Pneumonia Classification

Rachel Rea D’souza

*Dept. of Computer Science*

*(Coventry University)*

*Middle East College*

Student ID – 21F21817

**I. *Abstract***— **A respiratory disease called pneumonia is caused by lung tissue inflammation, which is typically brought on by bacterial, viral, or fungal infections. This dataset which focuses on chest X-ray images of pneumonia, includes several images showing people with pneumonia and people with healthy lungs. This dataset is intended to aid in creating algorithms for the binary classification of X-ray images of the chest into normal X-rays and pneumonia X-ray categories [3]. It is useful for studies on medical imaging analysis. Both the VGG16 model and a customized CNN were used, and their accuracy, precision, recall, and F1-score performance measures were thoroughly examined, analysed, and compared. The study concludes that VGG16 performs slightly better than CNN in identifying instances of pneumonia. While the overall accuracies for VGG16 and CNN were 61% and 59% respectively. The results provide detailed insights into each model, highlighting the need for more improvement to overcome difficulties, especially in the classification of normal X-rays. Overfitting was evident in both models, requiring the use of regularization strategies and data augmentation. The work emphasizes how complex medical image analysis is and how important it is to keep trying to increase model accuracy for more advanced medical uses.**

***Keywords — deep learning, pneumonia detection, chest x-ray images, binary classification, disease detection, medical image analysis, convolutional neural networks (CNN), VGG16, overfitting mitigation.***

# INTRODUCTION

Deep learning and Artificial intelligence (AI) have drastically changed the way we solve issues that arise in the real world. The development of artificial intelligence that can mimic human intellect and perform tasks that typically need human intelligence such as visual and speech recognition, computer vision, robotics, and Natural Language Processing(NLP) in healthcare diagnostics, is now being very efficiently handled by artificial intelligence [4].

A subdivision of AI known as "deep learning" is focused on building artificial neural networks utilizing human brain models. To find patterns, anticipate outcomes, and gain experience from past mistakes, these neural networks which consist of numerous interconnected layers of neurons (nodes), process and analyze vast amounts of data. The main goal is to simulate how the human brain makes predictions, recognizes patterns, and absorbs knowledge from previous experiences. This methodology has demonstrated remarkable efficacy in addressing complex issues that were previously thought to be unsolvable. [4].

The various layers that comprise these networks enable them to recognize hierarchical patterns and relationships within the data which give deep learning its "deep" reputation. Because of this depth, the models can identify complex patterns that may be difficult for conventional machine-learning techniques to identify. Deep learning's ability to recognize speech and images is one of its ground-breaking innovations.

AI and deep learning have revolutionized a wide range of sectors and industries. Below are a few notable examples:

* Image and speech recognition tasks, such as image categorization, object detection, and identification of facial features, have seen remarkable performance with deep learning models. These inventions have found use in a variety of industries, including safety systems, virtual reality, self-driving automobiles, and imaging for medical purposes.
* Natural Language Processing (NLP) has made it possible for machines to understand, interpret, and produce human language facilitates language-related operations, and improves human-machine communication. It is used in personal voice assistants, language translation, sentiment evaluation, chatbots, and data retrieval systems [8].
* Healthcare: Deep learning and AI have greatly enhanced it. They assist in the detection of disease, new medication discovery, picture analysis for medical purposes, customized therapy, planning individualized treatments, and surveillance of patients. These developments contribute to the improvement of the accuracy, efficacy, and accessibility in the field of medical care.
* Trading and finance: In the world of finance and trade the importance of deep learning beams through as AI algorithms automate trading strategies, predict market movements, and evaluate many financial datasets. Deep learning provides critical analysis in identifying fraud, risk evaluation, credit scoring, and effective portfolio management, offering valuable insights for making decisions and reducing human error.
* AI-driven autonomous systems are evolving across multiple domains, including robotics, drones, self-driving cars, and smart home appliances. These systems utilize deep learning models to interact with their surroundings, which allow autonomous navigation and decision-making.

These examples show how AI and Deep learning have a great deal of promise to solve problems in various real-world scenarios. Due to continuing research and tech advancement, future uses and societal influence may be very different [5].

# LITERATURE REVIEW

i. L. Kong and J. Cheng, “Based on improved deep convolutional neural network model pneumonia image classification,” *PLOS ONE*, vol. 16, no. 11, p. e0258804, Nov. 2021, doi: 10.1371/journal.pone.0258804. Available: <https://doi.org/10.1371/journal.pone.0258804>

Advances in pneumonia image categorization were examined in the research carried out by Kong and Cheng (2021) using enhanced deep convolutional neural network (CNN) models. According to their paper published in the PLOS ONE journal, the researchers sought to improve the accuracy of pneumonia diagnosis by the application of deep learning methods.

Kong and Cheng's study delves into the intricacies of classifying medical images, with a specific emphasis on chest X-ray images used to diagnose pneumonia. They made changes to the conventional CNN architectures, adding new functions for activation, pooling methods, and streamlined convolutional layers. The suggested deep CNN was made more robust by utilizing pre-trained models using methods based on transfer learning.

The study's evaluation of enhanced deep CNN models for pneumonia image classification demonstrated remarkable capability. A significant decrease in misclassification rates was evident from the results, which outperformed standards. With its demonstration of the potentially revolutionary effect of Deep Learning and artificial intelligence (AI) in pneumonia detection and its preparation for future advancements in healthcare applications, this work represents a major achievement in medical imaging.

ii. C. Szegedy *et al.*, “Going Deeper with Convolutions,” *arXiv (Cornell University)*, Sep. 2014, doi: 10.48550/arxiv.1409.4842. Available: <https://arxiv.org/abs/1409.4842>

The study’s specific goal was to improve the accuracy of image recognition, specifically using the ImageNet dataset Motivated by VGGNet's achievements, the authors aimed to tackle the difficulties associated with training incredibly complex networks. To address the problem of gradients disappearing, residual connections were included, which permit information to pass across shortcut links. This breakthrough greatly increased accuracy and made it easier to train networks with much deeper layers.

The study examined the significant influence of residual connections on network efficiency and presented experimental findings that showed increased accuracy over earlier designs. By using these connections, a 16-layer network called VGG16 was successfully trained and was able to produce cutting-edge results on ImageNet classification tests.

The results emphasized the role residual connections play in promoting the creation of deeper and more efficient CNNs. In addition to advancing the field of image classification, Szegedy et al.'s work had a significant impact on later research, encouraging the investigation of new architectures and deepening our grasp of the possibilities of CNN.

Building on VGG16, Szegedy et al.'s study adds to the ongoing investigation of deeper architectures and emphasizes the significance of creative connections in neural networks. Their approach deviates from traditional layer stacking, and because of that, the precision of image classification has advanced greatly. This key article inspired more research, demonstrating the growth of CNN architectures and the ongoing quest for enhanced image classification job performance.

# PROBLEM STATEMENT

One of the biggest problems in healthcare is identifying and categorizing pneumonia cases from chest X-ray pictures. Patient outcomes may be jeopardized by the laborious and error-prone manual examination of these pictures, which can delay diagnosis A computerized method that is both precise and effective is needed to classify pneumonia in chest X-rays. A system like this would speed up the diagnostic procedure, allowing for early identification and prompt and precise diagnoses of patients. As a rapid screening method, it can quickly identify patients who are at high risk, helping to stop the spread of the disease. Furthermore, by reducing the workload on medical institutions, this technique can help them manage pneumonia cases more successfully.

Additionally, a thorough examination of chest X-ray pictures using a computerized classification system provides insightful information for research and epidemiological analysis. This improvement in awareness of pneumonia patterns is essential for the advancement of medical expertise and the development of better therapy and management options for pneumonia [3].

# DATA DESCRIPTION

The Chest X-ray Pneumonia dataset, an extensive set of chest X-ray pictures meant to be studied for pneumonia categorization, will be used for our analysis.

The dataset categorizes pneumonia cases into normal and pneumonia cases using JPEG chest X-ray pictures. The dataset includes 1341 'NORMAL' and 3875 'PNEUMONIA' images in the 'train' database, 234 'NORMAL' and 390 'PNEUMONIA' images in the 'test' database, and 8 images in the 'validation' folder. Before any scaling, the image had a single channel (grayscale) and 224x224 pixels. Next, in the preprocessing stage, the images were resized to the same dimensions for the customized CNN model and the VGG16-based pre-trained model [3].

The original source of the data is the Kaggle platform; to access the dataset [3], click this link:

Mooney, P. T. (2022). Chest X-ray Pneumonia Dataset. Kaggle: Your Machine Learning and Data Science Community.

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

A screenshot of a computer

Description automatically generated

Fig.1. Pnemeunia chest X-ray images dataset

Approximately 5,863 pictures were accessible in the Chest X-ray Pneumonia dataset at the time of its release, offering an immense amount of data for study. It is essential to note that revisions can occur after the date of publication, and users are urged to visit the given website to see the most recent size and make-up of the dataset.

For the project, we may need to perform some effective preprocessing to make the best use of the dataset for our project. Methods like rescaling pixel values, splitting, grayscale conversion, data augmentation, validation split, and data feeding can be used to make sure that the classification model performs consistently and optimally, based on the specific needs of the selected model and the level of quality of the images [9].

# METHODOLOGY

Using convolutional neural networks (CNNs) and the VGG16 pre-trained model image classification model, a complete methodology is proposed to handle the urgent problem of recognizing and classifying pneumonia patients from chest X-ray images. Transfer learning, which uses an already trained algorithm to build on its learned features and eliminate the requirement for prolonged training on sparse data, will be the method used for this project.

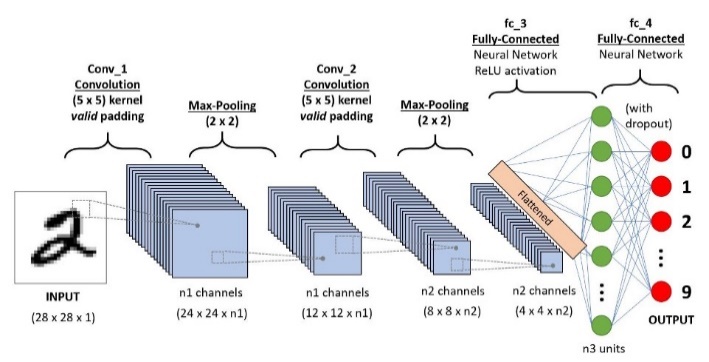
There are a few important steps in the process. After downloading the Chest X-Ray Pneumonia dataset from the given Kaggle link, will first divide it into training, testing, and validation sets before making it available for usage. The images are grayscale and resized to 224x224 pixels. We also incorporate important preprocessing methods such as scaling, standardization, and data augmentation to enhance the CNN model's functionality and generalization to efficiently extract pertinent characteristics from unprocessed image input.

The CNN model will be built upon the VGG16 architecture, which is well known for its ability to classify images. To customize the model for pneumonia categorization, fully connected layers will be modified and the convolutional layers of the pre-trained VGG16 will be kept, because of the use of transfer learning. Additionally, dropout layers will be utilized to lessen overfitting.

The CNN architecture comprises several convolutional layers, preceded by max-pooling layers. The number of filters in the model rises with each consecutive layer (64 and 128) starting with a convolutional layer that has six filters. To minimize spatial dimensions, max pooling comes after each convolutional layer. The last set of fully linked layers consists of a sigmoid activation for binary classification and two dense layers with ReLU activation function. The Adam optimizer is utilized to curtail binary cross-entropy loss during model training. The model's results on the validation set will be continuously monitored, and hyperparameter changes will be made for the best outcomes [7]. After the model has been trained to the point of completion, it is rigorously tested on the assigned test set to determine how well it can diagnose pneumonia cases. To assess the effectiveness of the model, evaluation measures such as accuracy, precision, recall, and F1 score will be carefully computed.

The transfer learning was based on the VGG16 model. VGG16 model employs a 16-layer network, and the pattern is consistent, and the network accepts input in thr form of 224 x 224 x 3 images [10]. The weights of the previously trained VGG16 model, which was trained on ImageNet, were frozen. The frozen VGG16 base was then covered in more layers to create a customized model. A convolutional layer containing 64 filters, a kernel sized at (3, 3), ReLU activation and max pooling were all included in this augmentation. After adding the frozen VGG16 base to the model, it was flattened, then two dense layers with ReLU activation (representing 128 and 64 neurons, respectively) and a last dense layer with a single neuron using a sigmoid activation to classify binary inputs. With the use of the Adam optimizer and binary cross-entropy loss, the model was assembled to develop a thorough architecture suitable for the assigned task.

A thorough grasp of techniques to reduce overfitting was demonstrated by the incorporation of regularization techniques like dropout layers. By using the VGG16 model, this method seeks to precisely categorize cases of pneumonia in chest X-ray images while reducing the need for extensive training on sparse data. The suggested CNN model ought to be utilised in tandem with suitable preprocessing and assessment techniques to improve the treatment of pneumonia patients. This will guarantee optimum accuracy in the recognition and categorization of pneumonia cases [3].

Fig.2. CNN model architecture

A diagram of a software development

Description automatically generated with medium confidence

Fig.3. VGG-16 CNN model architecture layer-wise

# RESULTS AND FINDINGS

To obtain the results, we utilized two models. By combining these two, we managed to determine the model's accuracy, confusion matrix, loss graphs, and underfitting/overfitting status. The GitHub link provides access to the code:

[MecRachel/Artificial\_Intelligence\_DL (github.com)](https://github.com/MecRachel/Artificial_Intelligence_DL)

As shown in Table 1, it can be observed that:

For VGG-16 Model:

Recall, F1-Score, and Precision: In class 0 (normal X-rays), the recall of 0.09 means that only 9% of the actual normal instances were detected, however, the precision of 0.40 means that 40% of the predicted normal cases were accurate. Class 0's F1-score was 0.14, indicating a moderate

balance between recall and precision. The VGG16 model showed a precision of 0.63 for class 1 (pneumonia cases), indicating that 63% of the projected pneumonia cases were accurate. Recall for class 1 was notably higher at 0.92, meaning that 92% of real pneumonia cases were effectively captured by the model. Class 1's F1 score was 0.75 which indicates a strong recall and precision performance.

Accuracy: The model's overall accuracy of 0.61 places it just slightly ahead of random guesswork. However, it is relatively low, which can mean that the model is having a problem identifying the underlying patterns in the data being analysed.

The observed high accuracy on the training set indicates overfitting. On the training set, the model performs well however, the reduced accuracy on the validation set suggests that this may not apply well to new, unseen data. When a model is overly complicated, it can lead to overfitting and learning both relevant patterns and the noise in the training set. After three epochs, the validation accuracy was 93.29% and the training accuracy was 93.94%. Overfitting can be prevented in the VGG16 model by using strategies like weight regularization in custom layers, adding dropout between the dense layers, and utilizing data augmentation to expose the model to a wider variety of instances during training.

From the VGG16 model confusion matrix, it can be observed that there are 214 false positives (FP) and 30 false negatives (FN) noted, along with 20 true negatives (TN) and 360 true positives (TP).

TABLE I.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | Findings | VGG-16 Model Output |  | CNN Model Output | | |  | | |  | | |  |  | | | Graphs of Model Accuracy and Model Loss |  | | |  |  | | |  | | | Confusion  Matrix |  | | | A diagram of a normal and pneumococcal disease  Description automatically generated |  | | | | Class - wise Performance  Report | A number of numbers in a row  Description automatically generated with medium confidence | | A number of numbers in a row  Description automatically generated with medium confidence | | | |  | | | |  | | | |

For CNN:

Precision, Recall, and F1-Score: In class 0 (normal X-rays), the recall of 0.16% indicates that only 16% of the actual normal instances were properly detected by the model. However, the precision of 0.39 implies that 39% of the predicted normal cases were accurate. The F1-core for class 0 was 0.22, suggesting a moderate balance between both recall and precision, like the VGG16 model. The precision of 0.63 for class 1 (pneumonia patients) in the CNN model means that 63% of the predicted pneumonia cases were accurate. With a recall of 0.85 for class 1, the model was able to accurately detect 85% of the actual cases of pneumonia. The F1-score for class 1 was 0.72, which is marginally lower than the VGG16 model but still indicates a strong recall and precision performance.

Accuracy: The model's overall accuracy is 0.59, which is relatively low, indicating that the model has trouble identifying the underlying patterns in the dataset. This decreased accuracy raises the possibility that the CNN model will have trouble correctly identifying X-ray pictures as normal or pneumonia cases.

The CNN model shows signs of overfitting, as evidenced by the significant discrepancy after three epochs between the high accuracy on the training set (97.56%) and the somewhat lower accuracy on the validation set (95.69%). To address overfitting in CNN techniques like adding dropout, weight regularization, and data augmentation on the training dataset by applying arbitrary image manipulations, like flipping, zooming, or rotating can be utilized.

From the CNN model confusion matrix, it can be observed that there are 197 false positives (FP) and 59 false negatives (FN) noted, along with 37 true negatives (TN) and 331 true positives (TP).

Comparison:

The accuracy of the VGG16 model was marginally higher than that of the CNN model, at 0.61. Both models, however, show accuracy levels that are not appreciably higher than chance, which emphasizes the need for additional improvement or different strategies to improve their performance in pneumonia chest X-ray picture classification.

Overfitting is evident in both models, highlighting the necessity of regularization strategies and data augmentation.

In terms of class classification, detecting normal X-rays (Class 0) presents difficulties for both models. While the CNN model demonstrates higher recall for normal X-rays and similar precision for pneumonia cases, the VGG16 model displays better performance in terms of recall for cases of pneumonia (Class 1) compared to CNN.

The disparities between precision and recall are illustrated by the confusion matrices, highlighting the necessity of a well-balanced evaluation metric that is tailored to the specific objectives of the application.

For the detection of pneumonia, VGG16 is better than the CNN model. In comparison to CNN (85%), VGG16 shows a higher recall (92%) for pneumonia cases, suggesting superior sensitivity in identifying genuine positive cases. Furthermore, VGG16 exhibits a balanced performance with moderate precision, which increases its dependability as a pneumonia classification option. Both models require more tuning to perform better, particularly when it comes to detecting normal X-rays.

# CONCLUSION & FUTURE RECOMMENDATIONS

In this study, the Chest X-ray Pneumonia database's chest X-ray images were classified using two models: CNN and VGG16. The study concludes that VGG16 performs better than CNN in identifying instances of pneumonia, exhibiting balanced precision and higher recall (92%). Still, both models need to be improved, especially in the difficult task of categorizing normal X-rays. The results highlight how difficult the endeavor is and how a sophisticated assessment metric that is adapted to the goals of the apps is required.

Utilizing pre-trained weights from transfer learning, the VGG16 model was enhanced, and the CNN architecture was chosen due to its efficacy in image classification tasks. It is recommended that in the future, comprehensive performance evaluations be carried out, data augmentation techniques be used, more dropout layers can be added, batch normalization be added, alternative pre-trained models like Resnet-50, MobileNet, etc. be investigated, ensemble learning techniques be used, real-time applications be created, and research models explaining ability methods be implemented. The models' functionality would be improved by these impending changes, which would also increase the range of possible applications for pneumonia chest X-ray image diagnosis and healthcare support.

The outcomes of this study demonstrate the effective use of VGG16 and CNN models for pneumonia chest X-ray image classification. Nonetheless, additional enhancements are suggested to provide comprehensive analysis and dependable functionality. By implementing the recommended future additions, researchers and healthcare practitioners can benefit from the models' capacity to enhance the precision of pneumonia diagnosis, encourage prompt decision-making, and provide valuable support in the medical field.

# REFERENCES

[1] L. Kong and J. Cheng, “Based on improved deep convolutional neural network model pneumonia image classification,” *PLOS ONE*, vol. 16, no. 11, p. e0258804, Nov. 2021, doi: 10.1371/journal.pone.0258804. Available: <https://doi.org/10.1371/journal.pone.0258804>

[2] C. Szegedy *et al.*, “Going Deeper with Convolutions,” *arXiv (Cornell University)*, Sep. 2014, doi: 10.48550/arxiv.1409.4842. Available: <https://arxiv.org/abs/1409.4842>

[3] “Chest X-Ray images (Pneumonia),” *Kaggle*, Mar. 24, 2018. Available: https://www.kaggle.com/datasets/paultimothymooney/chest-xray-pneumonia/data

[4] “Artificial intelligence (AI),” *New Scientist*. Available: <https://www.newscientist.com/definition/artificial-intelligence-ai/>

[5] M. Chatterjee, “20 Deep learning applications in 2024 across Industries | Great Learning,” *Great Learning Blog: Free Resources What Matters to Shape Your Career!*, Nov. 08, 2023. Available: <https://www.mygreatlearning.com/blog/deep-learning-applications/>

[6] “Fig. 8. VGG-16 CNN model architecture layer wise.,” *ResearchGate*. Available: <https://www.researchgate.net/figure/VGG-16-CNN-model-architecture-layer-wise_fig8_346259768>

[7] A. Rosebrock, “Convolutional neural networks (CNNs) and layer types - PyImageSearch,” *PyImageSearch*, Jun. 08, 2023. Available: <https://pyimagesearch.com/2021/05/14/convolutional-neural-networks-cnns-and-layer-types/>

[8] I. Lauriola, A. Lavelli, and F. Aiolli, “An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools,” *Neurocomputing*, vol. 470, pp. 443–456, Jan. 2022, doi: 10.1016/j.neucom.2021.05.103. Available: <https://doi.org/10.1016/j.neucom.2021.05.103>

[9] M. Patel, “The complete guide to image preprocessing techniques in Python,” *Medium*, Oct. 23, 2023. Available: <https://medium.com/@maahip1304/the-complete-guide-to-image-preprocessing-techniques-in-python-dca30804550c>

[10] D. Guan, “Classical architectures in CNN,” *Deep Learning*, Jul. 27, 2020. Available: <https://guandi1995.github.io/Classical-CNN-architecture/>

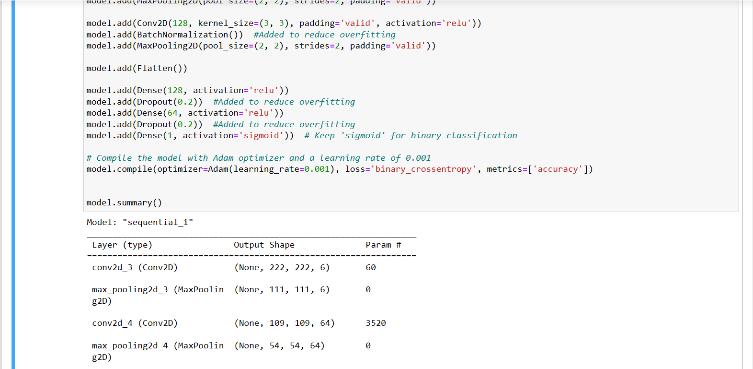
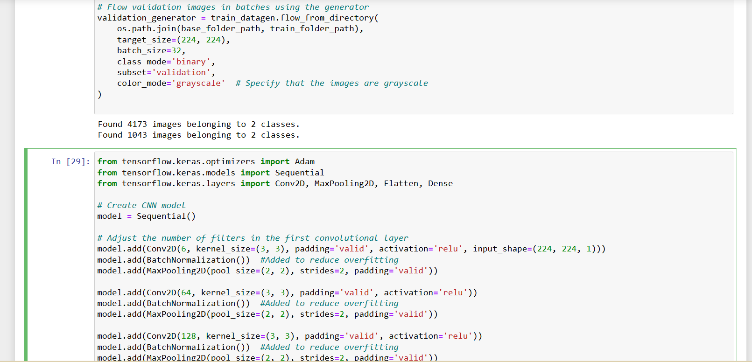
A screenshot of a computer

Description automatically generatedA screenshot of a computer code

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA graph with red and blue lines

Description automatically generatedA screen shot of a graph

Description automatically generatedA screenshot of a computer program

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer program

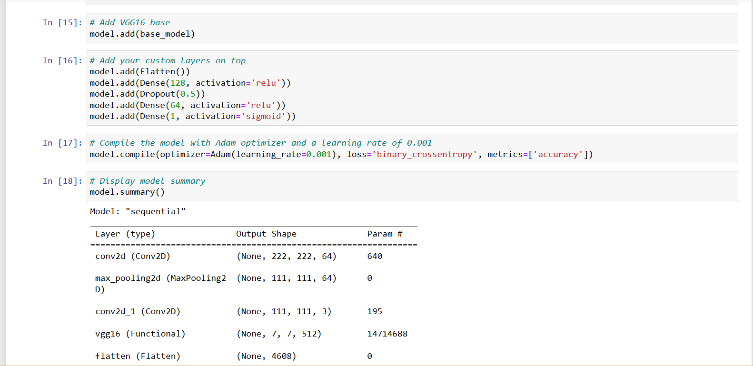
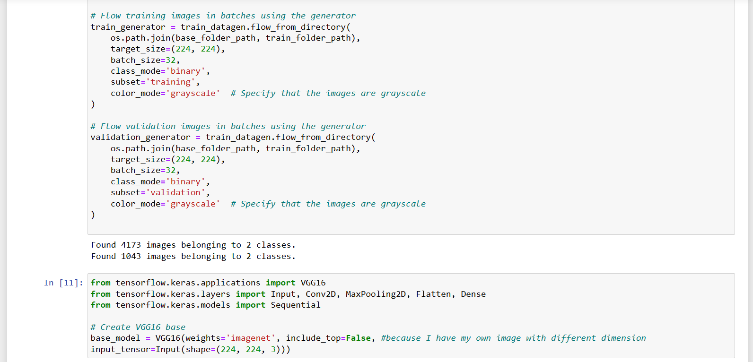
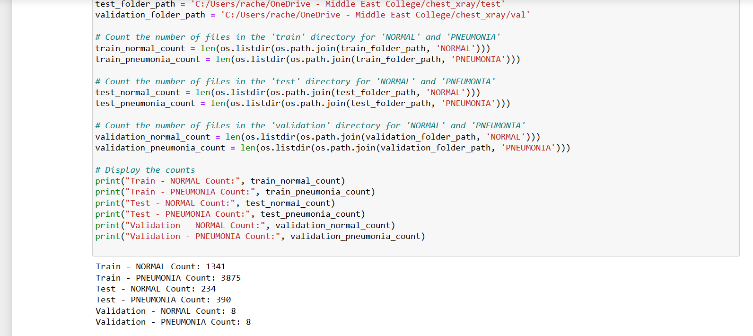
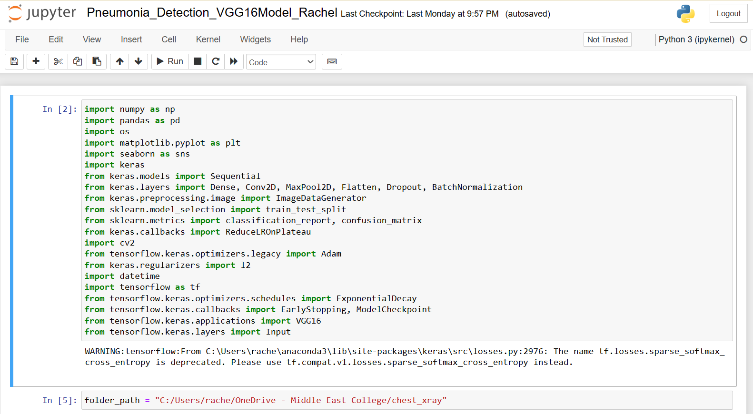
Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA close-up of a computer screen

Description automatically generatedA graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of a graph of

Description automatically generated

A screenshot of a computer

Description automatically generatedA graph with red and blue lines

Description automatically generatedA screen shot of a graph

Description automatically generated A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated A screenshot of a computer screen

Description automatically generated A screenshot of a computer code

Description automatically generatedA comparison of graphs with red and blue lines

Description automatically generated

The GitHub link provides access to the code:

[MecRachel/Artificial\_Intelligence\_DL (github.com)](https://github.com/MecRachel/Artificial_Intelligence_DL)