Creating an Accessible Convolutional Neural Network for Pediatric Pneumonia Diagnosis

Aadith Thiruvallarai
Computer Science Undergraduate Student
1300 Harvey Mitchell Parkway S, Apt. 523, College Station, TX 77840
aaditht@tamu.edu

1 Introduction

The World Health Organization stated in 2019 that "pneumonia account[ed] for 14% of deaths of children under 5 years old" resulting in the death of 740,180 children in 2019 (World Health Organization). Pneumonia has multiple causes: viruses, bacteria, or fungi. When any of these three enter the lungs, the immune system responds and causes inflammation of your alveoli, or air sacs (National Lung Heart and Blood Institute). It is inflammation that causes these air sacs to fill up with pus or liquid causing pneumonia symptoms.

1.1 Types of pneumonia

Bacterial pneumonia is common in adults, with the most common cause in the United States coming from *Streptococcus pneumoniae* (National Lung Heart and Blood Institute).Bacterial pneumonia can also develop after having a cold or the flu. This pneumonia in particular can be treated with antibiotics, but only one-third of children with pneumonia receive the antibiotics they need (World Health Organization).Viral pneumonia in adults is most often caused by the flu (influenza virus) or the common cold (rhinovirus). The most common cause of viral pneumonia in young children is respiratory syncytial virus (RSV). The virus that causes COVID-19, SARS-CoV-2, can also cause pneumonia. Fungi such as *Pneumocystis jirovecii* typically can cause pneumonia in those with weakened immune systems. Babies and children, 2 years old or younger, are considered at high risk for developing pneumonia symptoms because their immune systems are developing.

1.2 Diagnosis and evaluation

In order to diagnose and evaluate pneumonia, chest x-ray exams, CT scans of the lungs, ultrasound scans of the chest, MRI scans of the chest, and needle biopsies of the lung can be used to determine a myriad amount of factors involving an individual's pneumonia (RadiologyInfo.org). Chest x-ray exams in particular allow a doctor to see a patient's lungs, heart, and blood vessels to help determine if they have pneumonia. The radiologist is also looking for infiltrates, white spots in the lungs, that denote an infection. The x-ray exam can also help determine if an individual has any related complications such as abscesses, collection of pus, or pleural effusions, which is a build-up of fluid between the layers of tissue that line the lungs and chest cavity. A study published by the European Respiratory Journal also found that general practitioners (or G.P.'s) tend to overdiagnose pneumonia and that chest x-rays are useful to help prevent this (European Respiratory Journal).

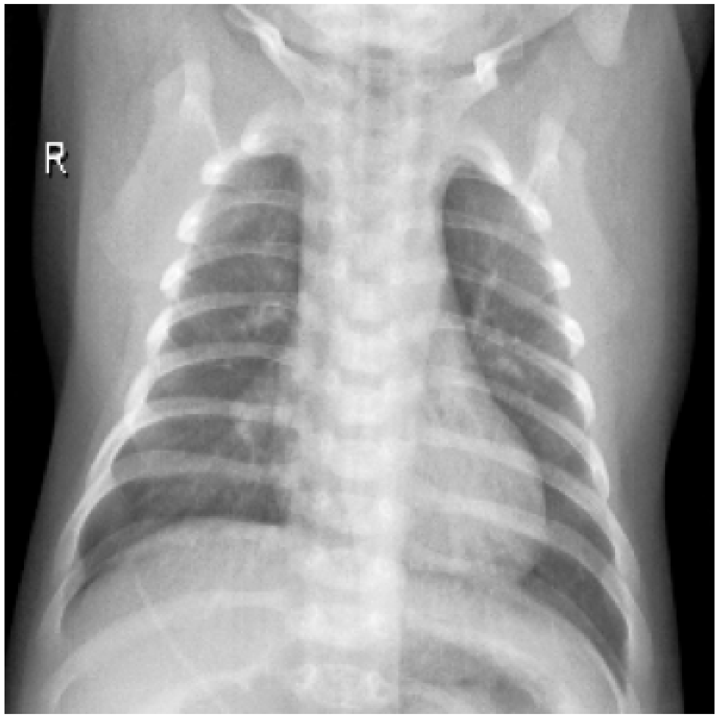

Figure 1: Patient without pneumonia x-ray

Figure 2: Patient with pneumonia x-ray

1.3 Visualization of pneumonia in chest x-rays

It is clear from Figures 1 and 2 that it can be difficult to discern the difference between x-rays of patients with and without pneumonia.

2 Literature review

The problem of diagnosing and evaluating pediatric chest x-rays at a rate better than trained medical professsionals already has solutions from multiple studies that utilize C.N.N.'s (convolutional neural networks). The most notable being from a joint-effort between Jadavpur University in India, the College of IT Convergence in South Korea, and Politechnika Slaska in Poland. The group designed a weighted average ensemble technique involving three convolutional neural network models: GoogLeNet, ResNet-18, and DenseNet-121 (PLOS ONE).

2.1 About the models used

The three models that the weighted average ensemble technique used are important to go over. The GoogLeNet is a C.N.N. with 22 layers and there are two provided pre-trained networks that come with it: ImageNet and Places365. ImageNet can classify images into 1000 classes: these include keyboards, mouses, pencils, animals, etc (MathWorks, GoogLeNet). Places365 is similar to ImageNet and can classify images into 365 classes of place categories: these include fields, parks, runways, lobbies, etc. The ResNet-18 is a C.N.N. with 18 layers and comes with a pre-trained ImageNet, similar to GoogLeNet's ImageNet. DenseNet-121 is a type of DenseNet with 112 convolutional layers (Towards Data Science). DenseNets simplify the connectivity pattern between layers by connecting every layer directly with each other. DenseNets, specifically, are constructed to utilize the network through feature reuse. DenseNets are a potential solution to the vanishing gradient problem that deep neural networks face.

2.2 Combining the models in the weighted average ensemble

The study's authors sought to implement transfer learning where a model already used for a task like image classification tend to do well with biomedical image classification. The ImageNet, with the architecture to identify between 1,000 classes is a popular choice for such a task. A weighted average ensemble machine learning approach combines predictions from multiple models, where the contribution of each model to the ensemble proportional to its capability (Machine Learning Mastery). The study employed a novel approach to calculate the weights: the precision, recall, f1-score, and area under the curve were fused to form the weight vector. The model reached accuracy rates of 98.81% and 86.85% and sensitivity rates of 98.80% and 87.02%, on two different datasets respectively. The developed model does exceedingly well with an AUC-ROC score ranging between 0.978 - 0.988 across 5 folds and a F1 score ranging from 98.03% to 99.66% across 5 folds.

3 Problem formulation

The problem of such a weighted average technique is the portability and usability for all individuals to use, which is the purpose of this study. The World Health Organization's data suggests that countries in the Global South are among those most affected by cases of pediatric pneumonia. This makes the accessibility of the model incredibly important.

4 Proposed Solution

With the context of this problem in mind and the development of previous models, it is proposed that a C.N.N. be used as the model architecture as a means of diagnosis. A C.N.N. is equipped to handle problems such as image classification because it utilizes convolutional and pooling methods where it evaluates local regions of each image. A C.N.N. can also be made to produce a picture that describes what it is looking for. Thus, there will be multiple methods by which we can evaluate the C.N.N. that is produced: precision, recall, accuracy, AUC-ROC, and AUC-PR. The baseline method by which the proposed approach can be evaluated will be based on the "Diagnostic Errors Are Common in Acute Pediatric Respiratory Disease: A Prospective, Single-Blinded Multicenter Diagnostic Accuracy Study in Australian Emergency Departments" where the precision score of professionally trained doctors is relatively low.

5 Data description

The data set used is the Pediatric Pneumonia Chest X-ray dataset, provided by LARXEL on Kaggle. Daniel Kermany, Kang Zhang, and Michael Goldbaum accumulated the data (http://dx.doi.org/10.17632/rscbjbr9sj.2) The dataset consists of a training set of 5,856 examples, each example is a grayscale image, associated with a label from 2 classes.

5.1 Data

5.1.1 File Organization

The data is split into two directories (pneumonia and normal). Each image's dimensions are different, so a decision was made to resize the image dimensions to a width and height of 256 pixels. This will be a potential hyperparameter to consider, because this affects the constructed neural network. Each pixel has a single pixel-value associated with it, indicating the lightness or darkness of that pixel, with higher numbers meaning darker. This pixel-value is an integer between 0 and 255.

5.1.2 Training and validation data

The images are split into a training dataset of 5585 images and a validation dataset of 1171 images.

5.1.3 Batches

Each epoch, or training iteration, takes in 32 images at a time.

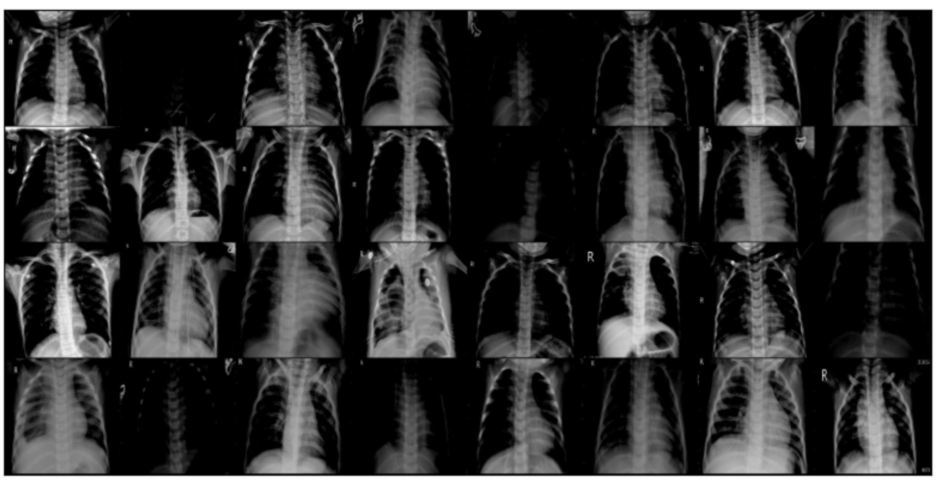

Figure 3: Batch of size 32

5.2 Labels

Each training and test example is assigned to one of the following labels:

Table 1: Labels

Name	Description
0	Normal
1	Pneumonia

6 Results

6.1 Old model

The first model created had three convolutional layers and had 100,000,000 parameters. Because it was incredibly large, it took too long to train and immediately overfit within the first and second epochs. Thus, dropout was intentionally added to increase the robustness of the model and help it generalize more to the training data. This did not change how the model overfitted and it was decided that the overall dataset was too small for how complex the model was.

```
==========================================================================================
Layer (type:depth-idx)                   Output Shape              Param #
==========================================================================================
ImageClassifierNet                       [32, 2]                   --
├─Conv2d: 1-1                            [32, 32, 254, 254]        320
├─ReLU: 1-2                              [32, 32, 254, 254]        --
├─MaxPool2d: 1-3                         [32, 32, 127, 127]        --
├─Conv2d: 1-4                            [32, 64, 125, 125]        18,496
├─ReLU: 1-5                              [32, 64, 125, 125]        --
├─MaxPool2d: 1-6                         [32, 64, 62, 62]          --
├─Conv2d: 1-7                            [32, 128, 60, 60]         73,856
├─ReLU: 1-8                              [32, 128, 60, 60]         --
├─MaxPool2d: 1-9                         [32, 128, 30, 30]         --
├─Dropout: 1-10                          [32, 128, 30, 30]         --
├─Flatten: 1-11                          [32, 115200]              --
├─Linear: 1-12                           [32, 1024]                117,965,824
├─ReLU: 1-13                             [32, 1024]                --
├─Linear: 1-14                           [32, 2]                   2,050
==========================================================================================
Total params: 118,060,546
Trainable params: 118,060,546
Non-trainable params: 0
Total mult-adds (G): 22.19
==========================================================================================
Input size (MB): 8.39
Forward/backward pass size (MB): 902.74
Params size (MB): 472.24
Estimated Total Size (MB): 1383.37
==========================================================================================
```

Figure 4: Summary of old model

6.2 New model

A new model was created with two convolutional layers, less overall features, and a dropout layer. The idea was that creating a less complex data would work better with the size of the overall dataset.

```
==========================================================================================
Layer (type:depth-idx)                   Output Shape              Param #
==========================================================================================
ImageClassifierNet_v2                    [32, 2]                   --
├─Conv2d: 1-1                            [32, 16, 254, 254]        160
├─ReLU: 1-2                              [32, 16, 254, 254]        --
├─MaxPool2d: 1-3                         [32, 16, 127, 127]        --
├─Conv2d: 1-4                            [32, 32, 125, 125]        4,640
├─ReLU: 1-5                              [32, 32, 125, 125]        --
├─MaxPool2d: 1-6                         [32, 32, 62, 62]          --
├─Dropout: 1-7                           [32, 32, 62, 62]          --
├─Flatten: 1-8                           [32, 123008]              --
├─Linear: 1-9                            [32, 256]                 31,490,304
├─ReLU: 1-10                             [32, 256]                 --
├─Linear: 1-11                           [32, 2]                   514
==========================================================================================
Total params: 31,495,618
Trainable params: 31,495,618
Non-trainable params: 0
Total mult-adds (G): 3.66
==========================================================================================
Input size (MB): 8.39
Forward/backward pass size (MB): 392.32
Params size (MB): 125.98
Estimated Total Size (MB): 526.69
==========================================================================================
```

Figure 5: Summary of new model

6.2.1 New model metrics

The model overfitted to the training dataset by the second and third epochs based on the accuracy and loss graphs.

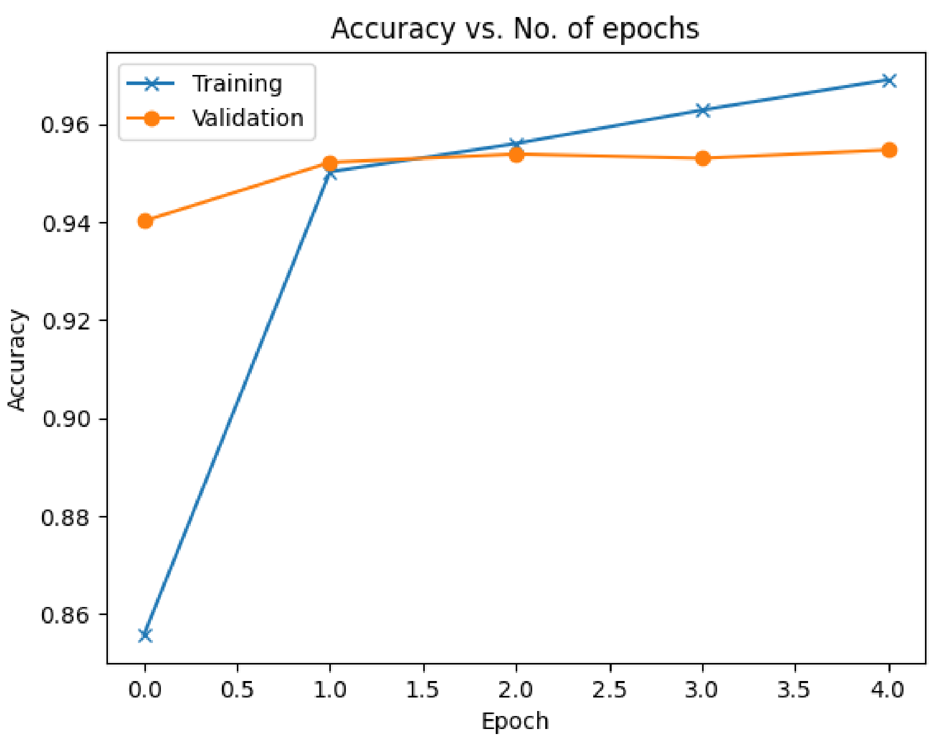

Figure 6: Accuracy of each epoch

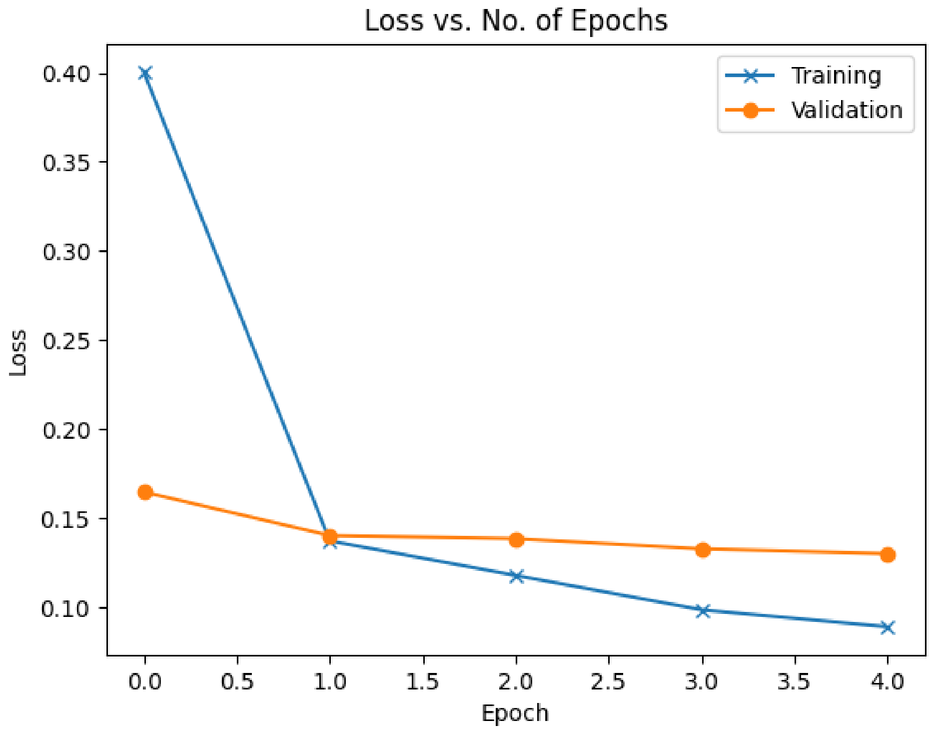

Figure 7: Loss of each epoch

Despite the overfitting of the model, the Precision-Recall AUC is .993 and the ROC-AUC is .985.

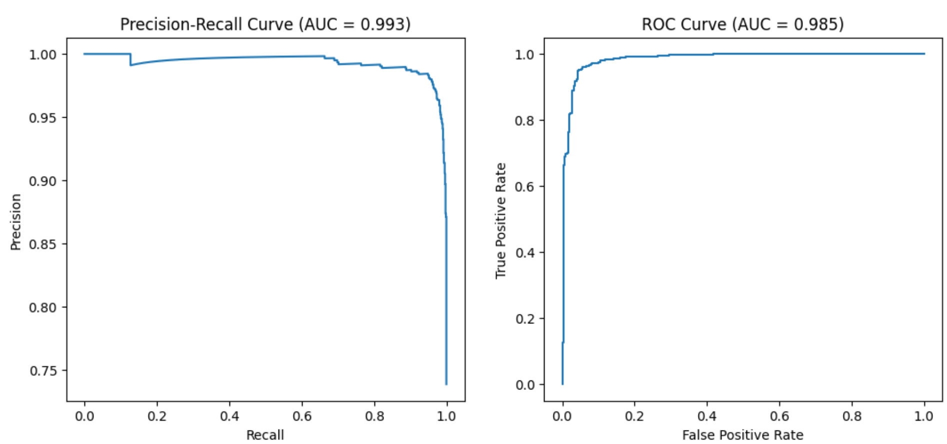

Figure 8: Receiving operator characteristic curve

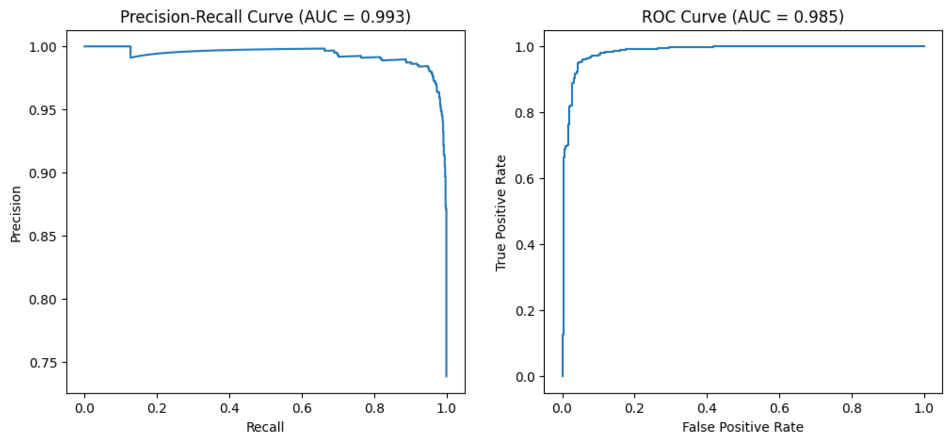

Figure 9: Precision-recall curve

7 Limitations and future directions

7.1 Data

In terms of data, the assumption was made that the dataset is representative of all chest x-rays. It is suspected that this isn't the case and in a future extension of the research, more data should be collected to create a better generalizable model.

7.2 Model

The architecture of a neural network is more closely related to the nature of the data than previously assumed, as further elaborated below.

7.2.1 Architecture

The model architecture is more closely related to the size of the data than was originally assumed. The models created seemed to be too complex for the size of the dataset and even the nature of the problem. It was assumed originally that because successful models used incredibly complex models, a complex model was required, which wasn't the case at all.

7.2.2 Training

The C.N.N.'s training is heavily dependent on the architecture of the device that is being used. A huge limitation of the study was inaccessibility to devices with specialized GPU's for neural network training and image processing. The time it took for tuning the model was further lengthened because of the long time it took to train the model once. A future extension of the research would utilize high performing GPU's in order to train the network faster and better tune hyperparameters.

8 Conclusions

It is safe to conclude that convolutional neural networks not be made incredibly complex to be successful. It is important to emphasize that the size of the dataset and the complexity of the model are closely related and must be properly tuned for in future extensions of the study. Hypertuning of parameters and proper experimental appraisal of model variations can be done with better resources as well.

References

[1] Brownlee, Jason. “How to Develop a Weighted Average Ensemble with Python.” MachineLearningMastery.com, Machine Learning Mastery, 7 May 2021, https://machinelearningmastery.com/weighted-average-ensemble-with-python/#: :text=Weighted%20average%20or%20weighted%20sum%20ensemble%20is%20an %20ensemble%20machine,related%20to%20the%20voting%20ensemble.

[2] “Causes and Risk Factors.” National Heart Lung and Blood Institute, 24 Mar. 2022, www.nhlbi.nih.gov/health/pneumonia/causes.

[3]“GoogLeNet Convolutional Neural Network.” GoogLeNet Convolutional Neural Network - MATLAB, MathWorks, 2017, https://www.mathworks.com/help/deeplearning/ref/googlenet.html.

[4] Kundu, Rohit, et al. “Pneumonia Detection in Chest X-Ray Images Using an Ensemble of Deep Learning Models.” PLOS ONE, vol. 16, no. 9, 2021, https://doi.org/10.1371/journal.pone.0256630.

[5] “Pneumonia in Children.” World Health Organization, 11 Nov. 2022, www.who.int/news-room/fact-sheets/detail/pneumonia.

[6] Radiological Society of North America (RSNA) and American College of Radiology (ACR). “Pneumonia.” Radiologyinfo.Org, 8 Feb. 2021, www.radiologyinfo.org/en/info/pneumonia.

[7] “ResNet-18 Convolutional Neural Network.” ResNet-18 Convolutional Neural Network - MATLAB, MathWorks, 2018, https://www.mathworks.com/help/deeplearning/ref/resnet18.html.

[8] Sarkar, Arjun. “Creating DenseNet 121 with TensorFlow.” Medium, Towards Data Science, 11 July 2020, https://towardsdatascience.com/creating-densenet-121-with-tensorflow-edbc08a956d8#: :text=DenseNet%20is%20a%20convolutional%20neural,4th%2C%205th%20and%20so%20on.

[9] Speets, A. M., et al. “Chest Radiography and Pneumonia in Primary Care: Diagnostic Yield and Consequences for Patient Management.” European Respiratory Journal, vol. 28, no. 5, 2006, pp. 933–938., https://doi.org/10.1183/09031936.06.00008306.