Advances in Computer Science and Information Technology (ACSIT)
p-ISSN: 2393-9907; e-ISSN: 2393-9915; Volume 5, Issue 2; April-June, 2018, pp. 104-119

http://www.krishisanskriti.org/Publication.html

Deep Learning-Based Framework for Pneumonia Diagnosis from Chest X-Ray Images

Shabana Nargis Rasool[a]

[a]*Department Of Computer Science, Islamic University of Science and Technology Kashmir India, shabana.nargis@islamicuniversity.edu.in*

Abstract

Pneumonia is a disease that affects the interstitial tissue in the lungs and is the primary cause of death among children below the age of five. In 2016, it caused the demise of roughly 880,000 children, making up around 16% of all fatalities in this age group, as per a UNICEF study. Most of those who were impacted were under the age of two. Early identification of pneumonia in children can accelerate the recovery process. To diagnose pneumonia, a radiologist interprets a chest X-ray (C-XRY). However, relying solely on human expertise to diagnose pneumonia has drawbacks, such as the cost and availability of experienced professionals. Thus, an automated method for detecting pneumonia from X-rays has become essential. In various fields, such as medical diagnosis tasks, deep learning (DL) methods are extensively employed to create strong classification models that demonstrate high efficacy. The objective of this study was to devise deep learning models that can spot pneumonia in chest X-ray images (C-XRY IMGs). Three models were constructed using basic convolutional neural networks (CNNs) and transfer learning (TL) methodologies. This study utilised a dataset of 4185 C-XRY images to train the models, and 1047 C-XRY images were used for validation. Subsequently, the models were tested using 624 additional C-XRY images, which yielded an impressive accuracy of 91 percent, f1-score of 0.93 and Auc of 97 percent.

Keywords: Deep learning, transfer learning, Pneumonia, Chest X-ray, Convolutionel Neural Networks.

1. Introduction

Pneumonia is a major contributor to mortality rates globally, particularly among vulnerable populations such as young children and the elderly. Early and accurate detection of pneumonia is crucial for effective treatment and improved patient outcomes. C-XRY imaging is a widely used tool for the diagnosis of pneumonia, but the interpretation of these images can be challenging, especially for inexperienced radiologists. Conventional methods of identifying pneumonia through C-XRY images depend on a radiologist‘s manual analysis. However, this approach is often limited by the subjectivity of the radiologist‘s interpretation, leading to variability in the accuracy and reliability of the diagnosis. Moreover, the manual interpretation of chest X-rays (C-XRYs) can be time-consuming, limiting the efficiency of the diagnosis process.

Automatic pneumonia detection using chest X-rays has emerged as a promising approach to address the shortcomings of conventional methods. The significance of automatic pneumonia detection using C-XRYs is multifold. Firstly, it can improve the accuracy and consistency of pneumonia diagnosis, enabling prompt and appropriate treatment, which can lead to better patient outcomes. Secondly, automatic detection systems can reduce the burden on radiologists, allowing them to focus on more complex and specialised tasks. Finally, these systems can increase the efficiency of the diagnosis process, enabling more patients to be diagnosed and treated in a shorter period of time, which can have significant implications for public health. However, the significant challenge in medical image analysis is to develop accurate and reliable methods for image classification. In traditional approaches, image classification is performed by extracting hand-crafted features from the images, which requires expert knowledge and is time-consuming. Furthermore, the manual feature extraction process is often prone to errors and can result in incomplete or inconsistent feature representations. Another limitation of traditional approaches is that they are not well-suited for large-scale datasets.

To overcome these limitations, DL methods have been increasingly used in medical image analysis. DL models, such as CNNs, have the ability to acquire feature representations from unprocessed images through automatic learning, eliminating the need for manual feature extraction [1]. Moreover, these models can handle large-scale datasets and have shown “state-of-the-art” performance in various medical imaging tasks.

In this paper, we investigate the effectiveness of DL models for automatic

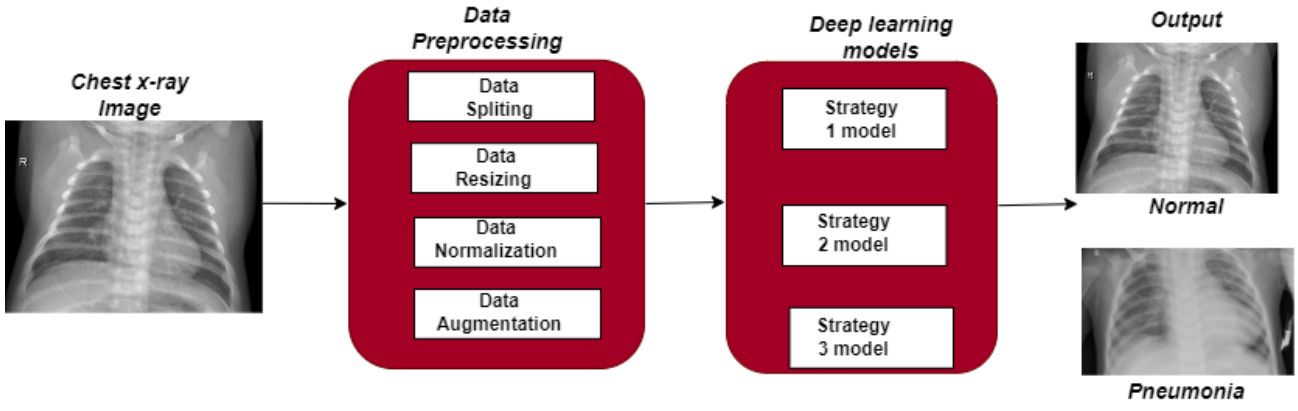


Figure 1: **The proposed deep learning framework for pneumonia diagnosis**

pneumonia detection using C-XRY IMGs. Our proposed deep learning framework can achieve high accuracy and efficiency, which could have significant implications for improving the diagnosis and treatment of pneumonia. The proposed deep learning framework is shown in Figure 1.

The remaining sections of the paper are organised as follows: Section 2 and 3 presented the related work and background of deep learning algorithms. Section 4 expounds on the methodology employed in the study, including details on the dataset, pre-processing procedures adopted to prepare the data for training, validation, and testing, and the proposed strategies adopted. The outcomes of the classification algorithms and the findings are discussed and presented in Section 5. Ultimately, Section 6 concludes the paper.

2. Related Work

Pneumonia is a severe respiratory disease that can lead to morbidity and mortality if not detected and treated in a timely manner. C-XRYs are a common diagnostic tool used for detecting pneumonia, and as a result, there are a variety of approaches that have been documented in academic literature for the automated detection of pneumonia through the use of C-XRY IMGs.Some traditional machine learning(ML) methods typically involve extracting handcrafted features from C-XRY IMGs using techniques such as edge detection, texture analysis, and shape analysis. Once the features are extracted, a machine learning algorithm such as “support vector machine”, “random forest”, or “logistic regression” is trained on the features to classify the images as either pneumonia-positive or pneumonia-negative. Whereas

DL methods automatically acquire characteristics from C-XRY IMGs, eliminating the requirement for explicit feature extraction. CNNs are the prevailing DL models employed to identify pneumonia, where the network is trained on a large dataset of C-XRY IMGs with corresponding labels (i.e., pneumonia-positive or pneumonia-negative). The trained CNN can then be used to predict whether a new C-XRY image contains signs of pneumonia. The researchers in [2] initially utilized logistic regression as a reference model to identify pneumonia in C-XRY images. However, the performance of the logistic regression model, as measured by the Area Under the Curve (AUC), was inadequate. To enhance the model's accuracy, they implemented a 121-layer DenseNet model, which was trained using an Adam optimizer. The DenseNet model achieved a slightly better AUC score of 0.609 compared to the logistic regression model's AUC of 0.60. Additionally, they mentioned that only 1% of the dataset contained C-XRY IMGs with pneumonia.

Pneumonia is a severe respiratory disease that can lead to morbidity and mortality if not detected and treated in a timely manner. C-XRYs are a common diagnostic tool used for detecting pneumonia, and as a result, there are a variety of approaches that have been documented in academic literature for the automated detection of pneumonia through the use of C-XRY IMGs.Some traditional machine learning (ML) methods typically involve extracting handcrafted features from C-XRY IMGs using techniques such as edge detection, texture analysis, and shape analysis. Once the features are extracted, a machine learning algorithm such as "support vector machine", "random forest", or "logistic regression" is trained on the features to classify the images as either pneumonia-positive or pneumonia-negative. Whereas DL methods automatically acquire characteristics from C-XRY IMGs, eliminating the requirement for explicit feature extraction. CNNs are the prevailing DL models employed to identify pneumonia, where the network is trained on a large dataset of C-XRY IMGs with corresponding labels (i.e., pneumonia-positive or pneumonia-negative). The trained CNN can then be used to predict whether a new C-XRY image contains signs of pneumonia. The researchers in [2] initially utilized logistic regression as a reference model to identify pneumonia in C-XRY images. However, the performance of the logistic regression model, as measured by the Area Under the Curve (AUC), was inadequate. To enhance the model's accuracy, they implemented a 121-layer DenseNet model, which was trained using an Adam optimizer. The DenseNet model achieved a slightly better AUC score of 0.609 compared to the logistic regression model's AUC of 0.60. Additionally, they mentioned

that only 1% of the dataset contained C-XRY IMGs with pneumonia.

The researchers mentioned in [3] developed a model called ChexNet that uses a 121-layer CNN to analyze C-XRY IMGs and determine the likelihood of pneumonia. The model is designed to classify the images as either showing the presence or absence of pneumonia and also to locate the pneumonia using a thermal map. The dataset used to train and test the ChexNet model was ChestX-ray14, which was provided by the authors [4]. The researchers faced challenges in developing the model due to the difficulty of accurately diagnosing pneumonia from C-XRY IMGs. Additionally, the quality of the images was not always clear, and access to patient files was restricted. To analyse the effectiveness of the ChexNet model, the researchers compared its results to those of four radiologists using the F1 score. The F1 score for the ChexNet model was 0.435, which is higher than the average F1 score of the radiologists (0.387). Overall, the ChexNet model developed by the researchers in [3] shows promise for accurately diagnosing pneumonia from C-XRY IMGs, and its performance surpasses that of a average radiologist. However, further research is needed to validate the model's performance on larger and more diverse datasets and to address the limitations and challenges faced during its development.

In [5] researchers overcome the limitations of traditional machine learning methods by exploiting the potential of transfer learning, which involves training a CNN on a large dataset from a related task, and then fine-tuning the network on a smaller dataset for a specific application. They used a multisource dataset, including CT scans and X-ray images of lungs, to train the CNN for lung pattern analysis. They then fine-tuned the network on a smaller dataset of CT scans to achieve improved performance.

In [6] authors used a large dataset of chest radiographs, and the deep learning model was trained to detect the presence of pneumonia. The authors evaluated the performance of the model on a separate validation set and found that it had high accuracy in detecting pneumonia.However, when the model was tested on a different set of data from a different institution, its performance was found to be significantly lower. The study found that the performance of the model was affected by differences in the distribution of data and the quality of the images used in training and validation.The results suggest that the generalization performance of deep learning models can vary depending on the data used for training and validation, and caution should be exercised when applying such models to new datasets or populations.

In [7] the authors describes a deep learning model that was developed

to classify childhood pneumonia as either bacterial or viral based on chest radiography images. The model was trained on a dataset of 4,281 images and achieved an accuracy of 86.5% on a test set of 300 images. The authors suggest that the model could be used to improve the speed and accuracy of pneumonia diagnosis, particularly in resource-limited settings where access to skilled radiologists may be limited.

In [8] authors present a method for training a CNN to classify chest radiographs into those with pulmonary tuberculosis and those without. Their results show that the deep learning algorithm has high accuracy and could potentially be used for screening chest radiographs for pulmonary tuberculosis.

In [9], the RSNA dataset, which was obtained from Kaggle, was utilized. The dataset consisted of 26,684 C-XRYs, with 6,000 of the images originating from pneumonia patients, and 20,000 images from normal patients. To reduce runtime, the images were transformed into PNG format and resized. They designed a sequential CNN model using "RGB images", which employed maximum pooling to capture the maximum pixel value within a defined area of interest. and subsequently flattened the output. The data were divided into three models, namely "Model 1, Model 2a, and Model 2b". These models were trained on varying datasets using the same architecture. Model 1 was utilized for classifying images as having pneumonia or not, while Model 2a was used for differentiating between normal and opaque images. Model 2b employed the output from Model 2a to reclassify images as opacity or pneumonia. The accuracy achieved by Model 1, Model 2a, and Model 2b were 78.5%, 68.5%, and 69.9%, respectively.

The authors mentioned in [10] developed a dual convolutional neural network (CNN) to automatically recognize front and lateral C-XRY IMGs from the "MIMIC-CXR dataset", which is currently the most extensive dataset of its kind. The CNN was designed to identify common thorax diseases. The dataset was partitioned into training, testing, and validation sets, with 70% allocated for training, 20% for testing, and 10% for validation. The model's average AUC for posteroanterior (PA) and anteroposterior (AP) views were 0.721 and 0.668, respectively. To further enhance the performance of the model in identifying common thorax diseases, the authors intend to implement data augmentation and pixel normalisation techniques in their workflow. These techniques are expected to assist the process of detecting common thorax disease.

3. Background of Deep Learning Algorithms

3.1. Convolutional Neural Network (CNN) Models

CNN models have gained popularity because of their ability to improve image classification performance. By utilizing convolutional layers and filters, spatial and temporal characteristics can be extracted from an image. Additionally, these layers use a weight-sharing technique that reduces the amount of computation required [11]. In terms of their architecture, CNN models are a type of feedforward artificial neural network (ANN) that adheres to two specific rules. Firstly, the neurons within a given filter are only linked to local sections of the image in order to maintain spatial structure. Secondly, the weights of these neurons are shared, which helps to minimize the total number of parameters in the model. CNN models are comprised of three key components: (i) a convolutional layer that enables the network to learn features, (ii) a max-pooling (subsampling) layer that reduces image dimensionality and computational demands, and (iii) a fully connected layer that gives the network the ability to classify data [12].

3.2. Deep Transfer Learning Models

Deep transfer learning models refer to a category of deep learning models that transfer knowledge acquired from one task to another. In transfer learning, a pre-trained model that has learned to classify images on a large dataset such as ImageNet is used as a starting point for a new task, which may involve a different dataset or a different set of classes. The pre-trained model is then fine-tuned for the new task to improve its performance. This eliminates the need to utilise a massive dataset and also decreases the lengthy training period, which is mandatory for deep learning algorithms when built from scratch [13] [14].

4. Methodology

4.1. Dataset

The dataset used in our proposed framework contains 5232 images for the purpose of training and 624 images for testing, with resolutions varying from 712 $\times$ 439 pixels to 2338 $\times$ 2025 pixels [15]. The dataset includes three distinct categories of images: Normal, Bacterial Pneumonia, and Viral Pneumonia. Figure 2 shows the images of the dataset.

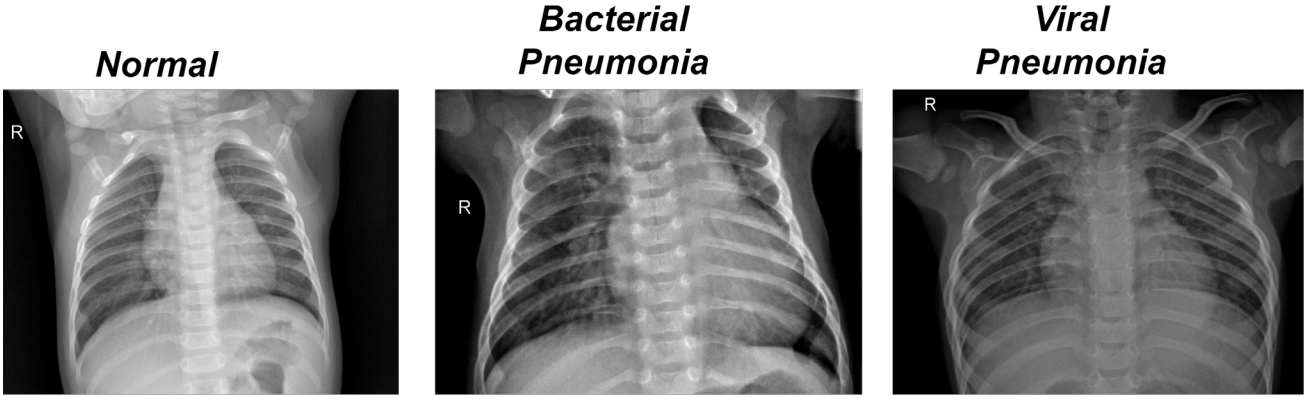


Figure 2: **Shows the Chest X-Rays dataset images of Normal, Bacterial Pneumonia, and Viral Pneumonia**

4.2. Preprocessing and Data Augmentation

Image pre-processing steps such as data splitting, resizing, normalization, and augmentation are involved in our experimentation. Here, 20% of the training dataset is used for validation. Then images of training, validation, and test sets are resized to 224 × 224 and normalized. To address the issue of overfitting and improve the model's ability to generalize during training, as well as enhance the size and quality of the dataset, we employed some data augmentation techniques. The specific parameters used for the data augmentation are outlined in Table 1. Here, we randomly apply a zoom range of 10% to the images during training. Then we apply a width shift to horizontally offset the images and a height shift to vertically offset the images, both at 10%.

Table 1: **summarizes the settings used for image augmentation**

Methods	setting
Zoom range	**0.1**
Width shift range	**0.1**
Height shift range	**0.1**

4.3. Strategies Adopted

Three distinct deep learning strategies were utilized.

4.3.1. Strategy 1

Convolutional neural networks (CNNs) are utilized for identifying pneumonia diagnosis in C-XRYs. The proposed CNN model consists of layers for input as well as layers for extracting features and performing classification. The input layer has dimensions of 224 x 224 x 3. The feature extraction component of the model is made up of three blocks of convolution, batch normalization, and ReLU layers. Each block may also include a maximum pooling and dropout layer. The result obtained from the feature extraction component is then passed through a flattening layer that converts the shape of the data into a single, one-dimensional vector that can be used by the classification dense layer. The dense layer is a typical and frequently used layer in neural networks, which connects every input to every output. This model employs a pair of dense layers along with three dropout layers, and the ultimate output is produced by a dense layer that utilizes a sigmoid activation function to classify the image output as either normal or pneumonia. The total number of model parameters is 2,621,089, of which 2,620,865 are trainable parameters and only 224 are non-trainable parameters.

4.3.2. Strategy 2

This strategy employed transfer learning, in which the pretrained Resnet-152V2 model was used as a feature extractor. The ResNet-152V2 architecture consists of residual blocks, which enable the network to be deeper without causing the vanishing gradient problem. Each residual block contains several convolutional layers and shortcut connections that bypass one or more layers, allowing the gradient to be easily propagated through the network. In ResNet-152V2, the residual blocks have a bottleneck design, which reduces the computational cost and memory usage of the network. Here we remove the top layer and include our one pooling layer, one dropout layer, and two dense layers for generating the output.

4.3.3. Strategy 3

In our strategy 3, we employed a technique called fine-tuning. Initially, we froze all the layers of the pre-trained Resnet-152V2 model and used the weights that were already calculated during its training on the Imagenet dataset. However, we then unfroze some of the last layers of the pre-trained model and continued training, tuning the weights of these layers based on our pneumonia detection dataset. The final output is generated from a dense

layer with a sigmoid activation function that categorises the output image as either normal or pneumonia.

5. Experimentation Results and discussion

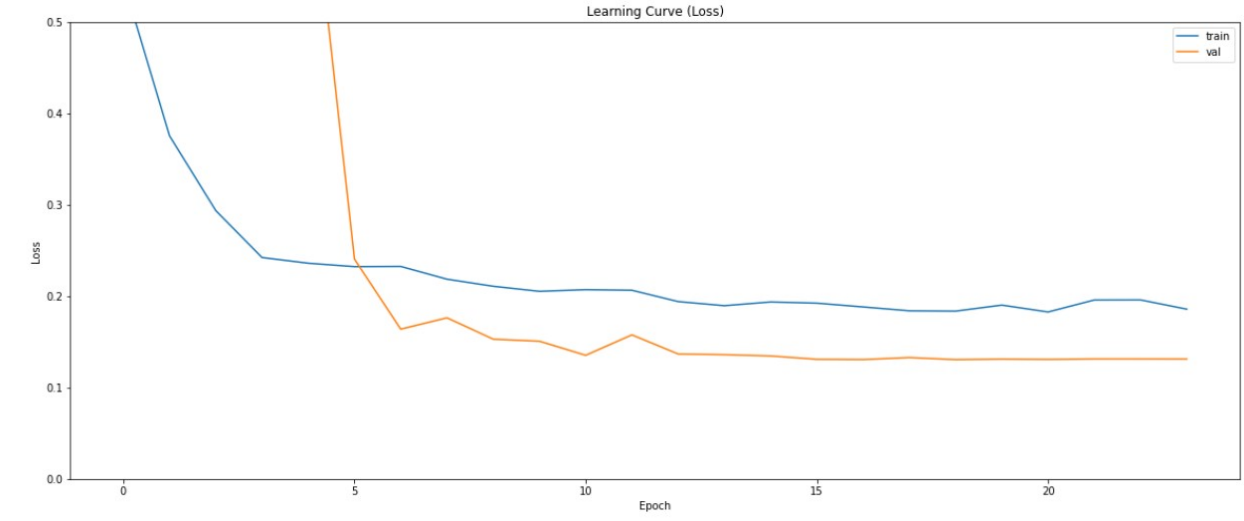
The experimentation was carried out on the chest X-ray imaging dataset. Three models were trained and tested on a Chest X-Ray Images (Pneumonia) dataset consisting of 5232 images for training and 624 images for testing the models. Then the training dataset was split into training and validation sets in a 80:20 ratio. A CNN strategy 1 model consisting of convolution layers, batch normalisation layers, max pool layers, dropout layers, and dense layers was trained from scratch on the training set. Binary Cross-Entropy Loss and the Adam Optimizer were employed for training. The model was trained for 50 epochs with a batch size of 32. After training, the strategy 1 model was evaluated on a separate test set and achieved an test accuracy of 84%, with a precision of 0.81, recall of 0.98, a specificity of 0.62, and F1-score of 0.89.

To improve the performance of pneumonia detection, transfer learning was employed, and the ResNet-152 pre-trained model was used as our base model in strategy 2, which was trained on the large-scale ImageNet dataset. We removed the top layers of the model and added our own custom dense layers on top for pneumonia detection. The weights of the ResNet-152 model were frozen and trained only on the custom dense layers for 50 epochs with a batch size of 32. After training, we evaluated the ResNet-152 model on the test set and achieved an accuracy of 88%, with a precision of 0.86, recall of 0.99, a specificity of 0.72, and F1-score of 0.92.

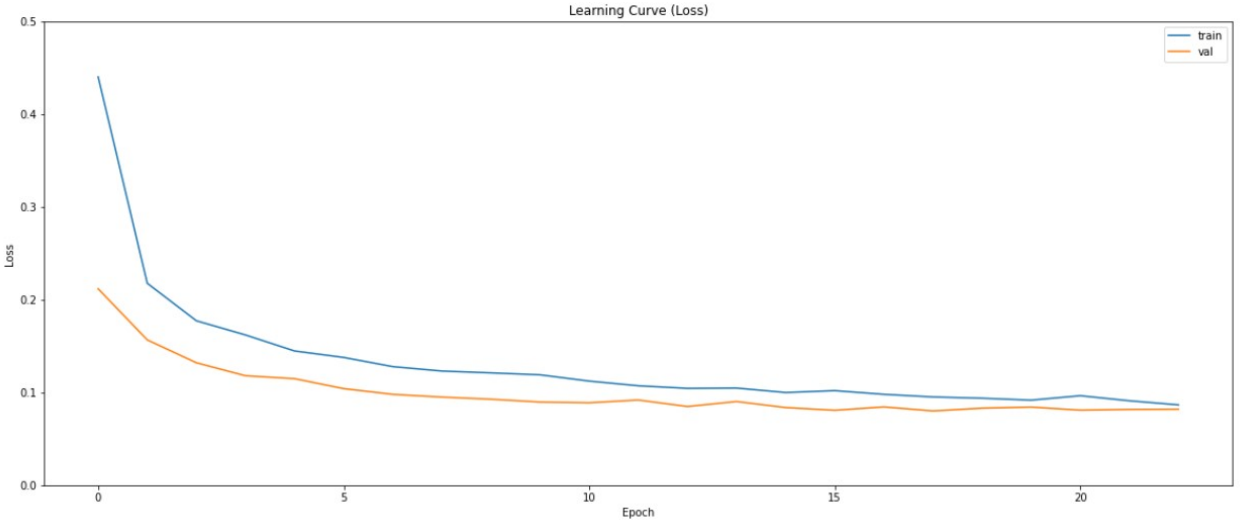
After that, we unfroze the weights of the last convolutional block of the ResNet-152 model in strategy 3 and fine-tuned the model for another 50 epochs with a lower learning rate of 5e-5. After training, the model was evaluated on a separate test set and achieved a test accuracy of 91%, with a precision of 0.89, a recall of 0.98, a specificity of 0.80, and F1-score of 0.93. The loss and accuracy graph of each strategy model is depicted in Figure 3 and 4 respectively .

To further evaluate the model's performance, we visualized the confusion matrix of each of our proposed strategy models.Figure 5 shows the confusion matrices of the models and their accuracy.

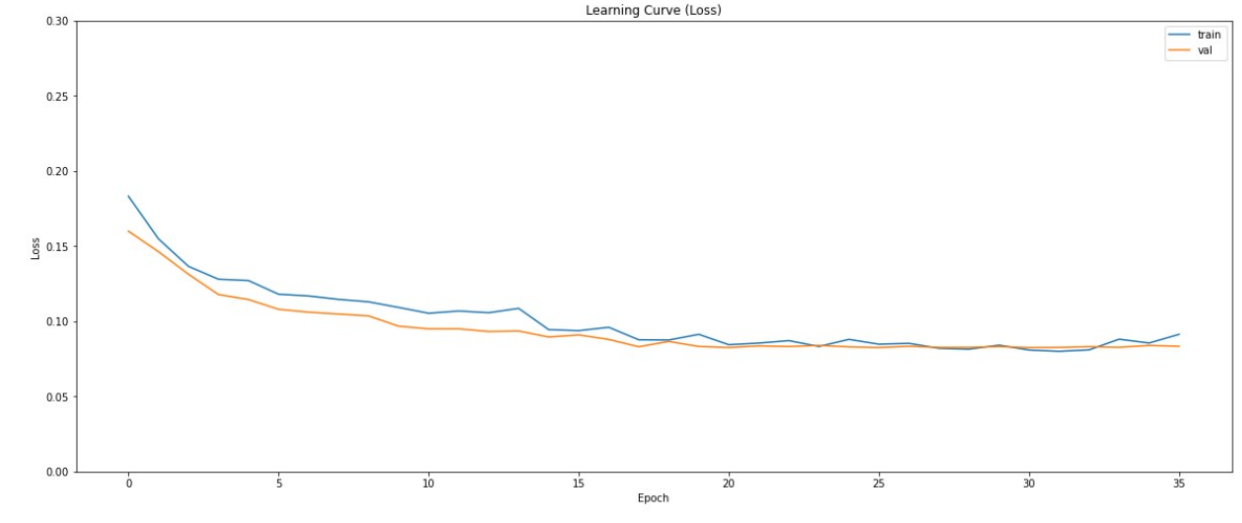
Figure 6 shows the Roc-Auc graph of strategy model 3, which outperforms the other two proposed models.



(a) **Strategy 1 model loss**

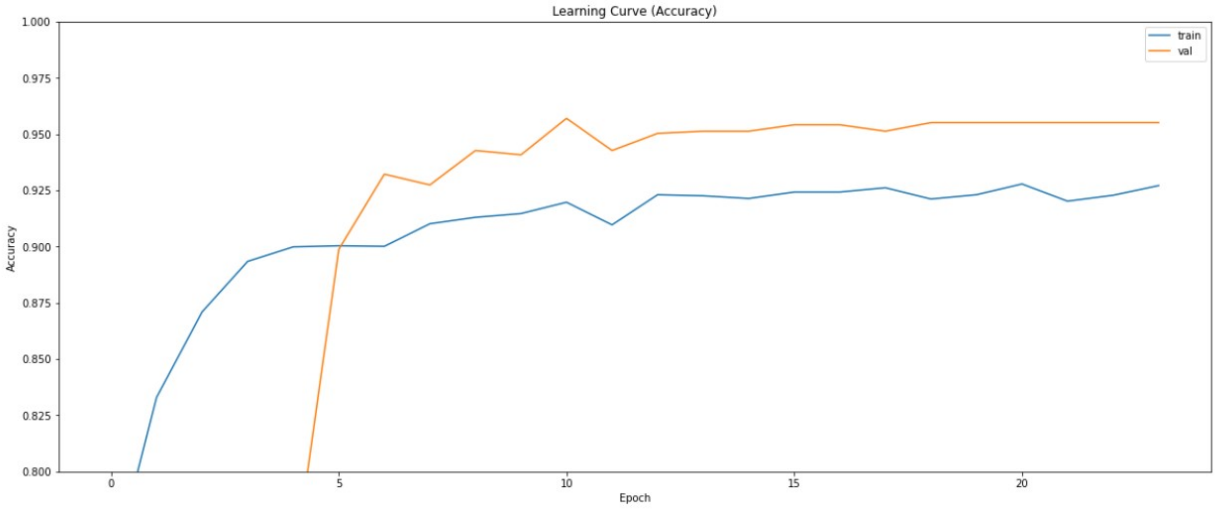


(b) **Strategy 2 model loss**

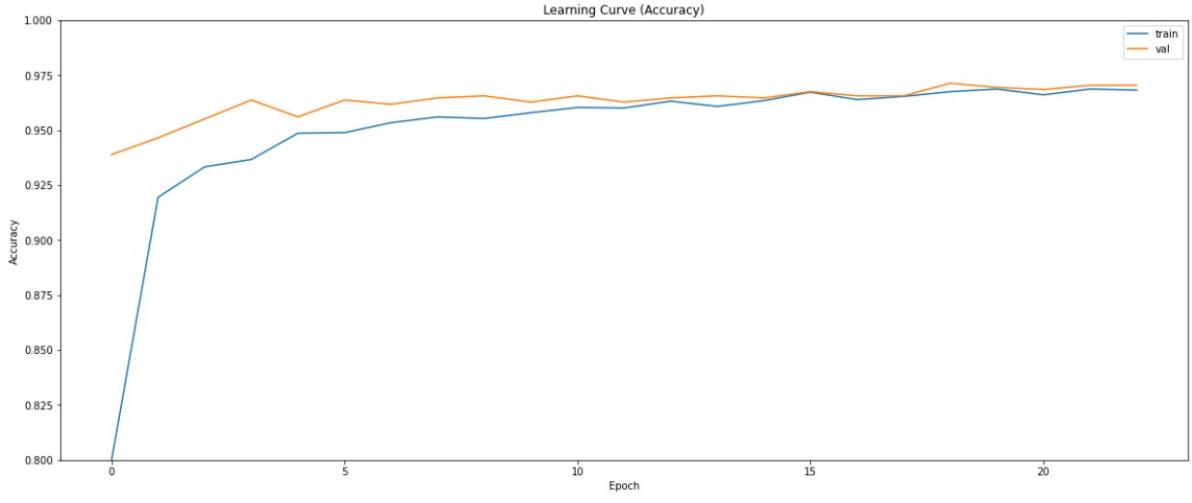


(c) **Strategy 3 model loss**

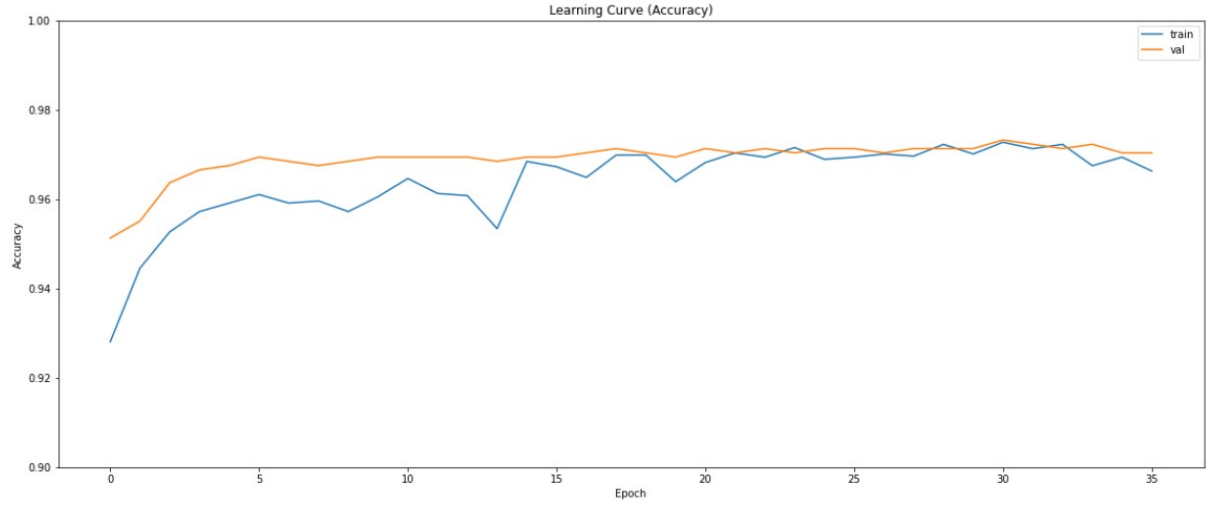
Figure 3: **Loss graphs of the proposed strategies**



(a) **Strategy 1 model accuracy**

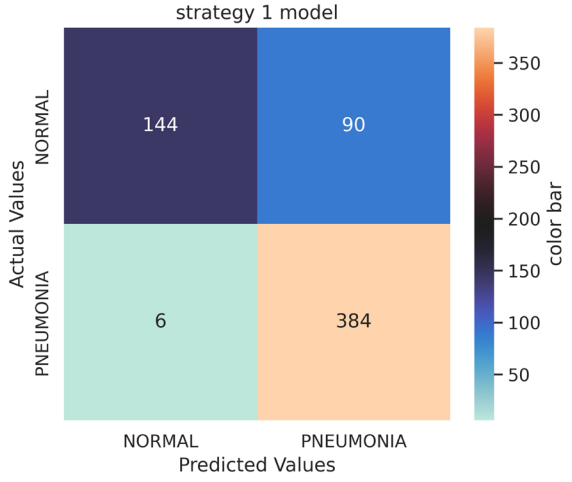


(b) **Strategy 2 model accuracy**

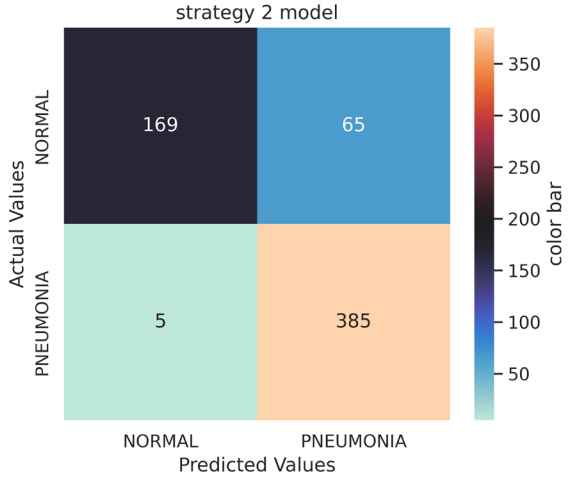


(c) **Strategy 3 model accuracy**

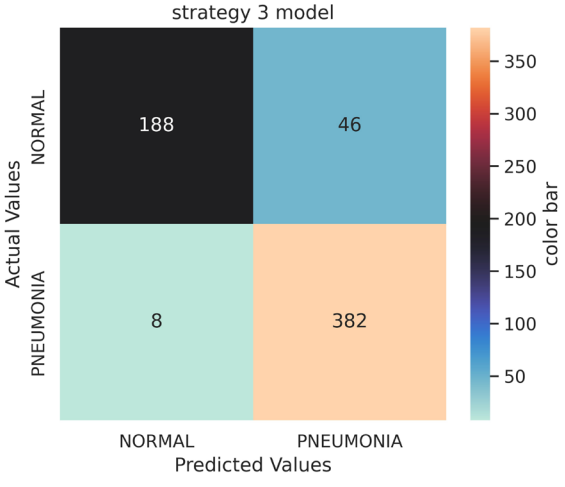
Figure 4: **Accuracy graphs of the proposed strategies**



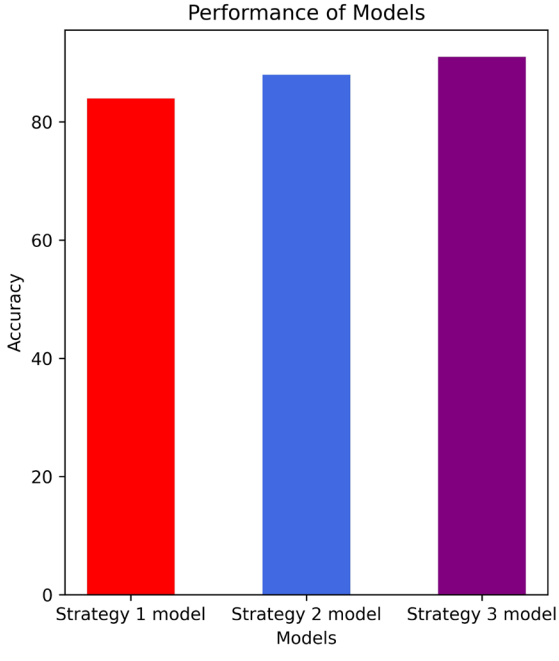
(a) **Strategy 1 model**



(b) **Strategy 2 model**



(c) **Strategy 3 model**



(d) **Performace of models**

Figure 5: **Confusion matrices and Accuracy**

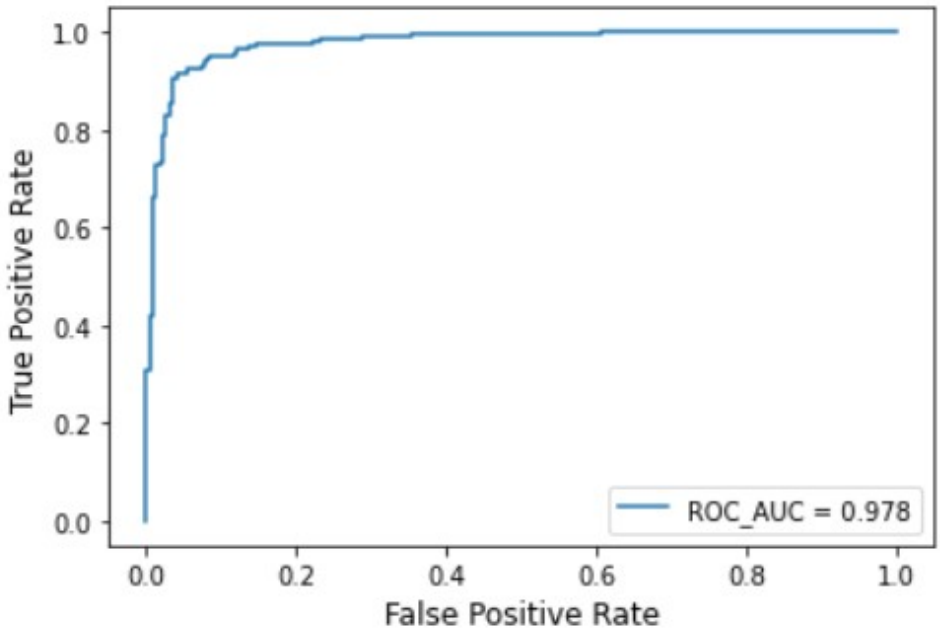


Figure 6: **Roc-Auc graph of Strategy model 3**

6. Conclusion and Future work

This paper introduces a deep learning framework that uses three different CNN models for classifying pneumonia. Two of these models are pre-trained, while the third model is developed from scratch. The results showed that our proposed deep learning framework achieved high values in all metrics, with the ResNet-152V2 model performing the best. Additionally, the other two proposed models achieved accuracy, recall, F1-score, precision, and AUC scores of over 80%.

In future work, we plan to explore the use of other CNNs and RNNs, including bidirectional LSTM architectures and pre-trained models, for detecting pneumonia in chest X-ray images.

References

[1] 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, 2017.

[2] B. Antin, J. Kravitz, and E. Martayan, "Detecting pneumonia in chest x-rays with supervised learning," *Semanticscholar. org*, vol. 2017, 2017.

[3] P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. Langlotz, K. Shpanskaya *et al.*, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *arXiv preprint arXiv:1711.05225*, 2017.

[4] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, "Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2097–2106.

[5] S. Christodoulidis, M. Anthimopoulos, L. Ebner, A. Christe, and S. Mougiakakou, "Multisource transfer learning with convolutional neural networks for lung pattern analysis," *IEEE journal of biomedical and health informatics*, vol. 21, no. 1, pp. 76–84, 2016.

[6] J. R. Zech, M. A. Badgeley, M. Liu, A. B. Costa, J. J. Titano, and E. K. Oermann, "Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: a cross-sectional study," *PLoS medicine*, vol. 15, no. 11, p. e1002683, 2018.

[7] X. Gu, L. Pan, H. Liang, and R. Yang, "Classification of bacterial and viral childhood pneumonia using deep learning in chest radiography," in *Proceedings of the 3rd international conference on multimedia and image processing*, 2018, pp. 88–93.

[8] P. Lakhani and B. Sundaram, "Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks," *Radiology*, vol. 284, no. 2, pp. 574–582, 2017.

[9] A. Donthi, A. Huang, and A. Tammanagari, "Detecting pneumonia with convolutional neural networks," *Semanticscholar Org.: Allen Institute for Artificial intelligence, Seattle, WA, USA*, 2018.

[10] J. Rubin, D. Sanghavi, C. Zhao, K. Lee, A. Qadir, and M. Xu-Wilson, "Large scale automated reading of frontal and lateral chest x-rays using dual convolutional neural networks," *arXiv preprint arXiv:1804.07839*, 2018.

[11] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *2017 international conference on engineering and technology (ICET)*. Ieee, 2017, pp. 1–6.

[12] C. Bailer, T. Habtegebrial, D. Stricker *et al.*, "Fast feature extraction with cnns with pooling layers," *arXiv preprint arXiv:1805.03096*, 2018.

[13] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, "Convolutional neural networks for medical image analysis: Full training or fine tuning?" *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.

[14] S. J. Pan and Q. Yang, "A survey on transfer learning ieee transactions on knowledge and data engineering," *22 (10)*, vol. 1345, 2009.

[15] P. MOONEY, "Chest x-ray images (pneumonia).[online] tersedia pada: https://www. kaggle. com/paultimothymooney/chest-xray-pneumonia," *Diakses pada tanggal*, vol. 17, 2018.