LSTM-Based Multi-Head Attention Lightweight CNN Models for Pneumothorax Classification using Chest X-ray Images

Oxford Brookes University in collaboration with Chengdu University of Technology

Supervised by Dr Happy Nkanta Monday

Author: Danyue Xu

Abstract

In this study, a lightweight CNN-LSTM hybrid network is proposed for pneumothorax X-ray detection in medical Settings in poor areas. Combining EfficientNetB2 feature extraction, LSTM spatial sequence modeling and multi-head attention mechanism, it achieves 86% accuracy, 93% recall and 0.91 AUC value through data augmentation on only 2,027 SIIM-ACR datasets, which outperforms baseline models such as ResNet-50. The ability of lesion localization was verified by Grad-CAM visualization, and a GUI interface adapted to low-resource environments was developed. The number of model parameters is reduced by 40% while maintaining diagnostic accuracy. The GDPR-compliant energy-saving design particularly emphasizes ethical compliance for deployment in developing countries, providing a reproducible technical solution for AI healthcare landing in poor areas.

Model Architecture

This paper proposes a hybrid deep learning model architecture, which combines CNN, LSTM and Transformer similar attention mechanism.

- **CNN part:** EfficientNetB2 is used as the basic model for feature extraction to retain spatial information.
- **LSTM part:** The features extracted by CNN are transformed into state evolution in the temporal dimension to capture long-distance dependencies.
- **Attention mechanism part:** A multi-head attention mechanism is used to enable the model to focus on different parts of the input sequence at the same time and break the one-way time limit of LSTM.

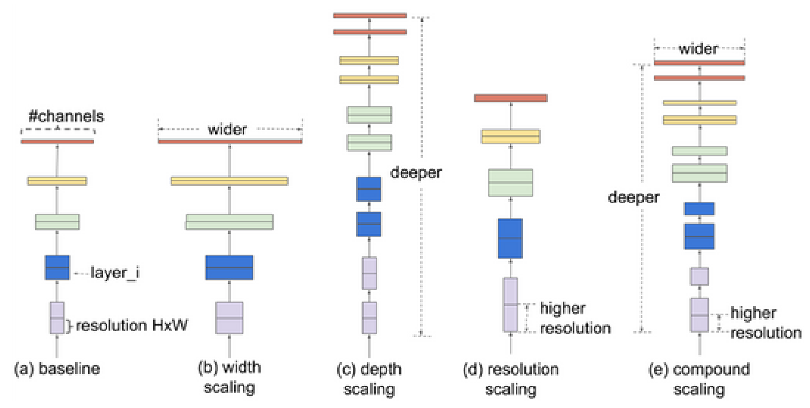

Figure 1 EfficientNetB2 model

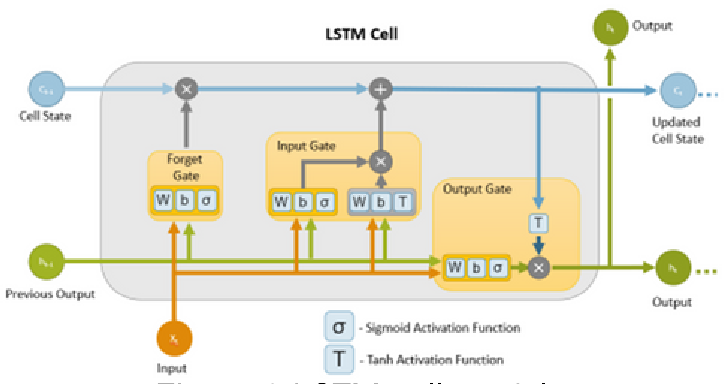

Figure 2 LSTM cell model

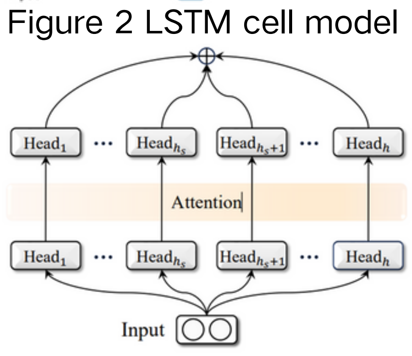

Figure 3 Multi-Head Mechanism model

Dataset Information

- **Dataset:** The SIIM-ACR Pneumothorax dataset containing 2027 chest X-ray images was used to perform a binary classification task (with/without pneumothorax).

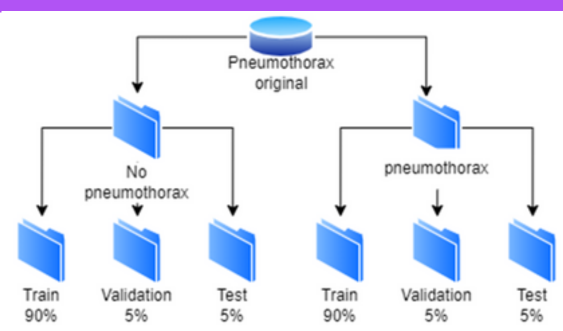

Figure 4 Segmentation of Dataset

- **Data Processing:** The dataset is fully preprocessed, including data augmentation, balancing, resizing, modifying color channels, and other steps to improve the generalization ability and performance of the model.

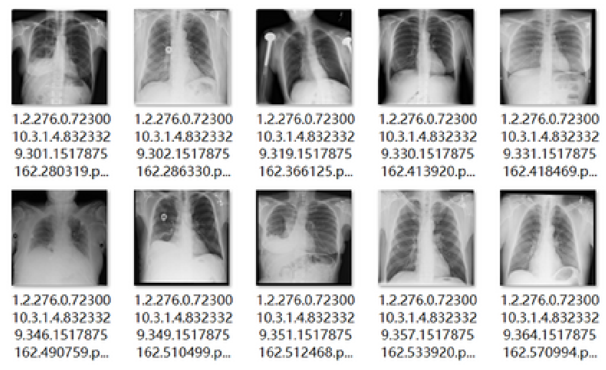

Figure 5 Original images

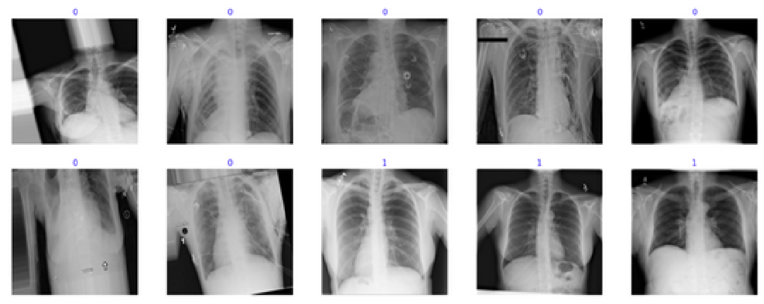

Figure 6 preprocessed images

Model Performance Analysis

1. **Classification performance index**

- The model has a 93% recall rate for pneumothorax cases, indicating a low risk of missed diagnoses, which is crucial for clinical settings.

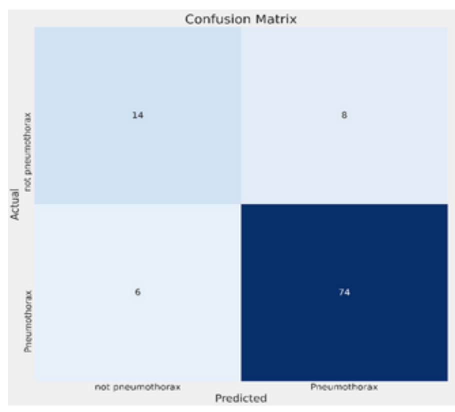

Figure 7 Confusion Matrix

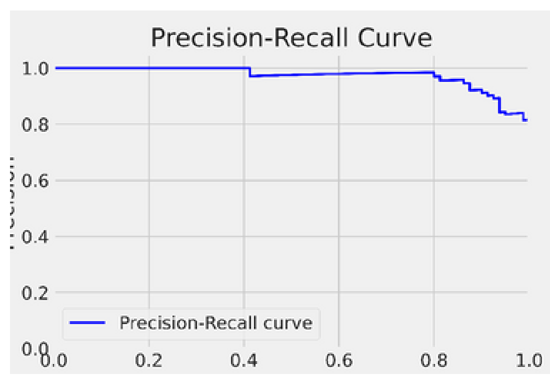

Figure 8 Precision-Recall curve

2. **Dynamic changes during training**

- During training, validation loss increased after the 11th round, and validation accuracy plateaued after the 3rd round. This shows the model reached performance saturation. Using an early stopping strategy can terminate redundant training rounds, improving training efficiency without sacrificing model performance.

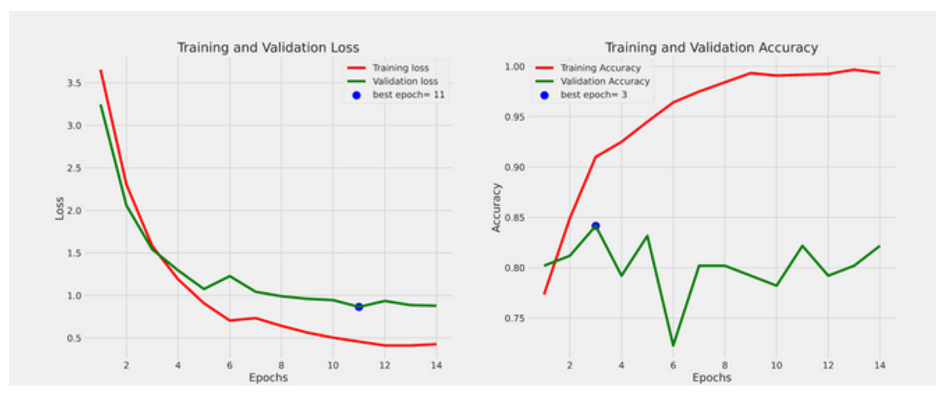

Figure 9 Training accuracy and loss curves

3. **Model Visualization:**

- Grad-CAM highlighted key regions the model focused on, boosting clinical interpretability.
- When predicting "negative", activation concentrated in the intact pleural line area, aligning with pneumothorax-negative signs, thus reinforcing the model's judgment.

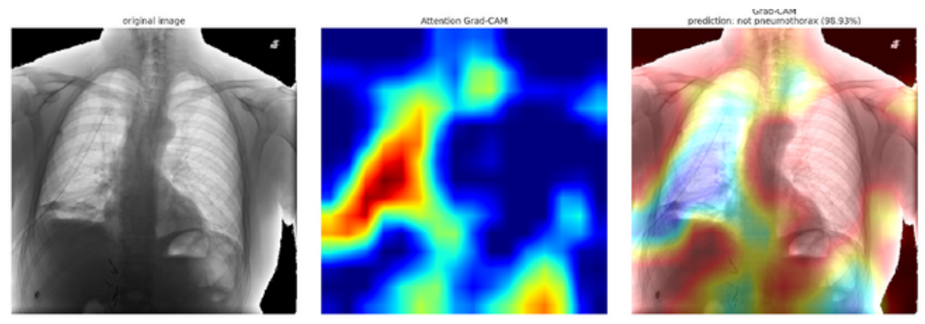

Figure 10 Example of Grad-CAM

Graphical User Interface

A GUI interface based on the Tkinter framework was developed to facilitate users to quickly analyze chest X-ray images and automate the process from image upload to report generation.

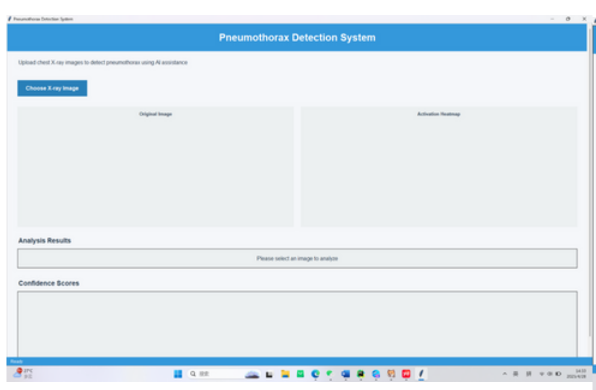

Figure 11 Homepage

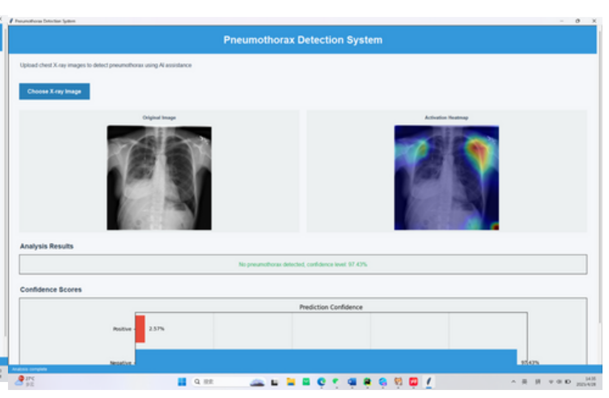

Figure 12 Detection page

Future Work

- **Conclusion:** This study successfully developed a lightweight CNN model that can efficiently and accurately detect pneumoperitoneum in resource-constrained environments, which has significant clinical application value.
- **Future work:** We plan to test the generalization ability of the model on more diverse datasets, further optimize the model architecture and training process to reduce the environmental footprint, and explore ways to incorporate patient clinical history or multi-view image information into the model.