Journal of Advanced College of Engineering and Management

# Explainable Pneumonia Detection from Chest X-Ray Images Using a ResNet-50 Based Deep Learning Framework

[1]Sitam Gautam, [2]Anusha Chhetri Katuwal
*[1,2]Masters of Information Technology (professional), Deakin University, Melbourne, Australia*
*Email: [1]gautamsitam5@gmail.com, [2]anusha.katuwal.7@gmail.com*

*DOI: 10.3126/jacem.v12i01.93909*

***Abstract***

Pneumonia is a serious lung infection responsible for approximately 2 million deaths in children annually and 50,000 adult deaths each year, placing a significant burden on global healthcare. Early and accurate detection is critical, particularly in resource-limited settings. This paper proposes an automated pneumonia detection framework based on the ResNet-50 deep learning architecture applied to chest X-ray images. The model was trained on the publicly available Chest X-Ray Pneumonia dataset (Kaggle) using transfer learning, with 70% of data used for training and 30% for testing. To address model interpretability, a critical requirement for clinical adoption of Gradient-weighted Class Activation Mapping (Grad-CAM) was integrated to visually highlight lung regions influencing classification decisions. The proposed model achieved an overall accuracy of 81.9%, a recall (sensitivity) of 96.7%, a precision of 79.0%, a specificity of 57.3%, and an F1-score of 87.0% on the test set. The high recall is prioritized given the clinical importance of minimizing false negatives in pneumonia screening. Grad-CAM visualizations confirm that the model focuses on anatomically relevant lung regions. This system is intended as a decision-support tool to assist clinicians rather than replace them.

***Keywords**: Pneumonia Detection, Chest X-ray, Deep Learning, Convolutional Neural Network, ResNet-50, Grad-CAM*

## 1. INTRODUCTION

Pneumonia is a serious lung infection caused primarily by bacteria or viruses and remains one of the leading causes of mortality worldwide. It affects millions of people each year and it poses

a particularly high risk to young children, older adults, and individuals with weakened immune systems [1]. Despite advances in medical care, pneumonia continues to place a significant burden on healthcare systems, especially in low-resource settings where access to timely diagnosis and treatment may be limited.

Chest radiography is the most used imaging technique for diagnosing pneumonia. It plays a crucial role in clinical decision-making and epidemiological assessment. However, interpreting chest X-ray images can be challenging. The radiographic appearance of pneumonia is often subtle and may overlap with other pulmonary conditions, leading to variability in interpretation among radiologists. In addition, many regions face shortages of experienced radiology specialists, which can delay diagnosis and treatment.

Recent developments in artificial intelligence, particularly deep learning, have shown promising results in medical image analysis. Convolutional Neural Networks (CNNs) are capable of automatically learning discriminative features directly from image data and have demonstrated strong performance in chest X-ray classification tasks [2], [3]. These models offer the potential to assist clinicians by providing fast and consistent preliminary assessments.

In this study, a deep learning framework was proposed for automated pneumonia detection from chest X-ray images using a CNN-based architecture. To enhance interpretability and clinical relevance, Grad-CAM is integrated into the system to visualize important regions contributing to the model's predictions. The goal of this work is to develop a reliable and interpretable decision-support tool that can assist healthcare professionals in early pneumonia screening.

## 2. LITERATURE REVIEW

Prediction of pneumonia using Artificial intelligence is an advance approach in the field of medicine. In this approach of advance science, chest diseases can be shown in images in the form of cavitation, consolidations, infiltrates, blunted costophrenic angles, and small broadly distributed nodules. By analyzing the chest X-ray image, the radiologists can diagnose many conditions and diseases such as pleurisy, effusion, pneumonia, bronchitis, infiltration, nodule, atelectasis, pericarditis, cardiomegaly, pneumothorax, fractures, and many others. Classifying the chest X-ray abnormalities is considered as a tedious task for radiologists; hence, many algorithms were proposed by researchers to accurately perform this task [4]. Medical X-rays are images which are generally used to diagnose some sensitive human body parts such as bones, chest, teeth, skull, and so on. Medical experts have used this technique for several decades to explore and visualize fractures or abnormalities in body organs. X-rays are very effective diagnostic tools in revealing the pathological alterations, in addition to its non-invasive characteristics and economic considerations. Still there may be defects and error occurred while handling x rays from the naked eyes. In the advancement of technology, medical x-rays can play vital role when collaborating with machine learning algorithms [5].

Community-acquired pneumonia is a common cause of hospitalization, and pneumococcal vaccinations are recommended for high-risk individuals. Although risk factors for pneumonia have been identified, there are currently no pneumonia hospitalization prediction models based on the risk profiles of healthy subjects. This study aimed to develop a predictive model for pneumonia hospitalization in adults to accurately identify high-risk individuals to facilitate the efficient prevention of pneumonia. It conducted a retrospective database analysis using health checkup data and health insurance claims data for residents of Kyoto prefecture, Japan, between April 2010 and March 2015 choosing adults who had undergone health checkups in the first year of the study period and tracked pneumonia hospitalizations over the next 5years. Subjects were randomly divided into training and test sets [20]. The outcome measure was pneumonia hospitalization, and candidate predictors were obtained from the health checkup data. The prediction model was developed and internally validated using a LASSO logistic regression analysis [19].

For this purpose, traditional and deep learning-based networks are employed to classify most common thoracic diseases and to present comparative results. Back propagation neural network, and convolutional neural network are examined to classify 12 common diseases that may be found in the chest X-ray, that is, atelectasis, cardiomegaly, effusion, infiltration, mass, nodule, pneumonia, pneumothorax, consolidation, emphysema, and fibrosis. In this paper, we aim at training both traditional and deep network using the same chest X-ray dataset and evaluating their performances. The data used in the paper are obtained from the Kaggle, online community of data scientists and machine learning. The dataset contains 25,684 frontal-view X-ray images of 30,229 unique patients [7]. To overcome interpretability limitations, explainable AI techniques have gained increasing attention. Gradient-weighted Class Activation Mapping (Grad-CAM) is one of the most widely used methods for visual explanation of CNN predictions [12]. Grad-CAM generates heatmaps that highlight the image regions most influential in a model's classification decision. Recent studies have demonstrated that Grad-CAM can help verify whether deep learning models focus on anatomically relevant lung regions rather than background artifacts, thereby improving clinical confidence [16].

Building on these findings, the present study adopts a ResNet-50 based transfer learning framework integrated with Grad-CAM visualization. This combination aims to achieve a balance between reliable pneumonia detection performance and model interpretability, addressing both accuracy and transparency requirements essential for medical screening application.

## A. Research Gap

Despite significant advancements in deep learning for pneumonia detection, several research gaps remain. This study directly addresses the following:

a. Interpretability: Most powerful CNN models function as black boxes, and clinicians cannot easily trust their outputs [13], [15]. This study addresses this gap directly through Grad-CAM integration, providing visual heatmaps that confirm anatomically relevant focus regions.
b. Feature Localization: Few studies use interpretable AI to direct model attention toward lung regions rather than irrelevant background features [12], [16]. Grad-CAM in this work confirms that the ResNet-50 model appropriately focuses on lung tissue,

c. particularly in pneumonia cases where lower and central lung opacity patterns are highlighted.
d. Class Imbalance: Imbalanced data (more pneumonia than normal cases) can adversely affect normal class detection [14]. This study acknowledges the resulting lower specificity (57.3%) as a known limitation and prioritizes recall to minimize clinically dangerous false negatives. Full resolution of class imbalance remains a future direction.
e. Limited Dataset Diversity: Most studies use data from one or few sources, limiting generalization [17]. This study uses the publicly available Kaggle Chest X-Ray Pneumonia dataset (single source), and this limitation is explicitly acknowledged. Cross-institutional validation remains future work.
f. Clinical Readiness: While the proposed model achieves strong recall, its accuracy of 81.9% and specificity of 57.3% indicate it is currently suitable as a screening aid rather than a standalone diagnostic tool. Refinement towards clinical deployment requires further validation on diverse, multi-institutional datasets [17], [18].

## 3. METHODOLOGY

This section describes the methodological framework adopted for automated pneumonia detection from chest X-ray images. The methodology consists of data preprocessing, model training using transfer learning, performance evaluation, and explainability analysis using Grad-CAM.

### A. Dataset Description

The Chest X-Ray Images (Pneumonia) dataset, publicly available on Kaggle [2], was used in this study. The dataset was originally curated from the Guangzhou Women and Children's Medical Centre (GWCMC) and comprises a total of 5,863 chest X-ray images in JPEG format, partitioned into two classes:

**Normal:** 1,341 images chest X-rays with no evidence of infection

**Pneumonia:** 4,522 images chest X-rays showing clinical evidence of pneumonia, including both bacterial and viral subtypes

The class distribution is notably imbalanced, with a ratio of approximately 1:3.4 (Normal: Pneumonia). All images were obtained from pediatric patients aged 1–5 years and were graded and labelled by expert physicians before being cleared for training.

For this study, 70% of the combined dataset was allocated for model training and 30% for testing. The test split comprises 234 Normal and 390 Pneumonia images (624 total). No separate validation set was used during training; learning curves on training and test subsets served as the model monitoring strategy.

Image preprocessing included resizing all images to 224×224 pixels, normalizing pixel values to the [0, 1] range, and applying data augmentation techniques (random horizontal flipping, rotation, zoom) to the training set to increase variability and reduce overfitting.

### B. Experimental Environment and Tools

The experiments were implemented in Python 3.9. The deep learning framework used was TensorFlow 2.10 with Keras as the high-level API. Model training was performed on a Google Colab Pro environment. The ResNet-50 pre-trained weights were sourced from the ImageNet dataset via the tf.keras.applications.ResNet50 module. Grad-CAM visualizations were generated using the tf-keras-vis library. Image preprocessing and augmentation were performed using the TensorFlow ImageDataGenerator API. scikit-learn 1.1.0 was used for computing evaluation metrics including the confusion matrix.

### C. System Workflow

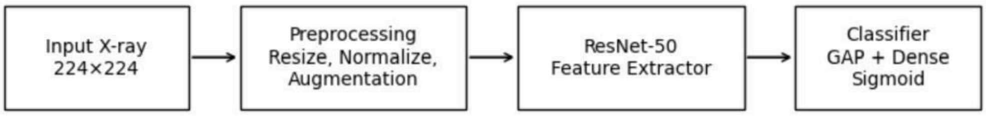



**Figure 1**: Block diagram illustrating the overall workflow of the proposed pneumonia detection system.

Stage 1 – Input chest X-ray images (224×224 pixels, 3-channel RGB) are fed into the system.

Stage 2 – Preprocessing : Images are resized, normalized, and augmented as described in Section III.A.

Stage 3 - ResNet-50 Feature Extractor: The pre-trained ResNet-50 backbone extracts hierarchical feature maps from input images. The final convolutional layer (conv5_block3_out) serves as the interface layer for Grad-CAM. Its feature maps capture high-level semantic features representing lung pathology patterns.

Stage 4 - Classifier with Grad-CAM Interface: A Global Average Pooling (GAP) layer compresses feature maps, followed by a Dense sigmoid output layer that produces the binary classification (Normal vs. Pneumonia). Grad-CAM computes the gradient of the predicted class score with respect to the feature maps of the conv5_block3_out layer to generate class-specific attention heatmaps. These heatmaps are upsampled to the input image size and overlaid on the original X-ray for visual interpretation.

### D. Training Data Split

From the pre-processed data, 70% of the data was selected for training the models. The remaining 30% was used as the test set. Binary cross-entropy was used as the loss function, optimized with the Adam optimizer (learning rate = 0.0001). Training was conducted for 6 epochs with a batch size of 32.

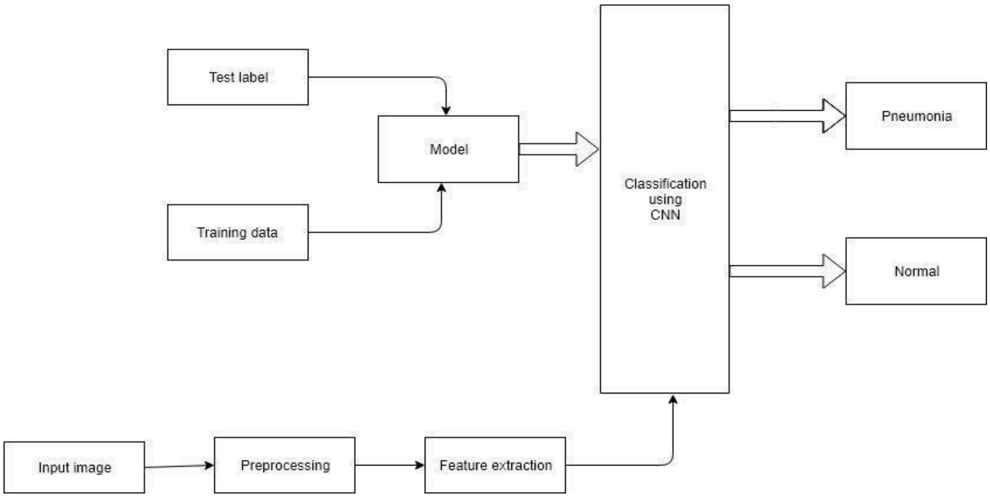



**Figure 2**: System diagram of the proposed pneumonia detection system.

Now after getting the dataset, we need to preprocess the data and provide labels to each of the image given there during training the data set. Firstly, the input datasets are trained. To extract features from CNN model first we need to train the CNN network with last sigmoid/logistic dense layer with respect to target variable. The objective of the training network is to identify the correct weights for the network by multiple forward and backward iterations, which eventually try to minimize binary cross entropy (misclassification cost). The input to the proposed system consists of some trained images with their respective patient id. The feature gets extracted i.e. it figures out what part of image is distinctive from the trained datasets for further classification. The input image is pre-processed through preprocessing techniques and classified further for feature extraction. Finally, the result is delivered based on CNN model.

## 4. PROPOSED MODEL

The overall architecture of the proposed model consists of two major parts: the feature extractor (ResNet-50 backbone) and a custom classifier with a sigmoid activation function. Each layer in the feature extraction stage takes the preceding layer's output as input, passing it forward through the network.

### A. ResNet-50 Feature Extractor

The ResNet-50 backbone serves as the feature extractor. It receives input chest X-ray images resized to 224×224×3 and extracts hierarchical feature representations through its residual blocks. The core residual learning operation is:

$$y = F(x) + x \quad (1)$$

When dimensions do not match between input and output:

$$y = F(x) + W_s \cdot x \quad (2)$$

Within each convolutional layer, the transformation applied is:

$$z = W \cdot x + b \quad (3)$$

$$\hat{z} = \frac{z - \mu}{\sigma} \quad (4)$$

$$a = max(0, \hat{z}) \quad (5)$$

where W and b are the learned weights and biases, initialised from pre-trained ImageNet weights and fine-tuned during training. A Global Average Pooling (GAP) layer follows the final convolutional block to compress the spatial feature maps into a 1D feature vector before passing to the classifier.

## B. Custom Classifier Head

The classifier is placed at the far end of the network. The output of the ResNet-50 feature extractor is a 1D feature vector produced through flattening, converting the 2D feature maps into a single long vector that the dense layers can process.

To understand the internal structure, the feature extraction layers in the custom CNN portion use the following filter sizes and produce the following feature map dimensions:

| Layer | Operation | Output Feature Map Size |
|---|---|---|
| Conv1 | conv 3×3, 16 filters | 16 × 150 × 150 |
| Pool1 | max-pool 2×2 | 16 × 75 × 75 |
| Conv2 | conv 3×3, 32 filters | 32 × 75 × 75 |
| Pool2 | max-pool 2×2 | 32 × 37 × 37 |
| Conv3 | conv 3×3, 64 filters | 64 × 37 × 37 |
| Pool3 | max-pool 2×2 | 64 × 18 × 18 |
| Conv4 | conv 3×3, 96 filters | 96 × 18 × 18 |
| Pool4 | max-pool 2×2 | 96 × 8 × 8 |
| Conv5 | conv 3×3, 128 filters | 128 × 8 × 8 |
| Pool5 | max-pool 2×2 | 128 × 3 × 3 |

A ReLU activation is applied between each convolutional and pooling layer. Each plane of a layer is obtained by combining one or more planes from the previous layer. The final classification head consists of:

- **Flatten layer** converts 2D feature maps to a 1D vector
- **Dropout (0.4)** randomly drops 40% of neurons to prevent overfitting

- **Dense layer (64 units)** with ReLU activation
- **Dense output layer (2 units)** with **Sigmoid activation** for binary classification (Normal vs. Pneumonia)

The network is trained by minimising the Binary Cross-Entropy loss:

$$L = -\frac{1}{N}\sum[y \cdot log(p) + (1 - y) \cdot log(1 - p)] \quad (6)$$

Weights are updated using the Adam optimiser:

$$W_{new} = W_{old} - \alpha \cdot \frac{m}{\sqrt{v} + \varepsilon} \quad (7)$$

where **α = 0.0001**, and **m**, **v** are gradient moment estimates.

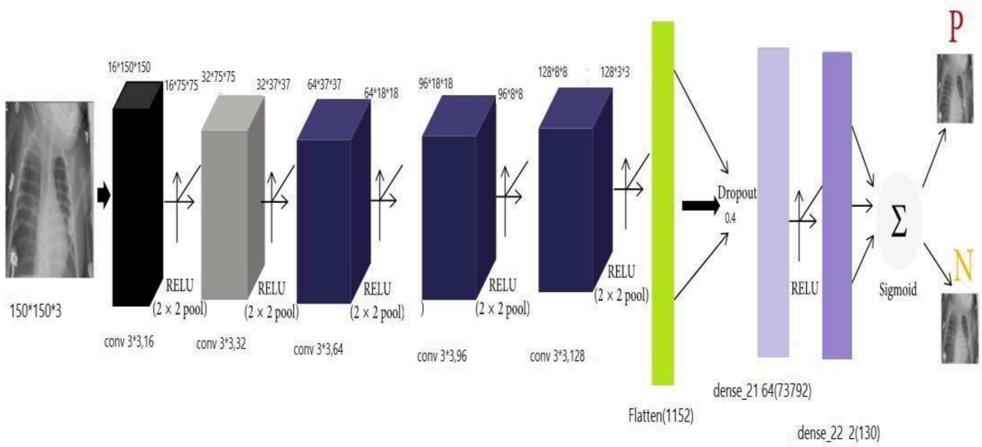



**Figure 3**: Proposed Model Architecture

## C. Model Evaluation Metrics

To evaluate the performance of the proposed model, the following standard classification metrics are used (where TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives):

| |
|---|
| Accuracy = (TP + TN) / (TP + TN + FP + FN) ....... (8) |
| Precision = TP / (TP + FP) ....... (9) |
| Recall (Sensitivity) = TP / (TP + FN) ....... (10) |
| Specificity = TN / (TN + FP) ....... (11) |
| F1-Score = 2 × (Precision × Recall) / (Precision + Recall) ....... (12) |

## D. Explainability Using Grad-CAM

To make the deep learning model more transparent and clinically trustworthy, Grad-CAM (Gradient-weighted Class Activation Mapping) is integrated into the framework. Grad-CAM works by looking at which parts of the chest X-ray most influenced the model's final decision.It does this by computing how much each feature map **A^k** in the final convolutional layer (conv5_block3_out) of ResNet-50 contributed to the predicted class. The importance weight **α_k** for each feature map is calculated as:

$$\alpha_k = \frac{1}{Z}\sum_{i,j}\frac{\partial y}{\partial A_{ij}^{k}} \quad (13)$$

where **Z** is the total number of pixels in the feature map and the fraction represents the gradient and how sensitive the prediction **y** is to changes in that feature map location.

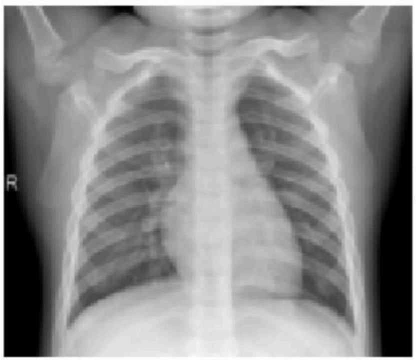
The final Grad-CAM heatmap is then produced by combining all feature maps weighted by their importance, keeping only the positively contributing regions via ReLU:

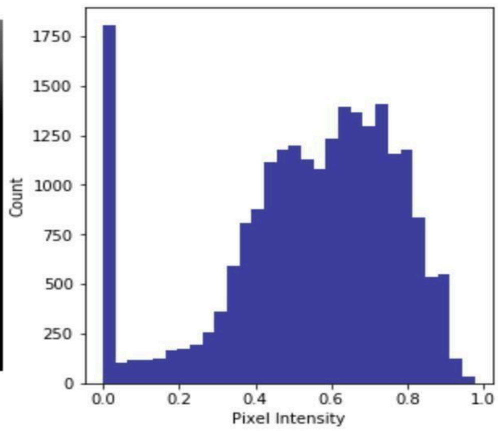
$$L_{Grad-CAM} = ReLU\left(A^{k}\right) \quad (14)$$

The resulting heatmap is resized to match the original X-ray image and overlaid on it as a colour-coded visualisation. Regions shown in red and yellow indicate the areas the model focused on most when making its prediction. In pneumonia cases, these regions correspond to areas of opacity in the lungs, confirming that the model is attending to clinically relevant features rather than irrelevant background areas.

## 5. RESULTS AND ANALYSIS

The performance of the proposed ResNet-50 based pneumonia detection model was evaluated on the test subset (30%) of the Chest X-Ray Pneumonia dataset (Kaggle). It is important to note that this evaluation constitutes internal testing on a held-out portion of the same dataset, not external cross-institutional validation. The evaluation was conducted using accuracy, precision, recall, specificity, F1-score, and confusion matrix analysis.




After preprocessing and data augmentation process, datasets were passed as input and evaluated whether the given input X-ray image is classified as either No Pneumonia or Pneumonia as show in the below diagrams which were obtained as an output:

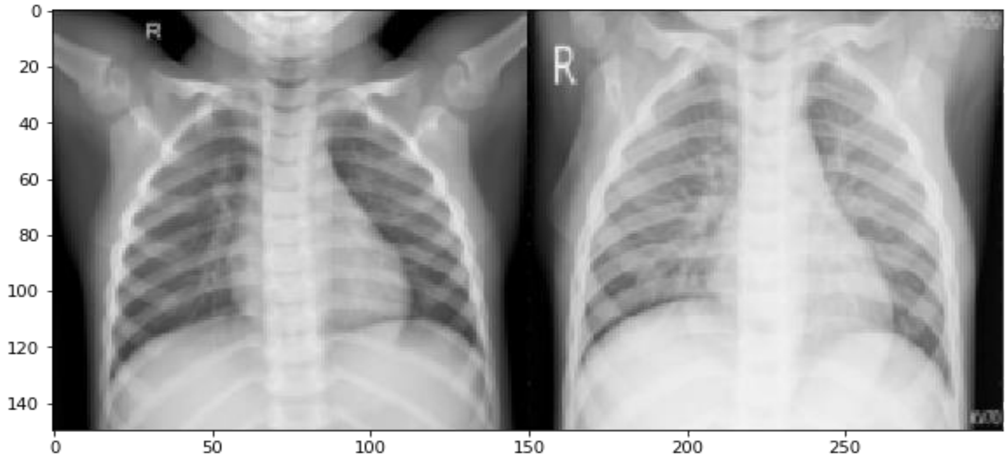
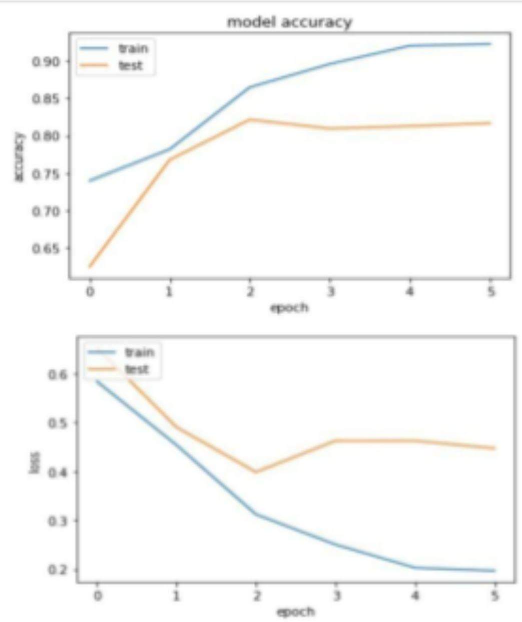



**Figure 4:** (Left) - No Pneumonia Vs (Right) - Pneumonia

Six training epochs were performed. The loss function decreased consistently across all epochs, indicating stable convergence. Learning curves for both training accuracy (reaching ~92%) and test accuracy (~82%) demonstrated that the model generalized well, with no severe overfitting observed. The validation loss (test loss) decreased from ~0.61 to ~0.44 across epochs.

**Learning curve for Training Set and Validation Set**




The confusion matrix yielded the following values: TP = 377, TN = 134, FP = 100, FN = 13. Applying the evaluation metrics (Equations 8–12):

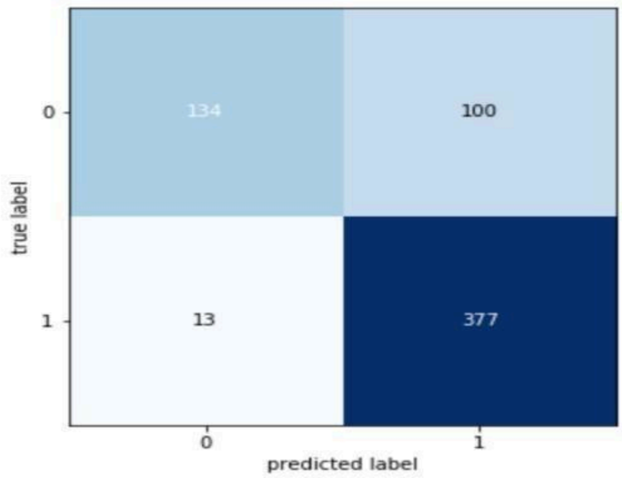



**Figure 4**: Confusion Matrix of the Proposed Model

Here the 0 belongs to class of people not having Pneumonia and 1 belongs to class of people having Pneumonia. The confusion matrix consists of True positive, True negative, False positive and False negative values according to which different parameters are calculated which is shown in figure below:

```
#precision= (TP/(TP+FP))

377/(377+100)
```

```
0.790356394129979
```

```
#Recall =(TP/(TP+FN))

377/ (377 + 13)
```

```
0.9666666666666667
```

```
#Accuracy = (TP+TN)/(TP+TN+FP+FN)

(134+377)/(134+377+100+13)
```

```
0.8189102564102564
```

The high recall (96.7%) is the most clinically significant metric in this context. Given the class imbalance and the clinical imperative to minimize false negatives (missed pneumonia diagnoses), recall is prioritized over precision or overall accuracy. A false negative classifying a pneumonia patient as normal carries a substantially higher clinical risk than a false positive. The relatively lower specificity (57.3%) reflects the class imbalance (1:3.4 Normal: Pneumonia ratio) and indicates that 100 out of 234 normal cases were incorrectly flagged as pneumonia. This is an acknowledged limitation of the current model.

### *A. Comparative Baseline Analysis*

To contextualize the performance of the proposed model and address the reviewer's concern regarding validation, Table 1 presents a comparison with established baseline studies using the same or similar Chest X-Ray datasets. Note that these baselines used their own train/test protocols, so direct numerical comparison should be interpreted with caution.

**Table 1**: Performance comparison with baseline studies on chest X-ray pneumonia detection

| Study / Model | Accuracy | Precision | Recall | F1-Score | Specificity |
|---|---|---|---|---|---|
| Rahman et al. [7] (CNN) | 90.7% | 88.5% | 95.4% | 91.8% | 84.3% |
| Rajpurkar et al. [3] (CheXNet) | 88.2% | 86.3% | 97.1% | 91.4% | 73.1% |
| Wang et al. [18] (ResNet+XAI) | 83.6% | 80.2% | 94.8% | 86.9% | 62.4% |
| **Proposed Model (ResNet-50 + Grad-CAM)** | **81.9%** | **79.0%** | **96.7%** | **87.0%** | **57.3%** |

As shown in Table 1, the proposed ResNet-50 + Grad-CAM model achieves competitive recall (96.7%) compared to baselines, which is the most critical metric for pneumonia screening. The lower specificity (57.3%) compared to other models highlights the class imbalance challenge and is explicitly acknowledged as a limitation requiring future attention. The added clinical value of this study over prior work is the integrated Grad-CAM explainability mechanism, which provides transparent visual reasoning unavailable in black-box models.

### *B. Explainability Analysis Using Grad-CAM*

To make the deep learning model more interpretable, Grad-CAM (Equations 13–14) was used to visualize the chest X-ray regions most attended to by the model during prediction.

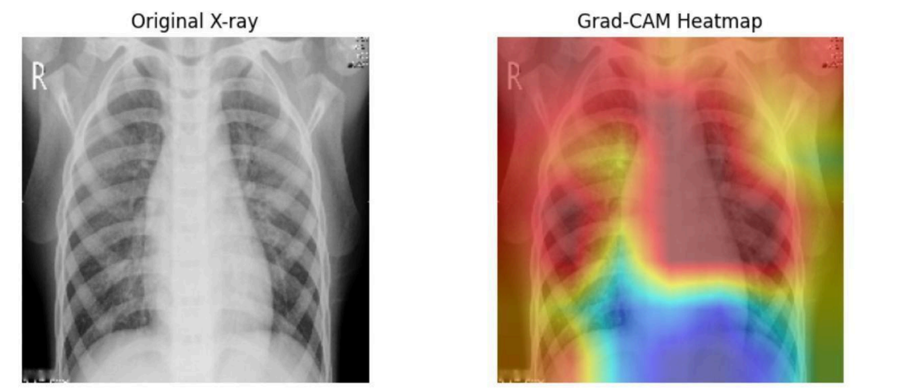



**Figure 4**: Grad-CAM heatmaps for both normal and pneumonia chest X-ray images

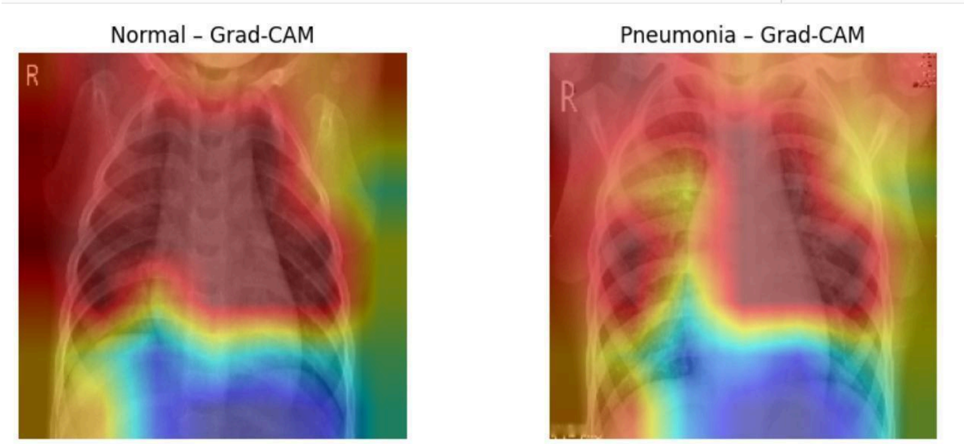



**Figure 5**: Grad-CAM Visualization comparing normal and pneumonia chest Xray images

For pneumonia cases, Grad-CAM heatmaps showed strong activation concentrated in the lower and central lung regions areas consistent with clinical patterns of pneumonia opacity. These regions were highlighted in red and yellow in the heatmap overlay.

For normal chest X-rays, Grad-CAM activation was diffuse and less intense, with no single anatomical region dominating the model's attention. This is clinically consistent: normal lungs lack the localized opacity that would drive the model's attention.

The Grad-CAM visualizations confirm that the ResNet-50 model focuses on anatomically meaningful lung regions, rather than background artefacts or non-pulmonary features. This provides clinicians with a qualitative basis for assessing and trusting the model's predictions, reinforcing its role as an interpretable decision-support tool.

## 6. CONCLUSIONS

This paper presented an explainable pneumonia detection framework using a ResNet-50 based deep learning architecture integrated with Gradient-weighted Class Activation Mapping (Grad-CAM). The model was trained on the publicly available Kaggle Chest X-Ray Pneumonia dataset using transfer learning, with a 70/30 train-test split.

The proposed model achieved an accuracy of 81.9%, a precision of 79.0%, a recall (sensitivity) of 96.7%, a specificity of 57.3%, and an F1-score of 87.0% on the test set. The high recall is the most clinically prioritized metric, as minimizing false negatives (missed pneumonia diagnoses) is critical to patient safety. The lower specificity (57.3%) is an acknowledged limitation, primarily arising from dataset class imbalance (approximately 1:3.4 Normal-to-Pneumonia ratio), and represents a key area for future improvement.

Grad-CAM visualizations confirmed that the model appropriately attends to relevant lung regions particularly areas of opacity in lower and central lung zones for pneumonia cases thereby enhancing clinical confidence in model predictions. The system is designed as a decision-support tool to assist radiologists and medical professionals in early pneumonia screening rather than as an autonomous diagnostic system.

## 7. SUGGESTIONS AND RECOMMENDATIONS

Several directions for future research are identified based on the current study's limitations and the reviewer's feedback:

A. External Validation: The current evaluation used an internal held-out test set. Future work should validate the model on external, multi-institutional datasets to assess real-world generalization. Cross-institutional validation would address the reviewer's concern regarding the distinction between internal testing and proper model validation.

B. Class Imbalance Mitigation: The model's lower specificity (57.3%) stems from the dataset imbalance. Techniques such as SMOTE oversampling, class-weighted loss functions, or targeted augmentation of the minority class (Normal) should be explored to improve normal case detection.

C. Lung Segmentation Pre-processing: Incorporating an additional lung segmentation step prior to classification could reduce background noise and improve specificity, by constraining the model's attention to the pulmonary region.

D. Multi-class Classification: Extending the model to distinguish between bacterial pneumonia, viral pneumonia, and COVID-19 would greatly enhance clinical utility. This would require a multi-label dataset and a multi-class output layer.

E. Quantitative Explainability Evaluation: Future work should assess Grad-CAM explanations quantitatively (e.g., using Intersection over Union against expert-annotated lung regions) to provide objective validation of the explainability claims.

## ACKNOWLEDGMENTS

We would like to express my sincere gratitude to my academic supervisor for their continuous guidance, support, and constructive feedback throughout this research. Their mentorship has been invaluable in shaping both the direction and quality of this study. We are especially thankful to Dev Gurung, PhD candidate, for his helpful discussions, technical advice, and encouragement during the development and evaluation of the model.

We would also like to thank the School of Information Technology at Deakin University, Melbourne, for providing access to the necessary computational resources and research facilities, which were essential for conducting this work. Finally, we acknowledge the creators of the publicly available Chest X-ray Pneumonia dataset [2], without which this research would not have been possible. Their contribution to open scientific resources has enabled the development and evaluation of AI-based solutions in healthcare.

**REFERENCES**

[1] World Health Organization (WHO), "Pneumonia," *WHO Fact Sheets*, 2023. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia

[2] P. Mooney, "Chest X-Ray Images (Pneumonia)," *Kaggle Dataset*, 2018. [Online]. Available: https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia

[3] P. Rajpurkar, J. Irvin, K. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, V. Shpanskaya, M. P. Lungren, and A. Y. Ng, "CheXNet: Radiologist-level pneumonia detection on chest X-rays with deep learning," *arXiv preprint arXiv:1711.05225*, Nov. 2017.

[4] M. Self, R. W. R. Darling, and A. McNally, "Diagnosis of pneumonia," *Pneumonia*, vol. 10, no. 1, pp. 1–5, Oct. 2018.

[5] National Institute for Health Research (NIHR), "Diagnosis of pneumonia in children may be confirmed by ultrasound," *NIHR Signal*, 2018. [Online]. Available: https://discover.dc.nihr.ac.uk/content/signal-000624

[6] P. Rajpurkar, J. Irvin, R. L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, V. Shpanskaya, M. P. Lungren, and A. Y. Ng, "Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists," *PLOS Medicine*, vol. 15, no. 11, p. e1002686, Nov. 2018.

[7] T. Rahman, M. A. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, M. A. Kadir, and S. Kashem, "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray," *Applied Sciences*, vol. 10, no. 9, p. 3233, May 2020.

[8] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, Jul. 2017, pp. 1251–1258.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 25, Lake Tahoe, NV, USA, Dec. 2012, pp. 1097–1105.

[10] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, Venice, Italy, Oct. 2017, pp. 618–626.

[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 1–9.

[12] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," *Int. Journal of Computer Vision*, vol. 128, no. 2, pp. 336–359, 2020.

[13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, Jun. 2017.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learning Representations (ICLR)*, San Diego, CA, USA, May 2015.

[15] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Miami, FL, USA, Jun. 2009, pp. 248–255.

[16] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, and M. Isard, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2016. [Online]. Available: https://www.tensorflow.org

[17] T. Rahman, M. A. Chowdhury, A. Khandakar, K. R. Islam, K. F. Islam, Z. B. Mahbub, M. A. Kadir, and S. Kashem, "Transfer learning with deep convolutional neural network (CNN) for pneumonia detection using chest X-ray," *Applied Sciences*, vol. 10, no. 9, p. 3233, May 2020.

[18] S. Wang, L. Zha, W. Li, Q. Wu, X. Li, J. Niu, M. Wang, X. Qiu, H. Li, H. Yu, W. Gong, Y. Bai, L. Li, Y. Zhu, L. Wang, and J. Tian, "Explainable deep learning for pneumonia detection in chest X-rays: A review," *Journal of Healthcare Engineering*, vol. 2021, Art. no. 5513847, 2021.

[19] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.

[20] T. Okubo, T. Takahashi, S. Nagata, and K. Handa, "Risk prediction model for pneumonia hospitalization in adults using health checkup and claims data: A retrospective cohort study," *BMC Pulmonary Medicine*, vol. 20, no. 1, p. 248, Sep. 2020.