

MULTITASK X-RAY IMAGE DIAGNOSIS USING DEEP LEARNING

T SATYANARAYANA MURTHY

STUDENT ID: 969196

MSc Artificial Intelligence & Machine Learning

Liverpool John Moores University

Thesis Report

AUGUST 2021

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my thesis supervisor Dr. Prasanna Porwal and Dr. Ahmed Kaky for their support and guidance throughout this research project.

I would express my special and graceful thanks to my family, who has been supporting me throughout this project.

ABSTRACT

X-ray imaging is commonly used and cost-effective radiology procedure for diagnosing multiple diseases, but availability of radiologists to meet demand for X-ray diagnosis is very less. This brings importance of need for automated X-ray diagnosis procedures to address radiologist scarcity or radiologist fatigue problem. Deep learning methods have shown prominence in automatic analysis of X-ray images with performance approximating to human radiologist level. However, analysis of deep learning methods that can concurrently analyse multiple tasks from X-ray images to provide additional supporting diagnosis information is less explored and would support radiologist in providing better patient diagnosis.

As part of this research, using chest X-ray records, multitask deep learning methods that can provide optimal performance concurrently on X-ray abnormality detection task and age prediction task are developed. With these methods radiologist would be able provide patient diagnosis in relation to age and not just based on pathology information, indicating better patient care. In this study, evaluation of multitask methods is made on chest X-ray records that are annotated by radiologist indicating use of this research in practical setting. Multitask methods developed in this study provides performance results approximating to state of art. Using a single multitask method, optimal performance is achieved on abnormality detection task with Area Under the Receiver Operating Characteristic Curve value as 92% and on age prediction task with Mean Absolute Error value as 5.8 years.

Additionally, as part of this research, multi-pathology multitask methods are developed to evaluate advanced use case of multitask methods developed in this study. This research adds significant contribution to automated medical image diagnosis field by providing methods that can analyse X-ray images from multiple perspectives, like a radiologist.

Keywords: Deep Learning, Multitask Learning, X-ray Image Analysis, Computer-aided Diagnosis, Computer Vision, Medical Imaging.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	i
ABSTRACT	ii
LIST OF FIGURES	vi
LIST OF TABLES	viii
LIST OF EQUATIONS	ix
LIST OF ABBREVIATIONS	x
CHAPTER 1: INTRODUCTION	1
1.1 Background of Study	1
1.1.1 Automating X-ray Diagnosis	3
1.2 Problem Statement	4
1.3 Research Questions	4
1.4 Aim and Objectives	5
1.4.1 Aim	5
1.4.2 Objectives	5
1.5 Significance of Study	5
1.6 Scope of Study	6
1.7 Structure of Study	7
CHAPTER 2: LITERATURE REVIEW	9
2.1 Introduction	9
2.2 Overall Healthcare Image Analytics	10
2.2.1 Overview of DL Applications in Healthcare imaging	11
2.3 Overall X-ray Image Analytics	12
2.4 Automating Chest X-ray Pathology Analysis	13
2.4.1 Tuberculosis Screening methods	13
2.4.2 Methods for analysis of multiple pathologies from Chest X-ray Images	15
2.4.3 Abnormal Chest X-ray detection	21

2.4.4 Chest X-ray Multitask Learning for Pathology Analysis	23
2.5 X-ray Age Prediction Analysis	25
2.6 Musculo-skeleton X-ray Image Analysis	27
2.7 Research Gaps Identified	29
2.8 Summary	29
CHAPTER 3: RESEARCH METHODOLOGY	31
3.1 Introduction	31
3.2 Dataset Description	32
3.2.1 Pathology information	32
3.2.2 Additional Metadata Description	33
3.3 Data Preprocessing and Transformation Methods	35
3.4 Model Development	36
3.4.1 Using Transfer Learning Methods for Analysis	37
3.4.2 Multitask methods for abnormality classification and age prediction	38
3.5 Model Evaluation	40
3.6 Approach for Multitask Investigation Analysis	41
3.7 Required Resources and Tools	42
3.7.1 Resources	42
3.7.2 Tools and Libraries	42
3.8 Summary	42
CHAPTER 4: ANALYSIS AND IMPLEMENTATION	44
4.1 Introduction	44
4.1.1 Overall Flow of Research Investigation	44
4.2 Data Processing	45
4.2.1 Approach for Data Splitting	45
4.2.2 Data Generators	45
4.3 Model Development	46

4.3.1	Methods for Abnormality Detection Task Analysis	46
4.3.2	Methods for Age Prediction Task Analysis	47
4.3.3	Methods for Multitask Analysis	51
4.4	Methods for Multi-Pathology Multitask Analysis	53
4.5	Hyperparameters And Other Considerations	54
4.6	Summary	54
CHAPTER 5: RESULTS AND DISCUSSION		56
5.1	Introduction	56
5.2	Performance Analysis	57
5.2.1	Multitask Analysis	57
5.2.2	Supporting Analysis	60
5.3	Further Improving Performance on Age Prediction Task	63
5.4	Multi-pathology Multitask Analysis	65
5.4.1	Comparative Analysis with CheXPert Methods	67
5.5	Summary	67
CHAPTER 6: CONCLUSION AND RECOMMENDATIONS		69
6.1	Introduction	69
6.2	Discussion and Conclusion	69
6.3	Contribution to Knowledge	70
6.4	Future Recommendations	71
REFERENCES		72
APPENDIX A: RESEARCH PROPOSAL		80
APPENDIX B: PSEUDO CODE		101
APPENDIX C: PROJECT PLAN		103

LIST OF FIGURES

Figure 1.1: (A) X-ray of Female patient having abnormality detected. (B) X-ray of Male patient without abnormalities detected (Irvin et al., 2019).....	2
Figure 1.2: (A) Figure indicating previous DL approaches (B) Figure indicating current proposed study of multitask DL approach.....	6
Figure 2.1 : History of ML in fields of computer vision and medical imaging (Suzuki, 2017).	9
Figure 2.2: Overall flow of literature study conducted.	10
Figure 2.3: Basic CNN Architecture with Input, Convolutional, Pooling, Fully Connected and Output layers (Esteva et al., 2019).	11
Figure 2.4: Residual connection used in ResNet (He et al., 2016).....	14
Figure 2.5: Overall framework of DL methods used for pathology classification on NIH chest X-ray dataset (Wang et al., 2017).....	16
Figure 2.6: CheXNet, a 121-layer model that takes X-ray image as input and provides probability of pathology as output (Rajpurkar et al., 2017b).	17
Figure 2.7: CheXPert for predicting probabilities of observations from multi-view chest X-ray images (Irvin et al., 2019).....	19
Figure 2.8: Figure indicating CNN-RNN-RNN model used for generating radiological reports (Liu et al., 2019)	20
Figure 2.9: Multitask neural network jointly trained for segmentation and abnormality detection (Gündel et al., 2019).	23
Figure 2.10: Deep Learning architecture used by top performing model in RSNA paediatric bone assessment challenge (Halabi et al., 2019).	26
Figure 2.11: Musculo-skeleton X-ray images of patients with both normal and abnormal label indication (Rajpurkar et al., 2017a).	28
Figure 3.1: Overall flow of research methodology in this study	31
Figure 3.2: (A) Frontal and lateral chest X-ray views of a patient. (B) Proportion of frontal and lateral X-ray views.....	34
Figure 3.3: (A) Age Distribution of patients in dataset. (B) Age distribution of patients with “No Finding” label in dataset.	34
Figure 3.4: Proportion of Male and Female patients with radiographic examination.....	35
Figure 3.5: (A) Uncertain cases marked as normal i.e., 0 (B) Uncertain cases marked as abnormal i.e., 1 (C) Uncertain cases left without change i.e., -1.....	36

Figure 3.6: Overall approach of using transfer learning for pathology analysis	37
Figure 3.7: Dense block with interconnected layers (Huang et al., 2017).	38
Figure 3.8: Overall modelling approach for multitask model to detect abnormal X-ray and predict age value from X-ray.....	39
Figure 3.9: Records indicating count of normal or abnormal X-rays available for testing the model.	40
Figure 4.1: Overall approach of analysis made for this study	45
Figure 4.2 : Overall flow of data processing before providing information to model during training process.....	46
Figure 4.3: Model architecture for abnormality detection from chest X-ray images	47
Figure 4.4 : (A) Figure indicating age distribution of all patients. (B) Figure indicating age distribution of patients with normal X-ray record.	48
Figure 4.5 : Figure indicating number of normal and abnormal X-ray images in NIH chest X-ray dataset.....	49
Figure 4.6 : Model architecture used for analysis of age prediction task.	49
Figure 4.7: Model architecture used for comparative multi-tasking analysis.	51
Figure 4.8: Overall approach of multitask methods using combination of pre-trained models.	52
Figure 4.9 : Methods for multi-pathology multitask analysis	53
Figure 5.1: Overview of different methods developed for analysis.	56
Figure 5.2: Performance comparison of multitask methods using different pre-trained models.	58
Figure 5.3: Performance comparison of different multitask architecture models.	59
Figure 5.4: Comparative performance analysis on abnormality detection task.	61
Figure 5.5 : Comparison of abnormality detection task performance against (Rajaraman et al., 2020).....	62
Figure 5.6 : Comparative analysis of performance on age prediction task	63
Figure 5.7: Comparative performance analysis on age prediction task between multitask methods using MSLE and MAE loss function.	65
Figure 5.8: ROC curve plot for multiple pathologies - uncertain as abnormal (5 prevalent pathologies)	66
Figure 5.9: Performance comparison of multi-pathology multitask methods developed in this study with CheXPert.....	67

LIST OF TABLES

Table 1.1: Table indicating summary of healthcare imaging modality characteristics (Panayides et al., 2020)	2
Table 2.1 Table indicating summary of study made on methods for automating TB screening.	15
Table 2.2: Table indicating summary of discussion made for studies using NIH Chest X-ray dataset	18
Table 2.3: Table indicating summary of discussion made for studies using CheXPert dataset	19
Table 2.4: Table indicating summary of methods developed for enabling automated radiology report generation using MIMIC-CXR dataset.	21
Table 2.5: Table indicating summary of discussion made on previous studies for abnormality detection task.	22
Table 2.6: Table indicating summary of study made on multitask methods	24
Table 2.7: Table indicating summary of discussion of studies that work on age prediction task	27
Table 2.8: Table indicating summary of study made on Musculo-Skeleton X-ray Analysis... Table 3.1: CheXPert dataset consisting of 14 Labelled Observations (Irvin et al., 2019).	28 33
Table 4.1: Table indicating age prediction performance results achieved on NIH Chest X-ray.	50
Table 4.2 : Comparative Study of Age Prediction Analysis Task.....	50
Table 4.3: Table indicating input image size for different pre-trained models	54
Table 5.1: Table indicating multitask analysis performance of methods using combination of pretrained models.	60
Table 5.2: Table indicating multitask performance results with MAE as loss for age prediction task.....	64
Table 5.3: Table indicating performance measures of multi-pathology multitask methods	66

LIST OF EQUATIONS

Equation 3.1	39
Equation 3.2	39
Equation 3.3	39
Equation 3.4	40
Equation 3.5	41
Equation 3.6	41

LIST OF ABBREVIATIONS

AP	Anteroposterior
AUC	Area Under Receiver Operating Characteristic Curve
AUPRC	Area Under Precision-Recall Curve
CAD	Computer Aided Diagnosis
CNN	Convolutional Neural Network
CT	Computed Tomography
DL	Deep Learning
FN	False Negatives
FP	False Positives
GAN	Generative Adversarial Networks
ID	Identifier
MAD	Mean Absolute Difference
MAE	Mean Absolute Error
ML	Machine Learning
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
MSLE	Mean Squared Logarithmic Error
MTL	Multitask Learning
PA	Posteroanterior
RNN	Recurrent Neural Network
SVM	Support Vector Machine
TB	Tuberculosis
TN	True Negatives
TP	True Positives

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Healthcare imaging is an important field of study that includes multiple acquisition domains like Magnetic Resonance Imaging (MRI), Computed Tomography (CT), X-ray, Ultrasound, Surgery, Microscopy and Mammography imaging techniques (Litjens et al., 2017). They are useful for diagnosis of multiple life-threatening pathologies like Musculo-skeleton problems, Dental problems, Foreign objects, Brain tumours, Pulmonary diseases, Cancers, Cardiac diseases, Eye diagnosis of fundus and retina etc. Additionally, Ultrasound imaging is also used for observing foetal position, heart rate etc.

Having multiple applications of radiology procedures is causing increasing demand for radiologist, while the number of radiologists per patient to meet demand is very less. For example, In US there is approximately one radiologist for every 10,000 population (Arora, 2014), in UK it is 7.5:100,000(The Royal College of Radiologists, 2016). While in populous country like India it is 1:100,000 (Arora, 2014) and in countries of Africa it is even lesser (Rosman1* et al., 2015; Ali et al., 2015). Hence radiologist must spend more time to meet the demand. Adding to this recently due to COVID-19 pandemic, there has been exponential increase in referred radiology procedures for diagnosis of patients. This causes fatigue for radiologist to meet increasing demand. Additionally, lesser availability of radiologist per person causes limited healthcare access to people. Hence, need for enabling Computer Aided Diagnosis (CAD) of radiology reports to assist radiologist in making faster and accurate diagnosis has become important problem to be addressed in near future.

Further X-ray imaging, part of radiology imaging, is low cost and quick in acquisition compared with other methods used for diagnosis. Table 1.1 indicates summary of characteristics of different radiology procedures available, where X-ray imaging is observed to be of low cost and useful. X-ray imaging helps in diagnosis of multiple organs and a valuable tool for wide variety of examinations and procedures (Panayides et al., 2020). Unlike CT which uses high doses of radiation, X-ray imaging uses relatively low dose of radiation making it safer to use

than CT. These characteristics makes X-ray imaging most referred imaging procedure for diagnosis of patients. Hence enabling automatic diagnosis of X-ray images will have significant impact in addressing the radiologist scarcity problem.

Table 1.1: Table indicating summary of healthcare imaging modality characteristics (Panayides et al., 2020)

Imaging Type	Anatomies	Dimensionality	Cost/Scan*
X-ray	Most Organs	2D,2D+t	\$15-385
CT	Most Organs	2D, 3D, 4D	\$57-385
Ultrasound	Most Organs	2D, 2D+t, 3D, 4D	\$57-1483
MRI	Most Organs	3D, 4D	\$32-691
Nuclear	All organs with radioactive tracker uptake	2D, 3D, 4D	\$182-1375
Microscopy	Primarily biopsies and surgical specimens	2D, 3D, 4D	\$248-1483

Among X-ray imaging procedures, chest X-ray imaging is widely adapted and commonly used procedure for pulmonary disease diagnosis. Hence enabling automated chest X-ray diagnosis procedures would be useful in practical setting. Chest X-ray images can be taken in frontal and lateral views. Frontal view X-rays can be taken with patient facing X-ray projector i.e., Anteroposterior view (AP) (or) against X-ray projector i.e., Posteroanterior view (PA). Lateral views are acquired with patient facing sideways to projector. Figure 1.1 indicates abnormal and normal X-ray images of patients taken from CheXPert dataset (Irvin et al., 2019).

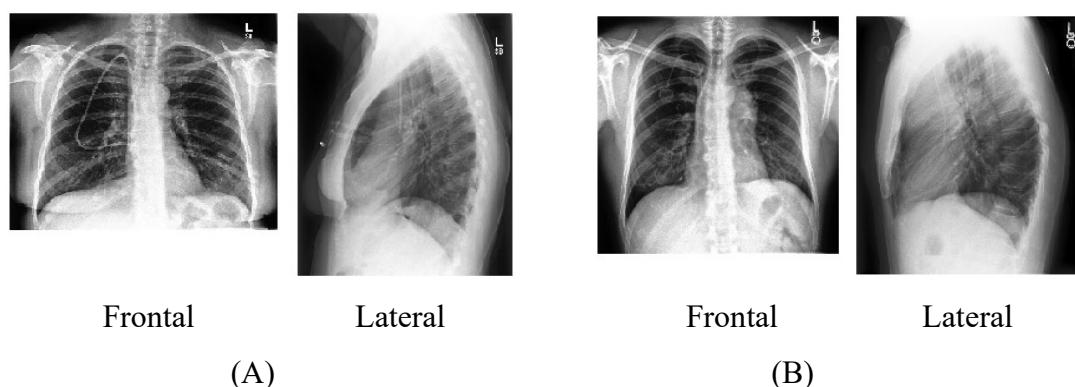


Figure 1.1: (A) X-ray of Female patient having abnormality detected. (B) X-ray of Male patient without abnormalities detected (Irvin et al., 2019).

1.1.1 Automating X-ray Diagnosis

Initially automated methods for analysis of X-ray images were proposed based on handcrafted features and smaller datasets. Later methods using X-ray images directly as input instead of features driven approaches are becoming prominent. Particularly in last decade, success of AlexNet (Krizhevsky et al., 2012) and release of large-scale database of X-ray images like NIH Chest X-ray (Wang et al., 2017) and CheXPert (Irvin et al., 2019) comprising comprehensive labelled diagnosis information of multiple pathologies enabled significant progress in field of automated X-ray diagnosis. Success of AlexNet enabled researchers to use Deep Learning (DL) methods and show performance exceeding previous studies. Most of the methods proposed used transfer learning methods, where they used pre-trained computer vision models as initialization methods and further optimized models to suit for analysis of X-ray images.

Multiple DL methods using Convolutional Neural Network (CNN) architecture were explored to analyze different X-ray imaging tasks like classification of X-ray images with pathologies, segmentation of organs from X-rays, indicating area of pathologies in X-ray images, generation of automated diagnosis report, determine skeletal maturity of pediatrics etc. These methods showed state of art results and sometimes performance exceeding that of human radiologist in multiple cases. Unlike traditional machine learning (ML) methods, CNN methods using transfer learning can easily enable analysis of input images from a different domain. This enabled faster and scalable algorithm development. Hence DL with CNN methods became dominant choice for exploring automated analysis of patient X-ray images.

However, X-ray images are complex to diagnose and hence they need DL models that can support more in-depth analysis. But increasing depth of DL model by adding more hidden layers to accommodate more comprehensive analysis can cause vanishing gradient descent problem preventing model from training further. Initial proposed CNN models suffer from this vanishing gradient descent problem during training. However, after success of models like ResNet (He et al., 2016) and DenseNet (Huang et al., 2017) on large-scale ImageNet database, multiple research methods were proposed using these pre-trained models for analyzing X-ray images. These models use residual connection mechanism and enabled model architectures to have more hidden layers without losing gradients during training. This enabled creation of DL models that can support more hidden layers and can support workflows that need comprehensive analysis. Mainly DenseNet (Huang et al., 2017), that use dense blocks where

all layers in dense block are connected to each other, became commonly used choice for analyzing X-ray images.

1.2 Problem Statement

Though current research methods using DL focus on analysis of X-ray images and show promising results, but they are mainly focused on analysis of a single task. These methods focus on analysing pathology detection, abnormality detection and age prediction tasks independently. Multiple multitask methods were proposed that analyse pathology detection and segmentation tasks together, but in general they also focus on accurate analysis of pathology information from X-rays.

Prior methods, as they focus on addressing a particular problem, don't consider comprehensive analysis of patient reports unlike a radiologist who can analyse patient reports from multiple perspective and provide accurate diagnosis. For automated patient diagnosis to work in practical conditions, identifying disease pathology alone is not sufficient, but additional parameters like patient body condition and body age are also needed for proper patient treatment. Getting such additional analysis information shall help in improved patient care, where patient diagnosis or treatment is not just based on pathology but also on patient body condition in relation to social context, which refers to idea of precision healthcare (Colijn et al., 2017).

In a practical setting, multitask analysis of identifying abnormal X-ray images in addition to predicting age would be extremely beneficial. It helps in initial screening or prioritising critical patients based on multiple parameters under consideration. This shall ease burden of radiologist and provide care to needed patients on priority. Additionally, such a tool shall also help radiologist to provide appropriate patient diagnosis relevant to social context than just based on pathology information.

1.3 Research Questions

What are the deep learning methods that can be used for concurrently analysing both abnormality classification and age prediction tasks from chest X-ray images?

1.4 Aim and Objectives

Below aim and objectives are considered for this research.

1.4.1 Aim

Aim of this research is to investigate methods that can concurrently detect abnormality and predict patient's age from chest X-ray images, to support radiologist in patient diagnosis.

1.4.2 Objectives

To meet aim of this study, research objectives formulated are as follows:

- To develop multitask DL methods that concurrently analyses chest X-ray images for abnormality detection based on pathology information and predict patient's age.
- To evaluate performance of multitask DL methods being developed to determine appropriate method for predictions.
- To comparatively evaluate multitask DL methods performance with performance of DL methods trained on these tasks separately.

1.5 Significance of Study

According to (World Health Organization, 2016) approximately 3.6 billion diagnostic medical examinations, such as X-rays, happen every year. With such a huge demand for medical procedures and lesser number of available radiologists, there is a need for smart procedures that can help in automating patient diagnosis to assist radiologist. (Irvin et al., 2019) suggests that “*Automated chest radiograph interpretation at the level of practicing radiologists could provide substantial benefit in many medical settings, from improved workflow prioritization and clinical decision support to large-scale screening and global population health initiatives*”.

To support this, current research focuses on predicting additional parameters i.e., age value in addition to abnormality detection based on pathology information from chest X-ray images. Though current thesis works on determining patient's age value as one of the tasks, this idea can be expanded to additional parameters like bone density, lung area etc. Having such supporting analysis information shall support in enabling precision healthcare, where patient

diagnosis is not just based on pathology but also based on his body condition and other parameters. Like a radiologist, this study provides automated multitask X-ray diagnosis framework that looks at patient diagnosis from multiple perspectives to support radiologist in providing right treatment to right patient at right time. For example, age value prediction task considered in this research can be used for preventive counselling of patients by comparing to average health status (Karargyris et al., 2019).

Figure 1.2 indicates overview of multitask approach that this thesis focuses on, where prediction of age value is made in addition to abnormal detection from chest X-ray images. Having such a tool which can diagnose patients from multiple perspectives, will help radiologist to perform at improved expertise level providing enhanced care and proper treatment to patients. In figure 1.2, I/P represents input data.

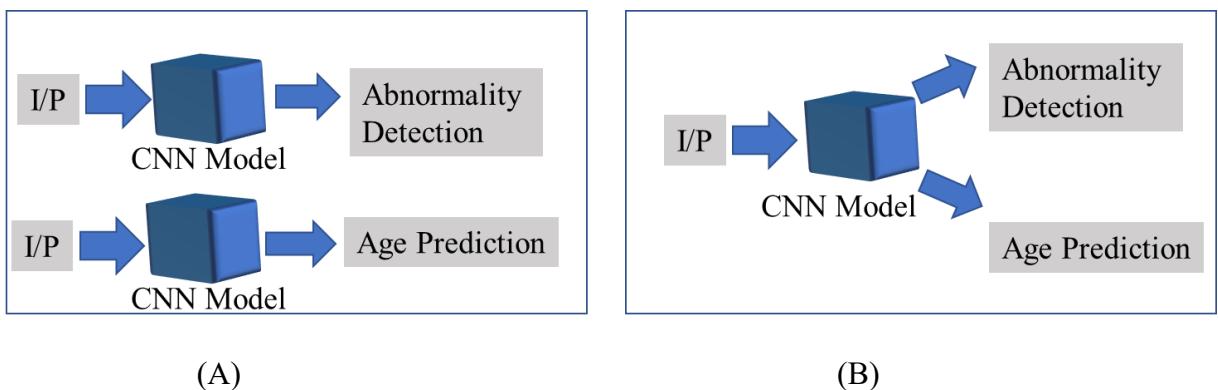


Figure 1.2: (A) Figure indicating previous DL approaches (B) Figure indicating current proposed study of multitask DL approach.

Additionally, this study also throws light on prior DL methods that works on automating medical images, particularly X-ray images. A systematic and thorough analysis of those methods is discussed in literature review section, helping to summarize important works made in field of medical image analysis with further focus on X-ray imaging methods.

1.6 Scope of Study

Though methods proposed in this study can be used for broad idea of automated analysis of healthcare images, scope of current study is limited to analysis of Chest X-ray images for determining abnormality and age value. Allowing these limits on scope of study enables this research to focus on in depth evaluation of problem statement within time constraints allowed for conducting this research. Since patient age is provided as metadata in chosen dataset,

prediction of age value is considered as additional task in this analysis. If more information can be acquired in future, those parameters can also be chosen as additional tasks.

1.7 Structure of Study

Structure of this thesis is organized as,

Chapter 1, Introduction, discusses about background of research domain and problem statement under consideration. It also discusses upon research questions, aim & objectives leading to this study. Further, significance and scope of study are described indicating importance of conducting this research and limitations.

Chapter 2, Literature Review, discusses about systematic study of previous works conducted in the field of automated analysis of healthcare images with focus on X-ray images. Prior studies on automated analysis of chest and Musculo-skeleton X-ray images with applications of pathology analysis, age prediction is discussed. Further, research gaps identified from the study are discussed.

Chapter 3, Research Methodology, discusses about research design process of multitask methods considered for this study. A brief description on dataset considered for analysis is made. Discussion is made on use of transfer learning methods for concurrent analysis of different tasks i.e., abnormality detection task and age prediction task. Further, overall approach of multitask DL methods considered for addressing research gaps is discussed.

Chapter 4, Analysis and Implementation, discusses on design analysis and implementation details of multitask methods developed for this study. Discussion is made on different methods that analyse abnormality detection task, age prediction task, and multitask methods that analyse both abnormality detection and age prediction tasks. Additionally, multi-pathology multitask methods that analyses multiple pathologies instead of abnormality detection task in addition to age prediction task are discussed.

Chapter 5, Results and Discussion, discusses on performance analysis results measured from multitask methods developed in this study. Comparative performance analysis of multitask

methods with methods that are trained on separate independent tasks is made. Further, brief evaluation of multi-pathology multitask methods is discussed.

Chapter 6, Conclusion and Recommendations, discusses on summary of research investigation made in this study. Further, contribution to knowledge and future recommendations indicating directions for next steps of this research are indicated.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Deep Learning (DL) is end-to-end Machine Learning (ML), as it can map input images to final output without need for computing features from input images. DL has gained momentum after Convolutional Neural Networks (CNN) won ImageNet competition, though prior to that some progress was made in this field (Suzuki, 2017). DL models achieved state of art performance approximating humans, for those tasks that are previously considered difficult tasks like automated medical image diagnosis. Since DL models are trained with real world clinical data to make more accurate diagnosis and these DL models additionally can also be optimized for individual treatment decisions, they shall enable different physicians to perform at improved level of expertise (Norgeot et al., 2019). Hence focus of this study is use of DL in enabling Computer Aided Diagnosis (CAD) for assisting radiologist. Figure 2.1 indicates progress of Machine Learning (ML) in computer vision and medical imaging, where DL architectures became dominant after success of AlexNet, using CNN on ImageNet database (Suzuki, 2017).

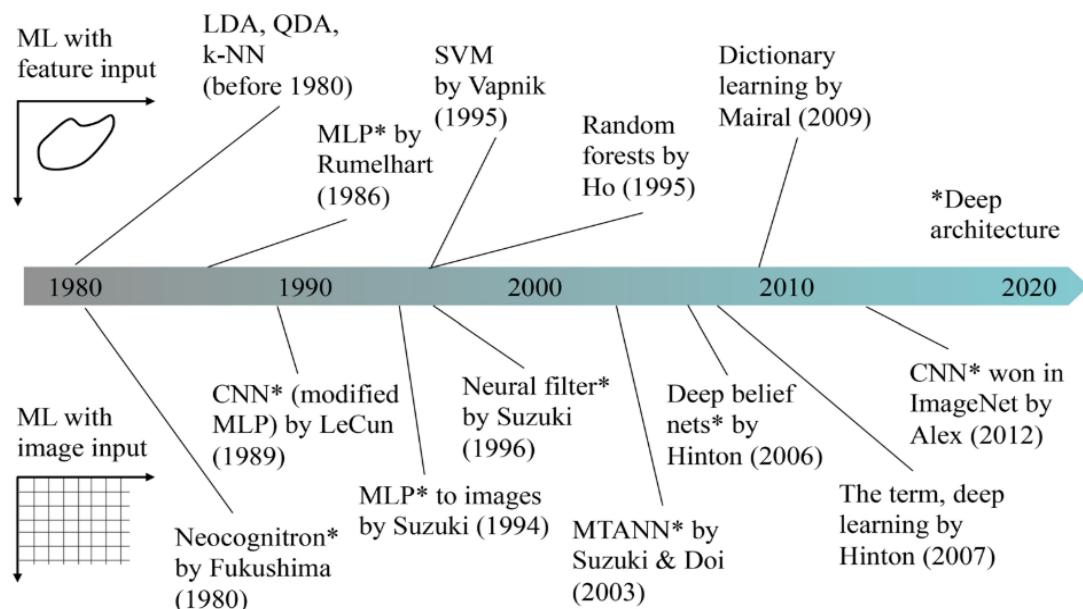


Figure 2.1 : History of ML in fields of computer vision and medical imaging (Suzuki, 2017).

Figure 2.2 depicts overview of systematic study of current literature conducted for this thesis, with focus on DL methods. Remaining sections of this chapter provides background and overview of current work in area of automated healthcare image analysis. Since medical image processing is a broader research area, to limit scope of study, further study narrows down to acquisition domain of X-ray image processing using DL methods. Discussion on methods addressing pathology analysis and age assessment from X-rays is made. Additionally, research gaps found from the study conducted is discussed.

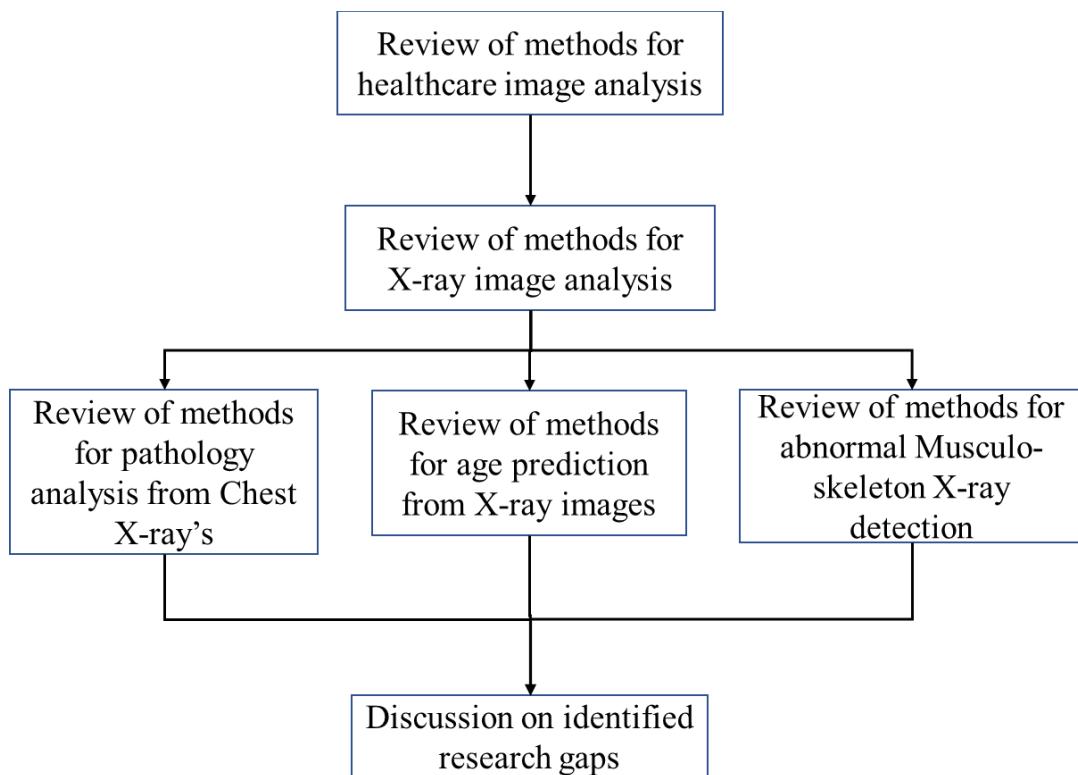


Figure 2.2: Overall flow of literature study conducted.

2.2 Overall Healthcare Image Analytics

(Suzuki, 2017) discusses on two major algorithms for processing of images as input i.e., Massive Training Artificial Neural Networks (MTANN) and Convolutional Neural Networks (CNN) architectures. Discussion was made on advantages of MTANN over CNN where, unlike CNN, MTANN helps in methods where availability of data is less like medical images. But, based on study conducted major progress in enabling automated processing of medical images is mainly driven by CNN with little focus on MTANN.

Driven by release of large scale public medical image datasets for research purposes and using transfer learning methods, state of the art results was achieved using CNN offsetting its limitations. Additionally, acceleration of parallelizable CNN algorithm calculations through GPU helped in enabling practical use of CNN (Greenspan et al., 2016). Figure 2.3 indicates basic CNN architecture where convolution layers followed by pooling layers are used to capture special and temporal representation of images, apart from input and output layers. (Esteva et al., 2019).

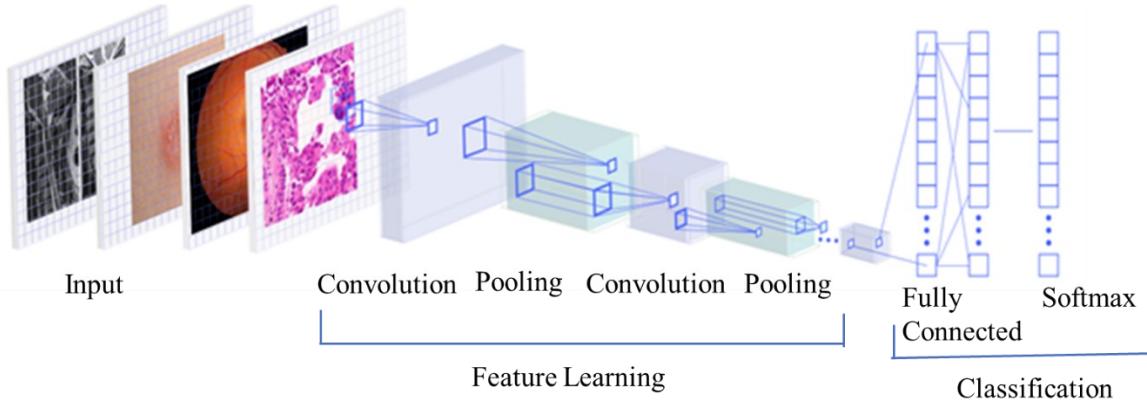


Figure 2.3: Basic CNN Architecture with Input, Convolutional, Pooling, Fully Connected and Output layers (Esteva et al., 2019).

In last decade, significant progress has been made in automation of medical imaging analysis with release of open access challenges on large scale datasets for benchmarking purposes. (Bram van Ginneken & Sjoerd Kerkstra, 2012) provides list of challenges that are organized within area of medical image analysis. Various applications of medical image analysis like segmentation, classification, image registration, content-based image retrieval (CBIR), explainable analysis interpretation and pathology detection for multiple acquisition domains like X-ray, CT, MRI etc., were explored using DL methods.

2.2.1 Overview of DL Applications in Healthcare imaging

DL methods have been extensively used for analysis of brain images. Multiple studies were made on classification of Alzheimer's disease and segmentation of brain tissues. Various applications of Mammography like breast cancer detection and classification, breast density estimation, localization and classification of masses were made. In analysis of thoracic or pulmonary diseases, CNN is used for segmentation and classification of lung nodules, predict

mortality rate, pathology classification and age regression (Litjens et al., 2017). In Genomics, DL has been extremely useful in identifying reliable imaging biomarkers and understanding genetics of disease to provide more accurate diagnosis (Lee et al., 2017). In Ophthalmology, DL methods were used for detection of retinal abnormalities and diagnosis of eye diseases (Razzak et al., 2018). In Surgery, Deep Reinforcement Learning (Deep RL) using computer vision models can help in robotic-assisted surgery by performing automation of repetitive and time-sensitive surgical tasks. But given that every surgery is unique, it will be difficult to automate this task and will be an interesting area for future research purposes (Esteva et al., 2019).

Apart from this significant progress was made in other research areas in healthcare imaging through DL i.e., Image reconstruction, multi-modality image analysis, precision medicine, generation of analysis interpretation statements (Panayides et al., 2020). Most of the approaches proposed use of DL methods with CNN for processing of medical images to achieve state of art performance and exceeding in few cases.

Adapting DL methods in practical setting has multiple challenges like interpretability, security, and bias of input data during training of models. Despite challenges in adopting DL methods to practice, significant progress was made since last 5 years in release of products that are approved for commercial use, enabling faster and accurate diagnosis (Benjamens et al., 2020). Based on discussion made so far, DL methods has not only been extremely useful for medical image analysis till now but also shows a promising future towards enabling automatic analysis of medical images helping radiologist. Since overall healthcare image analysis is a broad area of research including multiple acquisition domains, to limit scope of study, current thesis further focuses on automated X-ray image analysis.

2.3 Overall X-ray Image Analytics

Early use of CNN for analysis of medical images is made in 1995 for detection of lung nodules from Chest X-ray images (Lo et al., 1995). After that CNN was not much explored until new techniques for effective training of deep neural network and advances in core computer systems were made (Litjens et al., 2017). After success of AlexNet (Krizhevsky et al., 2012), multiple approaches were made on analysis of medical images using CNN models pre-trained on ImageNet classification. Various approaches such as pathology classification of X-ray images,

segmentation of images for organ and pathology detection, image reconstruction, multitask analysis for pathology segmentation and classification were proposed, using DL methods. These methods were able to produce state of art results approximating human performance and exceeding in few cases.

Based on study conducted, most of current literature was focused on analysis of Chest X-rays for pathology analysis. Also due to COVID-2019 pandemic outbreak and aiming for faster and effective diagnosis of patients, significant progress is being made on using DL with CNN for analysis of chest X-ray images to detect COVID patients (Jain et al., 2021). Apart from Chest X-ray image analysis, few other methods were also proposed where focus is on analysis of Musculo-skeleton X-ray's. Further sections discuss on DL methods used for analysis of Chest and Musculo-skeleton X-ray images.

2.4 Automating Chest X-ray Pathology Analysis

Using DL methods researchers were able to create automated chest X-ray image analysis methods whose performance approximates to radiologist. Though initial research focus was on enabling automatic detection of Tuberculosis(TB) from X-ray images, but later interest on detection of multiple pathologies from chest X-ray's is becoming dominant.

2.4.1 Tuberculosis Screening methods

Tuberculosis (TB) is potentially serious infectious disease that mainly effects lungs, which if not treat early can cause death. TB mainly effects pulmonary organs but it can affect other organs also (Burrill et al., 2007). Research focus for automating chest X-ray diagnosis increased with release of two public datasets, Montgomery, and Shenzhen Chest X-ray datasets, focusing mainly on detecting tuberculosis from chest X-rays. These datasets consist of normal and abnormal posteroanterior (PA) view chest X-rays with manifestations of tuberculosis (TB). They also include associated radiologist readings (Jaeger et al., 2014a). Further using these datasets, automatic tuberculosis screening methods using SVM were developed for practical use in Kenya. Authors developed models where they extracted features from images and provided those features to a binary classification model for tuberculosis detection. With approach followed, promising results were reported with Area Under Receiver Operating Characteristic Curve (AUC) of 88.5% for TB classification (Jaeger et al., 2014b). Though

reported model performance is better for benchmarking purposes, it is still less compared to radiologist performance and using DL methods can improve performance further.

(Pasa et al., 2019) proposed a novel DL model with CNN for TB classification with advantages of having low complexity and still achieving state of art performance suitable for real time deployment. However generalizability of such models on external data would be less due to bias in input data and using pre-trained models can still help in improving performance further. (Yadav et al., 2019) showed improved model performance on TB detection using transfer learning methods through pretrained GoogleNet model. They showed increased model performance from coarse to fine-tuned training methods where accuracy is improved from 50% at coarse to 94.89% at fine-tuned training. (Zhang et al., 2020) used a variant of ResNet model by replacing Global Average Pooling layer (GAP) with adaptive dropout and trained model on binary classification of multiple classes instead of multi-label classification, where they were able to achieve results approximating state of art performance. Figure 2.4 shows example of residual connection proposed in ResNet architecture, where extra identity(residual) connection from input to output shall help in increased depth of model offsetting vanishing gradients problem during training of neural network. ResNet uses residual connections which supports in increasing depth of model and thereby learning better feature representation of input (He et al., 2016).

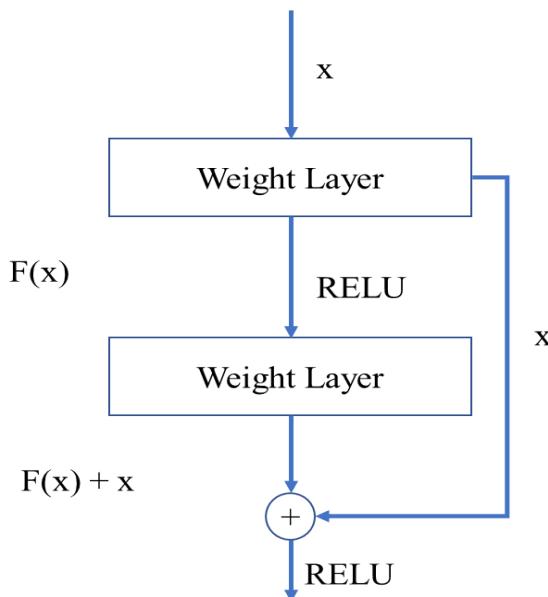


Figure 2.4: Residual connection used in ResNet (He et al., 2016)

Segmentation task in X-ray images is used for locating pathologies, organ structures in chest X-ray images to provide enhanced information to radiologist. Automatic segmentation task of X-ray images for patients with tuberculosis would be difficult because lung regions are affected by opacities or consolidations. Different DL methods were proposed for segmentation tasks where different metrics are used for evaluation of performance of methods like Dice score and Intersection Over Union (IOU) (Dai et al., 2018; Souza et al., 2019; Gordienko et al., 2019).

Table 2.1 indicates overall summary of discussion of previous studies on automated TB screening methods. Further, though analysis of X-rays images for TB classification shows potential impact of DL methods in radiology procedures, but analysis using information from multiple pathology labels would be more appropriate for enabling practical setting of automated radiology procedures.

Table 2.1 Table indicating summary of study made on methods for automating TB screening.

Reference	Remarks
(Jaeger et al., 2014a, 2014b)	Developed dataset for TB detection and developed methods for Tuberculosis detection with best AUC of 88.5%.
(Pasa et al., 2019)	Developed methods that does TB classification with AUC of 92.5%.
(Yadav et al., 2019)	Developed methods that does TB classification with accuracy of 94.89%, using methods based on GoogleNet model.
(Zhang et al., 2020)	Developed methods that does TB classification with accuracy of 87.7%, using methods based on ResNet model.
(Dai et al., 2018; Souza et al., 2019; Gordienko et al., 2019).	Developed methods for automatic segmentation of organs from images with TB.

2.4.2 Methods for analysis of multiple pathologies from Chest X-ray Images

Multiple research methods were proposed for analysis of pathologies from X-ray images. For automated analysis of X-ray images, multiple methods were proposed whose analysis is based on limited datasets. However, multiple other methods were also proposed whose focus is on analysis of DL methods based on large scale datasets released for benchmarking purposes. For reliable and systematic review of study being conducted, methods proposed based on large scale

datasets are considered for current study. In last decade, large scale datasets like NIH Chest X-ray, CheXPert and MIMIC-CXR consisting of information indicating multiple pathologies were released to public to foster research and development of methods for automated Chest X-ray image analysis.

2.4.2.1 NIH Chest X-ray dataset

Apart from binary classification of tuberculosis detection, additionally with release of large-scale public dataset NIH Chest X-ray 8 (Wang et al., 2017) (later extended to 14), significant progress has been made in this field. This dataset comprises of more than 100,000 frontal view X-ray images from 32717 patients, labelled with eight commonly observed pathologies i.e., Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia and Pneumothorax. Additionally, for benchmarking purpose, DL models using transfer learning approaches were developed to identify and localize pathologies in X-ray images. Figure 2.5 indicates overview of DL framework used for classification and generating heatmap for indicating pathologies (Wang et al., 2017). Multiple pre-trained models are used for analysis and evaluation is made based on AUC score. In this study, methods based on ResNet-50 model achieved better quantitative performance compared to other models, where “Cardiomegaly” with AUC of 0.814 and “Pneumothorax” with AUC of 0.789 are better recognized pathology groups compared to others (Wang et al., 2017). From this study, due to promising results obtained by use of transfer learning methods, use of pre-trained models became common approach for analysis of large-scale X-ray datasets.

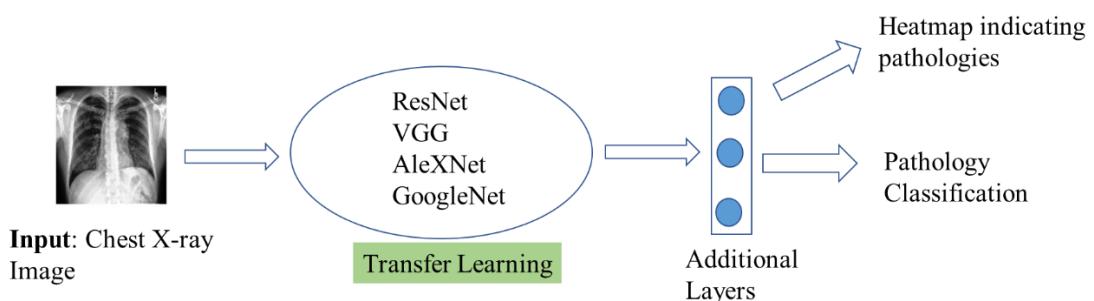


Figure 2.5: Overall framework of DL methods used for pathology classification on NIH chest X-ray dataset (Wang et al., 2017).

(Yao et al., 2017) worked on improving model performance across multiple labels further by using additional RNN decoder network and capturing inter label dependencies. (Rajpurkar et al., 2017b) developed CheXNet to mainly to classify pneumonia images using methods with

DenseNet 121-layer pretrained model and replacing last fully connected layer with a single output using sigmoid non-linearity. With this setup, authors were able to achieve performance significantly higher than radiologist on NIH Chest X-ray dataset. Further they extended model to classify multiple thorax diseases by replacing last layer with 14-layer output using element wise sigmoid non-linearity. With this change they were able to achieve state of art results on multiple thoracic diseases present in NIH Chest X-ray dataset. Figure 2.6 indicates overall approach of CheXNet model using DenseNet 121-layer pre-trained model, where input is chest X-ray image and output is binary prediction indicating presence or absence of Pneumonia (Rajpurkar et al., 2017b). After success of this research, use of DenseNet (Huang et al., 2017) pre-trained model to analyse radiology images became a common approach. However, analysis made in this study consider images acquired from a common environment setting. Hence for better generalizability of models developed, images acquired in a different hospital environment setting can be considered.

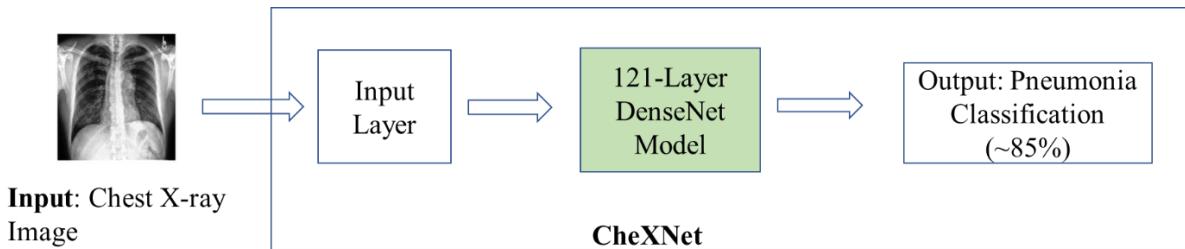


Figure 2.6: CheXNet, a 121-layer model that takes X-ray image as input and provides probability of pathology as output (Rajpurkar et al., 2017b).

(Zech et al., 2018) studied performance implications of CNN architectures on unseen external datasets acquired in different environment setting for chest radiograph images. They achieved better performance on models trained with combined datasets compared to models trained on individual datasets, using DenseNet pre-trained model. Further, they showed that for practical purpose of using CNN architecture, models need to account acquisition environment at different hospitals and train appropriately else models might be believed to be accurate than they truly are in the deployed context creating potential for patient harm. Table 2.2 indicates summary discussion of previous studies that work on automated chest X-ray analysis using NIH chest X-ray dataset.

Table 2.2: Table indicating summary of discussion made for studies using NIH Chest X-ray dataset

Reference	Remarks
(Wang et al., 2017)	Developed NIH Chest X-ray dataset and methods for detection of multiple pathologies with average AUC of 73.8%
(Yao et al., 2017)	Used methods based on CNN+RNN to improve pathology detection performance with average AUC of 79.8%.
(Rajpurkar et al., 2017b)	Used DenseNet 121-layer pretrained model with average AUC performance of 84.13%.
(Zech et al., 2018)	Used multiple datasets in combination to improve prediction accuracy for pneumonia detection.

2.4.2.2 CheXPert Dataset

Since labels in NIH Chest X-ray dataset are encoded using automatic labeller, it will be problematic to use it for benchmarking. Hence CheXPert, a large scale dataset is introduced that features radiologist-labelled test and validation sets to serve as strong reference standards (Irvin et al., 2019). On this dataset, authors experimented with multiple pre-trained models like ResNet152, DenseNet-121, Inception-v4, and SE- ResNeXt101 where they achieved optimal results using DenseNet 121-layer pre-trained model. Figure 2.7 shows overall modelling approach for determining multiple pathologies from X-ray images (Irvin et al., 2019). For evaluation metric of AUC, best and worst results are observed on Pleural Effusion (0.97) and Atelectasis (0.85) respectively. For evaluation metric of Area Under Precision-Recall Curve (AUPRC), best and worst results are observed on Pleural Effusion (0.91) and Consolidation (0.44) respectively.

(Allaouzi & Ben Ahmed, 2019) used DenseNet 121-Layer architecture for feature extraction and applied multiple problem transformation methods for classification and showed state of art result on both CheXPert and NIH Chest X-ray 14 datasets. (Majdi et al., 2020) used DL methods for detecting “Cardiomegaly” and “Lesion” pathologies on CheXPert dataset. DL methods used DenseNet 121-layer pretrained model and showed state of art results with AUC as 0.91 for “Cardiomegaly” and AUC as 0.71 for “Lesion” pathologies.

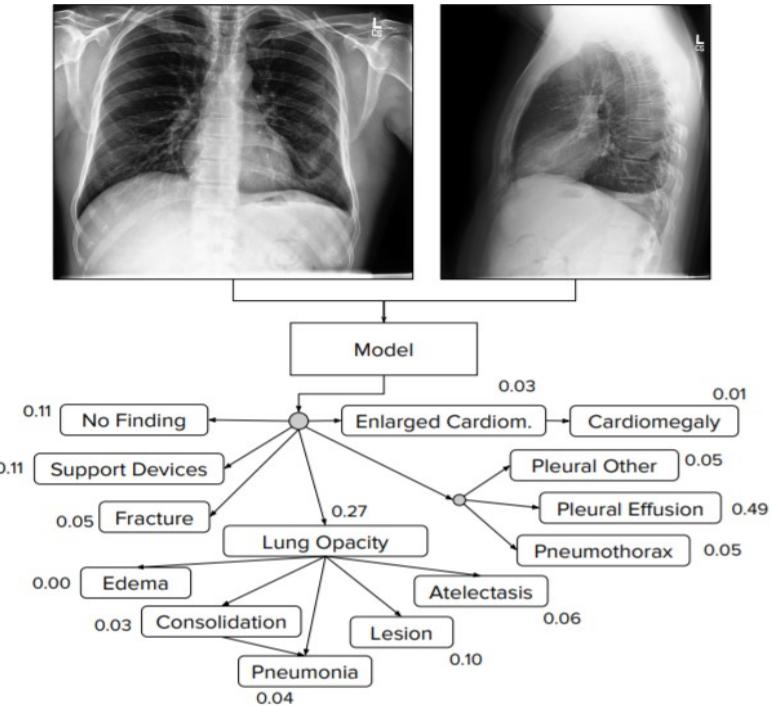


Figure 2.7: CheXPert for predicting probabilities of observations from multi-view chest X-ray images (Irvin et al., 2019).

(Pham et al., 2021) used ensemble of DL models, using information captured on inter label dependencies and using label smoothing mechanism to show average AUC value as 94%, which is promising. Based on study conducted, using pre-trained DenseNet model for initialization and further fine-tuning model for CheXPert dataset seems to commonly used approach for achieving state of art results. Table 2.3 indicates summary discussion of previous studies that work on automated chest X-ray analysis using CheXPert dataset.

Table 2.3: Table indicating summary of discussion made for studies using CheXPert dataset

Reference	Remarks
(Irvin et al., 2019)	Developed methods and CheXPert dataset for analysis of multiple pathologies with average AUC of 91% for competition tasks.
(Allaouzi & Ben Ahmed, 2019)	Developed methods that showed state of art performance with average AUC of 90% on competition tasks.
(Majdi et al., 2020)	Developed methods with AUC for Cardiomegaly as 92%.
(Pham et al., 2021)	Developed ensemble of methods that showed promising results with average AUC of 94% on competition tasks.

2.4.2.3 MIMIC-CXR Dataset

While previous studies discussed focuses on classification, segmentation, and detection of pathologies from images, MIMIC-CXR dataset is released with focus on radiology diagnosis report generation along with other analysis tasks (Johnson et al., 2019). Generating automated radiological reports has significant potential operationally and helps in improving patient care. (Liu et al., 2019) used CNN-RNN-RNN architecture for generating reports from images. First visual features are generated from input images, then topic vector is generated after which words are generated creating a report from input X-ray images. They also used Reinforcement learning methods for improving reliability of generated report. Figure 2.8 indicates overall flow from chest X-ray image as input to report generated as output (Liu et al., 2019). However, in this analysis images are considered separately and not in sequence of acquisition of images, which is a limitation of the study.

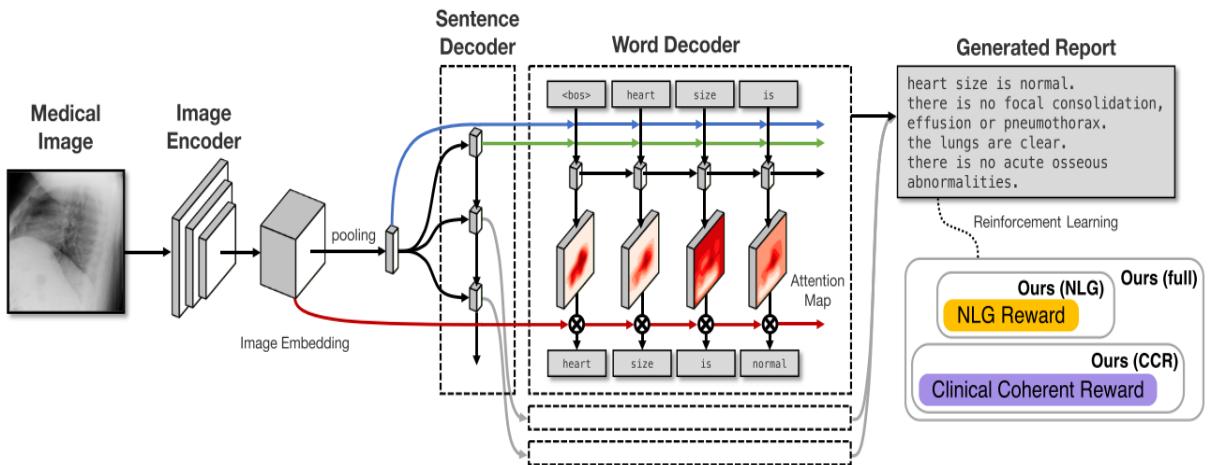


Figure 2.8: Figure indicating CNN-RNN-RNN model used for generating radiological reports (Liu et al., 2019)

(Boag et al., 2020) used simpler models for creating baseline methods to generate radiology reports. Authors observed that CNN+ Beam search is found to be extremely useful for report generation. Also, when good features are extracted from images, even simpler models like n-gram performed identical to using CNN+ Beam methods. However, since simpler models are considered for analysis, using advanced methods can help in getting better results. (Li et al., 2020) explored vision and language models like Visual BERT, Pixel BERT, LXMERT, UNIER for improving model performance using multi-modal learning.

Considering the challenges in generating proper and reliable automated radiology reports for practical setting, more advancements are yet to be made in this area. Table 2.4 indicates summary discussion of previous studies that work on radiology report generation from X-ray images.

Table 2.4: Table indicating summary of methods developed for enabling automated radiology report generation using MIMIC-CXR dataset.

Reference	Remarks
(Johnson et al., 2019)	Developed MIMIC-CXR dataset for enabling automated radiology report generation.
(Liu et al., 2019)	Developed methods to generate radiology reports using CNN-RNN-RNN methods.
(Boag et al., 2020)	Developed methods to generate radiology reports using CNN+ Beam Search approach and other simpler approaches like n-gram method.
(Li et al., 2020)	Comparative analysis of radiology report generation using different methods like BERT, is made

2.4.3 Abnormal Chest X-ray detection

In practical setting though identifying exact pathology from X-ray is useful, but abnormal chest X-ray detection would be also extremely useful. It shall help in reducing workload of radiologist without having to analyse every image. Automatic detection of abnormal X-ray images would help radiologist to focus more on pathology analysis. Several attempts have been made to address this problem.

(Tataru et al., 2017) explored using DL methods for abnormal chest X-ray identification. Authors used GoogleNet and ResNet models pretrained models for classification. They were able to improve accuracy of model by applying histogram equalization and other augmentation methods during training. They reported better accuracy with GoogleNet model compared with other models. (Tang et al., 2019) suggested on using Generative Adversarial Networks (GAN) to detect abnormal X-ray images. Authors trained GAN model on normal X-ray images, so that if abnormal X-ray images is provided as input model would fail to reproduce the image

appropriately and that image would be classified abnormal images. With this method, they were able to achieve AUC of about 84% on NIH chest X-ray dataset. Since these methods uses GAN for analysis, model performance can vary widely when input images are taken from a different setting and must be evaluated further.

(Bozorgtabar et al., 2020) proposed a model, SALAD, where it showed improved anomaly detection using Memory bank and aggregate learning methods on NIH Chest X-ray dataset. They used AUC and AUPRC as metrics for comparison with earlier models, where they were able to achieve performance exceeding state of art methods with AUC of about 91.68% and AUPRC of about 96.75%. However, since loss function used for analysis is complex and adapted for binary classification, generalizability of such methods to multiple pathologies would be difficult. (Rajaraman et al., 2020) developed methods using ensemble of models where they made coarse training on CheXPert dataset and finetuned it for predictions on RSNA chest X-ray dataset showing increased performance. However, since CheXPert dataset for abnormality detection has class imbalance problem, specificity of DL models developed is less and is not handled.

Most of proposed methods for abnormal X-ray detection use optimized loss functions for binary classification, but for future scalability purposes, considering model architectures that can scale to multiple pathologies and still achieving state of art results would be beneficial. Table 2.5 indicates summary discussion of previous studies on abnormality detection task.

Table 2.5: Table indicating summary of discussion made on previous studies for abnormality detection task.

Reference	Remarks
(Tataru et al., 2017)	Reported accuracy of 79.8% using GoogleNet model on chest X-ray dataset developed for this study.
(Tang et al., 2019)	Reported AUC of 84.1% on NIH chest X-ray dataset, using GAN methods.
(Bozorgtabar et al., 2020)	Reported AUC of 91.6% on NIH Chest X-ray dataset using optimized loss function.
(Rajaraman et al., 2020)	Reported AUC of 97.4% on RSNA chest X-ray dataset using ensemble of models.

2.4.4 Chest X-ray Multitask Learning for Pathology Analysis

Multitask Learning (MTL) is a method where model gets trained on multiple, but related tasks. MTL helps in improving generalization of model by sharing representation between tasks. In general, not necessary, when we try to optimize multiple loss functions it is called multitask learning (Ruder, 2017). They are multiple methods for approaching multitask learning like hard parameter sharing, soft parameter sharing, multi-stitch networks, sluice networks etc. Hard parameter sharing is commonly used approach for multi-task learning. However multitask learning using hard parameter sharing will break when tasks being learnt are not related, hence multiple approaches have been proposed focusing on auxiliary tasks that share and outperform hard parameter sharing approach (Ruder, 2017).

With MTL and DL methods, several attempts were made to train models on multiple tasks like localization, segmentation, and classification together for pathology classification from chest X-ray images. Performance metrics for pathology detection were shown to generalize better when multiple tasks are trained together. (Gündel et al., 2019) used multitask methods based on pre-trained DenseNet model for analysis on both segmentation and abnormality detection tasks together. Additionally, a combined dataset of NIH Chest X-ray and PLCO datasets is used for achieving improved AUC performance. Figure 2.9 indicates multitask model used for segmentation as well as abnormality detection.

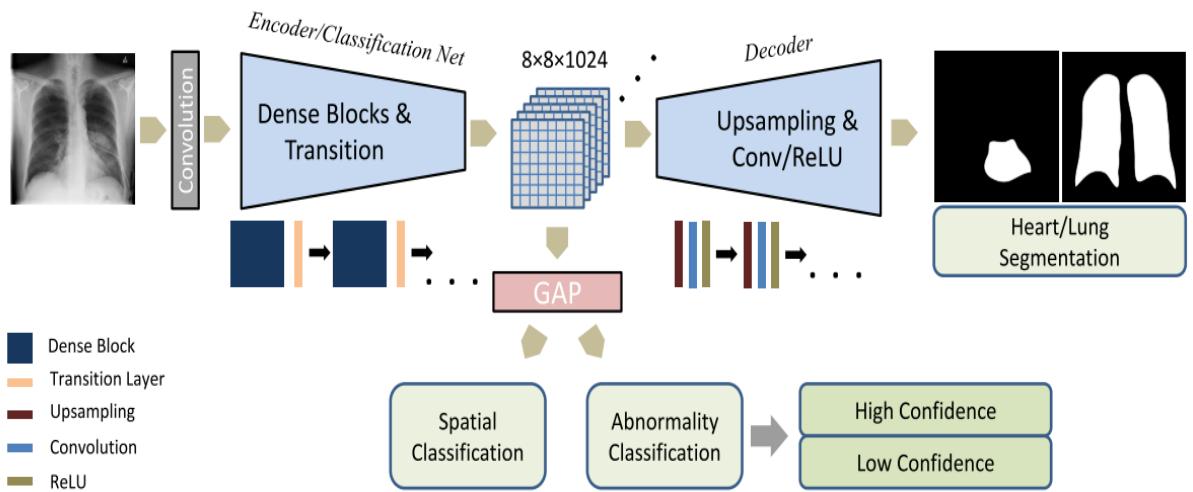


Figure 2.9: Multitask neural network jointly trained for segmentation and abnormality detection (Gündel et al., 2019).

(Imran & Terzopoulos, 2019) used novel multitask generative models for jointly learning on both classification and segmentation tasks. Complex models with new loss function are proposed for analysis of X-ray multitask methods. Use of such novel methods on large scale datasets for identifying multiple pathologies is not explored so far. MTL has also been explored to improve predictions on datasets with less samples like COVID detection. (Alom et al., 2020) used inception recurrent neural networks where they trained initial model on pneumonia classification where availability of data is more and used that model as transfer learning method for training upon COVID-19 X-ray and CT-scan methods. In addition, authors also worked on creating models for segmentation of radiographs to extract chest regions from X-ray images. With these methods, authors were able to achieve promising results of 84.7% and 98.8% test accuracy from X-ray and CT images respectively for detection of COVID-19. Though results are promising, since number of COVID-19 samples used for training and testing are less, the model wouldn't generalized well for a practical use.

MTL was also used by (Farag et al., 2020) to jointly train on multiple tasks like classification, detection and segmentation for detecting pathologies from X-ray images. Authors had used multiple stages of training due lack of available data to improve model performance. However, such methods cannot be used for direct practical use without further evaluations, due to lack of sound evaluation methodologies.

Based on study conducted, though multitask X-ray image analysis is proposed by previous studies, in general they still focus on a single task i.e., pathology analysis. Table 2.6 indicates summary discussion of previous studies on multitask methods discussed in this section.

Table 2.6: Table indicating summary of study made on multitask methods

Reference	Remarks
(Gündel et al., 2019)	MTL methods trained on classification and segmentation tasks for pathology analysis, achieving AUC of 0.883.
(Imran & Terzopoulos, 2019)	MTL model for classification and segmentation tasks using different sizes of same X-ray image as input with classification accuracy of 75.8% and segmentation Dice score as 0.976.

(Alom et al., 2020)	Model pre-trained on pneumonia detection task is trained further to detect COVID-19, with accuracy of 84.7%.
(Farag et al., 2020)	Model trained on classification, detection, and segmentation tasks for analysis on pathology information.

2.5 X-ray Age Prediction Analysis

DL has also been useful for tasks that require regression analysis from X-ray images in applications like predicting paediatric bone age from hand X-rays and biological age from chest X-rays. RSNA 2017 paediatric bone age prediction challenge was created for purpose of developing methods that can assist radiologist in automatically determining skeletal maturity of paediatric population, from X-ray images. In this challenge, Mean Absolute Difference (MAD) between reviewer's estimate and mean of all reviewer's estimates is used as metric for determining best method. (Iglovikov et al., 2018) used DL methods for predicting bone age. They used couple of approaches i.e., classification and regression methods. For classification methods they assumed bone age in month as class with 240 classes overall and used SoftMax as last layer for classification. For regression methods they used Mean Absolute Error (MAE) as loss for regression. Additionally, authors also used key point detection to identify specific regions for input to reduce computational complexity and improve model performance. However, the performance of best model reported is lesser than performance of top performing models reported in this challenge.

(Halabi et al., 2019) discusses in detail of RSNA paediatric bone age assessment challenge and performance of top models submitted. Top 3 contestants in the challenge used DL to achieve better accuracy, which shows importance of DL in analysis of different parameters from medical images apart from pathology analysis. Top 1 contestant used model ensemble during test time improved the overall performance and achieved mean absolute difference of 4.2 months for predicting bone age. Figure 2.10 indicates example of model architecture of top performing model, where X-ray image is provided as input and additionally gender information is also provided as input to Dense Layer. Further down the layers these two layers are combined to provide a age in months as single output value (Halabi et al., 2019). Since bone age growth for males will be faster than females, using gender information as additional input helped. Hence while analysing models using additional information based on problem statement can improve model performance further.

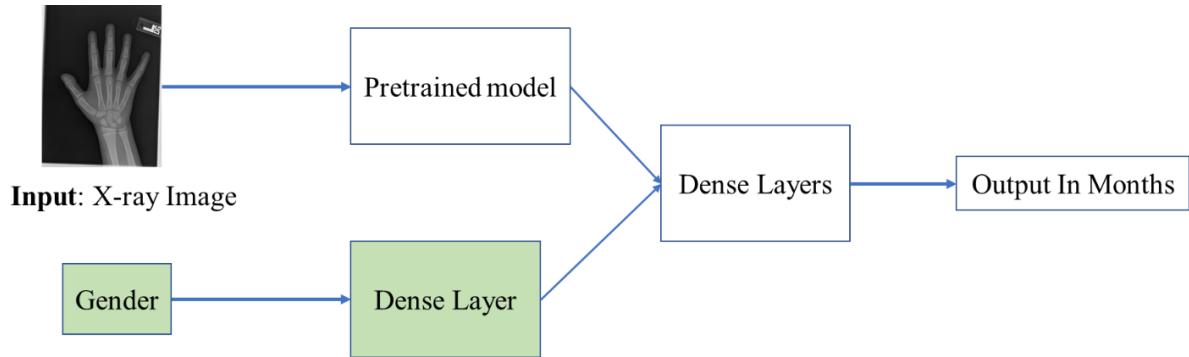


Figure 2.10: Deep Learning architecture used by top performing model in RSNA paediatric bone assessment challenge (Halabi et al., 2019).

(Pan et al., 2019) further explored on improving the model estimates reported in challenge by using different model ensemble strategies. They observed that performance has improved by using uncorrelated models. Using ensemble of different model results, they were able to reduce mean absolute difference to 3.79 months, which is significant. However, generalizability of these results on other data sources is not made, since the analysis is made on results submitted during competition and without reproducing results from models.

While most of methods proposed calculation of bone age from hand X-ray's, age prediction from chest X-ray is less explored area and can be used for preventive counselling of patients. (Karargyris et al., 2019) proposed DL methods to predict age value from chest X-ray. They used methods based on DenseNet for training on NIH Chest X-ray dataset to predict age value. Using proposed models, best recall value of 0.94 for posteroanterior (PA) view is achieved when +/- 9 years is considered as error range.

Above models considered regression methods for providing improved performance on predicting age value from X-ray images. However, creating models that can combine X-ray pathology analysis task in addition to age regression task would be extremely beneficial for proper treatment of patients than considering them as separate independent tasks. Table 2.7 indicates summary of discussion of previous studies on age prediction task.

Table 2.7: Table indicating summary of discussion of studies that work on age prediction task

Reference	Remarks
(Iglovikov et al., 2018)	Developed methods that provided mean absolute difference as 4.97 months on RSNA pediatric bone age challenge.
(Halabi et al., 2019)	Discussion is made on top performing models in RSNA pediatric bone age challenge.
(Pan et al., 2019)	Analysis is made using ensemble of models and achieved mean absolute difference of 3.79 months.
(Karargyris et al., 2019)	Developed methods that provides best recall value of 92% within error range of +/- 9 years.

2.6 Musculo-skeleton X-ray Image Analysis

Musculo-skeleton X-ray images are mainly used for determining bone fractures, which is one of most common injuries observed recently. According to (International Osteoporosis Foundation., 2018) approximately 2.68 million fractures occurs the EU6 nations, France, Germany, Italy, Spain, Sweden, and the UK. If bone related fractures are not treated on time they can lead to permanent damage or even death. (Tanzi et al., 2020) discusses on different DL methods proposed for determining different types of fractures i.e., proximal humerus fracture, orthopedic trauma, distal radius fracture, proximal and atypical femur fractures etc., from X-ray images. They further suggest upon use of large datasets like MURA and transfer learning methods for generating methods with improved performance.

MURA is a large-scale dataset consisting of Musculo-skeleton X-ray images having normal and abnormal bone images with corresponding labels manually labelled by radiologist. Figure 2.11 indicates X-ray images of some patients with normal and abnormal label indication. MURA dataset consists of X-ray images of multiple organs i.e., elbow, finger, hand, humerus, forearm, shoulder, wrist with multiple views and labels manually indicated (Rajpurkar et al., 2017a).

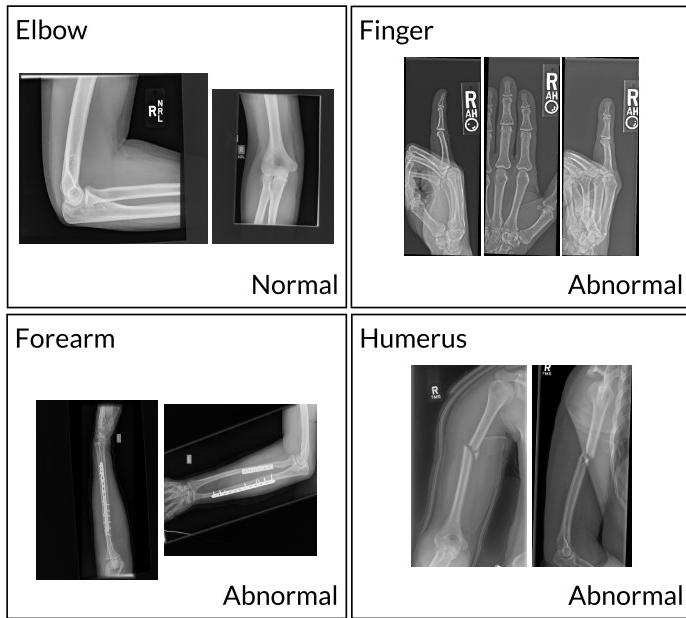


Figure 2.11: Musculo-skeleton X-ray images of patients with both normal and abnormal label indication (Rajpurkar et al., 2017a).

On this dataset additionally DL methods were developed to classify abnormal Musculo-skeleton X-ray images using DenseNet 169-layer pretrained model and weighted binary cross-entropy loss function. With this approach, authors were able to achieve AUC of 92.9% which is comparable to radiologist performance (Rajpurkar et al., 2017a). Though this analysis mainly focuses on detection of abnormal Musculo-skeleton X-ray images but combining such analysis with other X-ray images making it generic abnormal X-ray image detection would provide great benefit to radiologist. Table 2.8 indicates summary of discussion made on previous studies that work on Musculo-Skeleton X-ray Analysis.

Table 2.8: Table indicating summary of study made on Musculo-Skeleton X-ray Analysis

Reference	Remarks
(Rajpurkar et al., 2017a)	Developed MURA dataset and methods for detection of abnormal Musculo-Skeleton X-rays, with AUC of 92.9%.
(Tanzi et al., 2020)	Discussion on state of art methods for fracture detection from X-ray images.

2.7 Research Gaps Identified

Even though analysis of medical images is a difficult task to consider, performance of DL methods has shown to exceed prior state of art results. Several DL and CNN methods were proposed to identify pathologies, determine skeletal maturity, create radiology diagnosis report etc., exceeding state of art performance. This shows importance using DL methods in enabling automated diagnosis of X-ray images for assisting radiologist.

However, multiple research gaps are identified from the study conducted. Most of prior methods were attempting to automate analysis of specific task of X-ray image analysis i.e., either chest X-ray pathology analysis or age regression tasks are considered separately for analysis. Though multitask methods are proposed, their focus is generally on pathology analysis. They don't intend to understand patient diagnosis completely like radiologist who looks at process of patient diagnosis with multiple perspectives. Also, it is observed that performance of methods that work on generating radiology reports from X-ray images is not optimal and still need further improvement to be suitable for practical setting.

Focus of current thesis is addressing identified research gap of analysis of X-ray images from multiple perspectives enabling better diagnosis of patients and supporting radiologist to provide right treatment to right patient at right time. Like a radiologist analyzing X-ray images from multiple perspectives, current thesis works on investigation of methods that concurrent analyses different tasks from X-ray images. For this study, abnormality detection and age prediction tasks are considered for multitask analysis. Using such a tool which can predict age value in addition to abnormal X-ray detection will support radiologist in providing patient specific diagnosis.

2.8 Summary

So far in this chapter, detailed discussion is made on systematic study conducted outlining important research research works for automating medical image analysis, with further focus on X-ray imaging. Multiple methods that can analyse patient's X-ray for pathology information are discussed. Different methods are developed for analysis of applications such as Tuberculosis screening, Pneumonia detection, multi-pathology analysis, abnormal X-ray detection, radiology report generation etc. Multitask methods that analyse X-ray images to

provide improved pathology analysis are discussed. Further, DL methods that work on predicting age value from X-ray are discussed.

Finally, research gaps identified from study conducted is discussed. Though previous works discussed in this chapter doesn't consider handling patient diagnosis from multiple perspectives, yet they provide invaluable insight to our research by providing information on approaches like use of transfer learning methods, multitask methods, sample augmentation strategies for analysing large-scale medical image datasets. Further, research methodology section explains dataset and methods being considered in this thesis for addressing the research gaps identified.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

X-ray is commonly used diagnostic procedure compared to other imaging procedures. Currently deep learning (DL) methods are state of art approaches for exploring automatic radiograph analysis, whose performance approximates to radiologist. Based on literature study conducted in area of automated X-ray image analysis, as discussed in Chapter 2, it is observed that multiple deep learning methods were proposed focusing on analysis of different workflows like X-ray pathology analysis, abnormality detection, prediction of parameters like bone age from Musculo-skeleton X-rays and patient's age from chest X-rays as independent tasks. So far little research is made on methods that can concurrently determine both pathology analysis and Age prediction tasks together. Current thesis focuses on addressing this gap.

Current thesis proposes use of multitask deep learning methods for abnormality detection and age value prediction from X-rays. Additionally, CheXPert dataset will be used for investigation purposes for this research. Figure 3.1 discusses on overall flow of research methodology used for conducting this study. Idea is to process the CheXPert dataset to indicate abnormality label for X-ray record, where records with at least one pathology observed are marked as abnormal else normal. Further DL models are developed that can train on the processed dataset to concurrently determine abnormal X-ray records and predict age value.

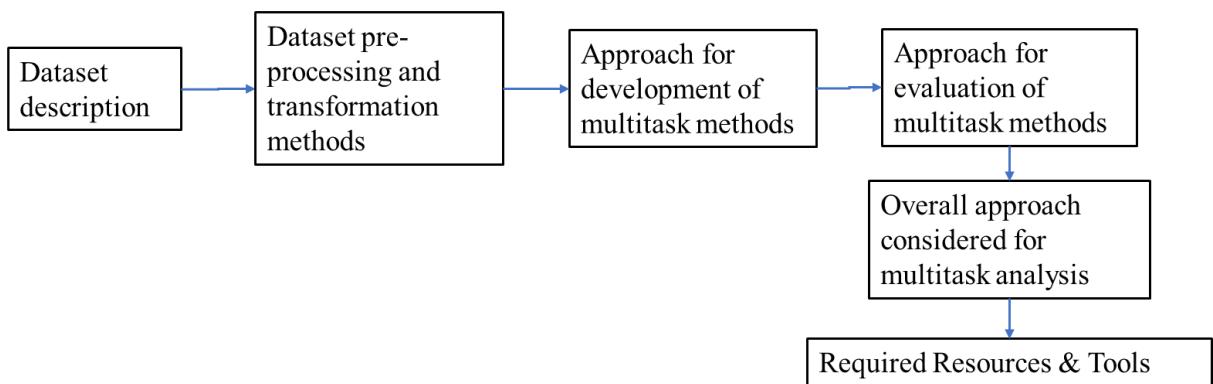


Figure 3.1: Overall flow of research methodology in this study

Models developed for this study consider accommodating additional parameters for future use cases like multi-pathology multitask analysis, hence loss functions and other hyperparameters are chosen accordingly. Further, methods developed for this study are evaluated to determine best possible model that can provide improved performance on both abnormality detection task and age value prediction task together.

3.2 Dataset Description

Large-scale datasets need to be considered for developing methods to provides performance results that are stable and can generalize well. CheXPert (Irvin et al., 2019) is a large-scale dataset that contains 224,316 chest radiographs of 65,240 patients collected from radiographic studies at Stanford Hospital between October 2002 and July 2017 along with associated radiology reports. It is a largescale chest radiograph dataset released for research purpose 1) with large database 2) having strong reference standards and with expert human performance metrics for comparison.

CheXPert dataset includes X-ray images of patients taken in multiple studies with multiple views i.e., frontal, and lateral views. These images also include AP and PA views. Also, as part of the released dataset information, radiology reports of patients are processed to indicate 12 pathologies in addition to “Supported Devices” and “No Finding”. Metadata information includes Age, Gender of patient along with each record.

3.2.1 Pathology information

CheXPert dataset includes pathology information of 12 commonly observed pulmonary diseases and additionally “Supported Devices” and “No Finding” observations. These observations were created by processing radiology reports corresponding to X-rays using a labeller. Table 3.1 indicates observations made on dataset against each pathology (Irvin et al., 2019). For each record, information specific to 14 observations is recorded as positive, negative, and uncertain. Blank indicates no information found for that observation. For example, a record with positive marked against “Cardiomegaly” label indicates definite finding of “Cardiomegaly” observation in the report, negative marked against “Cardiomegaly” label indicates no evidence of “Cardiomegaly” observation in the report and uncertain marked against “Cardiomegaly” indicates uncertainty of “Cardiomegaly” observation in the report.

Records which doesn't have pathology classification as positive or uncertain for all indicated pathologies are marked as positive against label "No Finding".

Table 3.1: CheXPert dataset consisting of 14 Labelled Observations (Irvin et al., 2019).

Pathology	Positive (%)	Uncertain (%)	Negative (%)
No Finding	16627 (8.86)	0 (0.0)	171014 (91.14)
Enlarged Cardiom.	9020 (4.81)	10148 (5.41)	168473 (89.78)
Cardiomegaly	23002 (12.26)	6597 (3.52)	158042 (84.23)
Lung Lesion	6856 (3.65)	1071 (0.57)	179714 (95.78)
Lung Opacity	92669 (49.39)	4341 (2.31)	90631(48.3)
Edema	48905 (26.06)	11571 (6.17)	127165(67.77)
Consolidation	12730 (6.78)	23976 (12.78)	150935(80.44)
Pneumonia	4576 (2.44)	15658 (8.34)	167407 (89.22)
Atelectasis	29333(15.63)	29377 (15.66)	128931 (68.71)
Pneumothorax	17313 (9.23)	2663 (1.42)	167665 (89.35)
Pleural Effusion	75696 (40.34)	9419 (5.02)	102526 (54.64)
Pleural Other	2441 (1.3)	1771 (0.94)	183429 (97.76)
Fracture	7270 (3.87)	484 (0.26)	179887 (95.87)
Support Devices	105831 (56.4)	898 (0.48)	80912(43.12)

3.2.2 Additional Metadata Description

Feature information like frontal/lateral view, gender, patient age is included in the dataset as additional metadata. This information is used for analysis where frontal/lateral X-ray views and gender can be used for filtering data or as additional inputs to DL models for better analysis. Age value used in this thesis is considered as an output variable.

3.2.2.1 Frontal and Lateral X-ray view

Frontal views are images acquired by directing X-rays on back/front of patient's chest. Lateral views are images acquired from side of patient body. Figure 3.2.A indicates frontal and lateral chest X-ray images of a patient in CheXPert dataset. Figure 3.2.B indicates proportion of total frontal and lateral views of a patient in the CheXPert dataset, indicating number of frontal views are significant compared to lateral views. Since this study also considers NIH chest X-ray

dataset for evaluating task performance and NIH chest X-ray dataset contains only frontal X-ray records, frontal records from the CheXPert dataset are only considered for this study.

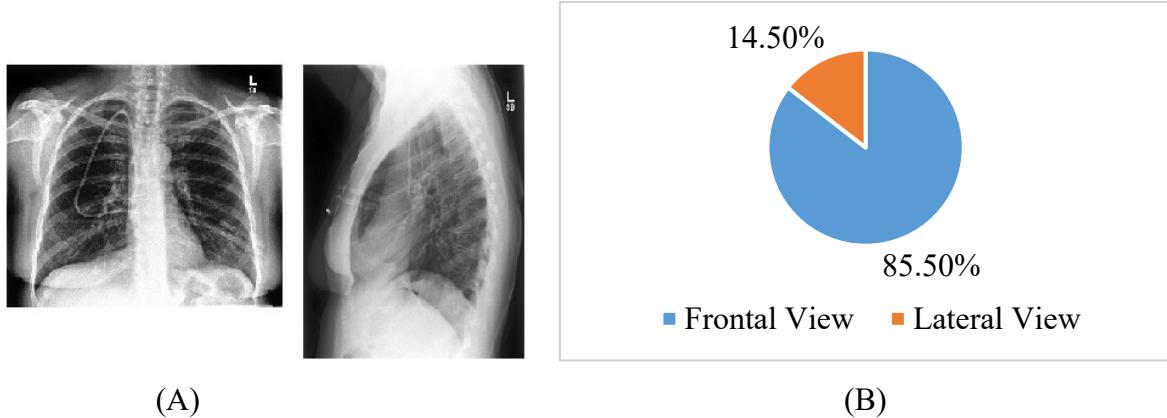


Figure 3.2: (A) Frontal and lateral chest X-ray views of a patient. (B) Proportion of frontal and lateral X-ray views.

3.2.2.2 Patient's Age

Metadata in the CheXPert dataset includes information about patients age for both normal and abnormal X-ray's. Figure 3.3.A shows age distribution of all patients in the dataset indicating most of patients who went for radiographic study are in 55-65 age group. Figure 3.3.B shows age distribution of patients with “No Finding” observation, which indicates proportion of patients with “No Finding” observation is more relevant in patients with less than 40 years.

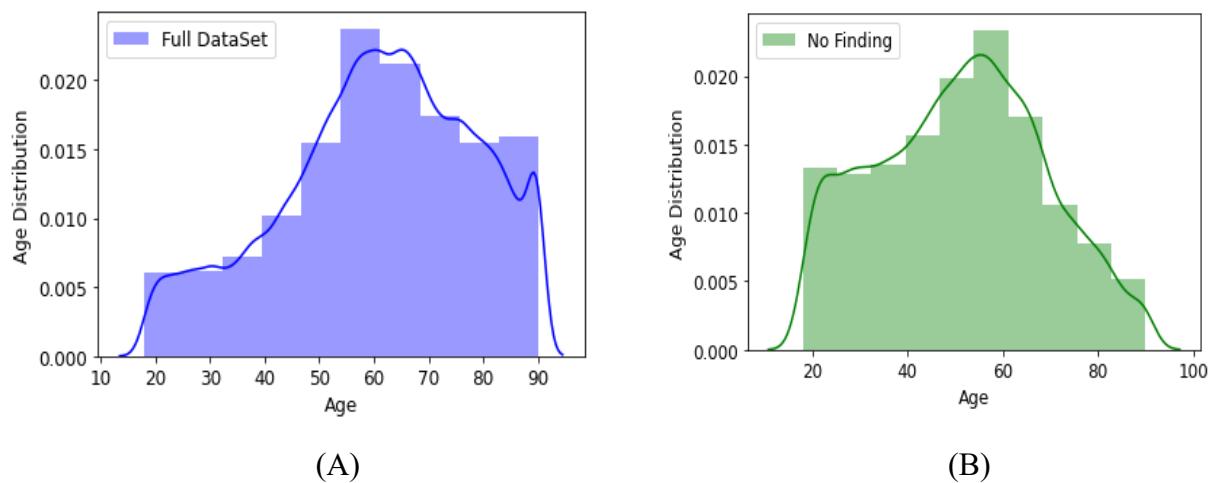


Figure 3.3: (A) Age Distribution of patients in dataset. (B) Age distribution of patients with “No Finding” label in dataset.

3.2.2.3 Gender

Gender information corresponding to each radiograph is indicated in the metadata. Figure 3.4 indicates proportion of male and female patients in the dataset. Proportion of records with male patients is more than female patients in CheXPert dataset.

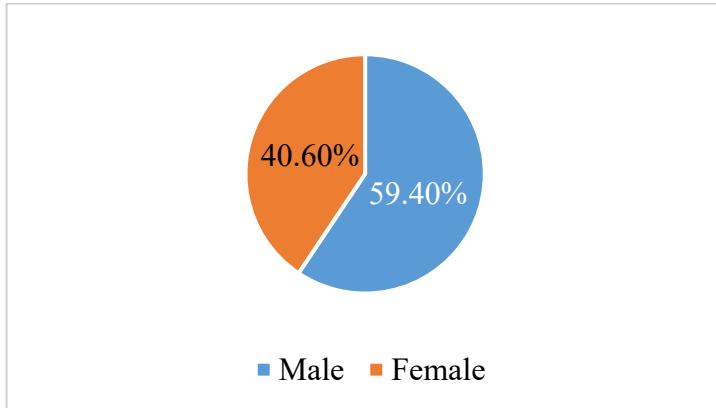


Figure 3.4: Proportion of Male and Female patients with radiographic examination

3.3 Data Preprocessing and Transformation Methods

One objective of this research is abnormality detection from chest X-rays. For abnormality detection, current pathology information is used to create a label indicating normal or abnormal X-ray records. If any of the pathology is marked as positive for an X-ray record then that record is marked as abnormal, indicating that X-ray needs to be further analysed by radiologist. If none of the pathologies are marked as positive, then that record is indicated as normal (Irvin et al., 2019). Pseudo code B.1 (in Appendix B) explains process of mapping pathology information in dataset to abnormal label. However, after marking records with abnormal label, additional processing is needed for records marked as uncertain case, where they can be taken as normal or abnormal or ignored (Irvin et al., 2019). Pseudo code B.2 (in Appendix B) indicates processing of mapping uncertain case to either normal or abnormal classes.

Figure 3.5 indicates count of ground truth labels formed after transformation of pathology information to abnormal label using different transformation methods and after filtering data with frontal view X-ray images. Uncertain cases are handled as normal (as 0's) or abnormal (as 1's) or as separate uncertain class (ignored as this study is a binary classification problem). In the CheXPert dataset, there are around 6000 records which doesn't have ‘No Finding’ label

marked as 1 though none of the pathologies are marked as 1 or -1. This is inconsistent with description of dataset (Irvin et al., 2019), hence removing those records to remove any ambiguities during analysis.

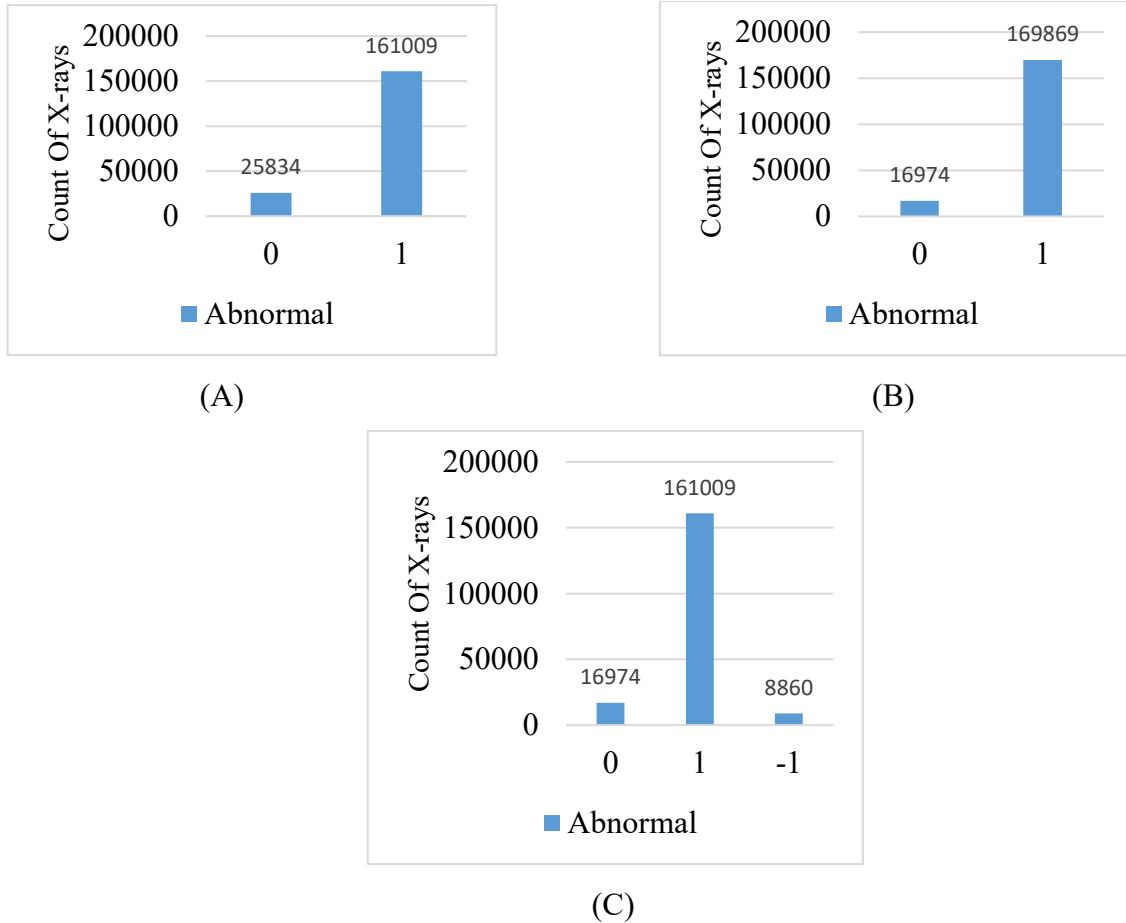


Figure 3.5: (A) Uncertain cases marked as normal i.e., 0 (B) Uncertain cases marked as abnormal i.e., 1 (C) Uncertain cases left without change i.e., -1

3.4 Model Development

Current thesis proposes creating multitask methods for analyzing both abnormality detection and age prediction tasks from patient X-ray images. As discussed in Chapter 2, transfer learning approach of using Convolutional Neural Network (CNN) methods, trained on largescale dataset like ImageNet, has produced state of art performance in analyzing medical images with performance approximating to humans. Current thesis proposes use of transfer learning methods for analysis. Since for this study multiple tasks are used for training, multitask DL methods are developed using transfer learning approaches for analysis of those tasks.

3.4.1 Using Transfer Learning Methods for Analysis

Transfer learning is a process where methods developed in one domain can be used for training in other domain, like methods developed for classification of natural images can be used for training of models to classify abnormal medical images. This helps in better scalability and faster training process. For example, methods like CheXNet (Rajpurkar et al., 2017b) successfully used transfer learning approach to get improved performance results on Pneumonia classification using NIH Chest X-ray dataset (Wang et al., 2017). Figure 3.6 indicates overall approach of transfer learning process at high level, where model trained on ImageNet database (Jia Deng et al., 2009) excluding input and output layers is used for training on target X-ray image dataset for pathology detection. As discussed in Chapter 2, pre-trained DenseNet (Huang et al., 2017) model is observed to commonly used transfer learning approach for medical image analysis. Since DL methods developed using DenseNet (Huang et al., 2017) have large parameters and are computationally intensive, methods using MobileNet (Howard et al., 2017) with fewer parameters are additionally considered for analysis to determine optimized approach. Hence, for this study, pre-trained DenseNet (Huang et al., 2017) and MobileNet (Howard et al., 2017) model will be used for developing and evaluating multitasking methods.

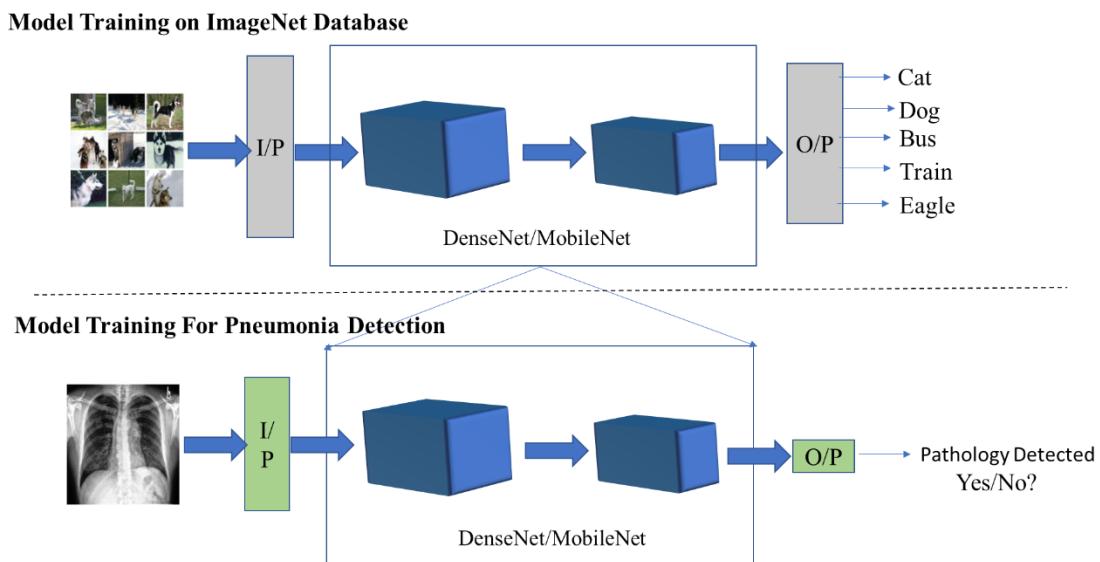


Figure 3.6: Overall approach of using transfer learning for pathology analysis

CNN's are efficient to train and accurate if shorter connections are made between layers. Based on this Dense Connected Convolutional Network (DenseNet) is developed where all layers in a dense block are connected. This helps in better flow of gradients during training and improved

feature representation. Figure 3.7 indicates a dense block with interconnected layers (Huang et al., 2017). With change in number of layers L and growth rate (K) i.e., number of feature maps in layer, model capacity can be increased with improved representation of input data.

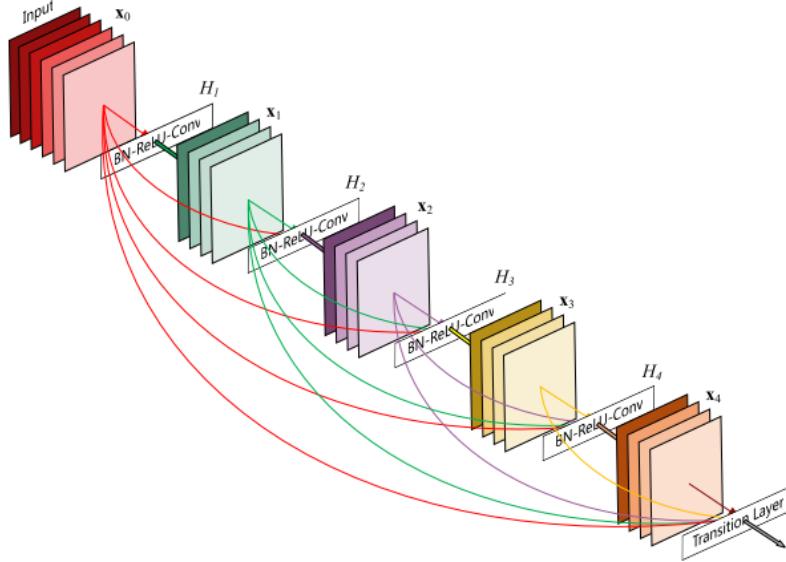


Figure 3.7: Dense block with interconnected layers (Huang et al., 2017).

3.4.2 Multitask methods for abnormality classification and age prediction

Multitask DL methods help in learning independent but related tasks together so that model performance is improved on balanced accuracy of all the tasks. Figure 3.8 shows overall modelling approach of multitask methods developed for this thesis. Two tasks i.e., abnormality detection and age prediction are jointly trained using DL methods. Unlike previous multitask approaches which work on pathology analysis, multiple tasks considered for this study are separate independent tasks, but related when considered from multi-perspective analysis of X-ray images. Hence, obtaining optimal balanced accuracy on these tasks using a single multitask method on a large scale X-ray image dataset is a challenging task. Multi-task methods being developed for this analysis shall use DenseNet pre-trained model or a simple CNN model like MobileNet with fewer parameters. This shall help in appropriate comparison of methods for determining optimal model addressing multitask needs considered in this study.

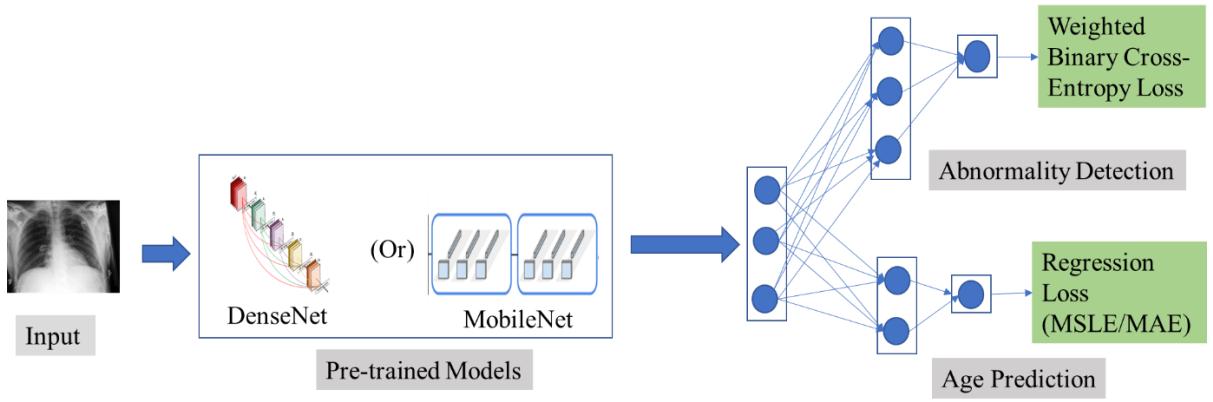


Figure 3.8: Overall modelling approach for multitask model to detect abnormal X-ray and predict age value from X-ray.

In this thesis, for multitask methods, weighted binary cross-entropy loss for abnormality detection task and regression loss functions for age prediction task are used for analysis. From pre-processing state discussed in section 3.3, it is observed that there is imbalance of records where more abnormal labels are present compared to normal labels. Hence weighted binary cross entropy loss is used for abnormal classification of X-ray images.

$$\text{Loss} = - \left[w^+ \cdot y \cdot \log p \left(Y = \frac{1}{X} \right) \right] - \left[w^- \cdot (1 - y) \cdot \log p \left(Y = \frac{0}{X} \right) \right] \quad \text{Equation 3.1}$$

where $p(Y = \frac{0 \text{ or } 1}{X})$ is the probability that the network assigns to the label 0 or 1 , $w^+ = \frac{|N|}{|P|+|N|}$ and $w^- = \frac{|P|}{|P|+|N|}$ with $|P|$ and $|N|$ are number of positive and negative abnormal cases respectively (Rajpurkar et al., 2017b).

Age value shall be trained using Mean Squared Logarithmic Error (MSLE) or Mean Absolute Error (MAE) loss functions.

$$\text{MAE}(Y, X) = \frac{1}{N} \sum_{k=0}^N (Y_k - X_k) \quad \text{Equation 3.2}$$

$$\text{MSLE}(Y, X) = \frac{1}{N} \sum_{k=0}^N (\log(Y_k + 1) - \log(X_k + 1))^2 \quad \text{Equation 3.3}$$

3.5 Model Evaluation

CheXPert dataset provides around 234 X-rays as validation data. Since test data from CheXPert dataset is not publicly available, this data shall be used for testing this model. This data is manually annotated by 3 board-certified radiologists and using this data for evaluation purposes makes this research reliable in relation to practical setting and helps for proper evaluation of performance of models being developed. Out of these X-ray images, 39 X-ray images doesn't have "No Finding" label set even though none of pathology information is positive (marked as 1) or uncertain (marked as -1) which is not consistent with description in (Irvin et al., 2019). Hence removing these records for evaluation purposes to avoid ambiguity. Further after processing, where only frontal X-ray records are considered, this data shall be used for testing the performance of models created. Figure 3.9 indicates count of records indicating normal or abnormal labels in data available for testing model performance.

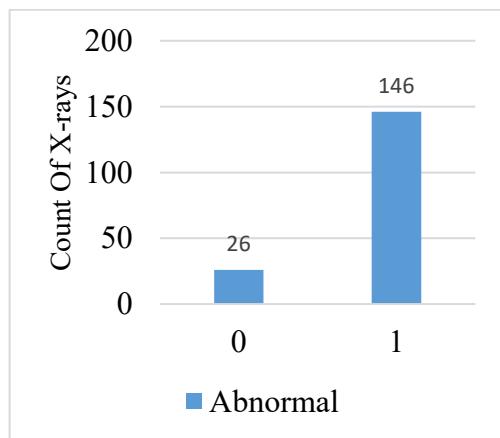


Figure 3.9: Records indicating count of normal or abnormal X-rays available for testing the model.

Model performance evaluation is based on balanced accuracy metrics of abnormality detection and age prediction, where metrics for

- Abnormal X-ray detection task is Area Under Receiver Operating Characteristic Curve (AUC) and Area Under Precision Recall Curve (AUPRC), Weighted Avg. F1 score, Weighted Avg. Recall.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{Equation 3.4}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$
Equation 3.5

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$
Equation 3.6

- Age prediction task is measured using Mean Absolute Error (MAE) and % of records with absolute prediction error within 9 years as reported by (Karargyris et al., 2019).

3.6 Approach for Multitask Investigation Analysis

Figure 3.10 indicates overall process followed for development of multitask methods in this study. Multiple methods are developed using different processed datasets and different pre-trained models for a proper evaluation and comparison. Additionally, as discussed in Chapter 4, multitask methods using combination of pre-trained models (DenseNet (Huang et al., 2017), MobileNet (Howard et al., 2017) and ResNet (He et al., 2016)) are developed for further enhancing performance on age prediction and abnormal detection tasks together, approximating state of art.

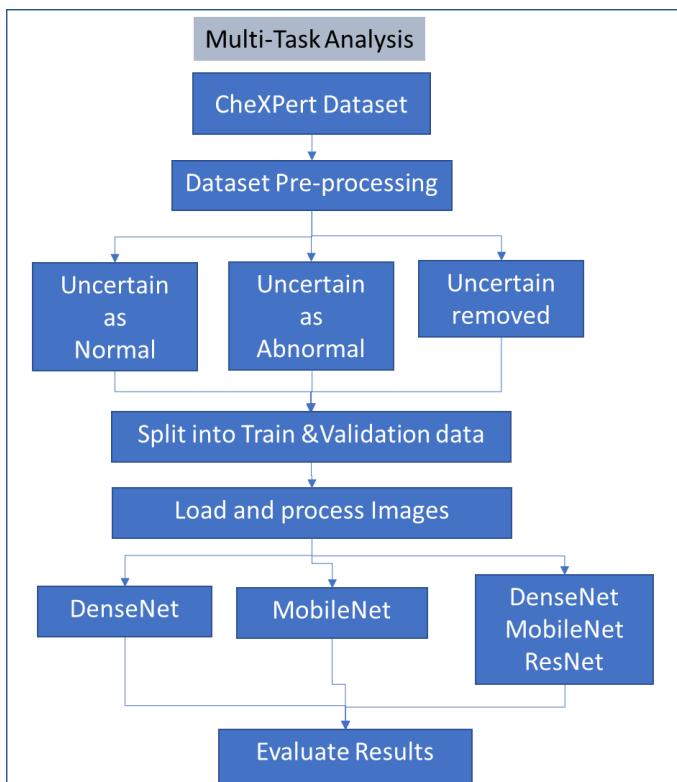


Figure 3.10: Figure indicating process of investigation followed for developing multitask methods.

3.7 Required Resources and Tools

This section discusses about required resources like datasets, hardware resources and tools used for this study.

3.7.1 Resources

CheXPert (Irvin et al., 2019), large-scale publicly available dataset is used for investigation purposes to address research gaps. Additionally, NIH Chest X-ray dataset (Wang et al., 2017) is used for comparative performance evaluation on age prediction task. For training DL models on large-scale datasets, resources with more computational capability are needed. Hence for this research investigation, CPU with 4 cores, at least 16GB RAM and GPU of Nvidia Tesla P100 is used.

3.7.2 Tools and Libraries

Python language and Jupyter notebook used are used for project development. Libraries like TensorFlow, Matplotlib, Seaborn, Pandas, NumPy, Scikit-learn are used for developing, training, and evaluating multitask methods in this study.

3.8 Summary

This chapter discusses on research methodology approach that is used for conducting investigation to address research gaps identified. Discussion is made on pathology information and metadata information present in the CheXPert dataset. Since records with uncertain class are present in the dataset, discussion is made on transformation methods applied for records with uncertain class, where records with uncertain class can be treated as normal or abnormal or ignored. These different processed datasets are used for training and evaluating multitask DL methods developed in this study.

Further, discussion is made on development of multitask methods. Transfer learning using different pre-trained models is considered for development of multitask methods to concurrently analyse abnormality detection and age prediction tasks. Discussion is made on different loss functions used for training multitask methods. Discussion is made on evaluation of multitask methods with different metrics used for evaluating abnormality detection and age prediction tasks performance.

Overall approach followed for development and evaluation of multitask methods is indicated along with required resources and tools used for development of multitask methods. Chapter 4 further discusses about investigation conducted along with implementation details for developing multitasking methods to meet aim and objectives of this study.

CHAPTER 4

ANALYSIS AND IMPLEMENTATION

4.1 Introduction

While Deep Learning (DL) methods has been extremely useful in analysis of independent tasks using medical image, using these methods for analyzing multiple different tasks from medical images is challenging and less explored. Chapter 3 discusses about research methodology approach that is used for investigation in this study to address research gaps identified. This chapter discusses on design and implementation of multitask methods being developed in this study, to meet proposed aim and objectives. Additionally, to evaluate advanced use of multitask methods developed in this study, multi-pathology multitask DL methods are developed that can analyze multiple pathologies along with age prediction task from X-ray images. Brief discussion is made on approaches for splitting data into train and validation sets and approach to process dataset for training multitask DL methods.

4.1.1 Overall Flow of Research Investigation

During experiments, it is observed that in multitask environment considered for this study, age value prediction from Chest X-ray images using CheXPert dataset became challenging and performance on age prediction task of the multi-task models developed is less. Since previous state of art approach (Karargyris et al., 2019) used NIH Chest X-ray dataset, age prediction task analysis is made on both NIH Chest X-ray (Wang et al., 2017) and CheXPert (Irvin et al., 2019) datasets for a proper comparative evaluation. Learnings from analysis on age prediction task and abnormal classification task is used as inputs for further developing multitask methods that can determine both abnormal X-rays and predict age value simultaneously. Further, multi-pathology multitask analysis methods are also discussed to evaluate advanced use of the multitask methods discussed in the thesis. Figure 4.1 indicates overall approach of analysis considered for this study.

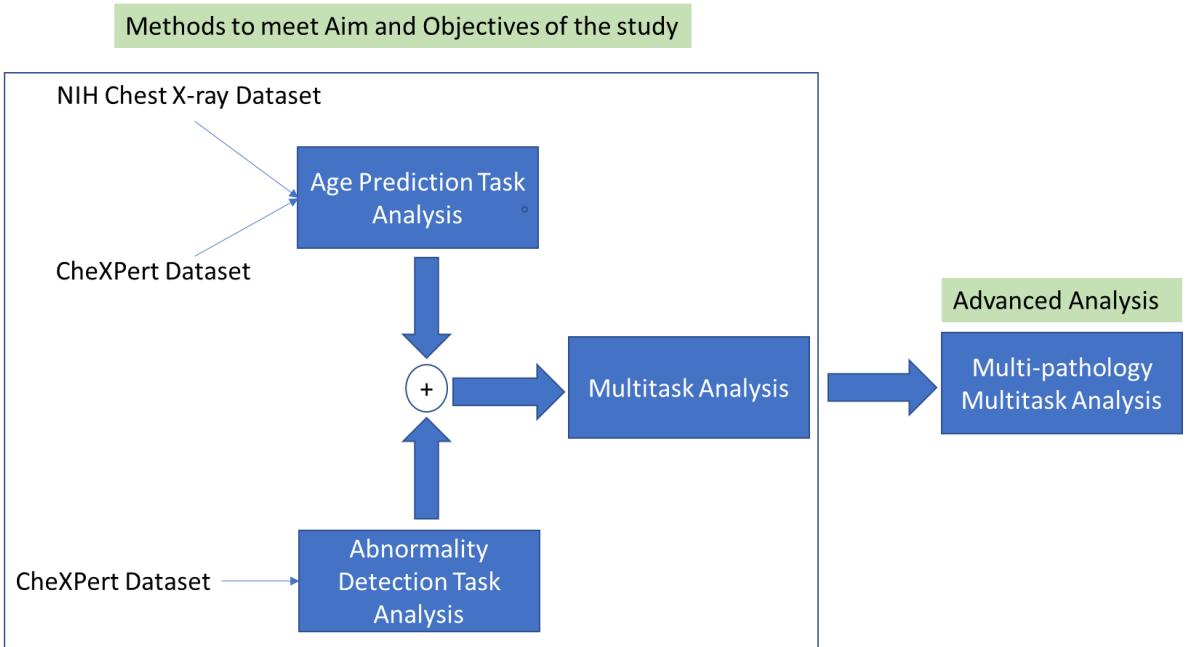


Figure 4.1: Overall approach of analysis made for this study

4.2 Data Processing

This section explains on approach for processing information for loading data to models being trained. Before sending data as input to models for analysis, dataset is split into train and validation sets. Data generators are used for loading patient X-ray images and supporting data into models being trained for predictions.

4.2.1 Approach for Data Splitting

CheXPert dataset has chest X-ray records of 64540 patients and with multiple studies present for each patient. For this study, splitting of dataset is made based on patient ID where there is no overlap between train and validation subsets (Tang et al., 2019). This helps in getting generalized models where performance of model is not biased by patients' records that are split between train and validation subsets. Pseudo code B.3 (in Appendix B) provides information on how the dataset is split into train and validation datasets based on patient ID.

4.2.2 Data Generators

For purpose of investigation in this study Tensor flow datasets are used. Figure 4.2 shows process followed for loading data to model during training. Patient X-ray Images are loaded

and resized before providing input to model during training process. Data is loaded, buffered, and shuffled, so that model doesn't get bias by input sequence of data. Then data is extracted in batches based on batch size specified and provided as input to model during training. However, for validation and test data generators, data shuffling is not applied, and input is provided to model in loaded data sequence.

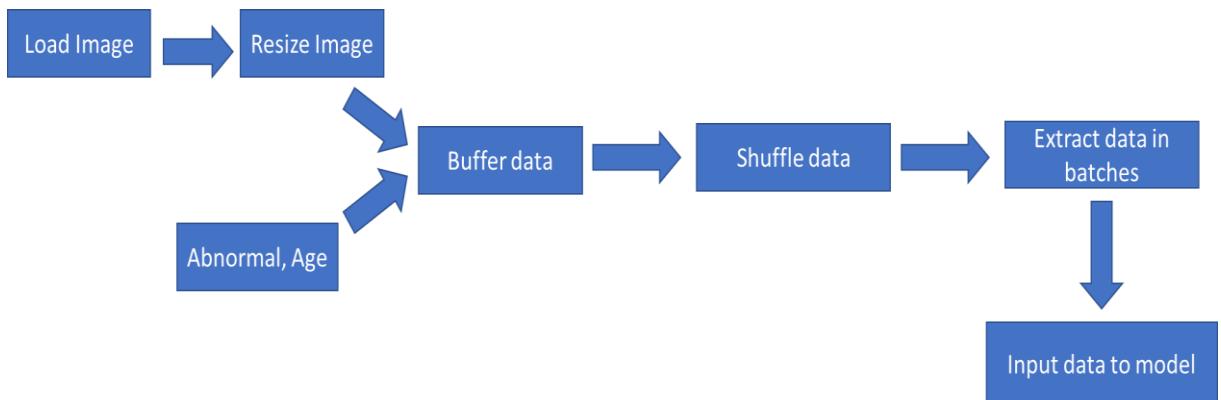


Figure 4.2 : Overall flow of data processing before providing information to model during training process.

4.3 Model Development

Multiple methods are developed for analysis purposes, where multitask methods that can analyze both age prediction and abnormal X-ray classification tasks are developed. Additionally, methods that analyze abnormal X-ray image classification and age prediction tasks independently are also developed for comparative analysis with multitask approaches being developed as part of this study. Further, in this section, discussion is made in sequence on design and implementation of methods for abnormality detection task, methods for age prediction task, methods for multitask approaches that concurrently work on abnormality detection and age prediction tasks, followed by methods for multi-pathology multitask analysis.

4.3.1 Methods for Abnormality Detection Task Analysis

As discussed in chapter 3, current investigation approaches use transfer learning methods through pre-trained DenseNet (Huang et al., 2017) or MobileNet (Howard et al., 2017) models for abnormal detection of X-ray images. Pre-trained models on ImageNet database from TensorFlow-Keras package are used, where the last layer is replaced with Dense Layer using

Sigmoid activation. Figure 4.3 shows visualization of methods developed for analysis on abnormality detection task from chest X-ray images.

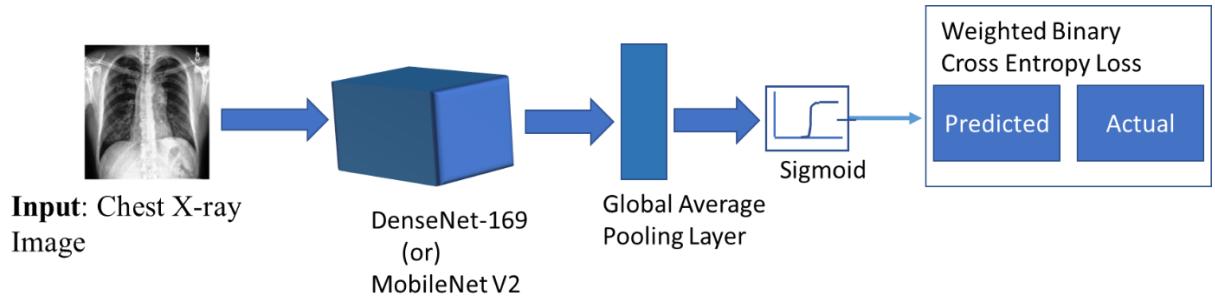


Figure 4.3: Model architecture for abnormality detection from chest X-ray images

When using pre-trained models for analysis, initial layers of model are frozen from training and parameters of last layers of model are set to be trainable. This approach is considered because in neural network models, initial layers work on generic features and layers at the end of model are specifically tailored for current problem under consideration. This helps in reducing training time, as all the layers need not be trained. Pseudo code B.4 (in Appendix B) indicates model development process where last layers of DenseNet model are set as trainable, while initial layers are frozen from training on X-ray abnormality classification task.

4.3.2 Methods for Age Prediction Task Analysis

Performance of multitask methods on age prediction task is observed to be less than reported state of art value, when CheXPert dataset is used for analysis. To analyze this problem, previous model architecture (Karargyris et al., 2019) that reported state of art, in predicting age value from chest X-ray images, is used and taken as benchmark for comparative analysis. As discussed in chapter 2, (Karargyris et al., 2019) developed DL methods that can predict age value from chest X-ray images on NIH chest X-ray 8 dataset (Wang et al., 2017). Authors reported more than 90% performance on age prediction task, with metric as % of records with absolute prediction error within 9 years. In this section, since CheXPert dataset description is made in detail in Chapter 3, a brief description on NIH chest X-ray dataset is also made. Further, discussion is made on methods developed for comparative performance analysis on age prediction task, using both NIH chest X-ray and CheXPert datasets.

4.3.2.1 NIH Chest X-ray Dataset description

For a fair comparison with analysis on CheXPert dataset which consists of information regarding 14 pathologies, NIH Chest X-ray 14 dataset is used. NIH Chest X-ray 14 dataset is based on NIH chest X-ray 8 dataset and extended to include pathology information for 14 thoracic diseases. This dataset contains approximately 112000 X-ray image records of 30805 patients with information regarding 14 pathologies and with patient age as a parameter in metadata of each record. Figure 4.4 indicates information on age distribution of patients in NIH chest X-ray 14 dataset.

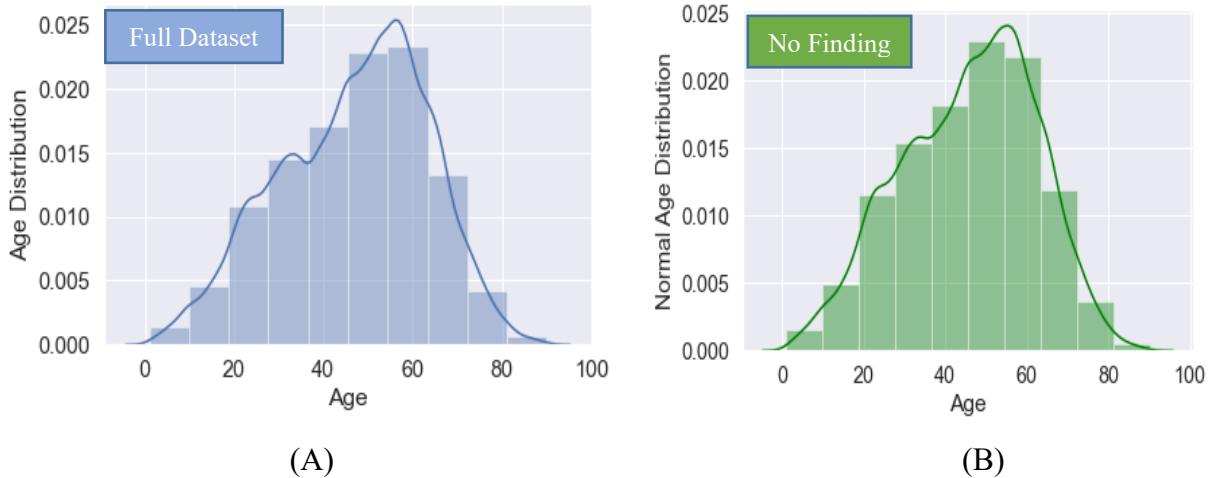


Figure 4.4 : (A) Figure indicating age distribution of all patients. (B) Figure indicating age distribution of patients with normal X-ray record.

For this study, though purpose of using NIH chest X-ray dataset is mainly for analysis of age prediction task, pathology information is needed for comparative analysis with multitask methods. Hence, as part of processing NIH chest X-ray dataset, pathology information is converted to abnormal label where records with “No Finding” label are indicated as 0 otherwise they are indicated as 1. Figure 4.5 indicates total number of normal and abnormal records present in NIH chest X-ray dataset, indicating better balanced classes compared to CheXPert dataset.

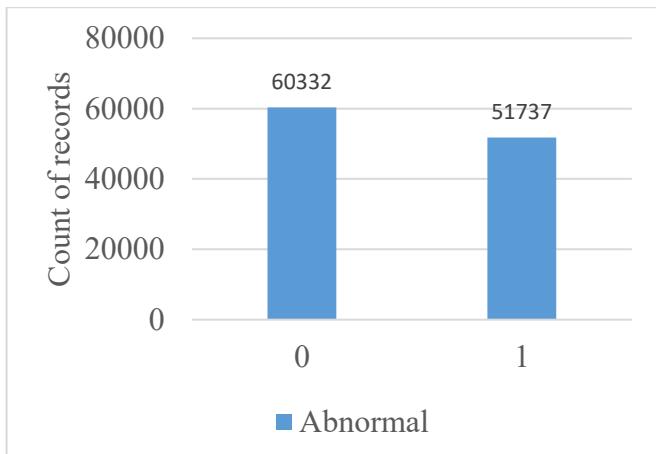


Figure 4.5 : Figure indicating number of normal and abnormal X-ray images in NIH chest X-ray dataset

4.3.2.2 Comparative Analysis Study on Age Prediction Task

(Karargyris et al., 2019) discusses on model architecture for predicting age value from NIH Chest X-ray 8 dataset. Authors used DenseNet 169-layer pre-trained model with output layer using Sigmoid activation and R^2 loss function. Authors also reported they didn't observe difference in performance when using Mean Squared Error Loss (MSE) function. Considering these inputs, experiments are conducted to develop DL models on NIH chest X-ray 14 dataset, reporting performance results approximating state of the art.

Figure 4.6 indicates model architecture developed for comparative study of age prediction task analysis. On NIH Chest X-ray 14, initial model architecture reported as per (Karargyris et al., 2019) along with additional Global Average Pooling Layer (GAP) and Mean Squared Logarithmic Error (MSLE) loss function gave improved performance results approximating to state of the art.

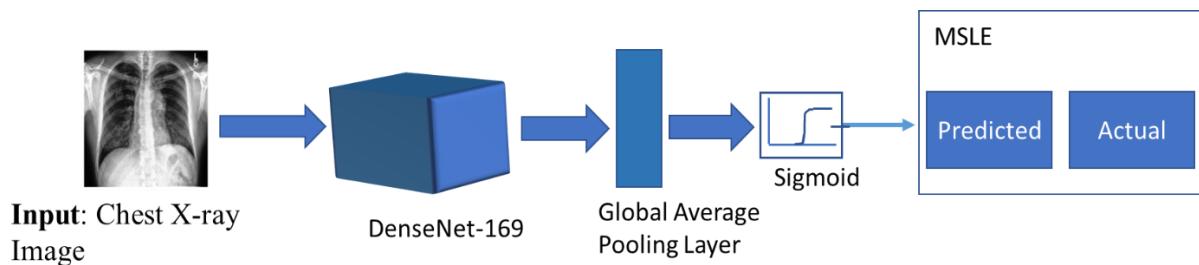


Figure 4.6 : Model architecture used for analysis of age prediction task.

Table 4.1 indicates comparable performance measurement results between state of art methods and methods developed for analysis of age prediction task based on figure 4.6. Note that slight difference of 4% performance would be attributed to regularization from use of increased dataset size NIH chest X-ray 14 compared to NIH chest X-ray 8 and additional hyperparameters not reported in original paper (Karargyris et al., 2019). These results indicate methods developed, based on figure 4.6, for age prediction task analysis are appropriate for further investigation.

Table 4.1: Table indicating age prediction performance results achieved on NIH Chest X-ray.

Method	Dataset	% of records with absolute prediction error within 9 years
Performance results as reported (Karargyris et al., 2019).	NIH Chest X-ray 8	92%
Performance of model as per figure 4.6.	NIH Chest X-ray 14	88%

For further analysis and comparative study, evaluation is made on both CheXPert and NIH chest X-ray 14 datasets. For a fair comparison, since NIH Chest X-ray dataset doesn't include uncertain labels, processed CheXPert dataset where uncertain labels are ignored is considered for this analysis. Additionally, as discussed in section 4.3.3, multitask methods are also developed for evaluation of impact on age prediction task performance due to multitask approaches. For multitask approaches, multitask methods that use sigmoid in output layer of age prediction task are used. Table 4.2 indicates Mean Absolute Error (MAE) measured on test data for age prediction task.

Table 4.2 : Comparative Study of Age Prediction Analysis Task

Analysis Type	CheXPert (Uncertain Class Ignored)	NIH Chest X-ray 14
Age Analysis only	5.9 Years	4.5 Years
Multitask model	6.5 Years	6.1 Years

From the results, it is observed that, when used in context of multitasking DL methods, even on NIH chest X-ray dataset performance of multitask model on age prediction task gets reduced. However, performance impact of multitasking methods with CheXPert dataset is less, which is due to regularization from using large number of records in CheXPert compared to NIH chest X-ray 14 dataset, thereby providing stable performance results. This indicates that performance of DL methods on CheXPert dataset is more stable across methods. For comparative analysis of different multitask approaches, sigmoid and RELU activation functions are used in output layer of age prediction task.

4.3.3 Methods for Multitask Analysis

Based on learnings from abnormal classification and age prediction tasks discussed in 4.3.1 and 4.3.2 respectively, multitask DL methods are developed for abnormality classification and prediction of age value from X-ray images of patients. Section 3.6 discusses on overall approach for developing multitask methods in this study. Multitask methods are developed using large pre-trained model DenseNet 169-layer with parameters 12,646,210 and MobileNet V2 with fewer parameters of 2,260,546, for comparative analysis of need for large DL models. Also as indicated in figure 4.7, for comparative analysis, different multitask methods are developed using sigmoid or RELU as activation function of output layer for age prediction task.

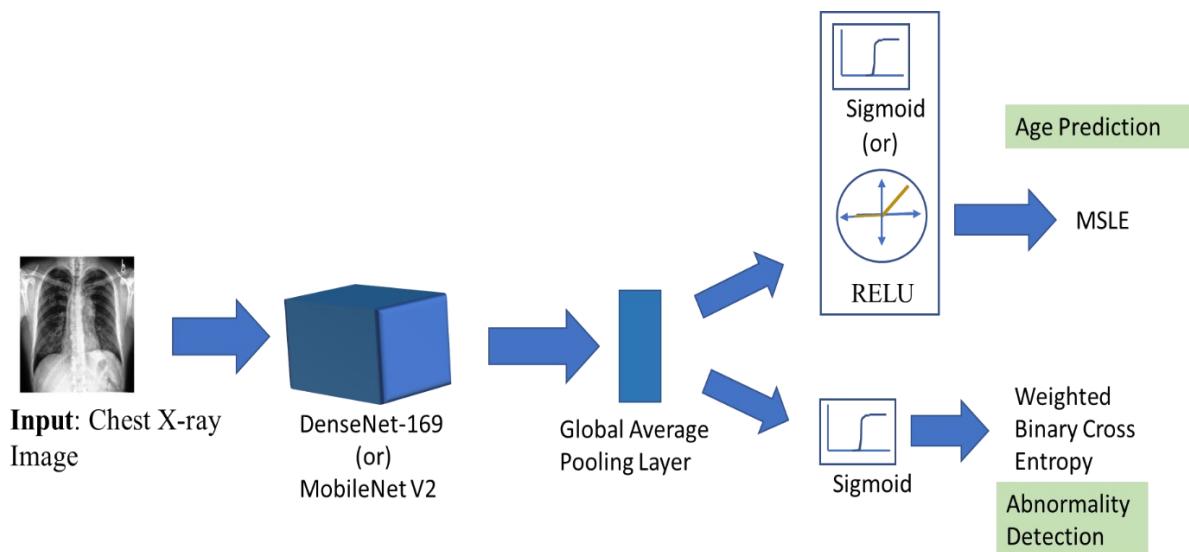


Figure 4.7: Model architecture used for comparative multi-tasking analysis.

However, multitask methods developed using single pretrained model gave comparatively less performance for different tasks. Hence, using a single multitask method to further enhance performance, combination of pretrained models is used for training on different tasks considered. Multitask methods using combination of pre-trained models DenseNet (Huang et al., 2017), ResNet (He et al., 2016) and MobileNet (Howard et al., 2017), are developed. Figure 4.8 indicates overview of multitask approach where pre-trained models are jointly used together for training and prediction of results. As reported in Chapter 5, RELU as an activation function in output layer of age prediction task provided optimal performance. Hence, for this study, multitask methods using combination of pre-trained models uses RELU instead of Sigmoid as activation in output layer of age prediction task. This improved performance on both tasks being trained, but at cost of increased model size. Each pre-trained model is provided with X-ray image as input and all models are concatenated before connecting to output layer. This helps in increased model performance, where different pre-trained models analyze input X-rays images with different features and finally model provides concatenated rich feature representation of input X-ray images. This method also helped in providing performance results approximating to state of art for multi-pathology multitask approaches. To improve performance on age prediction task, additional investigation is made using MAE instead of MSLE loss function for age prediction task.

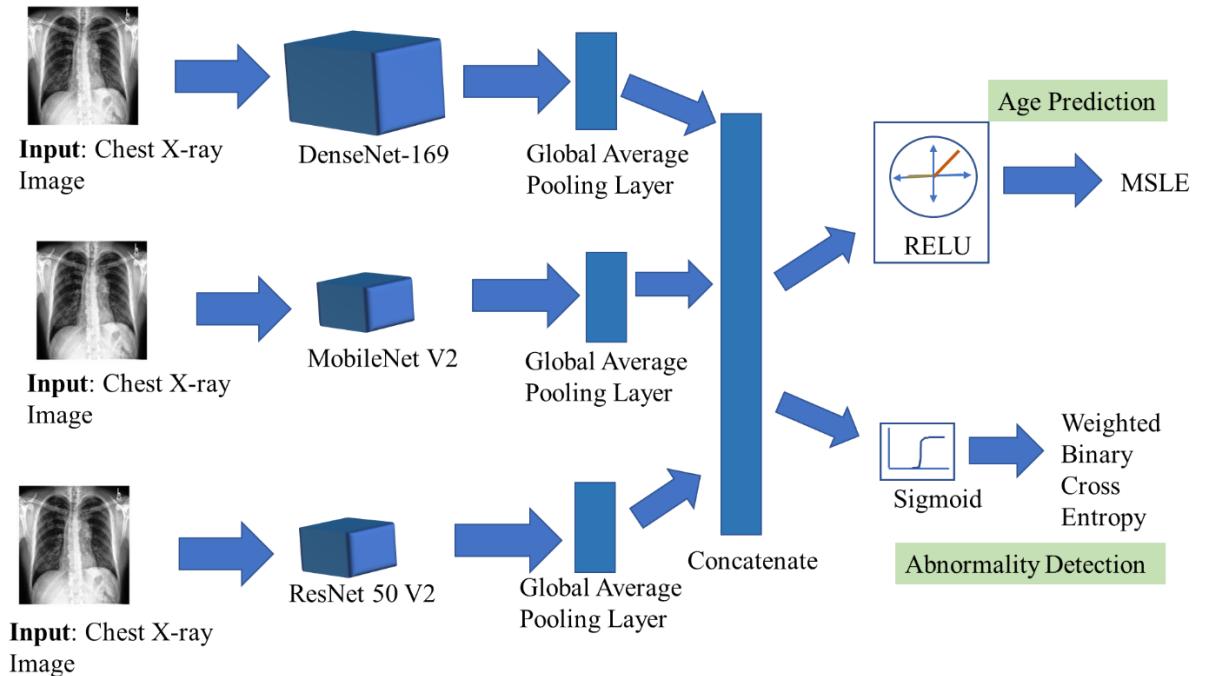


Figure 4.8: Overall approach of multitask methods using combination of pre-trained models.

4.4 Methods for Multi-Pathology Multitask Analysis

So far in this chapter, discussion is made on multitask methods developed for abnormal classification of X-ray images to meet aim and objectives of the study. But, to evaluate use of multitask approaches presented in this study on advanced use cases, it will be more beneficial to update abnormality detection task analysis to multi-pathology task analysis without significant impact on performance of age prediction task. This shall support radiologist to do patient diagnosis based on specific pathology information in relation to predicted age as advanced diagnosis tool. As mentioned in section 3.1, DL methods developed in this study consider accommodating additional parameters for future scalability purposes. Hence, unlike previous studies, multitask model architecture developed in this study are not tailored specifically for abnormal classification analysis. This helps in enhancing multitask model seamlessly from abnormal X-ray image detection to multi-pathology analysis from X-ray images. Figure 4.9 shows overview of model architecture for multi-pathology multitask methods based on learnings from multitask methods developed so far in this study. Abnormality detection task is updated with pathology prediction task of 14 labels (12 labels with pathology information, No Finding, Supported devices) present in CheXPert dataset. Further for evaluation purposes, AUC values of top 5 prevalent pathologies (Atelectasis, Cardiomegaly, Consolidation, Edema, Pleural Effusion) are used, which is a competition task as per (Irvin et al., 2019).

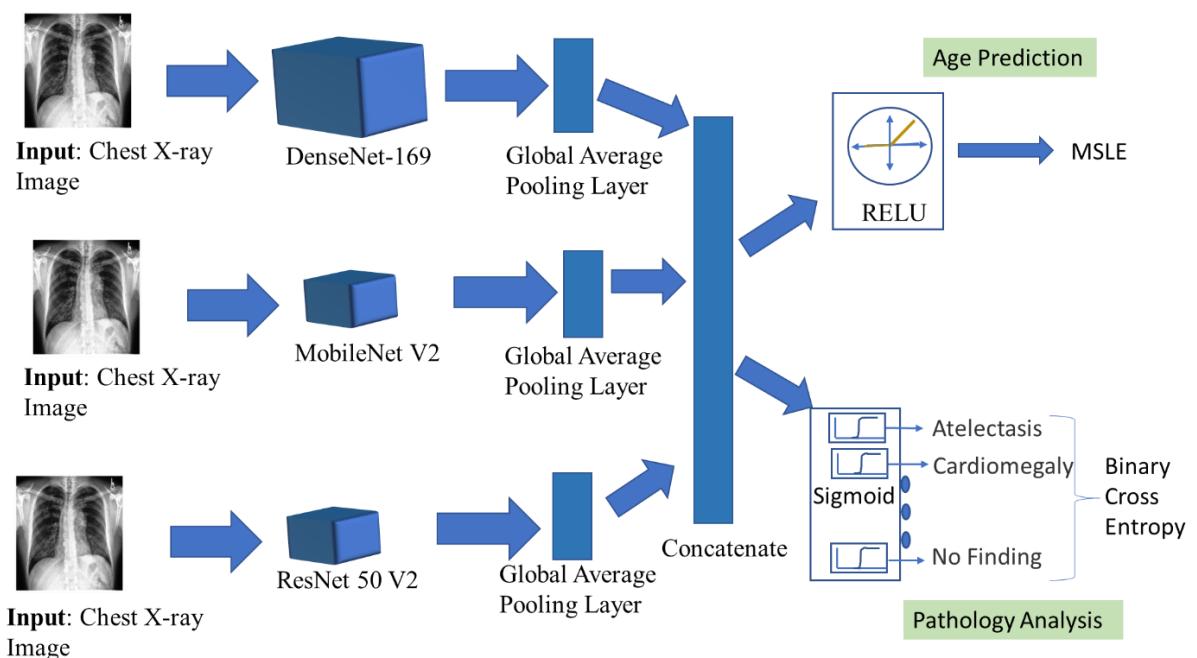


Figure 4.9 : Methods for multi-pathology multitask analysis

4.5 Hyperparameters And Other Considerations

To provide multiple perspectives of X-ray image analysis during training of methods with combination of pre-trained models, different pretrained models are provided with input of different image sizes as shown in table 4.3. Adam optimizer with default values is used for training models. Batch size is set to 60.

Table 4.3: Table indicating input image size for different pre-trained models

Pre-trained model	Input Image Size
DenseNet	(300,300)
MobileNet	(224,224)
ResNet	(334,334)

With methods using combination of pre-trained models with ResNet, errors are observed in model creation process by TensorFlow due to duplication of layer names. Hence, layers of ResNet model are appended with a string ‘EP_’ for uniqueness of layer names. This shouldn’t impact ResNet pre-trained model architecture, weights, and biases.

Since different methods need different number of epochs for convergence of models, epoch size is set accordingly. For example, MobileNet V2 model need more epochs to converge on age and abnormal detection task, hence 25 epochs are set. Multi-pathology multitask analysis need fewer epochs, where it converges within 7-8 epochs, hence 15 epochs are chosen for appropriate analysis. Additionally, in an experiment, since it would be hard to save best model as checkpoint automatically on multiple metrics of multitask model, models are saved for every epoch during training. Further, in each experiment, model with balanced optimal performance on validation data for both abnormal classification (or multi-pathology) task and age prediction task is considered for prediction on test data.

4.6 Summary

In this chapter, discussion is made on design and implementation of methods developed for analysis abnormal detection task and age prediction task separately. In multitask context, performance on age prediction task is observed to be less. Hence, benchmarking against prior state of art methods is made for appropriate comparative analysis. Both NIH chest X-ray 14 and

CheXPert datasets are used for analysis and comparison. Further, multitask methods are developed which can analyze both abnormal detection task and age prediction task concurrently. Additionally, for appropriate comparative analysis, multitask methods are developed with pre-trained models having large parameters like DenseNet and pre-trained models with fewer parameters like MobileNet to determine need for large complex models.

As performance of multitask methods with a single pre-trained model is less than methods trained on independent tasks, to boost performance of multitask methods combination of pre-trained models are used to get better feature representation of input X-ray images. This helped in improving performance of multitask methods, approximating to performance of methods trained on independent tasks. Further, DL methods are developed for multi-pathology multitask analysis to evaluate advanced use cases of multitask methods developed in this study.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 Introduction

In Chapter 4, discussion is made on design approach and implementation details of multitask methods developed for this study. This chapter further discusses about performance results from different methods developed in this study. Comparative analysis of results from abnormality detection task, age prediction task, multitask methods that can analyze both abnormality detection and age prediction task and multi-pathology multitask methods are discussed. Figure 5.1 provides information regarding overview of performance analysis discussion made in this chapter.

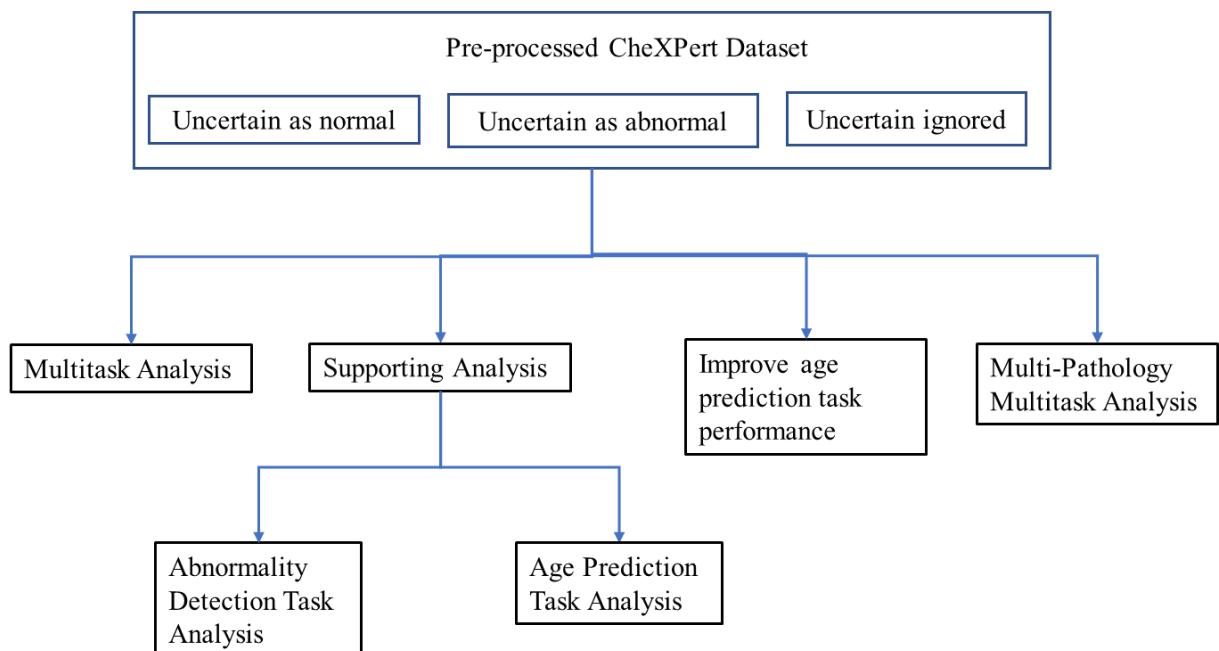


Figure 5.1: Overview of different methods developed for analysis.

Additionally, as indicated in figure 5.1, since performance results of multitask methods on age prediction task is less, discussion is made on multitask methods that showed improved performance results on age prediction task. As discussed in section 3.5 model evaluation metrics are reported on X-ray records that are annotated by radiologists. This helps in reliability of results from multitask methods developed in this study, for a practical environment setting.

In this chapter unless stated otherwise and for easy reference, DenseNet (Huang et al., 2017) indicates DenseNet 169-layer pre-trained model, MobileNet (Howard et al., 2017) indicates MobileNet V2 pre-trained model and ResNet (He et al., 2016) indicates ResNet 50 V2 pre-trained model. All pre-trained models used in this study consider weights pre-trained on ImageNet dataset.

5.2 Performance Analysis

In practical clinical setting, classifying an abnormal X-ray record as normal causes more harm than classifying normal X-ray record as abnormal. Since in later case, radiologist shall still later diagnose X-ray record if it is classified as abnormal thereby preventing misdiagnosis. In former cases making abnormal X-ray as normal causes radiologist to ignore critical condition of patient causing serious damage. Additionally, as discussed in section 3.3, CheXPert dataset has high class imbalance problem, where more records with abnormal class are present than normal class. Hence using AUC/AUPRC as only metric for comparison of performance is not completely reliable. Hence in this study, to provide better analysis view of abnormality detection task performance measures, Weighted Average Recall and Weighted Average F1 values are also considered in addition to AUC/AUPRC for evaluation of multitask methods developed as part of the study. For evaluating performance of multitask methods on age prediction task, % of records with absolute prediction error within 9 years and Mean absolute Error (MAE) are considered as metrics for this study.

5.2.1 Multitask Analysis

Based on discussion made in section 4.3.3, multiple methods are developed for multitask analysis, using pre-trained DenseNet and MobileNet models. Additionally, these methods are developed on different processed datasets of original CheXPert dataset i.e., considering uncertain class as normal, uncertain class as abnormal and ignoring uncertain class. For better visualization and understanding, Figure 5.2 indicates averaged performance results across different processed datasets. Overall, in current problem context, performance of multitask methods is higher in methods using DenseNet pre-trained model than methods using MobileNet pre-trained model, indicating need for models with large parameter space for analyzing chest X-ray images. From figure 5.2, it is observed that methods using sigmoid activation in output layer of age prediction task performs better on abnormality detection task and lesser on age

prediction task and methods using RELU activation in output layer of age prediction task perform better on age prediction task and less on abnormality detection task.

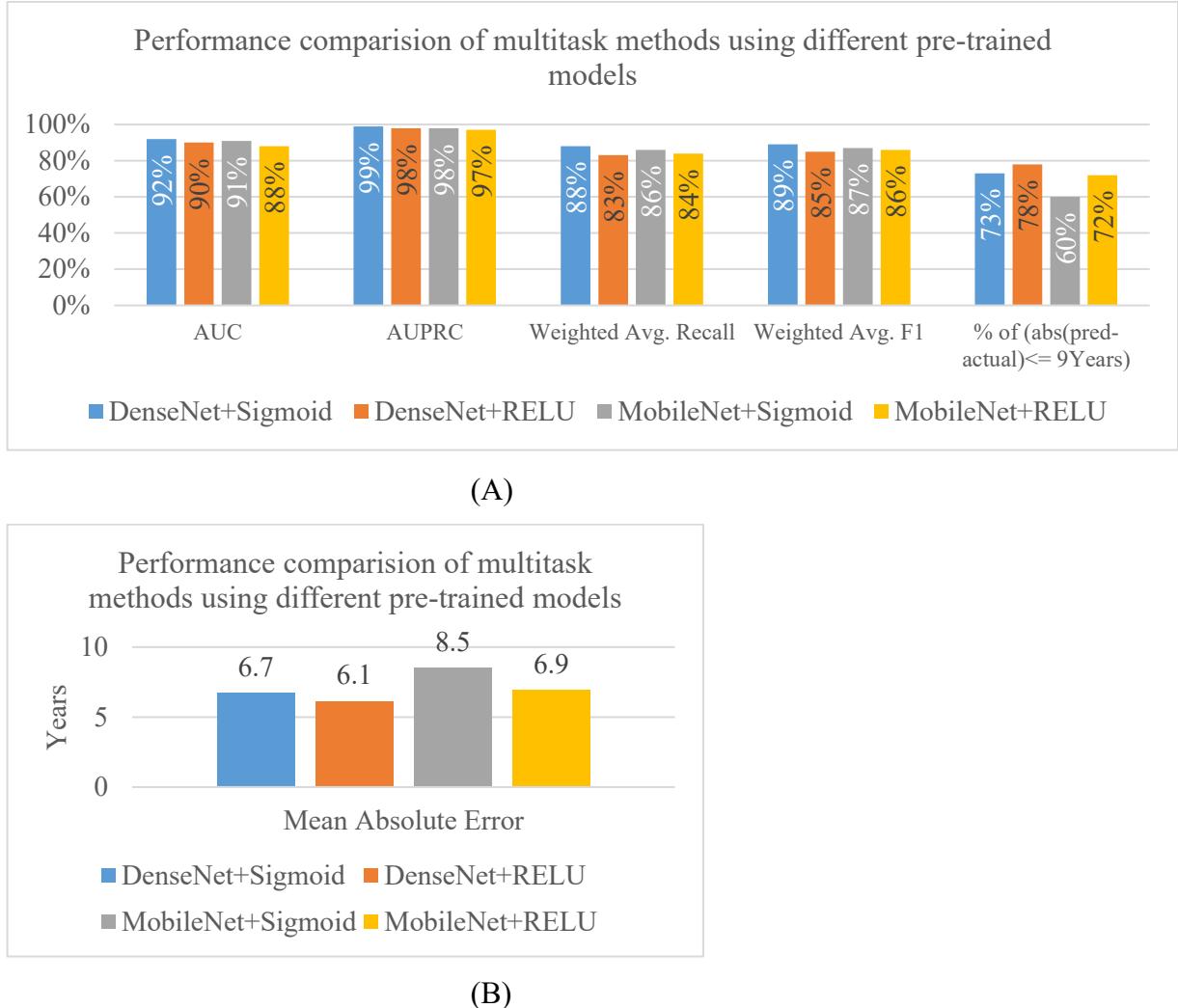
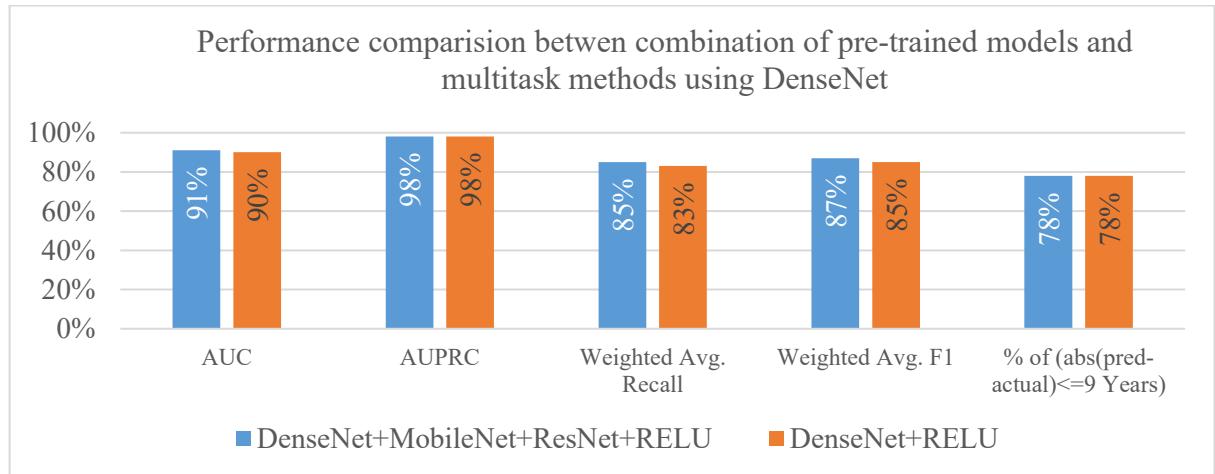


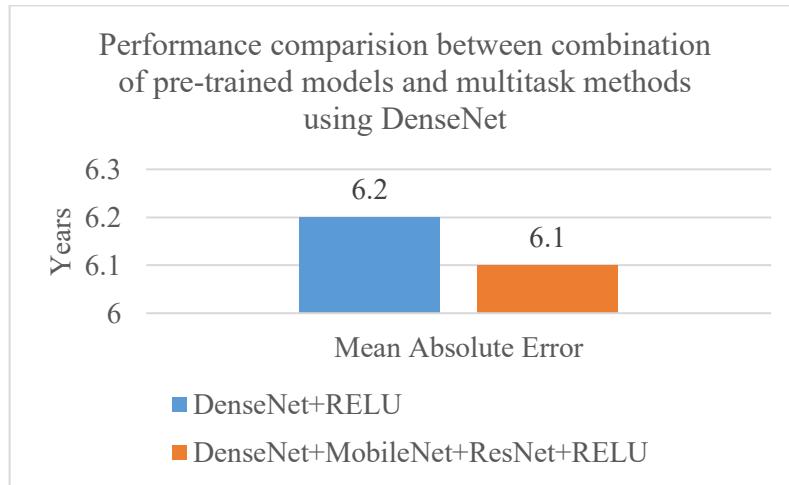
Figure 5.2: Performance comparison of multitask methods using different pre-trained models.

To support performance needs on multiple tasks with a single multitask method, advanced multitask methods using combination of multiple pretrained models, as indicated in figure 4.8, were developed enabling rich feature representation of input image and improving prediction performance on different tasks. Even with using advanced multitask methods that use combination of pre-trained models, performance on abnormality detection task improved and performance on age prediction task is not improved. Hence multitask methods with RELU activation layer is used for further study which provides optimal balanced performance on abnormality detection and age prediction tasks. Figure 5.3 indicates comparative performance results of multitask methods using combination of pre-trained models against multitask

methods using single pre-trained DenseNet model. The 0.1 Years difference in MAE is insignificant and would be due to rounding of errors in measurement values. Based on these results, performance on abnormality detection task is observed to increase with multitask methods that use combination of pretrained models, while performance on age prediction task mostly remains unchanged.



(A)



(B)

Figure 5.3: Performance comparison of different multitask architecture models.

Table 5.1 indicates performance results measured across different preprocessed datasets for multitask methods developed using combination of pretrained models as indicated in figure 4.8. From results in Table 5.1, performance results of multitask methods developed for different processed datasets doesn't vary widely. These results indicate reliable metrics from analysis made in this study.

Table 5.1: Table indicating multitask analysis performance of methods using combination of pretrained models.

Dataset	AUC	AUPRC	Weighted Avg. Recall	Weighted Avg. F1	MAE (Years)	% of records with absolute prediction error within 9 years
Uncertain_Ignore	0.91	0.98	0.85	0.87	5.9	79%
Uncertain_Abnormal	0.89	0.98	0.84	0.86	6.5	76%
Uncertain_Normal	0.93	0.99	0.87	0.88	6.3	78%

5.2.2 Supporting Analysis

So far comparative performance analysis of different multitasking methods is discussed to determine appropriate model for analysis. But comparative analysis of multitask methods with methods trained only on independent task would be more appropriate, like ablation studies. This will help in evaluating performance impact on different tasks due to multitask approaches.

5.2.2.1 Abnormality Detection Task Performance Analysis

In this section comparative performance analysis on abnormality detection task is made. Figure 5.4 indicates performance results averaged across different processed datasets. Comparative performance analysis is made between methods trained on only abnormality detection task as discussed in section 4.3.1 and proposed multitask methods. Based on results, it is observed that advanced multitask methods with combination of pre-trained models provides performance results approximating to methods trained only on abnormality detection task. This indicates need of advanced multitask methods using combination of pre-trained models, for concurrent analysis of abnormality detection and age value prediction from chest X-ray images.

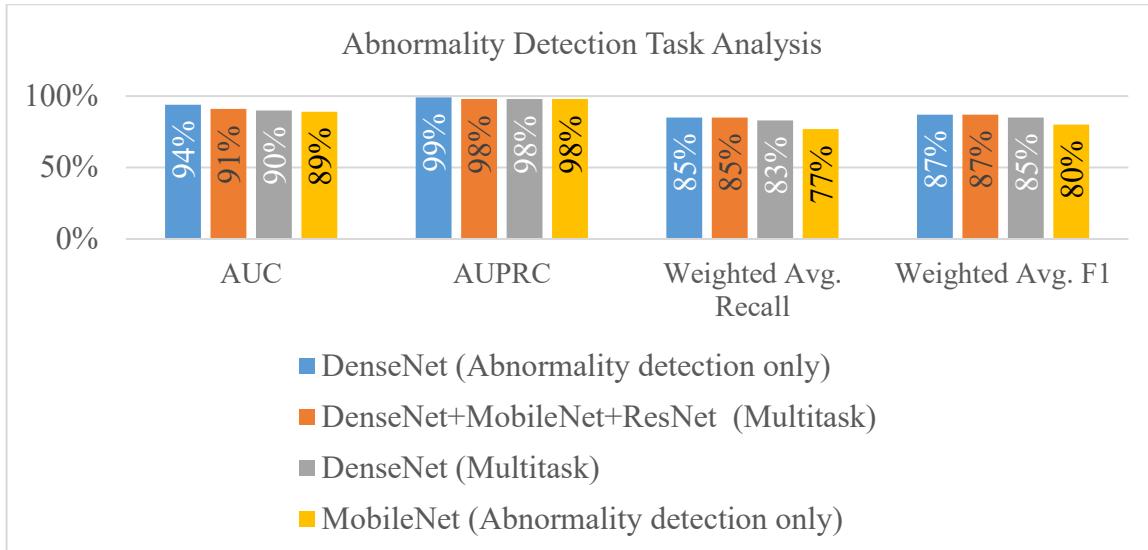


Figure 5.4: Comparative performance analysis on abnormality detection task.

Based on study conducted in Chapter 2, (Rajaraman et al., 2020) indicated state of art performance results for abnormality detection task on CheXPert dataset. Figure 5.5 indicates abnormality detection performance metrics comparison between best performance multitask method using combination of pre-trained models and state of art results as per (Rajaraman et al., 2020). CheXPert has class imbalance problem where number of records with abnormal class label is more than that of records with normal class label. Hence, DL methods developed on CheXPert has high tendency to achieve high sensitivity with compromise on specificity. Results as per (Rajaraman et al., 2020) indicates overfitting on abnormal class with more Sensitivity at compromised Specificity.

However, multitask approaches in this study consider models with balanced metrics and show improved F1 score compared to state of art results. Hence performance analysis results of proposed multitask methods developed as part of this study approximates to state of art performance in general on abnormality detection task and appropriate. This suggests on scalability to future use cases for proposed multitasking methods, which is discussed in section 5.4 for multi-pathology multitask analysis.

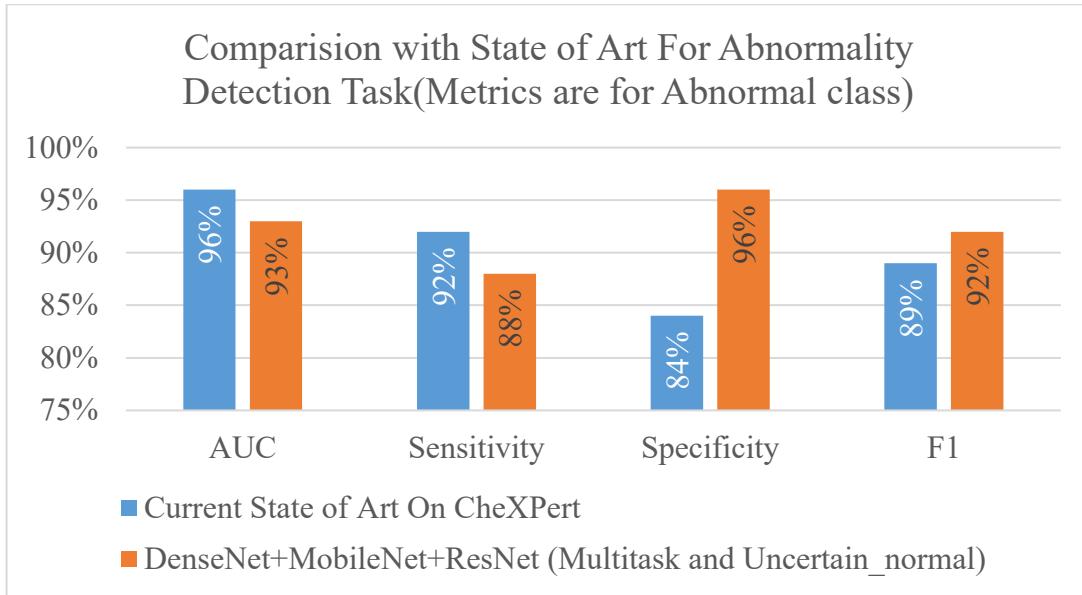
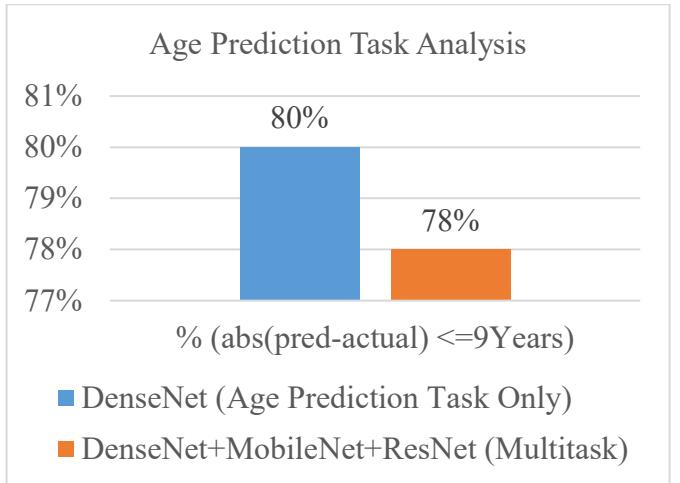


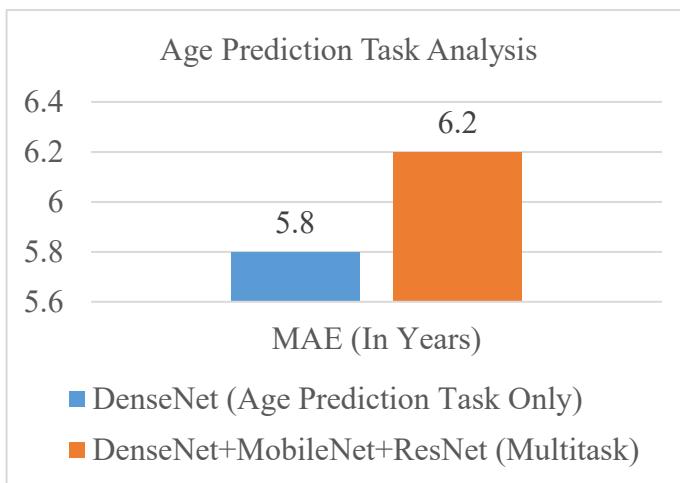
Figure 5.5 : Comparison of abnormality detection task performance against (Rajaraman et al., 2020)

5.2.2.2 Age Prediction Task Performance Analysis

So far based on study conducted, as discussed in section 2.5, performance analysis on age prediction task from chest X-rays using CheXPert dataset is not yet made. Hence for benchmarking purposes, using CheXPert dataset, DL methods that are trained only on age prediction task are developed for comparative analysis with multitask methods. Figure 5.6 indicates comparative performance analysis on age prediction task, where methods trained only on age prediction task are compared against methods trained on multitask methods. Results indicate methods trained on multitask approaches don't have significant performance difference (2% difference or 0.4 MAE in Years) when compared to methods trained only on age prediction task. Hence multitask methods using combination of pre-trained models, as indicated in figure 4.8 are appropriate for analysis of Chest X-ray images. Additionally, detailed comparative performance analysis with state of art on age prediction task is discussed in section 4.3.2.



(A)



(B)

Figure 5.6 : Comparative analysis of performance on age prediction task

5.3 Further Improving Performance on Age Prediction Task

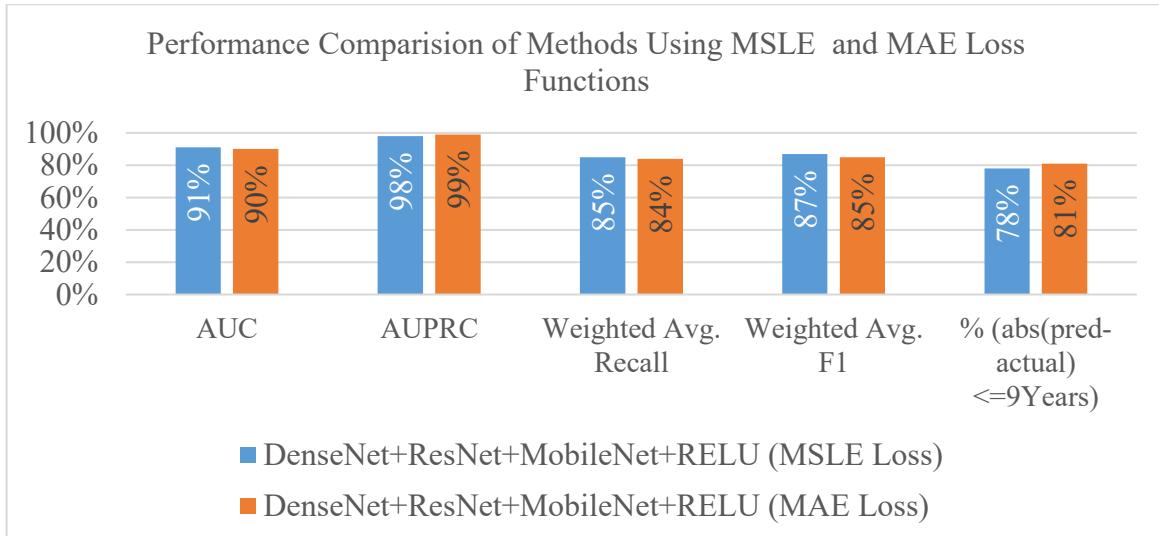
So far, multitask methods presented in this study indicates appropriate performance results for abnormality detection task, but performance of age prediction task is slightly less than 80% for metric indicated in state of art i.e., % of records with absolute age prediction error less than 9 years. These results indicate need for further improvement on performance of this task. Based on experiments further conducted, using Mean Absolute Error (MAE) as loss function for age prediction task helped in improving performance on age prediction task further, but with impact on abnormality detection task performance. Table 5.2 indicates performance results measured when MAE is used as loss function for age prediction task in multitask approaches.

Table 5.2: Table indicating multitask performance results with MAE as loss for age prediction task.

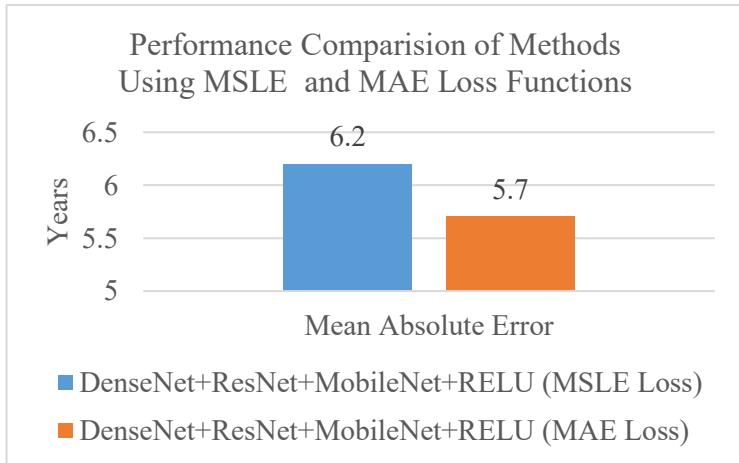
Dataset	AUC	AUPRC	Weighted Avg. Recall	Weighted Avg. F1	MAE (Years)	% of records with absolute prediction error within 9 years
Uncertain_Ignore	0.92	0.99	0.85	0.86	5.8	81%
Uncertain_Abnormal	0.87	0.98	0.83	0.84	5.5	83%
Uncertain_Normal	0.92	0.99	0.83	0.85	5.8	79%

For better clarity and understanding, figure 5.7 indicates comparative performance analysis between multitasking methods using Mean Squared Logarithmic Error (MSLE) and Mean Absolute Error (MAE) as loss functions for age prediction task. These results indicate, performance of multitask approaches on age prediction task has improved with MAE loss, but with impact on performance of abnormality detection task. This can be observed from lesser Recall and F1 scores with MAE as loss function for age prediction task, in figure 5.7.

However, predicting age value is auxiliary task while abnormality detection is main task and compromising on abnormality detection task performance with improved performance on age prediction task is not appropriate. Since objective of this study is to develop multitask method that can provide optimal performance on both abnormality detection task and age prediction task, further investigation of using ensemble of multitask methods that can improve abnormality detection performance with MAE as loss for age prediction task is future research scope of this study.



(A)



(B)

Figure 5.7: Comparative performance analysis on age prediction task between multitask methods using MSLE and MAE loss function.

5.4 Multi-pathology Multitask Analysis

Though goal of this study is multitasking analysis for abnormality detection and age prediction tasks, multi-pathology multitask methods are developed to evaluate advanced use cases of multitask methods developed in this study. Based on discussion made in 5.2.1, multitask DL methods that use combination of multiple pre-trained models showed optimal performance on both abnormality detection and age prediction task together. Further in this section evaluation of use of these methods for multi-pathology multitask analysis is made. Table 5.3 indicates average AUC performance results of 5 prevalent pathologies (Atelectasis, Cardiomegaly, Consolidation, Edema, Pleural Effusion), which is competition task as per (Irvin et al., 2019).

Table 5.3 additionally indicates performance results on age prediction task i.e., MAE value and % of records with age prediction error less than 9 years.

Table 5.3: Table indicating performance measures of multi-pathology multitask methods

Dataset	AUC	MAE(Years)	% of records with absolute prediction error within 9 years
Uncertain as abnormal	0.87	6.6	76%
Uncertain as normal	0.85	5.8	78%
Average	0.86	6.2	77%

From results in table 5.3, multitask methods developed using combination of pre-trained models can seamlessly be scaled to advanced use case of multi-pathology multitask analysis without significant performance impact to age prediction task. Figure 5.8 indicates AUROC curve plot for 5 prevalent pathologies on processed dataset considering uncertain class as abnormal. Results indicate performance of model is high on Edema with AUC of 93% and low on Atelectasis with AUC of 81%, compared with other pathologies.

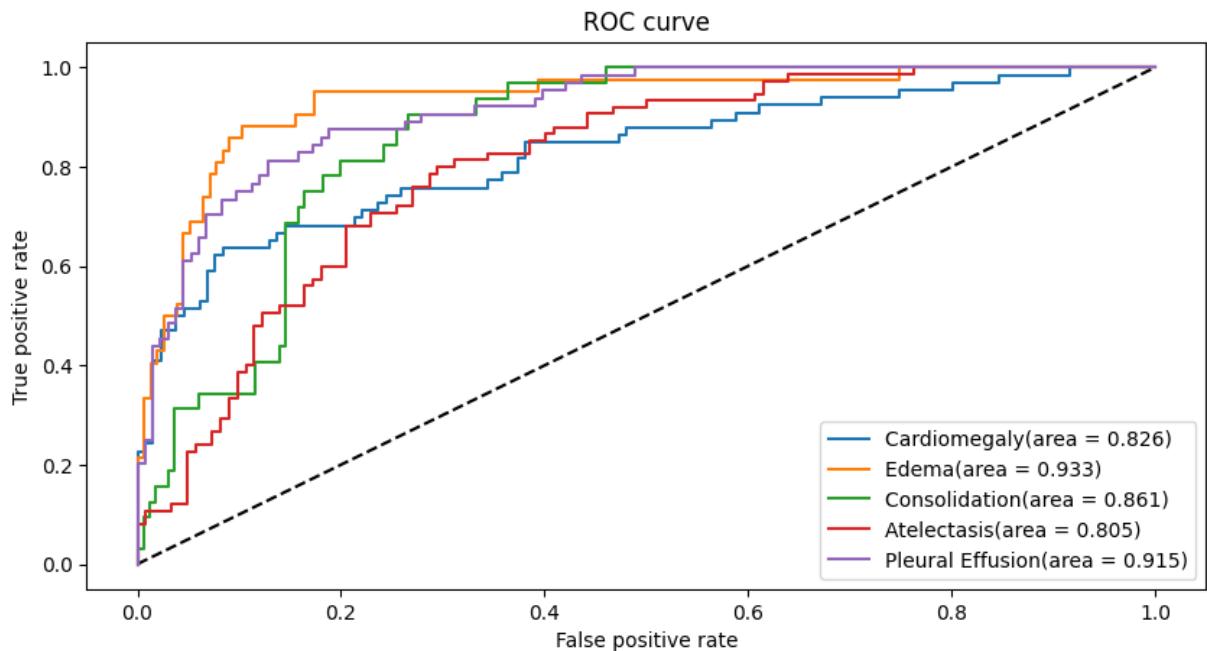


Figure 5.8: ROC curve plot for multiple pathologies - uncertain as abnormal (5 prevalent pathologies)

5.4.1 Comparative Analysis with CheXPert Methods

Figure 5.9 indicates comparison of pathology detection performance results of multi-pathology multitask DL methods developed in this study with metrics presented in (Irvin et al., 2019). Performance analysis results are provided for 5 different pathologies (Atelectasis, Cardiomegaly, Consolidation, Edema, Pleural Effusion), which is the competition task (Irvin et al., 2019). For pathology detection, performance of multi-pathology multitask methods developed in this study show, though slightly less, comparable results. For Cardiomegaly, methods proposed in this study has significantly lower performance compared to CheXPert, thereby impacting average AUC metric across 5 pathologies. But, for other pathologies AUC metric is comparable to CheXPert indicating use of methods proposed in this study for future scalability purposes. Further analysis to improve performance would be future research scope of this study.

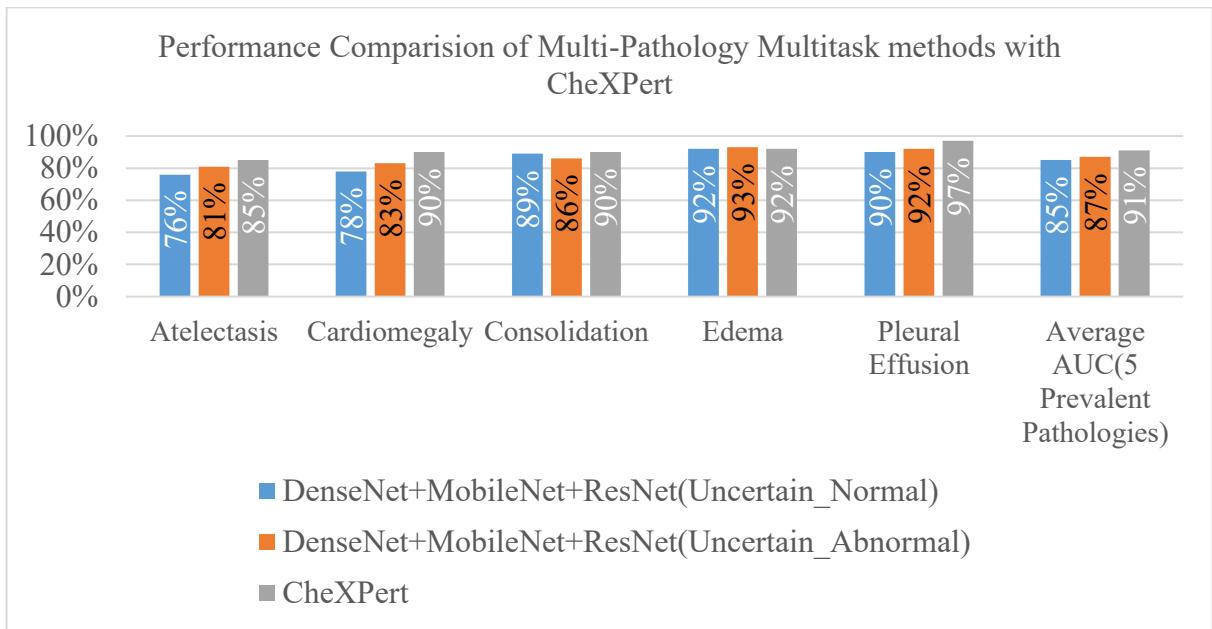


Figure 5.9: Performance comparison of multi-pathology multitask methods developed in this study with CheXPert.

5.5 Summary

In this chapter, discussion is made on comparative performance analysis of different multitask methods developed to meet aim and objectives of this study. Comparison of multitask methods using different pre-trained models is made for benchmarking purposes. However, since

performance of multitask model on different tasks is less, advanced multitask methods are developed using combination of different pre-trained models to provide rich feature representation of input images compared to methods using single pre-trained model. This helped in increased performance of different tasks considered for this study.

Further, to provide reliable analysis of proposed multitask methods, comparison is made with methods trained on independent tasks indicating performance impact due to novel multitask needs considered in this study. Here multitask methods developed using combination of pre-trained models provided performance metrics approximating to that of methods trained on independent tasks. This suggests possibility of use of advanced multitask framework developed in this study for future research purposes.

Further, since performance of multitask methods is less on age prediction task, comparative analysis of multitask methods with MAE loss for age prediction task is made. Results indicates improvement on age prediction task, but with compromise on performance on abnormality detection task.

Additionally, multi-pathology multitask methods are developed using proposed advanced multitask framework. These methods can analyze multiple pathologies from X-ray images in addition to prediction age value from X-ray images. Performance results indicates methods developed for multitask analysis can easily scale to multi-pathology multitask analysis with little or no impact to different tasks performance. This indicates significant future research scope of advanced multitask methods proposed in this study.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

6.1 Introduction

In this chapter discussion is made on analysis of proposed multitask methods in achieving aim and objectives of the study. To do this aim and objectives of this study are briefly discussed and how they are achieved with proposed multitask methods. Further, discussion is made on main contributions from this study. Finally, directions on future research of this study are discussed.

6.2 Discussion and Conclusion

Aim of this study is investigation of methods that can analyze chest X-ray images to concurrently determine presence of abnormality and predict age value. To meet aim of this study, objectives are formulated as development and evaluation of multitask deep learning methods that can concurrently analyze chest X-ray images for abnormality detection and age value prediction. As discussed in section 2.4.4, previous multitask approaches, in general, are related to improving performance of pathology detection problem. Hence improving performance on one task can have strong correlation to improve performance on another task. However, different tasks considered in this study, i.e., pathology analysis and age prediction, provide different analysis perspectives from the same X-ray image. Using these tasks, this multi-dimensional analysis view of X-ray image becomes challenging for training Deep Learning (DL) methods, such as improved performance on abnormality detection task can result in reduced performance on age prediction task and vice versa.

To improve performance on multiple tasks together, advanced multitask methods using combination of pre-trained models are proposed which perform well in analysis of both abnormality detection and age prediction tasks. Performance of these advanced multitask methods on abnormality detection and age prediction tasks approximates to state of art performance of methods trained on these tasks independently. Multitask models proposed from the analysis also provides stable results across different processed datasets for multiple metrics considered. This indicates that multitask methods developed in this study can analyze chest X-

ray images to concurrently detect abnormality and predict age value, thereby meeting aim and objectives of this study.

Further, investigation is made on scaling multitask methods to advanced use case of multi-pathology multitask analysis with less impact to performance on independent tasks. Hence, multitask methods developed in this study can seamlessly scale to advance analysis view of chest X-ray images, thereby supporting radiologist for better patient diagnosis.

6.3 Contribution to Knowledge

Unlike current state of art methods, this study proposes multitask DL methods that can analyze chest X-ray images from multiple perspectives, supporting advanced patient diagnosis. Different tasks considered in this study are abnormal X-ray detection based on pathology information and age value prediction. Using a single method to predict these tasks helps radiologist in understanding X-ray abnormality in relation to age, thereby helping in improved patient diagnosis.

Additionally, to evaluate multitask methods developed for advanced use cases, multi-pathology multitask methods are developed that can analyze multiple pathologies instead of abnormality detection from chest X-rays in addition to age prediction task. Performance results indicate multitask methods developed in this study are scalable to multi-pathology multitask analysis without significant impact to different tasks considered. This can help in improved patient diagnosis, where diagnosis is not just based on pathology information, instead pathology information in relation to patient's age.

Overall, this study provides multitask DL framework for patient diagnosis in context of precision health. In other words, this study provides multitask DL framework for patient diagnosis, which is not just based on pathology information but additionally consider other parameters, thereby providing advanced automated diagnosis information from X-rays.

6.4 Future Recommendations

Work presented in this study aims to develop framework of multitask DL methods that can analyse medical images from multiple perspectives like a radiologist, which represents broad research area of study. However, near future directions for this study involve,

1. Using ensemble of models, which shall help in increasing performance on pathology analysis task (Rajaraman et al., 2020; Pham et al., 2021). Hence, investigation of using ensemble of multitask and multi-pathology multitask DL methods for improving performance can be considered.
2. Enhancement of methods developed in this study to other clinical workflows can be considered, without limiting to analysis of chest X-ray images. Also, datasets for other clinical workflows like paediatric bone age assessment (Halabi et al., 2019), Musculoskeleton abnormality detection (Rajpurkar et al., 2017a) etc., are publicly available for research purposes. Hence including these additional clinical datasets in addition to chest X-ray image datasets can help in providing a generic X-ray diagnosis tool that can support radiologist.
3. Collection of additional metadata information such as lung area, heart area, bone density etc., can be considered and used as different tasks in multitask analysis, thereby providing rich patient diagnosis information.

REFERENCES

- Ali, F., Harrington, S., Kennedy, S. & Hussain, S. (2015). Diagnostic Radiology in Liberia: A Country Report. *Journal of Global Radiology*. [Online]. 1 (2). p.p. 6. Available from: <http://dx.doi.org/10.7191/jgr.2015.1020>.
- Allaouzi, I. & Ben Ahmed, M. (2019). A Novel Approach for Multi-Label Chest X-Ray Classification of Common Thorax Diseases. *IEEE Access*. [Online]. 7. p.pp. 64279–64288. Available from: <http://dx.doi.org/10.1109/ACCESS.2019.2916849>.
- Alom, M.Z., Shaifur Rahman, M.M., Nasrin, S., Taha, T.M. & Asari, V.K. (2020). COVID_MTNet: COVID-19 detection with multi-task deep learning approaches. *arXiv preprint arXiv:2004.03747*. [Online]. Available from: <https://arxiv.org/ftp/arxiv/papers/2004/2004.03747.pdf>.
- Arora, R. (2014). The training and practice of radiology in India: current trends. *Quantitative imaging in medicine and surgery*. [Online]. 4 (6). p.pp. 449–44950. Available from: <http://dx.doi.org/10.3978/j.issn.2223-4292.2014.11.04>.
- Benjamins, S., Dhunnoo, P. & Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digital Medicine*. [Online]. 3 (1). p.pp. 1–8. Available from: <http://dx.doi.org/10.1038/s41746-020-00324-0>.
- Boag, W., Hsu, T.-M.H., Mcdermott, M., Berner, G., Alsentzer, E., Szolovits, P., Dalca, A. V., Finlayson, S., Oberst, M., Falck, F. & Beaulieu-Jones, B. (2020). Baselines for Chest X-Ray Report Generation. *Machine Learning for Health NeurIPS Workshops*. [Online]. 116. p.pp. 126–140. Available from: <http://proceedings.mlr.press/v116/boag20a.html>.
- Bozorgtabar, B., Mahapatra, D., Vray, G. & Thiran, J.P. (2020). Salad: Self-supervised aggregation learning for anomaly detection on x-rays. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [Online]. 12261 LNCS. p.pp. 468–478. Available from: https://doi.org/10.1007/978-3-030-59710-8_46.
- Bram van Ginneken & Sjoerd Kerkstra (2012). *Grand Challenge*. [Online]. 2012. Available from: <https://grand-challenge.org/>. [Accessed: 25 June 2021].
- Burrill, J., Williams, C.J., Bain, G., Conder, G., Hine, A.L. & Misra, R.R. (2007). Tuberculosis: A radiologic review. *Radiographics*. [Online]. 27 (5). p.pp. 1255–1273. Available from: <https://pubs.rsna.org/doi/pdf/10.1148/rg.275065176>.

- Colijn, C., Jones, N., Johnston, I.G., Yaliraki, S. & Barahona, M. (2017). Toward precision healthcare: Context and mathematical challenges. *Frontiers in Physiology*. [Online]. 8 (MAR). p.p. 136. Available from: <http://dx.doi.org/10.3389/fphys.2017.00136>. [Accessed: 10 April 2021].
- Dai, W., Dong, N., Wang, Z., Liang, X., Zhang, H. & Xing, E.P. (2018). Scan: Structure correcting adversarial network for organ segmentation in chest x-rays. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [Online]. 11045 LNCS. p.pp. 263–273. Available from: https://doi.org/10.1007/978-3-030-00889-5_30.
- Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S. & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*. [Online]. 25 (1). p.pp. 24–29. Available from: <http://dx.doi.org/10.1038/s41591-018-0316-z>.
- Farag, A.T., El-Wahab, A.R.A., Nada, M., El-Hakeem, M.Y.A., Mahmoud, O.S., Rashwan, R.K. & El Sallab, A. (2020). MulticheXnet: A multi-task learning deep network for pneumonia-like diseases diagnosis from X-ray scans. *arXiv preprint arXiv:2008.01973*. [Online]. Available from: <https://arxiv.org/pdf/2008.01973.pdf>.
- Gordienko, Y., Gang, P., Hui, J., Zeng, W., Kochura, Y., Alienin, O., Rokovyi, O. & Stirenko, S. (2019). Deep learning with lung segmentation and bone shadow exclusion techniques for chest X-ray analysis of lung cancer. *Advances in Intelligent Systems and Computing*. [Online]. 754 (Advances in Computer Science for Engineering and Education. ICCSEEA 2018.). p.pp. 638–647. Available from: <https://ui.adsabs.harvard.edu/abs/2017arXiv171207632G>.
- Greenspan, H., Van Ginneken, B. & Summers, R.M. (2016). Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging*. [Online]. 35 (5). p.pp. 1153–1159. Available from: <https://ieeexplore.ieee.org/document/7463094>.
- Gündel, S., Ghesu, F.C., Grbic, S., Gibson, E., Georgescu, B., Maier, A. & Comaniciu, D. (2019). Multi-task learning for chest X-ray abnormality classification on noisy labels. *arXiv preprint arXiv:1905.06362*. [Online]. p.pp. 1–10. Available from: <https://arxiv.org/pdf/1905.06362.pdf>.
- Halabi, S.S., Prevedello, L.M., Kalpathy-Cramer, J., Mamonov, A.B., Bilbily, A., Cicero, M., Pan, I., Pereira, L.A., Sousa, R.T., Abdala, N., Kitamura, F.C., Thodberg, H.H., Chen, L., Shih, G., Andriole, K., Kohli, M.D., Erickson, B.J. & Flanders, A.E. (2019). The rSNA

- pediatric bone age machine learning challenge. *Radiology*. [Online]. 290 (3). p.pp. 498–503. Available from: <https://doi.org/10.1148/radiol.2018180736>.
- He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. [Online]. 2016-Decem. p.pp. 770–778. Available from: <https://arxiv.org/pdf/1512.03385.pdf>.
- Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. [Online]. Available from: <https://arxiv.org/pdf/1704.04861.pdf>.
- Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K.Q. (2017). Densely connected convolutional networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. [Online]. 2017-Janua. p.pp. 2261–2269. Available from: https://openaccess.thecvf.com/content_cvpr_2017/papers/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.pdf.
- Iglovikov, V.I., Rakhlin, A., Kalinin, A.A. & Shvets, A.A. (2018). Paediatric bone age assessment using deep convolutional neural networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [Online]. 11045 LNCS. p.pp. 300–308. Available from: https://doi.org/10.1007/978-3-030-00889-5_34.
- Imran, A.A.Z. & Terzopoulos, D. (2019). Semi-supervised Multi-task Learning with Chest X-Ray Images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [Online]. 11861 LNCS. p.pp. 151–159. Available from: <https://arxiv.org/pdf/1908.03693.pdf>.
- International Osteoporosis Foundation. (2018). *Broken Bones, Broken Lives: A Roadmap to Solve the Fragility Fracture Crisis in Europe*. [Online]. 2018. International Osteoporosis Foundation. Available from: https://share.osteoporosis.foundation/EU-6-Material/Reports/IOF Report_EU.pdf. [Accessed: 15 June 2021].
- Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., Seekins, J., Mong, D.A., Halabi, S.S., Sandberg, J.K., Jones, R., Larson, D.B., Langlotz, C.P., Patel, B.N., Lungren, M.P. & Ng, A.Y. (2019). CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *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.* [Online]. p.pp. 590–597. Available from: <https://arxiv.org/pdf/1901.07031.pdf>.
- Jaeger, S., Candemir, S., Antani, S., Wang, Y.-X.J., Lu, P.-X. & Thoma, G. (2014a). Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative imaging in medicine and surgery.* [Online]. 4 (6). p.pp. 475–477. Available from: <https://doi.org/10.3978/j.issn.2223-4292.2014.11.20>. [Accessed: 28 March 2021].
- Jaeger, S., Karargyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., Xue, Z., Palaniappan, K., Singh, R.K., Antani, S., Thoma, G., Wang, Y.X., Lu, P.X. & McDonald, C.J. (2014b). Automatic tuberculosis screening using chest radiographs. *IEEE Transactions on Medical Imaging.* [Online]. 33 (2). p.pp. 233–245. Available from: <https://ieeexplore.ieee.org/abstract/document/6616679>.
- Jain, R., Gupta, M., Taneja, S. & Hemanth, D.J. (2021). Deep learning based detection and analysis of COVID-19 on chest X-ray images Rachna. *2021 International Conference on Information and Communication Technology for Sustainable Development, ICICT4SD 2021 - Proceedings.* [Online]. p.pp. 210–214. Available from: <https://doi.org/10.1007/s10489-020-01902-1>.
- Jia Deng, Wei Dong, Socher, R., Li-Jia Li, Kai Li & Li Fei-Fei (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition.* [Online]. p.pp. 248–255. Available from: <https://doi.org/10.1109/CVPR.2009.5206848>.
- Johnson, A.E.W., Pollard, T.J., Berkowitz, S.J., Greenbaum, N.R., Lungren, M.P., Deng, C. Ying, Mark, R.G. & Horng, S. (2019). MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific Data.* [Online]. 6 (1). p.p. 317. Available from: <https://doi.org/10.1038/s41597-019-0322-0>. [Accessed: 3 April 2021].
- Karargyris, A., Kashyap, S., Wu, J.T., Sharma, A., Moradi, M. & Syeda-Mahmood, T. (2019). Age prediction using a large chest x-ray dataset. In: K. Mori & H. K. Hahn (eds.). *Medical Imaging 2019: Computer-Aided Diagnosis.* [Online]. 2019, SPIE, pp. 468–476. Available from: <https://doi.org/10.1117/12.2512922>.
- Krizhevsky, A., Sutskever, I. & Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1.* [Online]. 25. p.pp. 1097--1105. Available from: <https://dl.acm.org/doi/10.5555/2999134.2999257>.

- Lee, J., Jun, S., Cho, Y., Lee, H., Kim, G.B., Seo, J.B. & Kim, N. (2017). Deep learning in medical imaging: general overview. *Korean Journal of Radiology*. [Online]. 18 (4). p.pp. 570–584. Available from: <https://doi.org/10.3348/kjr.2017.18.4.570>.
- Li, Y., Wang, H. & Luo, Y. (2020). A Comparison of Pre-trained Vision-and-Language Models for Multimodal Representation Learning across Medical Images and Reports. *Proceedings - 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2020*. [Online]. p.pp. 1999–2004. Available from: <https://ieeexplore.ieee.org/abstract/document/9313289>.
- Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B. & Sánchez, C.I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*. [Online]. 42 (December 2012). p.pp. 60–88. Available from: https://www.sciencedirect.com/science/article/pii/S1361841517301135?casa_token=i4Jfj1hb0FsAAAAA:VRLvfzkbvg9Y07VAmcTSmEiLFILhhSxl5qOEbp5X8fh5a_rrl9M5zDP52Tsv_MtbCdnljc0HLy8Rzw.
- Liu, G., Hsu, T.-M.H., McDermott, M., Boag, W., Weng, W.-H., Szolovits, P. & Ghassemi, M. (2019). Clinically Accurate Chest X-Ray Report Generation. *Proceedings of the 4th Machine Learning for Healthcare Conference*. [Online]. 106. p.pp. 249–269. Available from: <http://proceedings.mlr.press/v106/liu19a.html>.
- Lo, S.-C.B., Lou, S.-L.A., Lin, J.-S., Freedman, M.T., Chien, M. V & Mun, S.K. (1995). Artificial convolution neural network techniques and applications for lung nodule detection. *Ieee Transactions on Medical Imaging*. [Online]. 14 (4). p.pp. 711–718. Available from: <https://doi.org/10.1109/42.476112>.
- Maier, A., Syben, C., Lasser, T. & Riess, C. (2019). A gentle introduction to deep learning in medical image processing. *Zeitschrift fur Medizinische Physik*. [Online]. 29 (2). p.pp. 86–101. Available from: <https://doi.org/10.1016/j.zemedi.2018.12.003>.
- Majdi, M.S., Salman, K.N., Morris, M.F., Merchant, N.C. & Rodriguez, J.J. (2020). Deep Learning Classification of Chest X-Ray Images. *2020 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*. [Online]. p.pp. 116–119. Available from: <https://ieeexplore.ieee.org/document/9094612>.
- Norgeot, B., Glicksberg, B.S. & Butte, A.J. (2019). A call for deep-learning healthcare. *Nature Medicine*. [Online]. 25 (1) p.pp. 14–15. Available from: <https://www.nature.com/articles/s41591-018-0320-3>. [Accessed: 28 May 2021].
- Pan, I., Thodberg, H.H., Halabi, S.S., Kalpathy-Cramer, J. & Larson, D.B. (2019). Improving

- Automated Pediatric Bone Age Estimation Using Ensembles of Models from the 2017 RSNA Machine Learning Challenge. *Radiology: Artificial Intelligence*. [Online]. 1 (6). p.p. e190053. Available from: <https://doi.org/10.1148/ryai.2019190053>.
- Panayides, A.S., Amini, A., Filipovic, N.D., Sharma, A., Tsafaris, S.A., Young, A., Foran, D., Do, N., Golemati, S., Kurc, T., Huang, K., Nikita, K.S., Veasey, B.P., Zervakis, M., Saltz, J.H. & Pattichis, C.S. (2020). AI in Medical Imaging Informatics: Current Challenges and Future Directions. *IEEE Journal of Biomedical and Health Informatics*. [Online]. 24 (7). p.pp. 1837–1857. Available from: <https://doi.org/10.1109/JBHI.2020.2991043>.
- Pasa, F., Golkov, V., Pfeiffer, F., Cremers, D. & Pfeiffer, D. (2019). Efficient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization. *Scientific Reports*. [Online]. 9 (1). p.pp. 2–10. Available from: <https://doi.org/10.1038/s41598-019-42557-4>.
- Pham, H.H., Le, T.T., Tran, D.Q., Ngo, D.T. & Nguyen, H.Q. (2021). Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels. *Neurocomputing*. [Online]. 437. p.pp. 186–194. Available from: <https://arxiv.org/pdf/1911.06475.pdf>.
- Rajaraman, S., Kim, I. & Antani, S.K. (2020). Detection and visualization of abnormality in chest radiographs using modality-specific convolutional neural network ensembles. *PeerJ*. [Online]. 8. p.p. e8693. Available from: <https://peerj.com/articles/8693/>.
- Rajpurkar, P., Irvin, J., Bagul, A., Ding, D., Duan, T., Mehta, H., Yang, B., Zhu, K., Laird, D., Ball, R.L., Langlotz, C., Shpanskaya, K., Lungren, M.P. & Ng, A.Y. (2017a). MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs. *arXiv preprint arXiv:1712.06957*. [Online]. p.pp. 1–10. Available from: <http://arxiv.org/abs/1712.06957>.
- Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Ball, R.L., Langlotz, C., Shpanskaya, K., Lungren, M.P. & Ng, A.Y. (2017b). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv preprint arXiv:1711.05225*. [Online]. Available from: <https://arxiv.org/pdf/1711.05225.pdf>.
- Razzak, M.I., Naz, S. & Zaib, A. (2018). Deep learning for medical image processing: Overview, challenges and the future. *Lecture Notes in Computational Vision and Biomechanics*. [Online]. 26. p.pp. 323–350. Available from: https://doi.org/10.1007/978-3-319-65981-7_12.
- Rosman^{1*}, D., Nshizirungu², J.J., Emmanuel Rudakemwa², Crispin Moshi³, Jean de Dieu Tuyisenge³, E.U. & Louise Kalisa (2015). Imaging in the Land of 1000 Hills: Rwanda Radiology Country Report. *Journal of Global Radiology*. [Online]. 1 (1). p.pp. 1–7.

- Available from: <https://doi.org/10.7191/jgr.2015.1004>.
- Ruder, S. (2017). An Overview of Multi-Task Learning in Deep Neural Networks. *arXiv preprint arXiv:1706.05098*. [Online]. Available from: <https://arxiv.org/pdf/1706.05098.pdf>.
- Souza, J.C., Bandeira Diniz, J.O., Ferreira, J.L., França da Silva, G.L., Corrêa Silva, A. & de Paiva, A.C. (2019). An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks. *Computer Methods and Programs in Biomedicine*. [Online]. 177. p.pp. 285–296. Available from: <https://doi.org/10.1016/j.cmpb.2019.06.005>.
- Suzuki, K. (2017). Overview of deep learning in medical imaging. *Radiological Physics and Technology*. [Online]. 10 (3) p.pp. 257–273. Available from: <https://link.springer.com/article/10.1007/s12194-017-0406-5>. [Accessed: 31 March 2021].
- Tang, Y., Tang, Y., Han, M. & Summers, R.M. (2019). Abnormal Chest X-Ray Identification With Generative Adversarial One-Class Classifier Imaging Biomarkers and Computer-Aided Diagnosis Lab , Radiology and Imaging Sciences , National Institutes of Health (NIH) Clinical Center , Bethesda , USA ; Ping An Te. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. [Online]. (Isbi). p.pp. 1358–1361. Available from: <https://doi.org/10.1109/ISBI.2019.8759442>.
- Tanzi, L., Vezzetti, E., Moreno, R. & Moos, S. (2020). X-Ray bone fracture classification using deep learning: A baseline for designing a reliable approach. *Applied Sciences (Switzerland)*. [Online]. 10 (4). Available from: https://share.osteoporosis.foundation/EU-6-Material/Reports/IOF Report_EU.pdf.
- Tataru, C., Yi, D., Shenoyas, A. & Ma, A. (2017). *Deep Learning for abnormality detection in Chest X-Ray images*. [Online]. Available from: <http://cs231n.stanford.edu/reports/2017/pdfs/527.pdf>.
- The Royal College of Radiologists (2016). Clinical radiology UK workforce census 2016 report [BFCR(17)6]. *Www.Rcr.Ac.Uk*. [Online]. (October). p.pp. 1–54. Available from: https://www.rcr.ac.uk/system/files/publication/field_publication_files/cr_workforce_census_2016_report_0.pdf.
- Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M. & Summers, R.M. (2017). ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. [Online]. 2017-Janua. p.pp. 3462–

3471. Available from: https://openaccess.thecvf.com/content_cvpr_2017/html/Wang_ChestX-ray8_Hospital-Scale_Chest_CVPR_2017_paper.html.
- World Health Organization (2016). *To X-ray or not to X-ray?* [Online]. 2016. World Health Organization-NewsLetter. Available from: <https://www.who.int/news-room/feature-stories/detail/to-x-ray-or-not-to-x-ray->. [Accessed: 17 June 2021].
- Yadav, O., Passi, K. & Jain, C.K. (2019). Using Deep Learning to Classify X-ray Images of Potential Tuberculosis Patients. *Proceedings - 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018*. [Online]. p.pp. 2368–2375. Available from: <https://doi.org/10.1109/BIBM.2018.8621525>.
- Yao, L., Poblenz, E., Dagunts, D., Covington, B., Bernard, D. & Lyman, K. (2017). Learning to diagnose from scratch by exploiting dependencies among labels. *arXiv preprint arXiv:1710.10501*. [Online]. p.pp. 1–12. Available from: <https://arxiv.org/pdf/1710.10501.pdf>.
- Zech, J.R., Badgeley, M.A., Liu, M., Costa, A.B., Titano, J.J. & Oermann, E.K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study A. Sheikh (ed.). *PLOS Medicine*. [Online]. 15 (11). p.p. e1002683. Available from: <https://dx.plos.org/10.1371/journal.pmed.1002683>. [Accessed: 3 April 2021].
- Zhang, Q., Bai, C., Liu, Z., Yang, L.T., Yu, H., Zhao, J. & Yuan, H. (2020). A GPU-based residual network for medical image classification in smart medicine. *Information Sciences*. [Online]. 536. p.pp. 91–100. Available from: <https://doi.org/10.1016/j.ins.2020.05.013>.

APPENDIX A: RESEARCH PROPOSAL

Multitask X-ray Image Diagnosis Using Deep Learning

Satyanarayana Murthy T

Research Proposal

Program: MSc Artificial Intelligence and Machine Learning.

Thesis Supervisor: Prasanna Porwal

Abstract

X-ray imaging is most common and cost-effective radiology procedure for diagnosing multiple diseases. But availability of radiologists to meet demand is very less. This brings importance of need for automating the diagnosis using X-ray procedures for addressing radiologist scarcity or radiologist fatigue problem.

Recently deep learning methods have shown prominence in automatic analysis of X-ray images whose performance approximates to human radiologist level, but focus is on analysis of improving metrics for pathology analysis. Analysis of deep learning methods that can analyse multiple tasks like a radiologist, is less explored.

Current research focuses on developing multitask deep learning methods for analysis of chest X-ray reports. Multitask deep learning methods shall be used for improving on combined accuracies of X-ray abnormality detection and age prediction. As part of research, literature review shall be conducted to gain systematic understanding of the research area in focus. Furthermore, chest X-ray dataset shall be used for developing and evaluating multitask deep learning methods.

This research shall add significant contribution to medical imaging field by providing framework for automated comprehensive analysis of X-ray reports to assist radiologist in providing enhanced care to patients.

Keywords: Multitasking, Deep Learning, X-ray Analysis, Computer-aided Diagnosis (CAD), Computer Vision, Medical Imaging

Table Of Contents

Abstract	80
List of Figures	82
Acronyms	82
Introduction	82
Proposal Outline:	7
Background and Related Works	83
Chest X-ray pathology Analysis:	13
Chest X-ray multitask pathology analysis:	23
Chest X-ray Age Regression Analysis:	25
Summary:	29
Research Questions	4
Aim and Objectives	5
Aim:	5
Objectives:	5
Significance of Study	5
Scope of Study	6
Research Methodology	31
Dataset Description:	32
Pathology information:	32
Age Information:	88
Data Preparation and Transformation:	35
Model Development:	36
Model Evaluation:	40
Risks and Contingency Plan:	90
Expected Outcome	91
Required Resources	91
Research Plan	91

References 93

List of Figures

Figure 1: CheXNet (Rajpurkar et al., 2017).	84
Figure 2: CheXpert dataset consisting of 14 Labelled Observations (Irvin et al., 2019)	88
Figure 3: (A)Frontal and Lateral chest X-ray view of a patient. (B) Proportion of frontal and lateral views in dataset.	88
Figure 4: (a) Age Distribution of patients in dataset. (b) Age distribution of patients with “No Finding” label in dataset.	34
Figure 5: (a) Uncertain cases marked as normal i.e., 0 (b) Uncertain cases marked as 1 i.e., 1 (c) Uncertain cases left without change i.e., -1	90
Figure 6: Overall modelling approach for multitask model to detect abnormal X-ray and Age value	39

Acronyms

GAN	Generative Adversarial Networks
DL	Deep Learning
CNN	Convolutional Neural Network
SVM	Support Vector Machine
CAD	Computer Aided Diagnosis
PA	posteroanterior
MTL	Multitask Learning

Introduction

X-ray imaging, part of radiology imaging, is used for diagnosing multiple life-threatening diseases like Tuberculosis, pneumonia, Bone Fractures, foreign objects etc. But availability of radiologists to meet demand is very less. For example, In US there is approximately one radiologist for every 10,000 population (Arora, 2014), in UK it is 7.5:100,000(The Royal College of Radiologists, 2016). While in populous country like India it is 1:100,000 (Arora, 2014) and in countries of Africa it is even lesser (Rosman1* et al., 2015; Ali et al., 2015). Because of less ratio of available radiologists to patients and increasing demand, radiologist fatigue and limited reach of diagnostic availability to people is happening(Arora, 2014). Hence need for automating the diagnosis of X-ray procedures, for assisting radiologist became an important problem to be addressed in immediate future.

Prior work in field of automated X-ray analysis mainly focuses on improving pathology identification from chest X-rays. But, for automated patient diagnosis to work in practical considerations identifying disease pathology alone is not sufficient, additional parameters like patient body condition, body age is also needed for proper patient treatment.

Current research focuses on investigation of approaches using multitask deep learning methods for pathology analysis as well as Age prediction from chest X-rays. Getting such sophisticated analysis information shall help in improved patient care, where patient diagnosis or treatment is not just based on pathology but also on patient body condition in relation to social context, which refers to idea of precision healthcare (Colijn et al., 2017).

Proposal Outline:

Remainder chapters of this proposal are organized as follows. Related works section describes systematic study of prior work along with research gaps identified. Research aim & objectives along with significance & scope of study will be described. Research methodology section proposes investigation of approaches to address identified research gaps. Expected outcome and needed resources to conduct research shall be indicated. Finally, research plan for project management shall be indicated.

Background and Related Works

X-ray imaging, part of radiology imaging, is used for diagnosing multiple life-threatening problems like Tuberculosis, pneumonia, Bone Fractures, Dental problems, foreign objects etc. Having multiple applications and limited availability of radiologists is creating limited access to radiologists, hence enabling Computer Aided Diagnosis (CAD) to assist radiologist became important problem to be addressed.

In last decade deep learning has gained momentum after CNN won ImageNet competition though prior to that some progress was made in this field (Suzuki, 2017). With success of CNN, research for automating different tasks of medical imaging using deep learning has gained focus (Maier et al., 2019; Greenspan et al., 2016; Lee et al., 2017).

For purpose of this study, scope is limited to investigation of automated diagnosis methods for chest X-ray. Chest X-ray diagnosis is commonly used procedure for diagnosing life-threatening diseases. This makes it an important area in X-ray imaging for need of automating analysis. So far, mainly pathology analysis (classification, pathology segmentation and detection) has been focus area of automating chest X-ray analysis, but with limited research focus on age regression analysis.

Chest X-ray pathology Analysis:

In last decade, research focus for automating chest X-ray diagnosis increased with release of two public datasets focusing on detecting tuberculosis from chest X-rays (Jaeger et al., 2014a). Automatic method for tuberculosis screening using SVM has been developed for practical use in Kenya (Jaeger et al., 2014b). Multiple approaches like augmenting samples, transfer learning methods have been explored. These methods showed better results, approximating human performance, in automation of tuberculosis detection (Pasa et al., 2019; Yadav et al., 2019; Imran & Terzopoulos, 2019; Zhang et al., 2020).

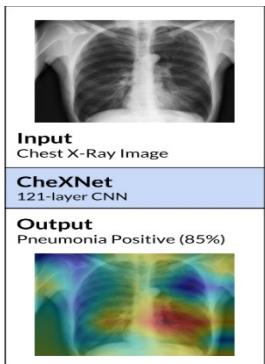


Figure 1:CheXNet (*Rajpurkar et al., 2017b*).

Apart from only tuberculosis detection, additionally with release of large scale public dataset Chest X-ray 8 (later extended to 14) (Wang et al., 2017), pointing to multiple common thorax diseases, significant progress has been made in this field. (Rajpurkar et al., 2017b) developed model CheXNet (shown in figure 1) using DenseNet 121-layer model achieving state of art result and even exceeding in few cases. (Yao et al., 2017) worked on improving model performance across multiple classes by using additional RNN decoder network and capturing inter label dependencies. (Zech et al., 2018) studied performance implications of CNN architectures on unseen external datasets acquired in different environment setting for chest radiograph images. (Bozorgtabar et al., 2020) showed improved anomaly detection using Memory bank, aggregate learning methods on Chest X-ray 8(14) dataset.

Since labels in Chest X-ray 8 dataset are encoded using automatic labeller, it will be problematic to use it for benchmarking. Hence CheXpert, a large scale database is introduced that features radiologist-labelled test and validation sets to serve as strong reference standards (Irvin et al., 2019). On this dataset (Irvin et al., 2019) proposed use of DenseNet-121 layer for which they got best results in their experiments. (Allaouzi & Ben Ahmed, 2019) also used DenseNet-121 Layer architecture for feature extraction and applied multiple problem transformation methods for classification and showed state of art result on both CheXpert and Chest X-ray 14 datasets.

From above study, it is observed use of deep learning methods has been in state of art methods in enabling automated X-ray diagnosis.

Chest X-ray multitask pathology analysis:

Multitask Learning (MTL) is a method where model gets trained on multiple, but related tasks. MTL helps in improving generalization of model by sharing representation between tasks. In general, not necessary, when we try to optimize multiple loss functions it is called multitask learning (Ruder, 2017).

Using MTL, several attempts were made to train models on multiple tasks like Localization, Segmentation, and classification together for pathology classification from chest X-ray. Performance metrics for pathology detection were shown to generalize better when multiple tasks are trained together (Gündel et al., 2019; Imran & Terzopoulos, 2019). MTL is also shown to improve predictions on datasets with less samples like COVID detection (Alom et al., 2020; Farag et al., 2020) thereby indicating improved generalizability of model performance.

Chest X-ray Age Regression Analysis:

Deep Learning has also been useful in regression analysis of X-ray images, in applications like predicting bone age from hand X-ray and biological age from chest X-ray applications.

RSNA paediatric bone age prediction challenge was created for purpose of assisting radiologist in automatically determining skeletal maturity of paediatric population, using X-rays. Top 3 contestants in the challenge used deep learning to achieve better accuracy, which shows importance of deep learning in CAD for medical imaging (Halabi et al., 2019).

Age regression from chest X-ray is less explored area, where it can be used for preventive counselling by comparing to average health status (Karargyris et al., 2019). Age value prediction is made on Chest X-ray 8 dataset and achieved best recall value of 0.94 for posteroanterior (PA) view when +/- 9 years is considered error range (Karargyris et al., 2019).

Summary:

After conducting above literature review, below gaps are identified:

Several attempts were made to improve automated analysis of specific task of X-ray imaging i.e., either chest X-ray pathology analysis or Age regression separately. Above methods doesn't consider idea of precision healthcare i.e., diagnosing patient based on his condition as whole. Though multitask methods were proposed, they were also focused on pathology analysis specifically.

Above methods work as solutions to specific problem statements. They don't intend to understand patient diagnosis completely like radiologist who looks at process of patient diagnosis with multiple parameters.

Study conducted outlines literature review and explains important research works for automating chest X-ray diagnosis and throws light on research gaps identified. Though these studies don't consider handling patient diagnosis from multiple perspectives, but they provide invaluable insight to our research. Further, Research methodology section explains methods being considered for addressing the research gaps.

Research Questions

What methods can be used for concurrently determining both disease abnormality classification and age regression from chest X-ray images?

Aim and Objectives

Aim:

Aim of this research is to investigate methods that shall support analysis of patient's chest X-ray to concurrently determine presence of abnormality as well as patients age, for assisting radiologist by providing a comprehensive diagnosis information.

Objectives:

Based on aim of this study, research objectives formulated are as follows:

To conduct literature study for gaining systematic understanding of approaches that does automated analysis patient's chest X-ray.

To develop multitask deep learning method that can analyse chest X-ray for abnormality detection and predict patient's age.

To evaluate performance of multitask method being developed.

Significance of Study

Current research focuses on predicting additional parameters, Age value in this research, from chest X-ray in addition to pathology analysis. This comprehensive analysis information shall

support in enabling precision healthcare where, patient diagnosis is not just based on pathology but also based on his body condition and other parameters.

Age value prediction can be used for preventive counselling of patients by comparing to average health status (Karargyris et al., 2019).

It provides X-ray diagnosis framework which works like a radiologist looking at multi-dimensional view of patient diagnosis.

Scope of Study

Though ideas suggested in this study are applicable to broader healthcare imaging, scope of current study is limited to:

Current study focuses on chest X-ray analysis.

Current study focuses on multitask analysis for abnormality detection and Age prediction from chest X-rays.

Research Methodology

Chest X-ray is commonly used and cost-effective imaging procedure for diagnosis of many life-threatening diseases. (Irvin et al., 2019) suggests that “*Automated chest radiograph interpretation at the level of practicing radiologists could provide substantial benefit in many medical settings, from improved workflow prioritization and clinical decision support to large-scale screening and global population health initiatives*”. Currently deep learning methods are state of art for exploring automatic chest radiograph analysis, whose performance approximates to radiologist.

Based on literature study conducted, most of deep learning methods are mainly focused on X-ray pathology analysis. Very few methods are focused on prediction of other parameters like Age value based on chest X-ray (Karargyris et al., 2019). So far very little research is made on multitasking methods where both pathology analysis and Age prediction are made.

This research proposes usage of multitask deep learning methods to determine Abnormality detection and Age value prediction from X-rays.

CheXpert dataset is used for investigation purpose in this research. It is a largescale chest radiograph dataset released for research purpose 1) with strong reference standards 2) with expert human performance metrics for comparison(Irvin et al., 2019). Metadata of this dataset includes Patient Age, Gender along with other Diagnosis related information.

Dataset Description:

CheXpert is “*a large dataset that contains 224,316 chest radiographs of 65,240 patients*”(Irvin et al., 2019), collected from Stanford Hospital.

Pathology information:

Dataset includes information of 14 observations that include 12 pathologies along with “Supported Devices” and “No Finding” observations. For each patient, information specific to 14 observations is recorded as Positive, Negative and Uncertain. Blank indicates no information found for that observation. Observations are indicated as positive for “No Finding” label, when there is no pathology classified as positive or uncertain for that image (Irvin et al., 2019). Figure 2 indicates observations made on dataset against each pathology.

Pathology	Positive (%)	Uncertain (%)	Negative (%)
No Finding	16627 (8.86)	0 (0.0)	171014 (91.14)
Enlarged Cardiom.	9020 (4.81)	10148 (5.41)	168473 (89.78)
Cardiomegaly	23002 (12.26)	6597 (3.52)	158042 (84.23)
Lung Lesion	6856 (3.65)	1071 (0.57)	179714 (95.78)
Lung Opacity	92669 (49.39)	4341 (2.31)	90631 (48.3)
Edema	48905 (26.06)	11571 (6.17)	127165 (67.77)
Consolidation	12730 (6.78)	23976 (12.78)	150935 (80.44)
Pneumonia	4576 (2.44)	15658 (8.34)	167407 (89.22)
Atelectasis	29333 (15.63)	29377 (15.66)	128931 (68.71)
Pneumothorax	17313 (9.23)	2663 (1.42)	167665 (89.35)
Pleural Effusion	75696 (40.34)	9419 (5.02)	102526 (54.64)
Pleural Other	2441 (1.3)	1771 (0.94)	183429 (97.76)
Fracture	7270 (3.87)	484 (0.26)	179887 (95.87)
Support Devices	105831 (56.4)	898 (0.48)	80912 (43.12)

Figure 2: CheXpert dataset consisting of 14 Labelled Observations (Irvin et al., 2019)

Additional information on Frontal or Lateral is also indicated. Frontal X-rays are images acquired by X-rays directed on either back/front of patient’s chest, while lateral X-rays are images acquired from side of patient body (shown in figure 3-A). Figure 3-B indicates proportion of frontal and lateral views of patient in the dataset.

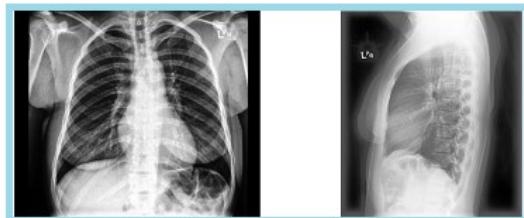


Figure A

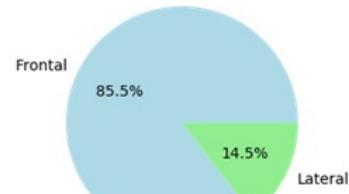


Figure B

Figure 3: (A)Frontal and Lateral chest X-ray view of a patient. (B) Proportion of frontal and lateral views in dataset.

Age Information:

CheXpert Dataset also includes information on patients Age for both normal and abnormal X-ray’s. Figure 4-A indicates age distribution of patients, where most of patients who went for examination are in 55-65 age group. Figure 4-B indicates out of total patients, “No Finding” observation is more relevant in patients with less than 40 years.

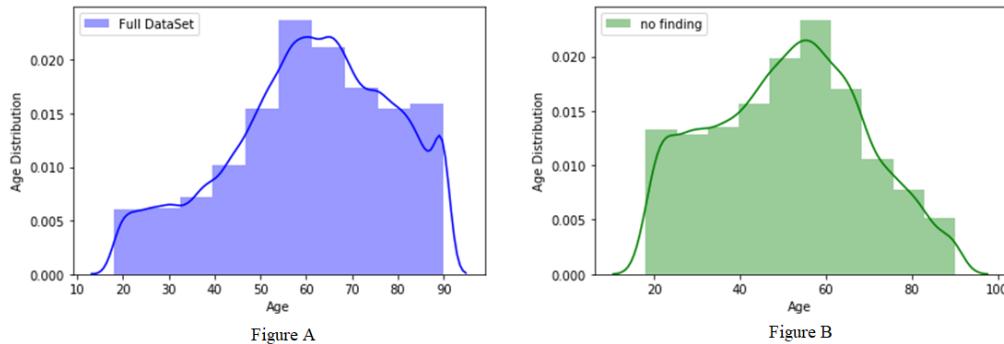


Figure 4: (a) Age Distribution of patients in dataset. (b) Age distribution of patients with “No Finding” label in dataset.

Data Preparation and Transformation:

One objective of this research is abnormality detection of chest X-rays. For abnormality detection, current pathology information of 12 variables is converted to binary label, where if any of disease is positive then the X-ray is marked as Abnormal.

Figure 5 show count of ground truth labels formed after transformation of multi-class labels to binary class of Abnormality with different transformation methods. Uncertain cases are handled as Normal (treated as zeros), Abnormal (treated as ones) or as separate uncertain class (Irvin et al., 2019). Modelling process shall be applied on 3 different cases to determine best possible result.

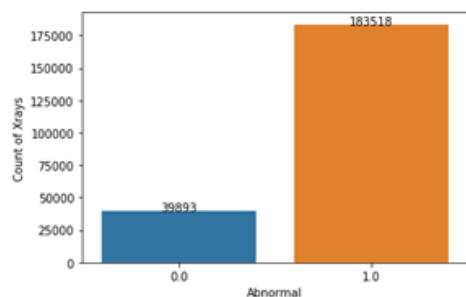


Figure A

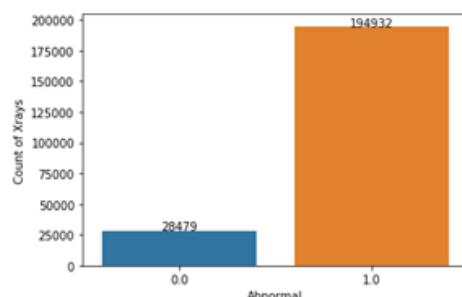


Figure B

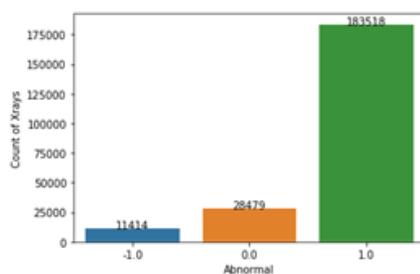


Figure C

Figure 5: (a) Uncertain cases marked as normal i.e., 0 (b) Uncertain cases marked as 1 i.e., 1
(c) Uncertain cases left without change i.e., -1

Model Development:

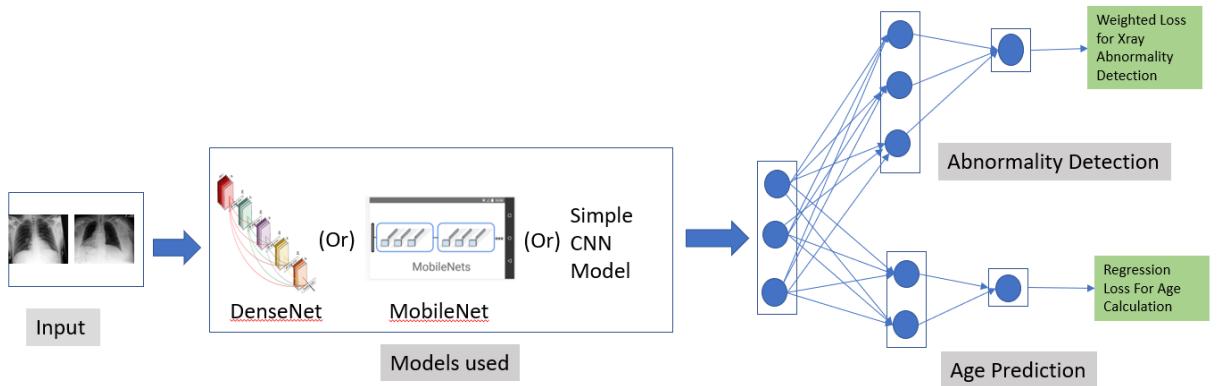


Figure 6: Overall modelling approach for multitasked model to detect abnormal X-ray and Age value

Figure 6 indicates overall approach of Multitask modelling for prediction of both Abnormal X-ray and Age value from chest X-ray images.

For feasibility study, DenseNet (Huang et al., 2017) or MobileNet (Howard et al., 2017) or a simple CNN architecture shall be explored.

For training model, combination of below loss functions is considered for multitask analysis,

From pre-processing state, it is observed that there is imbalance of image classes. Hence weighted binary cross entropy loss shall be used for Abnormal classification of X-ray images.

$$“L(X, y) = -[(w+) \cdot y \log p(Y = 1|X)] - [(w-) \cdot (1 - y) \log p(Y = 0|X)]”$$
 (Rajpurkar et al., 2017b)

“where $p(Y = i|X)$ is the probability that the network assigns to the label i , $w+ = |N|/(|P|+|N|)$, and $w- = |P|/(|P|+|N|)$ with $|P|$ and $|N|$ are number of positive and negative Abnormal cases respectively” (Rajpurkar et al., 2017b).

Age value shall be trained using Mean Squared Error (MSE) or Mean Absolute Error (MAE) or Sigmoid Loss functions.

Model Evaluation:

CheXpert dataset provides 200 X-rays as validation data. This information shall be used for evaluating models developed.

Model evaluation for performance is based on balanced accuracy metrics of

Model classification accuracy for abnormal X-rays i.e., accuracy, Precision, Recall.

Age prediction accuracy from X-rays i.e., R2, Adjusted R2.

Risks and Contingency Plan:

CheXpert is large dataset and might need heavy access to computation resources that are not foreseen during proposal.

Will still be able meet aim and objectives properly by analyzing on dataset where split is based on pathology observations.

Complexity of data may not allow time to explore multiple methods as part of this research

Will still be able to meet aim and objectives by working on one method that can analyse and make predictions.

Expected Outcome

Expected outcome of this research shall be,

A study of systematic understanding of previous automated chest X-ray analysis methods.

A model that can predict below outcomes by providing a comprehensive analysis of X-ray image input.

Predict if input chest X-ray is abnormal or not.

Predict Age value from input chest X-ray.

Required Resources

CheXpert: Publicly available dataset for research purposes(Irvin et al., 2019).

CPU with 16GB RAM

GPU with at least 4GB RAM access for training neural network models

Keras, TensorFlow, python and other visualization libraries.

Research Plan

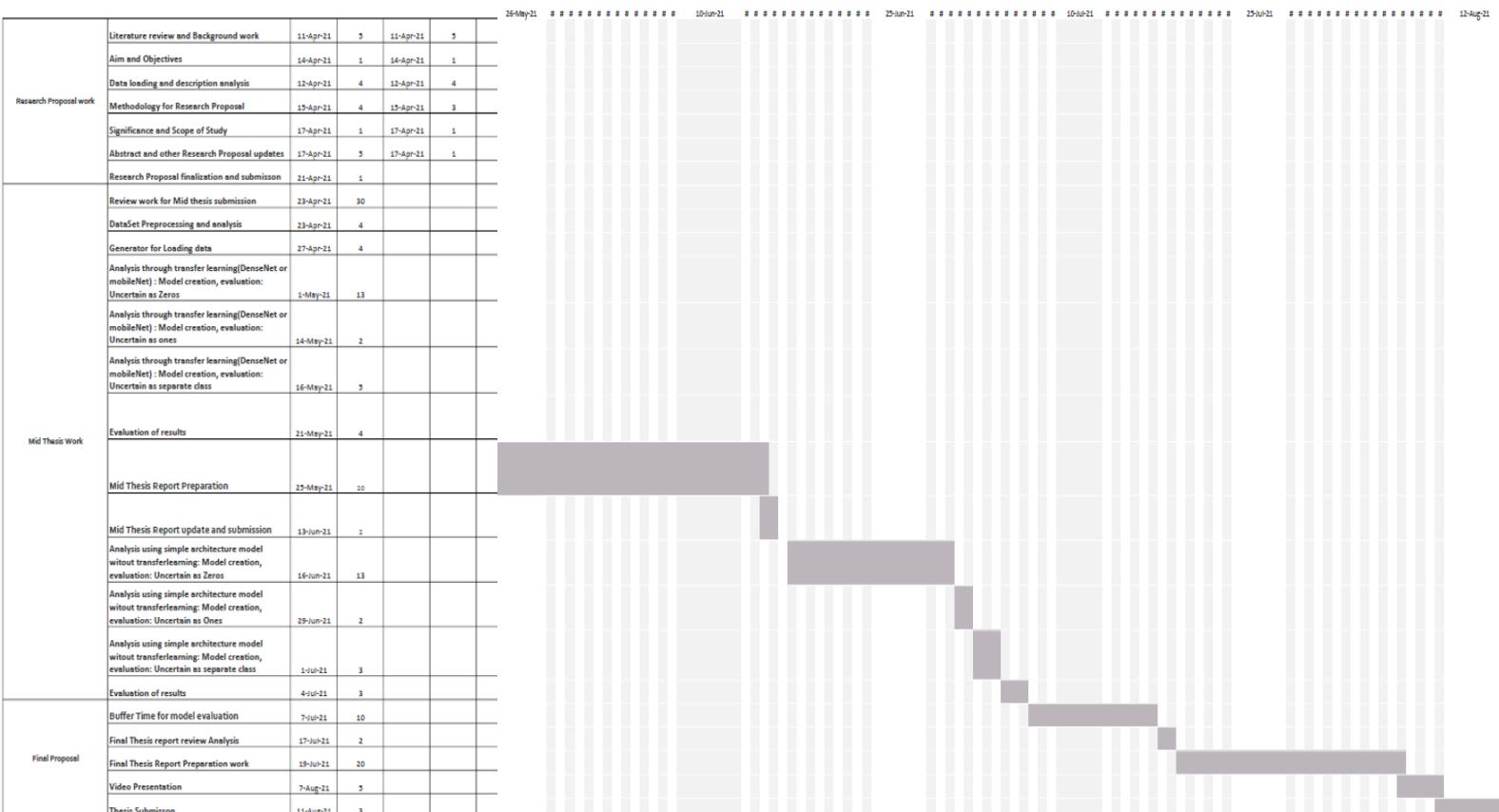
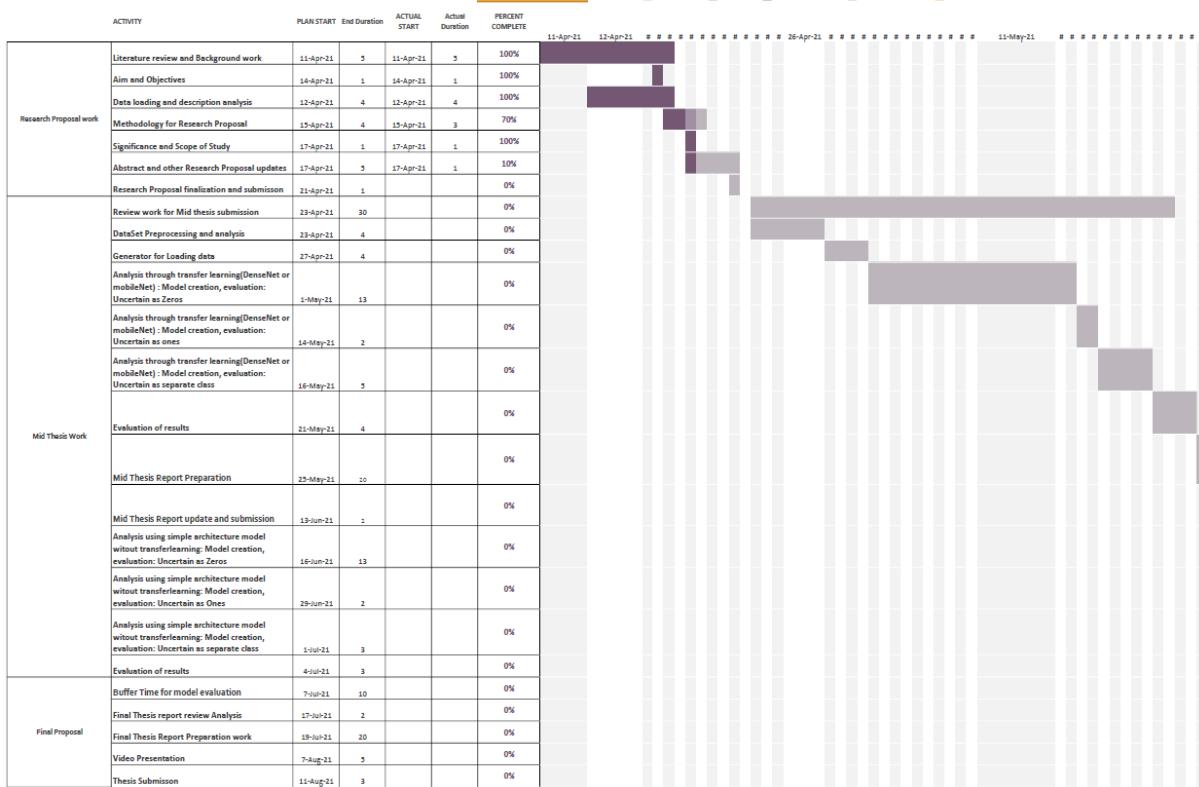
Project Planner

Select a period to highlight or right. A legend describing the charting follows.

Period Highlight:

Plan Duration Actual Start % Complete

Actual (beyond plan) % Complete (beyond plan)



References

- Ali, F., Harrington, S., Kennedy, S. & Hussain, S. (2015). Diagnostic Radiology in Liberia: A Country Report. *Journal of Global Radiology*. [Online]. 1 (2). p.p. 6. Available from: <http://dx.doi.org/10.7191/jgr.2015.1020>.
- Allaouzi, I. & Ben Ahmed, M. (2019). A Novel Approach for Multi-Label Chest X-Ray Classification of Common Thorax Diseases. *IEEE Access*. [Online]. 7. p.pp. 64279–64288. Available from: <http://dx.doi.org/10.1109/ACCESS.2019.2916849>.
- Alom, M.Z., Shaifur Rahman, M.M., Nasrin, S., Taha, T.M. & Asari, V.K. (2020). COVID_MTNet: COVID-19 detection with multi-task deep learning approaches. *arXiv preprint arXiv:2004.03747*. [Online]. Available from: <https://arxiv.org/ftp/arxiv/papers/2004/2004.03747.pdf>.
- Arora, R. (2014). The training and practice of radiology in India: current trends. *Quantitative imaging in medicine and surgery*. [Online]. 4 (6). p.pp. 449–44950. Available from: <http://dx.doi.org/10.3978/j.issn.2223-4292.2014.11.04>.
- Benjamens, S., Dhunnoo, P. & Meskó, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digital Medicine*. [Online]. 3 (1). p.pp. 1–8. Available from: <http://dx.doi.org/10.1038/s41746-020-00324-0>.
- Boag, W., Hsu, T.-M.H., Mcdermott, M., Berner, G., Alsentzer, E., Szolovits, P., Dalca, A. V, Finlayson, S., Oberst, M., Falck, F. & Beaulieu-Jones, B. (2020). Baselines for Chest X-Ray Report Generation. *Machine Learning for Health NeurIPS Workshops*. [Online]. 116. p.pp. 126–140. Available from: <http://proceedings.mlr.press/v116/boag20a.html>.
- Bozorgtabar, B., Mahapatra, D., Vray, G. & Thiran, J.P. (2020). Salad: Self-supervised aggregation learning for anomaly detection on x-rays. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [Online]. 12261 LNCS. p.pp. 468–478. Available from: https://doi.org/10.1007/978-3-030-59710-8_46.
- Bram van Ginneken & Sjoerd Kerkstra (2012). *Grand Challenge*. [Online]. 2012. Available from: <https://grand-challenge.org/>. [Accessed: 25 June 2021].
- Burrill, J., Williams, C.J., Bain, G., Conder, G., Hine, A.L. & Misra, R.R. (2007). Tuberculosis: A radiologic review. *Radiographics*. [Online]. 27 (5). p.pp. 1255–1273. Available from: <https://pubs.rsna.org/doi/pdf/10.1148/rg.275065176>.

- Colijn, C., Jones, N., Johnston, I.G., Yaliraki, S. & Barahona, M. (2017). Toward precision healthcare: Context and mathematical challenges. *Frontiers in Physiology*. [Online]. 8 (MAR). p.p. 136. Available from: <http://dx.doi.org/10.3389/fphys.2017.00136>. [Accessed: 10 April 2021].
- Dai, W., Dong, N., Wang, Z., Liang, X., Zhang, H. & Xing, E.P. (2018). Scan: Structure correcting adversarial network for organ segmentation in chest x-rays. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [Online]. 11045 LNCS. p.pp. 263–273. Available from: https://doi.org/10.1007/978-3-030-00889-5_30.
- Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., Cui, C., Corrado, G., Thrun, S. & Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*. [Online]. 25 (1). p.pp. 24–29. Available from: <http://dx.doi.org/10.1038/s41591-018-0316-z>.
- Farag, A.T., El-Wahab, A.R.A., Nada, M., El-Hakeem, M.Y.A., Mahmoud, O.S., Rashwan, R.K. & El Sallab, A. (2020). MulticheXnet: A multi-task learning deep network for pneumonia-like diseases diagnosis from X-ray scans. *arXiv preprint arXiv:2008.01973*. [Online]. Available from: <https://arxiv.org/pdf/2008.01973.pdf>.
- Gordienko, Y., Gang, P., Hui, J., Zeng, W., Kochura, Y., Alienin, O., Rokovyi, O. & Stirenko, S. (2019). Deep learning with lung segmentation and bone shadow exclusion techniques for chest X-ray analysis of lung cancer. *Advances in Intelligent Systems and Computing*. [Online]. 754 (Advances in Computer Science for Engineering and Education. ICCSEEA 2018.). p.pp. 638–647. Available from: <https://ui.adsabs.harvard.edu/abs/2017arXiv171207632G>.
- Greenspan, H., Van Ginneken, B. & Summers, R.M. (2016). Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging*. [Online]. 35 (5). p.pp. 1153–1159. Available from: <https://ieeexplore.ieee.org/document/7463094>.
- Gündel, S., Ghesu, F.C., Grbic, S., Gibson, E., Georgescu, B., Maier, A. & Comaniciu, D. (2019). Multi-task learning for chest X-ray abnormality classification on noisy labels. *arXiv preprint arXiv:1905.06362*. [Online]. p.pp. 1–10. Available from: <https://arxiv.org/pdf/1905.06362.pdf>.
- Halabi, S.S., Prevedello, L.M., Kalpathy-Cramer, J., Mamonov, A.B., Bilbily, A., Cicero, M., Pan, I., Pereira, L.A., Sousa, R.T., Abdala, N., Kitamura, F.C., Thodberg, H.H., Chen, L., Shih, G., Andriole, K., Kohli, M.D., Erickson, B.J. & Flanders, A.E. (2019). The rSNA

- pediatric bone age machine learning challenge. *Radiology*. [Online]. 290 (3). p.pp. 498–503. Available from: <https://doi.org/10.1148/radiol.2018180736>.
- He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. [Online]. 2016-Decem. p.pp. 770–778. Available from: <https://arxiv.org/pdf/1512.03385.pdf>.
- Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. & Adam, H. (2017). MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*. [Online]. Available from: <https://arxiv.org/pdf/1704.04861.pdf>.
- Huang, G., Liu, Z., Van Der Maaten, L. & Weinberger, K.Q. (2017). Densely connected convolutional networks. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. [Online]. 2017-Janua. p.pp. 2261–2269. Available from: https://openaccess.thecvf.com/content_cvpr_2017/papers/Huang_Densely_Connected_Convolutional_CVPR_2017_paper.pdf.
- Iglovikov, V.I., Rakhlin, A., Kalinin, A.A. & Shvets, A.A. (2018). Paediatric bone age assessment using deep convolutional neural networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [Online]. 11045 LNCS. p.pp. 300–308. Available from: https://doi.org/10.1007/978-3-030-00889-5_34.
- Imran, A.A.Z. & Terzopoulos, D. (2019). Semi-supervised Multi-task Learning with Chest X-Ray Images. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. [Online]. 11861 LNCS. p.pp. 151–159. Available from: <https://arxiv.org/pdf/1908.03693.pdf>.
- International Osteoporosis Foundation. (2018). *Broken Bones, Broken Lives: A Roadmap to Solve the Fragility Fracture Crisis in Europe*. [Online]. 2018. International Osteoporosis Foundation. Available from: https://share.osteoporosis.foundation/EU-6-Material/Reports/IOF Report_EU.pdf. [Accessed: 15 June 2021].
- Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., Seekins, J., Mong, D.A., Halabi, S.S., Sandberg, J.K., Jones, R., Larson, D.B., Langlotz, C.P., Patel, B.N., Lungren, M.P. & Ng, A.Y. (2019). CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *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.* [Online]. p.pp. 590–597. Available from: <https://arxiv.org/pdf/1901.07031.pdf>.
- Jaeger, S., Candemir, S., Antani, S., Wang, Y.-X.J., Lu, P.-X. & Thoma, G. (2014a). Two public chest X-ray datasets for computer-aided screening of pulmonary diseases. *Quantitative imaging in medicine and surgery.* [Online]. 4 (6). p.pp. 475–477. Available from: <https://doi.org/10.3978/j.issn.2223-4292.2014.11.20>. [Accessed: 28 March 2021].
- Jaeger, S., Karargyris, A., Candemir, S., Folio, L., Siegelman, J., Callaghan, F., Xue, Z., Palaniappan, K., Singh, R.K., Antani, S., Thoma, G., Wang, Y.X., Lu, P.X. & McDonald, C.J. (2014b). Automatic tuberculosis screening using chest radiographs. *IEEE Transactions on Medical Imaging.* [Online]. 33 (2). p.pp. 233–245. Available from: <https://ieeexplore.ieee.org/abstract/document/6616679>.
- Jain, R., Gupta, M., Taneja, S. & Hemanth, D.J. (2021). Deep learning based detection and analysis of COVID-19 on chest X-ray images Rachna. *2021 International Conference on Information and Communication Technology for Sustainable Development, ICICT4SD 2021 - Proceedings.* [Online]. p.pp. 210–214. Available from: <https://doi.org/10.1007/s10489-020-01902-1>.
- Jia Deng, Wei Dong, Socher, R., Li-Jia Li, Kai Li & Li Fei-Fei (2009). ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition.* [Online]. p.pp. 248–255. Available from: <https://doi.org/10.1109/CVPR.2009.5206848>.
- Johnson, A.E.W., Pollard, T.J., Berkowitz, S.J., Greenbaum, N.R., Lungren, M.P., Deng, C. Ying, Mark, R.G. & Horng, S. (2019). MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific Data.* [Online]. 6 (1). p.p. 317. Available from: <https://doi.org/10.1038/s41597-019-0322-0>. [Accessed: 3 April 2021].
- Karargyris, A., Kashyap, S., Wu, J.T., Sharma, A., Moradi, M. & Syeda-Mahmood, T. (2019). Age prediction using a large chest x-ray dataset. In: K. Mori & H. K. Hahn (eds.). *Medical Imaging 2019: Computer-Aided Diagnosis.* [Online]. 2019, SPIE, pp. 468–476. Available from: <https://doi.org/10.1117/12.2512922>.
- Krizhevsky, A., Sutskever, I. & Hinton, G.E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1.* [Online]. 25. p.pp. 1097--1105. Available from: <https://dl.acm.org/doi/10.5555/2999134.2999257>.

- Lee, J., Jun, S., Cho, Y., Lee, H., Kim, G.B., Seo, J.B. & Kim, N. (2017). Deep learning in medical imaging: general overview. *Korean Journal of Radiology*. [Online]. 18 (4). p.pp. 570–584. Available from: <https://doi.org/10.3348/kjr.2017.18.4.570>.
- Li, Y., Wang, H. & Luo, Y. (2020). A Comparison of Pre-trained Vision-and-Language Models for Multimodal Representation Learning across Medical Images and Reports. *Proceedings - 2020 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2020*. [Online]. p.pp. 1999–2004. Available from: <https://ieeexplore.ieee.org/abstract/document/9313289>.
- Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M., van Ginneken, B. & Sánchez, C.I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*. [Online]. 42 (December 2012). p.pp. 60–88. Available from: https://www.sciencedirect.com/science/article/pii/S1361841517301135?casa_token=i4Jfj1hb0FsAAAAA:VRLvfzkbvg9Y07VAmcTSmEiLFILhhSx15qOEbp5X8fh5a_rrl9M5zDP52Tsv_MtbCdnljc0HLy8Rzw.
- Liu, G., Hsu, T.-M.H., McDermott, M., Boag, W., Weng, W.-H., Szolovits, P. & Ghassemi, M. (2019). Clinically Accurate Chest X-Ray Report Generation. *Proceedings of the 4th Machine Learning for Healthcare Conference*. [Online]. 106. p.pp. 249–269. Available from: <http://proceedings.mlr.press/v106/liu19a.html>.
- Lo, S.-C.B., Lou, S.-L.A., Lin, J.-S., Freedman, M.T., Chien, M. V & Mun, S.K. (1995). Artificial convolution neural network techniques and applications for lung nodule detection. *Ieee Transactions on Medical Imaging*. [Online]. 14 (4). p.pp. 711–718. Available from: <https://doi.org/10.1109/42.476112>.
- Maier, A., Syben, C., Lasser, T. & Riess, C. (2019). A gentle introduction to deep learning in medical image processing. *Zeitschrift fur Medizinische Physik*. [Online]. 29 (2). p.pp. 86–101. Available from: <https://doi.org/10.1016/j.zemedi.2018.12.003>.
- Majdi, M.S., Salman, K.N., Morris, M.F., Merchant, N.C. & Rodriguez, J.J. (2020). Deep Learning Classification of Chest X-Ray Images. *2020 IEEE Southwest Symposium on Image Analysis and Interpretation (SSIAI)*. [Online]. p.pp. 116–119. Available from: <https://ieeexplore.ieee.org/document/9094612>.
- Norgeot, B., Glicksberg, B.S. & Butte, A.J. (2019). A call for deep-learning healthcare. *Nature Medicine*. [Online]. 25 (1) p.pp. 14–15. Available from: <https://www.nature.com/articles/s41591-018-0320-3>. [Accessed: 28 May 2021].
- Pan, I., Thodberg, H.H., Halabi, S.S., Kalpathy-Cramer, J. & Larson, D.B. (2019). Improving

- Automated Pediatric Bone Age Estimation Using Ensembles of Models from the 2017 RSNA Machine Learning Challenge. *Radiology: Artificial Intelligence*. [Online]. 1 (6). p.p. e190053. Available from: <https://doi.org/10.1148/ryai.2019190053>.
- Panayides, A.S., Amini, A., Filipovic, N.D., Sharma, A., Tsafaris, S.A., Young, A., Foran, D., Do, N., Golemati, S., Kurc, T., Huang, K., Nikita, K.S., Veasey, B.P., Zervakis, M., Saltz, J.H. & Pattichis, C.S. (2020). AI in Medical Imaging Informatics: Current Challenges and Future Directions. *IEEE Journal of Biomedical and Health Informatics*. [Online]. 24 (7). p.pp. 1837–1857. Available from: <https://doi.org/10.1109/JBHI.2020.2991043>.
- Pasa, F., Golkov, V., Pfeiffer, F., Cremers, D. & Pfeiffer, D. (2019). Efficient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization. *Scientific Reports*. [Online]. 9 (1). p.pp. 2–10. Available from: <https://doi.org/10.1038/s41598-019-42557-4>.
- Pham, H.H., Le, T.T., Tran, D.Q., Ngo, D.T. & Nguyen, H.Q. (2021). Interpreting chest X-rays via CNNs that exploit hierarchical disease dependencies and uncertainty labels. *Neurocomputing*. [Online]. 437. p.pp. 186–194. Available from: <https://arxiv.org/pdf/1911.06475.pdf>.
- Rajaraman, S., Kim, I. & Antani, S.K. (2020). Detection and visualization of abnormality in chest radiographs using modality-specific convolutional neural network ensembles. *PeerJ*. [Online]. 8. p.p. e8693. Available from: <https://peerj.com/articles/8693/>.
- Rajpurkar, P., Irvin, J., Bagul, A., Ding, D., Duan, T., Mehta, H., Yang, B., Zhu, K., Laird, D., Ball, R.L., Langlotz, C., Shpanskaya, K., Lungren, M.P. & Ng, A.Y. (2017a). MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs. *arXiv preprint arXiv:1712.06957*. [Online]. p.pp. 1–10. Available from: <http://arxiv.org/abs/1712.06957>.
- Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., Ding, D., Bagul, A., Ball, R.L., Langlotz, C., Shpanskaya, K., Lungren, M.P. & Ng, A.Y. (2017b). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning. *arXiv preprint arXiv:1711.05225*. [Online]. Available from: <https://arxiv.org/pdf/1711.05225.pdf>.
- Razzak, M.I., Naz, S. & Zaib, A. (2018). Deep learning for medical image processing: Overview, challenges and the future. *Lecture Notes in Computational Vision and Biomechanics*. [Online]. 26. p.pp. 323–350. Available from: https://doi.org/10.1007/978-3-319-65981-7_12.
- Rosman^{1*}, D., Nshizirungu², J.J., Emmanuel Rudakemwa², Crispin Moshi³, Jean de Dieu Tuyisenge³, E.U. & Louise Kalisa (2015). Imaging in the Land of 1000 Hills: Rwanda Radiology Country Report. *Journal of Global Radiology*. [Online]. 1 (1). p.pp. 1–7.

Available from: <https://doi.org/10.7191/jgr.2015.1004>.

Ruder, S. (2017). An Overview of Multi-Task Learning in Deep Neural Networks. *arXiv preprint arXiv:1706.05098*. [Online]. Available from: <https://arxiv.org/pdf/1706.05098.pdf>.

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

Suzuki, K. (2017). Overview of deep learning in medical imaging. *Radiological Physics and Technology*. [Online]. 10 (3) p.pp. 257–273. Available from: <https://link.springer.com/article/10.1007/s12194-017-0406-5>. [Accessed: 31 March 2021].

Tang, Y., Tang, Y., Han, M. & Summers, R.M. (2019). Abnormal Chest X-Ray Identification With Generative Adversarial One-Class Classifier Imaging Biomarkers and Computer-Aided Diagnosis Lab , Radiology and Imaging Sciences , National Institutes of Health (NIH) Clinical Center , Bethesda , USA ; Ping An Te. *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. [Online]. (Isbi). p.pp. 1358–1361. Available from: <https://doi.org/10.1109/ISBI.2019.8759442>.

Tanzi, L., Vezzetti, E., Moreno, R. & Moos, S. (2020). X-Ray bone fracture classification using deep learning: A baseline for designing a reliable approach. *Applied Sciences (Switzerland)*. [Online]. 10 (4). Available from: https://share.osteoporosis.foundation/EU-6-Material/Reports/IOF Report_EU.pdf.

Tataru, C., Yi, D., Shenoyas, A. & Ma, A. (2017). *Deep Learning for abnormality detection in Chest X-Ray images*. [Online]. Available from: <http://cs231n.stanford.edu/reports/2017/pdfs/527.pdf>.

The Royal College of Radiologists (2016). Clinical radiology UK workforce census 2016 report [BFCR(17)6]. *Www.Rcr.Ac.Uk*. [Online]. (October). p.pp. 1–54. Available from: https://www.rcr.ac.uk/system/files/publication/field_publication_files/cr_workforce_census_2016_report_0.pdf.

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M. & Summers, R.M. (2017). ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. [Online]. 2017-Janua. p.pp. 3462–

3471. Available from: https://openaccess.thecvf.com/content_cvpr_2017/html/Wang_ChestX-ray8_Hospital-Scale_Chest_CVPR_2017_paper.html.
- World Health Organization (2016). *To X-ray or not to X-ray?* [Online]. 2016. World Health Organization-NewsLetter. Available from: <https://www.who.int/news-room/feature-stories/detail/to-x-ray-or-not-to-x-ray->. [Accessed: 17 June 2021].
- Yadav, O., Passi, K. & Jain, C.K. (2019). Using Deep Learning to Classify X-ray Images of Potential Tuberculosis Patients. *Proceedings - 2018 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018*. [Online]. p.pp. 2368–2375. Available from: <https://doi.org/10.1109/BIBM.2018.8621525>.
- Yao, L., Poblenz, E., Dagunts, D., Covington, B., Bernard, D. & Lyman, K. (2017). Learning to diagnose from scratch by exploiting dependencies among labels. *arXiv preprint arXiv:1710.10501*. [Online]. p.pp. 1–12. Available from: <https://arxiv.org/pdf/1710.10501.pdf>.
- Zech, J.R., Badgeley, M.A., Liu, M., Costa, A.B., Titano, J.J. & Oermann, E.K. (2018). Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study A. Sheikh (ed.). *PLOS Medicine*. [Online]. 15 (11). p.p. e1002683. Available from: <https://dx.plos.org/10.1371/journal.pmed.1002683>. [Accessed: 3 April 2021].
- Zhang, Q., Bai, C., Liu, Z., Yang, L.T., Yu, H., Zhao, J. & Yuan, H. (2020). A GPU-based residual network for medical image classification in smart medicine. *Information Sciences*. [Online]. 536. p.pp. 91–100. Available from: <https://doi.org/10.1016/j.ins.2020.05.013>.

APPENDIX B: PSEUDO CODE

Pseudo code B.1 indicates information for processing of pathology information to abnormal label, where records with no pathology observations marked as positive or uncertain are marked as 0 indicating normal X-ray record otherwise it is marked as 1 indicating abnormal X-ray record.

```
PRECONDITION: main_df := Load CheXPert dataset
FOR every row in main_df
    IF any of pathologies in the row is marked as 1
        THEN set Abnormal column for the row as 1
    ELSE IF any of pathologies in the row is marked as -1
        THEN set Abnormal column for the row as -1
    ELSE
        THEN set Abnormal column for the row as 0
    ENDIF
```

Pseudo Code B.1: Pseudo code for processing of pathology information to abnormal label

Pseudo code B.2 indicates information for processing of uncertain classes in Abnormal label to either 1 (abnormal class) or 0 (normal class), based on chosen dataset processing approach.

```
For every row in main_df
    IF Abnormal column for the row is -1
        IF we are mapping uncertain to abnormal class
            THEN mark Abnormal column for the row as 1
        ELSE IF we are mapping uncertain to normal class
            THEN mark Abnormal column for the row as 0
        ELSE
            THEN drop the row
```

Pseudo Code B.2: Pseudo code for pre-processing of uncertain class labels

Pseudo code B.3 indicates approach for splitting patient data into train and validation subsets based on patient ID. This helps in ensuring there is no overlap of patient information across train and validation subsets.

```
main_df['PatientID'] := Extract Patient ID from Path column  
Group-by-PID := Group main_df based on patient ID  
Train-PID , Validation-PID := Split “Group-by-PID“ into train and validation datasets  
Train-dataset := Extract records from main_df using PID’s in Train-PID  
Validation-dataset := Extract records from main_df using PID’s in Validation-PID
```

Pseudo Code B.3: Pseudo code for splitting the patient data into train and validation datasets

Pseudo code B.4 indicates model development approach used for transfer learning through DenseNet 169-layer model, where initial layers till 350 are frozen from training and last layers are considered for training on abnormality classification task.

```
def GetModel():  
    DenseNetModel := Load DenseNet 169-Layer model  
    FOR every layer in DenseNetModel  
        IF layer number is greater than 350  
            THEN set layer as trainable  
        ELSE  
            THEN layer is not set as trainable  
        ENDIF  
    ENDFOR  
    DenseNetModel := DenseNetModel + Global Average Pooling Layer  
    DenseNetModel := DenseNetModel +Dense(activation := Sigmoid)  
    Model := Create TensorFlow model using DenseNetModel  
    Return Model  
  
Model := GetModel()  
Compile Model using “Weighted Binary Cross-entropy” loss and Adam Optimiser
```

Pseudo Code B.4: Pseudo code extract of DL model used for abnormal X-ray classification

APPENDIX C: PROJECT PLAN

Activity	PLAN START	End Duration	ACTUAL START	Actual Duration	PERCENT COMPLETE	29-Jun-21	10-Jul-21	17-Jul-21	25-Jul-21	10-Aug-21	15-Aug-21	16-Aug-21	17-Aug-21
Research Proposal work	Literature review and Background work	11-Apr-21	5	11-Apr-21	5	100%							
	Aim and Objectives	14-Apr-21	1	14-Apr-21	1	100%							
	Data loading and description analysis	12-Apr-21	4	12-Apr-21	4	100%							
	Methodology for Research Proposal	15-Apr-21	4	15-Apr-21	4	100%							
	Significance and Scope of Study	17-Apr-21	1	17-Apr-21	1	100%							
	Abstract and other Research Proposal updates	17-Apr-21	5	17-Apr-21	5	100%							
	Research Proposal finalization and submission	21-Apr-21	1	21-Apr-21	1	100%							
Mid Thesis Work	Review work for Mid thesis submission	23-Apr-21	30	23-Apr-21	30	100%							
	DataSet Preprocessing and analysis	23-Apr-21	4	23-Apr-21	4	100%							
	Generator for Loading data	27-Apr-21	4	27-Apr-21	4	100%							
	Analysis using DenseNet: Ignore Uncertain class	1-May-21	13	1-Jun-21	30	100%	■						
	Analysis using DenseNet: Uncertain as Ones	14-May-21	2	30-Jun-21	2	100%							
	Analysis using DenseNet: Uncertain class as zeros	16-May-21	5	4-Jul-21	5	100%							
	Evaluation of results	21-May-21	4	8-Jul-21	4	100%	■						
	Mid Thesis Report Preparation	25-May-21	20	1-Jun-21	20	100%							
	Mid Thesis Report update and submission	13-Jun-21	2	17-Jun-21	2	100%							

Implemen tation Phase	Analysis using MobileNet: Uncertain as Zeros	16-Jun-21	13	13-Jul-21	2	100%		
	Analysis using MobileNet model: Uncertain as Ones	29-Jun-21	2	15-Jul-21	2	100%		
	Analysis using MobileNet model: Uncertain as separate class	1-Jul-21	3	17-Jul-21	2	100%		
	Evaluation of results	4-Jul-21	3	19-Jul-21	1	100%		
Final Proposal	Evaluation of advanced multitask models on different datasets	7-Jul-21	10	20-Jul-21	15	100%		
	Improve age predicton task and multipathology multitask analysis	17-Jul-21	2	3-Aug-21	10	100%		
	Final Thesis Report Preparation work	19-Jul-21	20	3-Aug-21	18	100%		
	Video Presentation	7-Aug-21	5	21-Aug-21	1	100%		

