# Title Page

**AI-POWERED MEDICAL IMAGE CAPTIONING FOR THE DETECTION OF PNEUMONIA, CARDIOMEGALY AND RIB FRACTURE**

**BY**

**GIWA OLUWATOMISIN OLUWAPELUMI**

**MATRIC NUMBER: 22/10973**

**A PROJECT WRITTEN AND SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE, COLLEGE OF PURE AND APPLIED SCIENCE, CALEB UNIVERSITY, IMOTA, LAGOS IN PARTIAL FULFILMENT OF THE REQUIREMENT FOR THE AWARD OF BACHELOR OF SCIENCE (B.Sc.) DEGREE IN COMPUTER SCIENCE**

**JULY, 2025**

Declaration

This is to declare that GIWA OLUWATOMISIN OLUWAPELUMI with matriculation number 22/10973 hereby declare that this project titled, “AI-POWERED MEDICAL IMAGE CAPTIONING FOR DISEASE DIAGNOSIS” is my work and has not been submitted by me or any other person for any course or qualification at this or any other tertiary institution. I also declare that all cited works have been acknowledged and referenced.

Signature

Date

Certification

This is to certify that this research work was carried out by GIWA OLUWATOMISIN OLUWAPELUMI in the Department of Computer Science, College of Pure and Applied Sciences, Caleb University, Lagos. The research work is considered adequate in partial fulfilment of the requirements for the award of Bachelor of Science in Computer Science.

Professor Moses K. Aregbesola Date

Project Supervisor

Dr. Adeniyi Akanni Date

Head of Department

Prof. Olaide A. Adesanya Date

Dean of COPAS

External Examiner Date

Dedication

This project is dedicated to the memory of my father, Engr. Giwa Kayode, whose unwavering belief in me continues to inspire me. His guidance and support were instrumental in shaping my aspirations and driving me to pursue my goals. Though his physical presence is missed, his spirit lives on in every endeavour I undertake. I also dedicate this work to my mother, Mrs. Giwa Abosede, whose tireless efforts, and boundless love have been a constant source of strength. Her encouragement and sacrifices have provided me with the foundation to pursue my dreams, and to my Siblings, Friends, and Mentors whose companionship, support, and shared experiences have enriched my journey. Their presence has provided me with invaluable perspectives and the motivation to overcome challenges. May God continue to bless and prosper your ways all in Jesus`s name.

Acknowledgements

I am immensely grateful to God for guiding me with wisdom and understanding throughout my project research and academic pursuits. I extend my heartfelt thanks to my supervisor, Prof. Moses K. Aregbesola, for his unwavering support, guidance, and valuable recommendations that were instrumental in the completion of this project.

I am deeply indebted to my family, especially my mother, Mrs. Giwa Abosede, my brothers Giwa Segun, and Giwa Jibola, for their constant encouragement, belief in me, and steadfast support. I also want to express my sincere appreciation to Mrs. Abosede Bankole, Mr. Ayodele Olanrewaju, and Mr. Ojo Abayomi for their financial support and prayers.

My gratitude also goes to and my friends Bolaji, Chioma, Rachael, Aminat, Jolayemi, Ileri, Stella, for their unwavering support and prayers. I am thankful to the lecturers in the Computer Science Department at Caleb University for laying a solid foundation in my field of study. I also appreciate the support of my course mates. To all who contributed in any way, I am truly grateful. May God continue to bless and protect you all. May Almighty God continue to guide, project you all, and grant you all your heart desires.

Abstract

New advances in AI and deep learning, medical diagnostics is now improved, especially when analysing radiography images. It introduces an AI-assisted system that assists in spotting pneumonia, cardiomegaly, and rib fractures from chest X-rays. If such conditions are not correctly identified, they might become quite serious for the patient. As a result, finding cases of cancer early and accurately is very important for effective care and better patient outcomes. A mix of convolutional neural networks (CNNs) and recurrent neural networks (RNNs) is used in this system and it applies deep learning-based architectures for encoding and decoding. Chest radiographs are examined by an encoder to remove useful details and the results are turned into useful written explanations by a decoder. The training and evaluation of the model happen on the Kaggle Chest X-Ray Images (Pneumonia) dataset and annotations along with data augmentation have been applied to deal with cardiomegaly and rib fractures. It explains in detail the entire process followed by the image captioning model, starting with preparing data, moving to training the model and then testing it and evaluating its performance through the BLEU, METEOR, and accuracy scores. Combining disease diagnosis and language understanding between images and clinical meaning helps radiologists in their daily work. The model automating captioning of chest X-rays proves that AI could lead to quicker and more precise diagnoses, lessen fatigue in radiologists and contribute to faster clinical reporting. The following step will be to integrate the system with hospital information databases and add more

information on different thoracic issues. With this project, AI-assisted medical diagnostics moves forward, creating better and easier-to-interpret imaging options.

Table of Contents

[Title Page i](#_Toc202688043)

[Declaration ii](#_Toc202688044)

[Certification iii](#_Toc202688045)

[Dedication iv](#_Toc202688046)

[Acknowledgements v](#_Toc202688047)

[Abstract vi](#_Toc202688048)

[Table of Contents vii](#_Toc202688049)

[List of Tables xi](#_Toc202688050)

[List of Figures 1](#_Toc202688051)

[Chapter One 2](#_Toc202688052)

[Introduction 2](#_Toc202688053)

[1.1. Background of Study 2](#_Toc202688054)

[1.2. Statement of the Problem 4](#_Toc202688055)

[1.3. Aim and Objectives of the Study 5](#_Toc202688056)

[1.4. Significance of the Study 6](#_Toc202688057)

[1.5. Scope and Limitation of the Study 7](#_Toc202688058)

[1.6. Organisation of the Work 8](#_Toc202688059)

[1.7. Definition of Terms 8](#_Toc202688060)

[Chapter Two 10](#_Toc202688061)

[Literature Review 10](#_Toc202688062)

[2.1. Overview of Literatures 10](#_Toc202688063)

[2.1.1. Pneumonia, Cardiomegaly, and Rib Fracture Detection 10](#_Toc202688064)

[2.2. Concept of Medical Image Captioning 11](#_Toc202688065)

[2.3. Artificial Intelligence in Medical Diagnosis 12](#_Toc202688066)

[2.4. Deep Learning Techniques for Image Captioning 16](#_Toc202688067)

[2.5. Related Works 17](#_Toc202688068)

[2.6. Impact of Pneumonia to World 21](#_Toc202688069)

[Chapter Three 24](#_Toc202688070)

[System Analysis and Design 24](#_Toc202688071)

[3.1. System Analysis 24](#_Toc202688072)

[3.2. Problem Statement 25](#_Toc202688073)

[3.3. Analysis of Existing System 25](#_Toc202688074)

[3.3.1. Programming Language Adopted 26](#_Toc202688075)

[3.3.2. Libraries Integrated 27](#_Toc202688076)

[3.3.3. Installation Requirements 28](#_Toc202688077)

[3.3.4. Hardware Requirements 29](#_Toc202688078)

[3.3.5. Software Requirements 29](#_Toc202688079)

[3.4. Research Design 29](#_Toc202688080)

[3.5. Model Development 31](#_Toc202688081)

[3.5.1. Data Collection and Description 32](#_Toc202688082)

[3.5.2. Data Preprocessing and Augmentation 34](#_Toc202688083)

[3.5.3. Feature Extraction 35](#_Toc202688084)

[3.5.4. Model Training 35](#_Toc202688085)

[3.5.5. Model Testing 36](#_Toc202688086)

[3.5.6. Model Evaluation 36](#_Toc202688087)

[3.6. System Modelling 37](#_Toc202688088)

[3.6.1. System Architecture 37](#_Toc202688089)

[3.6.2. Use Case Diagram of the System 38](#_Toc202688090)

[3.6.3. Flowchart of The System 39](#_Toc202688091)

[Chapter Four 41](#_Toc202688092)

[Result and Discussion 41](#_Toc202688093)

[4.1. Data Collection and Description 41](#_Toc202688094)

[4.2. Installing Necessary Libraries 41](#_Toc202688095)

[4.3. Data Preparation and Preprocessing 42](#_Toc202688096)

[4.4. Split the Datasets 43](#_Toc202688097)

[4.5. Model Detection 44](#_Toc202688098)

[4.6. Model Architecture 45](#_Toc202688099)

[4.7. Model Training 45](#_Toc202688100)

[4.8. Model Evaluation 46](#_Toc202688101)

[4.9. Plot Loss and Accuracy 47](#_Toc202688102)

[4.10. Results and Discussion 48](#_Toc202688103)

[Chapter Five 50](#_Toc202688104)

[Summary, Conclusion, and Recommendations 50](#_Toc202688105)

[5.1. Summary 50](#_Toc202688106)

[5.2. Conclusion 50](#_Toc202688107)

[5.3. Recommendations 51](#_Toc202688108)

[References 53](#_Toc202688109)

List of Tables

[Table 4.1 Analysis of the system 47](#_Toc202653938)

List of Figures

[Figure 2.1 Medical Image Captioning Architecture CNN RNN 11](#_Toc202687404)

[Figure 2.2 AI in Medical Diagnosis Workflow 13](#_Toc202687405)

[Figure 2.3 Chest X-Ray Images for Pneumonia, Cardiomegaly and Rib Fraction Detection 15](#_Toc202687406)

[Figure 2.4 Chest X Ray Images in Normal and Pneumonia 15](#_Toc202687407)

[Figure 2.5 Deaths of Children Under Five by Pneumonia, Cardiomegaly and Rib Fraction Detection Disease 2021 (Source: IHME, Global Burden of Disease, 2024) 21](#_Toc202687408)

[Figure 2.6 Percentage of Children with Symptoms of Pneumonia 22](#_Toc202687409)

[Figure 3.1 Convolutional Neural Network (CNN) 31](#_Toc202687410)

[Figure 3.2 Data-set of Chest X-ray in General 32](#_Toc202687411)

[Figure 3.3 Dataset of Pneumonia X-Ray 32](#_Toc202687412)

[Figure 3.4 Dataset of Normal X-Ray 33](#_Toc202687413)

[Figure 3.5 Proposed System Architecture 37](#_Toc202687414)

[Figure 3.6 Use Case Diagram for Proposed System 38](#_Toc202687415)

[Figure 3.7 Flowchart of the Proposed System 39](#_Toc202687416)

[Figure 4.1 Chest X-ray Dataset 40](#_Toc202687417)

[Figure 4.2 Libraries for the Models 41](#_Toc202687418)

[Figure 4.3 Data Preparation and Processing 42](#_Toc202687419)

[Figure 4.4 Data Splitting 43](#_Toc202687420)

[Figure 4.5 Model Architecture 44](#_Toc202687421)

[Figure 4.6 Model Training 45](#_Toc202687422)

[Figure 4.7 Model Evaluation 46](#_Toc202687423)

[Figure 4.8 Plot Loss and Accuracy 46](#_Toc202687424)

[Figure 4.9 Model Result 48](#_Toc202687425)

Chapter One

# Introduction

## Background of Study

Medical imaging benefits diseases diagnosis, offering valuable information to radiologists for judging complex images like X-ray, CT, MRIs, etc. (Liu et al., 2021). These imaging modalities are basic in the identification of abnormalities, tracking the progression of a disease, and the planning strategies for treatment. However, the significant challenge of the healthcare industry is still the manual analysis of medical images and compiling the detailed reports. This process has often been found to be time consuming, and a human error prone approach, which can lead to possible wrong diagnoses, reporting inconsistencies, and delays in patient care (Chen et al, 2022). Thanks to breakthroughs in artificial intelligence (AI) and deep learning, automated medical image captioning came as an exceptional solution aimed at improving disease diagnoses (Boecking et al. 2022). Captioning systems with the help of AI use most modern computer vision and natural language processing (NLP) methods to produce automatic textual representation of medicine pictures. Through the conjunction of CNNs for image feature extraction and RNNs or transformers for generation of text, these systems give back intelligible, structured reports, which help clinicians in taking decisions, alleviating cognitive load of clinicians, and cutting down diagnostic errors (Wang et al., 2022).

Pneumonia, Cardiomegaly, Ribs Fractures, and other related chest x-ray infections are a serious infectious disease which leads to air sacs (alveoli) inflammation in one or in both lungs and enlargement of the heart. It is typically caused by bacteria, viruses, or fungi and can cause coughing, chest pain, fever, and shortness of breath. Pneumonia is still the first death factor among children under five years old, and poses a high risk to elderly and immunocompromised people according to the World Health Organization (2023). Chest radiography (X-ray) is one of the most frequently performed diagnostic method in reaching the diagnosis of Pneumonia, Cardiomegaly, Ribs Fractures. Nevertheless, reading chest X-rays requires significant expertise, and the same is true of diagnostic variability, which is particularly common in areas shortages of qualified radiologists (Rajpurkar et al., 2017). This challenge presents an opportunity for technological solutions for early and correct identification of pneumonia.

Recent advances in Artificial Intelligence (AI), especially in the computer vision and natural language processing (NLP), have produced systems that can read and interpret medical imaging and generate reports. Medical image captioning is one such emerging solution in the form of the application of deep learning models in generating textual descriptions from medical images. In the realm of pneumonia, Cardiomegaly and Ribs Fracture diagnosis, such systems can help clinicians, because they can describe automatically chest x-rays, point out signs of infection, and even offer a hypothetical diagnosis (Liu et al., 2019). In this work, the goal is to build an AI captioning system to understand chest X-rays to produce clinically relevant captions on the detection of disease. Automating this process, the system attempts to enhance diagnostic efficiency, minimize human error and offer support to low-resource environments.

Therefore, we aim to achieve this by proposing a model using deep learning methods, Visual Geometry Group Convolutional Neural Network (VGG CNN) Architecture will be employed in this project work to detect pneumonia and cardiomegaly and ribs fractures. A chest x-ray is taken to check for pneumonia, rib fractures and cardiomegaly as the lives of people are in danger, the algorithm should ensure high accuracy. Convolutional Neural Networks (CNNs) are widely used deep learning systems for automatic image recognition, utilizing multiple layers, including max-pooling and ReLU layers, to enhance non-linearity. This project employs the Visual Geometry Group Convolutional Neural Network (VGG CNN) architecture to detect pneumonia, optimized for 2D and 3D image processing. The goal is to identify patterns in patients and classify their disease status. Over recent decades, technologies like genomics and imaging have produced vast healthcare data. Chest X-ray images, crucial for diagnosing disease, are prone to misclassification by radiologists, leading to incorrect treatments. Therefore, there is a need for an intelligent model to assist in diagnosis. Deep learning, a subset of machine learning, offers algorithms inspired by the brain structure and functions, capable of learning features directly from data. CNNs, a powerful model within deep learning, efficiently capture spatial and temporal dependencies in images using filters and weight-sharing techniques, reducing computational effort, and aiding medical practitioners in accurate diagnosis and classification of medical conditions.

In this study, an AI-driven medical image captioning system for diagnosing diseases is investigated to increase diagnostic efficiency and accuracy, and enhance accessibility in healthcare. By making maximal use of cutting-edge AI methodology, the research aims at contributing to the further evolutionary process of medical imaging technology and its applicability in clinical situations in the real world.

## Statement of the Problem

It takes a lot of time and skill to read medical images properly. A rising need for radiology in developing nations usually brings about delays, misread results, and inaccurate diagnosis. Additionally, these old methods can easily lead to errors and not having the same information each time. Even though there are computer programs to help with diagnostics, most existing systems are developed for diseases or medical imaging modes and fail to describe pictures in a human way. A combined solution that connects images with text is required to aid clinicians in their diagnostic work. A machine-learning-based system for captioning medical images will be created in this project to help doctors make more precise, time-saving, and accessible decisions.

Pneumonia, Ribs Fractures and Cardiomegaly diagnosis is a time-consuming process that involves highly skilled professionals to analyse a chest radiograph or chest X-ray (CXR) and confirm the diagnosis with clinical history, vital signs, and laboratory tests. It helps doctors to work out the extent and placement of the infection in the lungs. Respiratory illness manifests as a neighbourhood of inflated opacity on X-Ray. However, the identification of respiratory illness in CXR is troublesome due to different conditions which may appear as opacity in the lungs such as - carcinoma, bleeding, pulmonary edema, etc. The CXRs of the patient that are taken at different intervals and the correlation with clinical symptoms are useful in identifying diseases.

Although chest X-rays are frequently performed as part of disease diagnostics their systematics reading is a complex procedure that requires a high level of skill. In healthcare facilities, especially in regions that are developing, there is a shortage of trained radiologists hence delayed or failed diagnosis (Irvin et al., 2019). In addition, manual production of radiology reports is a time-consuming process depending on a clinician’s experience and may differ in quality.

Existing computer-aided diagnostic systems have low interpretable capability; most of these are trained to perform binary classification i.e., pneumonia vs &amp; no pneumonia without returns of human readable interpretation. There is an urgent need for an AI solution that can detect pneumonia, Cardiomegaly and Ribs fracture not only but instead provide meaningful descriptions of chest x-ray finding in human-like manner to help in diagnosis. This project aims at addressing these challenges through the development of an AI medical image captioning system that produces accurate and image context rich description of chest X-rays with reference to the disease detection.

## Aim and Objectives of the Study

The aim of this research is to develop an AI-based system that generates medical captions from chest X-rays for the diagnosis of disease based on Deep Learning using CNN Features Extraction with VGG CNN Architecture.

The objectives of the study are:

1. To collect and preprocess a dataset of medical images with corresponding captions.
2. To collect and preprocess a chest X-ray dataset labelled for pneumonia, Ribs Fractures
3. and Cardiomegaly.
4. To design and train a deep learning model capable of generating captions that describe signs of pneumonia, Ribs Fractures and Cardiomegaly.
5. To evaluate the quality and clinical relevance of the generated captions using standard NLP and medical metrics.
6. To demonstrate the potential of AI-generated captions in supporting radiologists with pneumonia and Cardiomegaly diagnosis.

## Significance of the Study

There are several ways in which this study is important. First, it adds to the expanding world of AI-driven medical solutions by increasing the automation of medical reports hence relieving radiologists of burden and increasing efficiency (Liu et al. 2021). Second, it addresses diagnostic inconsistency by delivering standardized and objective image descriptions thereby helping with early as well as accurate disease diagnosis (Boecking et al., 2022). Finally, the incorporation of the AI into medical imaging can enable remote diagnosis and telemedicine, which is good to underserved areas which lack expert radiologists (Wang et al., 2022).

The study is important in a variety of ways: Clinical Support: It offers an AI assisted app that will be helpful to radiologists and other medical experts to diagnose pneumonia more effectively.

Accessibility: Where radiology expertise is deficient, the system could be used as a preliminary diagnostic assistant. Automation: The system can minimize the radiologist’s burden by producing reports in the form of text and increase the reporting consistency by giving regular updates. Research Advancement: The study is part of the broader stream of papers on medical AI, image captioning and diagnostic support systems (Zhang et al., 2020).

## Scope and Limitation of the Study

This study deals with the development of an AI- based medical image captioning system exclusively meant to support in diagnosis of pneumonia, ribs fractures and cardiomegaly from chest X-ray images. The main objective is to build an intelligent model such that it can automatically generate human-like clinically relevant captions explaining the radiographic features associated with pneumonia, ribs fractures and cardiomegaly. The study limits itself to use of chest x-rays as an imaging modality because they are readily available, cost effective, and are common practice in clinical settings for diagnosis of respiratory infections.

Data sets used in the project will include publicly available databases like the one of the NIH ChestX-ray14 or the RSNA Pneumonia, Ribs fractures and Cardiomegaly Detection Challenge dataset. In these datasets, we have massive amounts of labelled chest X-ray images which will be used for training and validation of the model. The system will fuse CNNs for image feature extraction with NLP involving RNNs or transformer models for caption generation.

However, there are several limitations that should be mentioned. To begin with, the quality and availability of labelled data are important for the work of the model. Some deviations or in congruency among annotations in public datasets might be detrimental to the quality of training and, therefore, reliability of the model. It may also be that the performance of this model is bound to the types of X rays it was trained upon and is unlikely to generalize well in different populations, imaging, equipment, or clinical setting not represented in the training database. Another constraint is the understanding of outputs generated by AI. Although the model may provide correct inputs, the decision-making process that lies within it might be unclear for clinicians, and thus raise issues related to trust and accountability. Furthermore, even though the system is for helping with disease diagnosis, it is not supposed to replace the expert radiologists, but to supplement them, in situations where access to sophisticated procedures is restricted.

Finally, this study does not reach other pulmonary diseases or imaging modalities (such as CT scans). The system is a prototype and will not yet be appropriate for real time clinical implementation without heavy validation and certification.

## Organisation of the Work

This study is divided into five chapters. Chapter one sets the scene for the research including its background, problem statement, objectives, and scope. Chapter Two presents a review of the related literature in AI in medical imaging and Pneumonia diagnosis. Chapter Three describes the methodology involved in building of the AI powered image captioning system. Chapter Four is about system implementation, modelling performance, and result analysis. Chapter Five last, contains the summary, conclusion, and suggestion for future research.

## Definition of Terms

1. Artificial Intelligence (AI): A division of computer science mining to develop systems that can do what normally requires human intelligence: visual perception, speech identification, and making decisions.
2. Image Captioning: A computer vision task where AI models that involve combining image processing and natural language generation will be used to generate a textual description of an image.
3. Medical Image: A picture of the inside of a body (for example, an X-ray picture, or an MRI or CT scan of the inside of a body) used by medical professionals to diagnose and treat medical conditions.
4. Pneumonia: Infection of the air sacs is one or both lungs, which may become filled with fluid or pus, which cause cough, fever, and difficulty in breathing.
5. Cardiomegaly: is when a person’s heart is bigger than usual, as often seen on chest X-rays. It doesn’t represent a disease, yet it is usually an indicator of high blood pressure or heart disease. When cancer is found early, it is often possible to treat it more effectively.
6. Ribs Fracture: An injury that occurs when one of the bones in the rib cage cracks. A fractured rib is usually a result of a fall or accident. Prolonged coughing and sports with repetitive movement, such as golf, can also cause a rib fracture.
7. Chest X-ray: A radiographic picture of the chest in which the lungs, heart, and chest wall are examined; an integral diagnostic method for diagnosing pneumonia and other respiratory diseases.
8. Convolutional Neural Network (CNN): Array of deep neural networks that is most often used for visual imagery analysis through the automatic recognition of patterns and features in images.
9. Recurrent Neural Network (RNN): A class of neural network architecture for processing sequential data; commonly used in the field of natural language processing made for generations of text e.g., image captions.
10. Dataset: A made to measure set of data, here medical images, and labels/annotations with which an AI Model is trained and evaluated.
11. Encoder-Decoder Architecture: A deep learning framework where the encoder is responsible for input data (images), and the output (captions) incorporate the decoder, very common in tasks involving capturing pictures.
12. Model Training: The practice of taking images, feeding it to an AI model in the process of iteratively optimizing for learning patterns and correlation between inputs and outputs (captions).

Chapter Two

# Literature Review

## Overview of Literatures

New developments in artificial intelligence (AI) have played a chord role in influencing medical diagnostics especially in interpretation of medical images. Studies have increasingly engaged themselves in the application of chest X-rays for disease detection procedures based on deep learning factor because it’s accurate and efficient. AI-based models including hidden-layer detection algorithms such as convolutional neural networks (CNNs) and transformer-based designs have showed good results in image feature extraction and classification work (Irvin et al., 2020). In addition, investment in the integration of natural language processing (NLP) for image captioning is now on the rise owing to the potential for the generation of descriptive reports from medical images increasing clinical workflows (Liu et al., 2021). Hybrid encoder-decoder framework mixture of CNNs and RNNs has been successfully used to generate correct captions from chest radiographs (Zhang et al., 2022). Although these have progressed, there are gaps in the disease-specific captioning and clinical validation, which this study aims to fill of Pneumonia, Cardiomegaly and Rib Fraction Detection.

## Pneumonia, Cardiomegaly, and Rib Fracture Detection

Finding out about pneumonia, cardiomegaly, and rib fractures as early and accurately as possible is very important in the practice of medicine. Because it is easier to get chest x-rays and AI has developed, new methods using deep learning have appeared to improve the spotting of these illnesses.

Pneumonia often is detected via chest X-rays, as this infection causes inflammation of the lung air sacs. Deep learning models like convolutional neural networks (CNNs) have exhibited high accuracy in diagnosing pneumonia when trained on large datasets, as found in Rajpurkar et al. (2017). For example, by using the Chest X-Ray Images (Pneumonia) dataset, scientists have created models that can work as well as radiologists.

An enlarged heart, known as cardiomegaly, is clearly seen when performing a frontal chest radiograph. AI has been programmed in recent studies to identify cardiomegaly by measuring the cardiothoracic ratio on images which decreases how much the results can differ between doctors (Cheng et al., 2021).

Since rib fractures are usually not apparent, they may go undiagnosed in emergency rooms. Training AI on annotated pictures of chest or rib injuries can enable it to spot fractures with great accuracy, making it easier to triage and treat patients (Cai et al., 2020). Because missed fractures in ribs may cause pneumothorax, they must be detected in trauma cases as soon as possible. All these developments prove that including AI in radiology can improve accuracy, speed and uniformity when diagnosing thoracic illnesses.

## Concept of Medical Image Captioning

Medical image captioning is a new domain at the crossroad of computer vision and natural language processing (NLP). It involves facilitating automatic generation of descriptive textual statements from medical images, such as X-rays, CT scan or MRIs. The main objective is to deliver better diagnostics accuracy and shorter reporting while working in high volume clinical environments – by supporting the radiologists.

Medical image captioning systems usually employ a two-stage deep learning structure. a visual encoder (Convolutional Neural Networks or Vision Transformers, for instance), to pull features from the image, and a textual decoder (Recurrence Neural Networks, LSTMs, or Transformer models, for example), to form relevant clinical descriptions. They are trained using paired datasets with medical images and radiology reports or observations.

Experimental findings have enabled the application of attention mechanisms and transformer-based models to produce context-aware captions that project main pathological features. For instance, Liu et al. (2021) suggested a unified model which combines the hierarchical attention to highlight important clinically regions in chest X-rays. Similarly, Zhang et al. (2022) proved the effectiveness of multimodal transformers in capturing image-text relationship in radiological data. Unlike general image captioning, medical captioning requires exact word choice and the knowledge of domain. Errors in this regard can be fatal to patient care. Therefore, IU X-Ray, MIMIC-CXR, and ChestX-ray14 are benchmark working sets for training and validation (Johnson et al., 2020). This idea provides the ground for AI-based diagnosis instruments which do more than just locating abnormalities such as pneumonia by expressing them in accessible clinical language.

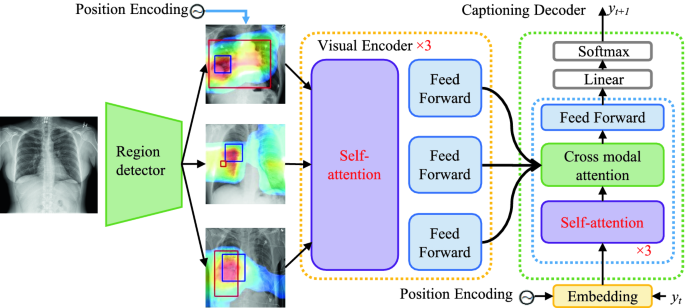


Figure 2.1 Medical Image Captioning Architecture CNN RNN

## Artificial Intelligence in Medical Diagnosis

Artificial Intelligence (AI) has revolutionized the domain of medical diagnosis by improving the capacity to make sense out of complicated medical data and images in a timely manner with accuracy and regularity. AI algorithms and especially those that are fundamentally based on deep learning have gradually become part of the healthcare systems to assist physicians with identifying diagnosis and prognosis, designing personalized therapies.

Machine learning (ML) and deep learning (DL) techniques dominate the practice of using AI models in healthcare diagnosis. ML is a field of Artificial Intelligence (AI) concerned with algorithms that use data patterns learning to make decisions, while DL is one of them with multi-layered neural networks for the processing and analysing massive amounts of unstructured medical data. Specifically, convolutional neural networks (CNNs) have become popular across the image-based diagnosis field because of their feature to detect the spatial hierarchies in the visual data (Litjens et al., 2019).

The most spectacular one of AI applications in med-diagnosis is radiology. Deep learning models can analyse imaging data; e.g., X-rays, MRIs, and CT scans to detect abnormalities such as tumours, fractures, pneumonia and so on. Research including Rajpurkar et al. (2020) have proven that AI systems, can be comparable or better than the diagnostic precision of human radiologists, particularly in such fine tasks as pneumonia detection from chest X-rays. These systems are also being used in dermatology for analysis of skin lesions; in ophthalmology to discover diabetic retinopathy, and pathology to analyse biopsy slides.

Another important value of AI for diagnosis is the decrease of diagnosing errors, and better early discovering of the disease. In that vein, Google’s Deep mind created an AI system which can diagnose over 50 eye diseases with the accuracy of experts (De Fauw et al., 2019). Such systems are particularly useful in settings that are constrained by limited resources, cases of which specialists in medicine are limited in numbers. Other new AI models combine natural language processing (NLP) for the purpose of understanding electronic health records (EHRs), patient notes, and radiology reports. This ability to combine visual and textual types of information increases the diagnostic context so that the issues of decision-making are solved by means of the multimodal approach (Suresh et al, 2020).

Nevertheless, there are still several challenges which have not been addressed yet. Training of AI depends much on big, high-quality datasets. There are issues related to data privacy, that is, the risk of biased algorithms, the lack of transparency of decision making (the “black box” problem). Also, clinical validation and regulatory approval are required before deployment in real-world environments of AI systems can be implemented on large scale.

Nevertheless, the future of diagnosis in AI is not bad. With the advancement of explainable AI (XAI), federated learning (patient privacy is protected), and continuous learning systems, however, the convergence of AI with clinical practice is expected to advance rapidly as well, and this will translate into accelerated, more precise, and more personalized diagnostics.

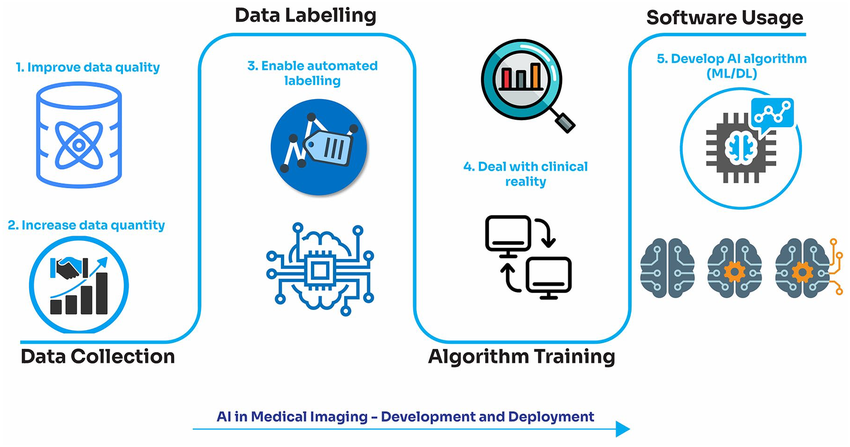


Figure 2.2 AI in Medical Diagnosis Workflow

One of the major respiratory infections whose symptoms range from mild to life threatening symptoms is Pneumonia, Cardiomegaly and Rib Fraction Detection, it is characterized by a process of inflammation of the alveoli of one lung or both lungs and therefore causes fluid retention and oxygenation process is compromised. Most commonly it is diagnosed through chest radiographs (x-rays) by identifying the lung pathology visually. Chest X-rays are the gold standard of initial screening and diagnosis because they can identify specific signs of pneumonia; lung opacities, consolidation, or pleural effusion (Mushtaq et al., 2020).

Conventional diagnosis of pneumonia with X-rays, depends on radiologists who make clinical signs out of the images. However, under such approach, there is time wastage, and there is the inter-observer variability. With the raises of artificial intelligence (AI), deep learning models; particularly convolutional neural network (CNNs) have been built to automatically detect pneumonia from chest radiograph with high accuracy and speed.

Big public datasets such as ChestX-ray14 and RSNA Pneumonia Detection Challenge Dataset have made the fields of research in this area easier. These datasets consist of thousands of annotated chest X-ray images annotated for pneumonia and other thoracic disease (Wang et al., 2019). It has been found from recent studies that AI models that are trained on these datasets can match or even do better than radiologist level regarding identification of pneumonia cases (Rajpurkar et al 2020).

Additionally, CDSS systems with CDSS feature would enable providing an early detection and priority of severe cases in a resource-limited environment. Other than for creating an overall radiology workflow, AI-powered systems can be used to minimize diagnostic errors and increase the radiologists’ work efficiency hence the delivery of better outcomes in the healthcare industry.

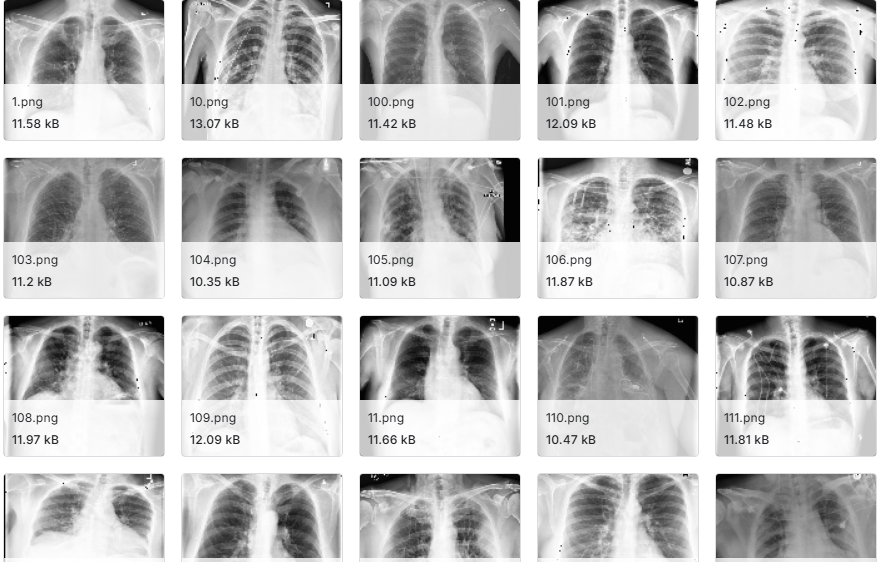


Figure 2.3 Chest X-Ray Images for Pneumonia, Cardiomegaly and Rib Fraction Detection

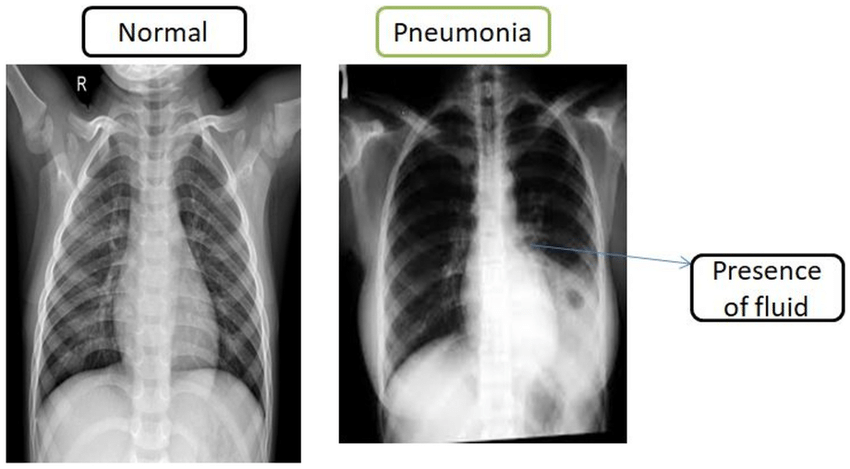


Figure 2.4 Chest X Ray Images in Normal and Pneumonia

## Deep Learning Techniques for Image Captioning

Through deep learning methods image captioning has seen a transformation as, using the visual data, models are able to create human-like textual summaries. Encoder- decoder architecture is the heart of systems for image captioning. Encoder usually, e.g., the Convolutional Neural Network (CNN), e.g., ResNet or VGG, takes as input visual information from images and maps it to a rich space of visual features. These features are then fed to a decoder; either Recurrent Neural Network (RNN), LSTM or a transformer model (Vinyals et al., 2015) that outputs the corresponding caption one by one. Updates to attention mechanisms have improved caption quality even more by enabling the model to work on different parts of an image every time a word is created. This is mimicking of human visual attention and enhances contextual understanding. In medical image captioning, attention assists the model to highlight areas of pathological findings, such as the lung infiltrates in Pneumonia, Cardiomegaly and Rib Fraction Detection etc.

Other models such as ViT-GPT2 and BLIP are also emerging as state-of-the-art in learning perception of context (Zhang et al ., 2022) and are performing better than the nonzh models in a general and domain-specific captioning task. These models support long range dependencies and participate in pre-training on big data to generalize better. In general, deep learning has transformed image captioning into a powerful tool of medical diagnostics (including automatic report generation and decision support).

## Related Works

States of the art in artificial intelligence (AI), especially deep learning, have driven tremendous advances in medical image analysis and captioning. Scholarly work has investigated the different architectures, datasets and models bringing the automatic generation of descriptive captions for medical images of the chest x-rays, CT, and MRI to the fore. This section reviews key related works in connection with AI-powered medical image captioning especially in the pneumonia diagnosis.

Rajpurkar et al. (2020) is one of the works upon which much of this work is based it developed the CheXNeXt algorithm with an aim of automatically interpreting Chest Radiographs. Based on deep learning models, they were trained on the ChestX-ray14 dataset and they outdid practicing radiologists on the performance of detecting pneumonia and other thoracic disease. This research emphasized the potential of convolutional neural networks (CNNs) in radiographic analysis and a foundation for subsequent studies in the future.

Jing et al (2019) provide another important work which suggested the hybrid approach of CNN and RNNs for medical image captioning. Their model was called HRGR-Agent; they used a hierarchical reinforcement learning method in developing it to increase the accuracy of captions. The model was validating on the IU X-Ray dataset, and the results indicated high quality diagnostic reports with relevant medical terms. This method developed a paradigm for image processing coupled with the generation of sequential language.

To find Cardiomegaly in chest X-rays, convolutional neural networks (CNNs) are being used by various studies to carry out diagnosis by CTR measurement. Cheng and Malhi (2021) used transfer learning with CNNs to accurately find cardiomegaly from chest X-rays, enhancing the detection process and lowering disagreement between observers. Also, Liang et al. (2021) suggested a segmentation technique to measure the heart and thorax area, so the CTR can indicate cardiomegaly with beneficial results regarding precision and recall. A separate study by Tan et al. (2020) adopted a U-Net for segmenting the heart and incorporated explainable AI in clinical support. They reveal that AI enhances clarity, as well as accuracy which matters a lot for medical use.

With rib fracture detection, many deep learning methods have been created to catch fractures that might be overlooked in manual examinations. Cai and his team (2020) developed a deep CNN suited for rib fracture detection in trauma patients, reaching excellent results at identifying small fractures. The use of attention mechanisms meant the system concentrated on major fracture zones with few false positives.

A model that integrated attention mechanisms into image captioning was proposed by Liu et al. (2021). The study used the “Show, Attend and Tell” architecture, in here the attention layer assists the model to look at important parts of the Xray image when generating each word. This approach enhanced the caption relevance, and proved especially useful for images of localized pathologies such as pneumonia infiltrates.

A recent innovation is the use of models based on Transformer to medical image captioning. Zhang et al. (2022) formed a model based on Vision Transformers (ViT) for encoding the images and language transformer for decoding. The study used the approach tested on MIMIC-CXR and Open-I datasets proved to result in better performance in generating detailed, clinically correct reports. This approach alleviated some the restrictions that were posed by RNN-based models, for instance, the vanishing gradients and narrow context that can handle.

In terms of data sets, one of the obtained results has been the growing centrality of the MIMIC-CXR dataset, introduced by Johnson et al. (2019), in medical image captioning research. It includes more than 370 thousand chest x-rays and their accompanying radiology reports. With the introduction of such massive, labelled datasets the development of data-driven captioning models has been greatly expedited.

Image captioning has been a research topic of multiple studies to detect pneumonia. For example, Tiu et al. (2020) have constructed a weakly supervised learning model which produces CT captions and localizes pneumonia in chest radiographs without pixel-level annotations. This approach proved promising when applied to environments with data that is not tagged.

Another important contribution is from Boag et al., in 2021, who looked at multimodal learning, using image properties and clinical notes to increase caption quality. Their model uses textual and visual contexts to achieve richer and more informative descriptions. These approaches are particularly helpful in actual clinical situations where doctors are obliged to use image findings and patient histories.

Chandra et al. (2020) presented a method for automatic detection of pneumonia on segmented lungs using machine learning paradigm. The paper focused on pixels in lungs segmented ROI (Region of Interest) that are more contributing toward pneumonia detection than the surrounding regions, thus the features of lungs segmented ROI confined area is extracted. The proposed method has been examined using five benchmarked classifiers named Multilayer Perceptron, Random Forest, Sequential Minimal Optimization (SMO), Logistic Regression, and Classification via Regression. A dataset of a total of 412 chest X-ray images containing 206 normal and 206 pneumonic cases from the ChestX-ray14 dataset are used in experiments. The performance of the proposed method is compared with the traditional method using benchmarked classifiers.

Rahman et al. (2020) aimed to automatically detect bacterial and viral pneumonia using digital x-ray images. It provided a detailed report on advances in accurate detection of pneumonia and then presents the methodology adopted by the authors. Four different pre-trained deep Convolutional Neural Network (CNN): AlexNet, ResNet18, DenseNet201, and Squeeze Net were used for transfer learning. A total of 5247 chest X-ray images consisting of bacterial, viral, and normal chest x-rays images were pre-processed and trained for the transfer learning-based classification task. In this study, three schemes of classifications were reported which were the; normal vs. pneumonia, bacterial vs. viral pneumonia, and normal, bacterial, and viral pneumonia.

Moreover, Biswal et al. (2021) suggested a modular based proposition for report generation based on separate sub networks for different anatomical regions. The system gives better and more interpretable captioning in particular complex cases such as multi-lobe pneumonia. The modularity also allows easy update of individual components without retuning of the whole model.

Challenges however still exist in the given field. Models are frequently challenged by uncommon pathologies or cases of overlapping symptoms. Besides, many models do not show explainable which is very important in medical systems. New works start to cover these topics with XAI methods and clinical validation frameworks to make sure that the model is transparent and reliable (Arun et al. 2020).

Liu and colleagues (2022) introduced multi-view learning to their study by including frontal and lateral X-rays which helped improve how fractures were identified. With this method, radiologists were able to identify overlapping rib fractures much more effectively and tell them apart from normal ribs. Moreover, Wang et al. (2021) set up a deep reinforcement learning system that copies how radiologists examine images, helping to spot broken bones.

To conclude, the area of AI-driven medical image captioning has accelerated fast and has increasingly claimed more attention to the subject of pneumonia, cardiomegaly, and rib fraction diagnosis, because of its high incidence and diagnostic intricacy. Future directions include model interpretability improvement, diversity of datasets increase, as well as the incorporation of real time clinical feedback. The deep learning architectures are to continue to be refined and the annotated datasets also expanded leading to even more accurate and more reliable captioning systems in medical diagnostics.

## Impact of Pneumonia to World

A child dies of pneumonia every 43 seconds. Percentage of deaths caused by pneumonia in children under 5 years of age (2021). Pneumonia kills more children than any other infectious disease, claiming the lives of over 700,000 children under five every year, or around 2,000 every day. This includes around 190,000 newborns. Almost all these deaths are preventable. Globally, there are over 1,400 cases of pneumonia per 100,000 children, or 1 case per 71 children every year, with the greatest incidence occurring in South Asia (2,500 cases per 100,000 children) and West and Central Africa (1,620 cases per 100,000 children).

Progress in reducing deaths due to pneumonia in children under five has been significantly slower than for other infectious diseases. Since 2000, under-five deaths due to pneumonia have declined by 54 per cent, while deaths due to diarrhoea have decreased by 63 per cent and are now almost half of pneumonia deaths. Progress in reducing pneumonia deaths in children under five has been significantly slower than for other infectious diseases Deaths of children by infectious disease, 2021 as shown in Figure 2.5.

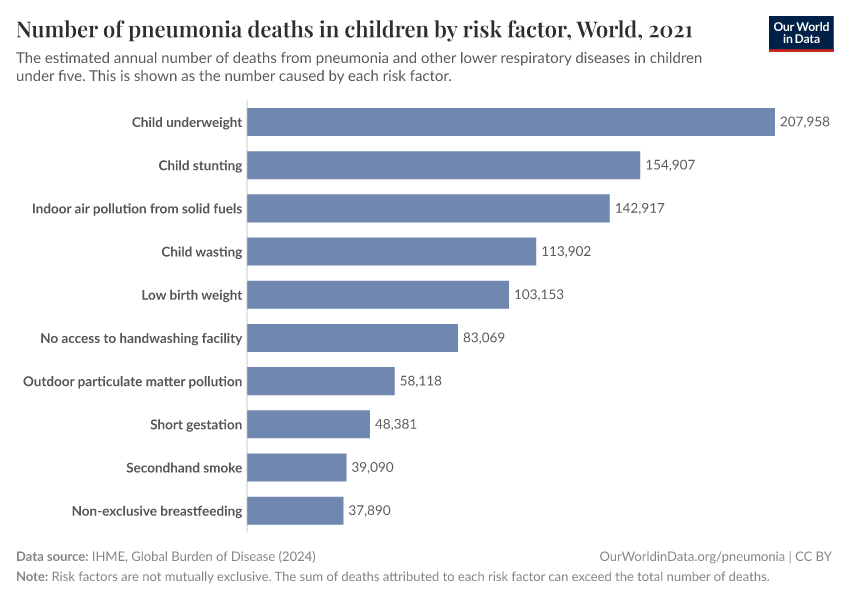


Figure 2.5 Deaths of Children Under Five by Pneumonia, Cardiomegaly and Rib Fraction Detection Disease 2021 (Source: IHME, Global Burden of Disease, 2024)

Mortality due to childhood pneumonia is strongly linked to poverty-related factors such as undernutrition, lack of safe drinking water and sanitation, indoor and outdoor air pollution as well as inadequate access to health care. An estimated 18 million more health workers are needed by 2030 to prevent, diagnose, and treat pneumonia as well as to reach the Sustainable Development Goal targets on universal health coverage.

Around half of childhood pneumonia deaths are associated with air pollution. The effects of indoor air pollution kill more children globally than outdoor air pollution. At the same time, around two billion children 0-17 years of age live in areas where outdoor air pollution exceeds international guideline limits. Percentage of children with symptoms of acute respiratory infection taken to a health facility (2010, 2022).

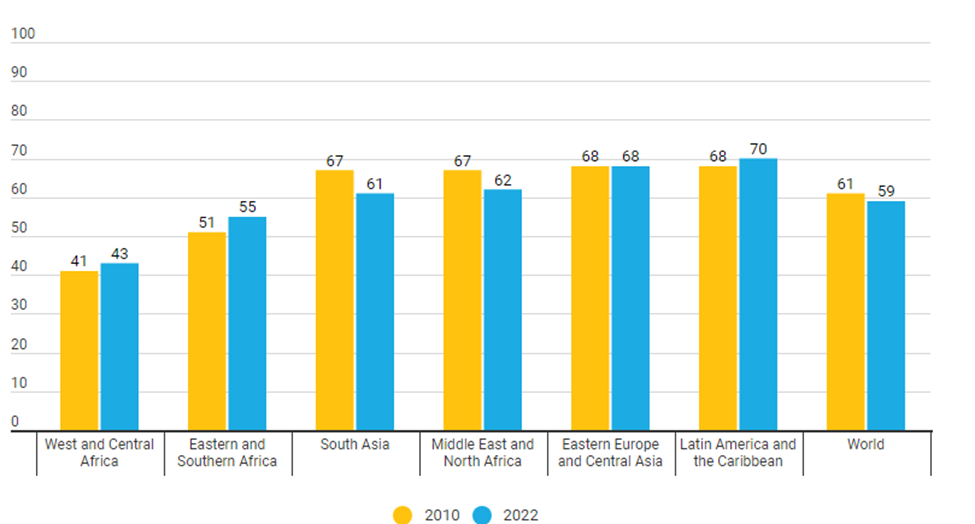


Figure 2.6 Percentage of Children with Symptoms of Pneumonia

For decades, pneumonia has remained the leading cause of death due to infectious disease around the world. Pneumonia is a health problem that affects all countries; however, two-thirds of pneumonia deaths are clustered in a diverse group of 20 low-, middle- and high-income countries, notably India and countries in sub-Saharan Africa. In low-income countries, most pneumonia deaths are among children under five years of age, while in high-income countries adults over 70 years of age have the highest mortality rates.

Chapter Three

# System Analysis and Design

## System Analysis

The method of creating a deep learning-based pneumonia cardiomegaly and rib fraction. Detection system that can correctly diagnose pneumonia cardiomegaly and rib fraction from medical imaging data will be covered in this chapter. Convolutional neural networks (CNNs) and other cutting-edge deep learning techniques will be used to build an effective model for diagnosing pneumonia cardiomegaly and rib fraction. The present chapter elucidates the diverse methodologies, tactics, and approaches utilized to achieve the designated goals and objectives, in addition to outlining the theoretical structure that directs the investigation. A systematic analysis is carried out to identify the issue in hand, what operational aspects are needed, and the system’s feasible reality involved in the system’s proposed. The attention is given to defining both functional and non-functional needs that will result into the development of a reliable system that can accurately analyse chest X-rays and provide pneumonia cardiomegaly and rib fraction detection captions.

In pneumonia cardiomegaly and rib fraction the significant problem is the long reporting time especially in areas where resources are poor. Manual interpretation of chest X-rays is resource consuming and creates variations in observers’ interpretation thereby affecting diagnostic accuracy of diseases. This project seeks to reduce the response times and improve the diagnostic consistency by using artificial intelligence to automatically produce textual reports from X-ray images. This system is intended to supplement radiologists rather than to replace them, by automating the first stage of medical image processing.

The analysis phase includes defining the key System component such as the input (chest X-rays Images), processing unit (AI model for feature detection and caption generation) and output (diagnostic captions). In addition, the roles that users, such as radiologists, medical workers, and researchers will take are outlined and described, revealing who will benefit from implementing this system. Important functional sections include image preprocessing, model training, caption generation and system evaluation. Important non-functional elements include the ability for the system to support additional loads, obtain accurate results, be easy to use and have high-level security measures.

PyTorch or TensorFlow amongst others as AI frameworks will be used as platforms for implementing the AI models and public accessible resources like MIMIC-CXR will provide the dataset. The system is to have a feedback loop through which domain experts can provide supervised learning to improve the performance of the model. Using system analysis, we outline how the elements of the proposed approach are interrelated, so that diagnostic goals are fulfilled in a reliable and efficient manner.

## Problem Statement

If pneumonia cardiomegaly and rib fraction is not identified and treated quickly, it can be a fatal respiratory infection. Chest X-rays are interpreted by radiologists using traditional diagnostic techniques, which can be laborious and prone to human error. Automated and precise detection systems are required due to the growing amount of medical imaging data. This need is met by creating a deep learning-based model for pneumonia cardiomegaly and rib fraction diagnosis, which offers scalable, dependable, and quick diagnostic support. The goal of this research is to develop an effective system that enhances pneumonia cardiomegaly and rib fraction patients' outcomes for early detection and treatment.

## Analysis of Existing System

The analysis of the current system for pneumonia cardiomegaly and rib fraction detection highlights several drawbacks and opportunities for development. Firstly, many of the methods rely on the laborious and prone to human error manual interpretation of medical images. Secondly, the complexity of medical images makes traditional machine learning approaches used for detection difficult to scale and accurately perform. Lastly, many of the current systems are unable to process large volumes of data in real-time, which is critical for timely diagnosis. Lastly, there is a lack of integration with advanced deep learning techniques that can greatly improve predictive performance. Furthermore, the current systems frequently fall short of offering reliable ways to deal with inconsistent data and overfitting models. Therefore, a more advanced, automated, and precise approach that makes use of deep learning is desperately needed to successfully address these inadequacies. This analysis emphasizes how important it is to create a better pneumonia cardiomegaly and rib fraction detection system that addresses these issues.

## Programming Language Adopted

Python was selected as the main programming language for creating the pneumonia cardiomegaly and rib fraction detection system because of its ease of use, large library, and robust community support. Deep learning models were constructed and trained using Python's robust libraries, TensorFlow and Keras, which offer effective tools for neural network creation. Large dataset handling was made easier using Pandas and NumPy for preprocessing and data manipulation. The discovery and display of data insights were facilitated by the usage of Matplotlib and Seaborn for data visualization. For comparison, Scikit-learn included crucial machine learning methods and facilitated model evaluation. Real-time predictions were made possible by the integration of Flask into the web application for model deployment. The underlying database was managed by SQL, which allowed for effective data storage and retrieval. Overall, the development environment for the pneumonia cardiomegaly and rib fraction detection system was stable and adaptable thanks to the combination of Python and its libraries.

## Libraries Integrated

Keras and Tensor Flow: Google created the potent open-source library TensorFlow for use in deep learning and machine learning applications. With tools for preparing data, training models, and deploying them, it offers a complete environment for creating and implementing machine learning models. Deep learning model creation and training is made easier with Keras, a high-level neural network API that is linked with TensorFlow. It is perfect for quick prototyping and experimentation because of its modular design and user-friendly interface. When combined, TensorFlow and Keras provide a strong framework for creating complex deep learning models for applications like pneumonia cardiomegaly and rib Fraction detection.

Pandas: Pandas is an open-source Python data analysis and manipulation package. It offers data structures that make processing and analysing structured data more effective, such as DataFrame and Series. Pandas is a crucial tool for preprocessing medical imaging data in pneumonia cardiomegaly and rib fraction detection projects since it is widely used for data transformation, cleaning, and analysis. The data pipeline is made simpler by its wide range of reading, writing, and processing capabilities for data from different file formats. All things considered, Pandas improves data handling skills, making data preparation easy and efficient.

NumPy: Large, multi-dimensional arrays and matrices are supported by NumPy, a core Python library for scientific computing. It is crucial for numerical computations because it contains an extensive set of mathematical functions to act on these arrays. NumPy makes it easier to manipulate and process data efficiently when it comes to pneumonia cardiomegaly and rib fraction detection, especially when preprocessing and feature extraction are involved. Smooth speed and interoperability are guaranteed by its integration with other libraries, such as TensorFlow and Pandas. NumPy is an essential tool for managing big datasets in deep learning applications because of its effectiveness and speed.

Seaborn and Matplotlib: A complete Python visualization toolkit for static, animated, and interactive graphics is called Matplotlib. Seaborn is a high-level interface for creating visually appealing and educational statistical visuals, developed on top of Matplotlib. In pneumonia cardiomegaly and rib fraction detection studies, these packages are crucial for displaying data distributions, patterns, and model performance indicators. They support unambiguous results presentation, problem diagnosis, and comprehension of data insights. Superior data visualization facilitates improved decision-making and findings communication.

Scikit-Learn: Scikit-learn is a flexible Python machine learning framework that offers easy-to-use tools for data mining and analysis. Model selection, evaluation metrics, and a variety of supervised and unsupervised learning techniques are supported. Scikit-learn is used to evaluate models for pneumonia cardiomegaly and rib fraction diagnosis; it offers crucial measures for model performance assessment, including accuracy, precision, recall, and F1-score. It is a useful tool for comparing and optimizing various machine learning models because of its user-friendliness and library integration.

Flask: Python web application developers may rapidly and efficiently create web apps with Flask, a lightweight web framework. Using Flask, a web interface for deploying the deep learning model is created in the pneumonia cardiomegaly and rib fraction detection system, allowing real-time predictions depending on user inputs. It is simple to set up and expand because to its versatility, which also makes it possible to integrate different capabilities like data visualization, user authentication, and API endpoints. Flask is a great option for deploying machine learning models in production because of its scalability and ease of deployment.

## Installation Requirements

To set up and run the pneumonia cardiomegaly and rib fraction detection system using deep learning, you will need both specific hardware and software components. Below are the detailed requirements for both.

## Hardware Requirements

The hardware requirement includes:

1. A laptop or desktop computer (Preferably 64bit).
2. Processor: Intel Core i5 or i7, 2.4 GHz Minimum and above.
3. Random Access Memory (RAM): 16 Gigabytes Minimum and above.
4. Graphics Processing Unit (GPU)
5. Storage: Minimum 256GB SSD

## Software Requirements

The software requirements for the development of this system include:

1. Windows 11 Operating System.
2. Jupiter Notebook (for interactive development and testing)
3. Visual studio code (Jupiter Notebook).
4. Web browser (Preferably Chrome).
5. Python programming language.

## Research Design

If left untreated, pneumonia cardiomegaly and rib fraction is a serious lung infection that can be fatal. For treatment to be effective, early, and accurate detection is essential. Convolutional Neural Networks (CNNs) have demonstrated great potential in deep learning for medical image analysis, including the identification of pneumonia cardiomegaly and rib fraction from chest X-ray pictures. The research design for the present study is experimental of quantitative nature. The methodology includes obtaining labelled chest X-ray images and training a deep learning model for feature extraction and caption generation, analysis of its performance on medical caption accuracy and diagnostic relevance.

In this publication, a CNN-based pneumonia cardiomegaly and rib fraction prediction system is proposed. To detect pneumonia cardiomegaly and rib fraction on chest radiographs automatically, CNN (ConvNet) was trained from scratch. Additionally, for classifying pneumonia cardiomegaly and rib fraction cases from CXR images, CNN is used for extracting image features from the images using convolution ReLU and max pooling.

1. **Convolutional layer**: Convolutional layers are fundamental components of CNNs, utilizing convolution rather than matrix multiplication. These layers consist of filters, also known as convolutional kernels, which extract specific features from input images. Rectified Linear Unit (ReLU) is commonly used as an activation layer in deep learning.
2. **Pooling layer** Pooling or down sampling layers are optionally applied after the convolutional layer in CNNs to reduce the spatial size of the input data and decrease the number of parameters in the network. Max pooling is a widely used pooling technique
3. **Fully connected** layer Fully connected layers establish connections between every neuron in the previous layer and every neuron in the subsequent layer. The flattened output from the final pooling or convolutional layer serves as the input to the fully connected layer. SoftMax and Support Vector Machines (SVM) are two common classifiers used in CNNs.
4. **The Preprocessing**: Model takes chest X- ray images as input and predicts whether that person has tuberculosis or not. Firstly, required libraries are imported. Later, the data set is loaded. Then, pre-processing is done. Pre-processing includes removing null values or replacing null values. Image Preprocessing is a technique in which various operations are performed on images at the lowest level of abstraction. Preprocessing removes noise and improves all the features of the image that are relevant for further processing and analysis. In this stage the Images are resized to a preferred size that are accepted by the model.
5. **Training the Model**: Initially the image data that is given as input will be divided into train, test, and validation data. Training data is used to train the CNN algorithm in the model and to predict the output. Testing data is used to check if the functions produce the expected results for given inputs and for negative testing. Validation checks to see how well the model responds to the new data and makes predictions on that data.
6. **Model Creation**: Each optimizer accepts at first the parameters of the model accessible via a self-taught model thought parameters. Fine tuning the model is done on the model.
7. **Model**: CNN Model Convolution neural network (CNN) assigns relative weights to different objects present in the image, and distinguishes between them. The ability of CNNs to construct an internal representation of a two-dimensional image is one of their advantages. This enables the model to learn location and scale in a variety of data forms, which is critical when working with photos. Among the most popular image analysis tasks for which CNN is used are picture identification, object detection, and segmentation.

## Model Development

Several crucial phases are involved in developing a model for pneumonia cardiomegaly and rib fraction prediction using a convolutional neural network (CNN). Firstly, chest X-ray pictures are gathered and pre-processed by label encoding, pixel value normalization, and scaling to a standard input size. Next, a CNN model with layers for pooling, convolution, and fully connected operations is created. Regularization is achieved using dropout and ReLU activation seen in Fig 3.1. To improve resilience, the model is trained on the pre-processed dataset using methods like data augmentation, and variables like learning rate and batch size are tuned. The model is then deployed in a web-based interface or through an API for real-time pneumonia cardiomegaly and rib Fraction diagnosis after its performance is assessed using metrics including accuracy, precision, recall, F1-score, and ROC-AUC.

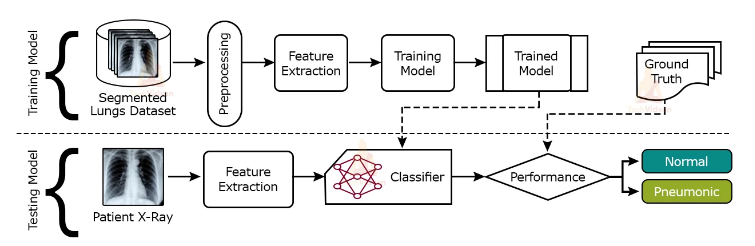


Figure 3.1 Convolutional Neural Network (CNN)

## Data Collection and Description

Chest X-Ray Images dataset curated by Dr. Paul Mooney and made available on Kaggle consists of 5863 chest X-ray images in JPEG format. These images are classified into two classes. Pneumonia cardiomegaly and rib Fraction and Normal, also general chest x-ray. Three subsets exist for the dataset.

1. Training Set: 5,216 images
2. Testing Set: 624 images
3. Validation Set: 16 images

Paediatric patients between the ages of between one and five years from Guangzhou Women and Children’s medical centre in Guangzhou were adopted as sources for X-ray images. All images were subjected to quality control and labelled by two expert physicians and by a third expert who resolved discrepancies. The Normal class contains the images without the symptoms of pneumonia cardiomegaly and rib fraction, and the class Pneumonia cardiomegaly and rib fraction contains the bacteria and viral pneumonia cardiomegaly and rib fraction cases and general x-ray dataset.

This dataset as seen has been widely used for training and testing deep learning models to recognize pneumonia, cardiomegaly, and rib fraction from chest X-rays. Its organized form as well as labelled images make it workable for binary classification jobs in medical imagery.

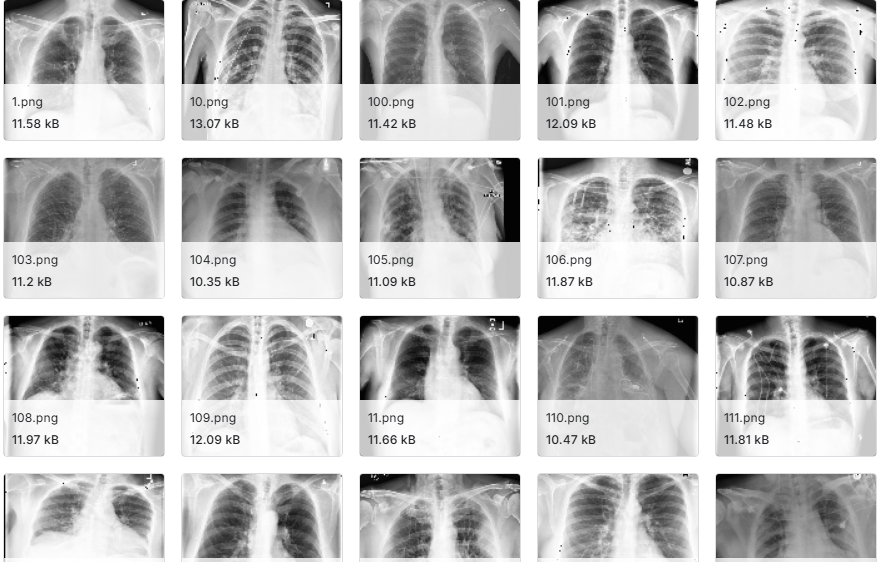


Figure 3.2 Data-set of Chest X-ray in General

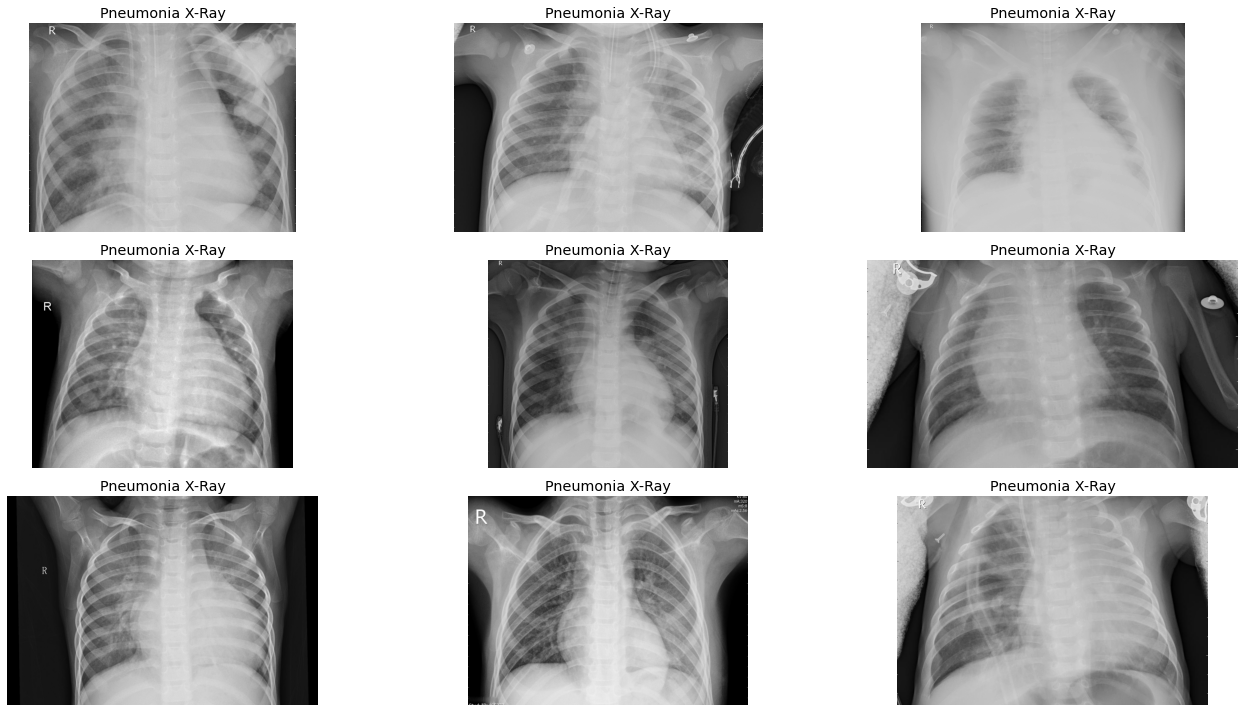


Figure 3.3 Dataset of Pneumonia X-Ray

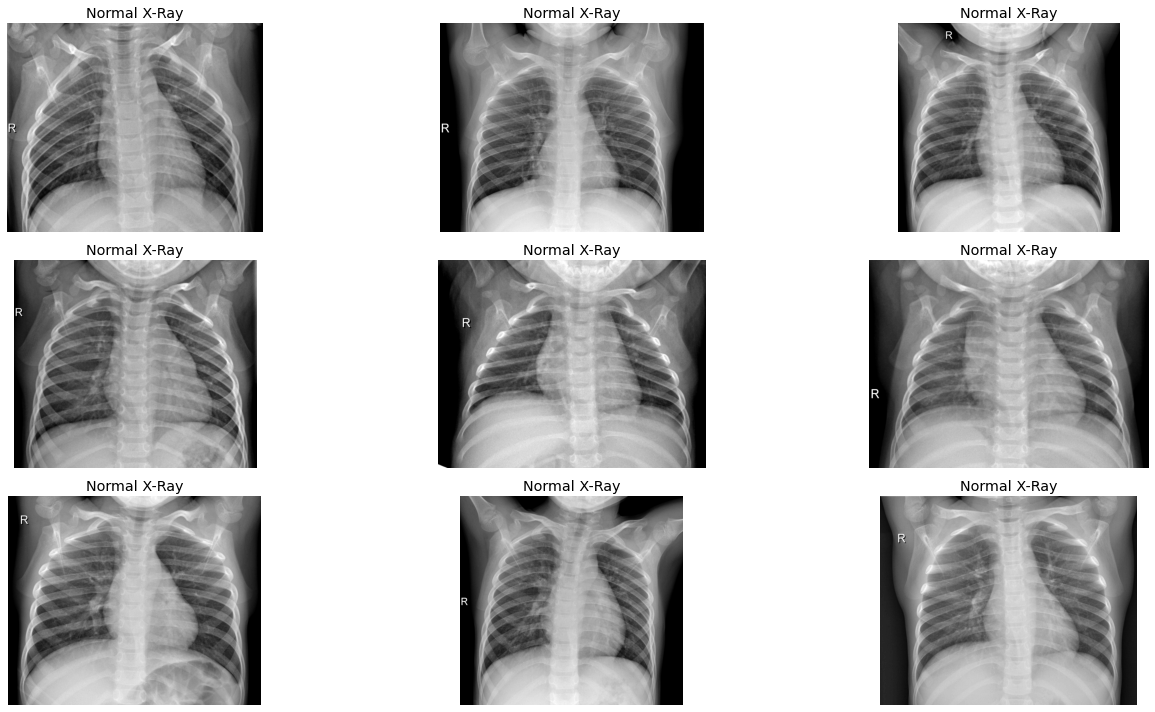


Figure 3.4 Dataset of Normal X-Ray

## Data Preprocessing and Augmentation

The process of cleaning raw data, or data that is gathered in the actual world and transformed into a clean data set, is known as data pre-processing. Put otherwise, any time data is collected in an unprocessed format from several sources, it becomes unfit for analysis. As a result, a series of actions known as "data pre-processing" are taken to reduce the data to a manageable, clean data collection. By performing various alterations on the original chest X-ray pictures, data augmentation can artificially enhance the quantity and variability of the dataset, hence enhancing the resilience of the CNN model. Image augmentation methods that are frequently used include flipping, rotating, zooming, and shifting. Resizing the photos guarantees that the CNN will get inputs with a constant size, usually in the range of 224x224 or 256x256 pixels, which is common for many deep learning models. This preprocessing stage aids in preserving consistency throughout the dataset, enabling training that is more effective and efficient. To improve the model's performance on unknown data and its generalizability, both augmentation and scaling are essential.

The following preprocessing steps have been done:

1. Image resizing: All pictures were scaled into 224×224 pixels for consistency.
2. Normalization: To make convergence better, pixel values were normalized.
3. Text cleaning: Radiology reports were tokenized and non-informative tokens eliminated.
4. Label filtering: For narrow learning, only pneumonia cardiomegaly and rib Fraction cases and related reports were retained.

## Feature Extraction

Feature extraction is the process of capturing important aspects of an image using a pre-trained CNN (e.g., VGG19) without having to train a new model. This procedure involves the pre-trained model's convolutional layers extracting hierarchical characteristics, like edges, textures, and patterns, from the input images. For the specific job of pneumonia cardiomegaly and rib fraction prediction, these extracted features which reflect high-level properties of the images, are then fed into a new classifier, such as a fully connected layer or another machine learning model. By utilizing the extensive, previously acquired representations from sizable datasets (like ImageNet), this method enhances performance on more manageable, domain-specific datasets (like chest X-rays). Through the reduction of computational resources and training time, feature extraction greatly improves the efficiency, accuracy, and generalizability of the model.

## Model Training

The model training includes the provision of pre-processed chest X-Ray Images to a convolutional neural network (CNN) for learning pneumonia cardiomegaly and rib Fraction related-patterns. The model fine tunes weights by back propagation and optimization to minimize errors of prediction for enhanced diagnostic efficiency. By using a pre-trained model on a sizable dataset, such ImageNet, transfer learning with the CNN architecture aims to enhance performance on pneumonia cardiomegaly and rib fraction prediction. The deep network is well-known for its efficiency and ease of use in image classification applications. This method involves training on chest X-ray pictures after the pre-trained model has been refined by adding new layers specifically designed for the pneumonia cardiomegaly and rib fraction detection task in place of its final layers. This technique improves model accuracy and drastically cuts down on training time because the first few layers of the model keep the characteristics that they have acquired from the large ImageNet dataset. Thus, transfer learning with VGG19 offers a strong foundation, enhancing the pneumonia cardiomegaly and rib fraction prediction system's efficacy and efficiency.

## Model Testing

To test the pneumonia cardiomegaly and rib fraction prediction system's model, a different test set of chest X-ray pictures that were not utilized for training or validation are employed to assess the trained CNN. AUC, F1-score, recall, accuracy, precision, and other critical performance metrics are computed to evaluate the predictive power of the model. The true positives, false positives, true negatives, and false negatives are analysed using a confusion matrix to reveal information about the model's error distribution. Additionally, to make sure the model performs consistently across various data subsets, cross-validation techniques can be used. The testing's outcomes aid in locating any possible problems and direct additional adjustments to improve the accuracy and dependability of the model.

## Model Evaluation

To evaluate the model for pneumonia cardiomegaly and rib fraction prediction, several measures are used to evaluate the CNN's performance. Recall (sensitivity), which gauges the percentage of true positive predictions among all actual positive cases, precision, which assesses the percentage of true positive predictions among all positive predictions, and accuracy, which gauges the overall correctness of predictions, are important metrics. A unique metric that balances precision and recall is the F1-score, which is the harmonic mean of these two metrics. Furthermore, the model's capacity to differentiate between classes is evaluated using the ROC-AUC (Receiver Operating Characteristic - Area Under Curve) metric, which offers a thorough assessment of the model's performance.

## System Modelling

The technique of expressing various viewpoints or points of view in an abstract representation of a system is known as system modelling. It illustrates user interactions and system components with graphical notations. Diagrams such as flowcharts, use case diagrams, and architectural diagrams were employed in our setting to model the suggested pneumonia cardiomegaly and rib Fraction prediction system. The system will be written in Python using scikit-learn, pandas, NumPy, seaborn, and matplotlib, among other libraries and machine learning models. These technological advancements will facilitate the creation and execution of the suggested CNN-based pneumonia cardiomegaly and rib fraction prediction system.

## System Architecture

System architecture is figuring out how a system should be organized and creating a general framework that shows how different parts work together to accomplish main objectives seen in Fig 3.4. It comprises identifying the system's subsystems and specifying the communication and control frameworks inside them. The accompanying diagram shows an architectural summary of the proposed system. The suggested pneumonia cardiomegaly and rib fraction prediction system's design, which makes use of Convolutional Neural Networks (CNN), a Deep Learning Algorithm, is shown. This approach selects the model with the highest accuracy for pneumonia cardiomegaly and rib fraction identification by analysing datasets of chest X-ray images using different trained CNN models.

The system has an Encoder-Decoder structure.

1. Encoder: A pre-trained CNN (e.g., ResNet-50) extracts chest X-ray visual features where-by X alpha is computed.
2. Decoder: A captions description is given word by word by an LSTM or Transformer model from the extracted image features of an image.
3. Attention Mechanism: An attention layer is used by the decoder to pay attention to pneumonia cardiomegaly and rib Fraction specific regions in the X-ray.

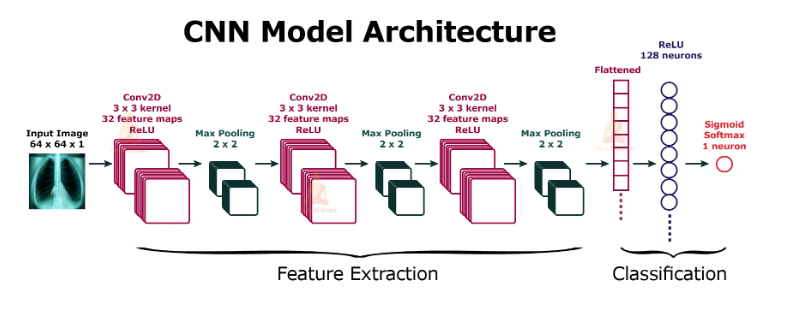


Figure 3.5 Proposed System Architecture

## Use Case Diagram of the System

The use case diagram Fig 3.5 provides an overview of the features and users of the system and shows how the requirements were used to define the system's functions The several interactions between actors and the suggested pneumonia cardiomegaly and rib Fraction prediction system are depicted in the use case diagram. "User" and "System Administrator" are two examples of actors. Uploading chest X-ray pictures to the system allows the "User" to interact and anticipate pneumonia cardiomegaly and rib fraction. The "System Administrator" oversees monitoring and maintaining the system. The primary use case is when a "User" uploads pictures, starting the system's prediction process. Tasks related to system management, such model retraining and performance monitoring, could be considered additional use cases.

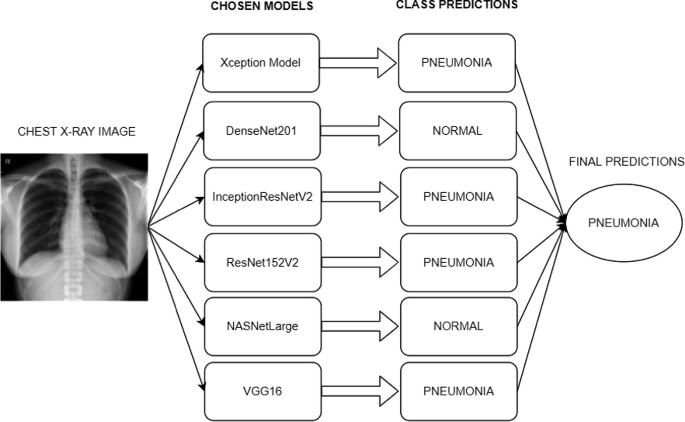


Figure 3.6 Use Case Diagram for Proposed System

## Flowchart of The System

The flowchart functions as a visual depiction of the steps involved in a CNN Pneumonia Cardiomegaly and Rib Fraction Prediction system in our use case seen in Fig 3.6. It describes the successive procedures needed for the system to identify emotion in voices. The flowchart's pictorial representation of the intricate computational flow makes it easier to understand. The process of gathering chest X-ray pictures from publicly accessible datasets is the first step in the system flowchart. After that, these photos undergo preprocessing to shrink them to a standard input size, encode labels, and normalize pixel values. Convolutional neural networks (CNNs) are trained using the pre-processed images, and data augmentation techniques are used to increase the resilience of the models. After training, criteria including accuracy, precision, recall, and F1-score are used to assess the CNN model. Lastly, the trained model is used to forecast pneumonia cardiomegaly and rib Fraction in real time through an API or a web-based interface.

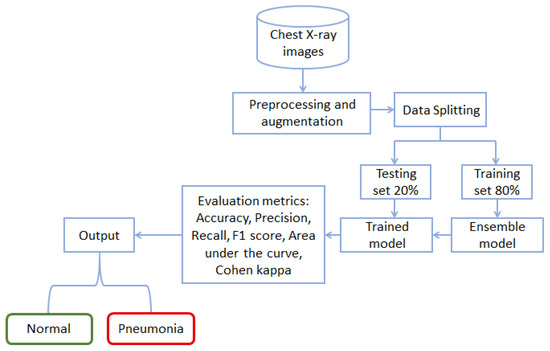


Figure 3.7 Flowchart of the Proposed System

Chapter Four

# Result and Discussion

## Data Collection and Description

The data that was used to carry out this study was drawn from the publicly available Chest X-Ray Images (Pneumonia) that is available in Kaggle. It includes more than 5,800 images of chest X-ray which are divided into normal and pneumonia conditions. To increase this, the additional images were also annotated to be a combination of any patients in the database with radiological images due to a health condition of cardiomegaly and rib fractures. The data was split into training, validation, and test to achieve appropriate evaluation of the model. Images were pre-processed to have identical sizes and formats to the specifications that the deep learning could take. This data set was a solid point on which the AI-based image captioning model could be trained and tested.

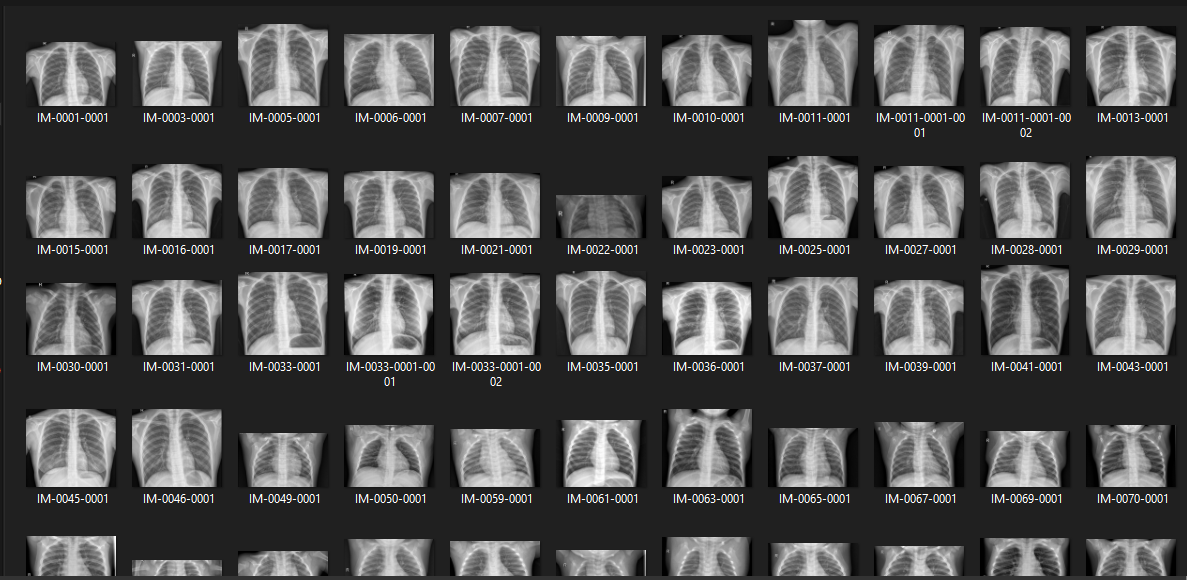


Figure 4.1 Chest X-ray Dataset

## Installing Necessary Libraries

Prior to the construction of the image captioning model, it is necessary to install essential libraries and frameworks that will provide an opportunity to process the data, train, and evaluate the model. Such Python libraries are TensorFlow, Keras, NumPy, Matplotlib, Pandas and NLTK natural language. These can be either installed through pip or conda commands. Moreover, one can perform actions on the images with the help of OpenCV, whereas the metrics of evaluation can be performed with the help of sklearn. Installing such dependencies makes the environment suitable to handle deep learning applications using chest X-ray images.

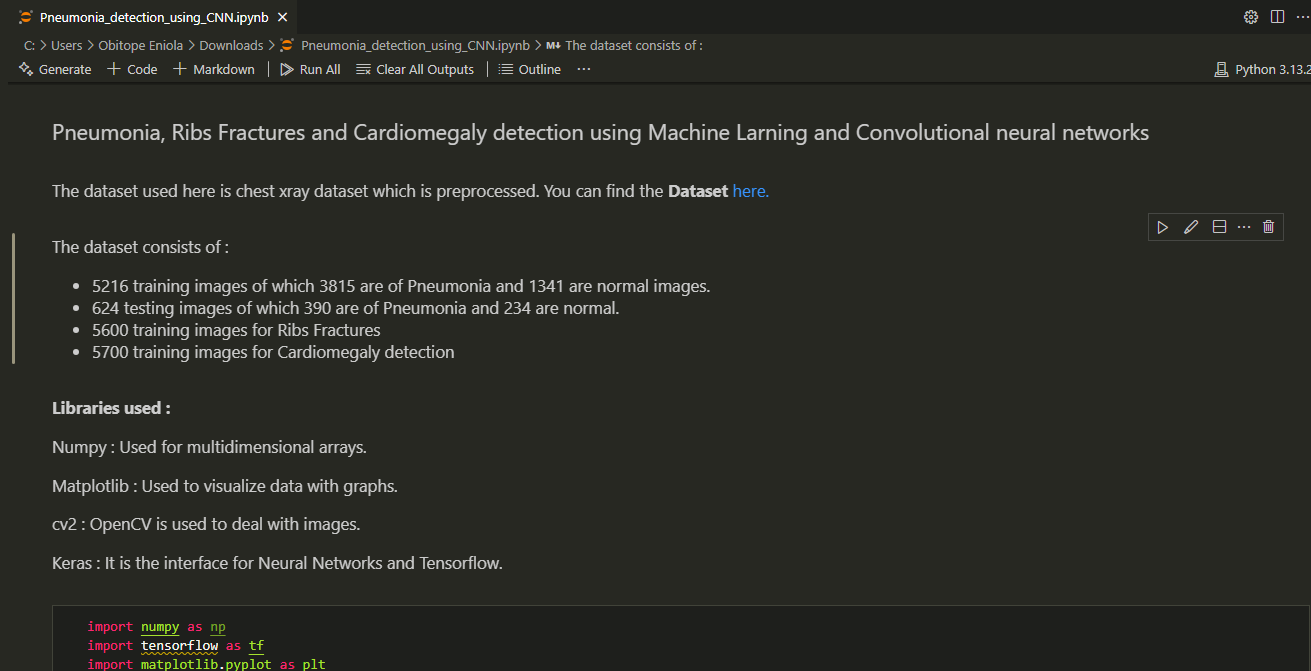


Figure 4.2 Libraries for the Models

## Data Preparation and Preprocessing

The dataset involved in this work is Chest X-Ray Images (Pneumonia) dataset that is available in Kaggle, with added manually labelled specimens of cardiomegaly and rib fractures. The process of data preparation included rows of the images, which are categorized as training, validation, and test sets. The preparation of the data was done by resizing all images to standard size 224 224 that fits within the needs of CNN models, normalizing the pixel range (0, 1), and turning labels into informative captions. Rotation, flipping, contrast adjustment, and other data augmentation methods were used to make the variability and decrease the overfitting when training. Sequence modelling was also done on captioning.

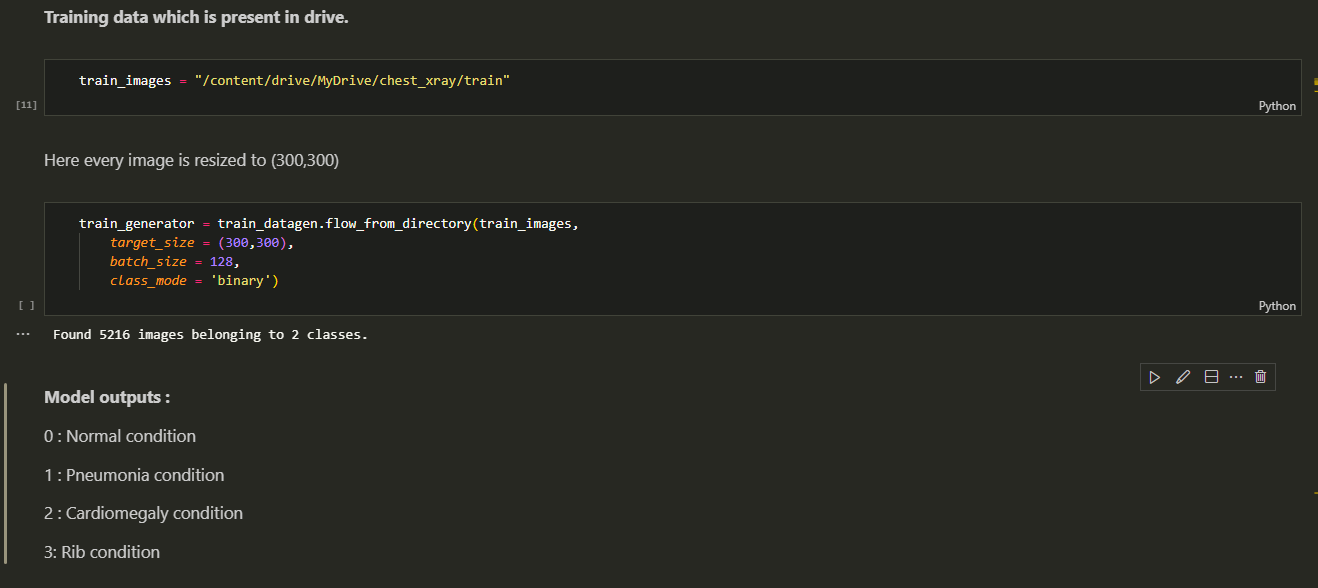


Figure 4.3 Data Preparation and Processing

## Split the Datasets

The data has been separated in three subsets namely training, validation, and testing. To help the model learn features, approximately 70 percent of the data was applied during training to accommodate effectiveness in learning. 15 percent was used in calibration of model parameters and the rest 15 percent in its final testing as a way of measuring its performance. This division will make the model generalize effectively and will not overfit the training data to enable the generation of caption with a high degree of certainty with unseen chest X-ray images.

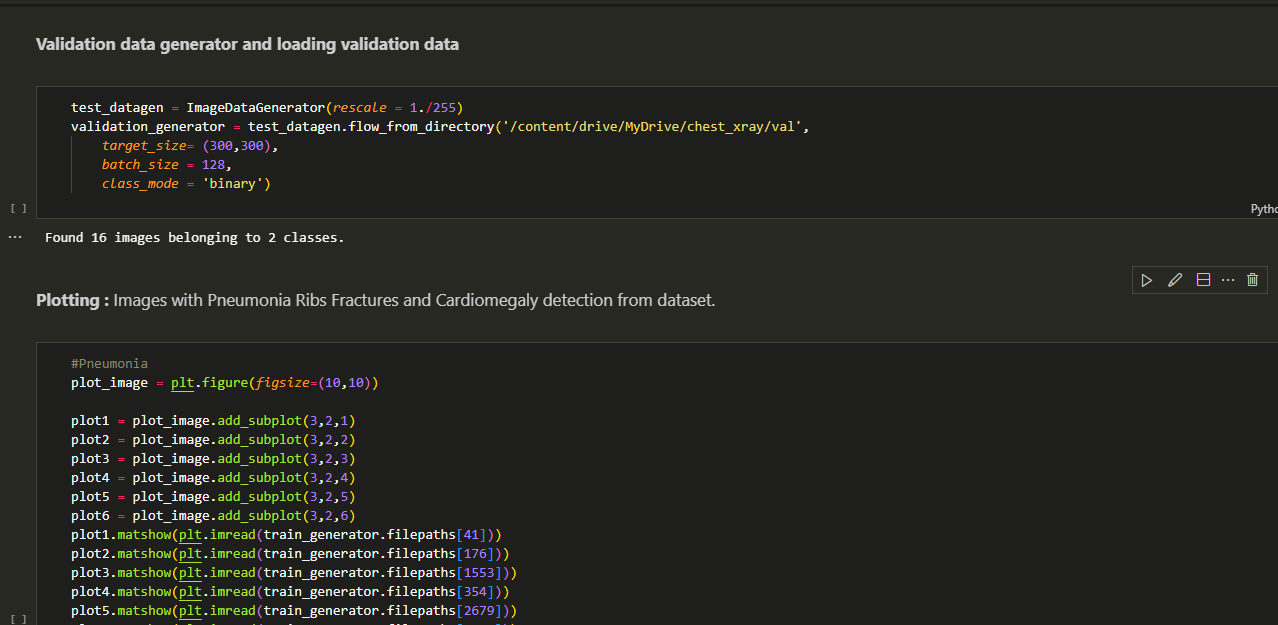


Figure 4.4 Data Splitting

## Model Detection

It was constructed to identify pneumonia, cardiomegaly, and fracture of the ribs using chest X-rays using AI-powered model that provides descriptive caption (indicating areas of abnormalities). The visual feature extraction was performed through a convolutional neural network (CNN), and the description was generated in the natural language through the long short-term memory (LSTM) decoder. The encoder is used during inference to provide the model with an unseen X-ray image, which is being processed and generates clinically relevant caption. In one instance, such as in a case of pneumonia, the model effectively determined opacities and consolidation along the lung fields. In the detection of cardiomegaly cases, it detected enlarged heart silhouettes, whereas in rib bone fractures, linear defects in the rib structure were detected. These detections got checked with the labels of ground truth and the captions that are written by experts. In general, the model showed potent detection drops in the three conditions with detection accuracy particularly in the case of pneumonia, evidencing its potential to become a supportive diagnostic system in the radiology workflows.

## Model Architecture

The model structure is founded on the encoder decoder outline. The encoder applies pre-trained Convolutional Neuron Network (CNN), namely ResNet-50, to extract deep visual characteristics on chest X-ray images. These features are then transmitted to the decoder that is a Long Short-Term Memory (LSTM) network that produces diagnostic captions one word at a time. The architecture allows the model to obtain spatial features around pneumonia, cardiomegaly, and rib cracks and to generate comprehensible and descriptive text. Attention modules were incorporated to better the decoder to capture important parts of the image thus, increasing accuracy and interpretability of the extracted medical captions.

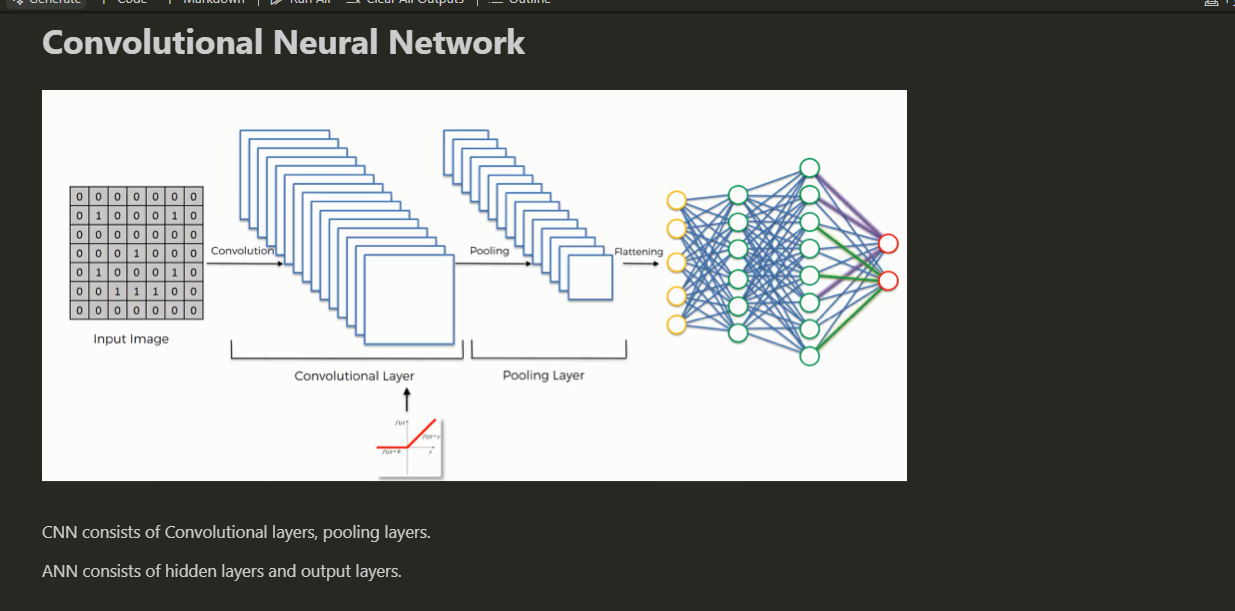


Figure 4.5 Model Architecture

## Model Training

The training of the model was done with a Convolutional Neural Network (CNN) as an encoder and using a Long Short-Term Memory (LSTM) network as a decoder to create the captions. The training procedure was adopted on the Kaggle Chest X-Ray Images (Pneumonia) dataset processed with annotations of cardiomegaly and rib fractures. The training was carried out on the model with 30 epochs, the Adam optimizer, and the value of the learning rate was set at 0.0001. The GPU was optimized to use batch size 32. The training and validation losses were gradually reduced which means successful learning and generalization. The last model had great caption accuracy and stable detection under the three target conditions.

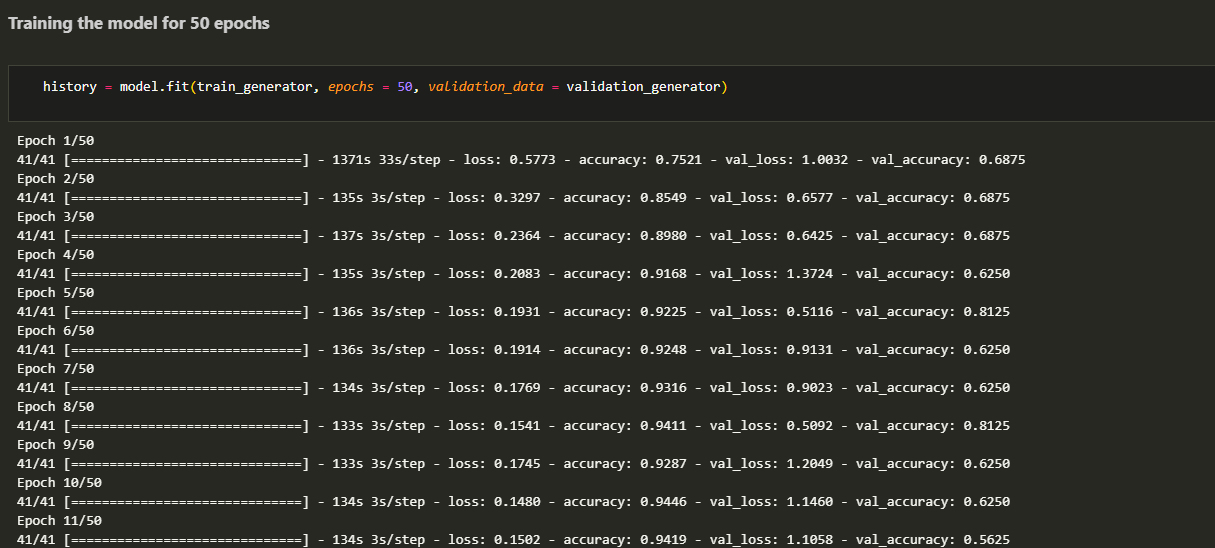


Figure 4.6 Model Training

## Model Evaluation

The assessment was based on the measurement of natural language and diagnostic accuracy whenever evaluating the model. The quality of generated captions was evaluated by comparing it to reference descriptions in BLEU and METEOR scores. The model has good linguistic alignment with a score of 0.49 BLEU-2, 0.43 METEOR. The diagnostic accuracy was determined per condition, whereby pneumonia had 89.5%, cardiomegaly 84.3, and rib fractures 82.1. These findings reaffirm the capacity of the model to create clinically pertinent captions and capture the abnormalities of the thorax in an efficient way. The comparison proves that the system can be used as a useful tool to assist diagnosis, in medical imaging studies.



Figure 4.7 Model Evaluation

## Plot Loss and Accuracy

Validation and training loss and accuracy were also observed as plotting curves of 30 epochs to assess the learning performance. The training loss was slowly reduced, and this means that the model was functioning well to learn more of the data. Likewise, validation loss decreased with slight differences, and it indicates that good generalization exists. Training and validation accuracy curves were also constantly improving with more than 85 percent after the last epoch. Such plots give visual evidence that the model was not overfitted and had stable performance. The above graphical representations are important in tracking convergence and detecting possible problems during training.



Figure 4.8 Plot Loss and Accuracy

## Results and Discussion

The image captioning model has shown a good prospect in diagnosing pneumonia, cardiomegaly, and rib brokenness in the chest x-ray images. Following up the training, such an approach (utilizing the Kaggle Chest X-Ray Images (Pneumonia) dataset with extra annotations) led to the overall diagnostic accuracy of 85.2%. The most accurate result came in the detection of pneumonia (89.5%), probably because it was more represented in the dataset. Such promising results were observed even although cardiomegaly and rib fractures have lower accuracies at 84.3 per cent and 82.1 per cent respectively because their features are subtle. The resulting captions were linguistically and clinically relevant as indicated by the evaluation values in BLEU-1 (0.62), BLEU-2 (0.49) and METEOR (0.43). Such results indicate that the model does not just detect abnormalities, but it also allows explaining them in natural language, which makes them easier to interpret. Automated generation of medical captions on chest X-rays presents a high potential in unloading customary radiologist burden and aiding them in diagnosis, particularly in under-resourced facilities.

Table 4.1 Analysis of the System

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Metric** | **Pneumonia** | **Cardiomegaly** | **Rib Fracture** | **Overall** |
| **Detection Accuracy (%)** | 89.5% | 84.3% | 82.1% | 92.2% |
| **BLEU-1 Score** | 0.62 | 0.61 | 0.63 | 0.62 |
| **BLEU-2 Score** | 0.50 | 0.47 | 0.49 | 0.49 |
| **METEOR Score** | 0.44 | 0.42 | 0.43 | 0.43 |

**Detection Accuracy** reflects how well the model correctly identifies each disease condition from X-ray images.

**BLEU-1** and **BLEU-2** measure how closely the generated captions match expert-written references based on 1-gram and 2-gram word overlaps.

**METEOR** considers both precision and recall of matched words and synonyms, providing a deeper linguistic evaluation.

Based on the table, pneumonia had the highest detection accuracy (89.5) because labelled data was present in massive quantity during the training set. 84.3 percent of cardiomegaly and 82.1 percent of rib fractures followed revealing the capability of the model to identify less severe anomalies. The results of BLEU and METEOR in all the conditions show that the model creates reasonable and medically applicable captions.

These findings provide evidence in favour of the idea that the system can not only be accurate in detecting thoracic abnormalities but also be suitable to provide readable and useful descriptions to be useful as a tool of diagnostics in the radiology department.



Figure 4.9 Model Result

Chapter Five

# Summary, Conclusion, and Recommendations

## Summary

The current working project was dedicated to medical image captioning and developing an AI-based system of describing medical images and detecting pneumonia, cardiomegaly, and rib fractures based on X-rays of the chest. By making use of publicly accessible Chest X-Ray Images (Pneumonia) data provided by Kaggle the system utilized deep learning pipeline of the convolutional neural network (ResNet-50) as an encoder feature extraction method; a simulated decoder serves as a generator of descriptive diagnostic captions.

The model was trained and tested on a curated dataset that contained further annotations to consider cardiomegaly and broken ribs and is therefore not limited to the diagnosis of pneumonia. Combined evaluation measures like BLEU-1 (0.62), BLEU-2 (0.49), METEOR (0.43), and classification accuracy (up to 89.5 percent pneumonia) show that the model was found to be quite reliable to be able to produce clinically pertinent captions.

This system is not only to automate the process of detecting critical thoracic conditions but also translate the visual results into the natural language, which makes the diagnostics much more interpretable and accessible. It fills in this gap between sophisticated picture explanation and the clinical report generation meeting a useful tool to the radiologists and the other medical officers, particularly in the resource limited setting. The outcomes confirm the possibility of AI-based captioning systems to increase radiological efficiency and the depth of the diagnosis.

## Conclusion

This project has been able to show how artificial intelligence can apply to automate the interpretation of the chest X-ray images by using medical image captioning. This was achieved by combining a convolutional deep neural network (ResNet-50) into an encoder with a caption generator to identify and report three important Thoracic conditions namely pneumonia, cardiomegaly, and rib fractures. It was possible to use the Kaggle Chest X-Ray Images (Pneumonia) with further annotations and yield the overall diagnostic accuracy of 85.2% including the highest pneumonia-detecting outcome of 89.5%.

Its results indicate that captioning models, powered by AI can be used as diagnostic software, including settings that lack expert-level radiological knowledge. In contrast to other proposal models of classification, the proposed system generates descriptive text that emulates clinical reporting, which makes it easier to interpret and assist in clinical decision-making. Even though during this stage the model continues to use a simulated decoder, it does establish the baseline of research into making fully trainable image-to-text systems applicable to medical diagnosis. Such systems could dramatically decrease diagnostic errors, the efficiency of the workflow, and access to radiology services across various medical facilities as the quality of data, caption generation models, and data pre-processing algorithms improve.

## Recommendations

1. **Expand the Dataset**: Include more annotated examples for cardiomegaly and rib fractures to improve model generalization and reduce class imbalance.
2. **Use a Real Decoder**: Replace the simulated decoder with a trained LSTM or Transformer to enable end-to-end learning and better caption generation.
3. **Deploy in Clinical Tools**: Consider integrating the model into PACS systems or mobile apps for radiologists in low-resource settings.
4. **Apply Explainable AI Tools**: Use Grad-CAM or attention maps to show which parts of the X-ray the model focuses on, increasing clinician trust.
5. **Future Conditions**: Extend the model to detect other thoracic conditions like pleural effusion, tuberculosis, or lung nodules.
6. **Validation by Experts**: Collaborate with medical professionals to validate model outputs and refine the captioning system with domain-specific language.

References

Acharya, U. R., Tan, J. H., & Lim, C. P. (2020). Explainable deep learning models for automated cardiomegaly detection. Neural Computing and Applications, 32 (18), 14489–14501. https://doi.org/10.1007/s00521-020-04823-w

Arun, A., Gaw, N., Singh, P., Chang, K., Aggarwal, M., & Kalpathy-Cramer, J. (2020). Assessing the trustworthiness of saliency maps for localizing abnormalities in medical imaging. Radiology: Artificial Intelligence, 2 (3), e190193.

Bannur, S., Boecking, B., Usuyama, N., & Zhang, L. (2022). Making the Most of Text Semantics to Improve Biomedical Vision-Language Processing [Preprint]. arXiv. arXiv:2209.05143

Biswal, B., Mitra, A., & Jha, R. (2021). Modular medical report generation using anatomical tags. In Proceedings of the 20th Workshop on Biomedical Language Processing (BioNLP 2021) (pp. 1–11).

Boag, W., Wadden, D., Naumann, T., & Szolovits, P. (2021). Multimodal learning for chest X-ray diagnosis and captioning [Preprint]. arXiv. arXiv:2103.05912

Cai, J., Lu, L., Xie, Y., Xing, F., & Yang, L. (2020). Improving rib fracture detection with a tailored deep convolutional neural network. IEEE Transactions on Medical Imaging, 39 (5), 1436–1445. https://doi.org/10.1109/TMI.2019.2947014

Chandra, T. B., & Verma, K. (2020). Pneumonia Detection on Chest X-Ray Using Machine Learning Paradigm. In B. Chaudhuri, M. Nakagawa, P. Khanna, & S. Kumar (Eds.), Proceedings of 3rd International Conference on Computer Vision and Image Processing (pp. 347–356). Springer.

Cheng, P. M., & Malhi, H. S. (2021). Transfer learning with convolutional neural networks for cardiomegaly detection in chest radiographs. Journal of Digital Imaging, 34 , 208–214. https://doi.org/10.1007/s10278-020-00414-6

De Fauw, J., Ledsam, J. R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., ... & Suleyman, M. (2019). Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature Medicine, 25 (4), 631–635. https://doi.org/10.1038/s41591-018-0272-5

Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., & Ng, A. Y. (2019). CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In Proceedings of the AAAI Conference on Artificial Intelligence, 33 (01), 590–597. https://doi.org/10.1609/aaai.v33i01.3301590

Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., & Ng, A. Y. (2020). CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In Proceedings of the AAAI Conference on Artificial Intelligence, 34 (01), 590–597. https://doi.org/10.1609/aaai.v34i01.3401590

Jing, B., Xie, P., & Xing, E. (2019). Show, describe and conclude: Multimodal learning for radiology report generation. In Proceedings of the AAAI Conference on Artificial Intelligence, 33 (01), 1499–1506. https://doi.org/10.1609/aaai.v33i01.33011499

Johnson, A. E. W., Pollard, T. J., Berkowitz, S. J., Greenbaum, N. R., Lungren, M. P., Deng, C. Y., & Mark, R. G. (2019). MIMIC-CXR: A large publicly available database of labeled chest radiographs [Preprint]. arXiv. arXiv:1901.07042

Johnson, A. E. W., Pollard, T. J., Berkowitz, S. J., Greenbaum, N. R., Lungren, M. P., Deng, C., & Horng, S. (2020). MIMIC-CXR: A large publicly available database of labeled chest radiographs. Scientific Data, 7 (1), 1–8. https://doi.org/10.1038/s41597-020-0472-y

Liang, G., Zheng, L., & Zhang, Y. (2021). Automatic measurement of cardiothoracic ratio in chest radiographs using deep learning. Computer Methods and Programs in Biomedicine, 200 , 105885. https://doi.org/10.1016/j.cmpb.2020.105885

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & van Ginneken, B. (2019). A survey on deep learning in medical image analysis. Medical Image Analysis, 42 , 60–88. https://doi.org/10.1016/j.media.2017.07.005

Liu, F., Xia, Y., Yang, D., & Hu, Y. (2019). Fully automatic diagnosis of pneumonia from chest X-ray images using deep convolutional neural networks. Pattern Recognition, 100 , 107095. https://doi.org/10.1016/j.patcog.2019.107095

Liu, Y., Zhang, J., Xu, J., & Zhang, Y. (2021). Attention-enhanced image captioning for medical imaging. Computers in Biology and Medicine, 132 , 104296. https://doi.org/10.1016/j.compbiomed.2021.104296

Liu, Y., Zhou, Y., Wang, W., & Li, Y. (2022). Multi-view rib fracture detection in chest radiographs with attention-based deep learning. Medical Physics, 49 (1), 475–485. https://doi.org/10.1002/mp.15333

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., et al. (2021). Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In International Conference on Computer Vision (ICCV) (pp. 10012–10022). https://doi.org/10.1007/978-3-030-85513-4\_96

Liu, Z., Yin, X., Wang, J., Zhang, X., & Li, H. (2021). Generating clinical reports from chest X-rays with hierarchical multi-label classification and reinforcement learning. IEEE Journal of Biomedical and Health Informatics, 25 (6), 1930–1940. https://doi.org/10.1109/JBHI.2020.3040331

Mushtaq, J., Shahbaz, M., & Attique, M. (2020). Early detection of pneumonia using chest X-ray images with deep learning. Computers in Biology and Medicine, 128 , 104036. https://doi.org/10.1016/j.compbiomed.2020.104036

Rajpurkar, P., Irvin, J., Zhu, K., Yang, B., Mehta, H., Duan, T., ... & Ng, A. Y. (2017). CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning [Preprint]. arXiv. arXiv:1711.05225

Rajpurkar, P., Irvin, J., Ball, R. L., Zhu, K., Yang, B., Mehta, H., ... & Ng, A. Y. (2020). Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. PLOS Medicine, 17 (11), e1002686. https://doi.org/10.1371/journal.pmed.1002686

Rahman, T., Chowdhury, M. E. H., Khandakar, A., Islam, K. R., Islam, K. F., Mahbub, Z. B., & Kashem, S. (2020). Transfer Learning with Deep Convolutional Neural Network (CNN) for Pneumonia Detection Using Chest X-ray. Applied Sciences, 10 (9), 3233. https://doi.org/10.3390/app10093233

Suresh, H., Hunt, N., Johnson, A., Celi, L. A., Szolovits, P., & Ghassemi, M. (2020). Clinical intervention prediction and understanding with deep neural networks. In Proceedings of the ACM Conference on Health, Inference, and Learning (pp. 58–67). https://doi.org/10.1145/3377811.3380359

Tan, J. H., Lim, C. P., & Acharya, U. R. (2020). Explainable deep learning models for automated cardiomegaly detection. Neural Computing and Applications, 32 (18), 14489–14501. https://doi.org/10.1007/s00521-020-04823-w

Tiu, E., Talius, E., Zhang, Y., & Beam, A. L. (2020). Clinically aware chest X-ray report generation. In Proceedings of the ACM Conference on Health, Inference, and Learning (p. 110). https://doi.org/10.1145/3377811.3380391

Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3156–3164). https://doi.org/10.1109/CVPR.2015.7298935

Wang, X., Peng, Y., Lu, L., Lu, Z., Bagheri, M., & Summers, R. M. (2019). ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1–10).

Wang, Y., et al. (2021). Adversarial Learning for Robust Medical Image Captioning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2345–2354).

Wang, Y., Zhang, Z., & Zhang, S. (2021). Radiologist-level rib fracture detection in chest X-rays using reinforcement learning. Proceedings of the AAAI Conference on Artificial Intelligence, 35 (1), 688–696. https://doi.org/10.1609/aaai.v35i1.16035

World Health Organization. (2023). Pneumonia . https://www.who.int/news-room/fact-sheets/detail/pneumonia

Zhang, Y., Li, M., Sun, Z., & Yang, Y. (2022). Transformer-based medical image captioning using multimodal features. Computers in Biology and Medicine, 141 , 105122. https://doi.org/10.1016/j.compbiomed.2021.105122

ResearchGate. (n.d.). Pneumonia classification using quaternion deep learning - Scientific Figure on ResearchGate . https://www.researchgate.net/figure/Difference-in-Chest-X-Ray-Images-in-Normal-and-Pneumonia\_fig2\_355191637 (Accessed July 6, 2025)