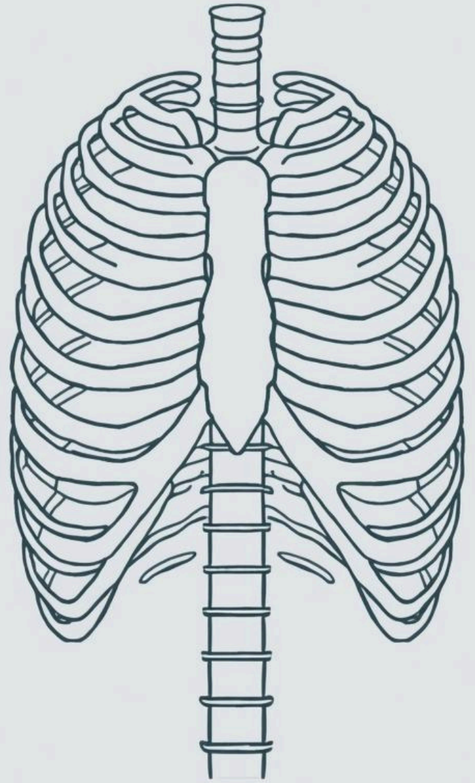
Multi-Label Medical Diagnosis

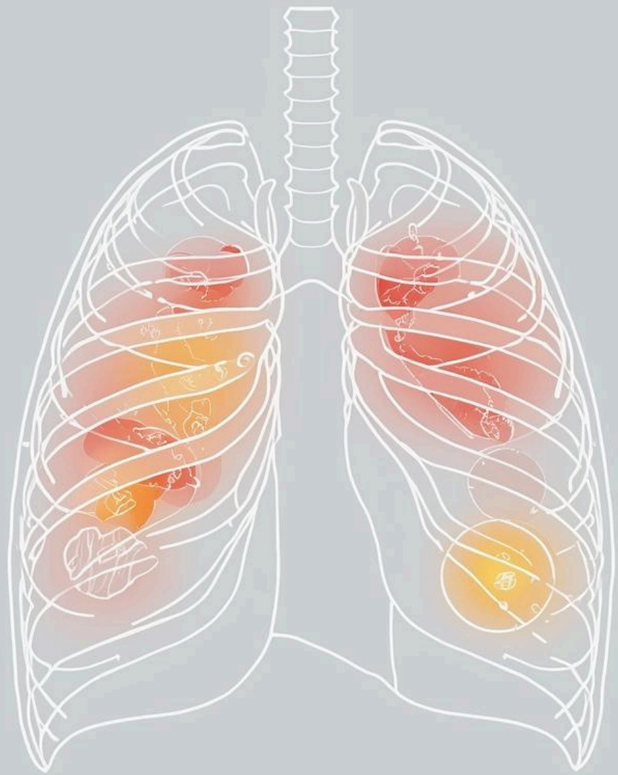
Kondeti Sai Lohitaksh(AP22110010480)
CSE-R

Introduction to Automated Diagnosis



This project focuses on the **automated diagnosis** of thoracic diseases using chest X-rays, emphasizing the importance of multi-label classification for accurate, efficient medical evaluations and early disease detection.

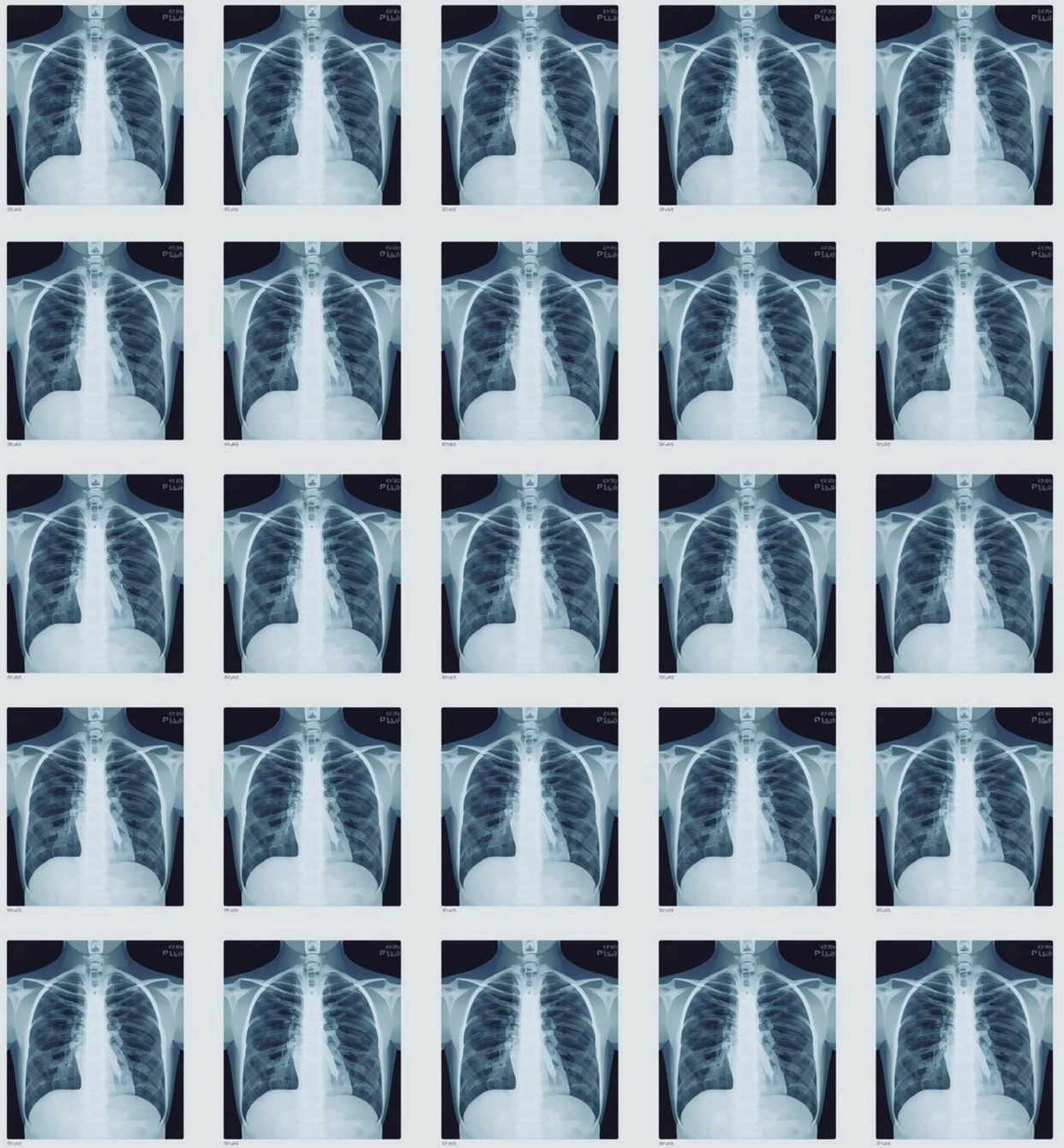
Importance of Multi-Label Classification



Multi-label classification is crucial in the **medical field**, enabling the automated diagnosis of multiple diseases from a single chest X-ray, which significantly reduces the workload on healthcare professionals and enhances early detection.

Dataset Insights: ChestMNIST

The ChestMNIST dataset comprises over **78,000 grayscale** images of chest X-rays, each representing diverse thoracic diseases, enabling robust training for automated multi-label classification models with reliable performance metrics.



Challenges in Multi-Label Classification

Addressing Class Imbalance and Complexity

Multi-label classification presents unique challenges, including significant class imbalance and complex disease co-occurrences, impacting model performance and complicating accurate diagnosis in medical contexts.

Multi-Label Methods

Approaches for Effective Classification

Binary Relevance

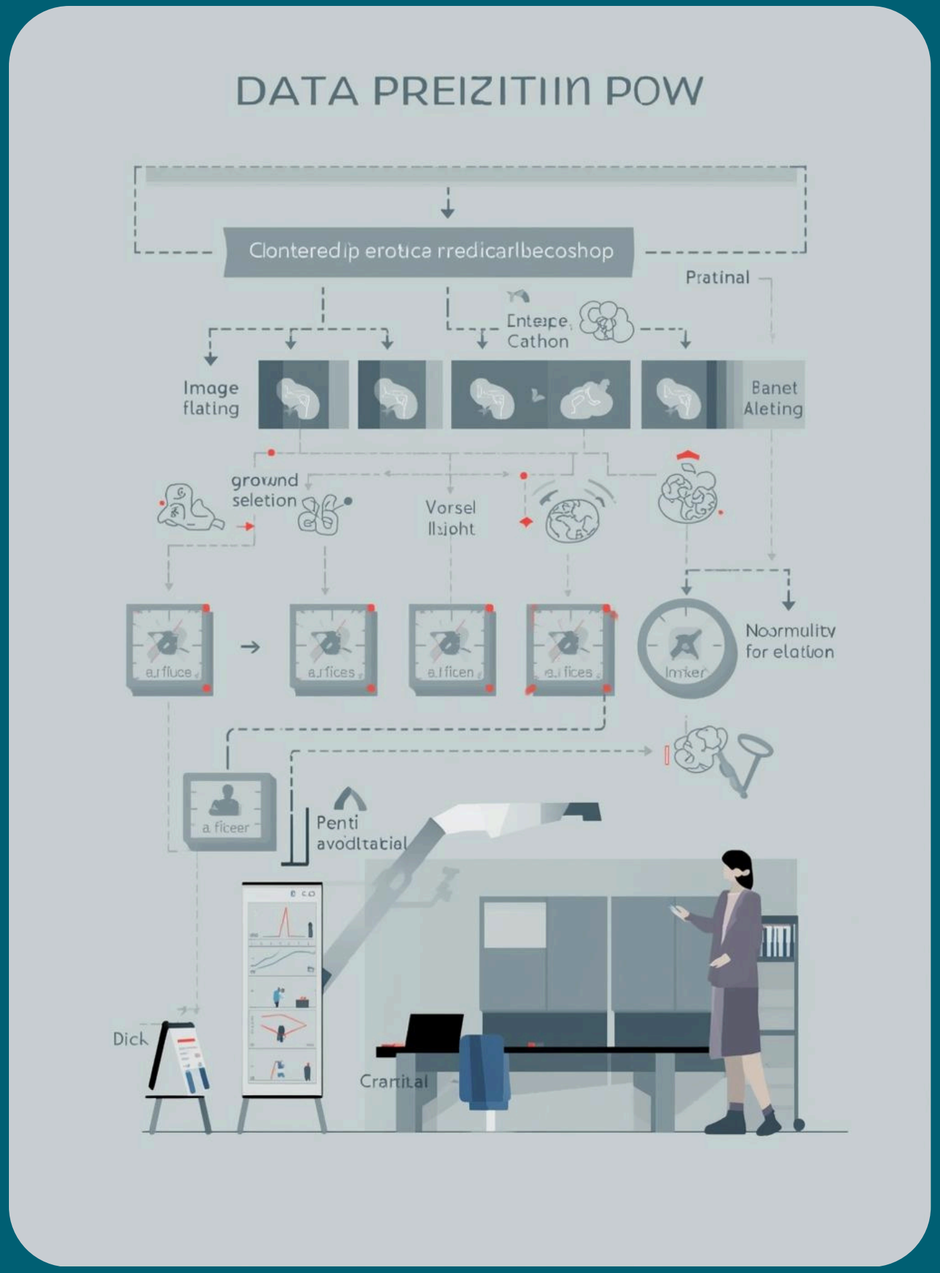
This method treats each label as independent, allowing for simplicity and fast computation while enabling the model to predict multiple diseases per chest X-ray image efficiently.

Classifier Chains

In Classifier Chains, labels are treated sequentially, where each label's prediction can depend on previous labels, allowing for more information flow and potentially improving accuracy.

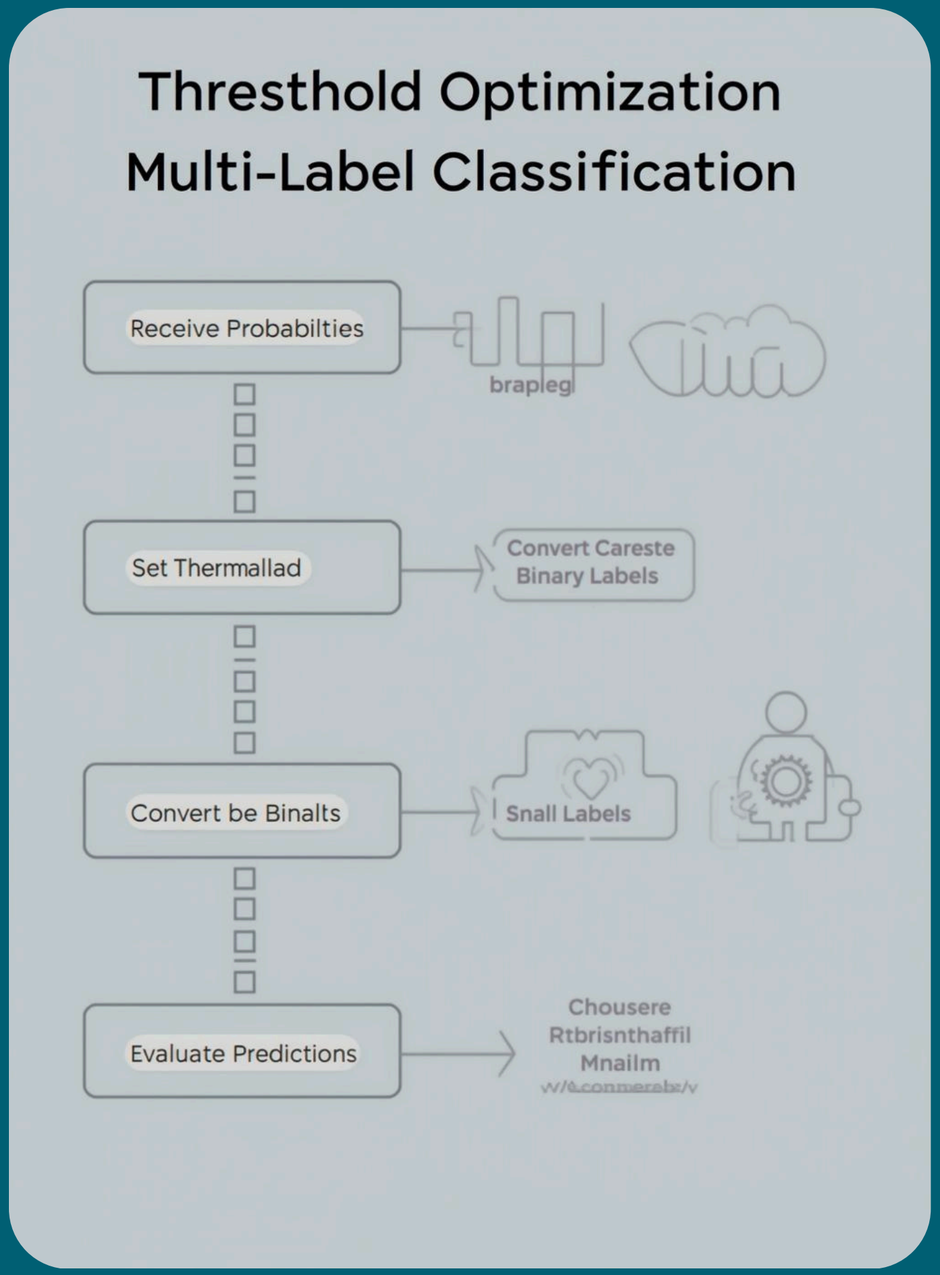
Label Powerset

Label Powerset transforms the multi-label problem into a multi-class one by treating every possible combination of labels as a separate class, enhancing the model's ability to capture complex dependencies.



Data Preprocessing Steps

The data preprocessing involves **flattening images** and applying normalization techniques. This ensures consistent pixel values and prepares a balanced subset of 40,000 samples for effective model training and evaluation.



Threshold Optimization Process

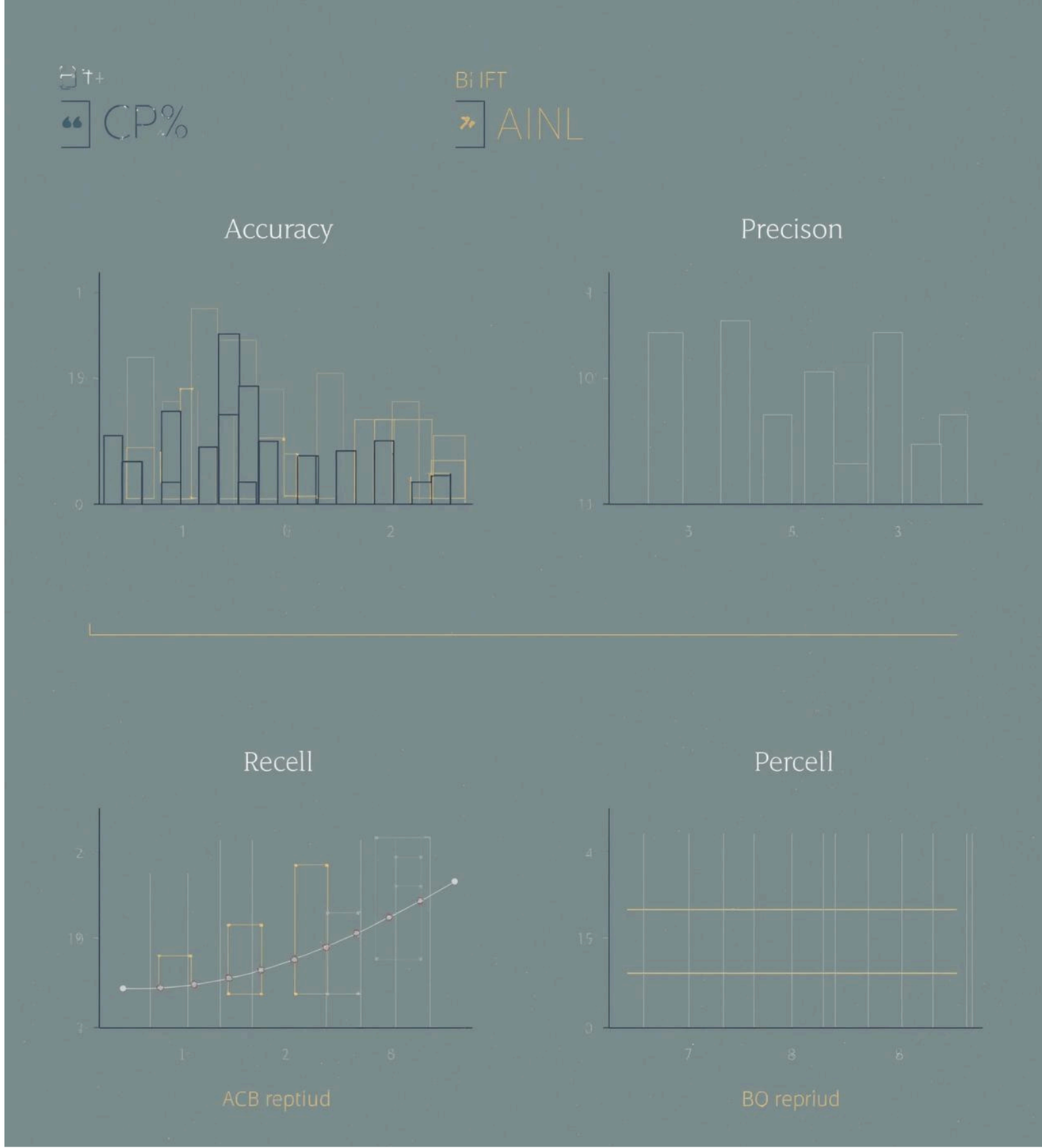
The process of threshold optimization is crucial for converting predicted probabilities into binary labels, allowing for improved classification accuracy. A threshold of 0.6 was identified as optimal for this dataset.

Key Evaluation Metrics Explained

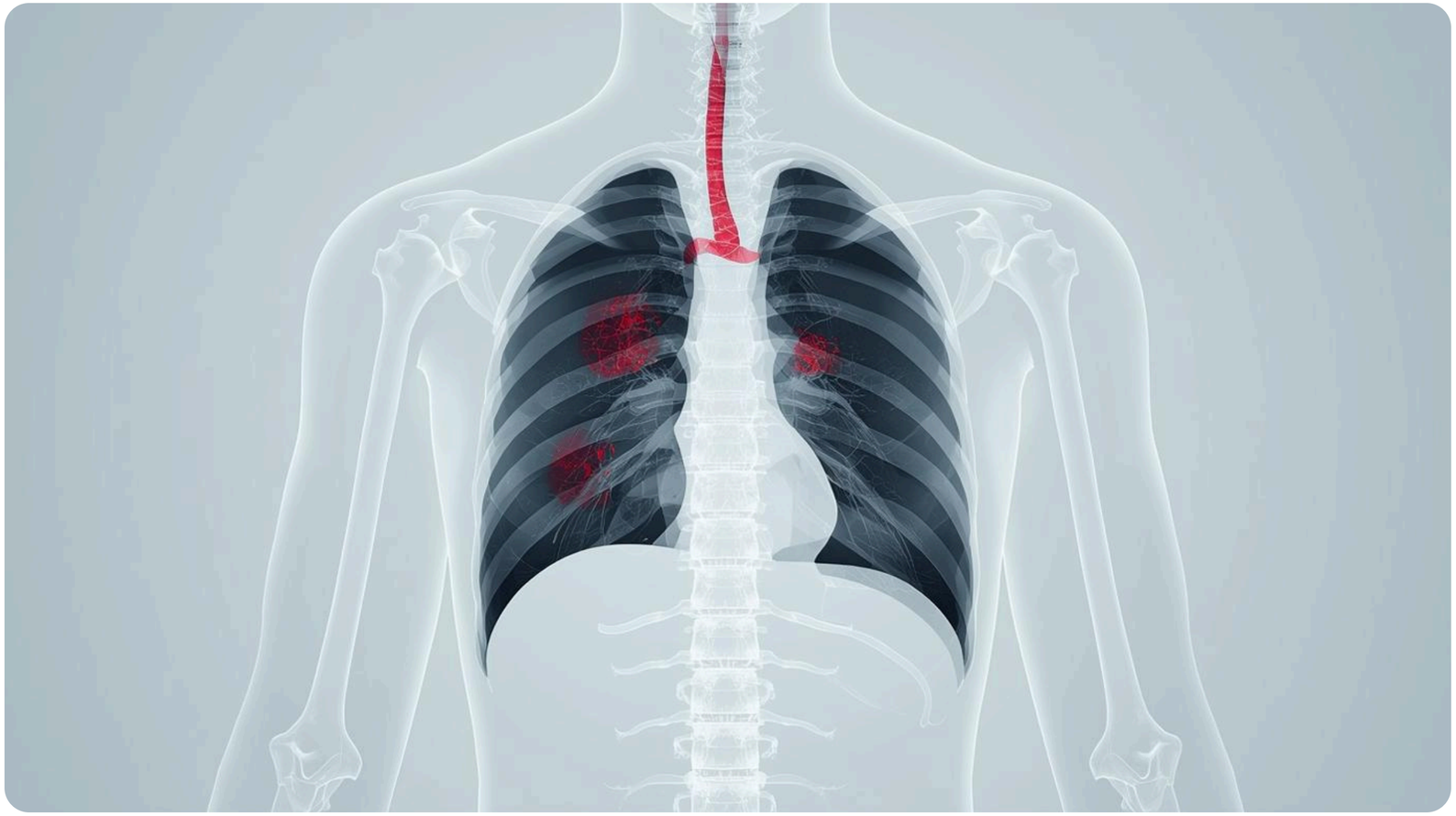
This section covers essential evaluation metrics including **Hamming Loss**, **F1-Micro**, and **F1-Macro**, crucial for assessing the performance of multi-label classification models in medical diagnosis contexts.

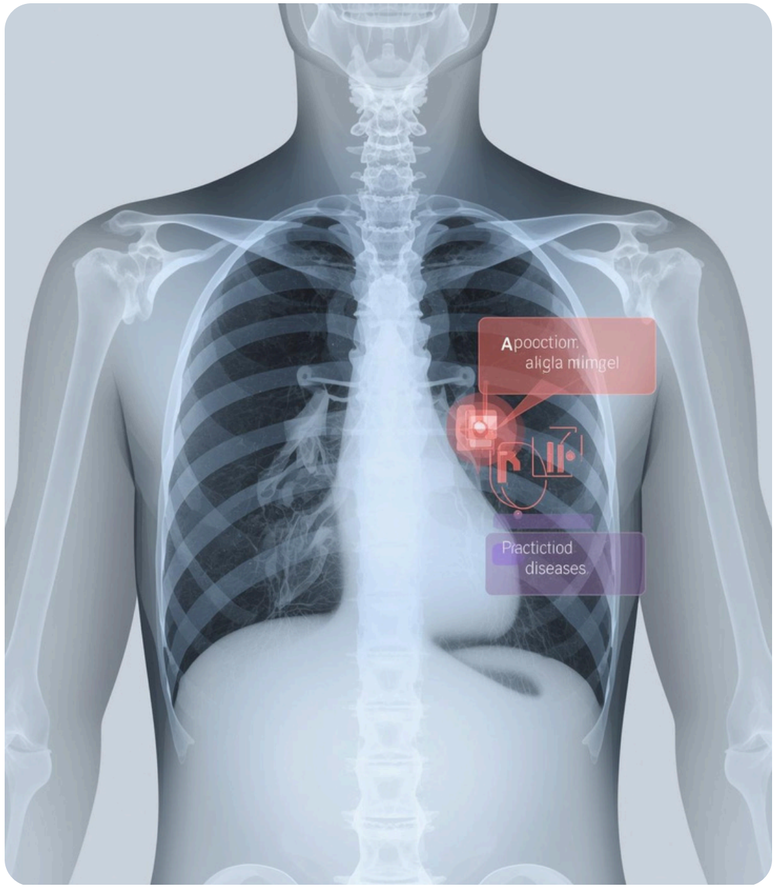
Best-Performing Model: Binary Relevance

The Binary Relevance method showed the **highest accuracy** in diagnosing multiple thoracic diseases simultaneously, effectively managing the challenges of class imbalance and enhancing diagnostic efficiency in clinical settings.



Sample Predictions on Chest X-Rays





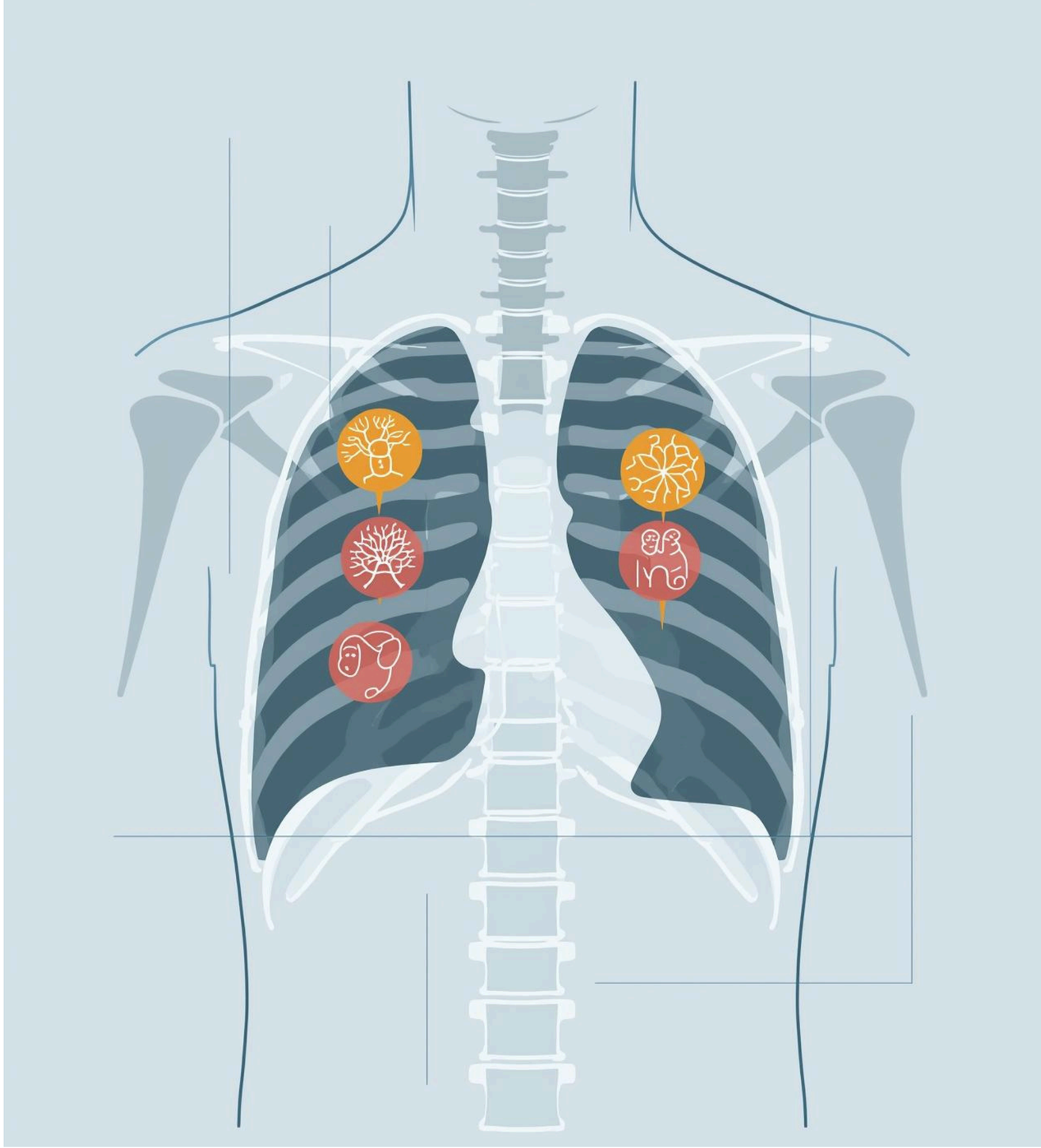
Challenges Faced

Addressing Limitations in Diagnosis Accuracy

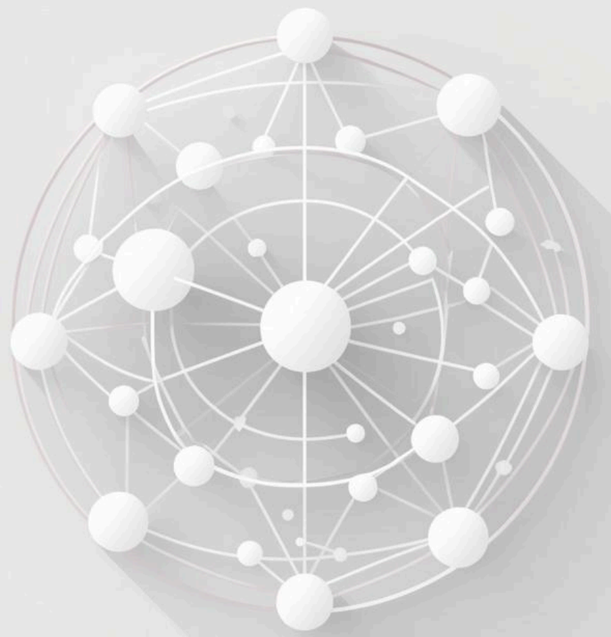
The analysis revealed significant **challenges** including low image resolution, class imbalance among diseases, and the inherent limitations of logistic regression in capturing complex relationships.

Conclusion: Insights on Multi-Label Classification

In conclusion, the **multi-label logistic regression** model effectively identifies thoracic diseases. Emphasizing the need for advanced techniques like CNNs and data augmentation can enhance diagnostic accuracy and clinical applications.



Future Directions in Medical Diagnosis



Future work should focus on **implementing CNNs**, enhancing data through augmentation, and exploring ensemble methods to improve classification accuracy and address existing challenges in multi-label medical diagnosis.

Thank You

Questions and Discussion Welcome