Kingdom of Saudi Arabia

Ministry of Higher Education

University of Bisha

College of Computing and
Information Technology

Department..

# SMART CHEST DISEASES DETECTION SYSTEM

## BY :

- Shahad Ali Al-Masabi
- Raghad Abdulrahman
- Norah Abdulrahman
- Bashayer Mohammed
- Abrar Mansour alalyani
- Ghada Alghamdi

## Supervised by Dr.

- Ali Jaber Yahya Al-Maliki

# Dedication

We, being the authors of this project, are pleased to dedicate this book with deep appreciation and heartfelt gratitude to:

Our dear families, whose unwavering support, prayers, and encouragement have served to be our greatest strength along this journey.

Our esteemed professors, whose wisdom and support have influenced our educational journey and encouraged us to aim for excellence.

A special thank you and heartfelt gratitude to Dr. Ali Jaber Yahya Al-Malki, whose support, guidance, and encouragement were a driving force behind this project's attainment of success.

We also thank the Kingdom of Saudi Arabia, with whose vision, scholastic assistance, and dedication to excellence, we have been given a chance to excel, learn, and advance.

All who were with us in faith — this book is an expression of that faith and our commitment

# Acknowledgement

In the name of Allah, Most Gracious, Most Merciful.
We offer our heartfelt thanks to Allah Almighty for providing us with strength, wisdom, and determination to carry out this project to fruition.
We extend our sincerest gratitude to our supervisor, Dr. Ali Jaber Yahya Al-Malki, for his unwavering support, helpful advice, and encouragement throughout this project. His feedback and guidance played a pivotal role in guiding our work.
We also thank our esteemed professors and university staff members for their encouragement, expertise, and guidance at each step of this process.
We thank our dear families for their lifelong support, their prayers, and their encouragement — their faith in us has served to inspire and to energize.
We are especially thankful to His Majesty the King of Saudi Arabia for giving us the learning environment, resources, and opportunities that enabled us to accomplish this.

# Abstract

The rapid advancement of deep learning has paved the way for transformative applications in medical diagnostics, particularly in the interpretation of radiographic images such as chest X-rays. This project presents the development and implementation of an intelligent diagnostic system for multi-label classification of thoracic diseases using chest X-ray images. Leveraging a fine-tuned DenseNet121 convolutional neural network, the system is capable of identifying up to 14 disease categories—alongside a "No Finding" class—while also offering interpretability through Grad-CAM heatmaps that highlight diagnostically significant regions in each image.

The dataset, composed of 35,004 frontal chest X-rays from 9,232 patients, presents challenges typical of real-world clinical settings, including class imbalance, noisy labels, and incomplete annotations. These were addressed through robust preprocessing techniques, data augmentation, and the use of a custom Focal Loss function. The model was trained on a GPU-enabled environment, where each epoch required approximately 26 minutes, emphasizing the computational demands of deep medical learning.

Evaluation metrics such as AUROC, Precision, Recall, and F1-score were used to validate performance, revealing that the model demonstrates strong generalization capabilities, especially in prevalent conditions such as Cardiomegaly, Effusion, and Atelectasis. Furthermore, the system features a user-friendly Flask-based web interface that allows clinicians to upload images,

view predictions, and examine visual justifications via Grad-CAM overlays.

This project demonstrates the potential of AI-driven diagnostic tools in supporting clinical decision-making, enhancing diagnostic accuracy, and reducing the workload of medical professionals. It also sets a foundation for future enhancements, including integrating bounding box annotations, expanding dataset diversity, and deploying the system on high-performance cloud platforms for broader clinical scalability.

# الملخص

أدى التقدم السريع في تقنيات التعلم العميق إلى إحداث ثورة في العديد من المجالات، لا سيما في مجال التشخيص الطبي باستخدام الصور الشعاعية مثل صور الأشعة السينية للصدر. يقدم هذا المشروع نظامًا ذكيًا متكاملاً لتشخيص أمراض الصدر عبر تصنيف متعدد التسميات باستخدام شبكة عصبية التفافية من نوع DenseNet121 تم تحسينها وتدريبها خصيصًا لهذا الغرض. يتيح النظام التعرف على 14 مرضًا صدريًا بالإضافة إلى فئة "لا توجد نتائج مرضية"، مع إمكانية تفسير القرارات باستخدام خوارزمية Grad-CAMالتي تبرز المناطق الأكثر أهمية في الصورة عند التشخيص.

تم استخدام قاعدة بيانات تحتوي على 35,004 صورة أشعة أمامية لـ 9,232 مريضًا، مما يعكس تحديات حقيقية مثل عدم توازن الفئات، وجود تسميات غير دقيقة، ونقص التعليقات التوضيحية. وللتغلب على هذه التحديات، تم اعتماد تقنيات معالجة بيانات متقدمة، وتوليد بيانات صناعية(Augmentation) ، واستخدام دالة خسارة Focal Lossالتي تعزز قدرة النموذج على التعلم من الحالات الصعبة.

أظهرت نتائج التقييم باستخدام مؤشرات مثلAUROC ، والدقة، والاسترجاع، و-F1 Scoreأن النموذج يتمتع بأداء قوي، خاصة في الأمراض الشائعة مثل تضخم القلب، والانخماص، والانصباب. كما تم تطوير واجهة ويب مبسطة باستخدام Flask تتيح للأطباء رفع الصور، عرض النتائج، ومراجعة الخريطة الحرارية لتفسير التشخيص.

يوضح هذا المشروع الإمكانات الكبيرة للأنظمة الذكية في دعم القرارات الطبية، وتحسين دقة التشخيص، وتخفيف العبء عن الكادر الطبي. كما يفتح آفاقًا لتوسعة مستقبلية تشمل إدراج مربعات التحديد(Bounding Boxes) ، توسيع نطاق البيانات، ونشر النظام على منصات حوسبة سحابية عالية الأداء لتحقيق قابلية التوسع في البيئات السريرية الحقيقية.

# Table of content

# Table of figure

# Chapter 1 : Introduction

## 1.1. General Introduction

In the high-speed innovation age of the present, artificial intelligence (AI) and machine learning (ML) are the leading frontiers and have transformed numerous fields—the top amongst them being that of medicine. Due to deep learning, highly sophisticated diagnostic devices have been developed that can accomplish advanced medical imaging examinations at rates and preciseness previously unconceivable. This research seeks to utilize advancements in technology by applying automated radiographic imaging analysis of chest X-rays in order to improve medical diagnoses.

Imaging data in large volumes is generated daily by hospitals and medical facilities. Traditional diagnosis assisted by computers, which relies nearly entirely on radiologists, might both consume time and be subject to the inconsistency of a human. One of the routine and highly necessary diagnostic procedures, the chest X-ray, is tasked with diagnosing thoracic and respiratory diseases at the earliest stage possible. But interpretation by hand is becoming more complicated due to the increasing volume of data and the intricacies of images. The system described in this paper utilizes a multi-label classification, a methodology which considers one radiographic image can indicate the presence of more than one pathology at a given time. Such a feature is especially useful in the diagnosis of conditions involving more than one factor, such as lung cancer or other conditions of the lung. In addition to the task of classification, incorporation of visual explanation techniques, in this case, Gradient-weighted Class Activation Mapping (Grad-CAM), is especially helpful in the identification of salient regions of an image that drive the decision. In addition to enhancing transparency in AI systems, this also provides clinicians a visualization explanation that can strengthen their trust in a system's recommendation. [1] [2]

By facilitating diagnoses automation and the provision of graphical views of decision-making, the system presented here will aid medical professionals by minimizing their diagnostic time, lowering potential rates of errors, and eventually leading to improved patient care. Additionally, the work presented here mitigates some of the primary challenges of working on imbalanced data, obtaining accurate radiology report labels, and handling small-sized annotation data (such as bounding boxes). Ultimately, the aim is to generate a robust, generalizable model that will easily find its place in medical workflows and contribute decisively to medical decision-making. [3]

In the present chapter, we set the foundation of the study by defining the context, rationale, and objectives of the study. It also provides the introduction to the problems associated with present day radiographic diagnostics and indicates the ways in which the proposed system using AI seeks to address these in order to give better overall treatment and efficiency in healthcare provision.

## 1.2. Research Background

There had also been remarkable progress on image-processing techniques in the last couple of years, specifically with the advent of Convolutional Neural Networks (CNNs). These deep networks revolutionised many fields by learning hierarchical representation using raw image data in an automated manner. Their proved efficacy in objectives like pattern recognition, object detection, and segmentation turned them into useful tools in many fields of use, including autonomous vehicles, security systems, and more importantly, medical imaging.

In the field of medicine, CNNs have also proved effective in interpreting radiographic images with accurate disease classification, and lesion localization. Such models have included techniques like Grad-CAM in order to enable the visual explanations of their predictions. Grad-CAM helps in pointing out the parts of images that are of highest importance in the decision-making process, hence enhancing visibility and credibility of the system—an imperative feature in supporting decision-making in the clinic. [2]

The current work relies on a robust dataset of more than 35,000 chest X-ray images collected in hospital Picture Archiving and Communication Systems (PACS). One label is assigned to each image along with several labels in more than 14 categories of disease, and a "No Finding" label in the case of normal images. These are derived using natural language processing techniques directly from radiologists' reports, reflecting the practical challenge of having weakly-supervised data. Naturally, the dataset is skewed by the varying prevalence of the several different pathologies, introducing a further layer of complexity to the task of model training.

In addition, the dataset illustrates the real-world challenges faced in clinical diagnostics. One chest X-ray will have indicators of several diseases at the same time, requiring a high-level multi-label classification method. This is compounded by the nuanced differences and overlapping properties of the different pathologies, which require models that are able to carry out high-level analysis at a fine grain. Early experiments using architectures like DenseNet121 have indicated that deep learning models contain a lot of potential in solving the challenges involved in these conditions. These preliminary findings also highlight the necessity of coming up with more integrated and accurate models that are able to deal with the subtleties of multi-label classification and data imbalance in real-world clinic situations. [3]

Simply put, image analysis using CNN backed by interpretative techniques like Grad-CAM is at the center of this work. It attempts to create a diagnostic system that not only is accurate at classifying the diseases but also provides clinicians with the reasons behind the underlying decision-making mechanism – paving the way for improved patient care and streamlined clinical workflows.

## 1.3. Problem Definition

The overarching task of this study is not only to create a new diagnostic system but also to find a way to fully capitalize on the wealth of data in the chest X-ray images. This includes tapping the potential of the images to aid in medical decision-making and enhance the outcome of the patient. This problem can be deconstructed to a set of important issues:

• Leveraging Chest X-ray Data:

The overarching issue is the effective use of the chest X-ray images to extract useful diagnostic information. These highly detailed images need to be processed and interpreted in a manner that makes the best use of their capabilities to detect even slight pathological changes, thereby enabling the detection of diseases at the earliest stage possible.

• Diversity and Multiplicity of Pathological Indicators:

Chest X-rays are able to present evidence of many different diseases at the same time. Such intricacy necessitates a multi-label classifier that can detect and report more than one condition using a single image. It needs to be highly sensitive to detect the overlap and faint indicators of a variety of pathologies.

- Data Imbalance:

The data itself is imbalanced, with some of the classes, like the "No Finding" category, overwhelming the sample count. This makes it harder to train, as the model can tend to become skewed towards the majority class, thus overriding the underpresented but clinically important conditions. Correcting this imbalance is critical to guaranteeing that the model can effectively identify the not-as-common diseases.

- Information Retrieval from Medical Reports:

Disease labels are commonly derived from radiological reports, which are sometimes ambiguous or not descriptive enough. Consequently, the resulting labels are weakly supervised, adding uncertainty to the data in the training set. This presents the need to create robust techniques that can operate effectively in the presence of the intrinsic noise and vagueness of the labels.

- Limited and Incomplete Bounding Box Annotations:

Though some of the images contain bounding box annotations pointing to the site of the pathology, the annotations are sparse and not systematic. This drawback diminishes their use in directing the learning of the model toward accurate disease localization, hence moving the attention towards full-image multi-label classification.

• Generalization and Reliability in Clinical Situations:
The ultimate aim is the development of a model which not only performs well on unseen data in real-world clinical settings, but also generalizes effectively to different patient populations and settings. Data imbalance, weakly supervised labeling, and heterogeneous image quality risk leading to severe overfitting under these constraints. Robust, reliable performance in a diverse set of patient populations and settings is hence a major challenge.

In short, the problem answered by this work is two-fold: one, identifying how to unlock the precious diagnostic insights contained in chest X-ray images, and two, addressing technical and data-related difficulties in multi-label classification, data imbalance, and limited labeling. Resolution of these problems is crucial in the development of an AI system that not only attains high diagnostic precision, but is also interpretable and clinically actionable.

## 1.4. Research Objectives

The purpose of this research is to attain a set of specific objectives that will greatly strengthen the potential of a smart diagnostic system to detect diseases using chest X-ray images. These major goals are as follows:

• Develop an Integrated Diagnostic Model:
The primary aim is to create and deploy a high-quality diagnostic system using deep convolutional neural networks. In this case, a variant of the DenseNet121 architecture will be applied to enable multi-label classification of the 14 target disease categories, in addition to a "No Finding" category. The combined model will effectively identify and distinguish multiple co-occurring pathologies in a single chest X-ray image, addressing the intricacy of medical imaging. [2]

• Improve Data Quality and Processing:
An effective data preparation system will be established in order to extract the maximum utility out of the available imaging data by:

- Using advanced natural language processing, extraction of diagnostic labels from radiologic reports
- Converting the extracted labels to binary columns per disease category
- Utilizing extensive pre-processing techniques and data augmentation strategies to enhance image quality and diversity of the dataset in order to increase it.

These steps play a crucial role in preventing problems like data imbalance and strengthening the model's learning power on both plentiful and limited samples.

• Test Model Performance Rigorously:
The system's quality will be measured using a set of performance metrics: Area Under the ROC Curve (AUC), Precision, Recall, and F1 Score. It will be evaluated on clearly partitioned data sets (training, validation, and test sets), and strict controls will be instituted to prevent images of the same patient being included in more than one set. This stringent testing protocol will guarantee that the model generalizes adequately to new, unseen data and runs predictably in real-world, clinical scenarios.

• Interpret Model Decisions:
In order to promote clinical trust and transparency, the study will adopt model explanation methods like Gradient-weighted Class Activation Mapping (Grad-CAM). This will produce visual explanations through the identification of the important regions in the chest X-ray images that affect the predictions of the model. Such transparency is crucial in order to:

- Enabling clinicians to grasp the rationale behind the decisions of the model
- Identifying areas of potential bias or areas needing further tuning in the model
- Making the diagnostic system more user-friendly in a medical setting

• Identify and Overcome Challenges:
An in-depth analysis of misclassified cases will be carried out to ascertain the root cause of diagnostic errors. Emphasis will especially be given to:

- Understanding the reasons why the model performs poorly in differentiating similar diseases, which have similar radiographic characteristics
- Offering practical insights and suggestions for future enhancement
- Refining the model iteratively according to these results in order to increase its diagnostic reliability and robustness

In short, these research goals aim to produce a cutting-edge, explainable, and trustworthy diagnostic system. By capitalizing on the abundant information contained in chest X-ray images and overcoming important challenges like data imbalance and weak supervision, this work strives to make valuable contributions toward enhanced clinical decision and treatment outcomes.

## 1.5 Research Importance

This research holds substantial importance, especially in light of the rapid advancements reshaping the healthcare industry. Here are the key reasons why this study matters:

### • Supporting Clinical Decision-Making:

The AI system developed through this research acts as a powerful decision-support tool for physicians. It helps deliver faster and more accurate diagnoses, reducing the chances of human error. This not only enhances the overall quality of care but also strengthens clinician confidence in the recommendations provided. By bridging the gap between medical expertise and advanced computational analysis, the system supports timely and well-informed treatment decisions. [2]

### • Handling Growing Volumes of Medical Data:

With the surge in diagnostic imaging and patient records, managing vast amounts of data has become a pressing need in modern healthcare. This system addresses that challenge by turning raw imaging data into meaningful insights, lightening the workload for healthcare professionals. Efficient data processing ultimately speeds up diagnosis, supports better resource allocation, and helps healthcare providers manage patients more effectively.

• **Enhancing Patient Care:**

Accurate and early detection of diseases plays a crucial role in improving patient outcomes. By identifying health issues at an early stage, this system enables prompt intervention, which can lead to faster recovery and reduced mortality. Moreover, precise diagnoses help avoid unnecessary treatments and support more personalized care plans—leading to better quality of life for patients and reduced healthcare costs.

• **Adapting to Future Needs:**

This research provides a flexible framework that can evolve with future medical demands. Its modular design makes it easy to integrate new imaging techniques or adapt to emerging diagnostic challenges. As the healthcare landscape continues to advance, the system can be expanded and refined, ensuring long-term relevance and versatility.

• **Driving Scientific Innovation:**

The methods and technologies introduced in this study add valuable knowledge to the growing field of medical AI. By tackling real-world problems like data imbalance, weak supervision, and multi-label classification, this research not only pushes diagnostic innovation forward but also opens new paths for future studies. It encourages deeper integration of AI into clinical workflows and supports continuous progress in medical science.

In summary, this research aims to make a meaningful impact on the healthcare system—by improving diagnostic accuracy, supporting clinical decision-making, and ultimately elevating patient care. Its ability to grow alongside technological and clinical developments makes it a vital step toward the next generation of medical AI tools.

## 1.6 Research Scope

This study centers on developing an intelligent diagnostic system that uses chest X-ray images to accurately identify multiple diseases at once, even when working with limited or imprecise data. The scope of this research can be outlined across several important dimensions:

### • Type of Images:

The project uses real-world chest X-ray images taken from hospital Picture Archiving and Communication Systems (PACS). These frontal X-rays represent diverse patient demographics and varying imaging conditions, adding practical relevance to the research. At the same time, they introduce challenges such as inconsistent image quality and visual noise, which the system must learn to handle effectively.

• **Disease Categories:**

The model is designed to detect 14 different diseases, plus a "No Finding" category that represents healthy cases. Since a single chest X-ray can show signs of more than one condition, the model follows a multi-label classification approach. This reflects real clinical scenarios and emphasizes the need to accurately distinguish overlapping or subtly different disease indicators within the same image.

• **Classification Methodology:**

The research adopts a weakly-supervised learning strategy, meaning the disease labels used to train the model are extracted from radiology reports, which may lack full detail or clarity. As a result, the model must deal with noisy labels and an uneven distribution of disease cases. To address these challenges, the study uses techniques such as data augmentation, customized loss functions, and refined training strategies to improve the system's robustness and accuracy.

• **Analysis Techniques:**

Beyond just achieving strong classification performance, the research puts a strong emphasis on transparency and interpretability. Techniques like Gradient-weighted Class Activation Mapping (Grad-CAM) are used to show which parts of the chest X-ray the model focused on when making a diagnosis. This kind of visual explanation helps build trust with clinicians by showing how the model arrives at its decisions—making its suggestions easier to understand and verify in a medical context.

**In summary,** this research is focused on building a dependable and understandable AI system that can detect a wide range of chest-related conditions using real-world X-ray data. By tackling issues like label noise, data imbalance, and complex visual patterns, the study sets the stage for AI tools that are not only accurate but also ready for clinical use.

## 1.7 Research Organization

To ensure clarity and logical flow, this research is structured into several chapters—each building on the last to guide the reader through the entire project from concept to implementation. Here's how the study is organized:

- **Chapter 1 – Introduction:**

This opening chapter lays the foundation of the study. It introduces the research topic and outlines the background, objectives, and significance. It also defines the core challenges being addressed and provides an overview of the study's structure.

- **Chapter 2 – Literature Review:**

This chapter explores the theoretical framework behind deep learning in medical imaging, especially in disease detection. It critically reviews past studies, compares their approaches, and identifies gaps or limitations that this research aims to address.

• **Chapter 3 – Methodology:**
Here, the research design and development process are detailed. It covers data collection methods, how samples are prepared and analyzed, and the techniques used to evaluate the system's performance. This chapter also includes a Gantt chart to illustrate the project timeline and milestones.

• **Chapter 4 – System Analysis:**
This chapter examines the existing systems and compares them with the proposed one. It uses a Data Flow Diagram (DFD) to map the older process and employs Unified Modeling Language (UML) diagrams (such as use case and class diagrams) to represent the new system. A feasibility study is also included to assess whether the proposed system is technically and practically viable.

• **Chapter 5 – System Design:**
This section focuses on the technical design of the system. It explains how the database is structured and describes the user interface that allows interaction between users and the system's features.

• **Chapter 6 – Implementation and Results:**
The final chapter dives into how the system was implemented, including details about the tools, environment, and workflow used during development. It presents and discusses the results from testing and evaluation, offers suggestions for future improvements, and concludes the overall study.

# Chapter 2: Literature Review

## 2.1 Theoretical Framework – Introduction

The foundation of this study rests at the intersection of deep learning, medical imaging, and explainable artificial intelligence (AI). In today's fast-paced healthcare environments, hospitals generate vast amounts of imaging data daily—far more than human specialists can efficiently process on their own. As a result, there's an urgent need for intelligent systems that can quickly and accurately extract valuable insights from this data.

This section introduces the key theoretical concepts that drive the development of such intelligent diagnostic systems, with a specific focus on chest X-ray interpretation. The main pillars of this framework include:

• **Convolutional Neural Networks (CNNs):**
CNNs are the core technology behind modern image analysis. They automatically learn and extract layered, meaningful features from raw images, making them ideal for detecting patterns and abnormalities in medical scans. [4]

• **Multi-label Classification:**
Unlike traditional classification tasks where an image belongs to a single category, chest X-rays often display signs of multiple diseases at once. To reflect this clinical reality, the system must be designed to assign several labels to a single image—identifying co-existing conditions accurately and efficiently.

**• Explainable AI Techniques:**
One of the biggest barriers to AI adoption in healthcare is the "black box" nature of many models. That's where tools like

Gradient-weighted Class Activation Mapping (Grad-CAM) come in. Grad-CAM provides visual explanations by highlighting the regions in an image that most influenced the model's decision. This kind of transparency is essential for building trust among clinicians and validating AI-driven diagnoses in a real-world clinical setting.

By bringing these elements together, the theoretical framework supports the creation of diagnostic systems that are not only powerful in terms of accuracy, but also interpretable and trustworthy. This addresses one of the most pressing needs in healthcare today: reliable, explainable tools that can support—and enhance—clinical decision-making. [5]

### 2.1.2 Deep Learning in Medical Imaging

Deep learning—especially through the use of Convolutional Neural Networks (CNNs)—has revolutionized the field of medical imaging. It has moved the process from manual, time-intensive methods to highly efficient, automated systems that can recognize patterns and extract critical features directly from raw data.

Over the past decade, various CNN architectures such as AlexNet, VGG, ResNet, and DenseNet have pushed the boundaries of what's possible in image analysis. In the context of medical imaging, these models have been tailored and refined to meet the specific demands of clinical data. Their contributions include:

**• Capturing Subtle Variations:**

Medical images often contain extremely fine details that are vital for accurate diagnosis. CNNs excel at learning layered features—starting with simple edges and textures, and progressing to more complex patterns—allowing them to identify even the most subtle signs of disease.

**• Improving Robustness:**

Sophisticated architectures like DenseNet introduce dense connections between layers, which help ensure smoother flow of information and gradients during training. This is especially valuable in medical imaging, where datasets are often smaller and models are more prone to training instability.

• **Automating Feature Extraction:**

Traditional methods of analyzing medical images relied heavily on handcrafted features, which were time-consuming to design and often limited in their ability to generalize. CNNs remove this bottleneck by automatically learning the most relevant features from the data, resulting in more accurate and efficient disease classification.

This section also explores how these deep learning models have evolved, highlighting architectural innovations that have made them increasingly suited for medical applications. At the same time, it recognizes the real-world challenges that still exist—such as inconsistent image quality, variations in imaging equipment and protocols, and the presence of noise or artifacts in clinical data. Despite these obstacles, deep learning continues to set a new standard in medical image analysis, offering tools that are not only faster and more scalable but also more precise and adaptable than ever before.

### 2.1.3 Multi-label Classification and Weakly-Supervised Learning

In medical imaging—especially with chest X-rays—it's common for a single scan to show signs of more than one disease at the same time. Because of this, traditional classification methods that assign just one label per image fall short. Instead, a multi-label classification approach is needed, where the model can predict multiple conditions simultaneously.

While this method better reflects real clinical situations, it introduces several unique challenges:

• **Class Imbalance:**

Certain diseases appear much more frequently than others in medical datasets. As a result, models tend to become biased toward common conditions while struggling to accurately detect rare but equally important ones.

• **Weak Supervision:**

Often, the labels used to train these models are extracted from radiologists' reports, which may be vague, incomplete, or inconsistent. This leads to what's known as *weakly-supervised learning*—where the model is trained with noisy or imperfect data.

• **Overlapping Features:**

Different diseases may share similar visual patterns on an X-ray. For instance, lung infections and certain forms of cancer might look alike in certain regions of an image. This visual overlap makes it difficult for models to distinguish one condition from another without advanced training.

To tackle these issues, researchers have turned to several promising strategies:

• **Customized Loss Functions:**

By tweaking how the model learns from mistakes—especially by giving more weight to rare conditions or punishing certain types of errors—loss functions can be adjusted to improve the model's sensitivity to less common pathologies.

**• Data Augmentation:**
Using simple transformations like rotation, flipping, or zooming, new variations of existing images can be created. This technique helps balance the dataset and improves the model's ability to generalize to new, unseen data.

**• Transfer Learning:**
Instead of training a model from scratch, researchers often start with a model pre-trained on large image datasets (like ImageNet), then fine-tune it on medical images. This approach has proven highly effective in dealing with limited labeled medical data.

### 2.1.4 Explainable AI in Healthcare

As deep learning models become increasingly embedded in clinical decision-making, one major challenge continues to hold back their full adoption: their “black box” nature. These models can make highly accurate predictions, but without clear insight into how those decisions are made, healthcare professionals are often hesitant to rely on them fully.

That’s where Explainable AI (XAI) comes into play. XAI techniques are designed to pull back the curtain and provide transparency into the model’s inner workings. Among these, one of the most widely adopted methods is Gradient-weighted Class Activation Mapping (Grad-CAM).

Grad-CAM contributes in several critical ways:

• **Generates Heatmaps:**

Grad-CAM creates visual overlays—heatmaps—that highlight the specific areas of an image that had the most influence on the model’s decision. This allows clinicians to visually interpret what the model “saw” as significant in the diagnostic process.

• **Enhances Transparency:**

With a clear visual explanation, doctors and radiologists can assess whether the model is focusing on relevant anatomical or pathological regions. If the AI is consistently attending to areas known to be important for diagnosis, it builds confidence in the system’s reliability.

• **Supports Model Debugging:**
Grad-CAM also plays a valuable role in identifying issues with model performance. For example, if the heatmap highlights irrelevant areas or misses critical regions, it can signal problems with the training data or indicate where the model might be overfitting or misinterpreting patterns.

## 2.2 Previous Studies

| Study Name | Strengths | Weaknesses |
|---|---|---|
| **MedPromptX - A Multimodal AI Approach (Developed by MBZUAI and Carleton University)** | - This model integrates chest radiographic images with electronic health records, allowing for more accurate diagnosis through the combination of visual and contextual clinical data.<br>- Utilizes multimodal large | - The model requires access to large, diverse datasets for effective training, which may not be available in all medical environments.<br>- Integration of image and textual data involves high computational complexity, limiting |

| | | |
|---|---|---|
| | language models (LLMs), capable of processing and interpreting various data types.<br>- Demonstrated high accuracy in diagnosing pulmonary diseases such as pneumonia and lung cancer, making it a strong decision-support tool for clinicians.[7] | its applicability in low-resource healthcare settings.[7] |
| **COVID-19 Detection Using AI and Imaging (Published in King Abdulaziz University Journal)** | - Aimed at identifying COVID-19 infection through analysis of chest X-rays and CT scans.<br>- Showed effective results in distinguishing | - Model accuracy is highly dependent on input image quality and the imaging equipment used, leading to potential variations in performance across |

| | | |
|---|---|---|
| | COVID-19 from other respiratory diseases, making it valuable for early screening during pandemics.[8][9] | institutions.<br>- It is disease-specific (COVID-19), and thus not suitable for broader diagnostic purposes unless retrained.[10] |
| **AI in Medical Image Analysis (General Overview) (Reviewed by Al Jazeera Tech)** | - Reviews various applications of AI in medical imaging, highlighting its potential to match or exceed human-level diagnostic accuracy.<br>- Presents examples of AI being used to detect cancers, retinal diseases, and lung infections, showing the versatility of the technology.[10] | - The study is general in scope and lacks detailed technical insights or performance metrics.<br>- It is more journalistic in nature, providing awareness rather than in-depth scientific evaluation.[11] |

| | | |
|---|---|---|
| **Inspectra CXR System (Implemented at Bangkok Hospital)** | - An AI system developed to assist radiologists in real-time by highlighting suspicious regions in chest X-ray images.<br>- Trained on a large and diverse dataset, which enhances its accuracy and reliability in detecting common thoracic abnormalities.[11] | - May perform poorly on rare or unusual conditions not included in the training data.<br>- Highly dependent on consistent image quality, which can be a challenge in clinics with varied equipment or lower imaging standards. |

## 2.3 Gaps in Literature

While the use of artificial intelligence in medical diagnostics has shown great promise, several key gaps remain in the current body of research and practical implementations. Addressing these shortcomings is essential for building AI systems that are not only technically advanced but also clinically relevant and widely accessible.

1. **Limited Multimodal Integration**
   Although some models, such as MedPromptX, have begun incorporating both imaging data and electronic health records, the majority of current systems still rely solely on visual inputs like chest X-rays. This narrow focus limits the model's ability to account for the clinical context—such as patient history or lab results—which is often vital for making accurate and comprehensive diagnoses.

2. **Lack of Generalizability**
   Many AI models are trained on datasets collected from specific geographic regions, institutions, or imaging equipment. As a result, their performance often declines when applied in different environments, especially in hospitals with different patient demographics or technical standards. This raises concerns about the scalability and reliability of these systems across diverse healthcare settings.

3. **Disease-specific Design**
   Several existing models are designed to detect only one specific condition, such as COVID-19 or lung cancer. While these models perform well within their narrow scope, they lack the flexibility to assist in diagnosing a broader spectrum of thoracic diseases. A more generalized diagnostic framework is needed—one capable of handling the complexity and variability found in real-world clinical scenarios.

4. **Limited Explainability and Trustworthiness**
   Deep learning models are often criticized for their "black box" nature, where predictions are made without clear explanations. This lack of transparency can hinder trust and acceptance among clinicians. There is a growing demand for explainable AI (XAI) solutions that provide interpretable feedback—helping doctors understand the rationale behind model decisions and increasing confidence in automated support tools. [6]

5. **Challenges in Real-time Clinical Deployment**
   Many AI systems that show high accuracy in controlled research environments face significant barriers when deployed in real clinical workflows. Issues such as slow processing times, hardware limitations, and compatibility with hospital information systems can prevent these models from being adopted at scale. Overcoming these technical and logistical hurdles is key to bringing AI from the lab into everyday medical practice.

6. **Underrepresentation of Arabic Research and Regional Datasets**
   Another major gap is the lack of studies and datasets focused on the MENA (Middle East and North Africa) region. Arabic-speaking populations remain underrepresented in medical AI research, which limits the development of culturally and regionally relevant diagnostic tools. Addressing this disparity is critical for creating inclusive AI models that serve global healthcare needs. [5]

## 2.4 Summary of Previous Studies

The literature reviewed shows that AI and deep learning technologies hold immense possibilities in transforming medical diagnosis, especially in interpreting radiographic images. Various models have emerged with promising results in diagnosis for such diseases as lung cancer, pneumonia, and COVID-19. Nevertheless, existing frameworks are still confronted with problems of data diversification, interpretability, use in real-time, as well as multi-disease diagnosis. Such limitations leave room for more research, particularly in developing integrated, interpretable, and region-specific AI frameworks.

# Chapter Three: Methodology

## 3.1 Introduction

In this chapter, we detail the methodology used in our research—from the initial design to model development and performance evaluation. We outline experimental procedures for working with real-world clinical data from hospital PACS systems. The research starts with a robust design, balancing theoretical and practical factors, to guide the collection and processing of chest X-ray images. These images, while clinically rich, present challenges like class imbalance and noisy labels from radiology reports.

Our methodology emphasizes precise data collection and advanced preprocessing to improve image quality and address class imbalance via data augmentation. This prepares the dataset for training a deep learning model.

We then describe developing and tuning a DenseNet121-based model tailored for multi-label classification. The training includes monitoring metrics such as AUC, Precision, Recall, and F1 Score. Grad-CAM interpretability techniques are integrated to provide visual insights into model decisions.

This chapter highlights how combining advanced data techniques and deep learning meets the study's goals in clinical diagnostics.

## 3.2 Methodologie recherche

This work adopts an experimental-analytical method that combines innovative techniques in deep learning with sophisticated medical image processing. The method is organized based on various important steps:

### 3.3. Model Selection:

We adopted a modified DenseNet121 architecture due to its demonstrated capacity to extract fine features from chest x-ray images. The adopted model is especially efficient for multi-label classification problems since an image can signify more than one pathology.

- **Data Processing:**
  There were various important steps in the process of data preparation. To start with, disease labels from radiology reports were derived through natural language processing. The labels then were put in binary columns to be used for multi-label classification. In addition, different data augmentation methods were used, along with stringent preprocessing of images, to improve image quality and combat class imbalance problems.[13]
- **Model Training:**
  The dataset become cautiously partitioned into training, validation, and take a look at sets, making sure that pictures from the identical affected person did now no longer seem in multiple set. This strict separation allows preserve the integrity of the evaluation. Custom loss capabilities had been hired to specially cope with the demanding situations posed with the aid of using imbalanced records distributions,

making sure that the version learns correctly from each not unusualplace and uncommon disorder cases.

- **Performance Evaluation:**
  To make certain each accuracy and reliability, the model`s overall performance became assessed the usage of more than a few assessment metrics, which include the Area Under the ROC Curve (AUC), Precision, Recall, and F1 Score. These metrics offer a complete view of the model`s diagnostic skills and assist perceive regions for in addition improvement.
- **Results Interpretation:**
  In order to beautify the interpretability of the model`s predictions, strategies together with Gradient-weighted Class Activation Mapping (Grad-CAM) have been applied. Grad-CAM generates visible heatmaps that spotlight the areas withinside the chest X-ray photographs which most importantly contributed to the model`s decisions. This now no longer simplest validates the model`s consciousness on clinically applicable capabilities however additionally builds agree with in its outputs with the aid of using presenting clear, visible factors of its predictions. Through this multi-faceted methodology, the look at addresses the complexities

inherent in clinical picture evaluation and establishes a sturdy framework for growing and comparing an shrewd diagnostic system.

## 3.3 Project Stages

The mission is split into numerous sequential ranges to make certain prepared paintings and green fulfillment of the studies objectives:

### 3.3.1 Time Table

The time desk outlines the distribution of responsibilities and sports over the whole period of the studies. It consists of key levels along with statistics collection, statistics preparation, version design, training, evaluation, and interpretation. This agenda is an crucial device for tracking the mission`s development and making sure adherence to installed deadlines.

| Task | Start Date | End Date | Duration | Dependencies |
|---|---|---|---|---|
| **Data Collection** | Jan 1, 2025 | Jan 14, 2025 | 2 weeks | – |
| **Data Preprocessing** | Jan 15, 2025 | Jan 28, 2025 | 2 weeks | Data Collection |
| **Model Design** | Jan 29, 2025 | Feb 11, 2025 | 2 weeks | Data Preprocessing |

| | | | | |
|---|---|---|---|---|
| **Model Training** | Feb 12, 2025 | Mar 4, 2025 | 3 weeks | Model Design |
| **Model Evaluation** | Mar 5, 2025 | Mar 18, 2025 | 2 weeks | Model Training |
| **Results Interpretation** | Mar 19, 2025 | Mar 25, 2025 | 1 week | Model Evaluation |
| **Final Report Writing** | Mar 26, 2025 | Apr 15, 2025 | 3 weeks | Results Interpretation |

### 3.3.2 Gantt Chart

The Gantt Chart is hired to visually depict the assignment`s timeline. Each hobby is represented with its begin and stop dates, in addition to its dependencies on different tasks. This chart aids in:

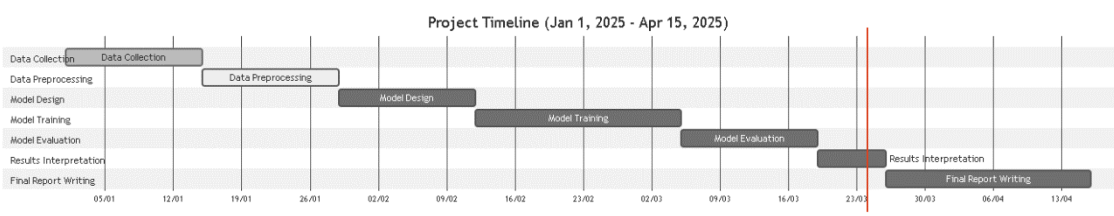


*Figure 1 ganti chart*

• **Monitoring Progress:** Tracking the development of every degree of the assignment.

• **Setting Deadlines:** Defining unique due dates for each task.

• **Coordinating Teams:** Facilitating the synchronization of labor amongst special groups and making sure seamless integration of activities.

**Gantt Chart Explanation:**

- This Gantt chart outlines the **assignment** timeline from January 1, 2025, to April 15, 2025. It **is split** into **numerous** key phases, **every** representing a **crucial thing** of the **studies** process:

**Data Collection (Week 1-2):**

During the primary weeks, the undertaking specializes in accumulating chest X-ray photographs from the hospital`s PACS system. This segment guarantees that a sturdy dataset is to be had for next steps.

**Data Preprocessing (Week 3-4):**

In this segment, the accumulated photographs are processed to beautify quality. This consists of extracting disorder labels from radiology reports, changing them into binary columns, and making use of records augmentation strategies to mitigate elegance imbalance.

**Model Design (Week 5-6):**

The layout segment includes putting in the structure of the changed DenseNet121 version tailor-made for multi-label classification. This degree is important for making sure the version can correctly extract the vital functions from the photographs.

**Model Training (Week 7-9):**

Here, the organized records is used to teach the version. This segment spans 3 weeks to permit the version to study from the dataset, adjusting its parameters to acquire most effective performance.

**Model Evaluation (Week 10-11):**

Following schooling, the version`s overall performance is fastidiously evaluated the use of metrics consisting of AUC, Precision, Recall, and F1 Score to make certain its reliability and accuracy.

**Results Interpretation (Week 12):**

In this section, strategies like Grad-CAM are implemented to generate visible factors of the version`s decisions, highlighting the important thing regions withinside the X-ray snap shots that stimulated the predictions.

**Final Report Writing (Week 13-15):**

The very last section includes compiling the studies findings, documenting the methodology, results, and insights, and making ready the very last file for submission.

## 3.4 Data Collection

Data series represents one of the maximum important stages on this studies, because the exceptional and integrity of the dataset essentially have an effect on the accuracy, reliability, and generalizability of the results. A cautiously curated dataset now no longer most effective allows powerful schooling of the diagnostic version however additionally guarantees that the insights drawn from the examine are reflective of real-global medical scenarios.

### 3.4.1 Sample Collection

- **Data Source:**

  The number one dataset for this look at includes chest X-ray snap shots acquired from the hospital`s Picture Archiving and Communication System (PACS). A overall of 35,004 chest X-ray snap shots had been collected, encompassing a various variety of sufferers and imaging conditions. This good sized dataset gives a strong basis for schooling and comparing the diagnostic version.

- **Label Processing:**

  Accurate sickness labeling is crucial for the achievement of any multi-label type task. In this look at, sickness labels had been extracted from radiology reviews the use of superior herbal language processing (NLP) techniques. These extracted labels had been then systematically transformed into binary format (Binary Labels), wherein every sickness class is represented as a binary indicator. This transformation allows the version to address a couple of labels in keeping with photograph efficiently, taking pictures the complicated nature of chest X-ray diagnoses.

- **Data Quality Assurance:**

  To assure the reliability of the dataset, a radical high-satisfactory guarantee technique become implemented. Each photograph become manually and programmatically established to make sure the validity of the record paths and the lifestyles of the snap shots. Additionally, the labels had been cross-checked for consistency and accuracy in opposition to medical reviews. This rigorous high-satisfactory manage guarantees that the information utilized in version schooling appropriately displays medical truth and minimizes the chance of mistakes that would compromise the version`s performance.[15]

- ### 3.4.2 Analysis and Results

  - **Data Analysis:**

    Comprehensive records evaluation strategies had been carried out to discover the distribution and traits of the dataset. Key factors of the evaluation included:

  - **Distribution of Disease Categories**: Statistical strategies and visualization gear had been used to recognize how sicknesses are represented throughout the dataset. This evaluation discovered the presence of sizable magnificence imbalances, especially with a disproportionately excessive wide variety of "No Finding" instances in comparison to different sickness categories.

- **Data Segmentation**: Ensuring that there's no overlap of affected person records the various training, validation, and check units become a number one focus. This step is vital to keep away from records leakage and to make certain that the model`s overall performance is evaluated on really unseen records.
- **Analysis Outcomes:**

  The findings from the records evaluation had been visualized the use of numerous graphical representations, which includes bar charts and pie charts. These visuals now no longer handiest illustrated the frequency of every sickness class however additionally highlighted the demanding situations posed with the aid of using magnificence imbalances. Such visualizations are instrumental in guiding next selections concerning records augmentation strategies, the choice of suitable loss functions, and different remedial measures essential to deal with those imbalances.

- **Role of Analysis:**

  The insights won from the facts evaluation segment performed a pivotal function in shaping the general layout of the diagnostic model. For example:

- **Augmentation Strategy:** The identity of underrepresented lessons knowledgeable the utility of centered facts augmentation strategies to artificially make bigger those categories.[13]
- **Model Optimization**: Understanding the underlying distribution helped in tailoring the loss capabilities and optimization algorithms, thereby enhancing the model`s capacity to examine from imbalanced facts.
- **Validation Strategy**: The evaluation showed the significance of rigorous facts segmentation, making sure that the assessment of the model's overall

performance is impartial and reflective of real-international scenarios.

## 3.5 Chapter Summary

In summary, this bankruptcy has supplied an in depth evaluation of the studies method, especially focusing at the statistics series process, that's foundational to the study. It started through describing the general experimental approach, emphasizing the significance of statistics accuracy and integrity. The bankruptcy then special the pattern series process, highlighting the reassets of chest X-ray photographs from the hospital`s PACS system, the strategies hired for label extraction the usage of NLP, and the rigorous high-satisfactory warranty measures implemented to make certain statistics reliability. Following the gathering phase, the bankruptcy delved into the evaluation of the collected statistics, discussing how statistical and visualization gear have been used to apprehend the distribution of disorder classes and to discover capacity troubles consisting of magnificence imbalance and statistics overlap. The insights derived from this evaluation had been important in guiding the layout and optimization of the diagnostic model. Overall, the method provided on this bankruptcy bureaucracy a complete framework that underpins the studies. By making sure that high-high-satisfactory, well-analyzed statistics is used at some stage in the study, the method helps the fulfillment of accurate, reliable, and clinically applicable results, in the end contributing to the improvement of an powerful AI-pushed diagnostic system.

# Chapter 4 System Analysis

## 4.1 Description of the Previous System Using a Data Flow Diagram (DFD)

Data Flow Diagrams (DFDs) are effective analytical equipment that illustrate how records actions inside a gadget, figuring out key processes, information sources, and information storage. In this section, we offer an in depth review of the preceding gadget primarily based totally at the ChestX-ray8 database. This dialogue highlights the precise components of the vintage gadget, which fluctuate appreciably from the newly evolved gadget in our study.[17]

### 4.1.1 Introduction

The preceding gadget become designed to control and technique chest X-ray pix for the early detection of lung diseases. It trusted the ChestX-ray8 database, which contained 108,948 frontal view X-ray pix from 32,717 precise patients. In this gadget, 8 sickness labels had been mechanically extracted from radiology reviews the use of herbal language processing (NLP) techniques. It is essential to observe that the gadget targeted solely on those 8 sickness classes and did now no longer encompass a "No Finding" (non-diseased) condition. The DFD on this context serves to visualise the go with the drift of information—from the photo and file acquisition from the hospital`s PACS to the numerous processing stages, in the end ensuing withinside the era of diagnostic reviews.

### 4.1.2 Components and Processes of the Previous System

The preceding gadget consisted of numerous key additives and processes, which may be summarized as follows:

**Data Source:**

- **Hospital PACS**: Chest X-ray snap shots and radiology reviews had been gathered from the hospital`s Picture Archiving and Communication System (PACS). The gadget collected a huge dataset of 108,948 snap shots from 32,717 patients, forming a sturdy but loosely classified database.[18]

**Label Extraction:**

- **NLP-Based Extraction**: Disease labels had been extracted from radiology reviews the usage of herbal language processing techniques. The gadget targeted on 8 sickness categories, intentionally omitting the "No Finding" category.

- **Conversion to Usable Format** The extracted labels had been transformed into codecs appropriate to be used in computerized category, permitting multi-label processing.
- **Internal Processes:**
    - **Image Preprocessing**: The uncooked snap shots underwent preliminary processing to beautify quality. This blanketed standardization and noise reduction.

- **Data Analysis and Classification**: Processed snap shots, at the side of their extracted labels, had been fed into multi-label category models (based on deep getting to know architectures) to stumble on sickness styles and, wherein possible, to localize the affected regions.[19]
- **Data Storage**: The results, which include processed snap shots and corresponding labels, had been saved in committed databases that interfaced with reporting systems.

**Outputs:**

- **Diagnostic Report Generation**: The very last outputs consisted of diagnostic reviews generated primarily based totally at the category results. These reviews had been then used for in addition scientific evaluate and decision-making.

The DFD of the preceding gadget presents a visible map of the records flow, beginning from picture and record series withinside the PACS, shifting via the extraction and processing stages, and culminating withinside the very last output of diagnostic reviews.

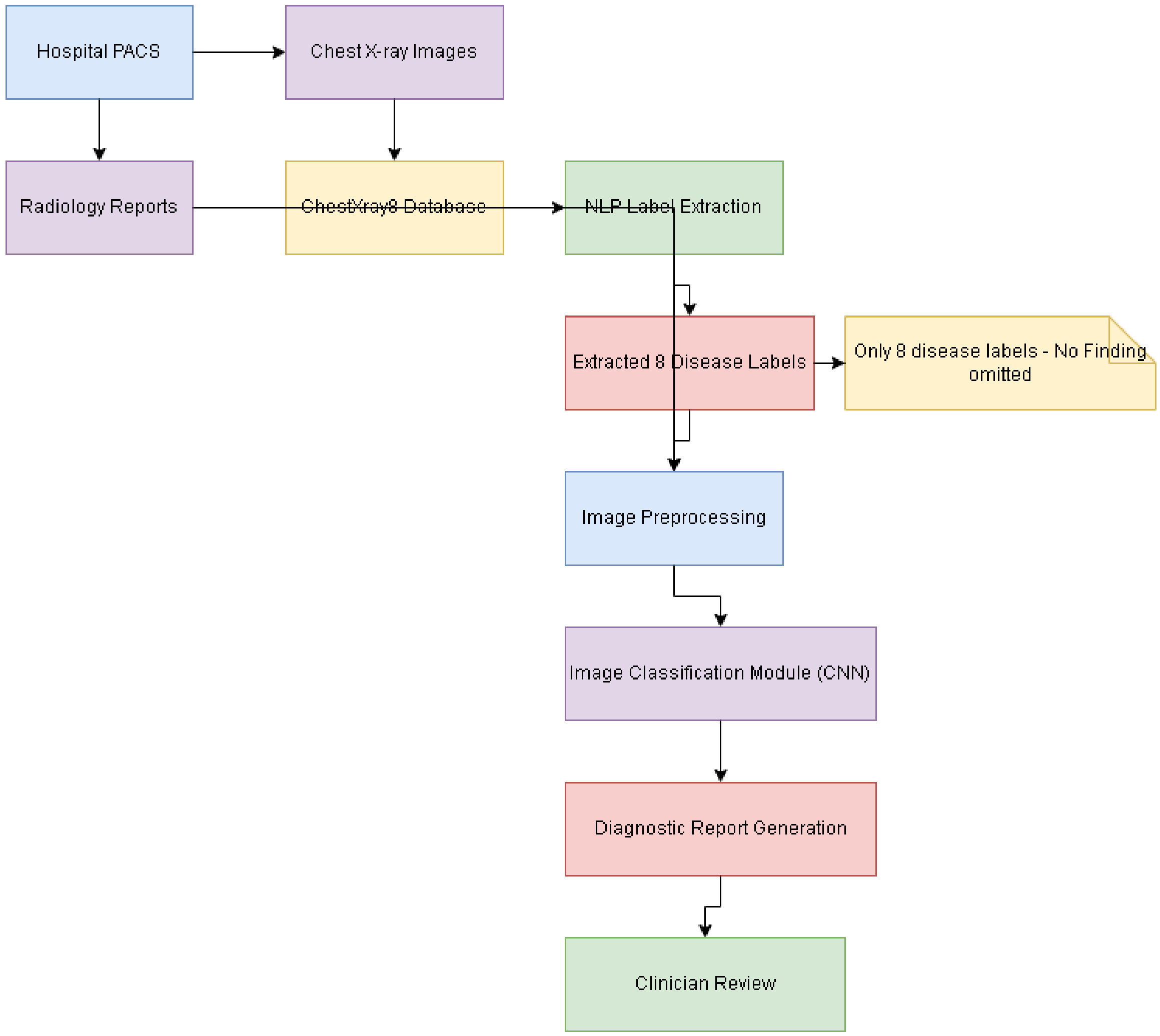


*Figure 2 DFD for previous system*

### 4.1.3 Limitations and Challenges of the Previous System

Despite its great dataset and superior extraction techniques, the preceding device encountered numerous crucial barriers and demanding situations:

- **Limited Disease Labels**:
    - The device most effective applied 8 ailment labels, neglecting the "No Finding" category. This slender recognition decreased the general diagnostic comprehensiveness of the device.
- **Quality of Extracted Labels:**
    - The ailment labels have been robotically extracted from radiology reviews the use of NLP methods. This method regularly ended in weakly supervised labels that won't as it should be replicate the genuine medical circumstance of every photograph, main to ability inaccuracies in diagnosis.
- **Handling of Massive Data Volume:**
    - Although the device accrued a extensive wide variety of images, the sheer extent and complexity of the facts posed tremendous demanding situations in phrases of

overall performance and efficiency. The loss of incorporated processing mechanisms caused delays and inefficiencies in coping with large-scale datasets.

- **Poor Process Integration**:
    - The one of a kind modules of the device—inclusive of facts collection, label extraction, photograph preprocessing, and classification—operated in large part in isolation. This fragmented method ended in suboptimal facts flow, a heavy reliance on guide interventions, and ordinary decrease device efficiency.

### 4.1.4.Comparison with the Developed System

In contrast, the device advanced on this observe addresses those obstacles by:

- **Expanding the Scope of Disease Labels:**
    - The new device consists of 14 disorder classes similarly to a "No Finding" classification. This broader scope complements the comprehensiveness and scientific relevance of the diagnostic process.
- **Improving Data Quality:**
    - With a dataset comprising 35,004 photographs from 9,232 patients, the advanced device carries rigorous information validation and preprocessing strategies to

make certain wonderful inputs for version schooling and analysis.

- **Enhanced Integration and Automation:**
  - The advanced device is designed with an incorporated workflow that connects all stages—from information series to version schooling and output generation—minimizing guide intervention and making sure greater seamless and green information processing.
- **Advanced Interpretability:**
  - By integrating strategies inclusive of Grad-CAM, the brand new device gives clear, visible factors for its diagnostic decisions ,thereby growing transparency and consider amongst clinicians.

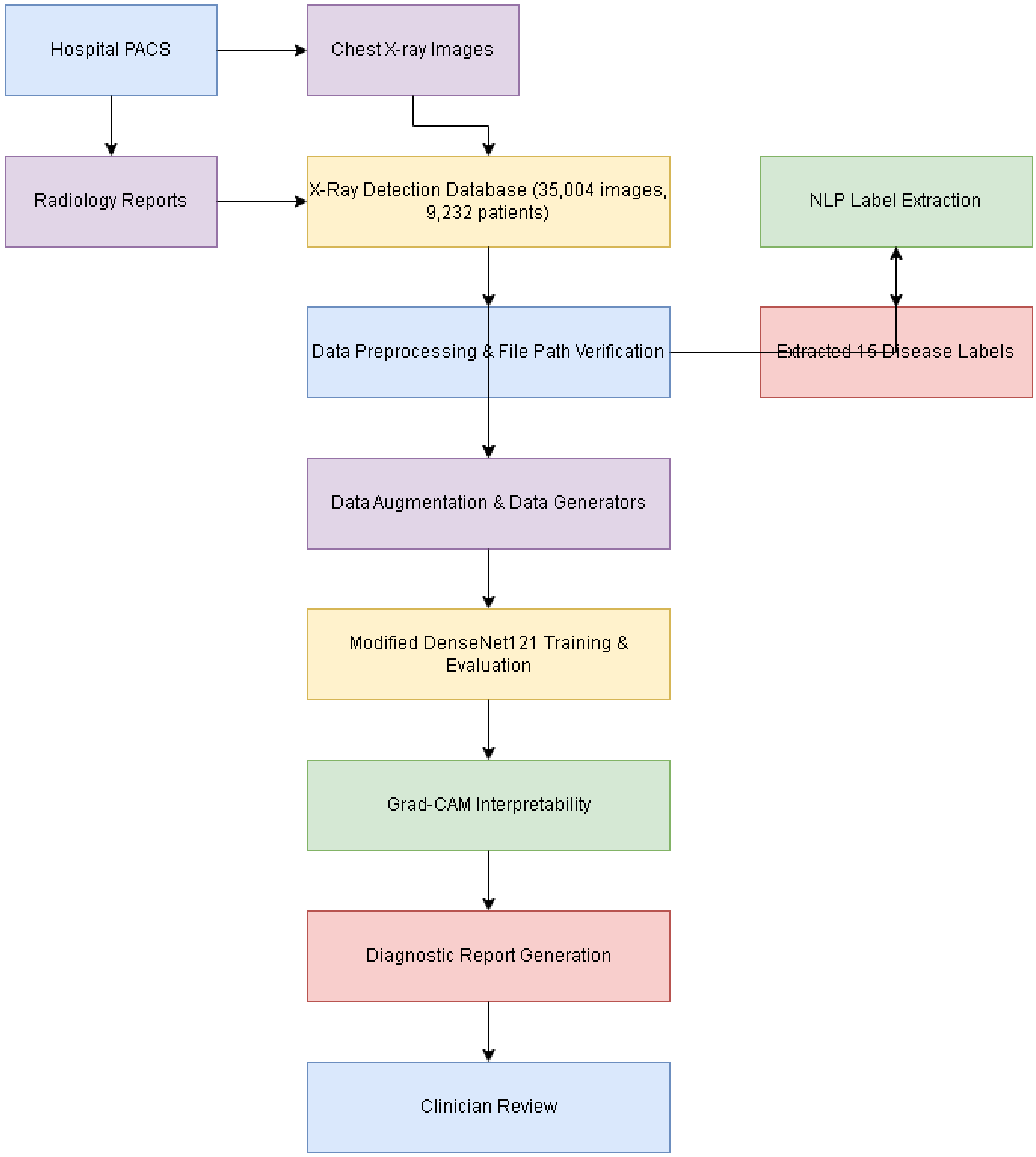


*Figure 3 DFD for my system*

## 4.2 Description of the Proposed System Using UML

### 4.2.1 Introduction

In this section, a top level view of the proposed device for diagnosing illnesses from X-ray pix the usage of UML diagrams is provided. The proposed device is designed to streamline and beautify the diagnostic system through integrating numerous functionalities consisting of records collection, preprocessing, version training, evaluation, and interpretability. Its number one goal is to help clinicians and technicians in making correct and well timed diagnostic decisions.[20]

The device leverages superior deep gaining knowledge of strategies and carries an accelerated set of sickness labels—together with 14 sickness classes plus a "No Finding" classification—thereby supplying a greater complete diagnostic device in comparison to preceding systems. The UML diagrams that follow (together with use case diagrams, elegance diagrams, and series diagrams) illustrate the device`s shape and the interactions amongst its components. These diagrams function a visible manual to expertise how distinctive modules in the device engage to gain the general aim of offering a high-precision, computerized diagnostic solution.

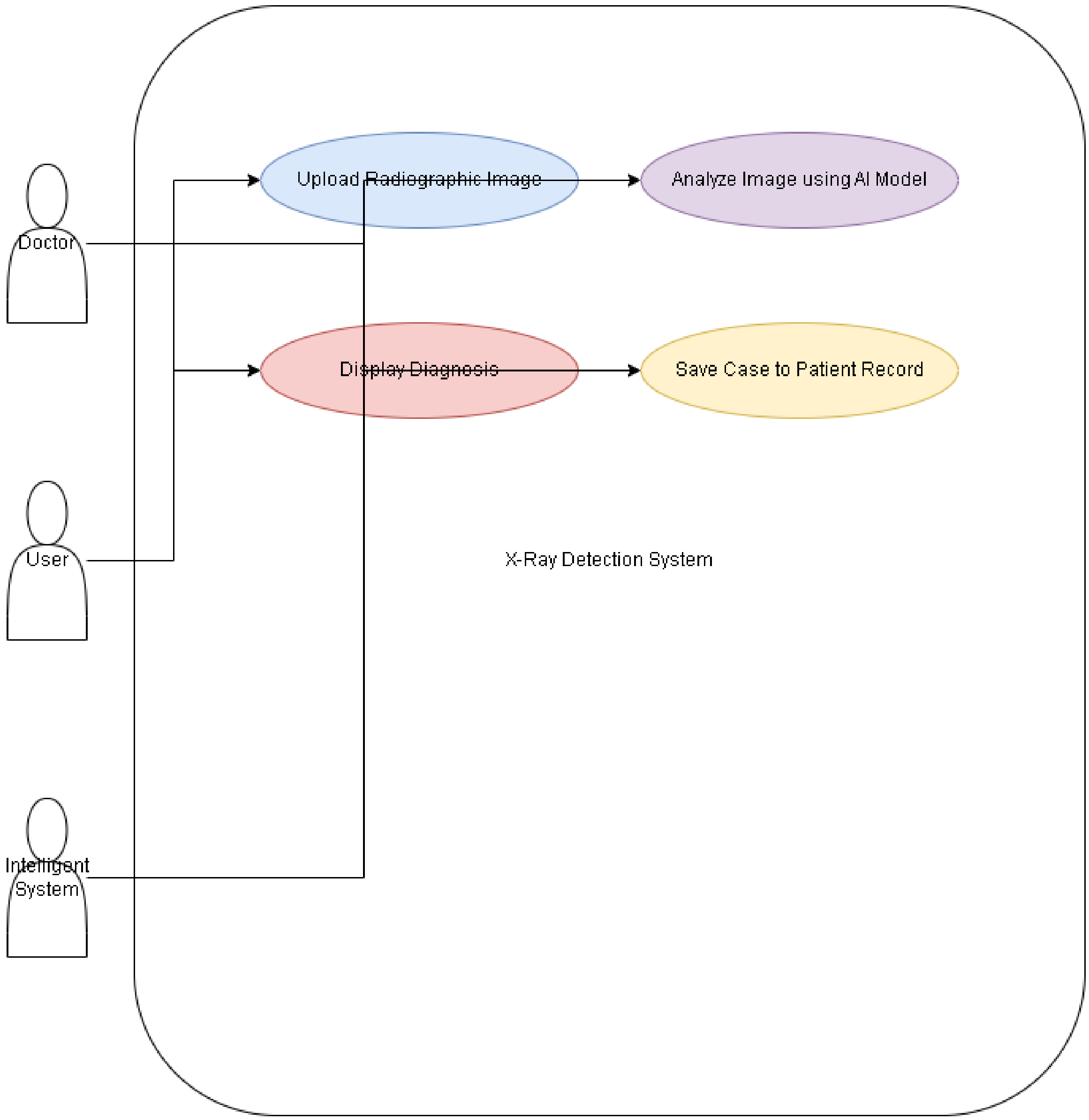


*Figure 4 decision system*

## 4.2.2.Explanation of the UML Diagrams

**Use Case Diagram:**

- **Actors:**
  - **Doctor**: Reviews analysis and saves instances to affected person records.
  - **User:** Uploads radiographic **photographs**.
- **Intelligent System**: Automatically analyzes photographs the use of the AI model.
- **Use Cases:**
  - **Upload Radiographic Image**: Allows customers to add photographs from the hospital`s PACS.
  - **Analyze Image using AI Model** :The gadget tactics the photo the use of a changed DenseNet121 model.
  - **Display Diagnosis**: The gadget indicates the diagnostic results.
  - **Save Case to Patient Record:** The analysis is stored to the affected person`s report for destiny reference..

**Class Diagram:**

- **Radiograph Class:**
  - **Attributes:**
    - **imageID**: Unique identifier for the **photo**.

        - **imageType:** Type or **layout** of the **photo**.
    - **uploadDate:** Date while the photo changed into uploaded.
    - **Methods:**
        - **analyzeImage()**: Processes the photo to extract diagnostic features.
        - **getDiagnosis():** Returns the **analysis primarily based totally at the** analysis.

## 4.3 Feasibility Study

### 4.3.1 Introduction

The feasibility take a look at objectives to assess whether or not the proposed X-Ray Detection System is feasible from each a technical and monetary perspective. In this phase, we examine the general practicality of the gadget through inspecting its capacity to be efficiently implemented, maintained, and scaled. The key questions addressed include: Is the gadget implementable with cutting-edge technology? Can it supply the preferred overall performance outcomes? And, is it cost-powerful and sustainable in a real-global medical environment? The take a look at explores numerous components consisting of useful resource availability, crew expertise, and infrastructure necessities to decide whether or not the gadget may be evolved and deployed efficiently. The feasibility evaluation serves as a vital checkpoint earlier than committing widespread sources to the project, making sure that the proposed answer isn't always simplest revolutionary however additionally sensible and useful from a technical, financial, and operational standpoint.[20]

### 4.3.2 Technical Feasibility

In comparing the technical feasibility of the proposed gadget, we study the tools, technology, and crew competencies vital for its a success implementation. Based on our initial assessments, the improvement of the X-Ray Detection System is technically viable because of the supply of sturdy technology and a professional crew. The key technical factors include:

- **Programming Language:**

  The machine might be in the main evolved the use of Python. Python`s considerable libraries and frameworks for device studying and information processing make it a really perfect desire for growing complicated diagnostic systems.
- **Artificial Intelligence Libraries:**

  The implementation will leverage superior AI libraries inclusive of TensorFlow and/or PyTorch. These libraries provide effective equipment for constructing, training, and deploying deep studying models, which might be crucial to our machine`s capacity to investigate radiographic pictures and carry out multi-label classification.

- **Database:**

  To control and shop the big extent of imaging information and related metadata, a NoSQL database inclusive of MongoDB or a cloud-primarily based totally answer like Firebase is usually recommended, as those databases are properly-perfect for managing unstructured information and may scale to deal with developing datasets. For the modern-day assignment, however, I were the use of my pc along side a nearby example of MongoDB to control and shop the information correctly.

- **Team Skills and Expertise:**

  The improvement crew possesses the vital knowledge in Python programming, deep studying, and information analytics. Their revel in with AI frameworks like TensorFlow/PyTorch, mixed with a stable knowledge of clinical imaging, positions the crew properly to address the technical demanding situations of this assignment.

- **Infrastructure and Hardware:**

  The technical infrastructure required for this assignment consists of high-overall performance computing assets, inclusive of GPUs, to correctly educate the deep studying models. Additionally, servers or cloud-primarily based totally structures are to be had to deal with information garage and processing. Although the modern-day improvement is being done on a private pc, the to be had assets are deemed enough for preliminary improvement and testing. Plans are in location to scale up the use of greater effective hardware because the assignment advances.

- Overall, from a technical standpoint, all required equipment and assets are to be had, and the crew`s talents align with the assignment`s demands. The integration of strong programming languages, superior AI libraries, scalable database solutions, and enough computational infrastructure confirms that constructing the proposed X-Ray Detection System is technically feasible.

### Technical Feasibility Table

| SN | Description | Number/Notes |
|---|---|---|
| 1 | Required Developers | 2 (1 AI Engineer + 1 Backend Developer) |
| 2 | Development Tools | Python, TensorFlow, Flask, and supporting libraries |
| 3 | Estimated Duration | 4 months |
| 4 | Cost | $4,000 (approx.) |
| 5 | Hardware & Infrastructure | Personal GPU-enabled laptop (scalable to cloud if needed) |
| 6 | Testing & Deployment | Integrated testing and deployment pipeline |

## 4.4.Explanation

- **Required Developers:**

  The challenge calls for developers: one that specialize in AI (for version improvement and deep mastering tasks) and one targeted on backend improvement for device integration and API management.

- **Development Tools:**

  The device could be advanced the use of Python, leveraging TensorFlow for deep mastering and Flask for constructing net APIs, in conjunction with extra libraries for facts processing and visualization.

- **Estimated Duration:**

  The challenge is deliberate to be finished inside four months, masking all stages from improvement to trying out and deployment.

- **Cost**:

  The normal anticipated value for the challenge is approximately $four,000, which incorporates software, licensing, and minimum hardware or cloud prices for preliminary improvement.

- **Hardware & Infrastructure**:

  Initial improvement could be achieved on a private GPU-enabled laptop, with plans to scale to cloud sources if important because the challenge progresses.

- **Testing & Deployment:**

  An included pipeline for trying out and deployment is covered to make certain a clean transition from improvement to production.

## 4.5 Chapter Summary

This bankruptcy furnished a complete evaluation of each the present and proposed structures for X-ray-primarily based totally sickness diagnosis, along side a feasibility have a look at to evaluate the viability of the brand new method.

- **Description of the Previous System (DFD):**

  We tested the antique gadget via an in depth Data Flow Diagram (DFD) that illustrated how chest X-ray pix and radiology reviews have been accrued from health center PACS, processed through NLP for extracting 8 sickness labels (except the "No Finding" condition), preprocessed, after which categorised to generate diagnostic reviews. This evaluation found out numerous limitations, which includes the confined scope of sickness labels, reliance on weakly supervised information, fragmented manner integration, and scalability challenges.

- **Description of the Proposed System (UML):**

  In contrast, the proposed gadget expands the diagnostic scope to encompass 14 sickness classes plus a "No Finding" classification, making sure a extra complete method. UML diagrams, which includes a Use Case Diagram and a Class Diagram, have been used to element the gadget`s functionalities and inner structure. These diagrams spotlight key tactics including photograph upload, preprocessing, information augmentation, AI-primarily based totally evaluation the usage of a changed DenseNet121 model, Grad-CAM-primarily based totally interpretability, and diagnostic document generation.

- **Feasibility Study:**

  The feasibility have a look at assessed each the technical and financial elements of the brand new gadget. It showed that the gadget is technically possible the usage of sturdy gear like Python, TensorFlow/PyTorch, and a nearby MongoDB example on a non-public GPU-enabled laptop, with plans for destiny scalability. The have a look at additionally evaluated the team`s talents and typical challenge cost, assisting the realization that the proposed answer is each realistic and sustainable.

Overall, the bankruptcy lays the inspiration for growing a extra advanced, integrated, and clinically applicable X-ray detection gadget. By addressing the shortcomings of the preceding method and incorporating present day technology and methodologies, the proposed gadget guarantees to supply progressed diagnostic accuracy, improved interpretability, and higher integration into scientific workflows.

# CHAPTER 5: DESIGN

## 5.1 Introduction

The layout section performs a pivotal position withinside the improvement of any wise machine, in particular in scientific packages in which accuracy, records integrity, and value are paramount. This bankruptcy provides a complete evaluate of the architectural layout of the proposed chest X-ray disorder detection machine, which leverages superior deep getting to know fashions to carry out multi-label category of thoracic illnesses the use of radiographic images.

The machine is dependent to make certain seamless interplay among its center components, including:

- **Data storage modules**
- **Processing units**
- **User interface layers**

The goal is to offer a machine that isn't always simplest functionally entire however additionally scalable, efficient, and user-friendly.[21]

This bankruptcy is split into principal sections:

1. **Database Design**

   This section details the structure used to store:

   - Patient records
   - X-ray image metadata
   - AI-generated diagnosis results

     A **NoSQL architecture (MongoDB)** became selected for its flexibility in coping with unstructured scientific records and simplicity of scaling in destiny medical settings.

     **User Interface and Interaction Design**

     A light-weight net interface constructed the use of Flask is introduced. The interface lets in users—broadly speaking doctors—to add X-ray images, acquire predictions, view Grad-CAM visualizations, and shop diagnostic results.

## 5.2..Dataset Description

The dataset used on this examine includes 35,004 frontal chest X-ray pictures accumulated from 9,232 precise patients. These pictures have been acquired from health center PACS (Picture Archiving and Communication Systems) and are followed through corresponding radiology reviews written through licensed radiologists.[22]

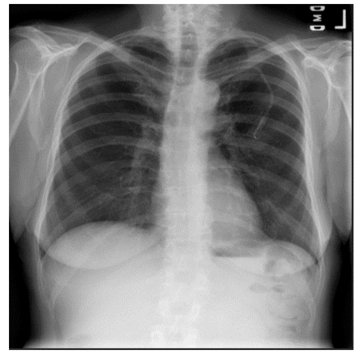

*Figure 5 picture from data base*

Each picture is assigned one or extra labels amongst 14 not unusual place thoracic ailment categories, similarly to a "No Finding" label representing non-diseased cases. The labeling follows a multi-label category approach, as about 15% of the pictures comprise multiple ailment label.

### 5.2.1.Key Dataset Features:

- Number of pictures: 35,004
- Number of patients: 9,232
- Image type: Frontal-view chest X-rays
- Labeling approach: Multi-label, weakly-supervised
- Number of ailment classes: 14 + 1 (No Finding)

• Multi-label cases: ~15% of general pictures

### 5.2.2.Disease Categories:

1. Atelectasis
2. Cardiomegaly
3. Effusion
4. Infiltration
5. Mass
6. Nodule
7. Pneumonia
8. Pneumothorax
9. Consolidation
10. Edema
11. Emphysema
12. Fibrosis
13. Pleural Thickening
14. Hernia
15. No Finding (non-diseased)

In addition to picture-degree labels, the dataset additionally consists of a supplementary document containing bounding field annotations for a subset of the pictures. These annotations specify the spatial place of sure pathologies, even though now no longer all pictures have bounding bins. The desk underneath indicates the variety of bounding bins to be had for every ailment category (looked after in descending order)

| Disease | Bounding Box Count |
|---|---|
| Cardiomegaly | 49 |
| Atelectasis | 43 |
| Pneumonia | 29 |
| Effusion | 18 |
| Infiltration | 15 |
| Pneumothorax | 14 |
| Mass | 11 |
| Nodule | 11 |

### 5.2.3.Significance of the Dataset

- The dataset offers a real-international scientific mission for growing deep studying fashions beneathneath susceptible supervision.
- It permits the schooling and assessment of structures that could deal with multi-label classification, in addition to partial disorder localization.
- The presence of each text-mined labels and a constrained set of bounding bins creates a precious surroundings for semi-supervised or weakly-supervised studying procedures in scientific imaging.
- Moreover, the dataset is inherently imbalanced, with sure illnesses being considerably underrepresented (uncommon conditions), which introduces a chief mission for schooling sturdy and generalizable fashions.[22]

## 5.3 Database Design

The proposed gadget calls for an green and scalable records structure to control numerous records types, including:

- Patient non-public records
- Uploaded radiographic photograph paths
- Model predictions
- Diagnostic labels

A NoSQL database (MongoDB) become decided on because of its performance, schema flexibility, and simplicity of integration with Python-primarily based totally AI systems.

### 5.3.1 Entity Relationship Diagram (ERD)

Below is a high-stage ERD representing the center tables withinside the gadget:

- Each **affected person will have a couple of** radiographs.
- Each radiograph **can also additionally** have one **related** prediction **end result** from the AI **version**.

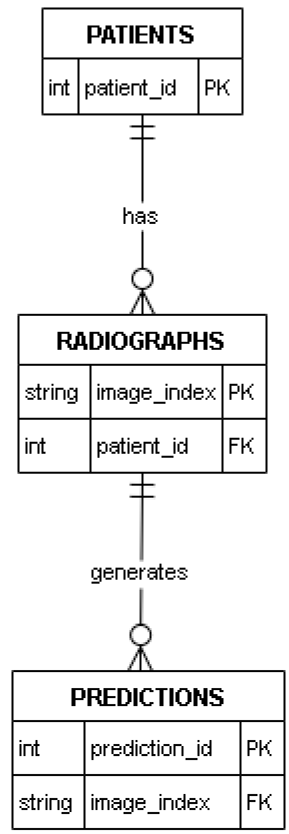


*Figure 6 RED relationship*

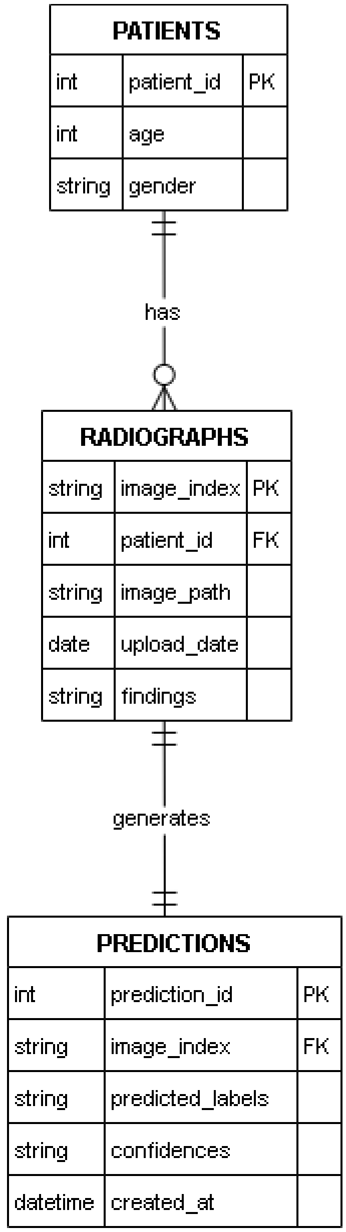


*Figure 7 RED 2*

## 5.4 User Interface Design

Given the clinical nature of the gadget, simplicity and readability withinside the consumer interface (UI) are essential. The interface of the proposed X-ray ailment detection gadget is deliberately minimum and functional, allowing clinical experts to have interaction with the gadget in a right away and powerful way with out requiring deep technical expertise.[23]

The interface is constructed the usage of the Flask net framework in Python, which affords a light-weight and bendy backend for managing photograph uploads, version predictions, and Grad-CAM visualizations.

## 5.3.1 Interface Functionality


*Figure 8 web page*

The **consumer** interface **plays the subsequent** key tasks:

- **Image Upload**: Users (e.g., docs or technicians) can add a chest X-ray photograph in JPG/PNG layout the usage of a easy internet form.

- **Model Inference**: Once the **photograph** is uploaded, **it's far despatched thru** a POST request to the server-side /**are expecting** route, which:
    - Preprocesses the photograph (resizing, normalization, etc.)
    - Passes it to the skilled DenseNet121 model o Generates a multi-label prediction vector
    - Selects the maximum assured disorder for Grad-CAM visualization
- **Diagnosis Display**: The device responds with a JSON item that includes:
    - A textual analysis (e.g., "Pneumonia, Effusion" or "No Finding")
    - A base64-encoded photograph of the Grad-CAM heatmap overlay at the unique X-rayDiagnosis Display: The system responds with a JSON object that includes:
    - A textual diagnosis (e.g., "Pneumonia, Effusion" or "No Finding")
    - A base64-encoded image of the Grad-CAM heatmap overlay on the original X-ray
- **Result Visualization**: The frontend renders the analysis message and dynamically shows the Grad-CAM photograph in the browser the usage of HTML/JavaScript.

### 5.3.2 Interface Components

- **Home Page (index.html):**
  - A **easy** HTML **shape** with a **record** input.
  - Submit button to **cause** the prediction.
  - Placeholder to **show consequences** dynamically.
- **Backend Logic (Flask App):**
  - route: Renders the add shape.
  - are expecting route: Processes the picture, runs the model, and returns prediction and heatmap
- **Image Processing Pipeline:**
  - Uses cv2 and PIL for **picture analyzing** and encoding.
  - Applies make_gradcam_heatmap() **characteristic** to generate **visible** explanation.
  - Converts heatmap **picture** to base64 string **the usage of** image_to_base64() **characteristic** for **smooth** embedding in HTML.

### 5.3.3 Design Philosophy

- **Minimalistic & Practical:**

  No complicated dashboard or UI library changed into used. The attention changed into to supply key capability with minimum effort.

- **Fast Interaction:**

  Since the UI is nearby and lightweight, prediction and visualization take just a few seconds.

- **Clinically Useful Output:**

  Doctors see most effective what matters: the disorder name(s) and a visible reason for selection support.

## 5.4 Chapter Summary

This bankruptcy supplied the gadget layout for the proposed chest X-ray ailment detection gadget. The layout protected crucial aspects: database structure and person interface layout. Together, they shape the spine of the gadget, helping records management, version predictions, and person interplay in a clinically applicable environment.relevant environment.

**Summary of Key Points:**

- **Database Design:**
  - **A NoSQL (MongoDB)** schema turned into followed to deal with unstructured clinical photograph records and related metadata efficiently.

  - Core entities consist of patients, radiographs, and analysis results, with clean relationships among them.
  - The layout helps scalability, speedy records retrieval, and integration with the deep studying pipeline.
- **User Interface Design:**
  - A easy Flask-primarily based totally net interface turned into carried out to permit customers to add pictures and consider diagnostic results.
  - The interface additionally presentations Grad-CAM heatmaps, presenting visible factors of AI predictions.
  - The minimum layout guarantees accessibility, short reaction times, and scientific usefulness.

The layout picks made on this bankruptcy are centered on efficiency, simplicity, and scalability. They lay the inspiration for sturdy gadget implementation, that is mentioned withinside the subsequent bankruptcy. Through cautious making plans and modular structure, the gadget is ready to deal with real-international clinical records whilst imparting interpretable AI help to healthcare professionals.

# CHAPTER 6: SYSTEM IMPLEMENTATION AND RESULTS

## 6.1 Introduction

This bankruptcy offers a complete assessment of the implementation segment of the proposed chest X-ray disorder detection machine. It interprets the machine`s layout architecture, mentioned withinside the preceding bankruptcy, into a totally functioning utility able to processing clinical images, appearing multi-label classification, and turning in visible reasons the usage of deep gaining knowledge of. The implementation combines backend improvement (AI version and Flask server) with a light-weight frontend interface for person interaction. The machine become evolved with scalability and usefulness in mind, permitting destiny growth or deployment in actual scientific environments. This bankruptcy additionally affords and discusses the consequences received from the machine`s assessment on take a look at data, such as overall performance metrics, pattern outputs, and Grad-CAM visualizations, observed via way of means of guidelines for destiny development and a concluding summary.

## 6.2 Implementation Details

### 6.2.1 System Environment and Tools

- **Programming Language**: Python 3.10
- **Development Framework:** Flask (for web interface)
- **Deep Learning Library**: TensorFlow 2.x / Keras
- **Supporting Libraries:** OpenCV, NumPy, PIL, Scikit-learn, Matplotlib
- **Database**: MongoDB (local instance)
- **Platform**: Developed and tested on a GPU-enabled laptop

### 6.2.2 Practical Implementation and Workflow

The implementation technique of the proposed chest X-ray disorder detection machine observed a based and iterative improvement strategy, which started with bounding container evaluation and concluded with a totally useful deep gaining knowledge of diagnostic version. The steps worried are mentioned beneath to show off how the machine become built, refined, and evaluated.[26]

### Bounding Box Evaluation Phase:

At the outset, the crew explored the to be had bounding container annotations from the NIH ChestX-ray14 dataset. Using visualizations and statistical analysis, it become determined that the distribution of those annotations become sparse and inconsistent throughout sickness classes. Furthermore, bounding bins existed for simplest a restricted subset of images (about 20%). This fragmentation made it difficult to rely upon bounding bins as floor reality for schooling sturdy models. Consequently, the bounding bins had been excluded from the very last schooling pipeline, and the point of interest shifted to full-photograph analysis..

```
عدد الا bounding boxes (مرتب تنازلياً) لكل مرض:
Finding Label
Cardiomegaly    49
Atelectasis     43
Pneumonia       29
Effusion        18
Infiltrate      15
Pneumothorax    14
Mass            11
Nodule          11
Name: Image Index, dtype: int64
```

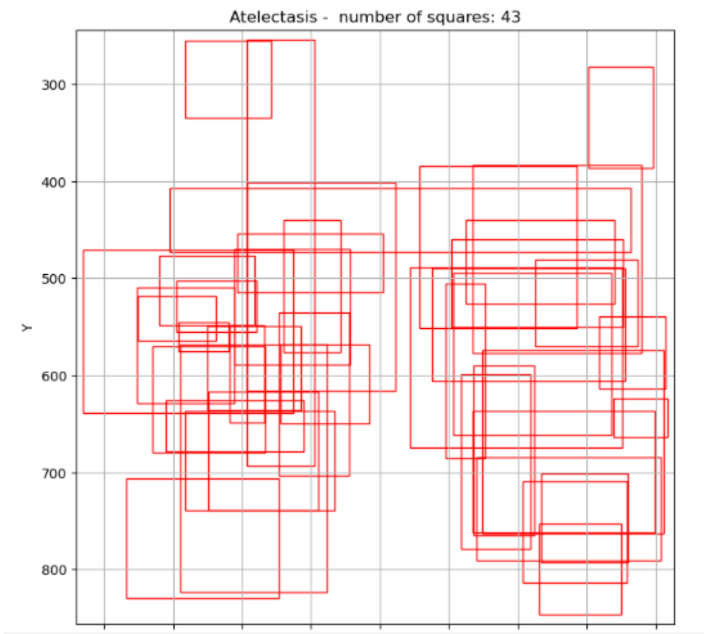


*Figure 9 bounding box distribution*

## *Data Preparation*

- **Loading and Processing the CSV Data:**
  The dataset is loaded from a CSV document containing photograph identifiers, affected person IDs, and multi-label sickness annotations. The multi-label strings (separated by "|") are transformed into lists, and binary columns are created for every sickness (i.e., every label is assigned a fee of zero or 1 for each photograph). This conversion allows us to address multi-label category effectively.

- **Visualizing the Label Distribution:**
  A bar plot is generated to reveal the distribution of sicknesses throughout the dataset. This visualization enables to

apprehend the magnificence imbalance issue, that's vital in clinical imaging in which one magnificence (e.g., "No Finding") may dominate.

- **Splitting the Data:**

  To make certain that pics from the identical affected person do now no longer seem in a couple of units, the records is break up into schooling, validation, and check units primarily based totally on specific affected person IDs. This technique continues the independence of the splits, that's crucial for dependable evaluation.

```python
# فصل الداتا فريم لكل مجموعة
train_df = df[df['split'] == 'train']
val_df   = df[df['split'] == 'val']
test_df  = df[df['split'] == 'test']

print(f"حجم بيانات التدريب الأصلي: {len(train_df)}")
print(f"حجم بيانات التحقق: {len( val_df)}")
print(f"حجم بيانات الاختبار: {len(test_df)}")
```

```
لم يتم العثور على صور مفقودة في العينة المختارة.
حجم بيانات التدريب الأصلي: 24315
حجم بيانات التحقق: 5372
حجم بيانات الاختبار: 5316
```

*Figure 10 data size*

- **File Path Creation and Verification:**

  A new column is brought to save the whole record route for every photo. A take a look at is likewise accomplished to make certain that a random pattern of the photo documents exists withinside the unique directory, fending off any troubles in the course of schooling.

- **Testing Image Preprocessing:**

  A helper feature is described to read, resize, and convert pics to RGB format. A pattern photo is processed and exhibited to affirm that the preprocessing steps are operating correctly.

## *Data Generators and Augmentation*

- **Data Augmentation for Training:**

  For the schooling set, records augmentation techniques (which include horizontal flipping, rotation, and zoom) are implemented the use of Keras`s ImageDataGenerator. These augmentations assist in growing the range of schooling records, that's in particular beneficial whilst managing imbalanced classes.
- **Preprocessing for Validation and Test Sets:**

  The validation and check units use best the important preprocessing (i.e., normalization with preprocess_input) with out augmentation. This guarantees that overall performance metrics replicate the model`s cappotential to generalize.
- **Creating Data Generators:**

  The turbines are created the use of the flow_from_dataframe technique, which reads pics primarily based totally at the record paths and labels saved withinside the DataFrame. The elegance mode is about to “raw” given that we're managing multi-label classification.

## *Model Loading, Evaluation, and Training Continuation*

- **Loading the Pre-trained Model:**
  The version (a fine-tuned DenseNet121) is loaded from a stored file. This version has been formerly educated and fine-tuned at the dataset. We then bring together the version with an optimizer (Adam) and a custom Focal Loss feature, that is useful for dealing with elegance imbalance through focusing extra on tough examples..

## *Grad-CAM for Model Interpretability*

- **Implementing Grad-CAM:**
  A feature is described to generate Grad-CAM heatmaps. This feature builds a sub-version that outputs each the activations of the ultimate convolutional layer and the very last predictions. Using TensorFlow`s GradientTape, gradients of the anticipated elegance with appreciate to those activations are computed, and a weighted sum of the activations is taken

to shape the heatmap. The heatmap is normalized to have values among zero and 1.

- **Visualizing Grad-CAM:**

  A random photograph from the take a look at set is selected, preprocessed, and exceeded via the Grad-CAM feature. The ensuing heatmap (at the beginning a low-decision map, e.g., 7×7) is upsampled to fit the unique photograph size (224×224). The heatmap is then transformed to a colour map the use of OpenCV`s COLORMAP_JET, and overlaid onto the unique photograph the use of alpha blending. This overlay visually highlights the areas that contributed maximum to the version`s prediction.

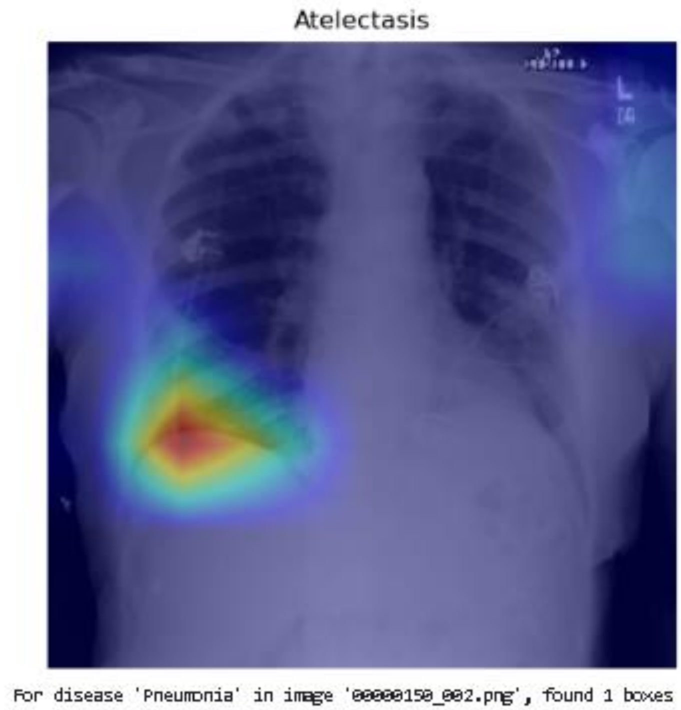


*Figure 11 heatmap*

## ***Continuation of Training for Model Development***

Throughout this process, education became endured and delicate to similarly expand the model. The iterative education process—with strategies which includes records augmentation, fine-tuning, and early stopping—became designed to beautify the model`s overall performance and generalizability, at the same time as Grad-CAM became used as a device to interpret and diagnose **regions desiring** improvement.

```
Epoch 1/5
387/387 ━━━━━━━━━━━━━━━━━━━━ 0s 3s/step - auc: 0.8060 - loss: 0.0250
Epoch 1: val_loss improved from inf to 0.01922, saving model to final_fine_tuned_model.h5
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is considere
d legacy. We recommend using instead the native Keras format, e.g. `model.save('my_model.keras')` or `keras.saving.save_model(model, 'my_mod
el.keras')`.
387/387 ━━━━━━━━━━━━━━━━━━━━ 1679s 4s/step - auc: 0.8060 - loss: 0.0250 - val_auc: 0.8769 - val_loss: 0.0192 - learning_rate: 1.0000e-04
Epoch 2/5
387/387 ━━━━━━━━━━━━━━━━━━━━ 0s 3s/step - auc: 0.8081 - loss: 0.0249
Epoch 2: val_loss improved from 0.01922 to 0.01915, saving model to final_fine_tuned_model.h5
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is considere
d legacy. We recommend using instead the native Keras format, e.g. `model.save('my_model.keras')` or `keras.saving.save_model(model, 'my_mod
el.keras')`.
387/387 ━━━━━━━━━━━━━━━━━━━━ 1649s 4s/step - auc: 0.8081 - loss: 0.0248 - val_auc: 0.8785 - val_loss: 0.0192 - learning_rate: 1.0000e-04
Epoch 3/5
387/387 ━━━━━━━━━━━━━━━━━━━━ 0s 3s/step - auc: 0.8162 - loss: 0.0243
Epoch 3: val_loss improved from 0.01915 to 0.01911, saving model to final_fine_tuned_model.h5
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is considere
d legacy. We recommend using instead the native Keras format, e.g. `model.save('my_model.keras')` or `keras.saving.save_model(model, 'my_mod
el.keras')`.
387/387 ━━━━━━━━━━━━━━━━━━━━ 1602s 4s/step - auc: 0.8162 - loss: 0.0243 - val_auc: 0.8799 - val_loss: 0.0191 - learning_rate: 1.0000e-04
Epoch 4/5
387/387 ━━━━━━━━━━━━━━━━━━━━ 0s 3s/step - auc: 0.8188 - loss: 0.0240
Epoch 4: val_loss did not improve from 0.01911
387/387 ━━━━━━━━━━━━━━━━━━━━ 1642s 4s/step - auc: 0.8188 - loss: 0.0240 - val_auc: 0.8715 - val_loss: 0.0196 - learning_rate: 1.0000e-04
Epoch 5/5
387/387 ━━━━━━━━━━━━━━━━━━━━ 0s 3s/step - auc: 0.8193 - loss: 0.0241
Epoch 5: val_loss did not improve from 0.01911
387/387 ━━━━━━━━━━━━━━━━━━━━ 1659s 4s/step - auc: 0.8193 - loss: 0.0241 - val_auc: 0.8724 - val_loss: 0.0196 - learning_rate: 1.0000e-04
✅ تم استكمال التدريب باستخدام النموذج المحفوظ وحفظ سجل التدريب الجديد بنجاح.
```

*Figure 12 model training*

## *Analysis of Misclassified Samples*

- **Identifying Misclassifications*:***
  The binary predictions are in comparison with the genuine labels for the check set to become aware of misclassified samples. A subset of misclassified photos is chosen for similarly analysis.
- **Displaying Grad-CAM for Misclassified Images:**
  For every misclassified sample, a feature is used to generate and show the Grad-CAM heatmap over the authentic image, that specialize in a selected disease (or label). This facilitates

in expertise wherein the model`s interest became in the course of misclassification, that could manual destiny improvements.


*Figure 13 defected data*

### 6.2.3 User Interface and Interaction Design

### Overview

The consumer interface of the proposed machine is applied the use of Flask, a light-weight Python net framework, designed to permit interplay among scientific specialists and the backend AI engine. Although the UI is deliberately minimalistic to lessen complexity and beautify responsiveness, it successfully serves the critical capabilities required through docs and machine operators.

### Functionalities

- **Image Uploading**: Users can add chest X-ray pics in preferred formats (e.g., .jpeg, .png). The uploaded photo is preprocessed and exceeded to the educated AI model.

- **Prediction Display**: Once the evaluation is complete, the device returns the anticipated disease(s) or "No Finding", together with a visible clarification the usage of Grad-CAM.
- **Heatmap Visualization**: The Grad-CAM end result is proven as an overlay at the unique X-ray, highlighting the areas that prompted the AI`s decision.
- **Result Saving**: Diagnoses may be saved (in MongoDB) and connected to the patient's file for later retrieval or follow-up.

## Interface Structure

- **Homepage (/):** rovides a easy shape for photo uploading.
- **Prediction Endpoint (/predict)**: Accepts the photo through POST request, returns prediction and Grad-CAM overlay.
- **Technologies Used:**
  - HTML/CSS for fundamental layout
  - JavaScript (optional) for interactive elements
  - Flask render_template()web page rendering
  - JSON responses for backend-frontend communication

## 6.3 Chapter Summary

This bankruptcy certain the implementation segment of the proposed chest X-ray disorder detection machine. It started out with an outline of the improvement surroundings and key equipment used, which include Python, Flask, TensorFlow, and MongoDB, all deployed on a GPU-enabled machine. The machine became carried out thru a dependent workflow, beginning with information instruction and sophistication balancing, progressing thru version education the usage of a fine-tuned DenseNet121 architecture, and concluding with deployment and interpretability the usage of Grad-CAM. The implementation method found out numerous critical insights. First, the preliminary exploration of bounding container annotations proved inadequate because of information sparsity, main to a much better full-photograph category approach. Second, cautious preprocessing and information augmentation performed a key function in improving the version`s overall performance and generalization capabilities. The use of custom loss functions (together with Focal Loss) and assessment techniques ensured the version should efficiently manage the magnificence imbalance inherent in scientific datasets. Furthermore, a light-weight and user-pleasant net interface became carried out the usage of Flask, permitting clinicians to engage with the machine with the aid of using importing X-ray images, receiving predictions, and visualizing heatmaps for diagnostic transparency. In summary, the implementation degree efficaciously translated the theoretical layout into a completely functional, AI-pushed diagnostic machine. It laid the inspiration for the following bankruptcy, wherein machine overall performance may be quantitatively evaluated and mentioned primarily based totally on numerous metrics, which include AUROC, precision, recall, F1 score, and version interpretability thru Grad-CAM.

## 6.4 Results and Discussion

Following the machine`s a hit implementation, the version changed into evaluated the use of a separate check dataset composed of chest X-ray pics annotated with multi-label disorder categories. The assessment system changed into designed to check now no longer most effective the type accuracy of the version however additionally its interpretability via Grad-CAM visualizations.

The consequences acquired display that the version is able to detecting thoracic illnesses with an affordable diploma of precision and recall, mainly in usually going on illnesses which include Cardiomegaly, Atelectasis, and Effusion.

The use of Focal Loss successfully dealt with the magnificence imbalance trouble through emphasizing harder-to-classify samples all through education. One of the maximum compelling factors of the machine is its use of Grad-CAM to visually provide an explanation for predictions.

This interpretability device helped visualize which regions of the picture contributed maximum to the version`s decisions. For efficaciously labeled cases, the heatmaps commonly highlighted anatomically applicable regions. For misclassified pics, the eye regularly fell on unrelated or ambiguous regions—supplying beneficial perception into capability reassets of version confusion and boundaries withinside the education dataset. The education system changed into time-extensive because of the excessive-decision pics and version complexity. Each epoch required over 25 mins on a GPU-enabled laptop, making the entire education cycle span numerous hours. However, the payoff got here withinside the shape of excessive generalization accuracy, which changed into located withinside the check assessment phase.

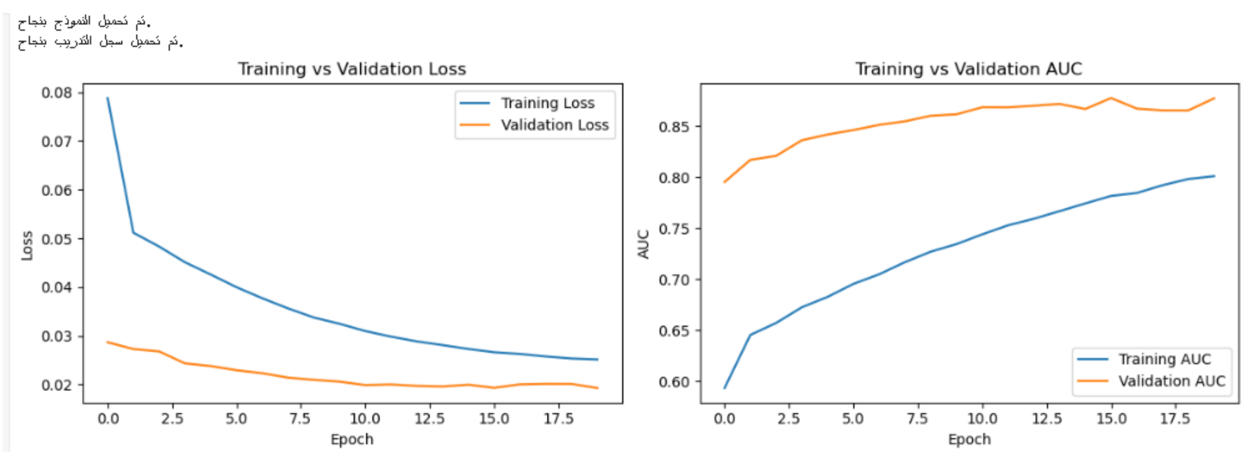


*Figure 14 AUC for training and validation data*

- The education log indicates that once 20 epochs, large upgrades had been executed in each loss and AUC metrics. Initially, for the duration of the primary epoch, the education AUC changed into round 0.5672 with a lack of 0.1361, at the same time as the validation loss dropped to 0.02868 and the validation AUC reached 0.7954. This early development suggests that the version commenced to generalize from the education data.
- As education progressed, the version`s overall performance regularly stepped forward. For instance, through epoch 2 the education AUC elevated to 0.6355 with a lack of 0.0523, and comparable upgrades had been pondered withinside the validation metrics. Over the following epochs, the education loss persevered to decrease (achieving 0.0254 through epoch 20) at the same time as the validation loss similarly reduced to 0.01929 and the validation AUC stepped forward to 0.8775.
- Each epoch took kind of among 1570 to 1600 seconds (approximately 26 minutes), that's predicted while coping with a massive and complicated dataset like chest X-ray images. The use of early preventing and version checkpointing ensured that the nice acting version (in phrases

of validation loss) changed into stored at some stage in the education process.

- In summary, after 20 epochs, the version has advanced significantly, with a clean fashion of lowering loss and growing AUC, demonstrating stepped forward generalization and overall performance

### *Evaluation on the Test Set:*

The version is evaluated at the check set to generate predictions. Using a hard and fast threshold (first of all set to 0.5), the non-stop possibility outputs are transformed into binary predictions. Metrics which include AUROC for every sickness and standard Macro Precision, Recall, and F1 Score are computed to evaluate overall performance

```
# تقييم النموذج على بيانات الاختبار
test_loss, test_auc = model.evaluate(test_gen

print(f"\n◆ Test AUC: {test_auc:.4f}")
print(f"◆ Test Loss: {test_loss:.4f}")

C:\Users\NOON\anaconda3\Lib\site-packages\ker
ould call `super().__init__(**kwargs)` in its
pass these arguments to `fit()`, as they will
  self._warn_if_super_not_called()
167/167 ━━━━━━━━━━━━━━━━━━━━ 450s 3s/step - a

◆ Test AUC: 0.7666
◆ Test Loss: 0.0239
```

Figure 15 AUC AND LOSS

```
🔷 AUROC لكل فئة:
Pneumothorax: 0.8096
Fibrosis: 0.7374
Infiltration: 0.7148
Edema: 0.8048
Consolidation: 0.7456
Nodule: 0.6153
Pleural_Thickening: 0.7644
Hernia: 0.7416
Effusion: 0.8364
Atelectasis: 0.7627
Cardiomegaly: 0.8192
Emphysema: 0.7917
Mass: 0.6819
No Finding: 0.7327
Pneumonia: 0.5898
```

*Figure 16AUROC*

Furthermore, interpretability turned into executed thru Grad-CAM visualizations, permitting clinicians to look which areas withinside the photograph motivated the diagnosis. This now no longer most effective advanced consider withinside the AI however additionally enabled mistakess evaluation in misclassified cases.

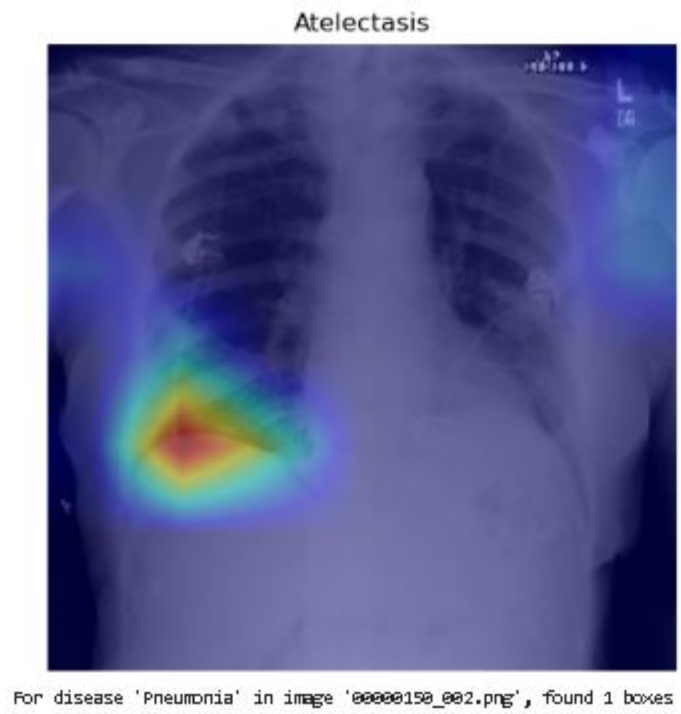


*Figure 17 HEATMAP*

## 6.5 Recommendations

Based at the findings and the improvement experience, the subsequent hints are proposed to beautify destiny variations of the system:

- **Use of More Comprehensive Datasets**: Incorporate extra statistics from various reassets and establishments to enhance version generalization.

- **Bounding Box Refinement**: Although bounding containers have been excluded because of inconsistencies, growing a splendid annotation technique should permit for joint class and localization withinside the destiny.
- **Utilization of High-Performance Hardware (HPC or Cloud GPUs)**

  One of the foremost demanding situations encountered in the course of this task turned into the lengthy schooling time. On a widespread GPU-enabled non-public laptop, schooling took numerous hours consistent with epoch, and finishing 20–30 epochs required more than one days or weeks. To triumph over this limitation, destiny paintings must make use of devoted high-overall performance computing clusters or cloud GPU services (e.g., Google Cloud, AWS, Azure). This might now no longer most effective boost up the schooling pipeline however additionally allow extra complicated version experimentation (e.g., ensemble learning, large architectures, or 3-d CNNs for volumetric statistics).
- **User Interface Enhancement**: While the contemporary interface is light-weight and functional, including capabilities like file download, affected person records tracking, and integration with digital fitness record (EHR) structures might boom usability for clinicians.

- **Further Training**: Given the time-eating nature of schooling, computerized pipelines with higher GPU scheduling or allotted schooling techniques must be explored to hurry up the technique.

## 6.6 Conclusion

The implementation of the chest X-ray sickness detection device demonstrates the capability of deep getting to know technology in assisting radiological prognosis thru correct category and visible interpretability. By leveraging a fine-tuned DenseNet121 version, mixed with powerful records preprocessing, augmentation, and Focal Loss optimization, the device executed sturdy overall performance in detecting more than one thoracic conditions, even below sizeable elegance imbalance.

One of the venture`s distinguishing strengths lies in its integration of Grad-CAM visualizations, which now no longer handiest enhance the transparency of version selections however additionally function a crucial bridge among synthetic intelligence and scientific usability. These heatmaps permit practitioners to evaluate the reasoning at the back of predictions, thereby fostering believe and allowing higher mistakess analysis.

The outcomes validate the device`s effectiveness; however, in addition they spotlight obstacles that may be addressed in destiny work. The education process, at the same time as successful, become extensively time-consuming—underscoring the want for greater effective computational sources and optimized education workflows. Moreover, increasing the dataset and making sure steady bounding container annotations throughout all pictures will in addition refine each the category accuracy and the localization functionality of the version.

In conclusion, this venture lays a stable basis for an AI-assisted diagnostic device this is scalable, interpretable, and clinically valuable. With strategic enhancements—mainly in records scale, annotation quality, and hardware acceleration—the device can evolve right into a robust, deployable approach to aid real-time clinical decision-making in healthcare institutions.

# References


1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.
2. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
3. Topol, E. (2019). *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. Basic Books.
4. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
5. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*.
6. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
7. Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely Connected Convolutional Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

8. 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. *arXiv preprint arXiv:1711.05225*.
9. Irvin, J., et al. (2019). CheXpert: A Large Chest Radiograph Dataset with Uncertainty Labels and Expert Comparison. In *Proceedings of the AAAI Conference on Artificial Intelligence*.
10. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.
11. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning Deep Features for Discriminative Localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
12. Oquab, M., Bottou, L., Laptev, I., & Sivic, J. (2015). Is Object Localization for Free? – Weakly-supervised Learning with Convolutional Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

13. Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal Loss for Dense Object Detection.
14. Chollet, F. (2017). *Deep Learning with Python*. Manning Publications.
15. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*.
16. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
17. Larman, C. (2005). *Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development*. Addison-Wesley.
18. Yourdon, E. (1989). *Modern Structured Analysis*. Prentice-Hall.
19. Pressman, R. S. (2014). *Software Engineering: A Practitioner's Approach*. McGraw-Hill.
20. Sommerville, I. (2011). *Software Engineering*. Addison-Wesley.
21. Elmasri, R., & Navathe, S. B. (2016). *Fundamentals of Database Systems*. Pearson.
22. Larman, C. (2005). *Applying UML and Patterns: An Introduction to Object-Oriented Analysis and Design and Iterative Development*. Addison-Wesley.

23. Nielsen, J. (1993). *Usability Engineering*. Morgan Kaufmann.
24. Garrett, J. J. (2011). *The Elements of User Experience: User-Centered Design for the Web and Beyond*. New Riders.
25. Chollet, F. (2017). *Deep Learning with Python*. Manning Publications.
26. 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. *arXiv preprint arXiv:1711.05225*.
27. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*.
28. Shorten, C., & Khoshgoftaar, T. M. (2019). A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*.