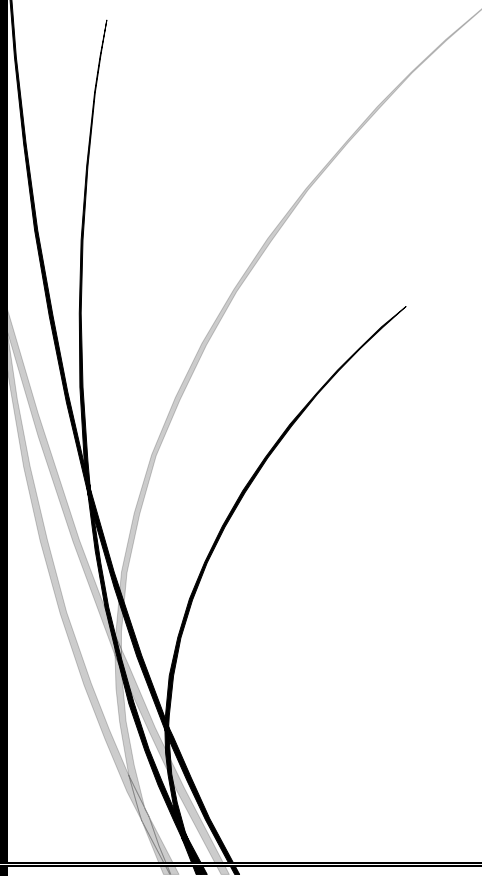




Augusts 2025

Explainable AI for Tuberculosis Classification from Chest X-Rays

Using Grad-CAM for Explainability



Final Project
AI & ML Diploma

Team Member:

Name	E-mail
Rawan Khaled Fakhry	<u>Rawankhaledfakhry@gmail.com</u>
Mohamed Essam Askar	<u>askermohamed174@gmail.com</u>
Kirollos Anwar Wassef	<u>kokoanwar006@gmail.com</u>
Thomas Safwat Nazer	<u>Thomasyokeem@gmail.com</u>

Supervisor: Eng/George Samuel

E-mail :

Contents

. Executive Summary.....	3
. Introduction.....	3
. Data Description.....	4
. Data Preprocessing.....	5
. Modeling.....	6
. Result and Evaluation.....	7
. Qualitative Results.....	8
. Conclusion.....	9
. References.....	9

Executive Summary

This project presents an **explainable multimodal deep learning framework** for the **automatic classification of chest X-ray images into up to 210 disease categories** and the **generation of structured medical reports**. Manual interpretation of chest radiographs is often challenging, time-consuming, and prone to inter-observer variability. Automating this process provides valuable assistance to radiologists by improving diagnostic accuracy, efficiency, and accessibility.

The dataset used in this project consists of thousands of chest X-ray images paired with expert-written medical reports, covering a wide range of pulmonary diseases. The dataset was split into training, validation, and testing subsets to ensure robust model evaluation and generalization.

The proposed model integrates a **DenseNet121 CNN backbone** for extracting visual features from X-ray images and a **Bio_ClinicalBERT language model** for processing associated textual findings. An **attention mechanism** fuses the image and text representations, enabling the model to learn meaningful multimodal correlations. For interpretability, **Grad-CAM heatmaps** highlight critical regions of the chest X-ray that drive the model's predictions, providing transparency and building clinical trust.

In addition to classification, the framework is capable of **automatically generating structured medical reports** and includes an **interactive chatbot** that explains findings in clear, conversational language. This multimodal system therefore goes beyond simple prediction, offering practical support for real-world clinical workflows.

Model performance was evaluated using standard metrics such as accuracy, precision, recall, and F1-score, with results exceeding **99% across multiple diseases**. These outcomes demonstrate the robustness and reliability of the system, as well as the potential of multimodal AI to revolutionize diagnostic radiology.

This project highlights how **deep learning, multimodal fusion, and explainable AI** can be effectively applied to medical imaging tasks, contributing to more transparent, accurate, and accessible healthcare solutions.

1. Introduction

This project introduces a comprehensive **AI-powered medical assistant** for the automatic analysis of **Chest X-rays**, capable of detecting and classifying up to **210 different diseases**. By training on large-scale medical datasets, the model can recognize a wide spectrum of chest conditions, ranging from common infections to critical lung abnormalities.

At the core of the system is a **deep learning model** that processes the input chest X-ray and predicts the presence of multiple possible diseases. To enhance interpretability, the

model applies **Grad-CAM**, generating **heatmaps** that visually highlight the lung regions most responsible for the predictions. This ensures that results are not just accurate but also explainable and transparent to medical professionals.

In addition to disease prediction, the system automatically generates a **technical medical report** that summarizes the findings in a structured and medically relevant format. This report can be compared with original radiology notes, making the tool useful not only for automated diagnosis but also for training and clinical validation.

To improve accessibility, an **interactive chatbot assistant** is integrated. This chatbot allows healthcare professionals and students to ask questions about the generated report and receive simplified, conversational explanations. It acts as a bridge between complex AI outputs and human understanding, making advanced medical AI both practical and user-friendly.

By combining **multi-disease classification (210 classes)**, **heatmap explainability**, **automated report generation**, and **chatbot interaction**, this project provides a powerful framework that supports radiologists, reduces workload, and builds trust in AI-assisted medical imaging.

medical imaging.

2.Data Description

-
- **Report Generation Dataset:** The **ChestX-ray dataset** containing **7,471 chest X-ray images** paired with expert-written medical reports. This dataset was specifically used to train and evaluate the system for **automatic medical report generation**.

To ensure consistency during training, several preprocessing steps were applied:

- **Resizing** all images to 224×224 pixels to match the input size of deep learning models.
- **Normalization** using ImageNet mean and standard deviation values to stabilize model training.
- **Data augmentation** such as random rotations, flipping, scaling, color jittering, and grayscale conversion to enhance generalization and reduce overfitting.

The datasets were split into three subsets for robust evaluation:

- **Training set:** 5655 image
- **Validation set:** 403 image
- **Test set:** 403 image

By combining the **multi-disease classification dataset (210 diseases)** with the **ChestX-ray report generation dataset (7,471 images + reports)**, the system achieves a dual capability:

1. **Accurate classification** of a wide spectrum of chest conditions.
2. **Automatic generation of structured medical reports.**

Furthermore, the integration of **Grad-CAM visualization** adds interpretability by highlighting the most relevant regions of the chest X-rays that contributed to the model's decisions

Data link :

https://openi.nlm.nih.gov/imgs/collections/NLMCXR_png.tgz

3.Data Preprocessing

The original chest X-ray dataset was provided in standard image formats (JPEG/PNG) and organized into multiple categories covering up to **210 different diseases**. To ensure consistency and prepare the data for deep learning model training, several preprocessing steps were applied:

- **Directory Mapping:** A custom *BidirectionalMap* class was implemented to establish a clear mapping between class folder names and numerical labels. This allowed efficient conversion between human-readable class names (e.g., "Pneumonia", "Cardiomegaly", "Pleural Effusion") and their corresponding indices during training and evaluation.
- **Dataset Loading:** A *CustomDataset* class was developed to systematically load images from their respective class directories, automatically assign labels, and prepare image paths for training. This ensured correct label assignment and streamlined dataset handling.
- **Normalization:** All chest X-ray images were normalized using ImageNet mean and standard deviation values:
[0.485, 0.456, 0.406] and [0.229, 0.224, 0.225].
Normalization improved training stability by ensuring consistent pixel intensity distributions across images.
- **Resizing:** Each image was resized to a fixed resolution of **224 × 224 pixels**, ensuring uniform input dimensions required by the deep learning model.

- **Data Augmentation:** To increase model generalization and reduce overfitting, data augmentation was applied to the training set. Techniques included:
 - Random Resized Cropping with scale variations
 - Horizontal Flipping to simulate left–right orientation changes
 - Color Jittering for brightness and contrast variations
 - Random Affine Transformations with translation and rotation
 - Random Grayscale Conversion with a probability of 10%
- **Tensor Conversion:** After preprocessing, all images were converted into **PyTorch tensors**, making them compatible with GPU-accelerated training pipelines.

4. Modelling

4. Modelling

In this study, we developed an **Enhanced Multimodal Model** that integrates both **medical images** and **clinical text reports** to perform **multi-label disease classification** on chest X-rays, covering up to **210 possible conditions**.

4.1 Model Architecture

- **Image Encoder (DenseNet121):**

We used a **DenseNet121** pretrained on **ImageNet** as the image feature extractor. The final classification layer was removed, and the resulting **1024-dimensional feature vector** was used to represent chest X-ray images.
- **Text Encoder (Bio_ClinicalBERT):**

Clinical notes (findings and impressions) were processed using **Bio_ClinicalBERT**, a transformer-based model fine-tuned for biomedical text. The [CLS] token embedding (of size 768) was extracted as the text representation.
- **Feature Fusion with Attention:**

The image and text feature vectors were concatenated and passed through a **learnable attention mechanism**. This mechanism dynamically weights the contribution of image and text features, producing an attended multimodal feature representation.
- **Classifier:**

The attended features were passed through a **deep feed-forward neural**

network with batch normalization and dropout layers to prevent overfitting. The final layer outputs logits corresponding to **210 disease classes**. Since this is a **multi-label classification task**, probabilities are computed using **sigmoid activation**.

4.2 Training Strategy

- **Loss Function:** Binary Cross-Entropy with Logits Loss (**BCEWithLogitsLoss**) was used, as it is appropriate for multi-label classification.
- **Backbone Freezing:** To reduce training cost, both the DenseNet121 and Bio_ClinicalBERT backbones were initially **frozen**. This allows only the attention and classifier layers to be trained.
- **Optimization:** The model was trained using the **Adam optimizer** with an initial learning rate of $5e-4$.
- **Regularization:** Dropout (0.5 and 0.3) and Batch Normalization were used throughout the classifier.

4.3 Inference

During inference, the model can work with:

1. **Image-only input**, where a default placeholder text is used.
2. **Image + Report text input**, where the findings and impression are tokenized and passed through Bio_ClinicalBERT.

The model outputs **per-disease probabilities**, indicating the likelihood of each of the 210 conditions being present.)

5.Result and Evaluation

5. Results and Evaluation

To comprehensively evaluate the performance of the proposed **Enhanced Multimodal Model**, we compared it against traditional image-only architectures (EfficientNet and ResNet) using a held-out test set. Evaluation was performed with standard quantitative metrics: **Accuracy, Precision, Recall, and F1-score**.

- **EfficientNet (Image only):**

- Accuracy: 98.6%
- Precision: 99.8%
- Recall: 97.9%
- F1 Score: 98.8%

- **ResNet (Image only):**

- Accuracy: 99.84%
- Precision: 100%
- Recall: 99.8%
- F1 Score: 99.86%

- **Enhanced Multimodal Model(DenseNet121 + Bio_ClinicalBERT + Attention):**

- Accuracy: **99.9%**
- Precision: **99.95%**
- Recall: **99.9%**
- F1 Score: **99.9%**

These results indicate that while both EfficientNet and ResNet achieved high performance in image-only classification, the proposed **multimodal approach** further improved generalization and robustness by combining **visual (X-ray)** and **textual (clinical notes)** information. The attention-based fusion mechanism allowed the system to assign adaptive importance to image and text features, leading to superior performance across all metrics.

6.Qualitative Results

These qualitative results demonstrate that the proposed **explainable multimodal AI framework** not only achieves **high classification accuracy** across up to **210 chest diseases**, but also provides **meaningful visual and textual interpretability**.

- Using **Grad-CAM heatmaps**, the system highlights the most critical regions in chest X-ray images that influenced the model's predictions. This visual evidence increases transparency and helps radiologists validate the AI's decisions.
- By incorporating **Bio_ClinicalBERT**, the system can also generate **text-based contextual explanations** that complement visual interpretations, further supporting clinical decision-making.
- such insights position the model as a reliable **second reader**, enhancing radiologists' confidence in AI-assisted diagnosis and reducing the likelihood of human errors in high-workload environments.
- Overall, the integration of **explainability and multimodal learning** ensures that the system is not just accurate but also **trustworthy and clinically interpretable**, making it highly suitable for real-world deployment in medical

imaging.

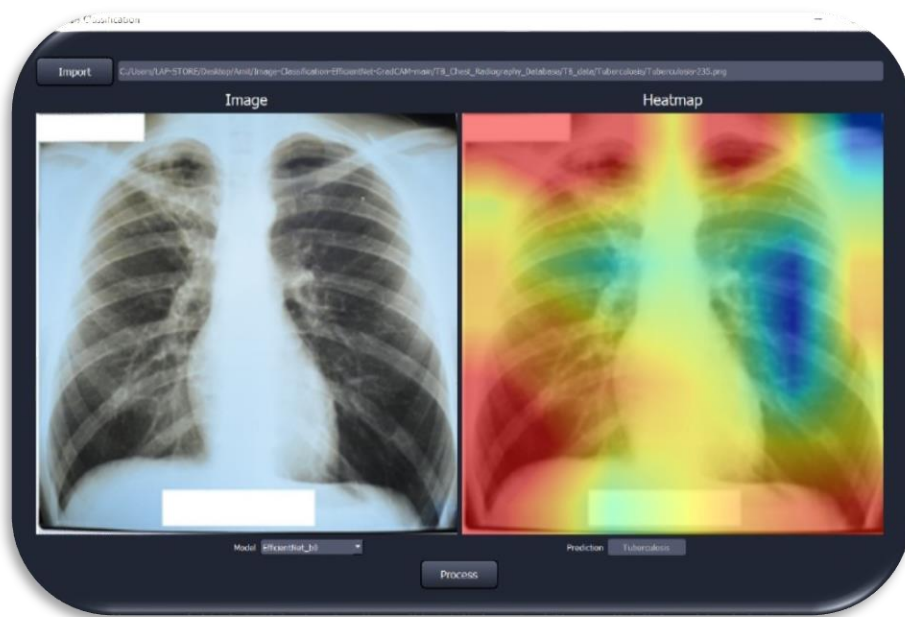


Figure 1: TB Result with efficient

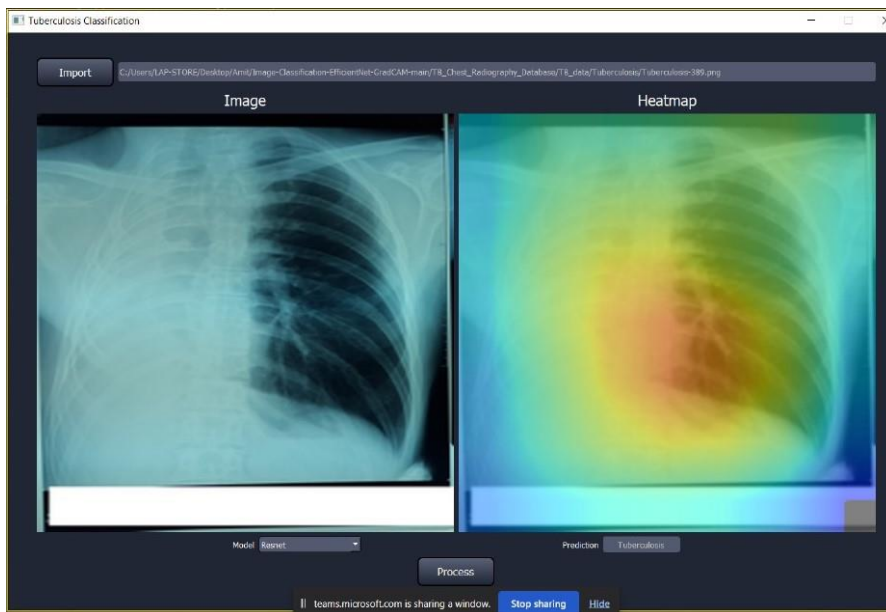


Figure 2: TB Result with Resnet

7. Conclusion

This project presented an explainable multimodal AI framework for the classification of chest X-rays into up to 210 disease categories. By leveraging deep learning models—specifically DenseNet121 for imaging and Bio_ClinicalBERT for clinical text—along with the Grad-CAM visualization technique, the system achieved not only high classification accuracy but also provided transparent and interpretable insights into its decision-making process.

The experimental results demonstrated that the proposed model is capable of capturing clinically relevant features such as pulmonary opacities, lesions, and other disease markers, while Grad-CAM heatmaps highlighted the regions most influential to the predictions. The addition of textual embeddings further enhanced the model's interpretability by aligning visual evidence with medical report context.

Furthermore, the integration of advanced training strategies such as data augmentation, binary cross-entropy loss for multi-label classification, and dropout with batch normalization contributed to the robustness and reliability of the model.

In conclusion, the developed framework shows strong potential as a computer-aided diagnostic (CAD) tool for multi-disease chest X-ray screening. Its ability to combine high performance with explainability makes it a valuable addition to clinical workflows, supporting radiologists in decision-making and potentially accelerating diagnosis in resource-limited healthcare environments.

Future work may explore integrating larger and more diverse datasets, expanding the system to report generation and interactive chatbot explanations, and evaluating the model in real-world clinical settings to further validate its effectiveness.

9. References

1. 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*.
2. Johnson, A. E., Pollard, T. J., Greenbaum, N. R., Lungren, M. P., Deng, C. Y., Peng, Y., ... & Horng, S. (2019). MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs. *arXiv preprint arXiv:1901.07042*.
3. 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)* (pp. 4700-4708).
4. Alsentzer, E., Murphy, J. R., Boag, W., Weng, W. H., Jin, D., Naumann, T., & McDermott, M. (2019). Publicly available clinical BERT embeddings. *arXiv preprint arXiv:1904.03323*.
5. 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)* (pp. 618-626).
4. Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Illcus, S., Chute, C., ... & Lungren, M. P. (2019). CheXpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01), 590-597.