosis classification from chest X-ray images. The contribution addresses the key challenges of this task including disease overlapping and mimic radiological features, limited annotated datasets, data imbalance, and computational resources constraints.

The target objective was to develop efficient CAD systems for Pneumonia and Tuberculosis detection to enhance diagnosis accuracy, while reducing time. Three experiments were conducted for this goal including:

- An improved hybrid CNN-XGboost for Pneumonia detection.
- An ensemble ResNet-50 and ViT-b16 for Tuberculosis and Pneumonia classification.
- A fine-tuned Vision Mamba model for Tuberculosis detection.

CNN-based architectures show promising results in binary classification, highlighting the model's clinical utility. However, the combination of CNN with vision transformer-based architecture enhanced the performance of multi-class classification tasks. On the other hand, the computational

efficiency of the proposed Vision Mamba, makes them suitable for resource-constrained clinical environments. Moreover, techniques such as data augmentation, class reweighting, and architectural fusion were shown to be critical in enhancing model generalization and robustness.

Collectively, these experiments demonstrate that deep learning can offer accurate, efficient, and scalable solutions to chest disease diagnosis, providing a practical and high-performing solutions for CAD systems to improve patient outcomes through early and accurate diagnosis of chest disease.

Conclusion & perspectives

This thesis entitled “X-ray Imaging System for chest diseases diagnostics” aims to address the limitations related to chest disease diagnosis, especially in under-resourced environments. Although chest X-ray images are the most used tool to detect these diseases, their accurate interpretation poses significant challenge due to the mimicry and overlapping radiological patterns and the shortage of trained radiologists.

The main goal was to deal with persistent challenges of diagnostic accuracy, efficiency, and accessibility that define the manual analysis of CXR images in modern healthcare. The research focused on building efficient deep learning systems capable of detecting Pneumonia and Tuberculosis using chest x-ray images. The work was hierarchically structured, starting with a CNN-XGboost to identify Pneumonia and differentiate between viral and bacterial cases. This differentiation is crucial for treatment decision, which is neglected in most of state-of-the-art studies, where the majority focused only in the classification of “normal” Vs “Pneumonia” images. The CNN-XGboost model achieved a commendable accuracy of 96.8% in binary Pneumonia detection. However, it slightly struggled with complex multi-class classification. To address this failing, an ensemble ResNet50-ViTb16 model was proposed by leveraging the power CNNs and Vision transformers. In this case the model excels in both binary and multi-class classification. It achieved an outstanding accuracy of 98.97% for Tuberculosis detection and maintained a robust 96.18% accuracy in the challenging four-class (Normal, Viral Pneumonia, Bacterial Pneumonia, and Tuberculosis) classification task.

While the ensemble ResNet50-ViTb16 model showed exceptional performance, its training and optimization were hampered by the significant computational and memory demands inherent to the quadratic complexity of the self-attention mechanism. This issue was tackled in the third contribution by designing a fine-tuned Vision Mamba-based model, which demonstrated its efficiency in Tuberculosis detection. The model achieved a notable accuracy of 94.17% while drastically reducing GPU memory consumption by approximately 80% compared to the ViT-b16 model. Its high recall and scalability make it a prime candidate for deployment in resource-constrained clinical settings.

In all experiments, data augmentation, and class weighting enhanced the model’s performance and generalization.

Conclusion & perspectives

In summary, this thesis demonstrates the ability of deep learning in addressing clinical limitations. Findings emphasize the value of combining architectural strengths, and confirm that AI-powered systems can serve as powerful tools to support clinicians, improve diagnostic workflows, and ultimately enhance patient outcomes in the global fight against respiratory diseases.

While the developed models have revealed promising results in detecting two overlapping pathologies (Pneumonia and Tuberculosis), our work does not end here. Several opportunities avenues for feature research including:

- Enhancing model interpretability and trust.
- Collaborate with clinicians' experts to validate the obtained results. his will help assess their real-world performance.
- Exploring the fusion of information from multiple modalities, such as integrating CXR data with clinical symptoms.
- Additionally, our current efforts are focused on completing the development of a real-time detection system based on YOLOv11 to accurately identify and localize pathology regions within CXR images. This work is in progress and expected to complement the classification models developed in this thesis.

Scientific contributions

International scientific journals

- Yousra HEDHOUD, Tahar MEKHAZNIA, Mohamed AMROUNE, “An improvement of CNN-XGboost model for pneumonia disease classification”, *Polish Journal of Radiology*, vol. 88, no. 1, pp. 483–493, 2023, doi: 10.5114/pjr.2023.132533.
- Yousra HEDHOUD et al., From Binary to Multi-Class Classification: A Two-Step Hybrid CNN-ViT Model for Chest Disease Classification Based on X-Ray Images,” *Diagnostics*, vol. 14, no. 23, Dec. 2024, doi: 10.3390/diagnostics14232754.

International communications

- Yousra HEDHOUD, Tahar MEKHAZNIA, Mohamed AMROUNE, “Vision Mamba for efficient Tuberculosis Detection based on Chest X-Rays: A comparative study with CNN and Vision transformer”, 7th International Conference on Pattern Analysis and Intelligent Systems (PAIS), Laghouat, Algeria, 2025.

National communications

- Yousra HEDHOUD, Tahar MEKHAZNIA, Mohamed AMROUNE, “Deep learning and vision transformers for chest diseases prediction and classification”, Week of Artificial Intelligence Trends (WAIT’2023), Tebessa, 2023.
- Yousra HEDHOUD, Tahar MEKHAZNIA, Mohamed AMROUNE, “The application of AI in chest diseases diagnosis using X-ray images”, The Second National Conference on the Digital Revolution in Healthcare: Applications and Impacts of Artificial Intelligence, Tebessa, 2024.

References

[1] "Radiology Workforce Shortage and Growing Demand Something Has to Give." Accessed: Jun. 11, 2025. [Online]. Available: https://www.acr.org/Clinical-Resources/Publications-and-Research/ACR-Bulletin/Radiology-Workforce-Shortage-and-Growing-Demand-Something-Has-to-Give

[2] "4 key trends in radiology at RSNA 2023." Accessed: Jun. 11, 2025. [Online]. Available: https://radiologybusiness.com/topics/professional-associations/radiology-associations/radiological-society-north-america-rsna/4-key-trends-radiology-rsna-2023

[3] "IHME, Global Burden of Disease (2024) – with minor processing by Our World in Data. ' 70+ year-olds' [dataset]. IHME, Global Burden of Disease, 'Global Burden of Disease - Deaths and DALYs' [original data].".

[4] "World Pneumonia Day - Global Initiative for Asthma - GINA." Accessed: Feb. 21, 2025. [Online]. Available: https://ginasthma.org/world-pneumonia-day/

[5] "Pneumonia - Our World in Data." Accessed: Feb. 25, 2025. [Online]. Available: https://ourworldindata.org/pneumonia

[6] K. Floyd, P. Glaziou, R. M. G. J. Houben, T. Sumner, R. G. White, and M. Raviglione, "Global tuberculosis targets and milestones set for 2016-2035: Definition and rationale," *International Journal of Tuberculosis and Lung Disease*, vol. 22, no. 7, pp. 723–730, Jul. 2018, doi: 10.5588/IJTLD.17.0835.

[7] A. , C. C. , N. M. S. et al. Torres, "Pneumonia," *Nat Rev Dis Primers*, vol. 7, no. 25, 2021.

[8] "Tuberculosis and Mycobacterial Pneumonia - Cancer Therapy Advisor." Accessed: May 19, 2025. [Online]. Available: https://www.cancertherapyadvisor.com/home/decision-support-in-medicine/hospital-medicine/tuberculosis-and-mycobacterial-pneumonia/

[9] R. Orofino-Costa *et al.*, "Pulmonary cavitation and skin lesions mimicking tuberculosis in a HIV negative patient caused by Sporothrix brasiliensis," *Med Mycol Case Rep*, vol. 2, no. 1, pp. 65–71, 2013, doi: 10.1016/j.mmcr.2013.02.004.

[10] M. Di Serafino *et al.*, "Point-of-Care Lung Ultrasound in the Intensive Care Unit—The Dark Side of Radiology: Where Do We Stand?," Nov. 01, 2023, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/jpm13111541.

[11] "Tips and techniques for decubitus and oblique chest x-rays | AuntMinnie." Accessed: Mar. 03, 2025. [Online]. Available: https://www.auntminnie.com/radiology-education/article/15559318/tips-and-techniques-for-decubitus-and-oblique-chest-xrays

[12] A. Bustos, A. Pertusa, J.-M. Salinas, and M. de la Iglesia-Vayá, "PadChest: A large chest x-ray image dataset with multi-label annotated reports," Jan. 2019, doi: 10.1016/j.media.2020.101797.

[13] S. Nazir, D. M. Dickson, and M. U. Akram, "Survey of explainable artificial intelligence techniques for biomedical imaging with deep neural networks," Apr. 01, 2023, *Elsevier Ltd*. doi: 10.1016/j.compbiomed.2023.106668.

[14] I. Pan, A. Cadrin-Chênevert, and P. M. Cheng, "Tackling the radiological society of North America pneumonia detection challenge," 2019, *American Roentgen Ray Society*. doi: 10.2214/AJR.19.21512.

[15] G. Shih *et al.*, "Augmenting the national institutes of health chest radiograph dataset with expert annotations of possible pneumonia," Jan. 01, 2019, *Radiological Society of North America Inc.* doi: 10.1148/ryai.2019180041.

[16] "Detect Pneumonia (Spring 2022) | Kaggle." Accessed: May 19, 2023. [Online]. Available: https://www.kaggle.com/competitions/detect-pneumonia-spring-2022/leaderboard
[17] "Chest X-rays (Indiana University)." Accessed: Oct. 15, 2024. [Online]. Available: https://www.kaggle.com/datasets/raddar/chest-xrays-indiana-university
[18] "TBX 11." Accessed: Jan. 06, 2025. [Online]. Available: https://www.kaggle.com/datasets/usmanshams/tbx-11
[19] "Tuberculosis Chest X-rays (Montgomery)." Accessed: Oct. 16, 2024. [Online]. Available: https://www.kaggle.com/datasets/raddar/tuberculosis-chest-xrays-montgomery
[20] S. Jaeger, S. Candemir, S. Antani, Y.-X. J. Wáng, P.-X. Lu, and G. Thoma, "Two public chest X-ray datasets for computer-aided screening of pulmonary diseases.," *Quant Imaging Med Surg*, vol. 4, no. 6, pp. 475–7, Dec. 2014, doi: 10.3978/j.issn.2223-4292.2014.11.20.
[21] Y. Wu, H. Gunraj, C. A. Tai, and A. Wong, "COVIDx CXR-4: An Expanded Multi-Institutional Open-Source Benchmark Dataset for Chest X-ray Image-Based Computer-Aided COVID-19 Diagnostics," Nov. 2023, [Online]. Available: http://arxiv.org/abs/2311.17677
[22] M. Pavlova, T. Tuinstra, H. Aboutalebi, A. Zhao, H. Gunraj, and A. Wong, "COVIDx CXR-3: A Large-Scale, Open-Source Benchmark Dataset of Chest X-ray Images for Computer-Aided COVID-19 Diagnostics," Jun. 2022, [Online]. Available: http://arxiv.org/abs/2206.03671
[23] J. Irvin *et al.*, "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," *33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*, pp. 590–597, 2019, doi: 10.1609/aaai.v33i01.3301590.
[24] X. Wang, "NIH Chest X-ray Dataset of 14 Common Thorax Disease Categories," *In Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2097–2106, 2017.
[25] H. Q. Nguyen *et al.*, "VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations," Dec. 2020, [Online]. Available: http://arxiv.org/abs/2012.15029
[26] H. H. Pham, N. H. Nguyen, T. T. Tran, T. N. M. Nguyen, and H. Q. Nguyen, "PediCXR: An open, large-scale chest radiograph dataset for interpretation of common thoracic diseases in children," *Sci Data*, vol. 10, no. 1, Dec. 2023, doi: 10.1038/s41597-023-02102-5.
[27] S. Shah, H. Mehta, and P. Sonawane, "Pneumonia Detection Using Convolutional Neural Networks," no. Icssit, pp. 933–939, 2020.
[28] A. Nasiri-Sarvi, M. S. Hosseini, and H. Rivaz, "Vision Mamba for Classification of Breast Ultrasound Images," Jul. 2024, [Online]. Available: http://arxiv.org/abs/2407.03552
[29] Q. H. W. W. L. L. Guangju Li, "Selective and multi-scale fusion Mamba for medical image segmentation," *Expert Syst Appl*, vol. 261, no. 125518, ISSN 0957-4174, 2025.
[30] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions," Aug. 01, 2017, *Springer New York LLC*. doi: 10.1007/s10278-017-9983-4.
[31] P. T. Chen *et al.*, "Pancreatic Cancer Detection on CT Scans with Deep Learning: A Nationwide Population-based Study," *Radiology*, vol. 306, no. 1, pp. 172–182, Jan. 2023, doi: 10.1148/radiol.220152.
[32] L. Brunese, F. Mercaldo, A. Reginelli, and A. Santone, "Explainable Deep Learning for Pulmonary Disease and Coronavirus COVID-19 Detection from X-rays," *Comput Methods Programs Biomed*, vol. 196, Nov. 2020, doi: 10.1016/j.cmpb.2020.105608.
[33] P. Rajpurkar *et al.*, "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning," pp. 3–9, 2017, [Online]. Available: http://arxiv.org/abs/1711.05225
[34] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," Jun. 2016, [Online]. Available: http://arxiv.org/abs/1606.04797
[35] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," Dec. 2014, [Online]. Available: http://arxiv.org/abs/1412.6980

[36] G. Hinton, N. Srivastava, and K. Swersky, “Neural Networks for Machine Learning Lecture 6a Overview of mini--batch gradient descent.”
[37] P. Beja-Battais, “Overview of AdaBoost : Reconciling its views to better understand its dynamics,” Oct. 2023, [Online]. Available: http://arxiv.org/abs/2310.18323
[38] Y. Lecun, L. Eon Bottou, Y. Bengio, and P. H. Abstract|, “Gradient-Based Learning Applied to Document Recognition.”
[39] A. Dosovitskiy *et al.*, “an Image Is Worth 16X16 Words: Transformers for Image Recognition At Scale,” *ICLR 2021 - 9th International Conference on Learning Representations*, 2021.
[40] L. Zhu, B. Liao, Q. Zhang, X. Wang, W. Liu, and X. Wang, “Vision Mamba: Efficient Visual Representation Learning with Bidirectional State Space Model,” Jan. 2024, [Online]. Available: http://arxiv.org/abs/2401.09417
[41] O. Russakovsky *et al.*, “ImageNet Large Scale Visual Recognition Challenge,” Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.0575
[42] M. H. A. K. M. M. T. P. N. S. M.A. Ganaie, “Ensemble deep learning: A review,” *Eng Appl Artif Intell*, vol. 115, no. 105151, 2022.
[43] “Bagging, Boosting & Stacking Made Simple [3 How To Tutorials].” Accessed: Mar. 19, 2025. [Online]. Available: https://spotintelligence.com/2024/03/18/bagging-boosting-stacking/
[44] I. Goodfellow *et al.*, “Generative adversarial networks,” *Commun ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.
[45] M. Hami and M. Jamebozorg, “ASSESSING THE IMPACT OF CNN AUTO ENCODER-BASED IMAGE DENOISING ON IMAGE CLASSIFICATION TASKS.”
[46] Y. Jo, S. Young Chun, J. Choi, and S. Korea, “Rethinking Deep Image Prior for Denoising.” [Online]. Available: https://github.com/gistvision/DIP-denosing
[47] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” Oct. 2016, doi: 10.1007/s11263-019-01228-7.
[48] H. Wang *et al.*, “Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks,” Oct. 2019, [Online]. Available: http://arxiv.org/abs/1910.01279
[49] D. Omeiza, S. Speakman, C. Cintas, and K. Weldermariam, “Smooth Grad-CAM++: An Enhanced Inference Level Visualization Technique for Deep Convolutional Neural Network Models,” Aug. 2019, [Online]. Available: http://arxiv.org/abs/1908.01224
[50] M. Nahiduzzaman, M. R. Islam, and R. Hassan, “ChestX-Ray6: Prediction of multiple diseases including COVID-19 from chest X-ray images using convolutional neural network[Formula presented],” *Expert Syst Appl*, vol. 211, Jan. 2023, doi: 10.1016/j.eswa.2022.118576.
[51] T. Sanida and M. Dasygenis, “A novel lightweight CNN for chest X-ray-based lung disease identification on heterogeneous embedded system,” *Applied Intelligence*, vol. 54, no. 6, pp. 4756–4780, Mar. 2024, doi: 10.1007/s10489-024-05420-2.
[52] B. U. Maheswari *et al.*, “Explainable deep-neural-network supported scheme for tuberculosis detection from chest radiographs,” *BMC Med Imaging*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12880-024-01202-x.
[53] I. Naskinova, “On Convolutional Neural Networks for Chest X-ray Classification,” in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1757-899X/1031/1/012075.
[54] N. Alrefai and O. Ibrahim, “Deep learning for COVID-19 diagnosis based on chest X-ray images,” *International Journal of Electrical and Computer Engineering*, vol. 11, no. 5, pp. 4531–4541, Oct. 2021, doi: 10.11591/ijece.v11i5.pp4531-4541.
[55] V. Kakani *et al.*, “Post-COVID Chest Disease Monitoring Using Self Adaptive Convolutional Neural Network Austin Journal of Pulmonary and Respiratory Medicine,” 2023. [Online]. Available: www.austinpublishinggroup.com

[56] Q. Li, “Convolutional Neural Networks for Pneumonia Diagnosis Based on Chest X-Ray Images,” in *022 International Conference on Big Data, Information and Computer Network (BDICN)*, Sanya, China: IEEE, 2022, pp. 717–720.

[57] I. Naskinova, “On Convolutional Neural Networks for Chest X-ray Classification,” in *IOP Conference Series: Materials Science and Engineering*, IOP Publishing Ltd, Feb. 2021. doi: 10.1088/1757-899X/1031/1/012075.

[58] S. Guefrechi, M. Ben Jabra, A. Ammar, A. Koubaa, and H. Hamam, “Deep learning based detection of COVID-19 from chest X-ray images,” *Multimed Tools Appl*, vol. 80, no. 21–23, pp. 31803–31820, Sep. 2021, doi: 10.1007/s11042-021-11192-5.

[59] K. Kansal, T. B. Chandra, and A. Singh, “ResNet-50 vs. EfficientNet-B0: Multi-Centric Classification of Various Lung Abnormalities Using Deep Learning ‘session id: ICMLDsE.004,’” in *Procedia Computer Science*, Elsevier B.V., 2024, pp. 70–80. doi: 10.1016/j.procs.2024.04.007.

[60] A. U. Ibrahim, M. Ozsoz, S. Serte, F. Al-Turjman, and P. S. Yakoi, “Pneumonia Classification Using Deep Learning from Chest X-ray Images During COVID-19,” *Cognit Comput*, vol. 16, no. 4, pp. 1589–1601, Jul. 2024, doi: 10.1007/s12559-020-09787-5.

[61] K. Kalaiselvi and M. Kasthuri, “Tuning VGG19 hyperparameters for improved pneumonia classification,” *The Scientific Temper*, vol. 15, no. 02, pp. 2231–2237, Jun. 2024, doi: 10.58414/scientifictemper.2024.15.2.36.

[62] G. M. M. Alshmrani, Q. Ni, R. Jiang, H. Pervaiz, and N. M. Elshennawy, “A deep learning architecture for multi-class lung diseases classification using chest X-ray (CXR) images,” *Alexandria Engineering Journal*, vol. 64, pp. 923–935, Feb. 2023, doi: 10.1016/j.aej.2022.10.053.

[63] F. J. M. Shamrat, S. Azam, A. Karim, K. Ahmed, F. M. Bui, and F. De Boer, “High-precision multiclass classification of lung disease through customized MobileNetV2 from chest X-ray images,” *Comput Biol Med*, vol. 155, Mar. 2023, doi: 10.1016/j.compbiomed.2023.106646.

[64] V. Ravi, V. Acharya, and M. Alazab, “A multichannel EfficientNet deep learning-based stacking ensemble approach for lung disease detection using chest X-ray images,” *Cluster Comput*, vol. 26, no. 2, pp. 1181–1203, Apr. 2023, doi: 10.1007/s10586-022-03664-6.

[65] S. Asif, M. Zhao, F. Tang, and Y. Zhu, “LWSE: a lightweight stacked ensemble model for accurate detection of multiple chest infectious diseases including COVID-19,” *Multimed Tools Appl*, vol. 83, no. 8, pp. 23967–24003, Aug. 2024, doi: 10.1007/s11042-023-16432-4.

[66] Q. S. Hamad, H. Samma, and S. A. Suandi, “Feature selection of pre-trained shallow CNN using the QLESCA optimizer: COVID-19 detection as a case study”, doi: 10.1007/s10489-022-04446-8/Published.

[67] I. Sirazitdinov, M. Kholiavchenko, T. Mustafaev, Y. Yixuan, R. Kuleev, and B. Ibragimov, “Deep neural network ensemble for pneumonia localization from a large-scale chest x-ray database,” *Computers and Electrical Engineering*, vol. 78, pp. 388–399, Sep. 2019, doi: 10.1016/j.compeleceng.2019.08.004.

[68] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal Loss for Dense Object Detection,” Aug. 2017, [Online]. Available: http://arxiv.org/abs/1708.02002

[69] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” Mar. 2017, [Online]. Available: http://arxiv.org/abs/1703.06870

[70] A. T. Sahlol, M. A. Elaziz, A. T. Jamal, R. Damaševičius, and O. F. Hassan, “A novel method for detection of tuberculosis in chest radiographs using artificial ecosystem-based optimisation of deep neural network features,” *Symmetry (Basel)*, vol. 12, no. 7, Jul. 2020, doi: 10.3390/sym12071146.

[71] D. S. Kermany *et al.*, “Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning,” *Cell*, vol. 172, no. 5, pp. 1122-1131.e9, 2018, doi: 10.1016/j.cell.2018.02.010.

[72] M. F. Hashmi, S. Katiyar, A. G. Keskar, N. D. Bokde, and Z. W. Geem, “Efficient pneumonia detection in chest xray images using deep transfer learning,” *Diagnostics*, vol. 10, no. 6, Jun. 2020, doi: 10.3390/diagnostics10060417.

[73] M. E. H. Chowdhury *et al.*, “Can AI help in screening Viral and COVID-19 pneumonia ?,” 2020.

[74] T. Chen *et al.*, "A vision transformer machine learning model for COVID-19 diagnosis using chest X-ray images," *Healthcare Analytics*, vol. 5, Jun. 2024, doi: 10.1016/j.health.2024.100332.
[75] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," May 2019, [Online]. Available: http://arxiv.org/abs/1905.11946
[76] H. Cai, J. Li, M. Hu, C. Gan, and S. Han, "EfficientViT: Multi-Scale Linear Attention for High-Resolution Dense Prediction," May 2022, [Online]. Available: http://arxiv.org/abs/2205.14756
[77] O. Uparkar, J. Bharti, R. K. Pateriya, R. K. Gupta, and A. Sharma, "Vision Transformer Outperforms Deep Convolutional Neural Network-based Model in Classifying X-ray Images," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 2338–2349. doi: 10.1016/j.procs.2023.01.209.
[78] S. Bharati, P. Podder, and M. R. H. Mondal, "Hybrid deep learning for detecting lung diseases from X-ray images," *Inform Med Unlocked*, vol. 20, Jan. 2020, doi: 10.1016/j.imu.2020.100391.
[79] S. Taslimi, S. Taslimi, N. Fathi, M. Salehi, and M. H. Rohban, "SwinCheX: Multi-label classification on chest X-ray images with transformers," pp. 1–10, 2022, [Online]. Available: http://arxiv.org/abs/2206.04246
[80] Z. Liu *et al.*, "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows".
[81] M. Chetoui and M. A. Akhloufi, "Explainable Vision Transformers and Radiomics for COVID-19 Detection in Chest X-rays," *J Clin Med*, vol. 11, no. 11, 2022, doi: 10.3390/jcm11113013.
[82] "SIIM-FISABIO-RSNA COVID-19 Detection | Kaggle." Accessed: Oct. 05, 2023. [Online]. Available: https://www.kaggle.com/c/siim-covid19-detection
[83] K. S. Krishnan and K. S. Krishnan, "Vision Transformer based COVID-19 Detection using Chest X-rays," *Proceedings of IEEE International Conference on Signal Processing,Computing and Control*, vol. 2021-Octob, pp. 644–648, 2021, doi: 10.1109/ISPCC53510.2021.9609375.
[84] C. Liu and Q. Yin, "Automatic diagnosis of COVID-19 using a tailored transformer-like network," *J Phys Conf Ser*, vol. 2010, no. 1, 2021, doi: 10.1088/1742-6596/2010/1/012175.
[85] L. Yuan, Q. Hou, Z. Jiang, J. Feng, and S. Yan, "VOLO: Vision Outlooker for Visual Recognition," *IEEE Trans Pattern Anal Mach Intell*, vol. 45, no. 5, pp. 6575–6586, 2023, doi: 10.1109/TPAMI.2022.3206108.
[86] X. Jiang, Y. Zhu, Y. Liu, G. Cai, and H. Fang, "TransDD: A transformer-based dual-path decoder for improving the performance of thoracic diseases classification using chest X-ray," *Biomed Signal Process Control*, vol. 91, p. 105937, May 2024, doi: 10.1016/J.BSPC.2023.105937.
[87] J. Ko, S. Park, and H. G. Woo, "Optimization of vision transformer-based detection of lung diseases from chest X-ray images," *BMC Med Inform Decis Mak*, vol. 24, no. 1, Dec. 2024, doi: 10.1186/s12911-024-02591-3.
[88] P. K. A. Vasu, J. Gabriel, J. Zhu, O. Tuzel, and A. Ranjan, "FastViT: A Fast Hybrid Vision Transformer using Structural Reparameterization," Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.14189
[89] C.-F. Chen, Q. Fan, and R. Panda, "CrossViT: Cross-Attention Multi-Scale Vision Transformer for Image Classification," Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.14899
[90] C. J. Ejiyi *et al.*, "ResfEANet: ResNet-fused External Attention Network for Tuberculosis Diagnosis using Chest X-ray Images," *Computer Methods and Programs in Biomedicine Update*, vol. 5, Jan. 2024, doi: 10.1016/j.cmpbup.2023.100133.
[91] S. Rajaraman, G. Zamzmi, L. R. Folio, and S. Antani, "Detecting Tuberculosis-Consistent Findings in Lateral Chest X-Rays Using an Ensemble of CNNs and Vision Transformers," vol. 13, no. February, pp. 1–13, 2022, doi: 10.3389/fgene.2022.864724.
[92] A. Iqbal, M. Usman, and Z. Ahmed, "Tuberculosis chest X-ray detection using CNN-based hybrid segmentation and classification approach," *Biomed Signal Process Control*, vol. 84, Jul. 2023, doi: 10.1016/j.bspc.2023.104667.
[93] C. C. Ukwuoma *et al.*, "A hybrid explainable ensemble transformer encoder for pneumonia identification from chest X-ray images," *J Adv Res*, vol. 48, pp. 191–211, 2023, doi: 10.1016/j.jare.2022.08.021.

[94] T. Wang *et al.*, "PneuNet: deep learning for COVID-19 pneumonia diagnosis on chest X-ray image analysis using Vision Transformer," *Med Biol Eng Comput*, vol. 61, no. 6, pp. 1395–1408, Jun. 2023, doi: 10.1007/s11517-022-02746-2.
[95] H. N. Monday, J. Li, G. U. Nneji, S. Nahar, M. A. Hossin, and J. Jackson, "COVID-19 Pneumonia Classification Based on NeuroWavelet Capsule Network," *Healthcare (Switzerland)*, vol. 10, no. 3, Mar. 2022, doi: 10.3390/healthcare10030422.
[96] M. Y. T. T. Ç. Şaban Öztürk, "HydraViT: Adaptive multi-branch transformer for multi-label disease classification from Chest X-ray images," *Biomed Signal Process Control*, vol. 100, no. Part A, 2025, 106959, 2025.
[97] "AI Chest X-Ray Classification | RSNA." Accessed: Jan. 19, 2025. [Online]. Available: https://www.rsna.org/news/2022/september/ai-chest-x-ray-classification
[98] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases," May 2017, doi: 10.1109/CVPR.2017.369.
[99] J. P. Cohen, P. Morrison, and K. Roth, "COVID-19 Image Data Collection : Prospective Predictions are the Future," pp. 1–38, 2020.
[100] "GitHub - ieee8023/covid-chestxray-dataset: We are building an open database of COVID-19 cases with chest X-ray or CT images." Accessed: Feb. 06, 2025. [Online]. Available: https://github.com/ieee8023/covid-chestxray-dataset
[101] R. T. K. A. M. R. K. M. M. Z. et al. Chowdhury MEH, "Can AI Help in Screening Viral and COVID-19 Pneumonia?," *IEEE Access*, vol. 8, pp. 132665–132676, Jul. 2020.
[102] T. Rahman *et al.*, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Comput Biol Med*, vol. 132, May 2021, doi: 10.1016/j.compbiomed.2021.104319.
[103] S. Rahman, S. Sarker, M. A. Al Miraj, R. A. Nihal, A. K. M. Nadimul Haque, and A. Al Noman, "Deep Learning–Driven Automated Detection of COVID-19 from Radiography Images: a Comparative Analysis," *Cognit Comput*, vol. 16, no. 4, pp. 1735–1764, Jul. 2024, doi: 10.1007/s12559-020-09779-5.
[104] T. Rahman *et al.*, "Reliable tuberculosis detection using chest X-ray with deep learning, segmentation and visualization," *IEEE Access*, vol. 8, pp. 191586–191601, 2020, doi: 10.1109/ACCESS.2020.3031384.
[105] "Belarus TB Database and TB Portal." Accessed: Feb. 09, 2025. [Online]. Available: https://grantome.com/grant/NIH/AAI12021001-1-0-5
[106] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Duong, and M. Ghassem, "COVID-19 Image Data Collection: Prospective Predictions are the Future," *Machine Learning for Biomedical Imaging*, vol. 1, no. December 2020, pp. 1–38, Dec. 2020, doi: 10.59275/j.melba.2020-48g7.
[107] "Neo Xrays." Accessed: Feb. 04, 2025. [Online]. Available: https://www.kaggle.com/datasets/ingusterbets/neo-xrays
[108] "COVID19+PNEUMONIA+NORMAL Chest X-Ray Image Dataset." Accessed: Feb. 05, 2025. [Online]. Available: https://www.kaggle.com/datasets/sachinkumar413/covid-pneumonia-normal-chest-xray-images
[109] "Tuberculosis (TB) Chest X-ray Database." Accessed: Jul. 17, 2024. [Online]. Available: https://www.kaggle.com/datasets/tawsifurrahman/tuberculosis-tb-chest-xray-dataset
[110] T. Rahman *et al.*, "Exploring the effect of image enhancement techniques on COVID-19 detection using chest X-ray images," *Comput Biol Med*, vol. 132, May 2021, doi: 10.1016/J.COMPBIOMED.2021.104319.
[111] Y. Liu, Y. H. Wu, Y. Ban, H. Wang, and M. M. Cheng, "Rethinking computer-aided tuberculosis diagnosis," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, 2020, pp. 2643–2652. doi: 10.1109/CVPR42600.2020.00272.
[112] W. E. A. E.-S. F. El-Shafai, "Extensive COVID-19 X-Ray and CT Chest Images Dataset," 2020.

[113] "RSNA Pneumonia Detection Challenge | Kaggle." Accessed: Feb. 03, 2023. [Online]. Available: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/
[114] "QaTa-COV19 Dataset." Accessed: Feb. 07, 2025. [Online]. Available: https://www.kaggle.com/datasets/aysendegerli/qatacov19-dataset
[115] A. Haghanifar, M. M. Majdabadi, Y. Choi, S. Deivalakshmi, and S. Ko, "COVID-CXNet: Detecting COVID-19 in frontal chest X-ray images using deep learning," *Multimed Tools Appl*, vol. 81, no. 21, pp. 30615–30645, Sep. 2022, doi: 10.1007/S11042-022-12156-Z.
[116] "COVID-19 Radiography Database." Accessed: Feb. 08, 2025. [Online]. Available: https://www.kaggle.com/datasets/tawsifurrahman/covid19-radiography-database
[117] S. Hao, X. Li, W. Peng, Z. Fan, Z. Ji, and I. Ganchev, "YOLO-CXR: A novel detection network for locating multiple small lesions in chest X-ray images", doi: 10.1109/ACCESS.2024.Doi.
[118] H. T. Nguyen, M. N. Nguyen, L. D. Phung, and L. T. T. Pham, "Anomalies Detection in Chest X-Rays Images Using Faster R-CNN and YOLO," *Vietnam Journal of Computer Science*, vol. 10, no. 4, pp. 499–515, Nov. 2023, doi: 10.1142/S2196888823500094.
[119] M. Jain, "Lung Diseases Image Segmentation using Faster R-CNNs," Sep. 2023, [Online]. Available: http://arxiv.org/abs/2309.06386
[120] W. Fan *et al.*, "A deep-learning-based framework for identifying and localizing multiple abnormalities and assessing cardiomegaly in chest X-ray," *Nat Commun*, vol. 15, no. 1, Dec. 2024, doi: 10.1038/s41467-024-45599-z.
[121] A. S. and N. D. V. Tiwari, "Detecting COVID-19 Opacity in X-ray Images Using YOLO and RetinaNet Ensemble," in *2022 IEEE Delhi Section Conference (DELCON)*, IEEE, 2022, pp. 1–5.
[122] T. Y. and C. L. L. mao, "Pneumonia Detection in chest X-rays: a deep learning approach based on ensemble RetinaNet and Mask R-CNN," in *2020 Eighth International Conference on Advanced Cloud and Big Data (CBD)*, Taiyuan, China: IEEE, 2020, pp. 213–218.
[123] M. S. Lee *et al.*, "Evaluation of the feasibility of explainable computer-aided detection of cardiomegaly on chest radiographs using deep learning," *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-96433-1.
[124] S. prasad Koyyada and T. P. Singh, "An explainable artificial intelligence model for identifying local indicators and detecting lung disease from chest X-ray images," *Healthcare Analytics*, vol. 4, Dec. 2023, doi: 10.1016/j.health.2023.100206.
[125] S. Rahimiaghdam and H. Alemdar, "Evaluating the quality of visual explanations on chest X-ray images for thorax diseases classification," *Neural Comput Appl*, vol. 36, no. 17, pp. 10239–10255, Jun. 2024, doi: 10.1007/s00521-024-09587-0.
[126] J. Zhao, "CrossEAI: Using Explainable AI to generate better bounding boxes for Chest X-ray images," Oct. 2023, [Online]. Available: http://arxiv.org/abs/2310.19835
[127] B. S. P. K. A. S. A. J. and J. J. V. Kadali, "Pneumonia Detection in Chest X-Ray Images by using Resnet-50 Deep Learning Algorithm," in *2023 Third International Conference on Artificial Intelligence and Smart Energy (ICAIS)*, Coimbatore, India: IEEE, 2023, pp. 1078–1084.
[128] O. Dokun *et al.*, "Deep Learning Model for COVID-19 Classification Using Fine Tuned ResNet50 on Chest X-Ray Images," *Machine Learning Research*, vol. 9, no. 1, pp. 10–25, May 2024, doi: 10.11648/j.mlr.20240901.12.
[129] K. H. Lee, J. W. Choi, C. O. Park, D. H. Han, and M. S. Kang, "A Development and Validation of an AI Model for Cardiomegaly Detection in Chest X-rays," *Applied Sciences (Switzerland)*, vol. 14, no. 17, Sep. 2024, doi: 10.3390/app14177465.
[130] R. Mehrotra, R. Agrawal, and M. A. Ansari, "Diagnosis of hypercritical chronic pulmonary disorders using dense convolutional network through chest radiography," *Multimed Tools Appl*, vol. 81, no. 6, pp. 7625–7649, Mar. 2022, doi: 10.1007/s11042-021-11748-5.
[131] M. K. U. Ahamed *et al.*, "DTLCx: An Improved ResNet Architecture to Classify Normal and Conventional Pneumonia Cases from COVID-19 Instances with Grad-CAM-Based Superimposed Visualization Utilizing Chest X-ray Images," *Diagnostics*, vol. 13, no. 3, Feb. 2023, doi: 10.3390/diagnostics13030551.

[132] K. V. Priya and J. D. Peter, "A federated approach for detecting the chest diseases using DenseNet for multi-label classification," *Complex and Intelligent Systems*, vol. 8, no. 4, pp. 3121–3129, Aug. 2022, doi: 10.1007/s40747-021-00474-y.
[133] M. S. Al Reshan *et al.*, "Detection of Pneumonia from Chest X-ray Images Utilizing MobileNet Model," *Healthcare (Switzerland)*, vol. 11, no. 11, Jun. 2023, doi: 10.3390/healthcare11111561.
[134] A. Souid, N. Sakli, and H. Sakli, "Classification and predictions of lung diseases from chest x- rays using mobilenet v2," *Applied Sciences (Switzerland)*, vol. 11, no. 6, Mar. 2021, doi: 10.3390/app11062751.
[135] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pp. 1–14, 2015.
[136] S. Sharma and K. Guleria, "A Deep Learning based model for the Detection of Pneumonia from Chest X-Ray Images using VGG-16 and Neural Networks," in *Procedia Computer Science*, Elsevier B.V., 2022, pp. 357–366. doi: 10.1016/j.procs.2023.01.018.
[137] L. Kong and J. Cheng, "Classification and detection of COVID-19 X-Ray images based on DenseNet and VGG16 feature fusion," *Biomed Signal Process Control*, vol. 77, Aug. 2022, doi: 10.1016/j.bspc.2022.103772.
[138] Z. P. Jiang, Y. Y. Liu, Z. E. Shao, and K. W. Huang, "An improved VGG16 model for pneumonia image classification," *Applied Sciences (Switzerland)*, vol. 11, no. 23, Dec. 2021, doi: 10.3390/app112311185.
[139] S. B. , P. P. H. , J. V. D. , V. J. P. Patel, "mprovised VGG16 CNN Architecture for Predicting Tuberculosis Using the Frontal Chest X-Ray Images," in *Smart Innovation, Systems and Technologies, vol 235*, Singapore: Springer, Sep. 2021.
[140] D. M. Ibrahim, N. M. Elshennawy, and A. M. Sarhan, "Deep-chest: Multi-classification deep learning model for diagnosing COVID-19, pneumonia, and lung cancer chest diseases," *Comput Biol Med*, vol. 132, May 2021, doi: 10.1016/j.compbiomed.2021.104348.
[141] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.
[142] S. Showkat and S. Qureshi, "Efficacy of Transfer Learning-based ResNet models in Chest X-ray image classification for detecting COVID-19 Pneumonia," *Chemometrics and Intelligent Laboratory Systems*, vol. 224, May 2022, doi: 10.1016/j.chemolab.2022.104534.
[143] H. Yoo, S. Han, and K. Chung, "Diagnosis Support Model of Cardiomegaly Based on CNN Using ResNet and Explainable Feature Map," *IEEE Access*, vol. 9, pp. 55802–55813, 2021, doi: 10.1109/ACCESS.2021.3068597.
[144] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," Aug. 2016, [Online]. Available: http://arxiv.org/abs/1608.06993
[145] V. T. Q. Huy and C. M. Lin, "An Improved Densenet Deep Neural Network Model for Tuberculosis Detection Using Chest X-Ray Images," *IEEE Access*, vol. 11, pp. 42839–42849, 2023, doi: 10.1109/ACCESS.2023.3270774.
[146] M. Usman, I. A. Nasir, R. Saeed, H. Nazir, and M. Asad, "A deep learning approach for multi-label chest X-ray diagnosis using DenseNet-121," *IET Conference Proceedings*, vol. 2024, no. 10, pp. 210–217, Oct. 2024, doi: 10.1049/icp.2024.3307.
[147] A. A. T. A. S. A. M. Z. A. B. A. K. A. H. AbuKaraki, "Pulmonary Edema and Pleural Effusion Detection Using EfficientNet-V1-B4 Architecture and AdamW Optimizer from Chest X-Rays Images," *Computers, Materials & Continua*, vol. 80, no. 1, p. 1055, 2024.
[148] M. Nawaz, T. Nazir, J. Baili, M. A. Khan, Y. J. Kim, and J. H. Cha, "CXray-EffDet: Chest Disease Detection and Classification from X-ray Images Using the EfficientDet Model," *Diagnostics*, vol. 13, no. 2, Jan. 2023, doi: 10.3390/diagnostics13020248.

[149] S. Kim, B. Rim, S. Choi, A. Lee, S. Min, and M. Hong, "Deep Learning in Multi-Class Lung Diseases' Classification on Chest X-ray Images," *Diagnostics*, vol. 12, no. 4, Apr. 2022, doi: 10.3390/diagnostics12040915.
[150] E. C. Too, D. G. Mwathi, L. K. Gitonga, P. Mwaka, and S. Kinyori, "An X-ray image-based pruned dense convolution neural network for tuberculosis detection," *Computer Methods and Programs in Biomedicine Update*, vol. 6, Jan. 2024, doi: 10.1016/j.cmpbup.2024.100169.
[151] A. G. Howard *et al.*, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," Apr. 2017, [Online]. Available: http://arxiv.org/abs/1704.04861
[152] O. N. Mohammed, "Enhancing Pulmonary Disease Classification in Diseases: A Comparative Study of CNN and Optimized MobileNet Architectures," *Journal of Robotics and Control (JRC)*, vol. 5, no. 2, pp. 427–440, 2024, doi: 10.18196/jrc.v5i2.21422.
[153] R. K. R. H. L. Nirupam Shome, "Detection of tuberculosis using customized MobileNet and transfer learning from chest X-ray image," *Detection of tuberculosis using customized MobileNet and transfer learning from chest X-ray image, Image and Vision Computing*, vol. 147, 2024.
[154] C. Szegedy *et al.*, "Going Deeper with Convolutions," Sep. 2014, [Online]. Available: http://arxiv.org/abs/1409.4842
[155] M. Mujahid, F. Rustam, R. Álvarez, J. Luis Vidal Mazón, I. de la T. Díez, and I. Ashraf, "Pneumonia Classification from X-ray Images with Inception-V3 and Convolutional Neural Network," *Diagnostics*, vol. 12, no. 5, May 2022, doi: 10.3390/diagnostics12051280.
[156] L. N. and S. G. S. Prasher, "Inception V3 model for Tuberculosis detection using chest x-ray images," in *2023 3rd International Conference on Intelligent Technologies (CONIT)*, Hubli, India: IEEE, 2023, pp. 1–5.
[157] J. H. Kim, "IMPROVEMENT OF INCEPTIONV3 MODEL CLASSIFICATION PERFORMANCE USING CHEST X-RAY IMAGES," *J Mech Med Biol*, vol. 22, no. 8, Oct. 2022, doi: 10.1142/S0219519422400322.
[158] "DA and DB - TB Chest X-ray Datasets." Accessed: Aug. 31, 2024. [Online]. Available: https://www.kaggle.com/datasets/vbookshelf/da-and-db-tb-chest-x-ray-datasets
[159] Y. Hedhoud, T. Mekhaznia, and M. Amroune, "An improvement of the CNN-XGboost model for pneumonia disease classification," *Pol J Radiol*, vol. 88, no. 1, pp. 483–493, 2023, doi: 10.5114/pjr.2023.132533.
[160] "Pneumonia | Johns Hopkins Medicine." Accessed: Apr. 27, 2025. [Online]. Available: https://www.hopkinsmedicine.org/health/conditions-and-diseases/pneumonia
[161] "Complications of Pneumonia You Should Know." Accessed: Apr. 27, 2025. [Online]. Available: https://www.webmd.com/lung/complications-pneumonia
[162] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and flexible image augmentations," *Information (Switzerland)*, vol. 11, no. 2, pp. 1–20, 2020, doi: 10.3390/info11020125.
[163] Y. Yue and Z. Li, "MedMamba: Vision Mamba for Medical Image Classification," Mar. 2024, [Online]. Available: http://arxiv.org/abs/2403.03849
[164] Z. Jiang, "Chest X-ray Pneumonia Detection Based on Convolutional Neural Networks," *Proceedings - 2020 International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering, ICBAIE 2020*, pp. 341–344, 2020, doi: 10.1109/ICBAIE49996.2020.00077.
[165] X. Gu, L. Pan, H. Liang, and R. Yang, "Classification of bacterial and viral childhood pneumonia using deep learning in chest radiography," *ACM International Conference Proceeding Series*, pp. 88–93, 2018, doi: 10.1145/3195588.3195597.
[166] Nillmani *et al.*, "Four Types of Multiclass Frameworks for Pneumonia Classification and Its Validation in X-ray Scans Using Seven Types of Deep Learning Artificial Intelligence Models," *Diagnostics*, vol. 12, no. 3, Mar. 2022, doi: 10.3390/diagnostics12030652.
[167] H. Thanh Nguyen, T. Bao Tran, H. Hoang Luong, T. Phuoc Le, and N. Cong Tran, "Viral and Bacterial Pneumonia Diagnosis via Deep Learning Techniques and Model Explainability." [Online]. Available: www.ijacsa.thesai.org

[168] T. Xu and Z. Yuan, “Convolution Neural Network With Coordinate Attention for the Automatic Detection of Pulmonary Tuberculosis Images on Chest X-Rays,” *IEEE Access*, vol. 10, no. July, pp. 86710–86717, 2022, doi: 10.1109/ACCESS.2022.3199419.
[169] G. I. Okolo, S. Katsigiannis, and N. Ramzan, “IEViT: An enhanced vision transformer architecture for chest X-ray image classification,” *Comput Methods Programs Biomed*, vol. 226, p. 107141, 2022, doi: 10.1016/j.cmpb.2022.107141.
[170] Z. Yang, J. Zhang, G. Wang, M. K. Kalra, and P. Yan, “Cardiovascular Disease Detection from Multi-View Chest X-rays with BI-Mamba,” May 2024, [Online]. Available: http://arxiv.org/abs/2405.18533
[171] Y. Chen, “Mamaba-Based Federated Learning for Medical X-Ray Detection,” in *2024 4th International Conference on Electronic Information Engineering and Computer (EIECT)*, Shenzhen, China: IEEE, Nov. 2024, pp. 388–391.

Glossary

A

Active Tuberculosis

Airspace Opacification A radiological term referring to the filling of alveolar spaces with fluid, pus, blood, or other material, resulting in a white area on chest X-rays.

ARDS A life-threatening condition where fluid builds up in the lungs' air sacs, making it hard for oxygen to get into the blood.

B

Bronchial walls The muscular and elastic walls of the bronchial tubes that carry air in and out of the lungs. These walls can become thickened or inflamed in diseases such as bronchitis, asthma, or infections like tuberculosis.

C

Cardiomegaly An abnormal enlargement of the heart, which can be detected on a chest X-ray. It may indicate underlying heart disease such as heart failure

Cavitation The formation of a gas-filled space within lung tissue, caused by infections such as tuberculosis or severe bacterial pneumonia. Appears as a hollow, radiolucent area on chest imaging.

Consolidation A radiological pattern where lung tissue becomes solid due to the accumulation of cellular debris, fluid, or pus within the alveoli. Seen commonly in bacterial pneumonia and tuberculosis.

Cynosis A bluish discoloration of the skin, lips, or nails resulting from insufficient oxygen in the blood. It is a clinical sign of respiratory or cardiac distress.

D

Disseminated Tuberculosis A form of tuberculosis where the infection spreads from the lungs to multiple organs or tissues throughout the body via the bloodstream. It may affect the liver, spleen, bones, brain, or lymph nodes and is more common in immunocompromised individuals.

Dyspnea Medical term for shortness of breath or difficulty breathing.

E

Erythrocyte Sedimentation Rate (ESR)	A blood test that measures how quickly red blood cells (erythrocytes) settle at the bottom of a test tube over a specific period. A faster-than-normal rate may indicate inflammation in the body, and it is commonly used to support the diagnosis of infections like tuberculosis or pneumonia
Ground-Glass Opacity (GGO)	A hazy area on an X-ray that does not obscure the underlying structures. It suggests partial filling of air spaces or interstitial thickening and is common in viral pneumonia and early-stage infections.
	I
Infiltrate	An area in the lung where substances like fluid, pus, or cells have entered and filled the air spaces, making that part of the lung appear cloudy or white on an X-ray.
Ionizing radiation	A form of energy that has enough power to remove electrons from atoms or molecules, potentially causing damage to living tissues. It is used in medical imaging to visualize internal structures but must be used carefully to minimize exposure risks, especially with repeated scans.
Inflammatory thickening	A process where tissues, such as the bronchial walls or lung linings, become swollen and thick due to inflammation.
	K
Klebsiella pneumoniae	A type of gram-negative bacteria that can cause severe lung infections, particularly in hospitalized or immunocompromised individuals. It is known for causing necrotizing pneumonia and is often associated with high antibiotic resistance, making treatment more difficult.
Latent Tuberculosis	A stage of tuberculosis that is not contagious, where the bacteria exist in the body in an inactive form and do not produce any symptoms. This condition can develop into active tuberculosis if the immune system is compromised.
Lobar Pneumonia	A type of bacterial pneumonia that affects one or more lobes of the lungs, often associated with lobar consolidation.
	M
Mycobacterium tuberculosis	The bacterium that causes tuberculosis. It primarily affects the lungs but can also spread to other parts of the body. It is a slow-growing, airborne pathogen that can remain latent in the body for years before becoming active.
	N
Necrotizing Pneumonia	A severe form of pneumonia where lung tissue is destroyed (necrotized) due to a severe infection, often caused by aggressive bacteria like *Staphylococcus aureus* or *Klebsiella pneumoniae*. It can lead to the formation of cavities in the

	lungs and may result in serious complications such as abscesses or respiratory failure.
Nodular opacities	Small, round or irregular areas on a CXR that appear whiter than normal lung tissue. They indicate regions where something (like infection, inflammation, or a growth) is blocking the passage of air, and are often seen in diseases such as post-primary tuberculosis or certain fungal infections.
	O
Occupational Exposure	In the context of chest diseases, it often refers to inhaling dust, fumes, or chemicals in workplaces like mines, factories, or construction sites, which can lead to conditions such as pneumoconiosis or chronic lung disease.
Pleural effusion	An abnormal accumulation of fluid between the layers of the pleura (the membranes surrounding the lungs), which can cause breathing difficulty and appears as a white area on the lower part of the chest X-ray.
Pulmonary Fibrosis	A chronic and progressive lung disease where lung tissue becomes damaged and scarred, leading to stiffness and breathing difficulty.
	S
Staphylococcus aureus	A bacterium that can cause a severe form of pneumonia known as necrotizing pneumonia. This infection often progresses rapidly and may result in cavitation and lung tissue destruction.
Streptococcus pneumoniae	A type of bacteria that is one of the most common causes of bacterial pneumonia. It can also cause other infections such as sinusitis, ear infections, and meningitis, especially in young children and elderly people.