Bahria University Islamabad

DEPARTMENT OF COMPUTER SCIENCES

Artificial Intelligence

Project Report

Title: **Chest X-ray Analysis using Deep Learning**

SUBMITTED BY:

Muhammad Awab Sial (01-134222-091)

Syed Amber Ali Shah (01-134222-146)

Submitted To: **Dr Arshad Farhad**

Contents

Chapter 1: Introduction

1.1 Introduction

Medical imaging plays a crucial role in the early diagnosis of diseases. However, radiologists often face challenges due to the complexity of X-ray images, variations in image quality, and the presence of multiple overlapping structures. This project leverages deep learning models to automate the detection of various lung diseases, reducing human error and speeding up the diagnostic process.

This project aims to develop an AI-powered system for automated analysis of chest X-ray (CXR) images. The system will assist radiologists in detecting abnormalities such as pneumonia, tuberculosis, and lung cancer. By leveraging deep learning techniques, the project seeks to improve diagnostic accuracy and reduce workload in medical imaging. The model will be trained and evaluated using the NIH Chest X-ray Dataset, ensuring high-quality medical data for robust predictions.

1.2 Existing Solutions & Their Limitations

Existing automated diagnosis tools often lack generalizability due to limited datasets and insufficient image diversity. Manual annotation also introduces subjectivity. Some tools require extensive computational resources, hindering real-time applications.

1.3 Scope

This project aims to build a robust AI-based diagnostic tool using ResNet-50 to classify thoracic diseases from chest X-rays. It includes model training, evaluation, and a simple interface for viewing predictions where the user can upload a chest X-ray image. Multi-label classification and Grad-CAM visualizations are within scope.

Out of Scope:

- Real-time X-ray scanning integration.
- Complex hospital system integration.

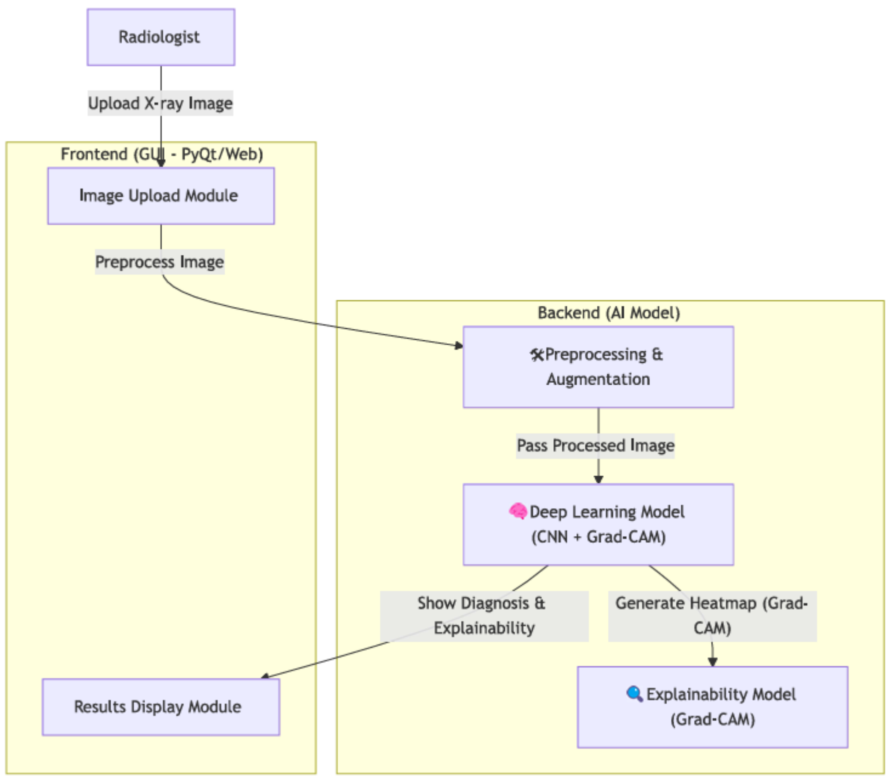
1.4 Objectives

The objective is to detect thoracic diseases using deep learning, enhancing diagnosis accuracy and reducing the burden on radiologists. We aim to use the NIH dataset and implement ResNet-50 with visual explanation via Grad-CAM.

1.5 Targeted Users

The primary users of this system will be radiologists, healthcare professionals, and medical researchers. The AI-assisted system will help in rapid diagnosis and decision-making, especially in resource-limited settings.

1.6 Architecture Diagram



This diagram represents the workflow of an AI-based chest X-ray classification system designed for radiologists. It consists of two primary components: the Frontend (GUI) 3 and the Backend (AI Model), which work together to process X-ray images, perform classification, and provide explainability through Grad-CAM.

Chapter 2: Dataset Specifications

2.1 Dataset Overview

NIH ChestX-ray dataset comprises 112,120 frontal-view X-ray images (PNG images in 1024*1024 resolution) of 30,805 unique patients with the text-mined fourteen disease image labels (where each image can have multilabels). Fourteen common thoracic pathologies include **Atelectasis, Consolidation, Infiltration, Pneumothorax, Edema, Emphysema, Fibrosis, Effusion, Pneumonia, Pleural_thickening, Cardiomegaly, Nodule, Mass, and Hernia.**

2.2 Strengths of NIH Dataset

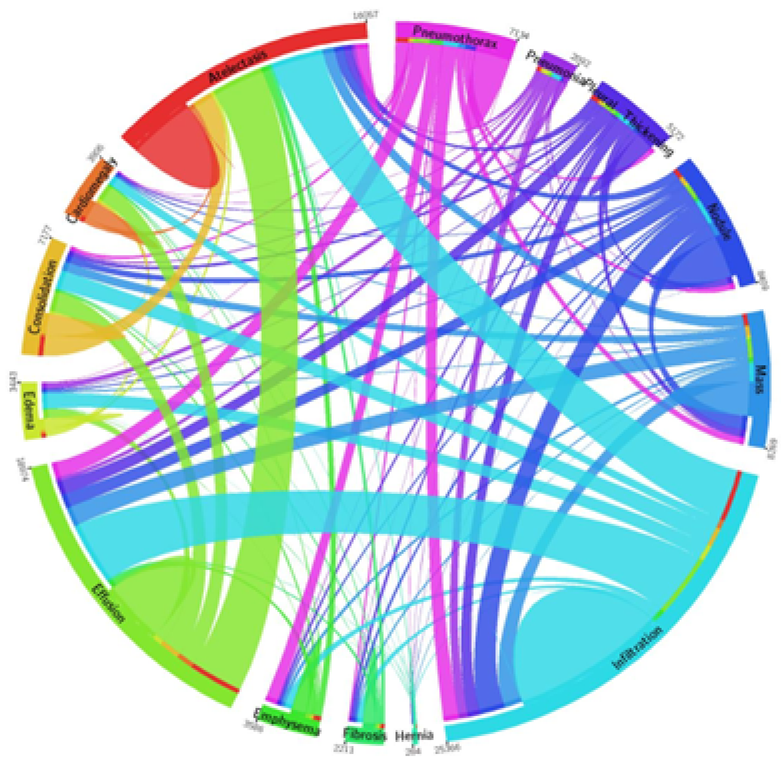
- Large-scale dataset ideal for deep learning.
- Multi-label annotations per image.
- Real hospital data for real-world applicability.

2.3 Weaknesses of NIH Dataset

- The image labels are NLP extracted so there would be some erroneous labels but the NLP labelling accuracy is estimated to be >90%.
- 2Very limited numbers of disease region bounding boxes.
- Chest x-ray radiology reports are not anticipated to be publicly shared. Parties who use this public dataset are encouraged to share their "updated" image labels and/or new bounding boxes in their own studied later, maybe through manual annotation.
- Dataset imbalance, which affects the training and testing of the model.

2.4 Dataset Details

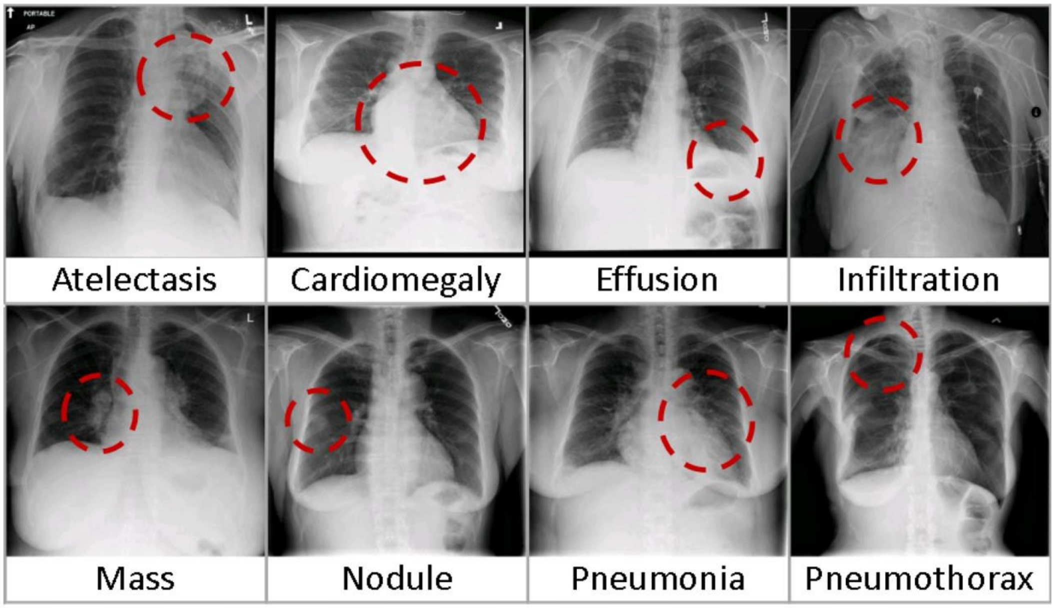
A. Distributions of 8 disease categories with co-occurrence statistics:



B. Co-occurrence matrix of the fourteen thorax diseases in this chest X-ray dataset

4212	369	3269	3259	727	585	243	772	1222	221	423	220	495	40
369	1094	1060	583	99	108	36	48	169	127	44	51	111	7
3269	1060	3959	3990	1244	909	253	995	1287	592	359	188	848	21
3259	583	3990	9552	1151	1544	571	943	1220	979	447	345	749	33
727	9	1244	1151	2138	894	62	424	602	128	212	115	448	25
585	108	909	1544	894	2706	63	340	428	131	115	166	410	10
243	36	253	571	62	63	307	34	114	330	21	11	45	2
772	48	995	943	424	340	34	2199	222	33	746	80	289	9
1222	169	1287	1220	602	428	114	222	1314	162	103	79	251	4
221	127	592	979	128	131	330	33	162	634	30	9	64	3
423	44	359	447	212	115	21	746	103	30	895	36	151	4
220	51	188	345	115	166	11	80	79	9	36	727	176	8
495	111	848	749	448	410	45	289	251	64	151	176	1127	8
40	7	21	33	25	10	2	9	4	3	4	8	8	110
11535	2772	13307	19871	5746	6323	1353	5298	4667	2303	2516	1686	3385	227

C. Eight visual examples of common thorax diseases



Chapter 3: Requirement Specifications

3.1 Functional Requirements

- Load and preprocess chest X-ray images on the front end.
- Train a ResNet-50 based classification model.
- Predict multiple disease labels per image.
- Visualize predictions using Grad-CAM.

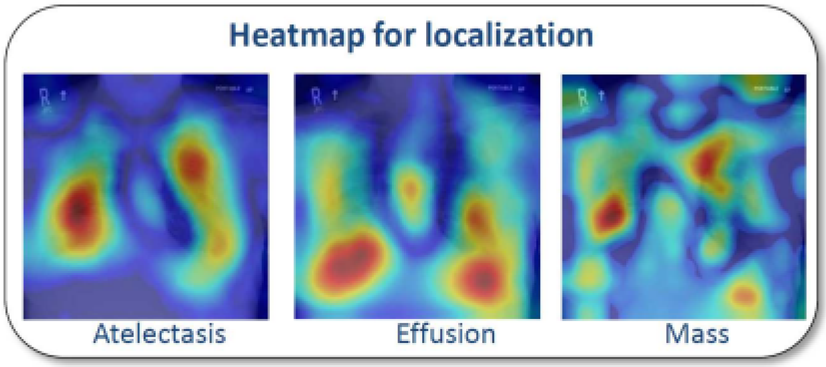
3.2 Non-Functional Requirements

- High model accuracy (>90%) on test data.
- Training time under 3 hours on GPU.
- User interface must be intuitive.

3.3 Use Case Descriptions

- Upload Image: Users upload chest X-ray images for classification.
- View Prediction: Users receive multi-label outputs with confidence scores.
- Explanation: Grad-CAM heatmap shows activated lung regions.

Grad-CAM visualizations were used to highlight regions in X-ray images that influenced predictions.



Chapter 4: Software Design & Implementation

4.1 Model Architecture

We use the ResNet-50 architecture, a deep convolutional neural network pre-trained on ImageNet. It is adapted for multi-label classification by replacing the final layer with a sigmoid-activated dense layer of 8 outputs. The model is fine-tuned using transfer learning.

4.2 Training Details

Training was done using the NIH Chest X-ray dataset.

Training configuration:

Parameter	Value
Optimizer	Adam
Learning Rate	1e-4
Batch Size	16
Epochs	5
Device	Kaggle T4 GPU
Checkpoint Saving	chestxray_model_resnet50.pth
Model	ResNet-50 (pretrained) + Conv2d (2048→1024) + LSEPooling + FC(1024→8)
Pooling	LSE Pooling (r=10, eps=1e-6)
Loss	Weighted BCE Loss
Image Size	512×512
Input Channels	3 (RGB)
Output	Sigmoid vector of size 8 (multi-label)
Transformations	Resize + ToTensor

4.3 Evaluation Metrics

The model was evaluated using the following metrics:
- Accuracy
- Precision, Recall, F1-Score (per class)
- ROC-AUC
- Confusion Matrix

4.4 Architectural Details

Final layers removed (children()[:-2]) to retain spatial feature maps for pooling.

Custom Layers Added

- **Transition**: nn.Conv2d(2048, 1024, kernel_size=1)
  - Reduces the channel dimension from 2048 (ResNet output) to 1024 for classification.
- **Pooling Layer**: LSEPooling(r=10)
  - Log-Sum-Exp pooling aggregates spatial features with a smooth approximation between max and average pooling.
  - Parameter:
    - r = 10 → Controls pooling sharpness (higher = more like max pooling).
    - eps = 1e-6 → Stability constant to avoid log(0).
- **Classifier**: nn.Linear(1024, 8)
  - Fully connected layer to output logits for **8 diseases**.

Output passes through torch.sigmoid() for **multi-label** probability scores (range 0–1 per disease).

Weighted Binary Cross-Entropy Loss

def weighted_bce_loss(output, target):

eps = 1e-8

pos_weight = (target == 0).float().sum() / ((target == 1).float().sum() + eps)

- **Purpose**: Handle class imbalance dynamically by giving higher weight to rare labels.
- **Mechanism**:
  - Calculates pos_weight for each batch.
  - Applies standard BCE with weighted positive class contribution.
  - Reduces with .mean() over all classes and samples.

Chapter 5: AI Model Training Results and Testing

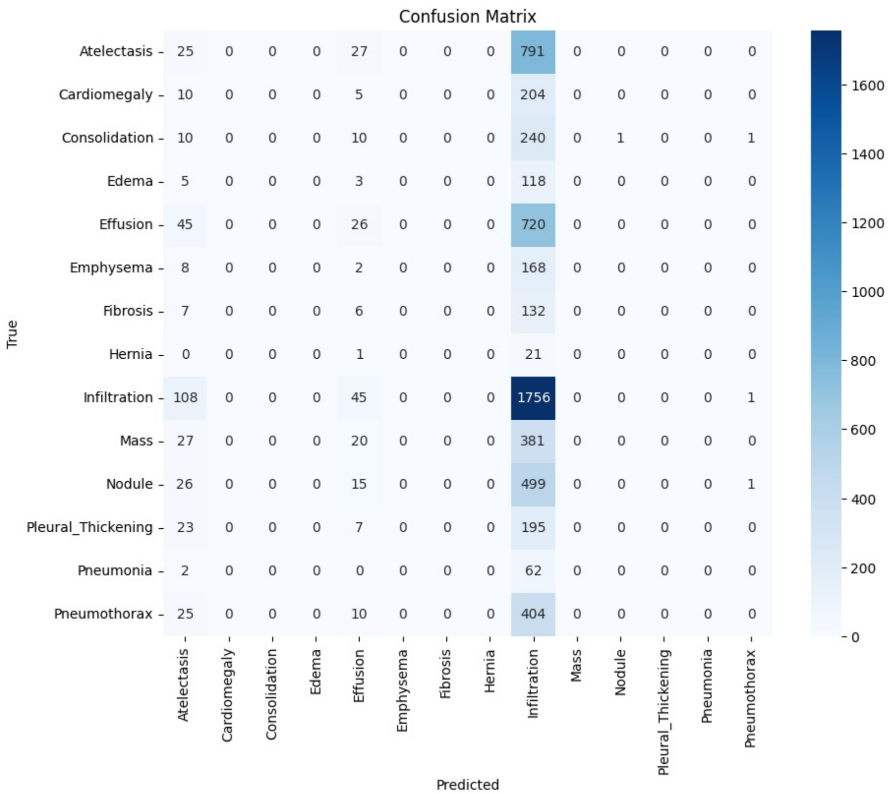
5.1 Different Models Comparison

Model Name	Average Loss	Epochs	Validation AUC	Accuracy
EfficientNet B1	**0.1475**	10	**0.7522**	0.53
Dense net	2.4056	10	0.3217	0.33
ResNet-50	0.7498	5	0.72	**0.78**

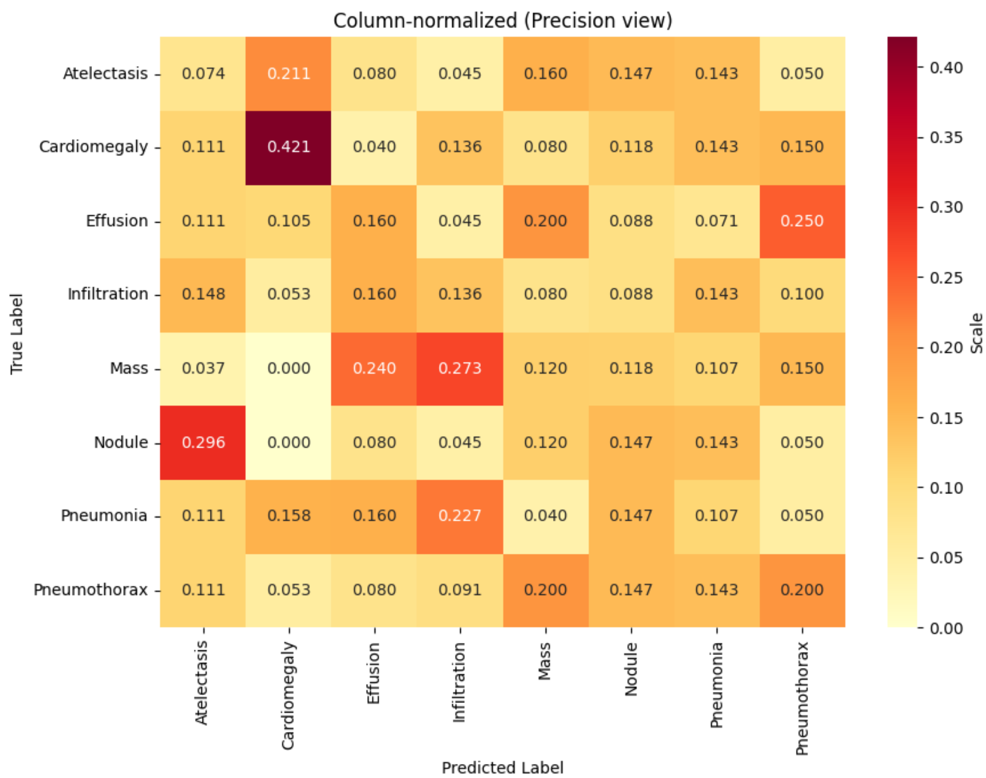
5.2 Evaluation Matrices

1. **DenseNet**

```
Classification Report:

                    precision    recall  f1-score   support

       Atelectasis       0.08      0.03      0.04       843
      Cardiomegaly       0.00      0.00      0.00       219
     Consolidation       0.00      0.00      0.00       262
             Edema       0.00      0.00      0.00       126
          Effusion       0.15      0.03      0.05       791
         Emphysema       0.00      0.00      0.00       178
          Fibrosis       0.00      0.00      0.00       145
            Hernia       0.00      0.00      0.00        22
      Infiltration       0.31      0.92      0.46      1910
              Mass       0.00      0.00      0.00       428
            Nodule       0.00      0.00      0.00       541
Pleural_Thickening       0.00      0.00      0.00       225
         Pneumonia       0.00      0.00      0.00        64
      Pneumothorax       0.00      0.00      0.00       439

          accuracy                           0.29      6193
         macro avg       0.04      0.07      0.04      6193
      weighted avg       0.12      0.29      0.16      6193
```


Confusion Matrix
True
Predicted
Atelectasis
Cardiomegaly
Consolidation
Edema
Effusion
Emphysema
Fibrosis
Hernia
Infiltration
Mass
Nodule
Pleural_Thickening
Pneumonia
Pneumothorax
25 0 0 0 27 0 0 0 791 0 0 0 0 0
10 0 0 0 5 0 0 0 204 0 0 0 0 0
10 0 0 0 10 0 0 0 240 0 1 0 0 1
5 0 0 0 3 0 0 0 118 0 0 0 0 0
45 0 0 0 26 0 0 0 720 0 0 0 0 0
8 0 0 0 2 0 0 0 168 0 0 0 0 0
7 0 0 0 6 0 0 0 132 0 0 0 0 0
0 0 0 0 1 0 0 0 21 0 0 0 0 0
108 0 0 0 45 0 0 0 1756 0 0 0 0 1
27 0 0 0 20 0 0 0 381 0 0 0 0 0
26 0 0 0 15 0 0 0 499 0 0 0 0 1
23 0 0 0 7 0 0 0 195 0 0 0 0 0
2 0 0 0 0 0 0 0 62 0 0 0 0 0
25 0 0 0 10 0 0 0 404 0 0 0 0 0
1600
1400
1200
1000
800
600
400
200
0

2. **ResNet-50**



Column-normalized (Precision view)
True Label
Predicted Label
Scale
Atelectasis
Cardiomegaly
Effusion
Infiltration
Mass
Nodule
Pneumonia
Pneumothorax
0.074 0.211 0.080 0.045 0.160 0.147 0.143 0.050
0.111 0.421 0.040 0.136 0.080 0.118 0.143 0.150
0.111 0.105 0.160 0.045 0.200 0.088 0.071 0.250
0.148 0.053 0.160 0.136 0.080 0.088 0.143 0.100
0.037 0.000 0.240 0.273 0.120 0.118 0.107 0.150
0.296 0.000 0.080 0.045 0.120 0.147 0.143 0.050
0.111 0.158 0.160 0.227 0.040 0.147 0.107 0.050
0.111 0.053 0.080 0.091 0.200 0.147 0.143 0.200
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00

Classification Report

	precision	recall	f1-score	support
Atelectasis	0.73	1.0	0.84	1.0
Cardiomegaly	0.81	0.0	0.0	0.0
Effusion	0.86	0.0	0.0	0.0
Infiltration	0.81	1.0	0.89	2.0
Mass	0.9	0.0	0.0	0.0
Nodule	0.8	0.0	0.0	0.0
Pneumonia	0.76	0.0	0.0	0.0
Pneumothorax	0.96	1.0	0.98	1.0
micro avg	0.27	1.0	0.42	4.0
macro avg	0.15	0.38	0.34	4.0
weighted avg	0.38	1.0	0.54	4.0

Chapter 6: Challenges Faced

Developing a deep learning model for chest X-ray disease classification using the NIH dataset involved several technical and resource-related challenges. This section outlines the key obstacles we encountered, along with the strategies employed to overcome them.

1. Dataset Imbalance

The NIH Chest X-ray dataset is heavily imbalanced across disease categories. Some diseases like Infiltration and Effusion have thousands of examples, while others like Mass or Nodule are significantly underrepresented. This imbalance caused the model to favor common diseases during training, reducing its sensitivity to rare conditions.

To address this, we implemented a custom Weighted Binary Cross-Entropy (BCE) Loss function, which calculates positive class weights dynamically for each batch. This helps compensate for the underrepresented classes during model updates.

2. Noisy and Weak Labels

The dataset labels were generated using NLP techniques applied to radiology reports, which can result in mislabeled or ambiguous samples. These noisy labels affect the model's ability to learn accurate decision boundaries.

Since manual relabeling was not feasible, we tackled this challenge by:

- Using a multi-label sigmoid output layer to allow soft predictions for co-occurring conditions.
- Focusing evaluation more on AUC and Grad-CAM visualization to validate correctness qualitatively.

3. Limited Hardware Capabilities

Training deep learning models on local machines was impractical due to the absence of a dedicated GPU and limited RAM. Our local setup lacked the computational power needed for large-scale medical image training.

To resolve this, we migrated training to Kaggle's free T4 GPU environment, which provided better speed and stability. However, Kaggle's platform has its own constraints:

- Session timeouts (after every 12 hours) and inactivity limits.
- Restriction on long-running jobs.
- I/O delays due to dataset mounting and image preprocessing.
- GPU usage is restricted to 30 hours per week.

These limitations required us to optimize training by:

- Reducing image resolution to 512×512.
- Lowering batch size and training for only 10 epochs.
- Using lightweight transforms and simplified data loading logic.

4. Evaluation and Monitoring Limitations

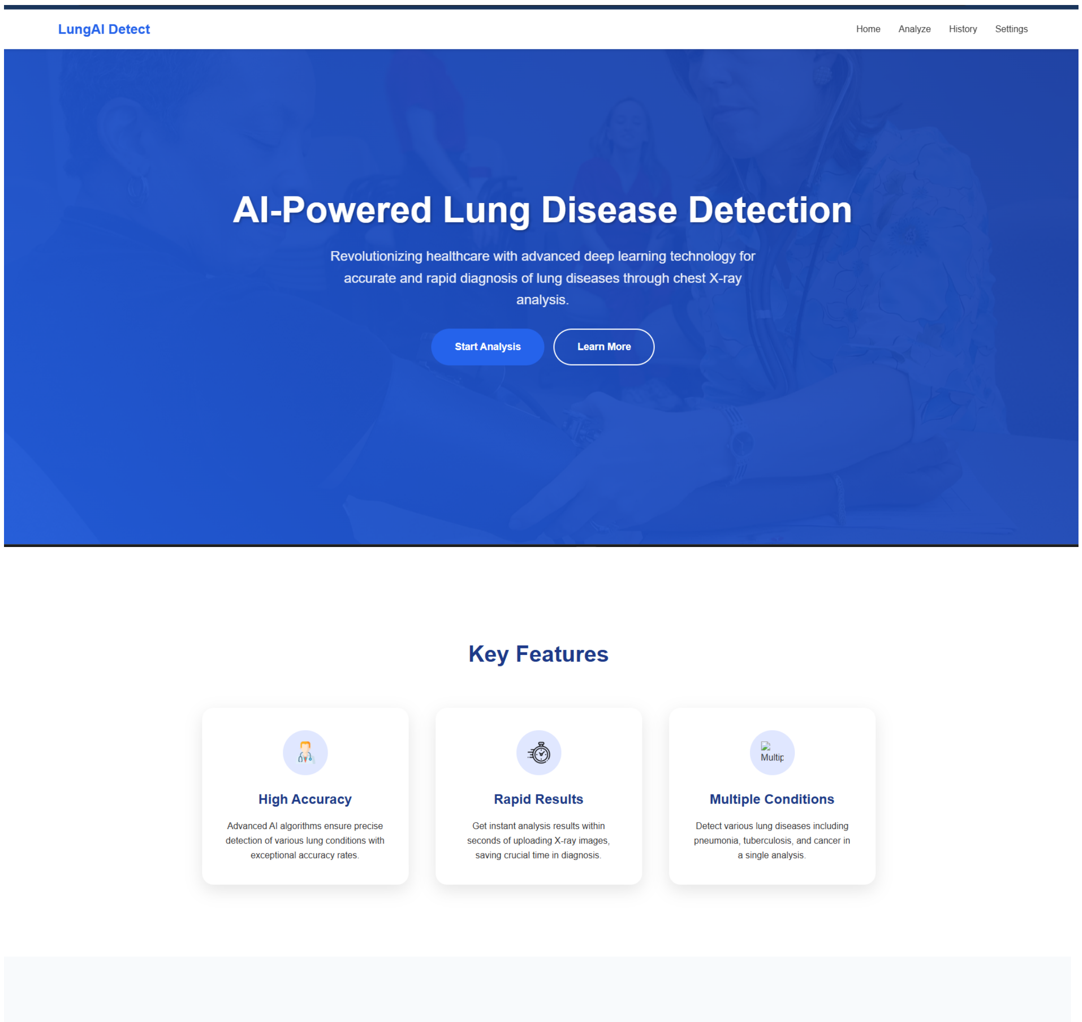
Due to time and GPU constraints, we were limited to training for only a few epochs and couldn't conduct exhaustive hyperparameter tuning or cross-validation. Moreover, the absence of a proper validation split during training restricted our ability to monitor generalization.

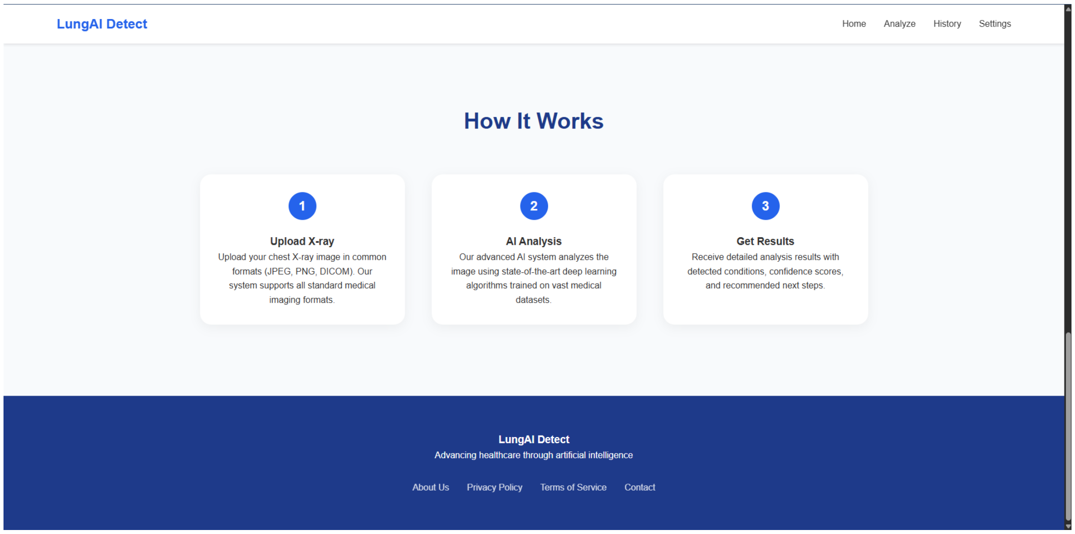
We focused instead on:

- Tracking training loss at regular intervals.
- Saving the best model using torch.save() after training.
- Performing post-training evaluation using confusion matrix, AUC, and Grad-CAM overlays.

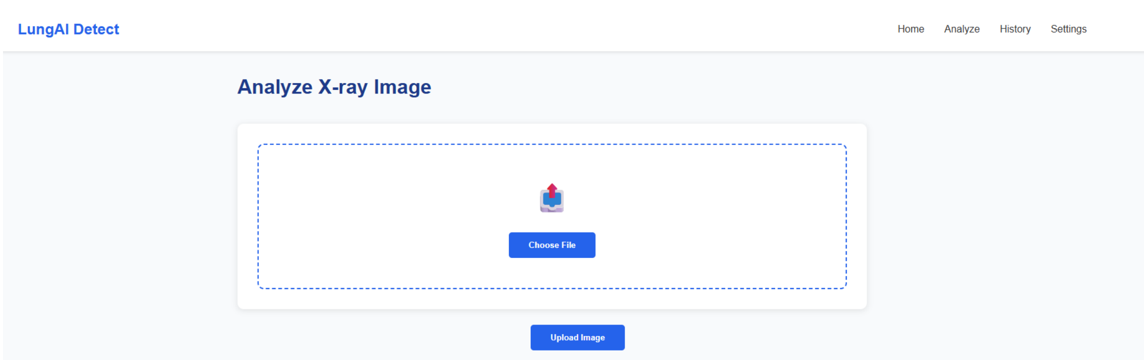
Chapter 7: Front End Screens

Home Page





Analysis Page



Analysis Output