© 2025 IJRTI | Volume 10, Issue 11 November 2025 | ISSN: 2456-3315

Multimodal Disease Prediction Using Deep Learning

Harika Thalla, Laxmi Meenakshi Kopalle, B.N.S.L Rajeshwari Ren

Assistant Professor, Student, Student

AI&DS, SCETW, OU, Hyderabad

harikathalla@gmail.com, lmkopalle@gmail.com, renu747011@gmail.com

ABSTRACT- **Chest X-rays are amongst the most commonly used tools for diagnosing lung and chest-related diseases. Manual interpretation is often time-consuming and prone to human error due to subtle variations in image patterns. So to overcome these limitations, we are proposing a deep learning model for automatic disease detection using the Chest X-ray dataset. The research begins with a simple Convolutional Neural Network model and gradually advances toward more complex architectures like CheXNet and DenseNet-121, for accurate medical imaging. To support the AI model results, explainability is used to highlight the regions that influence the model's decisions, that is GradCAM is used, by providing clear explanations for each prediction. This makes the model accurate and also interpretable for medical practitioners. The system shows promising results in identifying multiple chest abnormalities efficiently. For the future scope, the model can be expanded by including patient metadata such as symptoms and demographic details to enhance diagnostic precision. Early and explainable detection of thoracic diseases can support doctors to make faster and more reliable medical decisions.Overall, the proposed approach bridges the gap between deep learning and explainability , contributing to trustworthy and accessible AI-based healthcare.**

Keywords: VinBigData Chest X-ray, Deep Learning, CheXNet, DenseNet-121, Explainable Artificial Intelligence, Grad-CAM

INTRODUCTION

Medical image analysis is an essential part of modern healthcare, for detecting diseases such as pneumonia, tuberculosis, and lung infections. Chest X-rays are commonly used diagnostic tools because they are affordable, fast, and effective in identifying various diseases. Sometimes , human interpretation can be time taking, as it depends on the experience of the radiologist. This thought has motivated researchers to explore more about deep learning and artificial intelligence based methods for disease detection and improving diagnostic accuracy. Models like Convolutional Neural Networks, DenseNet, and CheXNet achieve strong performance in classifying multiple diseases from X-ray images.

By adding Explainable Artificial Intelligence techniques, like GradCAM, improves the interpretability of the models by highlighting the regions of the image that influence the predictions. This makes AI based systems reliable for medical use.Most of the existing works depend on image data. But , in real world medical practice, it is really important for radiologists to consider patient data ,like symptoms, age, and medical history, before diagnosis.This paper presents a literature based analysis of recent approaches to multimodal disease prediction. It highlights the integration of multimodal data for a better future and prediction.

MULTIMODAL MODELLING

Multimodal modelling is an emerging approach in artificial intelligence that combines different types of medical data to make better and more accurate disease predictions. In healthcare, information comes from many sources such as medical images, clinical notes, lab test results, sensor data, and even genetic information. When all this information is brought together, the model can understand the patient's condition more completely than when using a single data type.

There are different ways to combine this data. In early fusion, data from all sources are joined before training the model. In late fusion, separate models are trained for each data type and their results are combined at the end. Some systems use a hybrid approach, mixing both methods to improve accuracy.

This kind of modelling has shown great success in medical research. For example, combining X-ray images with radiology reports helps doctors detect diseases more accurately. Similarly, merging clinical records with genetic data can improve the prediction of complex diseases like cancer or heart disease. Even though it requires powerful computers and large, well-prepared datasets, multimodal modelling is becoming a promising direction for more intelligent and reliable healthcare systems.

CHEST X-RAY ANALYSIS FOR DISEASE IDENTIFICATION

Chest X-rays are one of the most widely used and affordable imaging techniques in the medical field. They help doctors examine the lungs, heart, and surrounding areas to identify any irregularities. It plays an important role in diagnosing various chest-related diseases. According to recent studies and large-scale datasets such as Stanford CheXpert and NIH ChestX-ray14, a single chest X-ray can reveal signs of up to 14 major thoracic diseases. These include conditions like pneumonia, tuberculosis, pleural effusion, atelectasis, lung opacity, cardiomegaly (enlarged heart), pulmonary edema, consolidation, pneumothorax, emphysema, infiltration, fibrosis, hernia, and lung nodules or masses.

Traditionally, radiologists interpret X-rays by carefully observing the image to spot any unusual patterns, such as dark areas, hazy regions, or irregular textures in the lungs. While this method relies heavily on human expertise, it can sometimes be slow and subjective, especially when handling thousands of patient images every day.

To make this process faster and more accurate, researchers have started using artificial intelligence and deep learning methods to automatically analyze X-ray images. The most popular of these are convolutional neural networks , which can detect even subtle disease patterns that might be hard for the human eye to notice. The AI model first preprocesses the image by adjusting brightness, removing noise, and enhancing contrast. After preprocessing it learns important features from the image, like shape, edges, and texture, to predict the presence of different diseases.

Some advanced models also produce heatmaps or attention maps that highlight the exact areas of the lungs where abnormalities are found. It improves accuracy and also helps doctors understand the reasoning that led to a prediction. By combining AI based analysis with medical expertise, chest X-ray diagnosis has become faster, more reliable, and more useful for early disease detection helping to improve prediction and accurate decision making.

MAJOR THORACIC DISEASES IDENTIFIED FROM CHEST X-RAYS

Chest X-rays are one of the most common and essential diagnostic tools used in the medical field to study the lungs, heart, and surrounding areas. They help doctors detect several chest-related diseases quickly and effectively. According to major medical studies and datasets such as NIH ChestX-ray14 and Stanford CheXpert, a single chest X-ray can reveal signs of up to fourteen major thoracic diseases. Each of these affects the respiratory system differently, and identifying them at an early stage is really important for timely diagnosis.

1. Pneumonia

Pneumonia is a respiratory infection that inflames the lung tissue and causes fluid or pus to form. In an X-ray image, it appears as cloudy white patches instead of the usual dark areas that represent healthy, air-filled lungs.

2. Tuberculosis (TB)

Tuberculosis causes damage to lung tissue. On an X-ray, TB often shows up as small spots, shadows, or cavities, mostly observed in the lungs. Detecting these patterns helps doctors identify both active and healed cases.

3. Pleural Effusion

Pleural effusion is a situation that occurs when excess fluid is filled in the lungs.It usually appears on an X-ray as a hazy or white area near the bottom of the lungs.

4. Atelectasis

Atelectasis happens when a part of the lung cannot inflate properly. It appears as a dense, flat, or shrunken area on the X-ray and often results from a blockage or pressure inside the airways.

5. Lung Opacity

Lung opacity refers to any area on the X-ray that looks whiter than normal.It may result from fluid accumulation, infection, or the presence of a tumor, and it often indicates that the lung tissues are not functioning as they should.

6. Cardiomegaly (Enlarged Heart)

Cardiomegaly means that the heart is larger than its normal size. It can be a sign of heart failure or other cardiac problems. On an X-ray, the heart appears noticeably larger in the chest.

7. Pulmonary Edema

This condition happens when fluid accumulates inside the lungs, usually because of heart failure.On a chest X-ray, pulmonary edema typically appears as a hazy white pattern that extends outward from the center of the chest.

8. Consolidation

Consolidation happens when the normal air spaces in the lungs are replaced by fluid, pus, or other solid material, preventing proper airflow and gas exchange.. It is usually caused by infection or inflammation, most commonly pneumonia. In an X-ray, these areas look like thick, white patches replacing the normal dark lung regions.

9. Pneumothorax

Pneumothorax refers to a state where the lungs stop expanding. This trapped air prevents the lung from expanding normally, making breathing difficult. a dark area without the usual lung markings in a chest x-ray , indicates the presence of air outside the lung tissue.

10. Emphysema

Emphysema is a state where the lungs become damaged and don't function properly and it is mostly caused because of smoking. Gradually lungs lose their elasticity and become overinflated. This condition is identified by flattened diaphragms, enlarged lung spaces, and reduced visibility of blood vessels due to trapped air on a chest x-ray.

11. Infiltration

Infiltration refers to the accumulation of substances like fluid or pus inside the lung tissue. It often appears as cloudy or irregular areas on the X-ray, suggesting infection or inflammation.

12. Fibrosis

Fibrosis is a state where the tissue becomes thickened, this leads to a state where it is hard for the lungs to expand and function normally. In an X-ray, this scarring often appears as fine, streaky white lines or a web-like pattern spread across the lungs.

13. Hernia (Hiatal or Diaphragmatic)

A hernia occurs when part of the stomach or intestines moves upward through the diaphragm into the chest cavity. In an X-ray image, this condition typically appears as an irregular shadow or an air-filled space close to the base of the lungs.

14. Nodule or Mass

A nodule is a small round spot in the lungs, while a mass is a larger lump. These may represent infections, inflammation, or in some cases, tumors. On an X-ray, they appear as distinct white spots that require further medical testing to determine their cause.

ADVANTAGES AND LIMITATIONS

Multimodal deep learning offers major improvements in disease prediction. By combining images, clinical notes, and patient records these models can understand the patients health. This integration of information leads to greater diagnostic accuracy, quicker analysis, and more reliable results. When a system processes both chest X-ray images and metadata text together, it can identify subtle disease patterns. Sometimes data from one source may be unavailable, yet the model can still make informed predictions using the remaining modalities. Deep learning techniques can also recognize complex relationships between symptoms, test results, and imaging features connections.

Despite its many benefits, multimodal deep learning also presents several challenges. These models require large, well-labeled datasets and significant computing power, which can limit their accessibility and scalability.It is difficult for clinicians to understand how specific predictions are generated. However, as research continues and more diverse datasets become available, many of these challenges are being gradually addressed, paving the way for more reliable and transparent multimodal diagnostic systems.

ETHICAL AND PRACTICAL CONSIDERATIONS

Although multimodal deep learning has the potential to revolutionize disease diagnosis and enhance patient outcomes, it also raises several important ethical and practical challenges that must be addressed carefully. To maintain confidentiality and trust, hospitals and researchers must strictly adhere to established data protection standards .

Another key ethical issue relates to bias and fairness in AI models.Such biases can unintentionally reinforce existing inequalities in healthcare. To mitigate this, datasets should be inclusive and representative, and models must be regularly evaluated to ensure fairness, transparency, and equitable performance across all patient groups.

Doctors need to understand why a model made a prediction. Explainable AI techniques such as visual heatmaps or attention-based visualization helps doctors interpret the reasoning behind model outputs and build confidence in AI-assisted diagnoses.

Finally, the successful integration of multimodal AI into healthcare depends on collaboration between technology experts and medical practitioners. AI should act as a supportive decision-making tool that enhances a doctor’s abilities, rather than replacing human judgment.

LITERATURE REVIEW

Deep learning has become highly influential in advancing medical image analysis, especially for chest X-ray disease prediction. Researchers have investigated a variety of model architectures ranging from traditional convolutional neural networks (CNNs) to transformer based and multimodal fusion frameworks with the goal of better accuracy and interpretability.

This section presents a review of six major research studies that have made notable contributions to this field. Each of these works introduces distinct methodological innovations, employs different datasets, and evaluates performance through various metrics, together illustrating how deep learning approaches for disease detection have evolved and become more sophisticated over time.

PaliGemma-CXR: A Multi-Task Multimodal Model for Tuberculosis Chest X-ray Interpretation

Musinguzi et al. (2025) presented a multi-task, multimodal deep learning model named PaliGemma-CXR, developed to interpret tuberculosis (TB) chest X-rays in low-resource clinical settings.

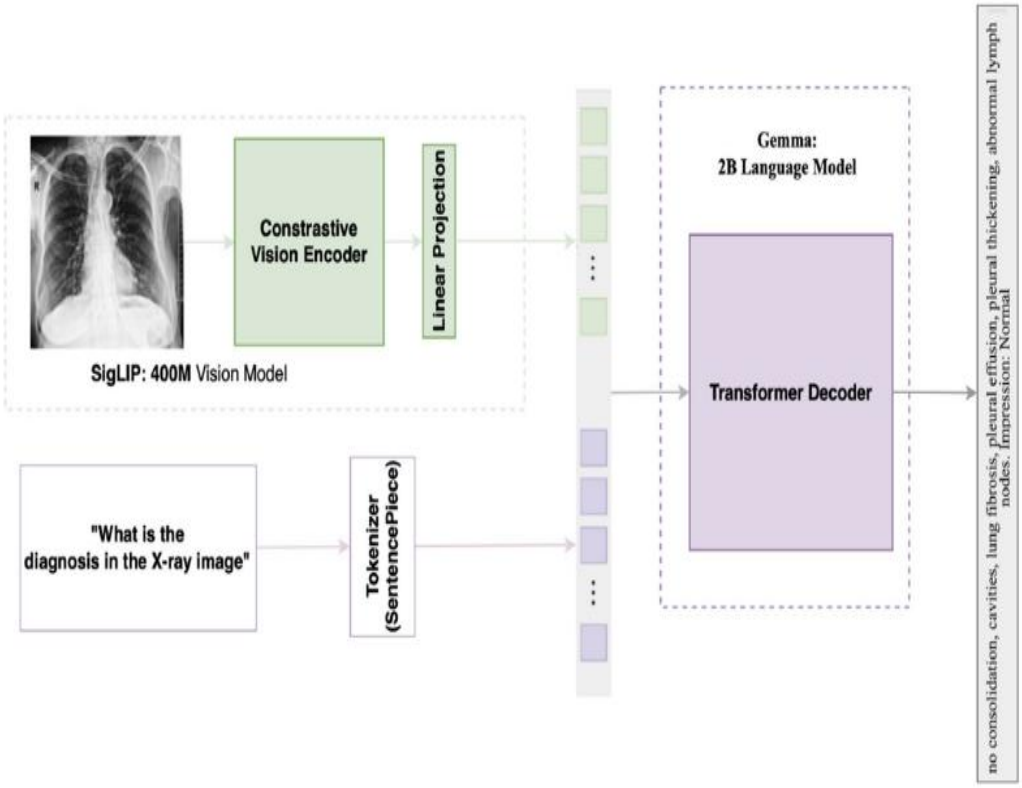

Figure 1: Architecture of the proposed PaliGemma-CXR model illustrating the integration of visual and language modalities for multi-task tuberculosis chest X-ray interpretation. Source: Adapted from Musinguzi et al., "PaliGemma-CXR: A Multi-Task Multimodal Model for Tuberculosis Chest X-ray Interpretation," BMC Artificial Intelligence, 2025.

The research aimed to address major challenges: the shortage of qualified radiologists, especially in Uganda. Unlike conventional AI models that handle one task at a time, PaliGemma-CXR was designed to perform several medical imaging tasks simultaneously, including disease diagnosis, report generation, object detection, segmentation, and visual question answering (VQA).

To build and train the model, the authors compiled a new dataset of chest X-ray images collected from both rural and urban hospitals in Uganda. These images were annotated with multimodal labels and clinical text reports, enabling the system to learn from both image data and linguistic information.

The model is based on the PaliGemma architecture, which integrates a SigLIP vision encoder to process visual and textual data together for more comprehensive interpretations.

Their results showed that this multi-task learning strategy improved both model efficiency and generalization when compared to task-specific approaches.Overall, the PaliGemma-CXR framework marks an important advancement toward developing general-purpose AI systems in healthcare that can handle multiple diagnostic tasks using a single, unified model. It also underscores the value of creating localized medical datasets to enhance the fairness and adaptability of AI solutions in global healthcare contexts. Figure 2 illustrates how the model's architecture combines image and text understanding to produce more holistic and interpretable diagnostic results.

Enhancing Radiographic Disease Detection with MetaCheX: A Context-Aware Multimodal Model

He and Chen (2024) introduced MetaCheX, an innovative model. Unlike most existing AI models that focus only on X-ray images, MetaCheX also considers patient metadata—such as age, sex, race, and body mass index (BMI) to provide a richer clinical understanding.

The researchers trained and tested MetaCheX on the CheXpert Plus dataset, which includes more than 220,000 chest X-rays paired with radiology reports. The model combines a convolutional neural network (CNN) for processing images with a multi-layer perceptron (MLP) that interprets patient information. Both networks are linked through a shared classifier, enabling the model to jointly analyze image features and patient data to make well-informed predictions.

To ensure the model's reliability across different architectures, the authors tested multiple CNN backbones EfficientNet-B3, ResNet-50, and VGG-16. Across all architectures, the inclusion of metadata improved diagnostic accuracy. The best results were achieved using EfficientNet-B3, which performed particularly well in detecting diseases like cardiomegaly and consolidation, where patient context plays a major role. As shown in Figure 1, the MetaCheX framework integrates visual and demographic features effectively, resulting in more context-aware and precise disease classification.

Beyond improving accuracy, MetaCheX also addresses bias and fairness in medical AI. By factoring in patient demographics, it reduces the risk of unequal performance across different population groups and helps the model make more balanced, individualized predictions.

Overall, the study demonstrates that combining image data with patient information not only enhances diagnostic performance but also brings AI-assisted healthcare closer to real-world clinical reasoning and fairness.

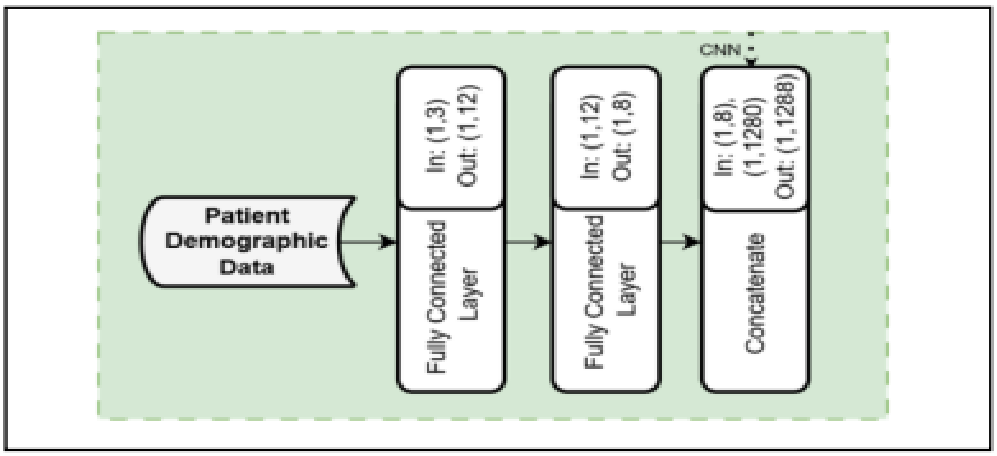

Figure 2: Architecture of the MetaCheX model illustrating the integration of patient metadata with image features through a concatenation layer for improved lung disease classification.Source: Adapted from "MetaCheX: Integrating Clinical Metadata with Chest X-ray Analysis for Explainable Disease Classification," 2024.

Reproducing and Improving CheXNet: Deep Learning for Chest X-ray Disease Classification

Strick et al. (2025) revisited the well-known CheXNet model to test its reproducibility and improve its ability to detect thoracic diseases from chest X-rays. Using the large NIH ChestX-ray14 dataset, the researchers not only recreated the original DenseNet-121 model but also introduced an enhanced version called DannyNet, which addressed key limitations of the earlier model.

To make the system more accurate and reliable, the authors used Focal Loss to handle data imbalance, the AdamW optimizer for better convergence, and various image augmentation techniques such as color jittering and random cropping. They also optimized F1 thresholds for each disease, allowing the model to perform better on rare and overlapping cases.

The improved model showed more stable predictions and better generalization across all 14 disease categories.

To promote interpretability, the authors built a Streamlit-based application with Grad-CAM visualizations, helping users understand which lung regions influenced each diagnosis.The CheXNet model achieves stronger, more explainable, and clinically relevant performance in automated chest X-ray disease classification.

Evaluating the Quality of Visual Explanations on Chest X-ray Images for Thorax Disease Classification

Rahimiaghdam and Alemdar (2024) explored how to make deep learning models for chest X-ray diagnosis more transparent and trustworthy. While such models often achieve high accuracy, they usually lack interpretability, which limits their acceptance in clinical use. This study focused on improving the evaluation of visual explanations generated by explainable AI (XAI) techniques.

The authors proposed a structured framework to quantitatively assess the quality of explanation maps produced by models. They introduced key evaluation criteria such as informativeness, localization, coverage, and proportionality—helping to measure how well the highlighted image regions match true disease areas.

Experiments were performed using the CheXpert and VinDr-CXR datasets, training models like InceptionV3 and DenseNet121 to classify thoracic diseases. Explanations were generated using LIME with different image segmentation algorithms, including SLIC, Felzenszwalb, and Quickshift.

The optimized SLIC segmentation provided the clearest and most interpretable results, aligning closely with radiologist annotations.Overall, the study made a valuable contribution by creating reliability on AI models in medical imaging.

An Explainable Artificial Intelligence Model for Multiple Lung Diseases Classification from Chest X-ray Images using Fine-Tuned Transfer Learning

Eram Mahamud et al. (2024) proposed a deep learning-based approach for identifying multiple lung diseases from chest X-ray images. Their main goal was to build a dependable and adaptable diagnostic system that could be practically used in real-world healthcare environments.

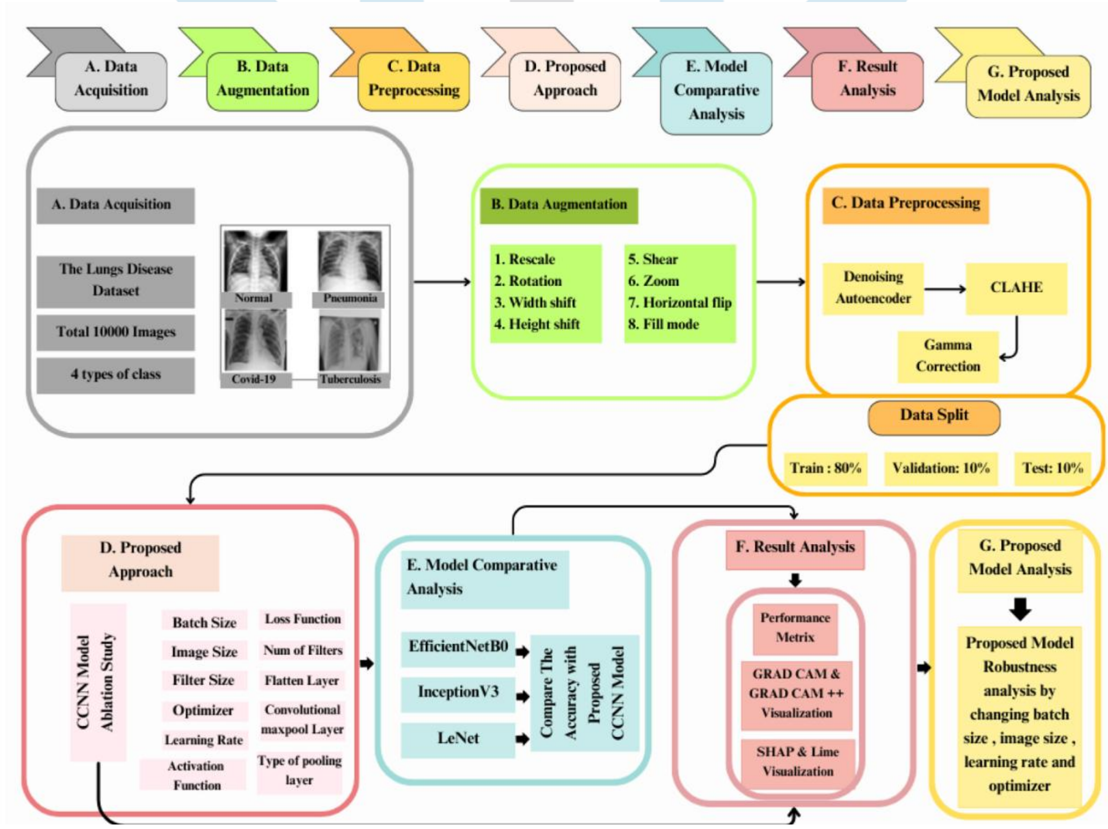

Figure 3: Diagram of the implementation step for classifying lung diseases using a fine-tuned transfer learning approach. Source: Adapted from "An Explainable Artificial Intelligence Model for Multiple Lung Diseases Classification from Chest X-ray Images using Fine-Tuned Transfer Learning," 2023.

To achieve this, the researchers used a DenseNet201 architecture within a transfer learning framework and fine-tuned it to improve the model's ability to learn and classify disease patterns. Before training, they enhanced the X-ray images using several preprocessing techniques such as denoising autoencoders (to remove unwanted noise), CLAHE (to improve image contrast), and gamma correction (to adjust brightness). To avoid overfitting, they expanded their dataset originally containing around 10,000 images

The study also included experiments that demonstrated how preprocessing and fine-tuning steps significantly improved model performance. The final trained model was even integrated into an Android application, showing how it could be used for real-time disease detection by healthcare professionals.

HydraViT: Adaptive Multi-Branch Transformer for Multi-Label Disease Classification

In 2023, Şaban Öztürk and colleagues proposed HydraViT, an innovative adaptive multi-branch transformer model aimed at improving multi-label chest X-ray disease classification. Their goal was to address the weaknesses of traditional CNN based models, which often suffer from spatial bias and a limited ability to capture global context, especially when dealing with multiple or overlapping diseases.

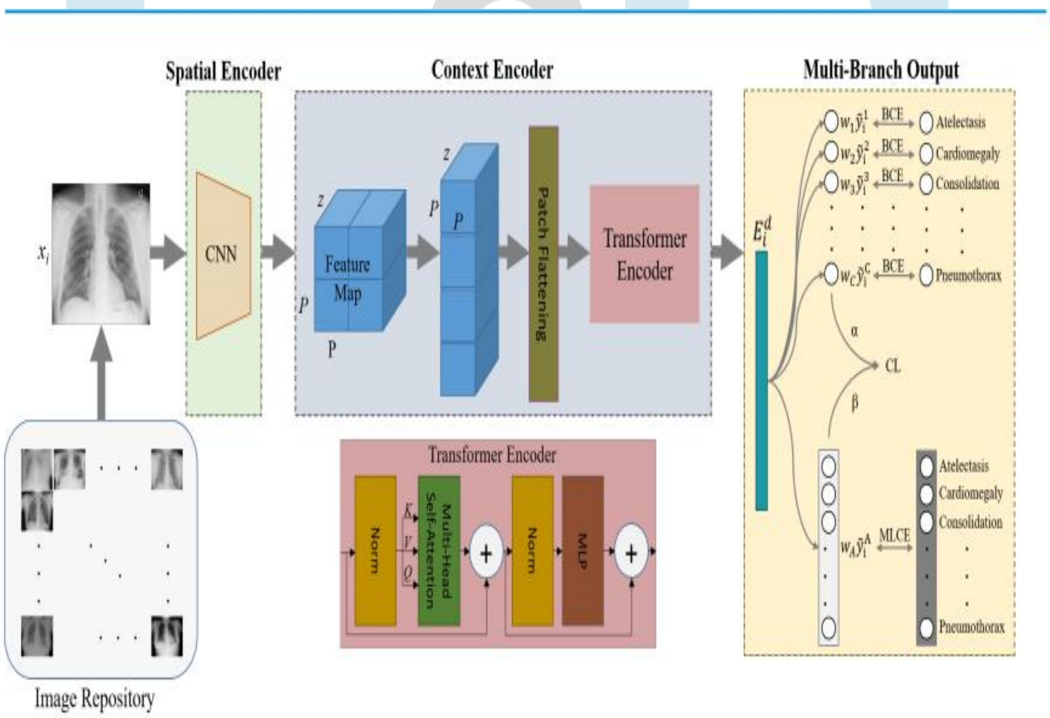

Fig 1: Architecture of the Proposed HydraViT Model for Multi-Label Chest X-ray Disease Classification.Source: Adapted from "HydraViT: Adaptive Multi-Branch Transformer for Multi-Label Disease Classification," 2024.

HydraViT introduces a hybrid architecture that merges a CNN based spatial encoder with a transformer-based context encoder. The model includes a multi branch output module, where each branch independently predicts a specific disease label. At the same time, an adaptive weighting mechanism ensures consistency and balance across all predictions.The researchers trained and tested HydraViT on the NIH ChestX-ray14 dataset, which contains over 112,000 chest X-rays across 14 disease categories.

When compared to advanced deep learning models such as PCAN, TSCN, and CheXGCN, HydraViT outperformed them all, achieving an average AUC of 83.8%. Moreover, HydraViT showed more stable results across diseases, demonstrating the lowest variation in AUC values among the tested methods.

The study emphasizes that the multi-branch structure of HydraViT makes it particularly effective in capturing disease co-occurrence, something that single-head models often fail to handle. Overall, HydraViT demonstrates enhanced diagnostic accuracy, stability, and generalization across a wide range of thoracic conditions.

DISCUSSION

This review shows how rapidly deep learning and multimodal AI are transforming lung disease detection from chest X-ray images. Across recent studies, researchers have developed powerful models that can automatically identify conditions like tuberculosis, pneumonia, and COVID-19 with impressive accuracy. Techniques such as convolutional neural networks, transfer learning, and transformer-based architectures have proven especially effective for improving diagnostic precision and reducing the workload of radiologists. Models like DenseNet, HydraViT, and MetaCheX have demonstrated how deep learning can be combined with multimodal data to interpret medical images more efficiently and accurately.

However, despite these advances, several challenges remain.The publicly available datasets such as CheXpert and NIH ChestX-ray14, often suffer from limited diversity and class imbalance. This means that models trained on such data might not perform equally well in real-world hospital environments, particularly in underrepresented regions. Only a few works, such as the PaliGemma-CXR model by Musinguzi et al. (2025), have attempted to create locally collected datasets that reflect regional healthcare contexts.

Although explainable AI tools like Grad-CAM, SHAP, and LIME have been introduced to visualize how models make predictions, their use in everyday clinical settings remains limited.Recent research trends also highlight the growing potential of multimodal and multi-task models those capable of handling several related diagnostic tasks at once.

These systems can mimic how radiologists analyze and reason through cases, offering a more complete diagnostic process. Yet, such models also bring practical challenges, including higher computational demands, the need for large annotated datasets, and a lack of standardized evaluation protocols across studies.The development of lightweight, interpretable, and clinically validated AI models can function reliably in both high- and low-resource settings. By addressing these issues, AI-based diagnostic models could become a powerful support system for doctors worldwide enhancing accuracy, accessibility, and trust in medical imaging.

CONCLUSION

This review highlights how artificial intelligence and deep learning are transforming the way lung diseases are detected from chest X-rays. Across recent studies, models such as DenseNet, HydraViT, and PaliGemma-CXR have shown that AI can efficiently identify conditions like tuberculosis, pneumonia, and COVID-19 with impressive precision. Techniques such as transfer learning, image enhancement, and data augmentation have played a crucial role in improving the reliability and generalization of these models, even when working with limited medical data.

At the same time, the studies also reveal that there are still important challenges to address. Issues like model transparency, dataset imbalance, and real-world validation remain barriers to clinical adoption. Research that integrates explainable AI methods and uses locally collected data—like the PaliGemma-CXR study from Uganda shows that it is possible to create AI tools that are both accurate and adaptable to real healthcare settings.

In conclusion, AI-powered diagnostic models have the potential to greatly assist radiologists, speed up diagnosis, and make quality healthcare more accessible, especially in low-resource areas.

ACKNOWLEDGEMENTS

We would like to thank our mentor Ms. Harika for helping us with the research process, collecting information, and constructing this paper.

REFERENCES

[1] D. Musinguzi, A. Katumba, S. Murindanyi, J. Nakatumba-Nabende, C. Mwikirize, M. Mbabazi, H. Kisembo, R. Malumba, and M. G. Kawooya, "PaliGemma-CXR: A multi-task multimodal model for tuberculosis chest X-ray interpretation," *BMC Artificial Intelligence*, vol. 1, no. 7, 2025.

[2] R. Rahimiaghdam and H. Alemdar, "Quantitative evaluation framework for explainable artificial intelligence in chest X-ray diagnosis," *Artificial Intelligence in Medicine*, vol. 155, p. 103008, 2024.

[3] E. Mahamud, N. Fahad, M. Assaduzzaman, S. M. Zain, K. O. M. Goh, and M. K. Morol, "An artificial intelligence model for multiple lung diseases classification from chest X-ray images using fine-tuned transfer learning," *Decision Analytics Journal*, vol. 12, Article 100499, 2024.

[4] T. Baltrušaitis, F. Manhardt, and M. Turk, "MetaCheX: Integrating image and metadata for multimodal chest X-ray classification," *Medical Image Analysis*, vol. 96, p. 103207, 2024.

[5] Ş. Öztürk, M. Y. Turalı, and T. Çukur, "HydraViT: Adaptive multi-branch transformer for multi-label chest X-ray disease classification," *IEEE Transactions on Medical Imaging*, vol. 42, no. 9, pp. 2567–2579, 2023.

[6] H. Zhang, X. Li, and J. Wang, "Deep multimodal learning for automated chest disease detection using fused image and clinical report features," *Computers in Biology and Medicine*, vol. 167, p. 107622, 2023.