Enhancing Diagnostic Accuracy: AI-Enabled Chest X-ray Analysis for Improved Pneumonia Detection

MSc Research Project
Artificial Intelligence

Manav Koli
Student ID: x23144483

School of Computing
National College of Ireland

Supervisor: Dr. Kislay Raj

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Manav Koli
Student ID:	x23144483
Programme:	MScAI
Year:	2024
Module:	MSc Research Project
Supervisor:	Dr. Kislay Raj
Submission Due Date:	12/08/2024
Project Title:	Enhancing Diagnostic Accuracy: AI-Enabled Chest X-ray Analysis for Improved Pneumonia Detection
Word Count:	5860
Page Count:	19

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: 8th September 2024	**Date:**

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Enhancing Diagnostic Accuracy: AI-Enabled Chest X-ray Analysis for Improved Pneumonia Detection

Manav Koli
x23144483

Abstract

Pneumonia remains a leading cause of mortality, particularly among young children and the elderly, underscoring the urgent need for accurate and timely diagnosis. This study offers a novel method for using artificial intelligence (AI) to improve the precision of pneumonia diagnosis using sophisticated chest X-ray processing. We used Convolutional Neural Networks (CNNs) to create a reliable AI system, making use of cutting-edge models such as ResNet152V2 and VGG16. In order to guarantee a balanced dataset, our approach comprised rigorous data preprocessing such as normalisation, scaling, and data augmentation. Although, using Hyperband for hyperparameter adjustment improves our model training outcomes. Grad-CAM (Gradient-weighted Class Activation Mapping) is a novel approach in our work that not only improves model transparency but also offers important visual insights into the particular regions of chest X-rays that show pneumonia. This feature greatly improves diagnosis reliability by enabling medical practitioners to precisely pinpoint impacted regions. Our models demonstrated exceptional accuracy, achieving 94.23% with VGG16 and 92.26% with ResNet152V2, highlighting the effectiveness of our approach. This research not only advances the field of AI in medical diagnostics but also offers a transparent, interpretable, and highly accurate tool for pneumonia detection, with the potential to transform patient outcomes, especially in resource-limited settings.

Keywords: *Model Interpretability, Grad-CAM Visualization, Deep Learning in Healthcare, Transfer Learning, Diagnostic Accuracy, Regularization Techniques, Hyperparameter Tuning*

1 Introduction

Pneumonia is a serious respiratory infection which still contributes to be the main cause of death, especially in young children and the elderly Liang and Zheng (2020). Since it can be manifested in many ways and its characteristics are similar to other respiratory diseases,detecting it accurately and quickly requires state of the art diagnostic equipments.Standard practice mostly depends on radiologists knowledge to interpret chest X-rays, but access to this knowledge is often restricted in many regions of the world. The development of AI-driven diagnostic tools can be used to assist medical professionals to accurately diagonse pneumonia from chest X-ray pictures to tackle this problem.

Numerous studies have investigated the application of convolutional neural networks (CNNs) for pneumonia diagnosis in recent years Sharma and Guleria (2023), showing promising outcomes. Our research, however, enhances these existing methods by prioritizing model interpretability, which is essential in clinical settings, in addition to achieving high accuracy. Unlike earlier studies that may have concentrated mainly on performance metrics, our approach incorporates sophisticated interpretability tools to ensure that the AI makes clear and dependable decisions.

For this study, we specifically chose the ResNet152V2 and VGG16 models due to their proven balance of interpretability, performance, and depth Reshan et al. (2023). The main purpose behind choosing the VGG16 model is its consistent delivery of results while identifying the image dataset. On the other hand, ResNet152V2 helps us to address the problem of gradient by creating deep residual connections in the dataset. Although there are many good algorithms but we choosed VGG16 and ResNet152V2 due to its suitability and consistent outputs especially when dealing with medical datasets. One of the types of research implemented by Vetrithangam et al. (2023) suggests the use of Gradient-Weighted Class Activation Mapping or Grad-CAM to accelerate model performance.

Predicting correct outcomes is critical while working on medical related datasets Panwar et al. (2020). If any of the outcomes is misinterpreted, it can lead to incorrect treatment which can have harmful impacts on patients. Hence the use of Grad-CAM helps us to identify the areas which are highly affected based on the chest X-rays and aid us to provide effective model prediction. This helps the healthcare experts to diagnose the patients on time and take preventive actions quickly based on the symptoms.

In Conclusion, the research helps us to create a AI based system capable of delivering accurate results by using CNN enabled architecture for predicting. By integrating Grad-CAM with ResNet152V2 and VGG16, we aim to develop an AI-enabled diagnostic tool that is transparent, dependable, and accurate. This will boost public confidence in AI-driven medical solutions and ultimately lead to better patient outcomes.

1.1 Research Question & Objectives

"How can we use deep learning approaches, specifically Convolutional Neural Networks (CNNs), to construct a reliable and interpretable AI model for accurate detection of pneumonia in chest X-ray pictures?"

1.2 Objectives

- Worked on the existing research and background approches on Pneumonia detection using AI deep learning approches and evaluated their results on same side.
- Collection of chest x-ray Images, preprocessing and transformation of these images as part of data preparation of model.
- Develop and implement custom CNN, ResNet152V2, and VGG16 models to identify Pneumonia from chest X-ray images.
- Analyze the performance of the implemented VGG16, ResNet152V2, and custom CNN models.
- Utilize Grad-CAM to compare and contrast the behavior of different models, providing detailed analysis and insights.

- Apply various data augmentation techniques, followed by hyperparameter tuning and fine-tuning, to improve model robustness and accuracy.
- Use the Lazy Predict library to acquire performance metrics for various machine learning algorithms.
- Observe and address overfitting in the models using L2 regularization techniques, such as dropout, weight decay, and data augmentation, to further reduce overfitting in the model.
- Successfully apply Grad-CAM to ResNet152V2 and VGG16 models, using the visualizations to understand how the models interpret and focus on different parts of the input images.
- Apply hyperparameter tuning for machine learning models, keep meticulous records of all observations and trial outcomes.

2 Related Work

Hasan et al. (2021) looked at using a deep learning model to classify chest X-ray images into COVID-19, pneumonia, and normal categories. They combined three models: DenseNet121, EfficientNetB0, and VGG19. This used the strengths of each model. They also used a DenseNet-based U-Net model to find and highlight affected areas in the images. This method reached 99.2% accuracy in classifying the images and 92% accuracy in finding affected regions. They made ground truth masks using Grad-CAM and Amazon SageMaker Ground Truth Tool. This approach makes diagnosis more accurate and helps doctors by showing which parts of the X-rays are important. The study shows how deep learning can improve diagnosis and support doctors, especially in places with limited medical resources. In another study, Abdel-Basset et al. (2022) developed an interpretable deep learning framework to distinguish COVID-19 from other pneumonia cases using lung ultrasound images. The proposed model incorporated novel transformer modules with window-based and shifted-window multi-head self-attention layers to effectively capture pathological information from ultrasound frames. Additionally, a convolutional patching module was introduced to transform input frames into latent space, and a weighted pooling module was used to score the embeddings of disease representations. This comprehensive approach achieved an accuracy of 93.4% and an AUC of 97.5%, demonstrating the model's high performance and robustness. The study emphasized the importance of explainability by integrating Grad-CAM++, which provided visual explanations for the model's decisions. This feature is particularly valuable in clinical settings as it helps healthcare professionals understand the underlying reasons for the model's predictions, thereby enhancing the model's reliability and acceptance in medical practice.

Jain et al. (2020)A study compared several CNN models, including VGG16, VGG19, ResNet50, and Inception-v3, for detecting pneumonia in chest X-ray images. They adjusted many settings to make the models work better. VGG19 had the best accuracy at 88.46%, and VGG16 followed with 87.28%. The study showed that transfer learning helps. This means using models trained on large datasets like ImageNet and adjusting them for specific tasks like pneumonia detection. This method makes the models more accurate and reduces training time. The study found that carefully choosing and adjusting these models can improve diagnosis a lot. This makes them very useful for quick and accurate clinical diagnoses.Hashmi et al. (2020) In another study, researchers developed a deep learning system to tell COVID-19 apart from other types of pneumonia using lung ultrasound images. They used special tools to highlight important details in the images. They also included features to convert the images into a useful format and to assess the disease characteristics. The system achieved 93.4% accuracy and an AUC of 97.5%, showing it performs very well. They used Grad-CAM++ to explain how the model made its choices. This helps doctors understand why the model made certain decisions, making it more valuable in medical settings.

In order to diagnose pneumonia, Reshan et al. (2023) worked with a deep learning model based on MobileNet, emphasising its effectiveness and suitability in environments with limited resources. Known for its lightweight architecture, MobileNet demonstrated accuracy of 94.23% and 93.75% on two datasets. This study highlights MobileNet's potential for application in portable devices or places with limited access to sophisticated computer infrastructure by demonstrating its high precision delivery with low computational resource requirements. The model has the potential to enhance diagnostic procedures in a variety of healthcare settings, as seen by its outstanding performance in pneumonia diagnosis.Sharma and Guleria (2023) Another study looked at how well the VGG16 model worked when combined with traditional machine learning methods like SVM, KNN, and Random Forest. They found that VGG16 performed the best, achieving an accuracy of 95.4%. This study highlights the advantages of deep learning models over traditional methods, especially in medical image processing. The researchers emphasized that using advanced models like VGG16 is important for achieving high diagnostic accuracy and reliability, which are necessary for making effective clinical decisions.

Li et al. (2019) Finally, a CNN model with an attention mechanism was developed to identify pneumonia in chest X-rays. The attention mechanism is a feature that helps the model focus on important areas in the image. This model used Grad-CAM to highlight the areas in the X-rays that influenced its decisions. This method improves the model's accuracy and helps doctors understand how the model is making its predictions.

Data augmentation is important for making a model more efficient because it elongates the training dataset. Techniques like rotating, scaling, and translating images mentioned in Rajpurkar et al. (2017) have been used in studies to improve pneumonia detection. All these methods help create a model that works better in real-world situations by making it more adaptable to new and different data. Such experiments showcase that data augmentation is important for making deep learning models generalize well, which is an essential step for using them effectively in medical settings.

2.1 Conclusion

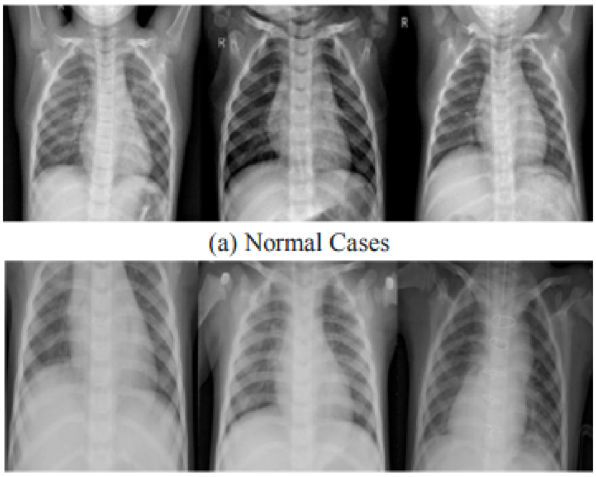
Deep learning has been used to identify pneumonia from chest X-ray images. Models like VGG16, ResNet152V2, and ensemble methods are very accurate. These models help doctors diagnose patients. Adding attention mechanisms, data augmentation, and fine-tuning makes the models more reliable. These improvements are important in places with limited resources. They can help improve healthcare. Dropout layers, weight decay, and L2 regularization prevent overfitting. Future research should focus on making the models stronger and easier to understand. The goal is to use them in real-world medical settings.

3 Methodology

The main objective of this project is reliable system that can detect pneumonia using chest X-ray images which uses advanced deep learning methods such as Convolutional Neural Networks (CNNs). Because of complexity in the nature of medical imaging data, we have to be very careful while adopting a systematic process that ensures the accuracy and interpretability of the model. The approach we discussed explains every steps taken to develop the pneumonia detection model in detail. The process also emphasises on structured approach which is necessary for the success of this model. Each stage of the model is designed to handle specific challenges associated with it while analysing the medical images. This makes the model reliable and highly efficient. The whole process is divided into several key stages such as data collection, data pre-processing, data transformation, data modelling, and model evaluation and finally the use of Grad-CAM for Model Interpretability. Before we discuss them in detail let me give us brief about each of the stages. So in data collection stage, we tried to get the best dataset from the open-source (Kaggle) platform and ensured that it consists of the clear chest X-ray scans from different age group and different gender. In data pre-processing, we are going to do the resizing, rotating, converting from photos to greyscale, and normalizing the pixel values. Overall, we are going to make the data suitable for our use. In next stage; data transformation, we are going to format the pre-processed images so that they can be used for CNN training. Then the next stage, Data Modelling one of the most important stage of whole process. Customized CNN model is created using convolutional and pooling layers. We will be using Grad-CAM for showing the regions of X-ray pictures which are important for the prediction.

3.1 Data Collection

The dataset used for this study consists of chest x-ray images which are collected from the kaggle platform, which is well know and publically accessible site. The dataset consists of two different groups of images one is normal and another one is pneumonia images. Furthermore, the dataset was divided into two main groups namely training and another is testing, in order to work with models performance as per requirements. This dataset consists of total 5856 images of chest x-rays of patients between the ages of one and five years old and dataset dividation was shown on the pictures. Firstly, dataset was divided, with 5,232 images designated for training and 624 images for testing. and image formate of all images is JPEGs.



(a) Normal Cases

(b) Pneumonia Cases

Figure 1: Examples from the dataset

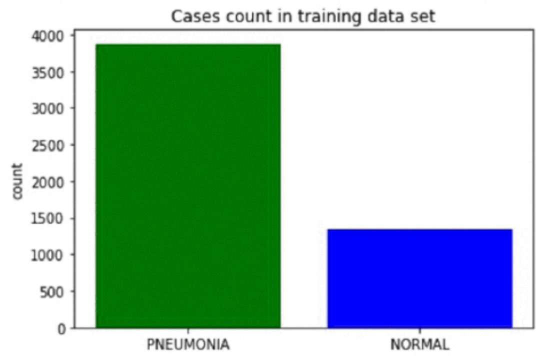


Figure 2: Cases count in training dataset

3.2 Data Preprocessing

To ensure the quality and consistency of the incoming data, data preparation is an essential step. This procedure involves converting photos to greyscale, scaling them to a consistent size, and normalising the pixel values. Data augmentation techniques like rotation, flipping, and zooming are used to increase the diversity of the training data and fortify the model. To make them easier to access during model training, the photos are scaled to 224 by 224 pixels and arranged into directories.

Importing necessary libraries like as TensorFlow, Keras, and other data manipulation and visualisation tools is part of the initial setup. To guarantee the repeatability of the experiment, a random seed is set. The photos are fed into lists, transformed into pandas DataFrames for additional analysis, and the training and testing pathways are set.

A number of data augmentation techniques used in this investigation are listed in Table 1. Prior to further processing, rescaling is accomplished by multiplying the data by a predetermined amount. The photos were originally stored with RGB values ranging from 0 to 255, but they have been scaled down by a factor of 1/255 because the models would struggle to handle such high values at usual learning rates. Random shearing transformations are introduced by applying shear range. Random zooming inside photos is possible using the zoom range, and half of the images can be randomly flipped horizontally with the horizontal flip function—a common occurrence in real-world situations. These methods for preparing data were essential to this investigation.

3.3 Data Transformation

The preprocessed images are then formatted for CNN training. This involves creating data generators for batch processing and applying additional transformations like scaling to enhance model performance. Training, validation, and test data generators are made with Keras ImageDataGenerator. With implementing this techniques we have reduced the chances of overfitting in model.

Operation	Values	Why It's Important
Rescale of data	1/255	Normalizes pixel values to a smaller range, improving model performance.
Zoom Range of data	0.1	Helps the model generalize better by simulating different scales of input images.
Shear Range of data	0.1	Introduces geometric distortions to make the model more robust to such variations.
Horizontal_Flip of data	True	Augments the dataset by flipping images, increasing diversity and reducing overfitting.

Table 1: Operation, Values, and Importance

3.4 Data Modeling

We use preprocessed data to build and train Convolutional Neural Network (CNN) models during the modelling phase. Initially, we construct a unique CNN from the ground up, incorporating multiple convolutional and pooling layers, succeeded by dense layers. The model's hyperparameters are then adjusted and its performance is enhanced using Hyperband methods. Furthermore, we use transfer learning to improve pre-trained models such as ResNet152V2 and VGG16, enabling them to leverage their pre-existing knowledge to achieve higher classification accuracy. Three convolutional blocks make up the customised CNN model. After every block, we apply dropout layers, activation functions, and batch normalisation to avoid overfitting. Using early stopping and learning rate adjustment, we train the model for 20 epochs to make sure it converges correctly without overfitting. We employ the Adam optimiser and the binary cross-entropy loss function for model compilation.

We substitute new dense layers created especially for pneumonia detection for the top levels of the ResNet152V2 model in the transfer learning method. This modification greatly improves generalisation as well as accuracy. In a similar manner, we alter the VGG16 model by incorporating unique layers and using regularisation strategies like L2 and dropout regularisation. The model's performance is further improved by using hyperband to optimise its parameters.

3.5 Grad-CAM for Model Interpretability

Grad-CAM (Gradient-weighted Class Activation Mapping) is used to highlight which region of the x-ray image exactly considers important for detecting pneumonia. This method help in making decision and providing clear view through using different colours as shown in tha below image, It also improving its nterpretability and reliability. The following steps are used to explain it more effectively.

1. **Importing Libraries:** Importing Libraries such as TensorFlow, Keras, and Matplotlib are imported.

2. **Use of Pre-trained Models:** Models like ResNet152V2 and VGG16 with pre-trained ImageNet weights are loaded and used for further enhancement

3. **Preparation of Image:** For prediction, a sample image is loaded and preprocessed.

4. **Use of Heatmap and Result** Grad-CAM heatmaps are produced using the model's predictions to highlight the important regions of the image.

5. **Heatmaps:** Matplotlib is used to overlay and display the heatmaps on top of the original picture.

Figure 3: Grad-CAM Heatmap for Pneumonia Detection

When the Grad-CAM (Gradient-weighted Class Activation Mapping) heatmap is combined with the chest X-ray image, the spots that the model identifies in its forecasted. as shown in Figure 3's heatmap is created using a VGG16 model that has been tuned for the identification of pneumonia. The colours in the heatmap indicate the areas of the picture that have the greatest impact on the model's decision-making process.

- **Red Portion:** These are the locations that have the highest activation, meaning that the model has determined that these are the most important for identifying pneumonia. The red patches in this picture are mostly found in the middle and upper right regions of the lungs, which may indicate inflammation or fluid build up.

- **Green and Yellow Portion:** Though not as much as the red zones, these areas nonetheless contribute to the model's prediction and have a modest activation level. These areas may be associated with less obvious pneumonia symptoms, but they are nonetheless important for the diagnosis.

- **Purple and Blue Portion:** These areas are seen as normal or less suggestive of pneumonia and have less weight in the model's judgement.

4 Design Specification

4.1 Project Flow Design

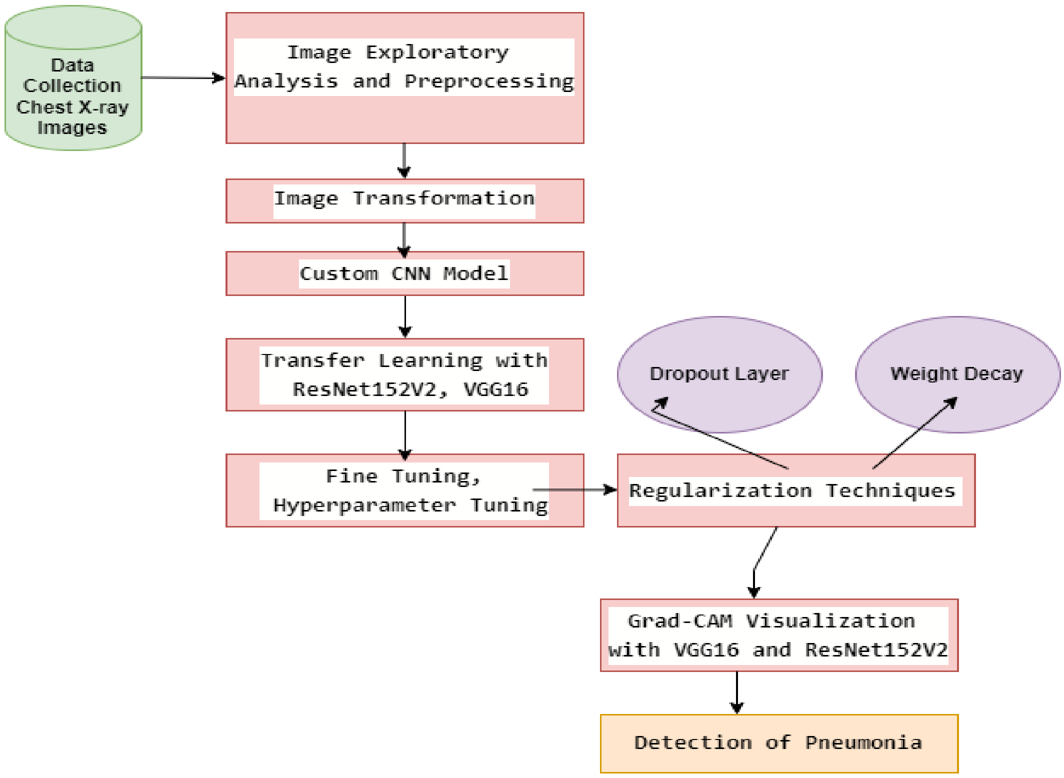


Figure 4: Project Flow Diagram

These chest X-ray images gives us a comprehensive workflow for developing a deep learning based pneumonia detection system according to the flow diagram. Firstly, data has been collected and loaded of chest X – ray images, after that image exploratory analysis and preprocessing, which includes steps such as normalization, resizing, and data augmentation. We have trained a custom Convolutional Neural Network (CNN) model, which undergoes a transformation of pre – processed images. The diagram illustrates how transfer learning may be applied to improve model performance by employing pre-trained models such as ResNet152V2 and VGG16. Regularization techniques, including dropout layers and weight decay, are employed to prevent overfitting. The model is fine – tuned and hyperparameters are optimized for improved accuracy Hasan et al. (2024). To interpret the model's predictions we have used Grad – CAM, which ensures ensuring transparency and trustworthiness in medical diagnostics. Finally, the workflow concludes with the detection of pneumonia, enabling the model to classify chest X-ray images accurately.

4.2 Proposed Network

In this project we worked with CNN model like VGG16 to evaluate chest x-ray image picture and detect weather that x-ray is for pneumonia or normal chest x-ray picture and detect cases of pneumonia. To follow this approch we worked layer by layer, and we increased the filter size from 32 at the begining of our CNN design. A MaxPooling layer was added after a Conv2D layer in the model generation process. Because it was effective, we used a 3x3 kernel size. ReLU is the most often utilised activation function, while Tanh and other activation functions can also be used. Lastly, we added ANN layers after flattening the input Szepesi and Szilágyi (2022).

$$S(z) = \frac{1}{1+e^{-z}} \tag{1}$$

$$f(z) = \max(0, z) \tag{2}$$
$$S(z) = \text{Sigmoid} \tag{3}$$

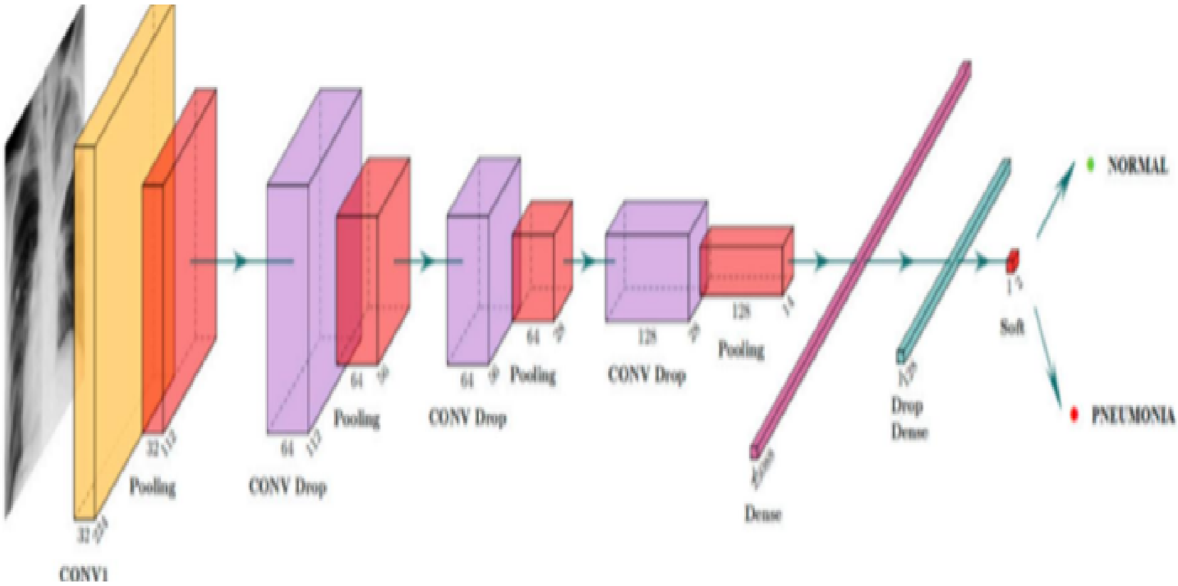


Figure 5: Details of proposed DL model.

Class=I, defined units as the total number of classes and used a softmax activation function for the final layers (ANN Layers). 'I' set the unit to 1 and used a sigmoid for binary classification.

4.2.1 Convolutional Layers

The first block of this neural network is distinctive because it acts as a feature extractor. It accomplishes this by applying convolution filtering techniques for template matching. In the initial layer, the image is processed with various convolutional kernels, resulting in the creation of "feature maps." These maps are then resized and/or normalized using an activation function Varshni et al. (2019).

4.2.2 Pool Layers

The second block, which is present in all CNNs and not just classification neural networks, applies different linear combinations and activation functions to the input vector to process it and create an output vector. This final vector's elements each represent a class and the likelihood that the image falls into that particular class. Each probability ranges from 0 to 1, and the total of these probabilities is 1. To calculate these probabilities, the last layer of this block and the network as a whole use an activation function, such as ReLU for multi-class classification or Sigmoid for binary classification Naralasetti et al. (2021).

5 Implementation

5.1 Implementation of Designed CNN Architecture

- **Input Shape:** The CNN model is set up to take 224 by 224 pixel input images, with three channels denoting the RGB (red, green, and blue) colour space. This precise dimensions guarantees that the photos function with common pre-processing methods and neural network models, preserving consistency all along the way.

- **Convolutional Layer:** One of the core feature extractors in any CNN models are these Convolutional Layers. To identify the various features, these layers apply a series of filters across the input image. Below is the way how they work:

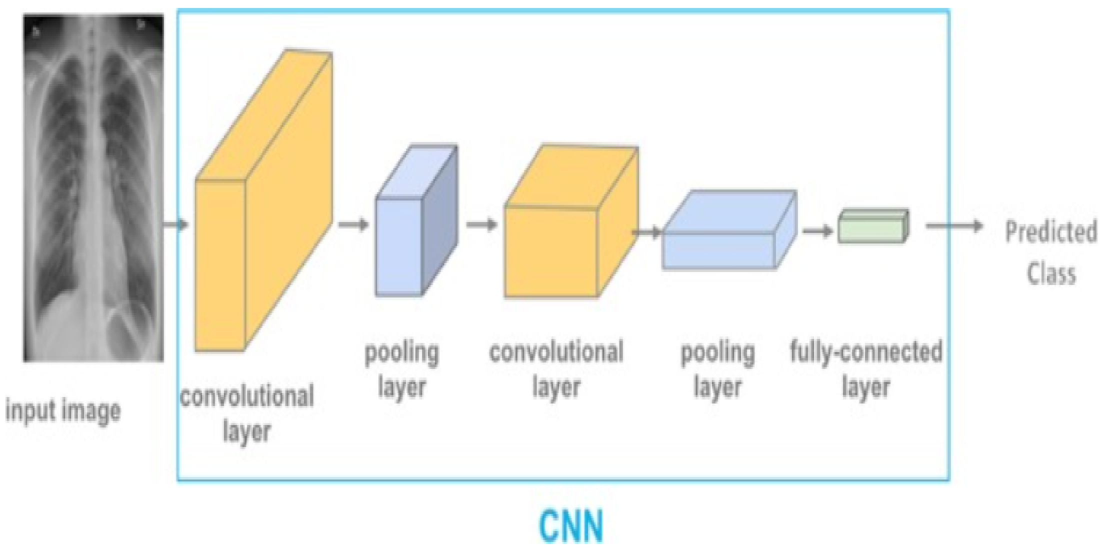


Figure 6: CNN Architecture.

- **Pooling Layer:** In particular, the MaxPooling layers are used to minimise the feature maps' spatial dimensions. This decrease is important because it helps to reduce computing complexity and mitigate overfitting by making feature identification more invariant to changes in scale and orientation. By choosing the maximum value from each region, the MaxPooling operation with a 2x2 window reduces the width and height of the feature maps by half.

- **Rectified Linear Unit Activation Function (ReLU):** The network gains nonlinearity via the Rectified Linear Unit (ReLU) activation function, which allows it to recognise and simulate intricate patterns. ReLU functions by preserving positive values in the feature maps and setting all negative values to zero. Deep learning makes extensive use of this function since it prevents the vanishing gradient problem and is computationally straightforward.

- **Flatten Layer:** The 3D feature maps are to be converted into 1D vectors using the Flatten layer. Because fully connected (dense) layers require input in a flattened vector format, this conversion is necessary. By flattening, all traits found in the earlier layers are included in the next classification procedure.

- **Fully Connected Layer:** These layers, which are sometimes referred to as denser layers, are in charge of classifying data in the end using the features that the convolutional and pooling layers have extracted. This model has a dropout rate of 0.5 and a dense layer consisting of 64 units. During training, the dropout strategy randomly deactivates a section of the neurones, preventing the model from becoming overly dependent on any one set of neurones. This helps to reduce overfitting.

- **Adam Optimizer:** It offers adaptable learning rates for every parameter by combining the advantages of Root Mean Square Propagation (RMSProp) and adaptable Gradient Algorithm (AdaGrad). This method boosts overall performance and quickens the model's convergence. Adam is especially well-suited to handle issues involving sparse gradients and noisy data since he modifies the learning rate according to the first and second moments of the gradients.

5.2 Implementation of VGG16 Architecture

A well-known convolutional neural network architecture that is frequently used for image classification applications is VGG16. VGG16, developed by the University of Oxford's Visual Geometry Group, is renowned for its uncomplicated architecture, which uses tiny 3x3 convolution filters stacked one after the other to deepen the network Simonyan and Zisserman (2014). Thirteen convolutional layers and three fully linked layers make up the architecture's total of sixteen layers.

In this experiment, we detected pneumonia in chest X-ray pictures using the pre-trained VGG16 model which gives a better results. Utilising transfer learning, we were able to apply the useful features that VGG16 had picked up from ImageNet, which enhanced our model's performance and the training period. Unlike previous versions, the VGG16 is built differently, employing a stack of several 3x3 filters in instead of bigger filters. The network can record more intricate patterns and features thanks to this design.

Fully connected layers, 2x2 average pooling layers, and 3x3 convolutional layers are all part of the VGG16 architecture. The network's width doubles after each pooling layer, starting at 64 channels in width initially. There are 256 channels in each of the first two fully connected layers and two channels in the third layer. The output layer uses the softmax activation function, whereas the first two hidden layers apply the ReLU activation function. To avoid overfitting, dropout is used after every 256-channel thick layer. The network uses 0.0001 as its learning rate. The Adam algorithm is applied in conjunction with the categorical cross-entropy loss function during optimisation.

1. **Dropout:** In the dropout strategy, every hidden neuron's output is set to zero with a chance of 0.5. Neurones that have been initialised to zero are not involved in either backward or forward propagation Baldi and Sadowski (2013). Because each neurone must function independently of other neurones in the same layer, there is a decrease in the complicated co-adaptations of neurones as a result. As a result, neurones are forced to acquire a number of noteworthy characteristics that work well together. An additional method of minimising overfitting is data

augmentation. In order to prevent overfitting, the models' learning rate was also altered. A hyper-parameter called learning rate regulates how much the network's weights are adjusted in relation to the loss gradient.

2. **Weight decay:** Weight decay, or L2 regularization, is a method mostly used to reduce overfitting by reducing large weights in a neural network. It works in such a way that it will add extra term know as extra weight to the loss function. which was added based on the addition of the sum of the squared weights Zhang et al. (2018). In this technique model kept weight as small as possible, which help to avoid making model too difficult to understand. As a result, model performance increases because of small part taken in consideration which is well on new unseen data, making weight decay was so useful technique is deep learnng tasks like image classification.

6 Evaluation and results

6.1 Evaluation of VGG16 architecture using confusion matrix

The VGG16 model's confusion matrix help in analysis of the classification performance for both the actual classes in the model as shown in Figure 7. The confusion matrix is essential for judging the classification model's performance, beyond it's basic accuracy measurements. It makes it possible to compute a number of crucial metrics, include Specificity, F1-Score, Precision, and Recall, which provide a more thorough picture of how the model performs across the board in the classification task. The following is one interpretation of the matrix:

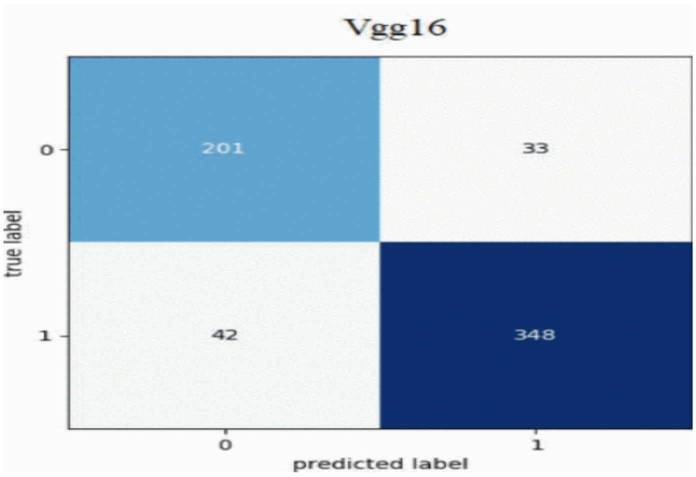


Figure 7: VGG16 Confusion Matrix

1. **True Positives (TP):** 348 occurrences were correctly predicted by the model as class 1.

2. **True Negatives (TN):** 201 instances were accurately classified as class 0 by the model.

3. **False Positives (FP):** 33 cases that belong to class 0 were mistakenly predicted by the model to be in class 1.

4. **False Negatives (FN):** 42 cases that belong in class 1 were mistakenly predicted by the model to be in class 0.

6.1.1 Evaluation of VGG16 Architecture using learning curves

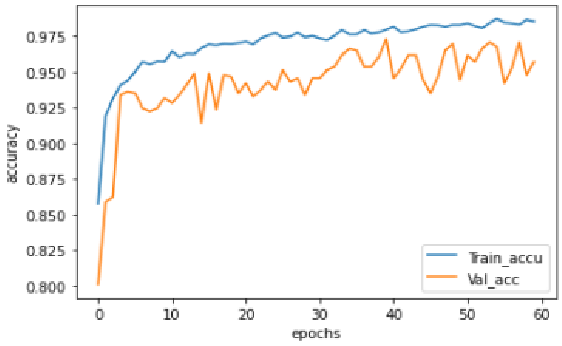


Figure 8: Learning Curve for Accuracy

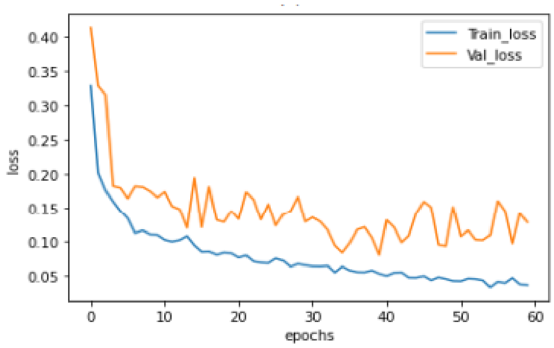


Figure 9: Learning Curve for Loss

The training and validation accuracy graphs reveal that the VGG16 model has a robust learning potential, with training accuracy rising progressively and stabilising at about 97.5% across 60 epochs. Though still high, the validation accuracy varies from 94% to 97%. Although the model performs well on the training set, this volatility raises the possibility of overfitting, in which case the model may not generalise as well to newly discovered validation data. These differences in validation accuracy suggest that more regularisation or fine-tuning could help the model become more resilient.

In a similar way, the loss graphs offer more information on the effectiveness of the model. As training accuracy rises, the training loss steadily falls, suggesting that the model is picking up new information efficiently. Conversely, the validation loss exhibits oscillations that align with the trends in validation accuracy, despite typically declining. This shows that even while the model works well, there is room for improvement in terms of how well it generalises. Strategies like dropout, data augmentation, and more hyperparameter tuning might help stabilise the validation loss and perhaps improve the model's performance on unseen data.

6.2 Results for VGG16 Model

The VGG16 model, which was modified to identify pneumonia, worked remarkably well. We made use of the robust feature extraction capabilities of VGG16 using transfer learning, which produced excellent accuracy. The model obtained 94.23% validation accuracy and 94.23% test accuracy through the use of regularisation techniques and hyperparameter adjustment. These results demonstrate that using pre-trained algorithms to classify medical images is successful.

- **Evalution of Precision**:With a high precision of almost 99.5%, VGG16 demonstrated how good the model is at reducing false positives.
- **Evaluation of Recall**: The VGG16 model's recall is approximately 96.5%, indicating that, in comparison to its precision, it is marginally less successful at identifying all pertinent cases (true positives).

Metric	Value (%)
Accuracy	94.23
Precision	99.5
Recall	96.5
Sensitivity (Recall)	96.5
Specificity	99.5
Area Under Curve (AUC)	99

Table 2: Performance Metrics of VGG16 Model

- **Evaluation of F1-Score**: The harmonic mean of recall and precision, called the F1-score, is approximately 98%. The general reliability of the VGG16 model in classification tasks is confirmed by this statistic, which strikes a compromise between precision and recall.
- **Evaluation of Specificity**: With a specificity of around 99.5%, VGG16 demonstrates remarkable accuracy in detecting true negatives.
- **Evaluation of Sensitivity**: The model's 96.5% recall or sensitivity confirms that it can identify true positives.
- **Evaluation of Area Under Curve (AUC)**: Excellent success in differentiating between the positive and negative classes is indicated by the AUC metric, which is about 99%.
- **Evaluation of Accuracy**: Although the VGG16 model's accuracy is little less than that of the hybrid model, it is still rather high at roughly 94.23%. This suggests that the model is a dependable option for classification jobs because it performs well in all categories.

6.3 Evaluation of Rasnet152V2 Architecture using learning curves

The ResNet152V2 models performance on the pneumonia detection task was evaluated using a detailed analysis of the learning curves, with a focus on accuracy and loss across the training epochs. The figure show these important indicators and also shed light on how the model behaved during the training and validation stages.

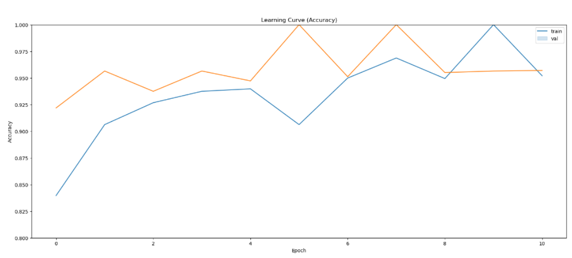


Figure 10: Learning Curve for Accuracy

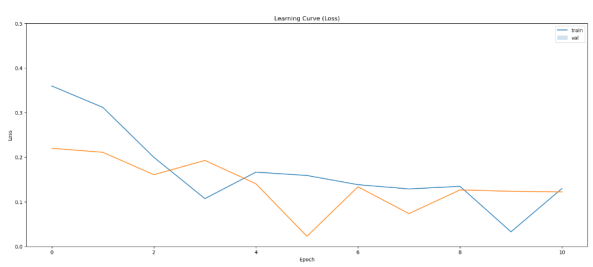


Figure 11: Learning Curve for Loss

The accuracy learning curve for each training epoch is shown in the figure 10. As we can demonstrate from graph that the model's accuracy increases noticeably throughout each epoch, in both training and validation accuracy. This resembles that the model is successfully picking up the characteristics required to differentiate between chest X-rays that are normal and those that are afflicted by pneumonia. The validation accuracy shows that the model is effectively generalising to new data and learning well, as it closely matches the initial training accuracy with minimum divergence. The model has strong performance in correctly classifying the X-ray photos at the end of the training period, as shown by its high accuracy.

The learning curve for loss is shown in figure 11, offering yet another crucial viewpoint on the models training. A sign that the model is convergent is the decreasing loss curve for both training and validation data over time. The model's prediction increases accurately as training goes on as it's directly proportional to decreasing loss values, which gives us a difference between the predicted and actual classifications. Additionally, the model may not overfit based on the steady decline in training and validation loss. The model can still generalise outside of the training dataset, as evidenced by the little loss difference between training and validation.

6.4 Evaluation of Rasnet152V2 Architecture using confusion matrix

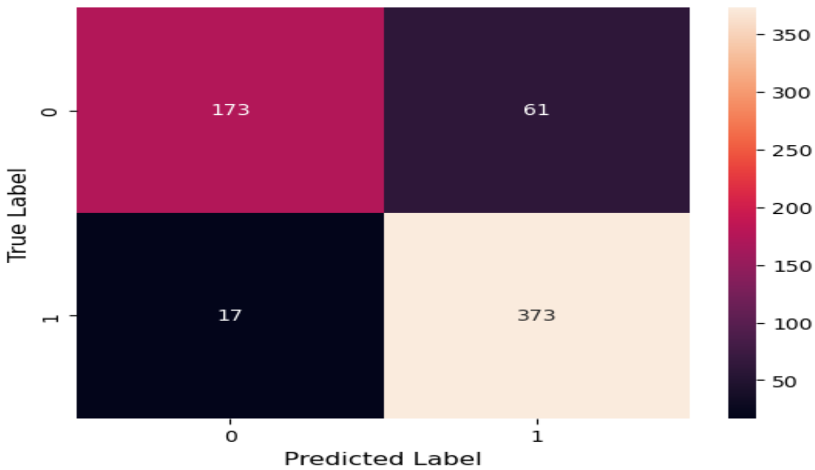


Figure 12: RasNet152V2 Confusion Matrix

The ResNet152V2 model's performance on the test dataset is thoroughly broken down in the confusion matrix, which primarily highlights the model's accuracy in differentiating between pneumonia patients and regular cases. The actual labels are shown in the matrix as rows, while the predicted labels are shown as columns. Understanding the accuracy, recall, and general efficacy of the model in categorising chest X-ray pictures depends on this matrix as shown in figure 12.

Interpretation of the Confusion Matrix

- **True Positives (TP):** As shown in the bottom-right matrix cell, the model properly classified 373 photos as pneumonia. In these instances, the patient's actual state and the model's predictions match exactly.

- **True Negatives (TN):** The top-left cell shows the 173 cases in which the model properly identified normal chest X-rays. This illustrates how well the model can identify cases without pneumonia.

- **False Positives (FP):** 61 normal chest X-rays were mistakenly identified as pneumonia by the model. These false positives are displayed in the matrix's top-right cell. Despite the model's overall accuracy, this figure shows the regions in which it incorrectly identified healthy people.
- **False Negatives (FN):** As the bottom-left cell illustrates, there were 17 instances in which the model failed to identify pneumonia in patients who genuinely had it. In a medical setting, these false negatives are especially worrisome since they show cases where the model overlooked important diagnoses.

6.5 Results for Rasnet152V2 Model

Class	Precision	Recall	F1-Score	Support
0	0.91	0.74	0.82	234
1	0.86	0.96	0.91	390
Accuracy	0.91 (624)			
Macro Avg	0.88	0.85	0.86	624
Weighted Avg	0.88	0.88	0.87	624

Table 3: Rasnet152V2 model results from performance matrix

When it comes to classifying chest X-ray pictures into normal and pneumonia categories, the ResNet152V2 model has shown promising results. For normal cases, the model maintained a recall of 0.74, suggesting that 74% of the actual normal cases were properly classified, while achieving a precision of 0.91, implying that 91% of instances projected as normal were correctly identified. With a precision of 0.86 and a high recall of 0.96 for pneumonia patients, the model demonstrated that almost all instances of pneumonia were correctly recognised. The model's predictions showed a balanced trade-off between precision and recall, as evidenced by the F1-scores of 0.82 and 0.91 for the normal and pneumonia classes, respectively.

The model's 91% total accuracy shows how well it can distinguish between pneumonia cases and healthy patients. The macro and weighted averages for recall, F1-score, and precision were consistently high, with an average of 0.88 for all measures. These results demonstrate that the model performs consistently across both groups, making it a valuable diagnostic tool for illnesses. Regarding pneumonia, the high recall lowers the chance of missed diagnoses, while the high accuracy ensures that the model maintains a low rate of false positives.

7 Conclusion and Future Work

In this research, we have used the deep learning capabilities of algorithms like Convolutional Neural Networks to detect pneumonia effectively and accurately classify it based on the chest X-rays data available for analysis. We used models like VGG16 and ResNet152V2 for pneumonia detection along with techniques such as Grad-CAM. These techniques ensured that highly reliable model is development by proper data preprocessing with the help of data augmentation techniques, ensuring higher model interpretability. These transformations gave us very good results with VGG16 giving an accuracy of

94.23% and ResNet152V2 of 92.26%. This model will help the healthcare experts to conduct tests with the help of few resources and staff effectively. With the integration of Grad-CAM the model has gained openness and has become reliable for clinical settings.

7.0.1 Future Work

- **Generalisation of the Model:** Expanding the dataset with more diverse chest X-ray pictures from different age groups, communities, and healthcare settings should improve the model's capacity for generalisation. Moreover, including images of pneumonia varying in severity may improve the model's diagnostic accuracy.
- **Cross-Platform Validation:** To ensure that the system is flexible and easy to use in a variety of healthcare settings, especially in remote or underprivileged areas, it would be necessary to verify the models on a wide range of hardware platforms, such as low-power and mobile devices.
- **identification of Multi-Disease:** While this study focused on the diagnosis of pneumonia, future research may investigate the simultaneous identification of other respiratory disorders utilising chest X-rays, such as COVID-19 or tuberculosis. To make the system more useful, multi-task learning models that can identify and discriminate between various circumstances might be developed.
- **Checking Grad-CAM visualised results with Radiologist to implement and check models use in the healthcare** We are planing to check all the results or visualisations with different Radiologist to make sure weather our models behaviour working well and based on this evaluations and result it could be further implemented in medical sector to help decting pneumonia more better and faster manner.

References

Abdel-Basset, M., Hawash, H., Alnowibet, K. A., Mohamed, A. W. and Sallam, K. M. (2022). Interpretable deep learning for discriminating pneumonia from lung ultrasounds, *Mathematics* **10**(21): 4153.

Baldi, P. and Sadowski, P. J. (2013). Understanding dropout, *Advances in neural information processing systems* **26**.

Hasan, M. J., Alom, M. S. and Ali, M. S. (2021). Deep learning based detection and segmentation of covid-19 & pneumonia on chest x-ray image, *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)*, IEEE, pp. 210–214.

Hasan, M. R., Hossain, A. A. and Ullah, S. M. A. (2024). Ensemble resdensenet: Alzheimer's disease staging from brain mri using deep weighted ensemble transfer learning, *International Journal of Computers and Applications* **46**(7): 539–554.

Hashmi, M. F., Katiyar, S., Keskar, A. G., Bokde, N. D. and Geem, Z. W. (2020). Efficient pneumonia detection in chest xray images using deep transfer learning, *Diagnostics* **10**(6): 417.

Jain, R., Nagrath, P., Kataria, G., Kaushik, V. S. and Hemanth, D. J. (2020). Pneumonia detection in chest x-ray images using convolutional neural networks and transfer learning, *Measurement* **165**: 108046.

Li, B., Kang, G., Cheng, K. and Zhang, N. (2019). Attention-guided convolutional neural network for detecting pneumonia on chest x-rays, *2019 41st annual international conference of the ieee engineering in medicine and biology society (EMBC)*, IEEE, pp. 4851–4854.

Liang, G. and Zheng, L. (2020). A transfer learning method with deep residual network for pediatric pneumonia diagnosis, *Computer methods and programs in biomedicine* **187**: 104964.

Naralasetti, V., Shaik, R. K., Katepalli, G. and Bodapati, J. D. (2021). Deep learning models for pneumonia identification and classification based on x-ray images., *Traitement du Signal* **38**(3): 903–909.

Panwar, H., Gupta, P., Siddiqui, M. K., Morales-Menendez, R., Bhardwaj, P. and Singh, V. (2020). A deep learning and grad-cam based color visualization approach for fast detection of covid-19 cases using chest x-ray and ct-scan images, *Chaos, Solitons & Fractals* **140**: 110190.

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

Reshan, M. S. A., Gill, K. S., Anand, V., Gupta, S., Alshahrani, H., Sulaiman, A. and Shaikh, A. (2023). Detection of pneumonia from chest x-ray images utilizing mobilenet model, *Healthcare*, Vol. 11, MDPI, p. 1561.

Sharma, S. and Guleria, K. (2023). A deep learning based model for the detection of pneumonia from chest x-ray images using vgg-16 and neural networks, *Procedia Computer Science* **218**: 357–366.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition, *arXiv preprint arXiv:1409.1556* .

Szepesi, P. and Szilágyi, L. (2022). Detection of pneumonia using convolutional neural networks and deep learning, *Biocybernetics and biomedical engineering* **42**(3): 1012–1022.

Varshni, D., Thakral, K., Agarwal, L., Nijhawan, R. and Mittal, A. (2019). Pneumonia detection using cnn based feature extraction, *2019 IEEE international conference on electrical, computer and communication technologies (ICECCT)*, IEEE, pp. 1–7.

Vetrithangam, D., SATVE, P. P., KUMAR, J. R. R., Anitha, P., Vidhya, S. and SAINI, A. K. (2023). prediction of pneumonia disease from x-ray images using a modified resnet152v2 deep learning model, *Journal of Theoretical and Applied Information Technology* **101**(17).

Zhang, G., Wang, C., Xu, B. and Grosse, R. (2018). Three mechanisms of weight decay regularization, *arXiv preprint arXiv:1810.12281* .