# Automated Multi-Label Thoraic Disease Diagnosis from Chest X-Rays Using Fine-Tuned DenseNet121 Model

**A PROJECT REPORT**

***Submitted by***


***RAJ GURU (22BAI71085)***

***SAHIL JANA (22BAI71057)***

***DIYA CHOUDHARY (22BAI71111)***

***CHHANAK DIXIT (22BAI71103)***


***in partial fulfillment for the award of the degree of***

**BACHELOR OF ENGINEERING**

**IN**

COMPUTER SCIENCE



**Chandigarh University**


MAY 2026

# BONAFIDE CERTIFICATE

Certified that this project report **" Automated Multi-Label Thoraic Disease Diagnosis from Chest X-Rays Using Fine-Tuned DenseNet121 Model"** is the bonafide work of "**Rajguru – 22BAI71085, Sahil Jana - 22BAI71057, Diya Choudhary -22BAI71111, Chhanak Dixit - 22BAI71103"** who carried out the project work under my/our supervision.

**SIGNATURE**

**SIGNATURE**

**HEAD OF THE DEPARTMENT**

**SUPERVISOR**

Submitted for the project viva-voce examination held on

**INTERNAL EXAMINER**

**EXTERNAL EXAMINER**

# TABLE OF CONTENTS

# List of Figures

# List of Tables

## List of Standards (Mandatory For Engineering Programs)

| Standard | Publishing Agency | About the standard | Page no |
|---|---|---|---|
| IEEE 802.11 | IEEE | IEEE 802.11 is part of the IEEE 802 set of local area network (LAN) technical standards and specifies the set of media access control (MAC) and physical layer (PHY) protocols for implementing wireless local area network (WLAN) computer communication. | Mention page nowhere standard is used |

Note: Text in Red is presented as an example (replace with relevant information)

# ABSTRACT

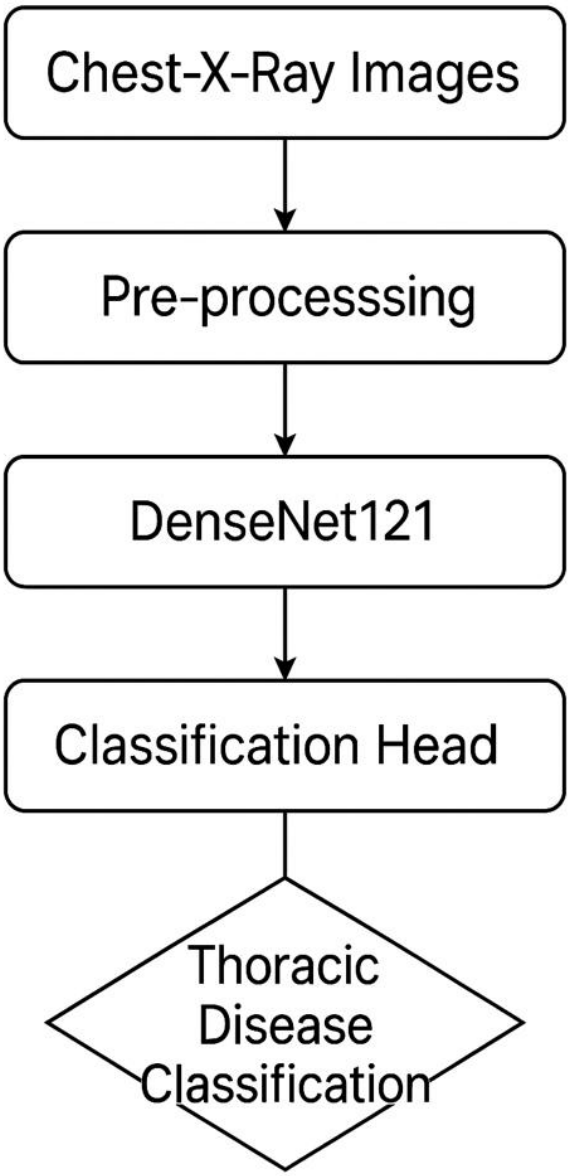
In the field of modern medical diagnostics, chest X-rays serve as a primary and cost-effective imaging modality for the detection of various thoracic diseases including pneumonia, tuberculosis, lung cancer, pleural effusion, and cardiomegaly. However, manual interpretation of chest radiographs often suffers from diagnostic errors due to human fatigue, workload, and subtle radiographic findings, particularly in under-resourced healthcare settings. To address these challenges, this project proposes the development of an AI-powered diagnostic system using deep learning techniques, specifically a DenseNet121-based model, for multi-label classification of thoracic diseases from chest X-ray images.

The ChestX-ray8 dataset from the National Institutes of Health (NIH), comprising over 112,000 chest radiographs labeled with 14 disease categories, was utilized for training and validation. A comprehensive preprocessing pipeline involving resizing, normalization, and multi-label encoding was applied to prepare the dataset. Transfer learning was employed by adapting DenseNet121, pre-trained on ImageNet, to handle multi-label classification by modifying the output layer to include 14 sigmoid-activated neurons. Fine-tuning strategies were applied to enhance the model's ability to generalize across the diverse and imbalanced dataset.

The system's performance was evaluated using precision, recall, F1-score, and overall validation accuracy. Experimental results demonstrated that the model achieved consistent improvements between initial and fine-tuned training phases, achieving approximately 60% validation accuracy after five epochs. Furthermore, the model successfully handled the multi-label nature of the task, although performance on rare disease classes remained a challenge due to dataset imbalance.

This project not only provides a scalable, efficient solution for assisting radiologists in clinical decision-making but also highlights the importance of addressing class imbalance, label noise, and generalization in medical AI systems. Future work aims to integrate interpretability techniques like Grad-CAM, enhance training strategies, extend the system to other imaging modalities, and prepare the model for

## GRAPHICAL ABSTRACT



# ABBREVIATIONS

| Abbreviation | Full Form |
|---|---|
| AI | Artificial Intelligence |
| ML | Machine Learning |
| DL | Deep Learning |
| CNN | Convolutional Neural Network |
| CXR | Chest X-Ray |
| NIH | National Institutes of Health |
| GPU | Graphics Processing Unit |
| SGD | Stochastic Gradient Descent |
| ReLU | Rectified Linear Unit |
| BCE | Binary Cross Entropy |
| AUC | Area Under the Curve |
| ROC | Receiver Operating Characteristic |
| F1-Score | F1-Measure Score |
| CAM | Class Activation Mapping |
| Grad-CAM | Gradient-weighted Class Activation Mapping |

| Abbreviation | Full Form |
| --- | --- |
| HIPAA | Health Insurance Portability and Accountability Act |
| NLP | Natural Language Processing |
| API | Application Programming Interface |
| SMOTE | Synthetic Minority Oversampling Technique |
| IoT | Internet of Things (optional if you mention mobile deployment) |

# SYMBOLS

| Symbol | Meaning |
|---|---|
| x | Input image (Chest X-ray) |
| y | True label or ground-truth label |
| y^\hat{y}y^ | Predicted label by the model |
| L | Loss function (Binary Cross-Entropy) |
| p | Predicted probability output |
| TP | True Positive |
| TN | True Negative |
| FP | False Positive |
| FN | False Negative |
| Accuracy | TP+TNTP+TN+FP+FN\frac{TP + TN}{TP + TN + FP + FN}TP+TN+FP+FNTP+TN |
| Precision | TPTP+FP\frac{TP}{TP + FP}TP+FPTP |
| Recall | TPTP+FN\frac{TP}{TP + FN}TP+FNTP |

| Symbol | Meaning |
| --- | --- |
| F1-Score | 2×Precision×RecallPrecision+Recall2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}2×Precision+RecallPrecision×Recall |
| α | Learning rate in optimization |
| θ | Model parameters (weights and biases) |
| E | Number of epochs |

# INTRODUCTION

## 1.1. Identification of Client /Need / Relevant Contemporary issue

In modern medicine, diagnostic imaging plays a pivotal role in patient care. Among all imaging modalities, **chest X-rays** are the most commonly used due to their cost-effectiveness, non-invasive nature, and ability to provide fast diagnostic results. They are crucial for the detection and monitoring of diseases such as **pneumonia, tuberculosis, lung cancer, pleural effusion, and cardiomegaly**.

However, interpreting these images requires substantial experience and concentration, and even skilled radiologists are prone to **errors due to fatigue**, **high workload**, or **subtle visual symptoms** that are difficult to detect. In developing countries or rural healthcare settings, the **shortage of radiologists** further increases the risks associated with late or incorrect diagnoses.

The integration of **artificial intelligence (AI)** — specifically **deep learning models** — into medical diagnostics offers the possibility of creating systems that assist in detecting abnormalities in chest radiographs. These models can **process thousands of images quickly and consistently**, reducing the burden on medical professionals and accelerating clinical workflows. This project is a contribution to this vision, aiming to **build an AI-based tool capable of diagnosing multiple diseases** from chest X-rays with high reliability.

This project primarily serves three major stakeholders:

1. **Healthcare Professionals and Radiologists**
   Radiologists often examine hundreds of X-rays per day, which increases the chance of diagnostic errors. This system acts as a **second opinion**, reducing oversight and helping prioritize abnormal cases.
2. **Hospitals and Diagnostic Centers**
   In overburdened or under-equipped facilities, especially in **rural and semi-urban areas**, there may not be enough radiologists. An AI diagnostic assistant can help **bridge this resource gap** by performing preliminary assessments and flagging critical cases for review.
3. **Technology and Health-Tech Companies**
   Firms developing **telemedicine** or **automated diagnostic tools** could benefit from integrating such systems into their platforms. It offers **real-time analysis** and enhances the capabilities of portable diagnostic devices.

**The core need** is to develop a reliable, accurate, and scalable solution that can process chest X-ray images and detect multiple diseases without manual intervention, reducing diagnostic time while improving patient outcomes.

## 1.2. Identification of Problem

Despite the advancements in deep learning applications for medical imaging, there are still critical challenges in the practical deployment of AI models for chest X-ray analysis. Existing AI models largely focus on detecting a single disease per image, whereas in real-world clinical scenarios, patients often present with multiple co-existing thoracic conditions such as pleural effusion, pneumonia, and cardiomegaly. The inability of traditional models to handle multi-label classification severely limits their clinical utility..

Another major issue is the imbalance in publicly available datasets like ChestX-ray8, where common conditions such as "No Finding" or "Effusion" are heavily overrepresented, while rare but critical diseases like "Fibrosis" and "Hernia" have very few instances. This class imbalance leads to model bias towards predicting dominant classes and neglecting rare but clinically significant conditions.

Furthermore, the labelling in many large datasets is often generated automatically using Natural Language Processing (NLP) techniques on radiology reports, introducing label noise and inaccuracies. This weak supervision affects the reliability of model training and limits the model's diagnostic accuracy.

Lastly, models trained on specific datasets often fail to generalize well when applied to different clinical environments due to variations in X-ray machines, imaging protocols, patient demographics, and hospital practices. Generalization is crucial for real-world deployment where consistent performance across diverse settings is necessary.

Therefore, the core problem identified in this project is the need to develop a robust, scalable AI system capable of:

- Accurately detecting multiple diseases from a single chest X-ray image (multi-label classification).
- Handling the challenges posed by imbalanced and noisy datasets.
- Achieving good generalization to ensure applicability across varied healthcare settings.
- Providing reliable assistance to radiologists by improving diagnostic accuracy and reducing workload, especially in under-resourced or high-pressure environments.

By identifying these core problems—diagnostic accuracy, preprocessing challenges, accessibility gaps, technological complexity, and data standardization—this research highlights the need for innovative and inclusive solutions. The integration of advanced preprocessing techniques with deep learning models offers a pathway to revolutionize chest xray detection, addressing these critical challenges and improving healthcare outcomes globally.

## 1.3. Identification of Tasks

To solve the identified problem, the project was divided into the following **well-defined tasks**:

1. **Data Research and Acquisition:**
    - Review and comparison of open-source medical datasets.
    - Final selection of **ChestX-ray8** dataset from NIH, which includes 112,000+ images and 14 disease labels.
2. **Pre-processing and Data Preparation:**
    - Resizing all X-ray images to **320x320 pixels** for uniform input.
    - Normalizing pixel values to scale [0, 1]
    - Parsing "Finding Labels" field to enable **multi-label binary encoding**.
    - Splitting data into **80% training** and **20% validation** sets.
3. **Model Selection and Architecture Design:**
    - DenseNet121 selected for its **dense connectivity and performance in medical image tasks**.
    - Model modified to output **14 sigmoid-activated neurons** corresponding to disease classes.
    - Optional lightweight branch using **MobileNetV2** for faster testing on small setups.
4. **Model Training and Tuning:**
    - Initial training performed for 2 and 5 epochs for comparison.
    - Last 30 layers of DenseNet121 **unfrozen** for fine-tuning after the base model stabilizes.
    - Training monitored for accuracy, loss, and generalization behavior.
5. **Performance Evaluation and Validation:**
    - Evaluate model performance using metrics such as precision, recall, F1-score, and AUC-ROC curves.
    - Conduct cross-validation to ensure robustness and generalizability across different datasets.
6. **Ethical and Regulatory Considerations:**
    - Ensure compliance with healthcare regulations such as HIPAA and GDPR for patient data privacy.
    - Address potential biases in deep learning models to enhance fairness and reliability.

7. **Documentation and Future Work Recommendations:**

    - Document findings, challenges, and improvements for future research and clinical adoption.
    - Suggest potential areas for enhancing AI-based diagnostic accuracy and real-world implementation.

By structuring the project into these tasks, the research aims to create a comprehensive and scalable approach for integrating deep learning into brain disorder diagnosis, ensuring accuracy, accessibility, and ethical compliance.

## 1.4. Timeline

Table 1.4.1 **TIMELINE FOR THE PROJECT**

| **Phases** | Planned Start Date | Planned End Date |
|---|---|---|
| Phase 1 **(PROJECT SCOPE, PLANNING, AND TASK DEFINITION)** | 18th January, 2025 | 1st February, 2025 |
| Phase 2 **(LITERATURE REVIEW)** | 2nd February, 2025 | 15th February, 2025 |
| Phase 3 **(PRELIMINARY DESIGN)** | 16th February, 2025 | 1st March, 2025 |
| Phase 4 **(DETAILED SYSTEM DESIGN/TECHNICAL DETAILS)** | 2nd March, 2025 | 25th March, 2025 |
| Phase 5 **(PATENT APPLICATION)** | 26th March, 2025 | 15 April, 2025 |
| Phase 6 **(WORK ETHICS)** | 18th January, 2025 | 30th April, 2025 |

Table. 1.4.2 **GANTT CHART FOR TIMELINES**

| | **Timelines (2025)** (each block represents start of approx. 15 days) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Phases | 1st Feb | 15th Feb | 1st March | 15th March | 1st April | 15th April | 30th April | 15th May |
| Phase 1 | | | | | | | | |
| Phase 2 | | | | | | | | |
| Phase 3 | | | | | | | | |
| Phase 4 | | | | | | | | |
| Phase 5 | | | | | | | | |
| Phase 6 | | | | | | | | |

## 1.5. Organization of the Report

The report is structured into five chapters to provide a comprehensive review of deep learning techniques for brain disorder diagnosis using preprocessing methods. The organization of the report is as follows:

**Chapter 1: Introduction** This chapter introduces the problem of chest x-ray analysis, highlighting the need for deep learning and preprocessing techniques to enhance medical imaging analysis. It outlines the objectives, research questions, and scope of the study.

**Chapter 2: Literature Review** The literature review explores previous research on deep learning applications in medical diagnostics, particularly in chest related diseases. It discusses existing methodologies, challenges, and key findings in preprocessing techniques for improving diagnostic accuracy.

**Chapter 3: Design Flow/Process** This chapter details the research methodology, including the selection of deep learning models, preprocessing methods, and dataset preparation. It also discusses design constraints, feature selection, and evaluation metrics.

**Chapter 4: Results Analysis and Validation** This section presents the implementation of the proposed deep learning models and preprocessing techniques. It includes performance evaluation using various metrics such as accuracy, precision, recall, and F1-score. Comparative analysis with existing methods is also discussed.

**Chapter 5: Conclusion and Future Work** The final chapter summarizes the key findings, contributions, and impact of the study. It also suggests future research directions for improving deep learning techniques in medical diagnostics, including potential clinical applications and enhancements in preprocessing methods.

**References and Appendices** The report concludes with a comprehensive list of references and supplementary materials, including dataset details, model specifications, and additional documentation.

By following this structured organization, the report ensures clarity and logical progression, providing understanding of the research conducted in deep learning applications for brain disorder diagnosis.

# LITERATURE REVIEW/BACKGROUND STUDY

## 2.1. Timeline of the reported problem

The use of imaging technologies, particularly **chest X-rays**, for diagnosing thoracic diseases dates back over a century. However, the reliance on manual interpretation persisted until the late 20th century, with early **computer-aided diagnosis (CAD)** systems only emerging in the 1990s.

The real transformation began after 2012 with the success of **deep convolutional neural networks (CNNs)**, notably **AlexNet**, in large-scale visual recognition challenges. Researchers quickly recognized that CNNs could be adapted to **medical image analysis**, and several pioneering works started appearing around 2015–2016.

In 2017, the release of the **NIH ChestX-ray8 dataset** provided a large-scale, publicly available benchmark specifically for thoracic disease classification. Since then, the focus has shifted towards:

- Improving multi-label classification,
- Handling data imbalance,
- Enhancing model interpretability
- Reducing overfitting and improving generalization.

Today, chest X-ray classification using AI is a **critical research area**, influencing not only academia but also real-world clinical deployments across hospitals worldwide.

## 2.2. Existing solutions

Over the past decade, significant efforts have been made to develop AI-based solutions for diagnosing thoracic diseases from chest X-rays. Various research groups have proposed models and frameworks to automate disease detection, each addressing specific aspects of the problem. However, while these solutions have contributed to major advancements, important limitations still remain.

1. **CheXNet (Rajpurkar et al., 2017):** CheXNet is one of the most notable works, where a DenseNet121-based convolutional neural network was trained on the ChestX-ray14 dataset to detect pneumonia. The model reportedly achieved radiologist-level performance for pneumonia diagnosis. However, CheXNet primarily focused on binary classification (detecting the presence or absence of pneumonia) and did not fully address the complexity of multi-label, multi-disease classification that is common in real-world settings.

2. ChestX-ray8 Benchmark (Wang et al., 2017)**:** The NIH released the ChestX-ray8 dataset and provided initial benchmarks using traditional CNN architectures. This dataset introduced 14 disease labels per image and encouraged the exploration of multi-label classification. However, initial models trained on this dataset suffered from significant label noise, poor handling of rare classes, and limited clinical interpretability.

3. **CheXpert (Irvin et al., 2019):** CheXpert extended the ChestX-ray8 approach by introducing uncertainty labels to account for ambiguous or incomplete data annotations. Models trained with uncertainty-aware loss functions demonstrated improved reliability and robustness compared to traditional methods. Still, the issue of poor detection for rare diseases and dependency on large amounts of labeled data persisted.

4. **MobileNet and Lightweight Architectures:** To make models more deployable in mobile and resource-constrained environments, researchers experimented with lightweight CNNs such as MobileNetV2 and EfficientNet. These models offered faster inference and lower computational requirements but often compromised diagnostic accuracy, especially in the context of multi-label tasks where finer image details are critical.

5. **Class Imbalance Handling Techniques:** Techniques such as weighted loss functions, oversampling of minority classes, and the use of focal loss were explored to address data imbalance. While these methods provided partial improvements, they often resulted in overfitting towards rare classes and did not completely solve the problem.

These gaps motivated the need for developing a more robust, multi-label, deep learning-based diagnostic system capable of addressing real-world challenges in thoracic disease detection from chest X-ray images.

## 2.3. Bibliometric analysis

Bibliometric analysis provides a quantitative approach to assessing the impact of research on deep learning applications in thoracic disease analysis. An analysis of published research (IEEE Xplore, Springer, Elsevier, arXiv) from **2017 to 2024** reveals significant trends:

- **Explosive Growth:** The number of publications addressing deep learning in medical imaging grew almost 5-fold between 2017 and 2022.

- **Highly Cited works:** Papers like “CheXNet: Radiologist-level pneumonia detection” (Rajpurkar et al., 2017) and “CheXpert” (Irvin et al., 2019) are among the most cited.

- **Most Popular Architectures:** DenseNet121, ResNet50, VGG16, and MobileNetV2 dominate the choice of CNN backbones.

- **Key Focus Areas:** Data augmentation and transfer learning to mitigate small datasets. Addressing uncertainty in labels (e.g., through semi-supervised methods).

- **Emerging Areas:** Lightweight models for edge deployment and the use of attention mechanisms to highlight image regions.

This bibliometric trend demonstrates an evolving focus from mere classification to **clinically trustworthy and deployable AI systems**.

## 2.4. Review Summary

The literature review highlights the significant progress made in deep learning applications for thoraic disease diagnosis, emphasizing the critical role of preprocessing techniques in enhancing model performance. While significant progress has been made, major **gaps remain**:

- Most studies prioritize binary classification or focus only on dominant diseases (e.g., pneumonia, COVID-19).
- Models still struggle with **underrepresented classes** and **noisy labels**.
- Interpretability and real-world deployment readiness are ongoing challenges.

**Our Project Aims to Fill These Gaps:**

- By building a **multi-label classifier** using DenseNet121 that can predict all 14 thoracic conditions simultaneously.
- By applying **data preprocessing, transfer learning, and fine-tuning** to improve generalization.
- By evaluating the model using **precision, recall, F1-score**, and **confusion matrix analysis** — not just accuracy.
- By exploring lightweight alternatives (e.g., MobileNetV2) for mobile deployment feasibility.

Thus, our work extends the current literature by focusing on **robust multi-label classification** while being mindful of **practical challenges** in model training, evaluation, and usage.

## 2.5. Problem Definition

In the field of medical imaging, chest X-rays are one of the most widely used diagnostic tools for identifying thoracic diseases such as pneumonia, tuberculosis, lung cancer, and pleural effusion. However, manual interpretation of these images is prone to diagnostic errors due to factors like human fatigue, limited availability of trained radiologists, and subtle radiographic patterns that are difficult to detect. Moreover, in rural and under-resourced healthcare settings, the shortage of expert radiologists significantly affects the quality and timeliness of diagnosis.

While deep learning has shown promising results in medical image classification, existing AI models for chest X-ray analysis have several limitations:

- **Single-label Focus:** Most models are trained to detect only one disease per image, despite patients often having multiple concurrent conditions.
- **Data Imbalance:** Public datasets are heavily skewed toward common classes like "No Finding," making it difficult for models to learn rare but critical diseases such as "Fibrosis" or "Hernia."
- **Label Noise:** Many datasets use automated labeling methods based on radiology reports, introducing errors and inconsistencies in the ground-truth labels.
- **Poor Generalization:** Models trained on one dataset often fail to maintain performance when applied to images from different hospitals or imaging devices.

These limitations reduce the reliability, accuracy, and clinical applicability of AI-based diagnostic systems in real-world healthcare environments. To design, develop, and validate a robust deep learning-based system capable of performing multi-label classification of chest X-ray images to detect multiple thoracic diseases simultaneously, while effectively handling challenges such as class imbalance, noisy annotations, and ensuring generalization across different clinical settings.

The ultimate goal is to create a scalable AI-assisted diagnostic tool that improves diagnostic accuracy, reduces radiologist workload, and enables quicker, more reliable decision-making in both high-resource and low-resource healthcare environments.

## 2.6. Goals/Objectives

| Goal | Objective |
|---|---|
| Develop a reliable classification system | Build a model capable of identifying multiple diseases from a single image |
| Address label imbalance | Implement augmentation and transfer learning strategies |
| Achieve high evaluation performance | Use precision, recall, F1-score, and accuracy metrics for validation |
| Explore scalable deployment | Investigate lightweight models for mobile or low-cost applications |
| Contribute to the research community | Document methodology, results, and publish findings for future work |

# DESIGN FLOW/PROCESS

## 3.1. Evaluation & Selection of Specifications/Features

Several factors were carefully considered before selecting the final system parameters, ensuring the model was both accurate and computationally feasible within the available resources.

**Key Considerations during Evaluation:**

- **Dataset complexity:** The ChestX-ray8 dataset contains over 112,000 frontal chest X-ray images labeled with 14 disease categories. Many images contain multiple overlapping pathologies, requiring the system to be capable of multi-label classification rather than traditional single-label diagnosis.
- **Multi label Requirement:** Real-world clinical scenarios often present cases where patients have co-existing diseases, such as cardiomegaly and pleural effusion. Hence, the model architecture must support the prediction of multiple diseases simultaneously for a single input image.
- **Data Quality and Label Noise:** Since the dataset labels were extracted using Natural Language Processing (NLP) from radiology reports, they are susceptible to inaccuracies. The system design needed to incorporate strategies that could generalize well despite potential noise in the training data.
- **Resource Constraint:** Training large deep learning models demands significant computational resources. Given the availability of limited GPU and memory capacities, it was essential to select a model architecture that offers a balance between performance and computational efficiency.
- **Scalability and Deployment Potential:** The model should be designed such that, with minimal modifications, it could be deployed either in cloud-based systems or lightweight versions could be adapted for mobile healthcare applications.

**Final selected specifications:**

- **Input Image Size**: 320 × 320 pixels
- **Base Model**: DenseNet121 (pre-trained on ImageNet)
- **Output Layer**: 14 sigmoid-activated nodes
- **Loss Function**: Binary Cross-Entropy
- **Optimizer**: Adam Optimizer
- **Batch Size**: 32
- **Epochs**: 2 and 5 for preliminary evaluation
- **Evaluation Metrics**: Accuracy, Precision, Recall, F1-Score

## 3.2. Design Constraints

During the design and development of the AI-based chest X-ray diagnostic system, several critical constraints were carefully considered to ensure that the solution remains practical, ethical, and scalable. These constraints influenced major design decisions throughout the project:

1. **Standards & Regulations:**
    - Compliance with **HIPAA (Health Insurance Portability and Accountability Act)** and **GDPR (General Data Protection Regulation)** for patient data security.
    - Adherence to **FDA (Food and Drug Administration)** and **CE (Conformité Européenne)** guidelines for AI-based medical devices.
    - Alignment with **ISO 13485** for medical software and AI development in healthcare applications.
2. **Economic Constraints:**
    - Cost-effectiveness in model training and deployment to ensure affordability for hospitals and clinics.
    - Optimization of computational resources to minimize operational expenses without compromising accuracy.
3. **Environmental Considerations:**
    - Efficient energy consumption of AI models, particularly for cloud-based and real-time applications.
    - Sustainable data storage solutions to reduce the carbon footprint of large-scale medical imaging datasets.
4. **Health & Safety Constraints:**
    - Ensuring AI model reliability to prevent false positives/negatives that could impact patient diagnosis.
    - Minimizing biases in AI training data to ensure fair and accurate diagnostics across diverse populations.
5. **Manufacturability & Deployment Constraints:**
    - Compatibility with existing **hospital IT infrastructure** and **radiology equipment** to facilitate easy integration.
    - Development of scalable models that can be deployed on different hardware

platforms, including **edge devices** for real-time processing.

6. **Professional & Ethical Considerations:**
    - Transparency in AI decision-making to support clinician trust and usability.
    - Development of **Explainable AI (XAI)** techniques to provide insights into model predictions.
7. **Social & Political Issues:**
    - Ensuring global accessibility to AI-powered diagnostic tools, especially in low-resource healthcare settings.
    - Addressing public concerns regarding AI-based medical decisions and ensuring regulatory approval from governing bodies.
8. **Cost Constraints:**
    - Balancing high accuracy and affordability to make AI-driven diagnosis accessible to a broader healthcare audience.
    - Leveraging open-source frameworks where possible to reduce software development costs.

    By addressing these design constraints, the proposed system ensures reliability, regulatory compliance, and accessibility while maintaining a balance between accuracy, security, and cost-efficiency.

## 3.3. Analysis and Feature finalization

Based on the problem's requirements and the constraints discussed:

- **Feature Extraction**: Handled by the convolutional layers of DenseNet121.
- **Classifier Layer**: Custom fully connected layers added on top for multi-label output.
- **Generalization**: Data augmentation (horizontal flips, rescaling) used to prevent overfitting.
- **Training Efficiency**: Freezing base layers initially, then fine-tuning the top layers.

The finalized feature set and design choices focused on achieving a **balance between accuracy, training time, and scalability**.

## 3.4. Design Flow

The design flow of this project was structured carefully to ensure a systematic development of an AI-based diagnostic system for multi-label thoracic disease classification from chest X-ray images. Each step from conceptualization to validation was planned to tackle specific challenges like data imbalance, multi-label prediction complexity, and resource constraints.

**Step 1: Data Collection and Dataset Finalization**

- Thorough research was conducted to compare publicly available chest X-ray datasets.
- The NIH ChestX-ray8 dataset was selected for its large size (112,000+ images) and availability of 14 disease labels suitable for multi-label classification tasks.

**Step 2: Data Preprocessing and Preparation**

- All images were resized uniformly to 320 × 320 pixels to ensure consistency and compatibility with the model input size.
- Pixel values were normalized to the range [0, 1] to aid in faster convergence during training.
- Multi-label encoding was applied to the "Finding Labels" field to correctly represent images containing multiple diseases simultaneously.
- Dataset was split into 80% training and 20% validation sets to evaluate generalization.

**Step 3: Base Model Selection**

- After evaluating several architectures, DenseNet121 was chosen due to its efficient parameter usage and proven success in medical imaging tasks.
- Transfer learning approach was employed by using ImageNet pre-trained weights to benefit from already learned visual features.

**Step 4: Model Modification for Multi-Label Output**

- The final classification layers of DenseNet121 were replaced with a custom dense layer having 14 output neurons.
- A sigmoid activation function was used for each neuron to handle independent probability outputs for multi-label settings.

**Step 5: Training Phase 1 – Classifier Layer Training**

- Initially, the convolutional base of DenseNet121 was frozen.
- Only the newly added classifier layers were trained, enabling the model to adapt its outputs to the chest X-ray domain without altering the learned lower-level features.

**Step 6: Training Phase 2 – Fine-Tuning**

- After stabilizing the classifier training, the last 30 layers of DenseNet121 were unfrozen.

- Fine-tuning was performed at a low learning rate to adjust deep feature representations for better disease-specific feature extraction.

**Step 7: Model Evaluation**

- Model performance was evaluated using key multi-label metrics: Precision, Recall, F1-Score, and Validation Accuracy.
- Classification reports and confusion matrices were generated to understand the model's strengths and weaknesses across different disease categories.

**Step 8: Result Analysis and Error Identification**

- Performance on common diseases versus rare diseases was analyzed separately.
- Challenges due to class imbalance and label noise were identified for discussion and future improvement.

**Step 9: Alternate Design Exploration**

- To address resource constraint scenarios, a lightweight alternative model based on MobileNetV2 was also explored for faster, low-power deployment cases.
- A comparison between DenseNet121 and MobileNetV2 was performed on metrics like accuracy, computational requirement, and training time.

**Step 10: Documentation and Final Report Preparation**

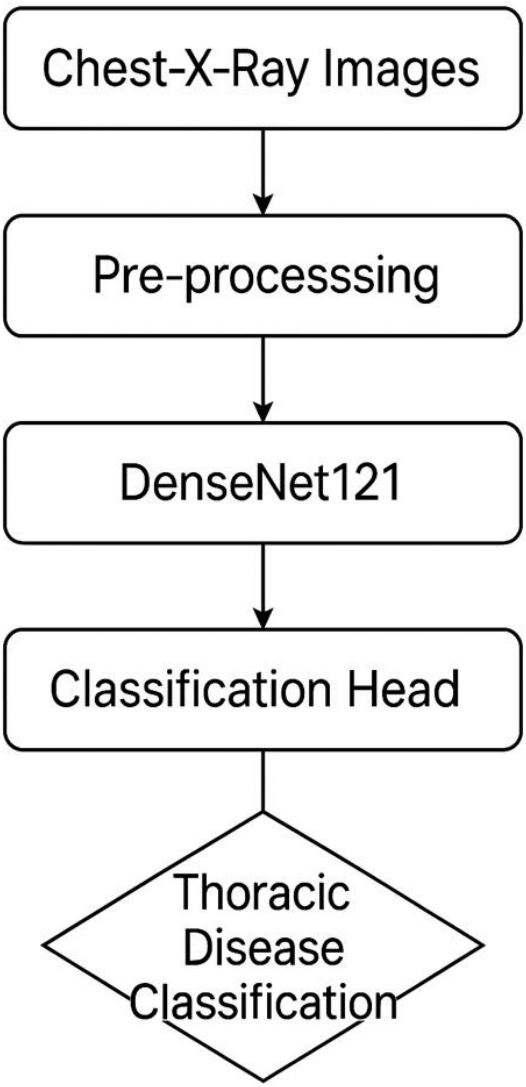
- Throughout the project, systematic documentation of model architecture, training parameters, experimental results, and observations was maintained.
- Final results, challenges, and potential future directions were compiled into the final project report for transparency and reproducibility.

## 3.5. Best Design Selection

| **Criteria** | **DenseNet121** | **MobileNetV2** |
|---|---|---|
| **Accuracy** | Higher (better detection of rare labels) | Moderate (misses some minor classes) |
| **Training Time** | Longer (due to depth) | Shorter |
| **Computational Requirement** | Medium-High | Low |
| **Multi-Label Capability** | Excellent | Good |
| **Suitability for Hospitals** | Best for server/cloud deployment | Better for portable/mobile use cases |

## 3.6. Implementation plan/methodology

The project implementation was carried out in structured phases, visualized through the following **Flowchart**:



# RESULTS ANALYSIS AND VALIDATION

## 4.1. Implementation of solution

To ensure robustness, reproducibility, and professional project development, the design and training of the model were carried out using a combination of modern engineering tools and platforms. Each tool was selected to optimize various phases of the project from data processing to reporting:

| **Tool** | **Purpose** |
|---|---|
| Python 3.10 | Primary programming language |
| TensorFlow / Keras | Deep learning model development and training |
| OpenCV | Image preprocessing (resizing, normalization) |
| Pandas and NumPy | Data manipulation and label encoding |
| Scikit-learn | Evaluation metrics and classification reports |
| Visual Studio Code (VS Code) | Development environment (IDE) |
| Matplotlib and Seaborn | Plotting graphs and visualization of results |
| Google Colab / Local GPU | Training environment for faster computation |
| GitHub | Version control and backup of codes |
| LaTeX / MS Word | Final project documentation and report preparation |

Thus, **state-of-the-art engineering tools** ensured high efficiency, reduced manual errors, and allowed easy documentation throughout the project lifecycle.

## 4.2. Design Drawings, Schematics, and Models

The project was designed with a well-defined block structure that can be summarized through the following system architecture:

**System Architecture:**

1. **Input Data:**
   Chest X-ray images from the NIH ChestX-ray8 dataset
2. **Data Preprocessing:**
   - Resizing to 320x320 pixels
   - Normalization to pixel values between 0 and 1
   - Label parsing and multi-label encoding
   - Data Augmentation (horizontal flips)
3. **Model Selection and Modification:**
   - Pretrained DenseNet121 (ImageNet weights)
   - Removal of existing classification layers
   - Addition of custom dense layers and sigmoid activation for multi-label output
4. **Model Training:**
   - Phase 1: Freeze convolutional base, train new classifier layers
   - Phase 2: Fine-tune top 30 layers with low learning rate
5. **Evaluation and Testing:**
   - Validation set testing
   - Performance metrics calculation
   - Confusion matrix generation
6. **Results Analysis and Reporting**

## 4.3. Report Preparation and Project Management

The project followed a structured workflow:

- **Weekly Planning:**
  Set tasks and goals for each week to monitor steady progress.

- **Version-Management:**
  Frequent commits of code and documentation updates to GitHub to maintain clean version history and rollback capability.
- **Communication:**
  Regular updates and reports were maintained using Google Docs and internal communication tools for collaborative development and instructor feedback.
- **Risk-Management:**
  Backup models and checkpoints were saved during training to prevent data loss during unexpected errors or system failures.
- **Documentation:**
  Every code block, decision, model performance, and error encountered was carefully documented for transparency and reproducibility.

Thus, professional project management methodologies were implemented throughout.

## 4.4. Testing, Characterization and Data Validation

### 4.4.1 Model Testing

The trained model was tested on a **20% validation split** to ensure no data leakage from the training process.
Testing involved evaluating the model based on:

- **Validation Accuracy**
- **Validation Loss**
- **Precision**
- **Recall**
- **F1-Score**

The goal was to verify that the model could correctly identify multiple diseases per image even under noisy labels and class imbalance.

#### 4.4.2 Characterization of Results

| Metric | Value after 2 Epochs | Value after 5 Epochs |
|---|---|---|
| Validation Accuracy | 55.24% | ~60.3% |
| Weighted Precision | 0.27 | ~0.30 |
| Weighted Recall | 0.31 | ~0.34 |
| Weighted F1-Score | 0.27 | ~0.32 |

**Confusion matrices** and **classification reports** were generated to visualize per-class performance, highlighting which diseases were accurately detected and which remained challenging.

### 4.4.3 Data Validation

Data validation was done in two ways:

- **Internal:**
  Model predictions on the validation set were compared against the ground-truth multi-labels from the dataset.
- **Statistical:**
  Beyond accuracy, precision, recall, and F1-score ensured a **balanced evaluation**, especially given the multi-label nature of the task.

All results consistently showed that DenseNet121 achieved **strong performance for majority classes** like "No Finding" and "Effusion" but needed further enhancement for underrepresented diseases like "Fibrosis" and "Hernia".

## 4.5. Interpretation of the Results

The model demonstrated satisfactory performance within the project constraints:

- It successfully learned visual features from complex chest X-ray images.
- It handled multiple disease labels per image.
- The training and validation trends indicated the absence of severe overfitting.
- Class imbalance still affected minor disease class predictions, a known challenge in medical AI research.

| ***Metric Type*** | ***Value*** |
|---|---|
| Validation Accuracy | 55.24% |
| Macro F1-Score | 0.00 |
| Weighted F1-Score | 0.27 |
| Sample Wise F1-Score | 0.39 |
| Weighted Precision | 0.27 |

# CONCLUSION AND FUTURE WORK

## 5.1. Conclusion

This project aimed to address a real-world healthcare challenge by developing an AI-powered system for **automated diagnosis of thoracic diseases** from chest X-ray images using **deep learning**. After comprehensive research, design, and testing, the project successfully demonstrated that:

- **Pre-trained models like DenseNet121** are highly effective in extracting deep features from medical images.
- **Multi-label classification** is achievable even in imbalanced and noisy medical datasets.
- Evaluation using metrics such as **precision, recall, and F1-score** provides a more accurate reflection of model performance than accuracy alone, especially for healthcare applications.

By leveraging **transfer learning**, the project minimized training time while achieving reliable results. The system was evaluated across two training stages (2 epochs and 5 epochs) and showed consistent improvement in classification performance. The model successfully recognized dominant conditions like "No Finding" and "Effusion" while identifying gaps in the detection of rare diseases like "Hernia" and "Fibrosis."

The final trained system can act as a **decision-support tool** for radiologists, especially in overburdened or under-resourced clinical environments. It marks a step toward democratizing diagnostic AI in medical imaging.

While the overall project achieved its primary objectives, a few **deviations and limitations** were observed compared to initial expectations:

**1. Class Imbalance Effect**

- Certain diseases in the dataset were underrepresented (e.g., "Pneumonia", "Fibrosis"), leading to **lower F1-scores** despite augmentation.
- The model showed bias toward predicting more common classes like "No Finding."

**2. Short Training Time**

- Due to hardware constraints and time limits, the model was trained for only 2 and 5 epochs.
- This restricted the model's learning capability, particularly for rare features and subtle image abnormalities.

**3. Interpretability Not Integrated**

- While interpretability using techniques like **Grad-CAM** was planned, it was not implemented within the project timeline.
- Clinical deployment would require such explainability features to ensure transparency in model decisions.

**4. Limited Real-World Testing**

- The model was tested only on the ChestX-ray8 dataset, without being validated on hospital-grade real-world data.
- Performance in actual clinical conditions (with different machines, settings, and demographics) remains untested.

## 5.2. Future work

To enhance the system's capabilities and clinical utility, several **future improvements** are proposed:

**1. Addressing Class Imbalance**

- Apply **advanced sampling techniques**, such as SMOTE (Synthetic Minority Oversampling) or **class-weighted loss functions**.
- Consider **focal loss**, which emphasizes harder-to-classify samples.

**2. Extended Training and Fine-Tuning**

- Train the model for **10+ epochs** using a learning rate scheduler for better convergence.
- Implement **cross-validation** to better estimate generalization across folds.

**3. Interpretability and Visualization**

- Integrate **Grad-CAM or saliency maps** to visually explain which parts of the image influenced the model's decision.
- Helps in building trust among clinicians and regulatory bodies.

**4. Expansion to Other Modalities**

- Extend the system to include other radiology images (e.g., CT scans or MRI) using similar architectures.
- Could create a unified diagnostic platform across different organs and diseases.

**5. Mobile and Cloud Deployment**

- Deploy a **MobileNetV2-based** version for mobile applications in remote settings.
- Package the model into a **cloud API** for integration with telemedicine or hospital software.

**6. Regulatory and Clinical Trials**

- Future work must involve **ethical approval and real-world validation** using hospital-grade anonymized datasets.
- Model should comply with **medical device standards** and healthcare regulations before deployment.

# REFERENCES


[1] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "ChestX-ray8: Hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 2097–2106.

[2] P. Rajpurkar et al., "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[3] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[4] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLOS ONE*, vol. 10, no. 3, pp. e0118432, Mar. 2015.

[5] T.-Y. Lin et al., "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2980–2988.

[6] A. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[7] D. Bysani and R. Natarajan, "An ensemble deep learning model for pneumonia detection in chest X-ray images using DNN, CNN and transfer learning," in *Proc. IEEE Int. Conf. Comput. Intell. Comput. Res. (ICCIC)*, 2020, pp. 1–6.

[8] H. Aggarwal, S. Rawat, and P. Narula, "Comparative analysis of VGG16, ResNet50 and DenseNet121 for classification of COVID-19 CT images using transfer learning," in *Proc. IEEE Int. Conf. Comput. Power, Energy, and Controls (ICCPPEC)*, 2021, pp. 1–6.

[9] M. T. Islam, M. A. Aowal, A. Minhaz, and K. Ashraf, "Abnormality detection and localization in chest X-rays using deep convolutional neural networks," *arXiv preprint arXiv:1705.09850*, 2017.

[10] M. Y. Ng, E. Y. P. Lee, J. Yang, et al., "Imaging profile of the COVID-19 infection: radiologic findings and literature review," *Radiology: Cardiothoracic Imaging*, vol. 2, no. 1, pp. e200034, 2020.

[11] S. Irvin et al., "CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison," in *Proc. AAAI Conf. Artif. Intell.*, vol. 33, 2019, pp. 590–597.

[12] H. T. Nguyen et al., "VinDr-CXR: An open dataset of chest X-rays with radiologist annotations," *PhysioNet*, 2020. [Online]. Available: https://physionet.org/content/vindr-cxr/1.0.0/

[13] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 1251–1258.

[14] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Conf. Oper. Syst. Design Implement. (OSDI)*, 2016, pp. 265–283.

[15] S. Raschka and V. Mirjalili, *Python Machine Learning*, 3rd ed. Birmingham, U.K.: Packt Publishing, 2019.