Enhancing Chest Xray Analysis: A Comparative Study of Deep Learning Models

Hari Priya Kandasamy

Introduction

Medical image analysis plays a crucial role in today's healthcare landscape. While traditional deep learning models have made significant strides in this field, understanding their inner workings, especially with complex images like chest X-rays, can be challenging. In our project, we compare the performance of cutting-edge deep learning models, such as Inceptionv4 and Vision Transformers, to more traditional models like ResNet50 to determine if they offer improved classification of chest X-ray images. We utilize two extensive datasets, NIH-CXR-LT and MIMIC-CXR-LT, to evaluate the models' performance. Furthermore, we aim to make these models more understandable and accessible by employing explainable AI techniques for visualizations. Our research seeks to contribute to the development of user-friendly and interpretable AI tools for medical imaging.

Data and Methodology

Datasets and Preprocessing: NIH-CXR-LT (100,000 images) and MIMIC-CXR-LT (380,000 images), with chest X-ray conditions including pneumothorax, edema, and fibrosis. Resized images to 224x224 pixels, normalized pixel values, and applied data augmentation techniques.

Deep learning models: Inceptionv4, Vision Transformers, and ResNet50, commonly used in medical imaging research.

Training and Validation: Cross-entropy loss, batch size of 256, Adam optimizer with a learning rate of 1e-4, trained for 50 epochs with early stopping after 15 epochs, evaluated using AUC Macro, accuracy disease-wise scores, and F1.

Explainable AI: Used GradCAM and SHAP techniques to identify biases or errors and improve interpretability. GradCAM generates heatmaps of prediction regions, while SHAP computes feature importance for each pixel in the image.

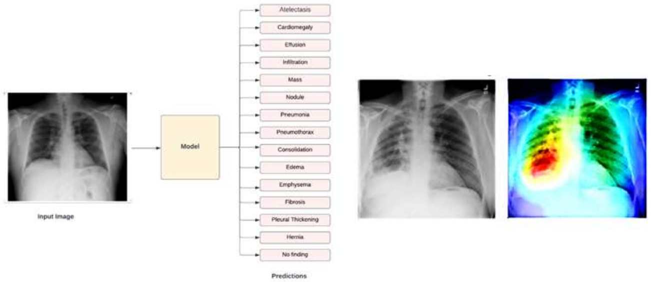

Fig 1. Model Overview

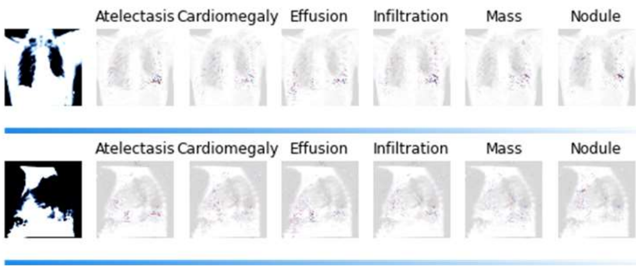

Fig 2. SHAP visualization

Table 1. NIH Chest X-rays results

Model	Train Loss	Val Loss	AUC Macro	F1 Macro
Resnet50	0.152962	0.238783	0.9357	0.6114
Inceptionv4	0.291863	0.2890	0.6460	0.0254
ViT	0.057522	0.369523	0.9932	0.8911

Table 2. MIMIC results

Model	Train Loss	Val Loss	AUC Macro	F1 Macro
Resnet50	0.232	0.458	0.888	0.484
Inceptionv4	0.3588	0.4787	0.6460	0.082
ViT	0.0882	0.8315	0.9886	0.8740

The University of Texas at Austin
School of Information
INF 388L: 2023 Spring Capstone

Results and Findings

- Inceptionv4 performed well on the NIH-CXR-LT dataset.
- ResNet50 had superior results on the MIMIC-CXR-LT dataset.
- Vision Transformer (ViT_base_patch16_224k) had competitive accuracy and the highest AUC values, making it the best overall performer.
- Grad-CAM and SHAP visualizations provided insights into the model decision-making process, supporting clinical adoption.

Overall, our study found that the Vision Transformer (ViT) had the best performance, with competitive accuracy and superior AUC values. ResNet50 outperformed Inceptionv4, particularly on the MIMIC-CXR-LT dataset. While Inceptionv4 showed strong performance on the NIH-CXR-LT dataset, it lagged behind ViT and ResNet50 on MIMIC-CXR-LT. The Grad-CAM and SHAP visualizations provided valuable insights into the model decision-making process, which will support the clinical adoption of these models in the future.

Conclusion and Future Work

Our study demonstrates the potential of Vision Transformer (ViT) for chest X-ray image analysis, outperforming traditional CNN models. The use of explainable AI techniques provided valuable insights into the model decision-making process which would increase the confidence of healthcare providers in using these models for clinical practice.

In future work, we aim to explore ensemble techniques, and integrate additional data sources to investigate the application of these models in other medical imaging modalities and clinical scenarios. Overall, we believe that the integration of AI models with explainable AI techniques holds great potential for improving healthcare outcomes and providing better patient care.