1 Methodology

Pneumonia detection is approached in this study using a deep learning methodology, specifically a convolutional neural network (CNN) on a Kaggle dataset of 5,863 chest X-ray images. The dataset, which was compiled from pediatric patients at Guangzhou Women and Children's Medical Center, is clinically relevant and diverse. This dataset was chosen because of its large size, clinical origin, and balanced distribution of Pneumonia and Normal cases. For accurate Pneumonia detection, the proposed CNN model aims to effectively distinguish between these classes.

1.1 Dataset

In this study, we utilized a dataset sourced from Kaggle , which is organized into subfolders for each image category (Pneumonia/Normal) and further structured into three folders (train, test, and val). The dataset encompasses a total of 5,863 X-ray images (JPEG format), with two distinct categories: Pneumonia and Normal. The images were acquired through standard clinical procedures for pediatric patients aged one to five, from retrospective cohorts at the Guangzhou Women and Children's Medical Center in Guangzhou. Notably, the dataset is characterized by an imbalanced distribution, with a ratio of 1,583 images labeled as Normal to 4,272 images labeled as Pneumonia.

In Figure.1 , The normal chest X-ray (left panel) depicts clear lungs without any areas of abnormal pacification in the image. Bacterial pneumonia (middle) typically exhibits a focal lobar consolidation, in this case in the right upper lobe (white arrows), whereas viral pneumonia (right) manifests with a more diffuse "interstitial" pattern in both lungs.

Figure 1: Chest X-ray comparisons—Normal (left), bacterial pneumonia (middle), and viral pneumonia (right)

1.2 Data Pre-Processing

Different Pre-processing techniques can be used to prepare the dataset for effective analysis this involves resizing all the images and normalizing the pixel values, ensuring standardized input for the neural network. Normalization improves an image, enlarging its brightness to fill the entire dynamic range to reduce the distribution of noise.

As part of the pre-processing phase, the dataset is divided into training, validation, and test sets. Subsequently, the model is trained using the training set and validated on the validation set.The model is trained to recognize patterns and features in chest X-ray images that indicate the presence or absence of pneumonia. It represents the first step in utilizing convolutional neural networks (CNNs) for pneumonia detection, emphasizing the significance of the pre-processing steps in preparing the data for effective model training and validation.

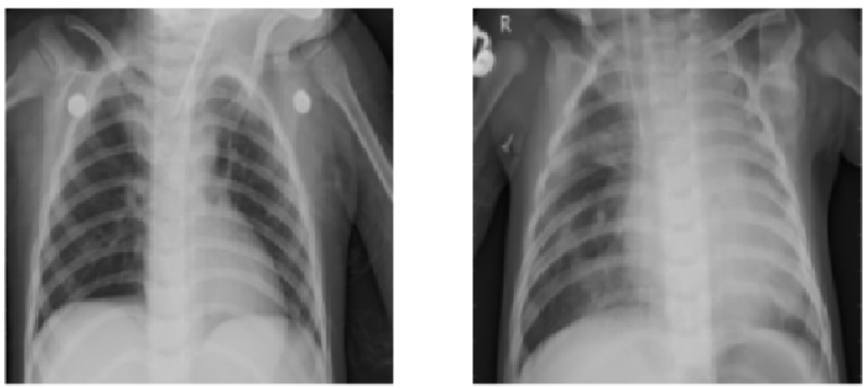

Figure 2: Example of Resized and Normalized Chest X-ray Normal (Left) Pneumonia (Right)

1.3 Model Architechture

The input data is fed into the model for feature extraction following the preprocessing stage. Features from the images can be extracted in this step and fed into the classification for use in the following prediction processes.Two deep learning models, RestNet50 [1] and VGG16 [2], have been used in this work.

Each component in the creation of these neural network models is critical to a powerful feature extraction process. Convolutional layers detect intricate patterns, beginning with the input layer for pixel representation, and pooling layers to minimize feature maps for efficiency. Dense layers allow for extensive data flow, which leads to final predictions from the output layer with softmax activation. Non-linearity is introduced by activation functions such as ReLU and sigmoid, and diverse optimizers such as Adam optimize weights during training. A learning rate of 0.0001 ensures gradual convergence. . These options represent a well-thought-out design for effective feature extraction in image classification.

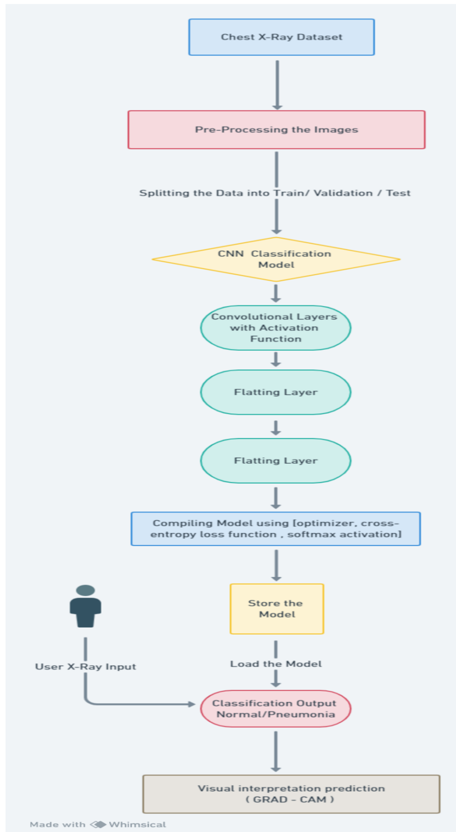

Figure 3: Proposed Framework Model for Pneumonia Prediction

1.4 Evaluation Metrics

the estimated results of the suggested model are based on a number of measures, including accuracy, precision, F1-score, and recall. Four concepts, ”false positive,” ”true positive,” ”false negative,” and ”true negative,” need to be clearly defined in order to use these measurements. The term ”false positive (FP)” describes samples that are expected to belong to positive classes but actually belong to negative classes. Samples that are both positive and members of the positive class are referred to as ”true positives (TP)”. Samples classified as ”false negative (FN)” are those that are expected to belong to the negative classes but really belong to the positive classes. Samples that are successfully predicted and correspond to the negative classes are referred to as ”true negatives (TN)”.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

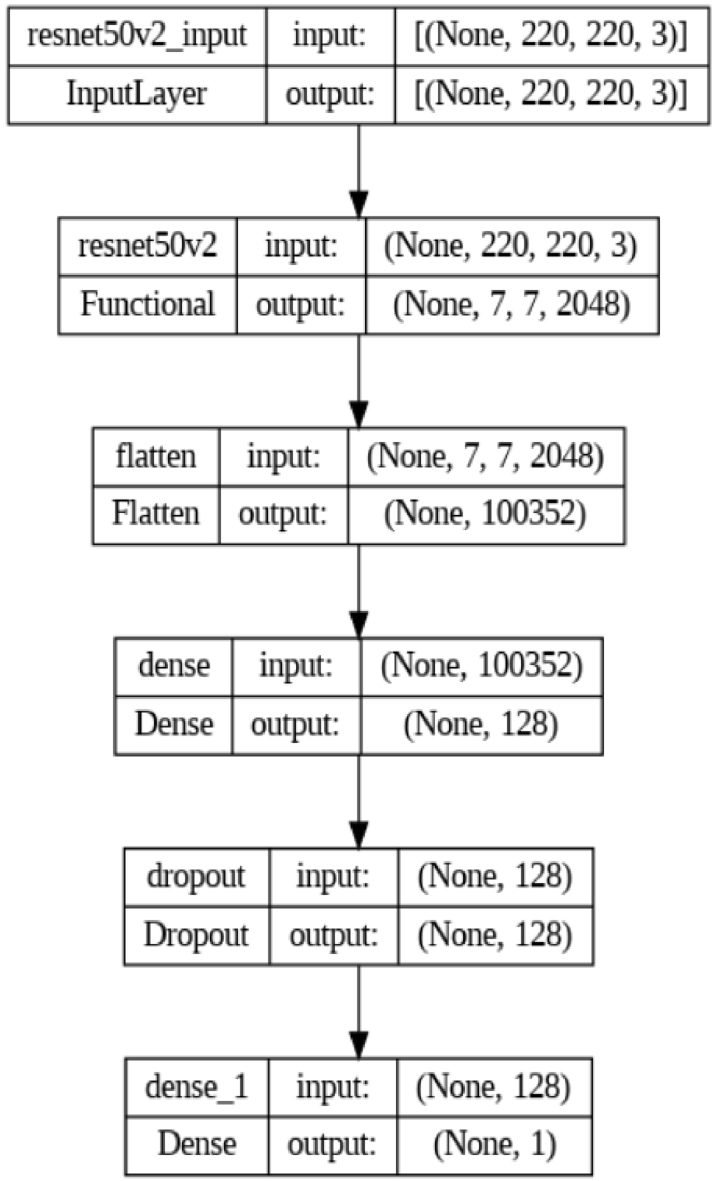

Figure 4: RestNet50 Model Architecture

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3}$$

$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

1.5 Prototype Results

In this comparative analysis, our proposed deep learning model, integrating RestNet50 architectures, emerges as a powerhouse in Pneumonia detection. The model show cases remarkable prowess with an accuracy of 97 percent, precision of 98 percent, signifying a low rate of false positives, recall of 96 percent, and an F1 score of 97 percent. These metrics underscore its robust capabilities in accurately classifying instances and achieving a harmonious balance between precision and recall. Comparatively, the first model, a CNN-based architecture, demonstrates competitive performance with a recall of 96 percent, precision of 92 percent, accuracy of 92 percent, and an F1 score of 94 percent. While this model exhibits commendable proficiency, the superiority of our proposed model is evident in its higher accuracy and precision.

Model	Evaluation Metrics			
	Accuracy	Precision	Recall	F1 Score
RestNet50	0.97	0.98	0.96	0.97

Table 1: Model Evaluation Results

The second model, employing VGG-19 and ResNet-50, presents mixed results. ResNet-50, with an accuracy of 82.8 percent, faces challenges in capturing intricate patterns, potentially limiting its utility in Pneumonia detection. On the other hand, an enhanced CNN model utilizing VGG-19 show cases the highest accuracy at 92.4 percent. This comparative analysis serves as a compelling testament to the power of our model, positioning it as a standout choice for Pneumonia detection. The exceptional accuracy and precision achieved underscore its potential for real-world applications, show casing its strength in making accurate predictions while minimizing false positives.

Model	Evaluation Metrics			
	Accuracy	Precision	Recall	F1 Score
CNN [3]	0.92	0.92	0.96	0.94
VGG-19 [4]	0.92	-	-	-
RestNet-50 [4]	0.82	-	-	-

Table 2: Model Comparison Results

For this dataset, the outcomes are also shown in a ROC graph, as shown in Fig. 5, which is helpful in displaying the degree of separatability. ROC highlights the correlation between recall and precision. It is used to graphically represent the evaluation of binary classification. Other ways produce a single value to represent performance. The ROC curve is a well-known curve for identifying a good classifier; the closer the ROC is to the upper left of the graph, the better the model.

The ROC curve is an important way for determining how well a binary classifier performs with different classification thresholds. The standard threshold is 50percent, but the ROC curve allows us to find an optimal threshold that balances sensitivity and specificity. To find the optimal threshold, look for the point on the ROC curve closest to the upper-left corner, or use metrics such as the Youden's Index. Choosing the appropriate threshold is critical, especially since false positives and false negatives have different consequences. This threshold, once determined, can be used to achieve a balanced trade-off between false positives and false negatives in model predictions.

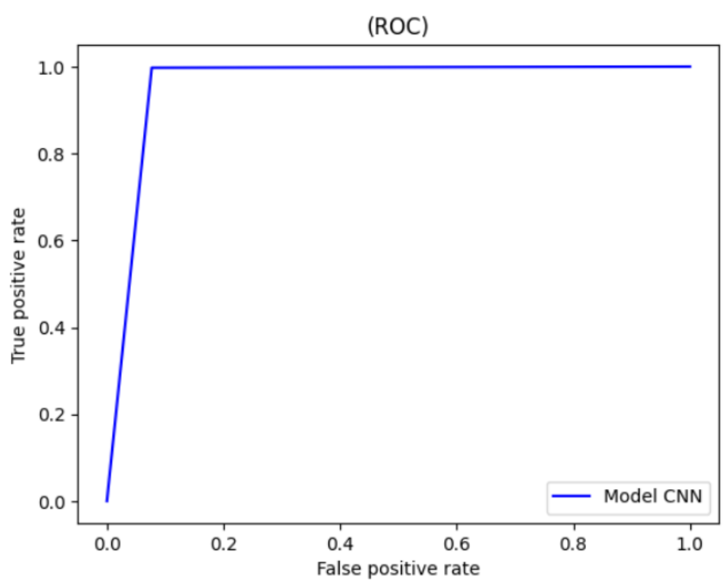

Figure 5: ROC Curve for RestNest50

Figure .6 We evaluate our models by computing accuracy and accuracy validation. It measures the ratio of correctly predicted instances to the total instances. Accuracy is a common metric for classification problems. In addition to Monitor the training and validation loss during the training process. The model is learning when the training loss decreases, but to make sure the model is not overfitting, it is crucial to keep an eye on the validation loss.

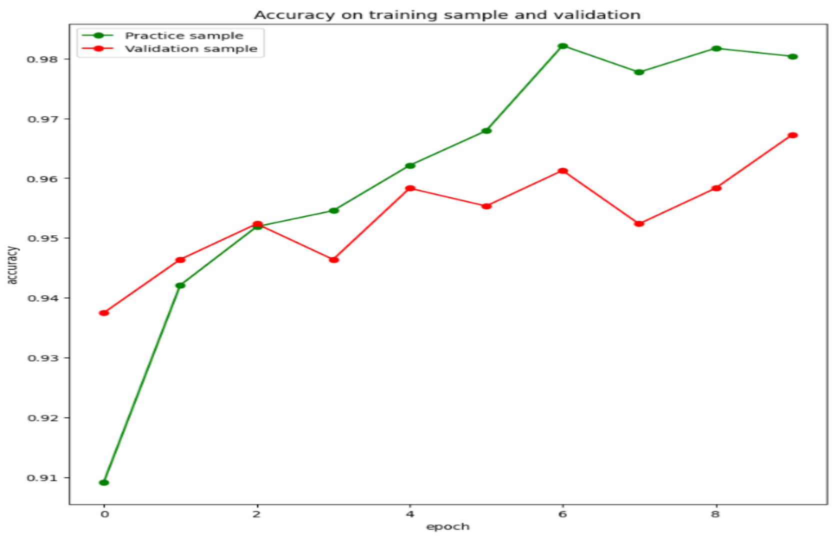

Figure 6: Accuracy and Loss Function on Training and Validation

1.6 Model Interpretation

Grad-CAM, or Gradient-weighted Class Activation Mapping, which stands for Gradient-weighted Class Activation Mapping, is a technique used in computer vision to visualize and understand the decisions made by a neural network, particularly in the context of image classification. It helps highlight the regions of an image that are influential in the network's decision-making process ,the key advantage of Grad-CAM is that it provides a fine-grained localization of important features in an image, making it a powerful tool for visualizing and interpreting deep learning models. It builds upon Class Activation Mapping (CAM), which generates a heatmap pinpointing regions contributing most to a class prediction, typically in the final convolutional layer of a CNN. Grad-CAM enhances CAM by incorporating gradients of the target class's output concerning the feature maps of the last convolutional layer. These gradients act as weights in a weighted sum, emphasizing regions crucial for the model's prediction. The resulting heatmap, reflecting influential areas, is then upsampled to the original input size, providing a visual representation of the image regions deemed most relevant by the model.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[2] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[3] Batyrkhan Omarov and Inkar Bazarkulova. A deep learning model for pneumonia detection from x-ray images. *Scientific Journal of Astana IT University*, 14, 2023.

[4] Shadi A Aljawarneh and Romesaa Al-Quraan. Pneumonia detection using enhanced convolutional neural network model on chest x-ray images. *Big Data*, 2023.