X-Ray Image Processing and Large Language Models for Enhanced Multi-Disease Detection and Automated Reporting

Dr. Raja Das
School of Advanced Sciences
Vellore Institute of Technology
Vellore, India
rajadas@vit.ac.in

Kartik Ranjan
School of Advanced Sciences
Vellore Institute Of Technology
Vellore, India
kartik.ranjan2023@vitstudent.ac.in

Mathew Kevin John
School of Advanced Sciences
Vellore Institute Of Technology
Vellore, India
mathew.kevin2023@vitstudent.ac.in

Udaiyar Vishnu Ganesan
School of Advanced Sciences
Vellore Institute Of Technology
Vellore, India
udaiyar.vishnu2023@vitstudent.ac.in

Abstract*— Chest X-rays are a fundamental diagnostic tool for identifying a wide range of thoracic diseases. However, the interpretation of these images is often time-consuming and subject to variability among radiologists. This project aims to enhance multi-disease detection and automate the reporting process by integrating advanced X-ray image processing techniques with large language models (LLMs). Utilizing a comprehensive Chest X-ray dataset comprising 112120 high-resolution (1024×1024) frontal-view PNG images across 14 common thoracic disease categories, we develop a robust deep learning framework for accurate disease identification and localization. The identified findings are input into a large language model to generate coherent and clinically relevant diagnostic reports automatically. This approach improves accuracy and efficiency in disease detection while reducing the workload on radiologists and minimizing human error.

I. Introduction

Chest X-rays remain one of the most widely utilized medical imaging modalities for diagnosing thoracic diseases. Despite their widespread use, the interpretation of X-rays presents challenges such as inter-reader variability and potential misdiagnosis. This project explores the potential of combining advanced image processing techniques with artificial intelligence, particularly large language models (LLMs), to enhance multi-disease detection and automate the generation of diagnostic reports.

II. Data Collection and Preprocessing

The NIH Clinical Center's publicly available Chest X-ray dataset, a comprehensive collection consisting of 112,120 frontal-view X-ray images at a resolution of 1024x1024 from 30,805 patients. Each image is annotated with one or more of 14 thoracic disease categories, including conditions like Atelectasis, Pneumonia, and Cardiomegaly. The dataset's multi-label annotations enable a sophisticated classification framework that reflects real-world clinical scenarios where multiple thoracic conditions may co-occur. In addition to disease labels, each image is accompanied by bounding box annotations, which help define areas of interest, and patient demographics, adding valuable clinical context for training and analysis.

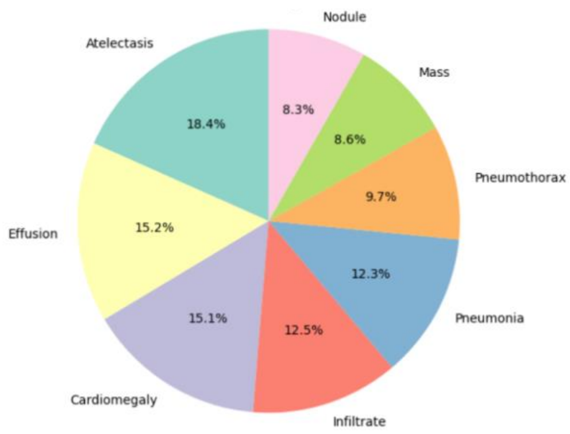

Fig.1.1. Distribution of Unique Diseases

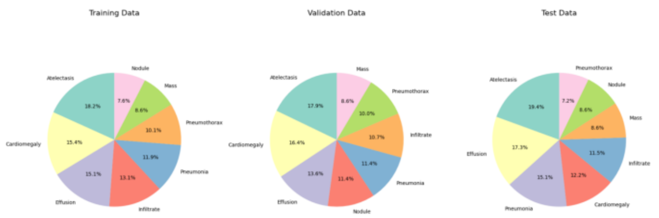

Fig.1.2. Distribution of unique diseases across training, validation and test

In order to support the Retrieval-Augmented Generation (RAG) framework for automated report generation, this study integrates an additional data layer by curating a disease-specific corpus from peer-reviewed research articles, clinical case studies, and diagnostic guidelines. This supplementary data allows the RAG model to dynamically access disease-specific insights and relevant terminology, thereby enriching the language model's diagnostic narrative with evidence-based detail. By incorporating a vast amount of contextual data beyond the primary image dataset, the RAG framework achieves a nuanced understanding of each disease, enabling it to generate clinically coherent and diagnostically accurate reports.

Image Preprocessing Pipeline

The preprocessing phase focuses on preparing the X-ray images for high-performance model training and feature extraction. Given the high-dimensionality of the dataset, images are standardized by normalizing pixel intensity values, which enhances contrast and assists in consistent disease marker identification across different samples. Each image is then vectorized, reshaping it into an (1024x1024, m) matrix format to streamline input for deep learning models, while bounding box annotations are adjusted to a consistent scale, ensuring precise disease localization across images.

To enhance image quality, several preprocessing techniques are employed: histogram equalization is used to adjust image contrast, making subtle pathological details more distinguishable, while Gaussian and median filters mitigate noise, preserving critical anatomical structures. Augmentation techniques, such as random rotations, flips, and translations, introduce variety to the dataset, enhancing the model's ability to generalize across diverse imaging scenarios. Further refinements include advanced contrast adjustments through Contrast Limited Adaptive Histogram Equalization (CLAHE) and edge detection filters, which amplify anatomical contours and disease features, ensuring that key details are prominently captured during feature extraction.

By combining the curated textual dataset with a rigorously processed X-ray dataset, this study optimally supports the RAG framework and deep learning models, facilitating robust disease detection and the generation of high-quality, AI-driven diagnostic reports. This dual-data approach not only maximizes model performance but also provides the RAG model with a solid foundation to produce clinically relevant, language-rich diagnostics, ultimately aligning with radiological standards and enhancing clinical workflows.

III. METHODOLOGY

A comprehensive AI-driven framework for thoracic disease detection and reporting, leveraging advanced image processing, object detection, and natural language generation techniques. The process is divided into three stages: Image Processing and Disease Detection, Analysis and Severity Assessment, and Automated Reporting.

Stage 1: Image Processing and Disease Detection

The NIH Clinical Center's Chest X-ray dataset, encompassing 112,120 high-resolution (1024x1024) frontal-view images across 14 disease categories, serves as the data foundation for this study. The multi-label dataset allows for identifying co-occurring thoracic conditions, creating a robust framework for multi-disease detection. To ensure unbiased evaluation, data is split by patient into training, validation, and test sets.

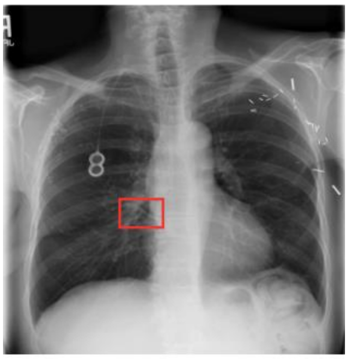

Fig 2.1. Original image with bounding box

Data Preprocessing

The preprocessing pipeline standardizes bounding box labels for consistent scale and position across images, optimizing localization accuracy. Grayscale images are vectorized, reshaping input vectors to (1024×1024, m) and bounding box annotations to (15×5, m), where m represents the number of images. Quality enhancement techniques, including histogram equalization, Gaussian and median blurring, enhance image contrast and reduce noise without sacrificing essential details. Data augmentation, comprising random rotations, flips, and translations, introduces variability, boosting model robustness. Additional preprocessing steps, such as Contrast Limited Adaptive Histogram Equalization (CLAHE) and edge detection (e.g., Canny or Sobel filters), emphasize anatomical structures to facilitate precise feature extraction.

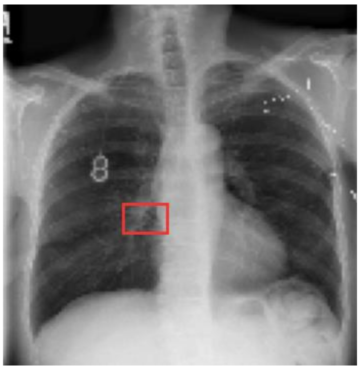

Fig 2.2. Resized image with bounding box

Model Selection and Training

The YOLO (You Only Look Once) model is employed for real-time, multi-label object detection, tailored for identifying multiple markers within a single X-ray. NasNet and EfficientNet are incorporated as complementary feature extractors, given their efficiency in processing high-dimensional chest X-ray data. These models are trained on the labeled images to maximize detection and classification accuracy across the dataset's disease categories. Validation and testing stages confirm the model's reliability in detecting anomalies and performing multi-label classification with precision.

Stage 2: Analysis and Severity Assessment

Following model training, detection outputs include bounding boxes and probability scores for each identified anomaly, with metadata on anomaly type, position, and size. This information supports severity analysis by quantifying disease extent and assessing its potential impact based on factors like size and position. The resulting data assists radiologists by providing quantitative insights, thereby enhancing diagnostic accuracy and supporting clinical decision-making.

Stage 3: Automated Reporting Using LLaMA3

Automated report generation is implemented using the LLaMA3 model within a Retrieval-Augmented Generation (RAG) framework, such as LangChain, to dynamically include relevant information for generating accurate, clinically relevant diagnostic reports. Detected anomalies, alongside patient metadata, are used to prompt the model to produce contextually aligned diagnostic language consistent with standard radiological practices.

Evaluation and Validation

Generated reports undergo rigorous evaluation against expert radiologist reviews, focusing on diagnostic accuracy, clinical relevance, and coherence. This assessment gauges the effectiveness of the combined image processing and LLaMA3 model approach in automating radiological workflows, reducing inter-radiologist variability, and enhancing diagnostic efficiency.

This methodology presents an integrated approach combining advanced image processing with LLaMA3 for automated report generation, contributing to improved diagnostic workflows, reduced radiologist workload, and AI-enhanced clinical decision support.

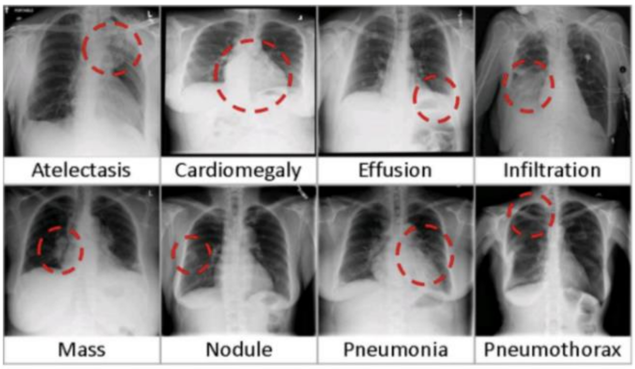

Fig.3. Thoracic Disease Visualizations

IV. RESULT AND DISCUSSION

Preliminary results indicate that the integration of X-ray image processing with LLM-based report generation shows promising outcomes. The multi-disease detection accuracy has reached competitive levels compared to existing models in the literature, and the generated diagnostic reports were found to be coherent and clinically relevant.

V. CONCLUSION

This research highlights the transformative potential of AI in medical imaging, particularly in automating the diagnostic process and reducing human error. The combination of deep learning models for image processing and LLMs for report generation has shown significant promise in improving diagnostic workflows, ultimately contributing to better patient outcomes.

ACKNOWLEDGMENT

We sincerely thank ICMR-NIRT Chennai for the insightful visit, with special appreciation to Dr. C. Ponnuraja for his informative session on tuberculosis and intern Krishna for his valuable insights on X-ray diagnosis, both of which have significantly guided our project team. Additionally, we acknowledge the wealth of information made available through online research platforms such as Radiopaedia and ScienceDirect, which provided numerous peer-reviewed articles and medical case studies. These resources have been instrumental in supplementing our dataset, particularly for the Retrieval-Augmented Generation (RAG) model, enabling a more comprehensive and evidence-based approach to our research.

REFERENCES

[1] S. Kumar Satapathy, S.-B. Cho, S. Mishra, S. Sah, Kamaldeep, and S. Nandan Mohanty, “A federated learning approach for classifying chest diseases from chest X-ray images,” *Biomedical Signal Processing and Control*, vol. 100, p. 107107, Feb. 2025, doi: 10.1016/j.bspc.2024.107107.

[2] D. Ueda *et al.*, “A deep learning-based model to estimate pulmonary function from chest x-rays: multi-institutional model development and validation study in Japan,” *The Lancet Digital Health*, vol. 6, no. 8, pp. e580–e588, Aug. 2024, doi: 10.1016/S2589-7500(24)00113-4.

[3] Mr. Z. He, Dr. H. A. Cheung, Dr. T. Hei, E. Wong, and Dr. J. S. Yoo, “Enhancing Chest X-Ray Report Generation in Radiology through Deep Learning: Leveraging Keywords from Existing Reports and Multi-Label Classification,” *Journal of Medical Imaging and Radiation Sciences*, vol. 55, no. 3, p. 101657, Oct. 2024, doi: 10.1016/j.jmir.2024.101657.

[4] R. Bigolin Lanfredi, P. Mukherjee, and R. Summers, “Enhancing chest X-ray datasets with privacy-preserving large language models and multi-type annotations: A data-driven approach for improved classification,” *Medical Image Analysis*, p. 103383, Nov. 2024, doi: 10.1016/j.media.2024.103383.

[5] R. Dharmesh Ishwerlal, R. Agarwal, and K. S. Sujatha, “Lung disease classification using chest X ray image: An optimal ensemble of classification with hybrid training,” *Biomedical Signal Processing and Control*, vol. 91, p. 105941, May 2024, doi: 10.1016/j.bspc.2023.105941.

[6] Y. Jin, H. Lu, W. Zhu, and W. Huo, “Deep learning based classification of multi-label chest X-ray images via dual-weighted metric loss,” *Computers in Biology and Medicine*, vol. 157, p. 106683, May 2023, doi: 10.1016/j.compbiomed.2023.106683.

[7] J. Lian, Z. Dong, H. Zhang, Y. Chen, and J. Liu, “Multifocal region-assisted cross-modality learning for chest X-ray report generation,” *Computers in Biology and Medicine*, vol. 183, p. 109187, Dec. 2024, doi: 10.1016/j.compbiomed.2024.109187.

[8] A. Giełczyk, A. Marciniak, M. Tarczewska, and Z. Lutowski, “Pre-processing methods in chest X-ray image classification,” *PLOS ONE*, vol. 17, no. 4, p. e0265949, Apr. 2022, doi: 10.1371/journal.pone.0265949.