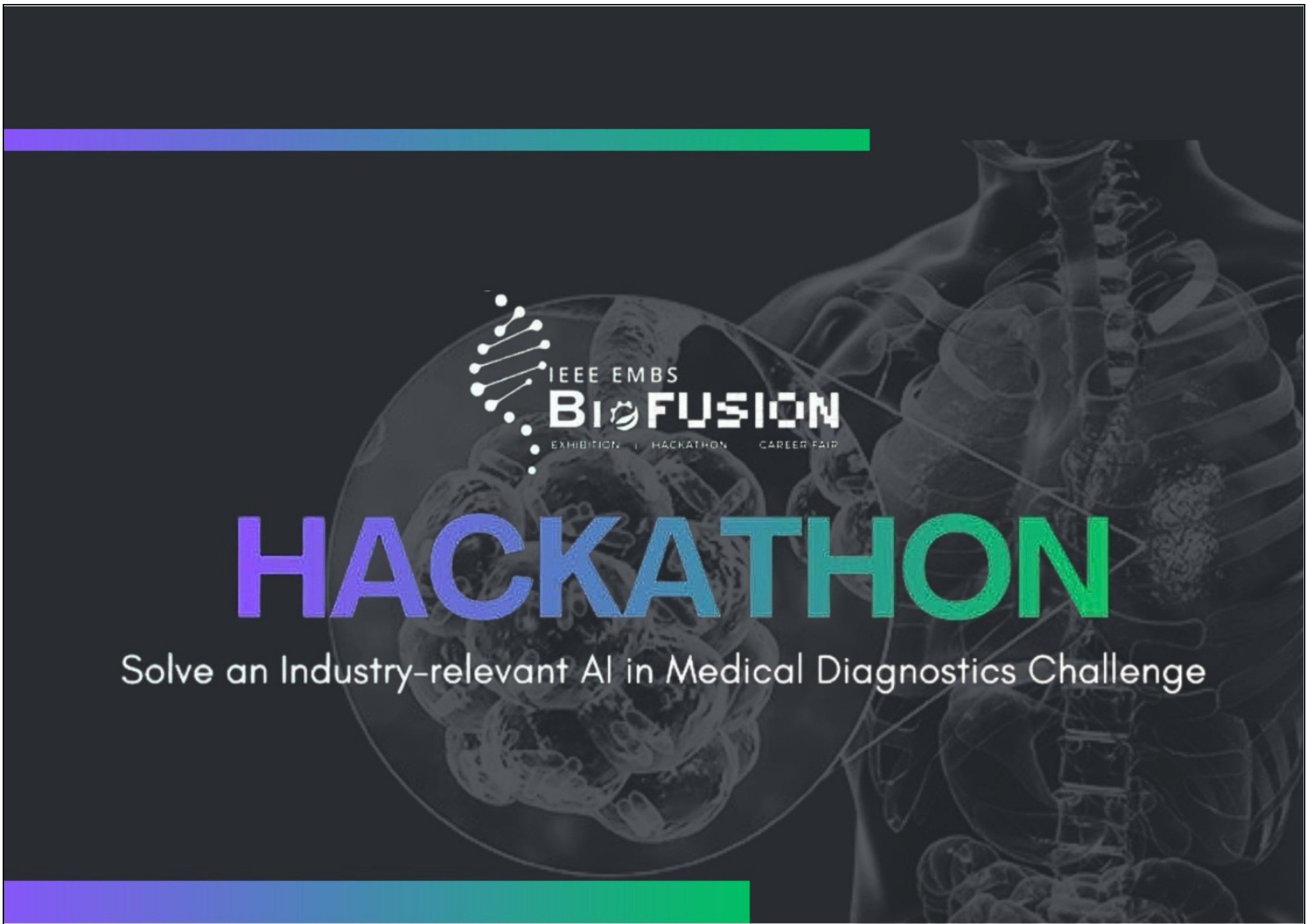


PNEUMONIA DETECTION FROM CHEST X-RAYS USING DEEP LEARNING

Team GMora

- Chanupa Hansaja
- Kusal Pabasara
- Pasindu Mihiranga
- Thanoj Buddhima
- Senuja Dilmith

UNIVERSITY OF MORATUWA

Automated Pneumonia Screening from Chest Radiographs

A ResNet-50 Approach

Abstract- This study presents an automated Pneumonia detection system using deep learning to address the accurate diagnosis of the disease as it remains a leading cause of mortality around the globe. Our team developed a decision support tool, trained on ImageNet. To enhance the clinical trust, we added Grad-CAM visualization that highlights the lung region and visual explanations ready for clinical deployment.

1. Literature Review – Deep learning is revolutionizing diagnosis of diseases. In "*Radiologist Level Pneumonia detection on Chest X-Rays with Deep Learning" by Rajpurkar et al,* the researchers developed a 121 DenseNet model (CheXnet) for Pneumonia detection using X-ray imagery with an F1 score of 0.435. They used transfer learning and implemented a weighted binary loss for imbalances in class. The limitation of the model is that it is not well designed for clinical adoption. In "*Hospital-scale Chest X-ray Database and Benchmarks by researchers" at National Library of Medicine, National Institutes of Health, Bethesda,* they introduced a large scale dataset with frontal view of X- ray imageries and evaluated multiple architecture such as VGG, ResNet. But multi label classification of their model reduces the accuracy of the model. Some researchers used transfer learning and achieved more than 90% accuracy on paediatric chest X-Ray detection (*Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. Kermany et al. (2018)*. The limitation of the model is that it didn't distinguish the type of Pneumonia. We identified several gaps in the solution. Clinical interpretability, sensitivity of the data, and resources constraints are the main gaps in this field. In order to bridge the research gap, we are introducing several solutions Mainly;

- Prioritization of Sensitivity to minimize false negatives
- Increase clinical adaptability using Grad CAM visualizations
- Architecture with reduced computational requirements and resource efficient
- Complete deployment ready to use pipeline
- Robust training with early stopping and data segmentation

2. Problem Identification

Affected Population – The most vulnerable groups are paediatric patients and elderly population. Healthcare workers face heavy workloads as well as limited access to radiologists in rural and under developed areas. Secondary stakeholders of this problem are Hospitals, Clinics and health care system.

Why is this problem important – According to *Google*, "*Pneumonia remains a major global health threat, causing millions of deaths annually, particularly among young children (under 5) and older adults (over 70), with recent estimates suggesting around 2.5 million deaths in 2023, making it the leading infectious killer of children*". So, this is a local as well as a global burden.

Specific unmet need in healthcare – There are several factors affecting the treatment and diagnosis of Pneumonia. The radiologist shortage across the globe, Diagnostics delays and variability in diagnosis and lack of early identification of the disease are major unmet needs in healthcare system.

Our solution addresses scalable deployment, constant and 24 hour diagnosis, decision support system and Automated pre screening.

3.Details about the dataset:

Name	Chest X-Ray Images (Pneumonia)
Source	Kaggle (Paul Mooney)
Total Images	5,863 chest X-rays (JPEG)
Original Publication	Kermany et al., 2018 - Mendeley Data
Institution	Guangzhou Women and Children's Medical Centre

Dataset Distribution:

Note; We created a proper validation split (15% of training data) due to the inadequate original validation set size.

	Normal	Pneumonia	Total
Train	1,341	3,875	5,216
Val	8	8	16
Test	234	390	624
Total	1,583	4,273	5,856

Why the selected dataset is appropriate – These are taken from a real clinical data and published in peer reviewed publication. Also, the dataset provides sufficient data for transfer learning and a widely used benchmark dataset.

4. Methodology

4.1 Data Processing pipeline

Image Preprocessing steps:

Stage	Operation	Description & Purpose
1	Resize	Purpose: ResNet50 requires a fixed input size. Variable raw X-rays are standardized to this square format.
2	RGB conversion	Process: Convert grayscale to RGB (3-channel). Purpose: Transfer learning models pre-trained on ImageNet expect 3 input channels.
3	Tensor conversion	Process: PyTorch tensor. Purpose: Converts image pixel data (0-255 integers) into floating-point tensors (0.0-1.0).
4	Normalization	Values: Mean [0.485, 0.456, 0.406], Std [0.229, 0.224, 0.225]. Purpose: Standardizes the pixel value distribution to match the ImageNet statistics used during the original pre-training.

5.Model Architecture

Resnet 50 architecture:

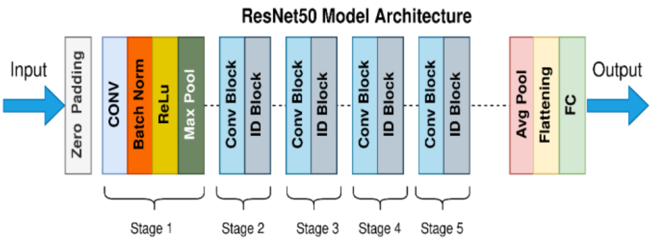


Figure 1:ResNet50 Model Architecture

Why ResNet50 - The ResNet50 prevents vanishing gradients in deep networks with sufficient 50 layer depth for feature extraction. It also consists with fewer parameters than VGG. Also has a proven medical imaging performance.

Training process – For training configuration we implemented 0.01 learning rate, 32 batch size,10 epochs early stopping parameter to 3 and seed to 42. In here we can use 64 batch size but it will consume more computational power and also, we used 10 epochs to successfully train the model. Increase in the epoch created more recall and overfit so, we decide to use epoch as 10. Given the significant class imbalance in the training set (Normal: 1341 vs. Pneumonia-3875), would bias the model toward the majority class (Pneumonia). To solve this, we employed a weighted function. Result is that the model pays nearly double the attention to normal cases to prevent overfitting to Pneumonia.**Training Loop** – For each epoch, there are 3 phases. First the model set to train and update each weight and record training loss and accuracy. In second Phase which is validation phase, set the model to eval mode and record validation loss and accuracy. Next the model checks the early stopping criteria and increment patient counter or reset it. So, if patience counter is greater than or equal to patience number (given earlier) it will stop training and load the next epoch. Finally, it outputs the best model as checkpoint.

Validation Strategy - Original data set has inadequate Val set. So, we use 85:15 model for training dataset to train and validate. The validation metrices tracked loss and accuracy and then saved when Val loss improves.

6.Pretrained Model Usage and Adaptation

Why we choose a pretrained model – We choose this model as our data set training from the scratch is insufficient and time consuming process. These pretrained models accelerates convergence and reduces training time. As we are using sensitive medical dataset, we inclined to use pretrained model as its accuracy and sensitivity is higher to meet our goal and also the model is peer reviewed.

Modifications - For an original resNet50 model there are many ImageNet classes. But in our model, we implemented two classes as Normal/Pneumonia which leads to SoftMax value to 2. Only the classifier head was replaced in layer adaptation. Our output is [P(Normal), P(Pneumonia)], Predicted class = argmax(output). **Training Strategy** – We employed feature extraction to fine tune the model. This will help as we have limited data but it is robust and faster training method. It also helps in preventing catastrophic forgetting of useful features. To optimizes the learning rate, we use Adam optimizer.

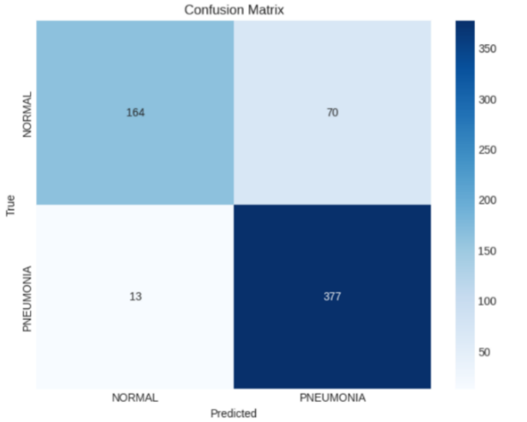
Risk and Bias discussion - Because the ResNet-50 backbone was pre-trained on ImageNet, a dataset of natural, RGB, object-centric images, which is very different from the grayscale, textural, and anatomical nature of chest radiographs and the use of Transfer Learning creates a fundamental domain mismatch. The model's deployment scope is strictly limited due to inherent dataset biases, which go beyond technical limitations. The model displays age and demographic bias because the training data comes from a single Guangzhou paediatric (ages 1–5), which makes it unsuitable for adult populations without outside validation. Additionally, the high prevalence of pneumonia in the training set (74%) introduced a bias, which we reduced using a weighted function to prevent the model from over-predicting the positive class.

7.Results

Test Set Performance Summary and Classification Report:

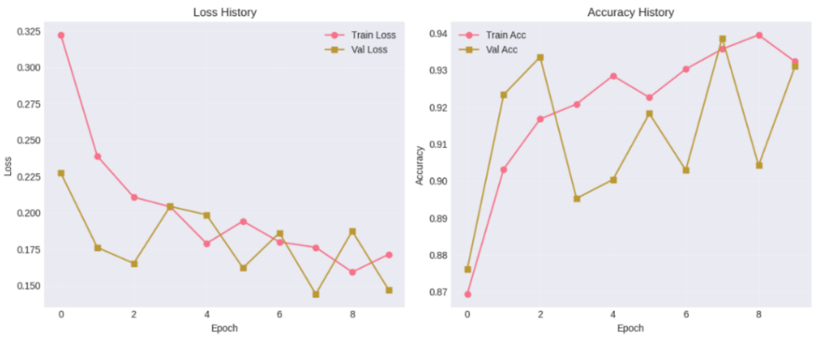
```
TEST SET METRICS
Accuracy:  86.70%
Precision: 0.8434
Recall:    0.9667
F1-Score:  0.9008
AUC-ROC:   0.9428
              precision    recall  f1-score   support

      NORMAL       0.93      0.70      0.80       234
   PNEUMONIA       0.84      0.97      0.90       390

    accuracy                           0.87       624
   macro avg       0.88      0.83      0.85       624
weighted avg       0.87      0.87      0.86       624
```

Confusion Matrix:

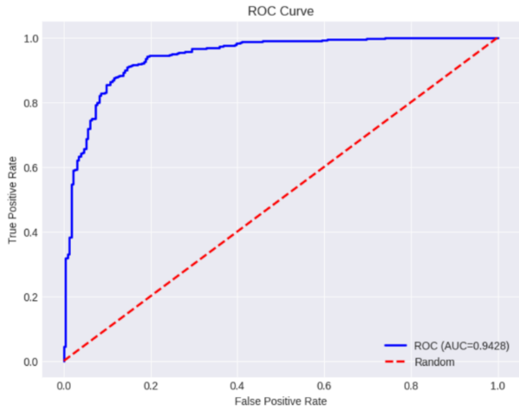


Negative cases: 164, False positives: 70, False negatives: 13, Positive cases: 377

Training History:



ROC Curve Performance:



Grad-CAM Attention Maps – These visualizations confirm the model focuses on lung fields and opacity regions. This technique produces a localization map highlighting the important regions in the image.

Error Analysis – There were several misclassifications in the analysis. There were false positives but fewer false negatives resulting additional reviews. These errors may include borderline cases with unclear classification and subtle positioning variations. We are recommending a threshold optimization for these errors will help to reduce the error.

Model Limitations – We identified several model limitations and they are as below:

Limitation	Description	Impact
Binary classification only	Cannot distinguish between types of Pneumonia.	Limits the specificity of treatment guidance.
Paediatric data	Trained on patients ages 1–5.	Performance on adult chest X-rays is likely unreliable.
Single institution	All data originates from a single hospital facility.	Model may not generalize well.
Image quality dependency	Performance is sensitive to noises.	Requires quality control.
No severity grading	Only detects the presence or absence of disease.	Cannot assess the progression of Pneumonia
Bias	Deliberately trades specificity for higher sensitivity.	Results in a higher false positive rate.

8.Real-World Application

Clinical Decision Support System - We are proposing a Clinical decision support system which input the patient's data or X-ray image and then processed with ResNet50 model and deliver a clinical dashboard to the radiologist or the doctor, improving the decision making. AI acts as a "second opinion" providing Grad-CAM visualizations to explain its reasoning.

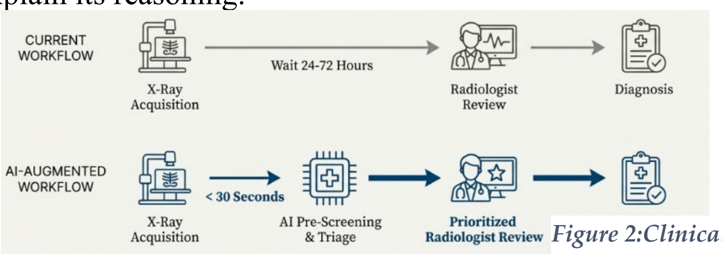


Figure 2:Clinical decision support system (Concept)

Target Audience & Value Proposition - We can integrate the healthcare workflow with adapting direct image retrieval and result storage, useful API for custom integrations and complete prediction history for institutions. For the deployment, we are proposing cloud based, scalable infrastructure for multi hospital network system. But we should address the data privacy and security issue. Apart from that we can implement this system to an edge device which is at the X-Ray machine. We can obtain instant result with no network dependency but hardware cost is a concern.

Risks & Limitations in Deployment - Because the ResNet50 backbone was pretrained on ImageNet, a dataset of natural, RGB images, textural and anatomical nature of chest radiographs, the use of Transfer Learning creates a fundamental domain mismatch. The model's deployment scope is strictly limited due to inherent dataset biases, which go beyond technical limitations. The model displays age and demographic bias because the training data comes from a single institute (ages 1–5), which makes it unsuitable for adult populations or diverse ethnic groups without outside validation.

Marketing & Impact Strategy - We identified several target groups to implement our solution. Primary target is resource limited health centres, rural hospitals and mobile health clinics. The secondary target group is high volume facilities such as urban hospital so they can reduce workload and faster response. The telemedicine providers are also a target group as this solution has remote diagnostic capabilities.

The implementation of an autonomous diagnostic system offers significant benefits. Some of them are; **higher sensitivity, reduced missed cases, faster diagnosis, decision support, radiologist productivity, 24/7 availability and resource optimization.**

9.Future Improvements

In order to turn this prototype into a Software as a Medical Device (SaMD) that is clinically feasible, we have established a roadmap that emphasizes data diversity, technical advancement, and regulatory compliance. Technically, the model will advance from ensemble techniques and instantaneous fine tuning to sophisticated multi modal, improving privacy and accuracy. In addition to these enhancements, we intend to add more photos to the dataset, focusing on a variety of demographics in order to reduce current biases.

Resources:

- ***Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning, Kermany, Daniel S. et al., Cell, Volume 172, Issue 5, 1122 - 1131.e9***
- ***Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning, https://doi.org/10.48550/arXiv.1711.05225***
- ***Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases, https://www.researchgate.net/publication/320068322_ChestX-ray14_Hospital-scale_Chest_X-ray_Database_and_Benchmarks_on_Weakly-Supervised_Classification_and_Localization_of_Common_Thorax_Disease***
- ***Dataset: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia/data***
- ***Jupyter Notebook: https://colab.research.google.com/drive/1G-9deBqRItV8JEFBGVJ4Wxym88MrvaqR?usp=sharing***