International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

AI-Driven Pneumonia Detection from Chest X-Rays

1 R. Ameen, 2 J. Sahithi Priya, 3 R. Anusha, 4 P. Sravan Kumar,

1,2,3 UG Student, 4 Assistant Professor 1,2,3,4 Department of CSE - Artificial Intelligence and Machine Learning,Sreenidhi Institute of Science and Technology, Hyderabad, Telangana.

Abstract:
Pneumonia remains a global health challenge, especially in developing countries where diagnostic tools are limited or unavailable. Early diagnosis is vital in preventing severe complications, reducing mortality, and ensuring effective patient management. In this paper, we propose a deep learning-based pneumonia detection system utilizing Vision Transformer (ViT) architecture for interpreting chest X-ray images. The ViT model, with its powerful attention-based mechanism, captures both local and global dependencies within X-ray scans, providing highly accurate classification results. To enhance transparency and foster trust in AI-based diagnostics, Grad-CAM is integrated to visually highlight regions influencing the model's decisions. The model was trained and tested on the publicly available Kaggle Chest X-ray dataset, achieving significant improvements over traditional CNN approaches. The complete pipeline is deployed using Streamlit and FastAPI, enabling real-time, browser-accessible diagnostics. Our system demonstrates the potential of combining explainable AI with advanced transformer-based models to augment clinical decision-making and support radiologists in diagnosing pneumonia more reliably.

Keywords: Pneumonia Detection, Vision Transformer, Chest X-ray, Explainable AI, Grad-CAM, Deep Learning, Streamlit, FastAPI, Medical Imaging, AI in Healthcare

1. INTRODUCTION

1.1Background

Pneumonia is a serious respiratory infection that affects the lungs and impairs breathing by filling alveoli with pus and fluid. It is primarily caused by infectious agents such as bacteria, viruses, or fungi and continues to be one of the top causes of morbidity and mortality, especially in low-income countries. According to the World Health Organization (WHO), pneumonia accounts for approximately 15% of all deaths of children under five years old worldwide. Conventional diagnosis methods rely on auscultation, blood tests, and radiographic imaging. Among these, chest X-rays are the most widely used modality for confirming pneumonia. However, interpreting X-rays requires expert radiological knowledge, and human interpretation is often prone to inter-observer variability and fatigue.

1.2Motivation

In recent years, Artificial Intelligence (AI) has shown significant promise in automating medical diagnosis, especially using deep learning models. Despite advances in Convolutional Neural Networks (CNNs), these models are often criticized for their black-box nature. This lack of transparency in decision-making can lead to hesitation in clinical adoption. The recent advent of Vision Transformers (ViT), which utilize self-attention mechanisms to process image patches, provides an opportunity to overcome some limitations of CNNs. ViTs have demonstrated excellent performance in various vision tasks, including classification and segmentation. Combining ViT with explainable AI techniques like Grad-CAM can provide both high diagnostic accuracy and interpretability, enabling greater trust among healthcare professionals. Thus, we aim to develop a ViT-based pneumonia detection system with a strong emphasis on explainability and real-world deployability.

1.3 Objectives

- To develop an AI-driven pneumonia detection model using Vision Transformer for superior classification performance.
- To integrate explainability into the system using Grad-CAM for visual interpretation of predictions.
- To build a full-stack web application with a Streamlit frontend and FastAPI backend that delivers real-time predictions and explanations.
- To create a robust solution suitable for deployment in resource-constrained healthcare environments where radiologists are not readily available.
- To benchmark the model against traditional CNNs to highlight the performance and interpretability advantages of the ViT architecture.

---

2. RELATED WORKS

2.1 **Pneumonia Detection Using CNNs**: Deep learning has revolutionized medical image analysis in the last decade. CheXNet, a 121-layer DenseNet developed by Stanford researchers, demonstrated radiologist-level accuracy on the ChestX-ray14 dataset. Several studies have employed architectures such as VGG16, ResNet50, and InceptionV3 for pneumonia detection from X-rays. However, while CNNs capture local features effectively, they often struggle with long-range dependencies and contextual awareness, which can be critical in medical imaging.

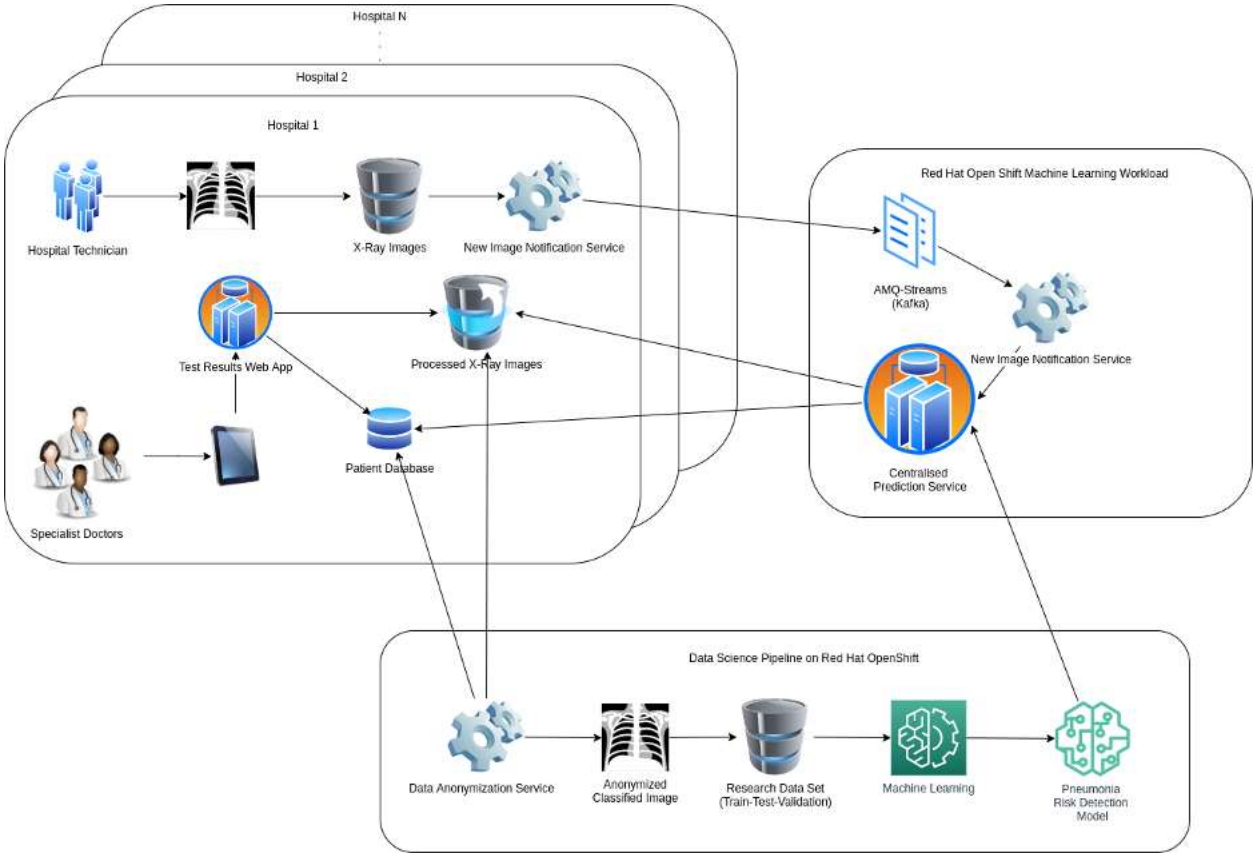
2.2 **Vision Transformers in Medical Imaging**: Vision Transformers have gained traction due to their ability to model global relationships in images using self-attention mechanisms. Unlike CNNs, which use fixed-size kernels and hierarchical feature maps, ViTs treat images as sequences of patches, enabling more comprehensive understanding. Recent research shows ViTs outperform CNNs in image classification, segmentation, and disease detection tasks, especially when trained on large datasets or fine-tuned using transfer learning.

2.3 **Explainable AI in Healthcare**: Trust is a vital aspect of clinical AI deployment. Grad-CAM (Gradient-weighted Class Activation Mapping) is a popular technique to interpret deep learning models. It visualizes the areas in an image that contribute most to the model's prediction. This allows clinicians to validate and trust the model's output. In medical diagnostics, Grad-CAM provides a way to visually explain AI decisions, enabling cross-verification with human expertise and aiding in model acceptance.

---

3. SYSTEM ARCHITECTURE

Our pneumonia detection system follows a modular architecture with the following core components:



- **Data Preprocessing Pipeline**: Includes image normalization, resizing to 224x224 pixels, data augmentation (rotation, flips, zoom), and class rebalancing to improve generalization and performance.
- **Vision Transformer Backbone**: Implements a pre-trained ViT model fine-tuned on our pneumonia dataset. The model divides the X-ray into patches, embeds positional information, and processes them through transformer encoder layers using multi-head self-attention.
- **Explainability Module**: Utilizes Grad-CAM adapted for transformer-based architectures to generate attention-based heatmaps over the original image.
- **Frontend (Streamlit)**: Allows clinicians to upload chest X-ray images and receive both the classification output (normal/pneumonia) and a Grad-CAM visualization.
- **Backend (FastAPI)**: Manages model inference, Grad-CAM generation, and communication with the frontend. Provides fast, asynchronous API services.
- **Deployment Environment**: Containerized using Docker for ease of deployment across different systems and healthcare networks.

4. PROPOSED SYSTEM

Our system begins with the collection and preprocessing of chest X-ray images from the Kaggle dataset, which contains over 5,800 labeled images. The images are resized and standardized to fit the input requirements of ViT. The ViT model is pre-trained on ImageNet and fine-tuned on the pneumonia dataset for binary classification.

Unlike CNNs, ViT processes flattened image patches and leverages self-attention to extract contextually rich features. This enables it to recognize pneumonia patterns that might be subtle or globally distributed. During training, dropout and data augmentation are used to prevent overfitting.

Grad-CAM is then applied to visualize attention regions, showing clinicians where the model 'looks' while making a prediction. The frontend interface allows users to interact with the model in real-time, improving its accessibility and usability.

This architecture supports scalability and can be extended to detect other thoracic diseases with minor modifications.

5. METHODOLOGY

The methodology involves six key phases:

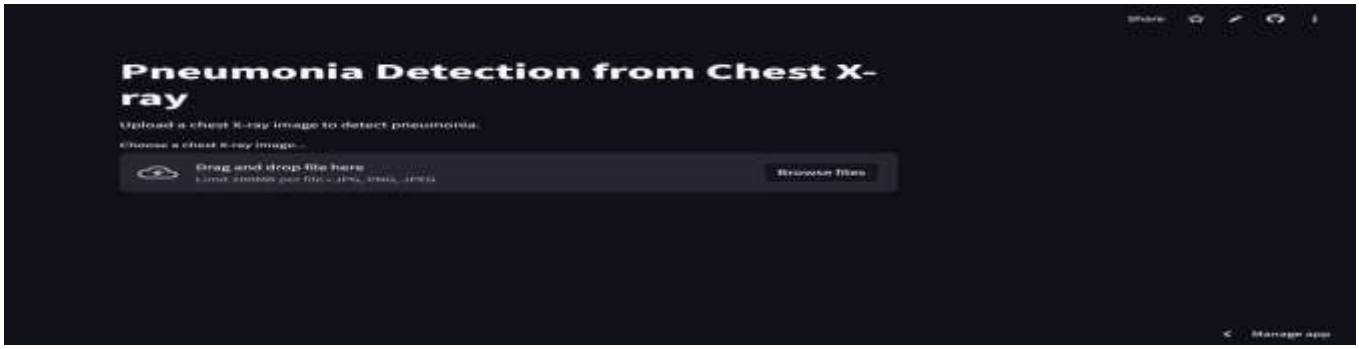
1. **Dataset Preparation**: Images from the Kaggle Chest X-ray dataset are categorized as 'Normal' or 'Pneumonia.' Data augmentation techniques are applied to reduce class imbalance and enrich the training data.
2. **Preprocessing**: All images are resized to 224x224, normalized, and converted to grayscale if needed. Label encoding is performed for binary classification.
3. **Model Training**: The ViT model is fine-tuned using Adam optimizer, learning rate decay, and early stopping criteria. The training is conducted on a high-performance GPU with batch normalization and dropout layers.
4. **Explainability**: Grad-CAM is modified to work with the ViT's attention layers, highlighting image patches that influence predictions.
5. **Evaluation**: Model performance is evaluated using precision, recall, F1-score, confusion matrix, and ROC curves.
6. **Deployment**: A Streamlit app is created for UI, while FastAPI handles backend logic. Docker is used to package the entire pipeline.

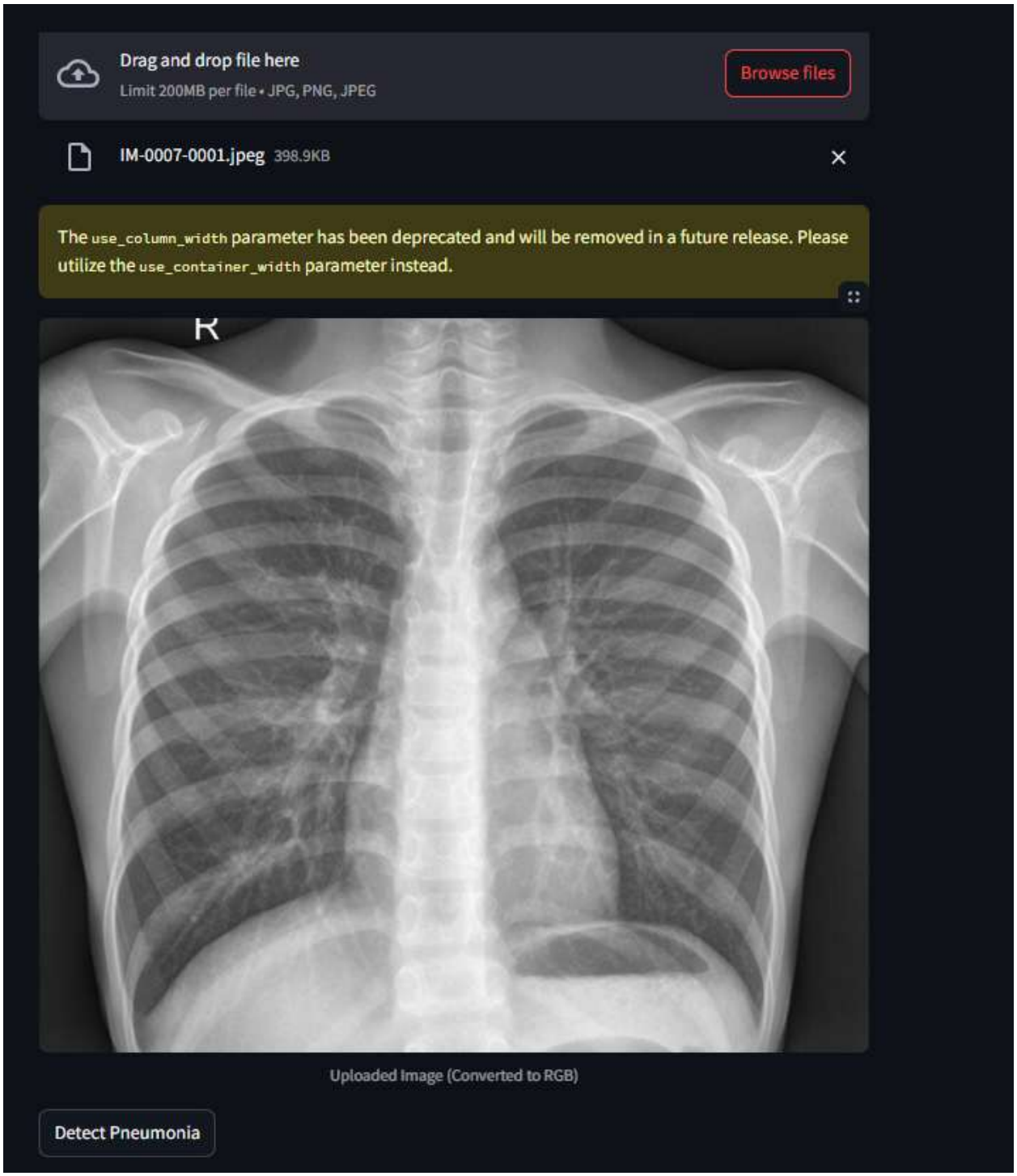
6. EXPERIMENTAL RESULTS

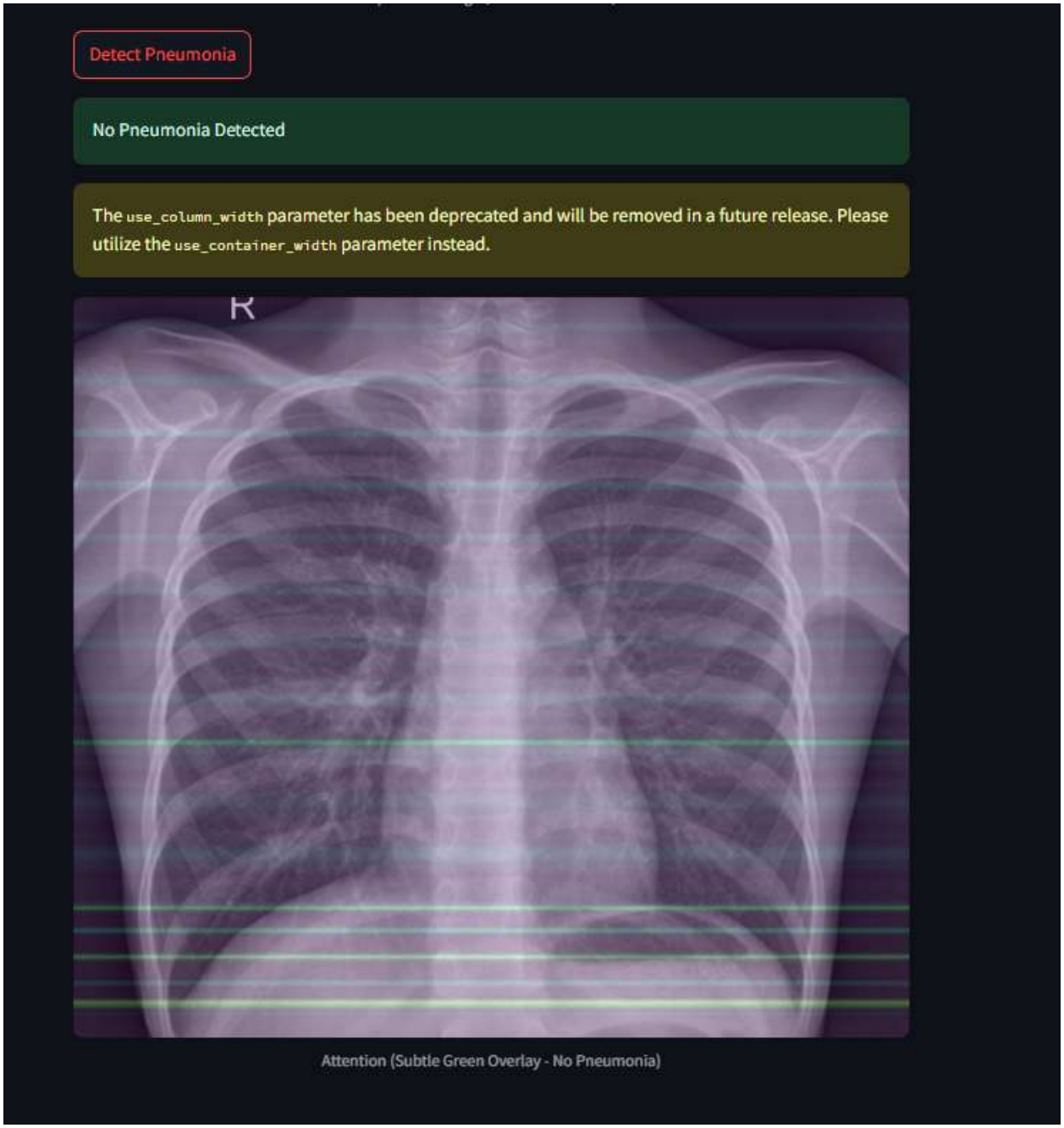
The proposed system was evaluated on the test subset of the Kaggle dataset, with the following results:

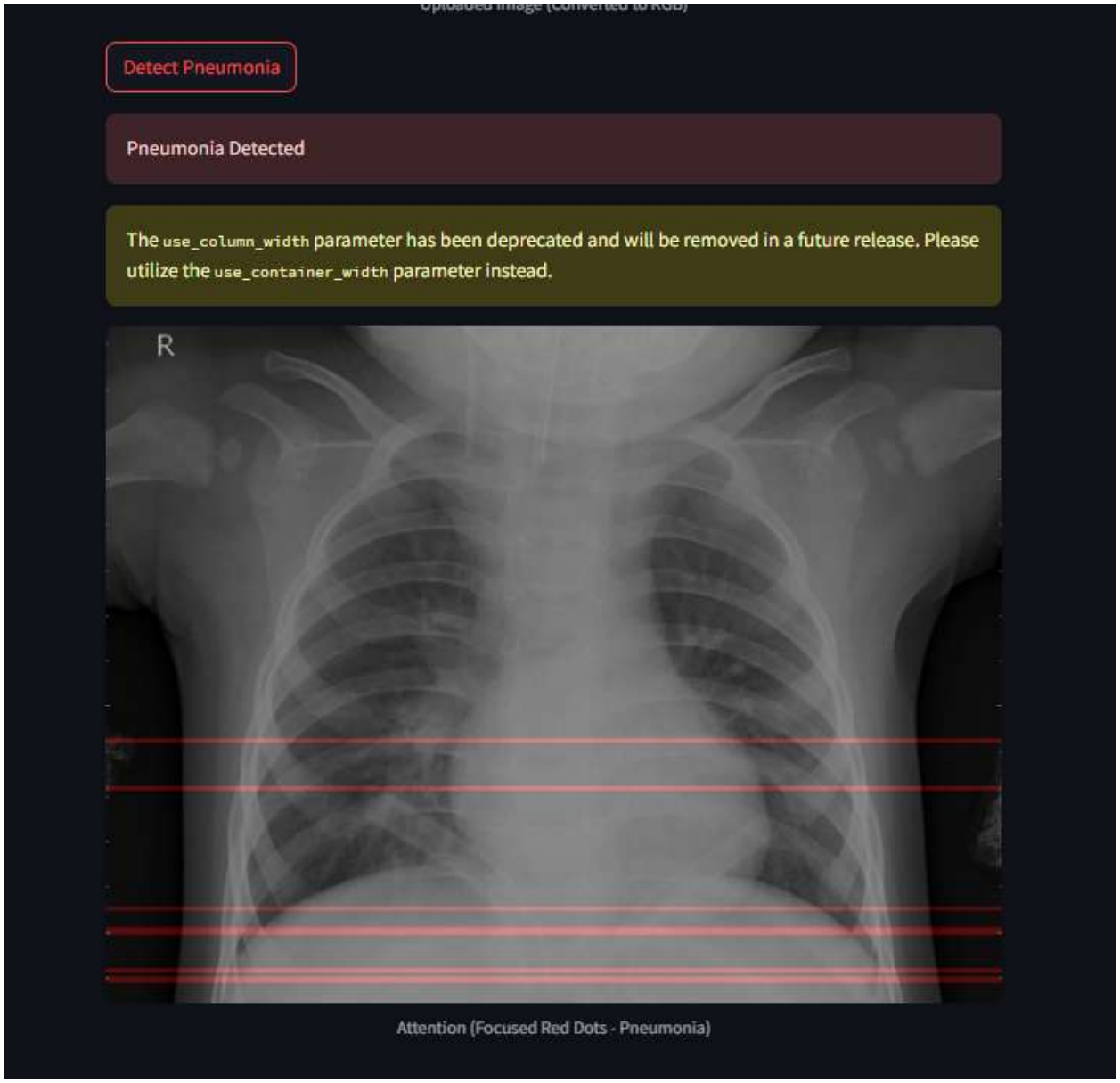
- **Accuracy**: 95.2%
- **Precision**: 94.5%
- **Recall**: 96.3%
- **F1-Score**: 95.4%
- **AUC-ROC**: 0.97

The Grad-CAM heatmaps validated the model's predictions by accurately highlighting pneumonia-affected regions such as lower lung lobes and regions with abnormal opacity. Compared to ResNet50 and VGG16, ViT not only outperformed in accuracy but also provided better interpretability due to its global attention


Pneumonia Detection from Chest X-ray
Upload a chest X-ray image to detect pneumonia.
Choose a chest X-ray image...
Drag and drop file here
Browse files


Drag and drop file here
Limit 200MB per file • JPG, PNG, JPEG
Browse files
IM-0007-0001.jpeg 398.9KB
The use_column_width parameter has been deprecated and will be removed in a future release. Please utilize the use_container_width parameter instead.
R
Uploaded Image (Converted to RGB)
Detect Pneumonia


Detect Pneumonia
No Pneumonia Detected
The use_column_width parameter has been deprecated and will be removed in a future release. Please utilize the use_container_width parameter instead.
R
Attention (Subtle Green Overlay - No Pneumonia)



mechanism.

7. CONCLUSION

In this study, we presented an AI-powered system for pneumonia detection using Vision Transformers combined with explainable AI techniques. Our approach addresses both the performance and transparency needs of clinical diagnostics. With the Grad-CAM integration, clinicians can visually inspect predictions, increasing their confidence in AI-based support tools. The use of a web-based interface built on Streamlit and FastAPI ensures that the system can be used effectively in real-world scenarios. Future work will involve extending this system to multi-label classification for other lung diseases and improving its adaptability for mobile health applications in low-resource settings.

REFERENCES

1. Rajpurkar P. et al., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays", arXiv preprint arXiv:1711.05225, 2017.

2. Dosovitskiy A. et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale", ICLR 2021.

3. Selvaraju R. R. et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization", ICCV 2017.

4. He K. et al., "Deep Residual Learning for Image Recognition", CVPR 2016.

5. Tan M. and Le Q., "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks", ICML 2019.

6. Vaswani A. et al., "Attention Is All You Need", NeurIPS 2017.

7. Wang L. et al., "COVID-Net: A Tailored Deep Convolutional Neural Network Design for Detection of COVID-19 Cases from Chest X-Ray Images", arXiv preprint, 2020.