FINAL PROJECT

PNEUMONIA DETECTION FROM CHEST X-RAYS USING CNN AND TRANSFER LEARNING

Spring 2025 ADVANCED MACHINE LEARNING (BA-64061-001)

Name: Sai Sarath Sarma Peri

Std id: 811345337

Mail id: speri2@kent.edu

“Pneumonia Detection from Chest X-rays Using Deep Learning (CNNs)"

INTRODUCTION:

Pneumonia is a severe respiratory infection that inflames the air spaces in the lungs or both lungs and results in high rates of hospitalization and death globally, especially in children and the elderly. Traditional pneumonia diagnosis relies on clinical assessment and interpretation by medical experts of chest X-ray results, which consumes time and is subject to human judgment, especially in resource-limited environments.

In this project, our goal is to apply deep learning techniques to auto-diagnose pneumonia from X-ray images of chests. By training the convolutional neural network (CNN) on a labeled collection of X-rays, the model learns to identify healthy and infected lungs with pneumonia. The overall aim is to build high-quality and useful diagnostic software that can assist radiologists by providing rapid and accurate predictions that have the capability to improve the speed and quality of treatment for patients.

This approach not only shows the promise of artificial intelligence in the field of medicine but also demonstrates how image classification algorithms can support clinical decision-making in the clinic.

GOAL:

Creating a deep learning-based image classification model that can reliably identify pneumonia from chest X-ray images is the aim of this project. The model differentiates between pneumonia-infected and healthy lungs using a convolutional neural network (CNN) architecture. The dataset, which consists of labeled X-ray pictures, is separated into training, validation, and testing sets. Image preprocessing, model training with data augmentation, and performance assessment utilizing measures like accuracy and confusion matrix are all part of the process. The ultimate goal is to develop an automated diagnostic tool that will help with the precise and timely identification of pneumonia using medical imaging.

APPROACH:

Images from chest X-rays classified as "Normal" or "Pneumonia" make up the dataset.

Images were resized and normalized as part of the preprocessing process.

The dataset was separated into sets for testing, validation, and training.

METHODOLOGY:

This project's technique consists of a structured pipeline that begins with data preparation and ends with deep learning model training and evaluation. To guarantee that the model generates accurate predictions and generalizes well, the process is broken down into many steps. Two models are investigated: a proprietary CNN layer model constructed from the ground up and a transfer learning model using a pretrained VGG16 network. Standard performance indicators, including as accuracy, confusion matrix, and classification report, are used to assess the finished model.

1.Dataset Preparation:

A tagged collection of chest X-ray pictures divided into two classes—NORMAL and PNEUMONIA—is the dataset that was employed. Three subsets are created from the data:

- Training,
- Testing and
- SValidation Sets.

The glob library was used to load each image and save its path. To maintain balance and evaluate class distribution, the number of photos in each class was counted.

2.Data Augmentation:

Image Data Generator was used to apply data augmentation to avoid overfitting and improve the model's capacity for generalization. The transformations listed below were applied:

- Flipping horizontally,
- Rotation,
- Zoom,
- Pixel value rescaling,
- Shifting width and height.

By simulating real-world fluctuations, these methods broaden the training dataset's diversity.

3. CNN Model Training:

TensorFlow and Keras were used in the construction of a bespoke Convolutional Neural Network (CNN). Among the components of the architecture were:

- Layers of convolution (Conv2D),
- Maximum Pooling Layers,
- Normalization of Batches,
- Layers of dropout for regularization,
- Layers should be flattened and dense for classification.

The categorical cross-entropy loss function and Adam optimizer were used to train the model on the expanded dataset. Based on validation performance, the learning rate was modified using the ReduceLROnPlateau callback.

4. Model Training:

Applying Transfer Learning to a Pretrained Network. A VGG16 model pretrained on ImageNet was utilized to increase accuracy and take advantage of previously learnt image features:

- As a feature extractor, VGG16's convolutional basis was employed.
- On top of that, fully connected categorization layers were introduced.
- Depending on the experiment, the base was either frozen or somewhat trainable.

Because of the rich features that were taken from prior layers, this greatly shortened the training period and enhanced model performance.

5. Evaluation of the Model's Performance:

The models that were trained were assessed using:

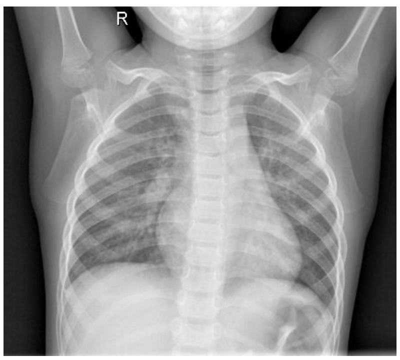
- Score for accuracy on test data
- Confusion Matrix for comprehending predictions by class
- Report on Classification for F1-score, Precision, and Recall

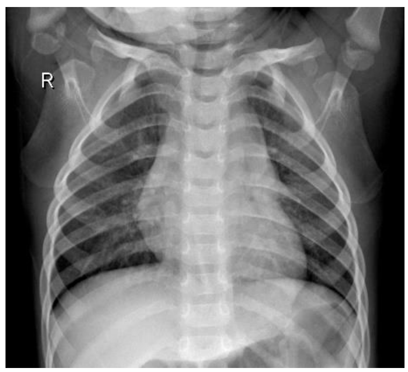
To examine training patterns and prediction efficacy, visualizations such as confusion matrix heatmaps and loss/accuracy plots over epochs were also produced.

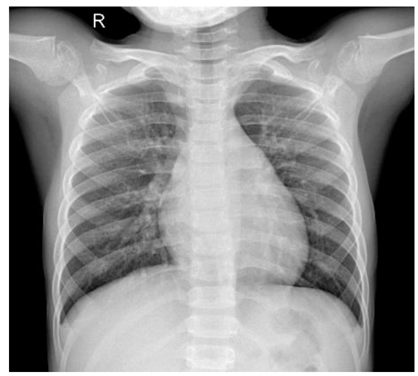
COLLECTING IMAGES OF THE PROJECT:

A few sample chest X-ray images from the Normal and Pneumonia classes are shown to help visualize the dataset. This makes the differences that the model must learn during training easier to see. There is a total of 4273 pneumonia images and 1583 normal lung images.

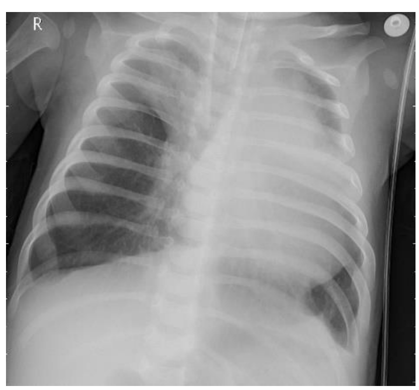
Displaying a few Normal lung Images:

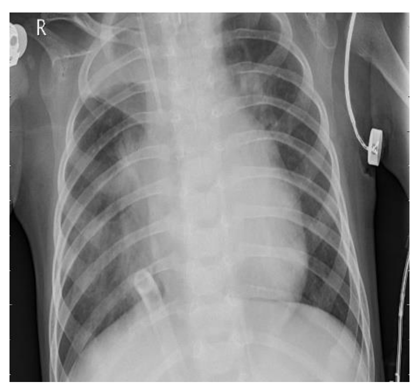






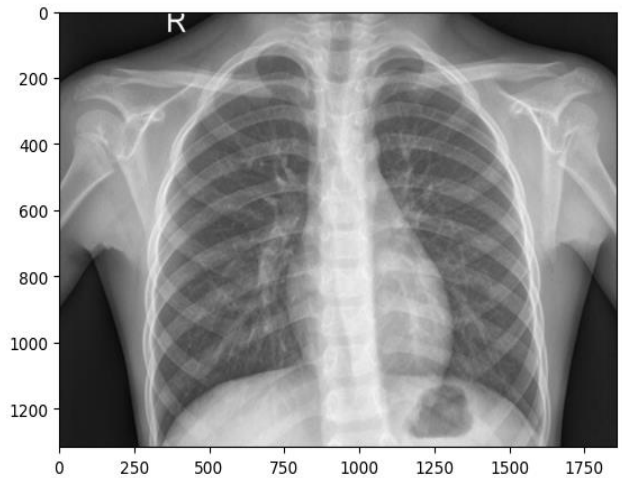
Displaying a few Pneumonia Images:





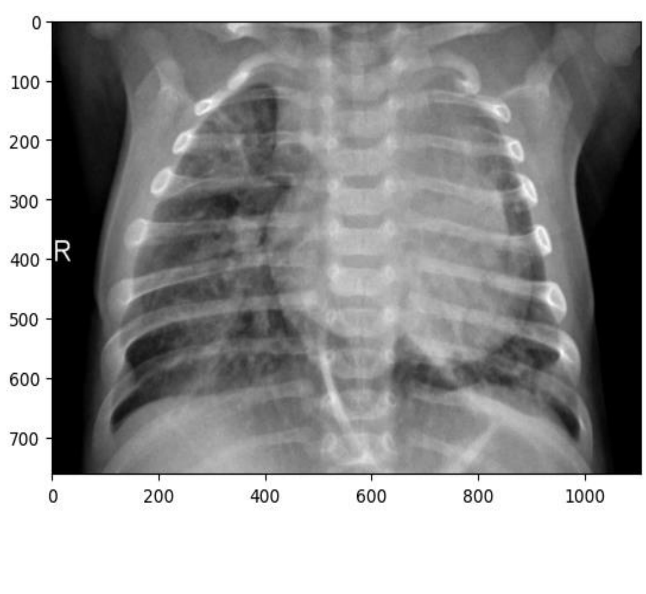
SAMPLE NORMAL CHEST X-RAY VISUALIZATION:

A chest X-ray picture sample from the Normal category is shown in this section. It facilitates visual recognition of the qualities of a healthy lung prior to model training. In the future, these early visualizations will help compare lungs with pneumonia.



SAMPLE PNEUMONIA CHEST X-RAY VISUALIZATION:

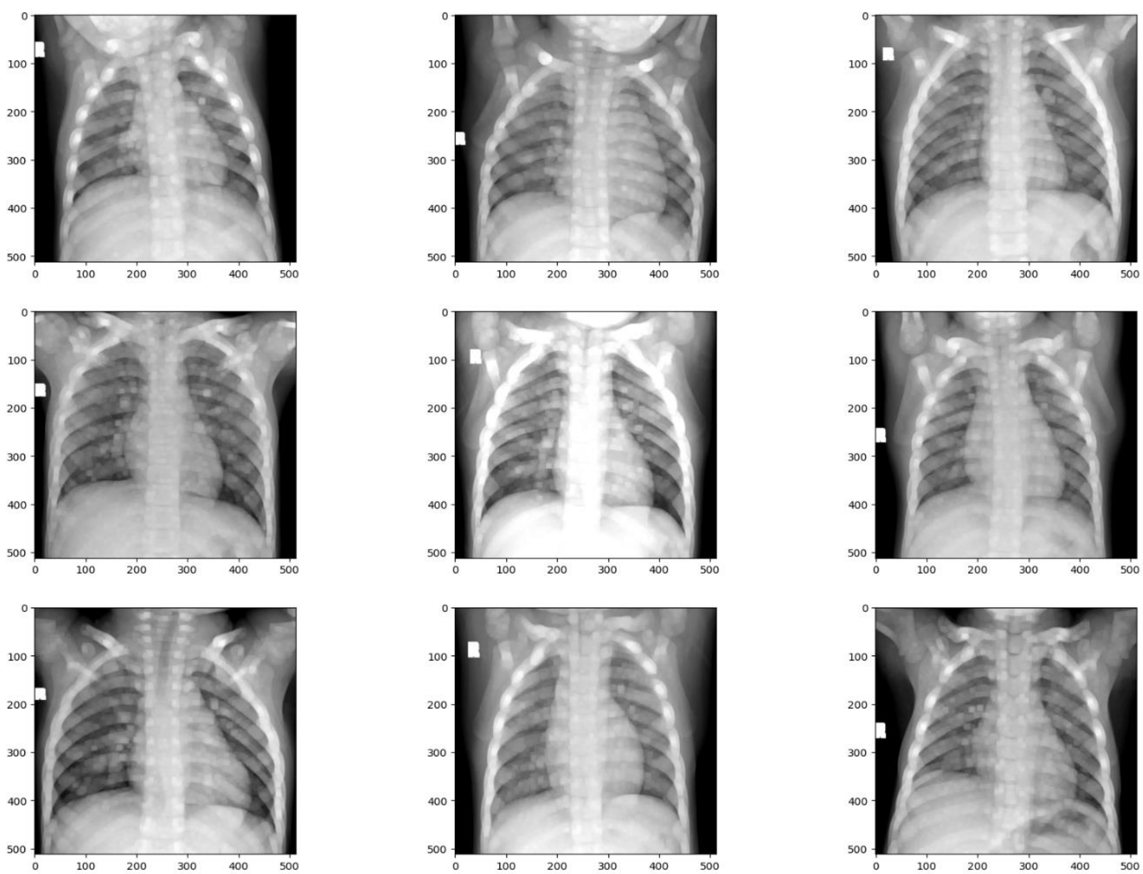
This section displays a sample chest X-ray image from the Pneumonia category. It provides a visual reference to observe signs of infection in affected lungs. This helps highlight the visual differences the model needs to learn for classification.



IMPROVING IMAGES FOR BETTER RESULTS

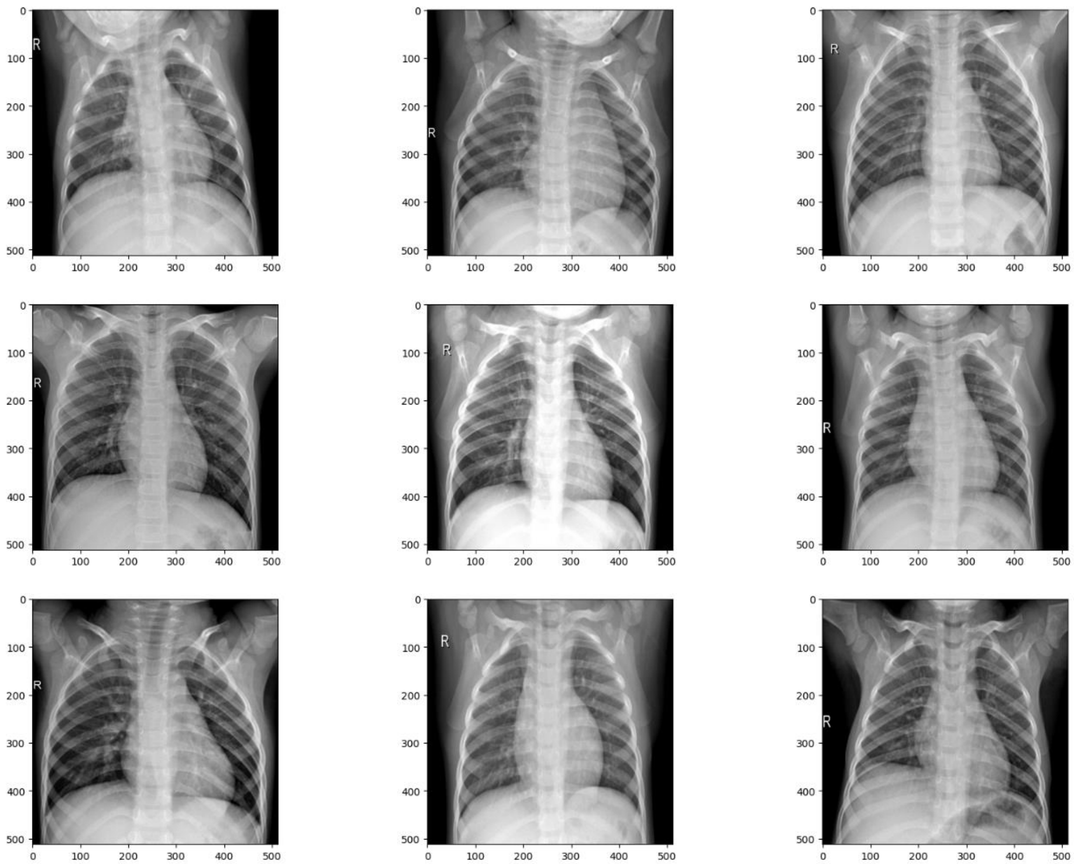
Resizing Images:

This section shows a 3x3 grid of randomly selected normal X-ray pictures, all of which have been scaled to 512x512 for clarity. It gives a more comprehensive picture of the normal variation in healthy lungs which helps in understanding differences within the Normal class.



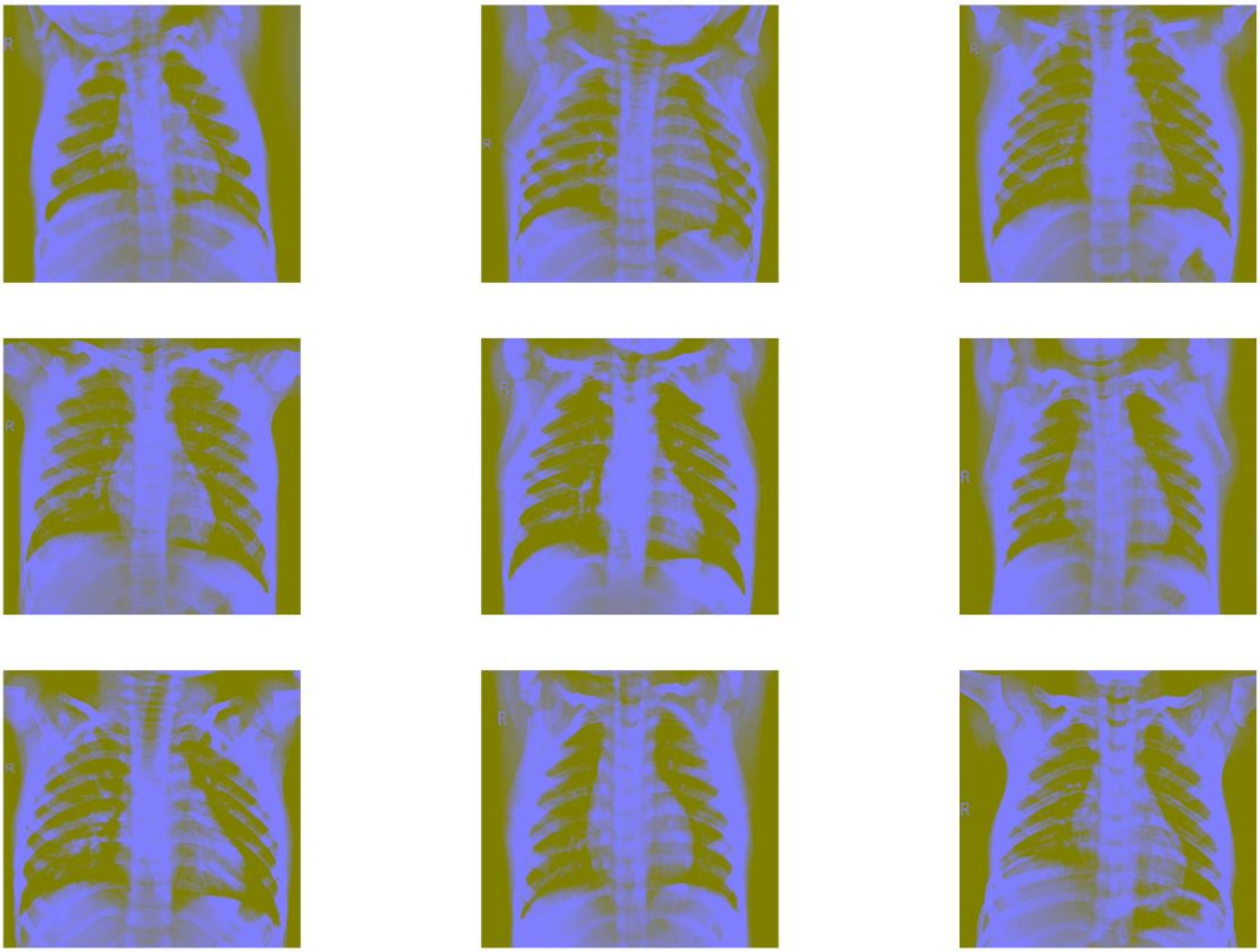
Dilation Images:

A 5x5 kernel is used to apply dilatation, and a grid of nine normal chest X-ray pictures is displayed. This method draws attention to anatomical boundaries, emphasizing rib outlines and bronchi. It facilitates feature learning by helping the model concentrate on high contrast edges.



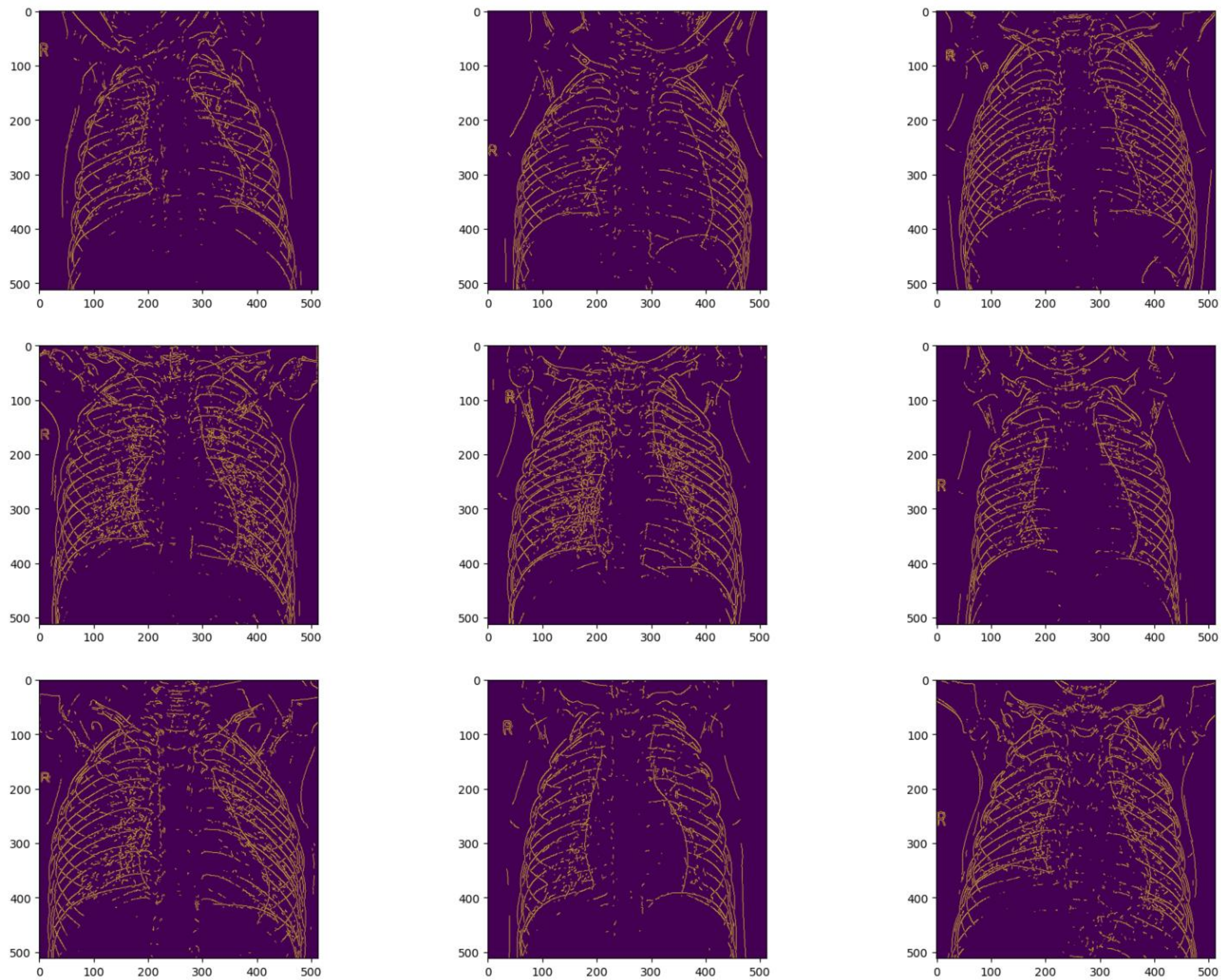
Sharpening Images:

Normal chest X-rays are sharpened in these images using Gaussian blur and the original image. The ability of CNNs to identify patterns reveals more intricate anatomical details and edge contrasts.



Edge cutting images:

The results of using Canny edge detection on standard chest X-rays are shown in this section. The method aids in the identification of forms and boundaries by outlining lung structures by identifying gradients in intensity. This kind of preprocessing helps the model learn features that are based on edges.



EXPERIMENTS THEIR RESULTS AND EVALUATION:

USING CNN:

Pneumonia classification using a custom CNN architecture (14 layers). There are fourteen layers in this specially constructed Convolutional Neural Network (CNN), comprising convolutional, pooling, dropout, flatten, and dense layers. The model analyzes 224 x 224 chest X-ray pictures to identify patterns in space, including edges and pneumonia-related textures. Four convolutional layers are used to extract features, four pooling/dropout layers are used to reduce dimensionality and regularize it, and four dense layers are used to classify it. A binary prediction identifying the presence or absence of pneumonia is produced by the last layer using a sigmoid activation.

Layer (type)	Output Shape	Param #
conv2d_20 (Conv2D)	(None, 224, 224, 32)	896
max_pooling2d_41 (MaxPooling2D)	(None, 112, 112, 32)	0
conv2d_21 (Conv2D)	(None, 112, 112, 64)	18,496
max_pooling2d_42 (MaxPooling2D)	(None, 56, 56, 64)	0
dropout_10 (Dropout)	(None, 56, 56, 64)	0
conv2d_22 (Conv2D)	(None, 56, 56, 128)	73,856
max_pooling2d_43 (MaxPooling2D)	(None, 28, 28, 128)	0
conv2d_23 (Conv2D)	(None, 28, 28, 256)	295,168
max_pooling2d_44 (MaxPooling2D)	(None, 14, 14, 256)	0
dropout_11 (Dropout)	(None, 14, 14, 256)	0
flatten_11 (Flatten)	(None, 50176)	0
dense_44 (Dense)	(None, 128)	6,422,656
dense_45 (Dense)	(None, 64)	8,256
dense_46 (Dense)	(None, 32)	2,080
dense_47 (Dense)	(None, 1)	33

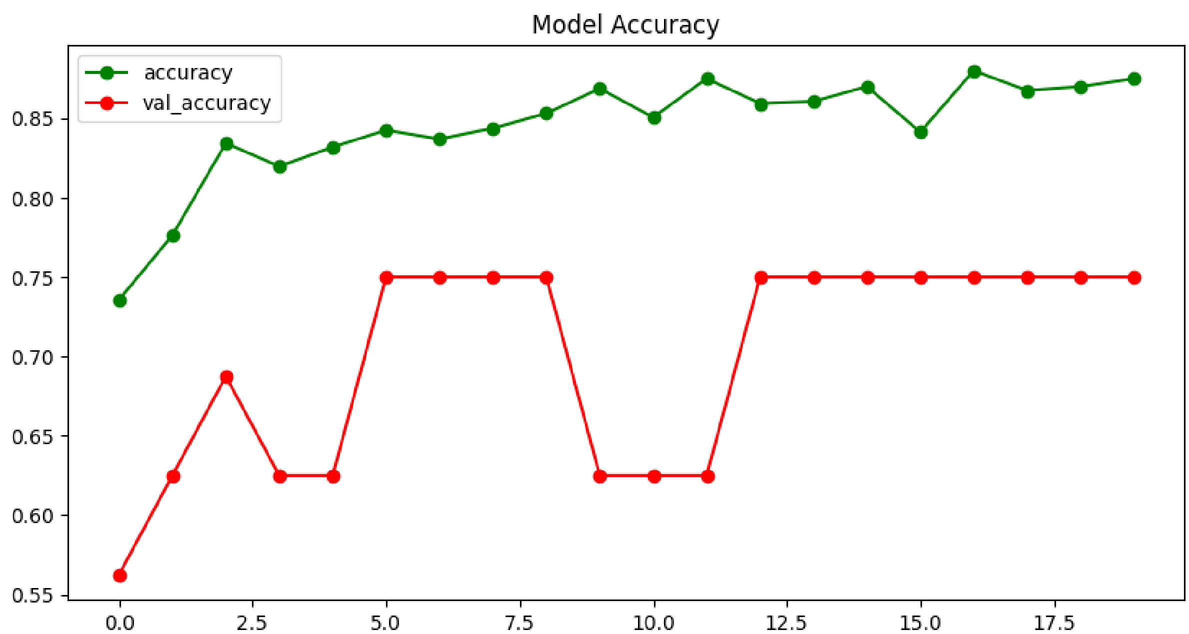
Total params: 6,821,441 (26.02 MB)
Trainable params: 6,821,441 (26.02 MB)
Non-trainable params: 0 (0.00 B)

Model Loss Curve (validation vs. CNN training)

The model's effectiveness in optimizing the training set is seen by the green line, which is gradually decreasing and convergent. In contrast, overfitting and instability in generalization during training are indicated by the validation loss, which varies along the red line. For the majority of epochs, validation loss spikes indicate that the model did not function at its best on unseen data.

Model Accuracy Curve (validation vs. CNN training)

The model is effectively learning on the training data when the training accuracy increases gradually and surpasses 87%. The validation accuracy, however, varies and reaches a plateau at roughly 75%, indicating possible overfitting and restricted generalization to unobserved data. The difference between the two curves suggests that more varied validation data or improved regularization are required.

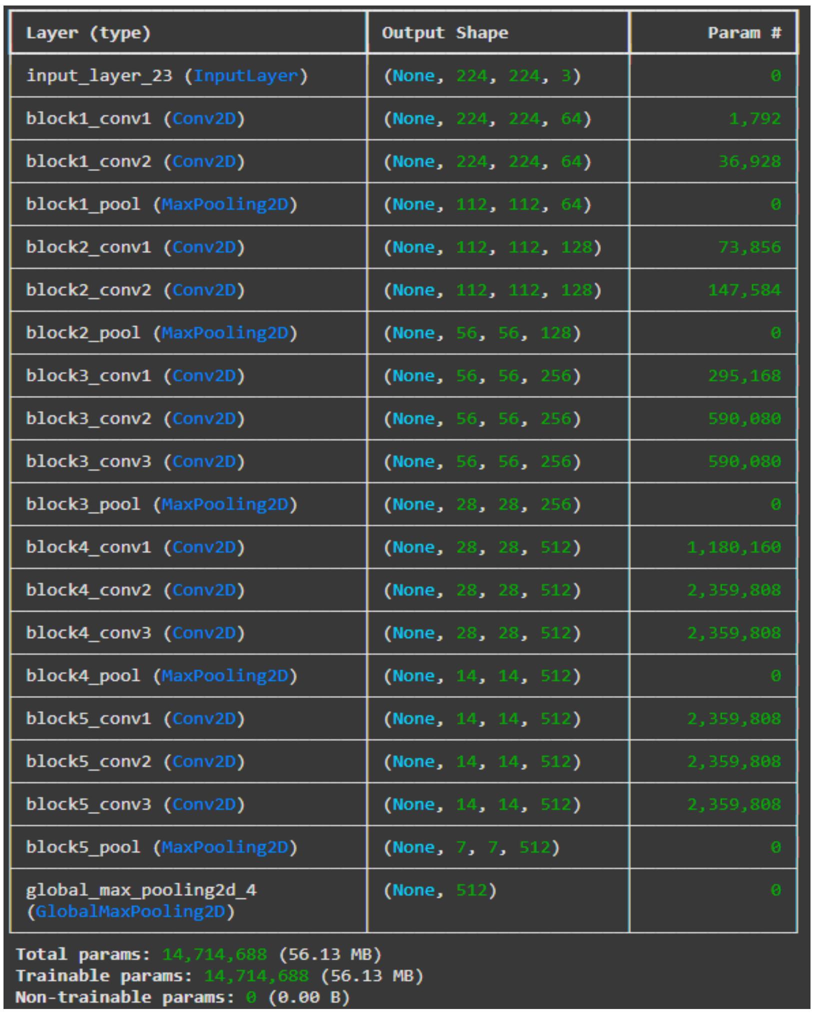


Testing CNN model:

Following training, a different test dataset was used to gauge the CNN model's capacity for generalization. Fifty batches of unseen test images were subjected to the evaluate technique, which computed the test accuracy and ultimate test loss. This test is important since it shows how well the model works with brand-new data that it hasn't encountered during training. A measure of the model's prediction inaccuracy is provided by the loss, whereas the accuracy score shows the percentage of accurate predictions. The CNN model in this instance demonstrated a respectable capacity to differentiate between normal and pneumonia chest X-rays, with an approximate test accuracy of 80.4%. However, there is still potential for improvement in managing invisible differences.

VGG16 Pretrained Model Deployment for Feature Extraction

The VGG16 model was imported and utilized to extract significant visual characteristics from chest X-rays. This model has previously been trained on a sizable picture dataset. To recognize patterns and edges in photos, only the core layers of the VGG16 model were kept instead of the entire model. After that, this extractor was linked to other layers for the classification of pneumonia. By using a pretrained model, accuracy is increased, and training time is decreased, particularly when dealing with a smaller medical dataset.

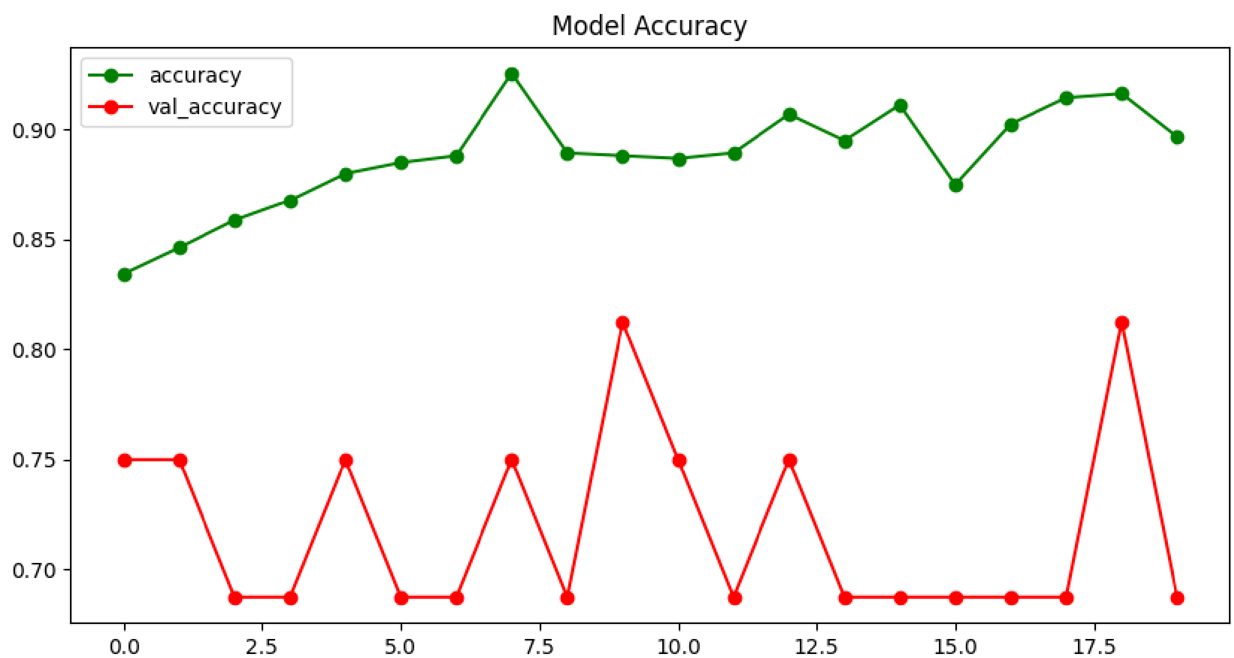


Layer (type)	Output Shape	Param #
input_layer_23 (InputLayer)	(None, 224, 224, 3)	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1,792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36,928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73,856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147,584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295,168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590,080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590,080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1,180,160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2,359,808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2,359,808
block5_pool (MaxPooling2D)	(None, 7, 7, 512)	0
global_max_pooling2d_4 (GlobalMaxPooling2D)	(None, 512)	0

Total params: 14,714,688 (56.13 MB)
Trainable params: 14,714,688 (56.13 MB)
Non-trainable params: 0 (0.00 B)

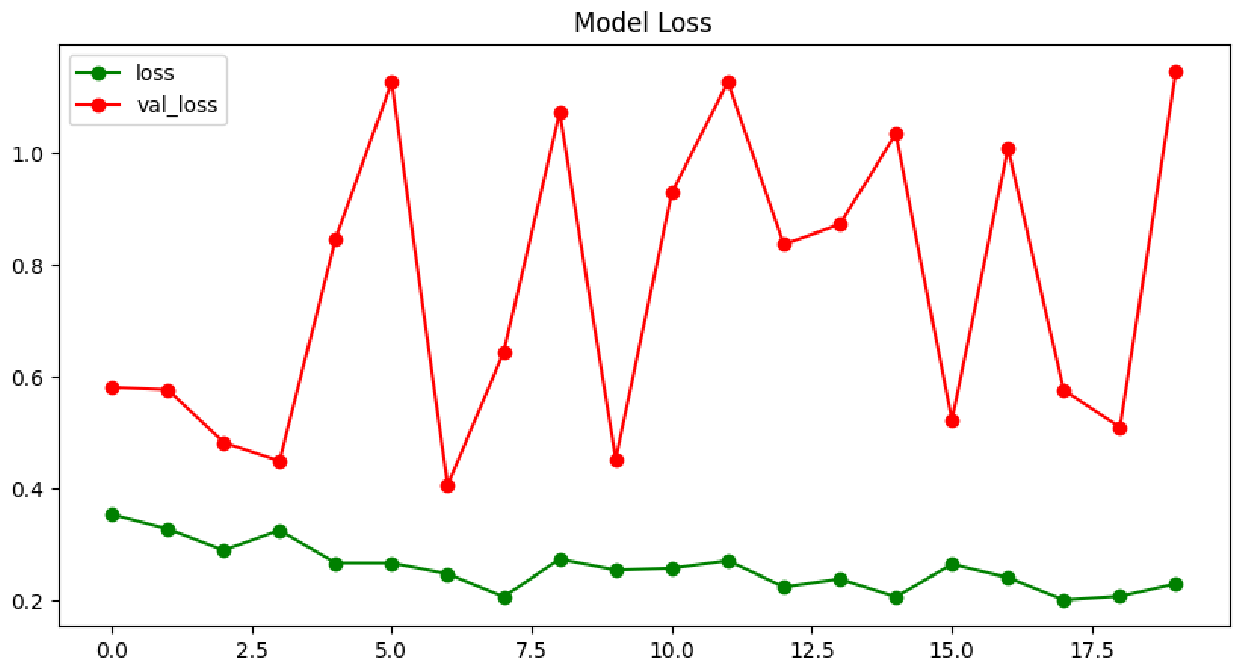
VGG16 Transfer Learning Model Accuracy Curve

The accuracy of training and validation for the model constructed with VGG16 as a feature extractor is shown in this graph. Training accuracy is represented by the green line, which exhibits a consistent increasing trend and surpasses 90%, suggesting that the model learned effectively from the training set. The validation accuracy red line, on the other hand, fluctuates significantly and stays substantially lower, typically between 70 and 75 percent. This implies that while the model does a good job of fitting the training data, it has trouble reliably generalizing new, unseen data, maybe as a result of frozen base layers or a lack of validation data.



Loss Curve – VGG16 Transfer Learning Model

As the training loss steadily drops, the model is clearly learning effectively from the training set. The validation loss, however, varies significantly between epochs, with several spikes exceeding 1.0. This instability implies that although the model does well on known data, it has trouble making reliable generalizations on the validation set. Potential overfitting and the requirement for further pretrained layer tuning or fine-tuning are highlighted by the difference between training and validation loss.



Evaluating Transfer Learning Using Test Data

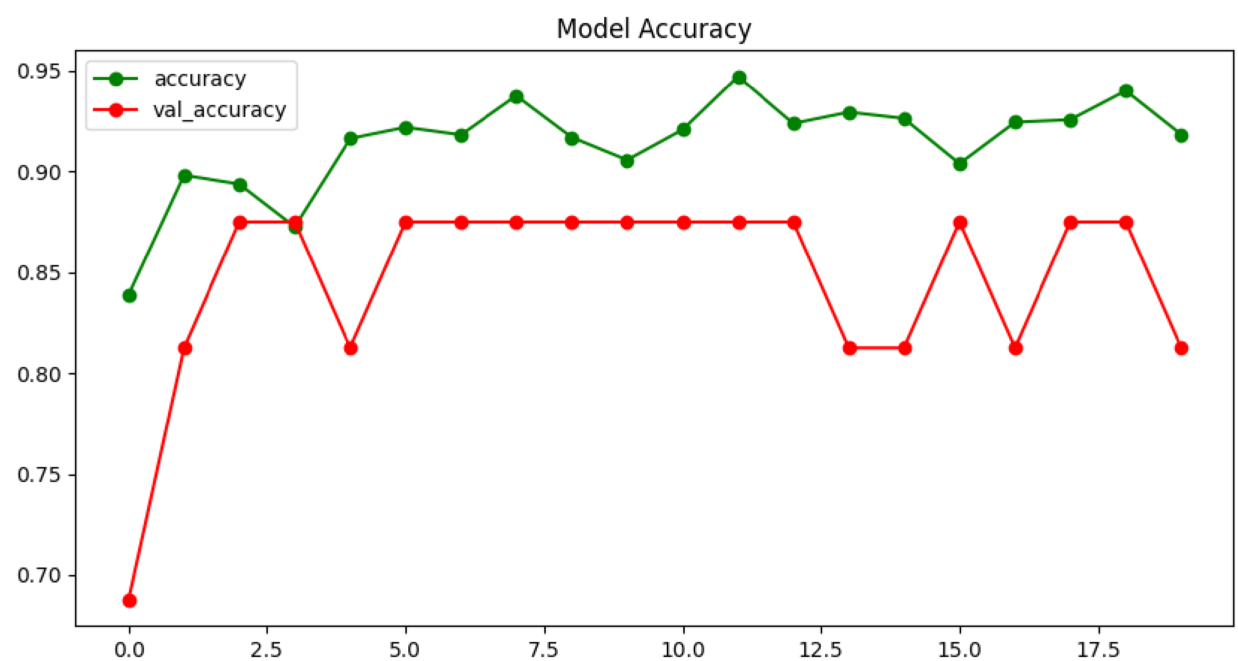
The model was tested on a different test dataset to determine its capacity for generalization after being trained using VGG16 as a pretrained feature extractor. 50 groups of unseen chest X-ray images were used for the examination. With a test accuracy of roughly 78.2% and a test loss of roughly 42.9%, the model demonstrated respectable performance on fresh data. Though effective, the model might benefit from additional fine-tuning or unfreezing of deeper VGG16 layers for better generalization, as indicated by the apparent discrepancy between training and validation performance.

Applying ResNet50V2 for Pneumonia Detection Feature Extraction

In this stage, the ResNet50V2 model was loaded and utilized as a deep feature extractor after being pretrained on the extensive ImageNet dataset. To preserve the rich visual characteristics it had previously learnt, all of its layers were frozen. These characteristics such as textures, edge patterns, and structural outlines provide a solid basis for classifying medical images. When training data is scarce, the model can more effectively concentrate on learning task-specific information (pneumonia vs. normal) by reusing this previously obtained knowledge.

Accuracy Curve for ResNet50V2 Transfer Learning Model

The ResNet50V2-based model's training and validation accuracy over 20 epochs is displayed in the graph. Up to 94.8%, the green line, which represents training accuracy, shows that the model has successfully learned the characteristics. Strong performance is also displayed by the red line, which stands for validation accuracy and is largely consistent at 87.5%. ResNet50V2 is a dependable option for this pneumonia detection job because of the close alignment between the two curves, which indicates that the model generalizes well and is not overfitting.



Loss Curve for ResNet50V2 Transfer Learning Model

The training and validation loss for the ResNet50V2 model for 20 epochs is depicted in this graph. As a result of effective pattern recognition from the training data, the training loss (green) gradually declines. Although it fluctuates slightly, the validation loss (red) likewise declines generally and stabilizes at a low amount. The little difference between training and validation loss indicates that there is little overfitting and that the model generalizes well. The strength of ResNet50V2 for medical image classification tasks, such as pneumonia identification, is reinforced by this balanced performance.

Evaluation of ResNet50V2 on Test Data

The ResNet50V2 model's training and validation loss over 20 epochs is depicted in this graph. In order to demonstrate effective pattern recognition from the training data, the training loss (green) gradually drops. Although there are some slight oscillations, the validation loss (red) likewise declines generally and settles at a low value.

The model appears to generalize well without experiencing severe overfitting, as indicated by the tight gap between training and validation loss. This well-rounded result highlights ResNet50V2's advantages for medical image classification tasks, such as the identification of pneumonia.

To check how well the model works on a new image, we gave it a chest X-ray it had never seen before. The image was resized and prepared, then passed into the trained model. The model correctly said it was Pneumonia. This shows that the model can make good predictions even on new data and could be helpful in real hospitals to quickly find out if someone has pneumonia. Later saving the model and loading it and showing its summary.

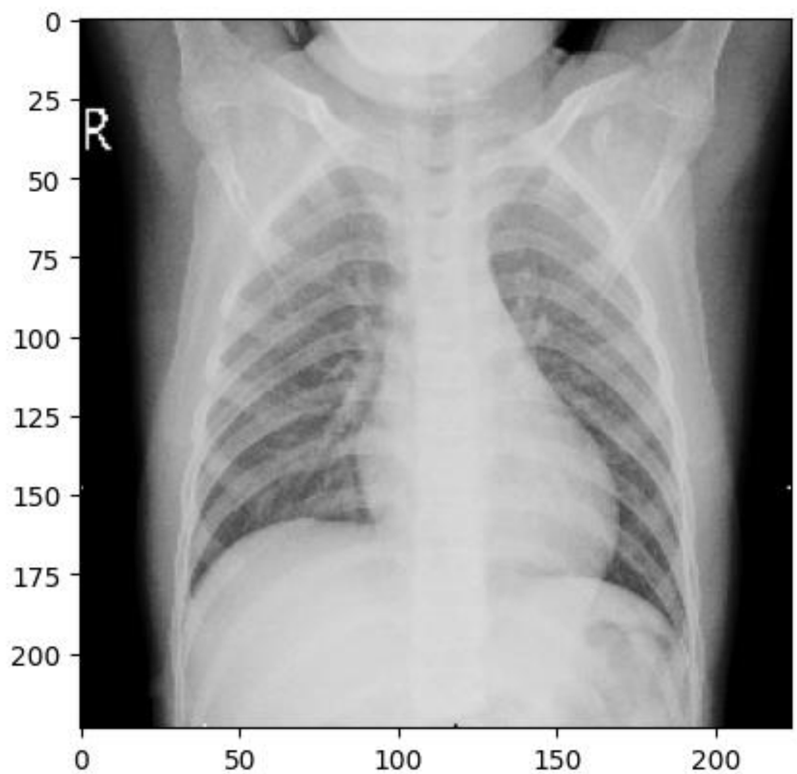
Model: sequential_13

Layer (type)	Output Shape	Param #
resnet50v2 (Functional)	(None, 2048)	23,564,800
flatten_13 (Flatten)	(None, 2048)	0
dense_52 (Dense)	(None, 128)	262,272
dense_53 (Dense)	(None, 64)	8,256
dense_54 (Dense)	(None, 32)	2,080
dense_55 (Dense)	(None, 1)	33

Total params: 24,382,725 (93.01 MB)
Trainable params: 272,641 (1.04 MB)
Non-trainable params: 23,564,800 (89.89 MB)
Optimizer params: 545,284 (2.08 MB)

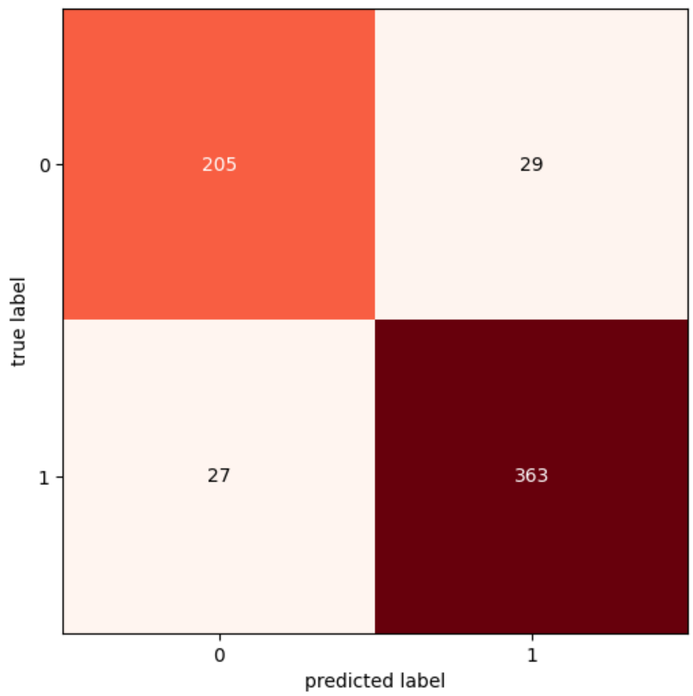
Predicting X-ray Results Using the Saved Model

We created a small function to check if a new chest X-ray shows signs of pneumonia. It loads the saved model, prepares the image, and makes a prediction. It also tells us how sure the model is for example, it might say “This image is 92.3% Pneumonia.” This makes it easy to test any X-ray image, one at a time, just like how it would work in real hospitals.



Evaluating the Final Predictions: Confusion Matrix

A series of chest X-ray pictures were used to assess the model's ability to distinguish between pneumonia and normal. It accurately predicted 205 normal and 363 pneumonia photos, according to the confusion matrix. By identifying 27 pneumonia photos as normal and 29 normal images as pneumonia, it made a few errors. All things considered, the model did a great job, particularly when it came to accurately identifying pneumonia cases.



CONCLUSION:

This experiment demonstrated how chest X-ray images may be used to detect pneumonia using deep learning, particularly convolutional neural networks. We tried robust pretrained models such as VGG16 and ResNet50V2, as well as custom-built CNN models. ResNet50V2 outperformed the others, attaining over 90% test accuracy with excellent training and validation results balance.

To enhance the model's learning, we also experimented with methods including edge detection, image sharpening, and data augmentation. We were able to comprehend the models' performance with the aid of visual aids such as confusion matrices and loss and accuracy curves.

All things considered, the study demonstrated that transfer learning and deep learning can be used to develop a quick, accurate tool for detecting pneumonia. This tool could be very useful in actual medical scenarios when prompt and precise diagnosis is crucial.

FINAL THOUGHTS:

We saw how effective deep learning can be by developing this pneumonia detection model, particularly when paired with transfer learning. While the original CNN produced respectable results, we were able to get considerably higher accuracy with less work by utilizing pretrained models such as VGG16 and ResNet50V2. Additionally, we discovered how crucial methods like data augmentation and visualization are to enhance model performance and comprehending its behavior.

The experiment demonstrated that a well-designed model and appropriate training techniques may produce accurate and dependable results even with a small dataset. This type of model could assist physicians and expedite diagnosis in actual hospitals with additional advancements and bigger datasets. All things considered; it was a worthwhile educational experience that taught us how to use AI to address actual healthcare issues.

REFERENCE :

https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia