HIGH SCHOOL EDITION
Journal of Student Research

Volume 12 Issue 3 (2023)

Utilizing Different Deep Learning Models to Accurately Detect Pneumonia in Chest X-ray Images

Hasika Oggi[1] and Rhea Jain[#]

[1]Branham High School
[#]Advisor

ABSTRACT

This study investigates how to diagnose pneumonia at early stages by using an AI model. Previous research has shown that AI is greatly increasing in its use in healthcare. Many doctors now rely on AI to diagnose diseases like cancer, heart disease, and even Alzheimer's. In this study, we use multiple models to test the accuracy of AI when it comes to diagnosing pneumonia through chest X-rays. In accordance with the results from previous studies, AI Is generally quite accurate when it comes to diagnosing health issues, such as pneumonia. However, it is also true that AI has its limitations, and we cannot continually rely on it to make decisions about a patient's wellbeing. Our findings indicate that AI, while it is a useful tool for healthcare, will be ultimately dependent upon the integrity of the data the model has been trained on.

Introduction

To understand what pneumonia really is and what is the purpose of the model, first one needs to understand the basics of the respiratory system. The lungs are the most essential part of the respiratory system. They are responsible for bringing oxygen into the body and releasing carbon dioxide in every inhale and exhale. Inside the lungs are little air sacs called alveoli [1]. Alveoli are where the lungs and the blood exchange oxygen and carbon dioxide during the process of breathing in and breathing out. However, these alveoli can become infected if an outside pathogen finds its way into the alveoli. As a result, the alveoli fill up with either fluid or pus, making it harder for the bloodstream to absorb the oxygen that it needs. This is the pathogenesis of pneumonia. There are different types of pneumonia (viral, bacterial, fungal, atypical, etc.). Symptoms include cough with phlegm or pus, fever, chills, rapid pulse, loss of appetite, and difficulty breathing [2]. One may also experience sharp pains in their chest or be extremely tired and have low energy throughout the day. There are many complications from contracting pneumonia, but they vary from age to age. People who are extremely young or over 65 years have a higher risk of being in a critical condition if they contract pneumonia. Common complications include bacteria in the bloodstream, sepsis (a life-threatening complication of an infection), pleural effusion, lung abscess (pus-filled holes in the lungs), and/or difficulty breathing [3]. Pneumonia can even cause death if it's not treated in time, especially if there's an underlying health condition. Fortunately, pneumonia (if diagnosed at an early stage) can be treated with a variety of antibiotics to cure the patient. Some forms of pneumonia can even be "prevented" with vaccines. However, there must first be a medical diagnosis before implementing any prescribed treatments because there are multiple symptoms of pneumonia that make it very easy to mistake it for the common cold or the flu. A diagnosis of pneumonia is most confirmed through multiple chest X-rays, although tests like a blood test can also be done to see if one's immune system is fighting an infection. However, X-rays emit an intense amount of radiation which could potentially harm the patient, damage their DNA, and lead to cancer later. This leads to the imminent question: how can one diagnose pneumonia in a potentially more innocuous way?

Method 1:

To begin our research, we first did an in-depth study on the different types of AI models. After some thought, we decided that our baseline models would be a logistic regression model and K-nearest neighbors (KNN) model.

With AI, learning is always about "labeling" or predicting some unknown value for a piece of data. Either the value will be a classification (into a category) or a regression (the value is a number). When we have some data and plot it on a graph, a best-fit line must be drawn in order to make a prediction about the data and test the fit. There are two types of way that one can compile this data. One is through Mean Absolute Error. This is found through averaging the distances of each point to the best-fit line (absolute value of residuals [4]). Another way to compile the data is through Mean Squared Error. If the scientist would like to heavily penalize large error, they can square the residuals before averaging them. After deciding the metric for a good or bad fit (mean absolute or mean squared error), we perform linear regression by solving an equation to find the line that minimizes this error. To solve for the line of best fit, you must utilize the form $\hat{y}=mx+b$, where "m" is the slope, "b" is the y-intercept, " $\hat{y}$" is the prediction, "x" is the input, and "y" is the true output. $(y-\hat{y})$ is the residual and $(y-\hat{y})^2$ is the squared residual. To find the Mean Squared Error, average $(y-\hat{y})^2$ across all data points and use linear algebra or calculus to find the line with the smallest mean squared error. To figure out the accuracy of the model, the coefficient of determination, R^2, describes how much of the variation in the data is explained/captured by the line (from 0 to 1). The higher the value, the better the fit. This is how a logistic regression model is created.

A KNN model is an approach to data classification that estimates the likelihood of a data point being in one set based on its distance from the group of points nearest to it [5]. This kind of algorithm does not generate a model of the data set beforehand; rather, it only makes calculations when it is asked to poll the data point's neighbors. This model sets up an imaginary boundary between classes and cannot be used for outliers, as they will not be placed into a group. It is typically used as classification algorithm, working off the assumption that similar points can be found close to each other [6], though it can also be used for regression problems.

When we compared the results between the logistic regression model and KNN model, we found that the logistic regression model had an accuracy of 67.25% while the KNN model had an accuracy of 70.25%. Thus, we concluded that KNN achieves higher accuracy than logistic regression and it also works for non-linear data. Logistic Regression is bounded mainly by linear boundaries. However, the accuracy of the KNN model depends on how long it is trained for - it takes much longer to train a KNN model than a logistic regression model.

METHOD 2:

After we received our results from our logistic regression model and KNN models, we decided to try a more accurate model, using dense and convolutional neural networks.

Simple classification algorithms like those mentioned above (logistic regression) often don't work well for computer vision. Computer vision is an AI field that enables "computers and systems to derive meaningful information from digital images, videos and other visual inputs — and take actions or make recommendations based on that information" [7]. For computer vision, we need to use neural networks since they can recognize complex patterns. This kind of algorithm was slightly inspired by the structure of the brain and how the neurons interact with one another. The neurons are stacked into layers of different sizes, thus leading to an organized neural network. Deep learning is the presence of multiple layers in the network.

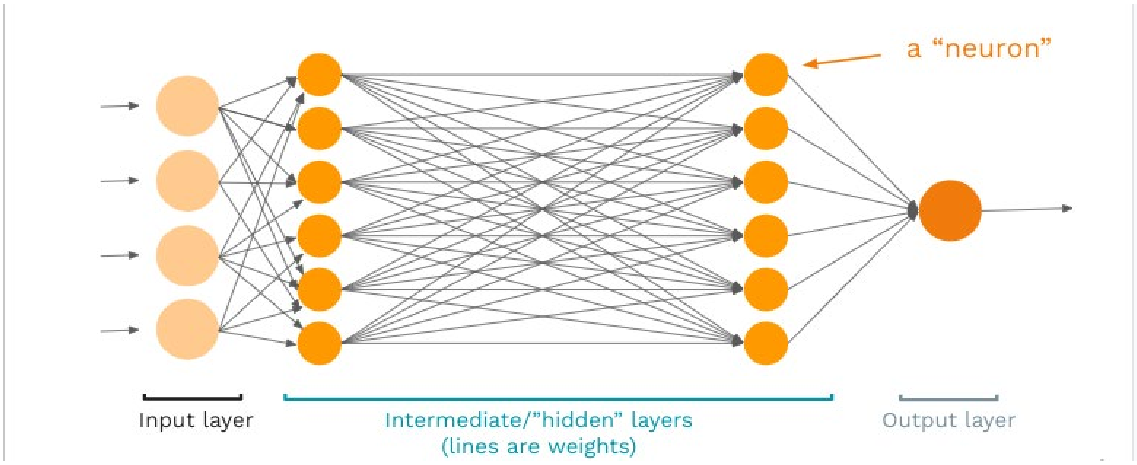


Figure 1. Depicts the basic structure of a neural network, with the input, output and multiple hidden layers.

A dense neural network is when all neurons in a single layer are very closely connected to the neurons of the layer beneath and above it. The dense layer's neuron in a model receives output from every neuron of its preceding layer. This is the layer that is most commonly used in AI neural networks. Dense neural networks are used for language translation, image searching, and detection of cells.

Convolutional neural networks (CNN) are good at finding patterns in images, similar to dense neural networks. However, this kind of network is able to recognize patterns anywhere, all at once. There are 3 tools used in this process: convolution (a mathematical operation that lets us find a pattern in a certain portion of an image) and kernel (the pattern we're looking for, represented as numbers). To perform a convolution, you need to multiple each element of an image tile by the corresponding element in a kernel (pattern). Different kernels represent different patterns, so these can alter the input. Finally, a convolutional layer: a neural network layer that lets us find the pattern anywhere at once. In this algorithm, we apply the convolution to each tile in the image at once, with the same kernel.

A basic CNN architecture involves the input, convolutional layer, pooling layer, fully connected layer, and finally the output. A pooling layer is one that divides the output into tiles and takes the maximum of each. This means there's a smaller model, so the computation is faster and there is less overfitting. A fully connected layer connects every input neuron to every output neuron. To adapt CNN for more complex problems, we need to add more layers as each layer learns a more complex shape. Compared to logistic regression, KNN, and dense neural networks, the main advantage of CNN is that it can automatically detect the important features of a photo without any human supervision. For example, given a group of photos with fish and dogs, the model will learn the distinctive features of each class by itself. This means that this kind of algorithm would be an ideal solution to computer vision and image classification problems.

When we compared the results between the dense neural network and CNN models, we found that the dense neural network had an accuracy of 72% with 100 epochs and 1 hidden layer. The CNN model had an accuracy of 78%. The number of epochs in a neural network is the number of times that the algorithm that's being trained will work through the given data set. A hidden layer is "a layer in between input layers and output layers, where artificial neurons take in a set of weighted inputs and produce an output through an activation function" [8]. Thus, we concluded that CNN achieves higher accuracy than dense neural network, especially for images.

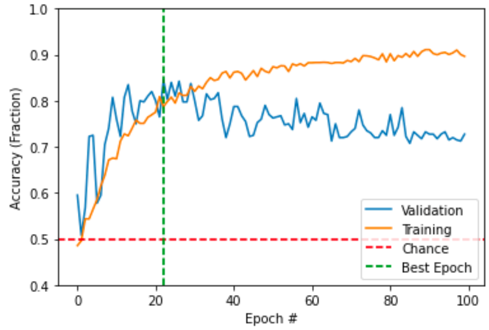


Figure 2. Depicts the Dense neural network: 72% with 100 epochs and 1 hidden layer.

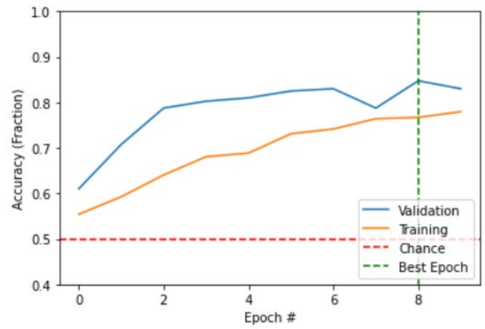


Figure 3. Depicts CNN: 78% with 100 epochs and 1 hidden layer.

Data Augmentation

After testing all these models, we decided to try a new approach: data augmentation. Data Augmentation in machine learning is training a model by altering training data in different ways to better familiarize the model with different situations images, features, and characteristics. This was done by changing the rotation (angle) of the image or changing the color (removing or altering color) of the image.

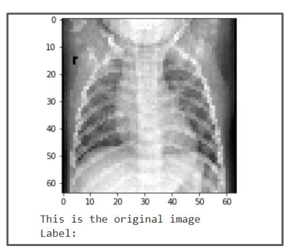
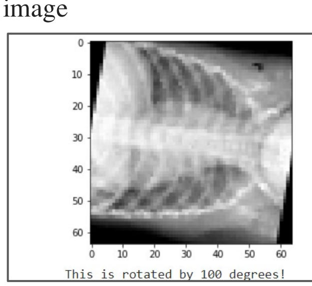
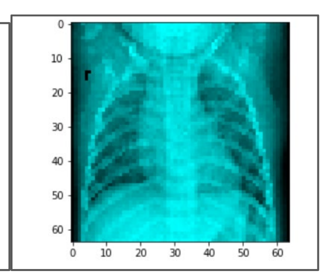


Figure 4 depicts the original image





Figures 5 and 6 depict the modified versions. Figure 5 is rotated, Figure 6 had color added

We found that rotation greatly increased the accuracy, but color did not significantly change the accuracy of the model. This was likely because rotation would allow the model to quickly identify any discrepancies vs. when the x-ray is in its normal position. In the end, our augmentation model/dataset consisted of: 5 Epochs (# of times the model runs through the data), + 2 hidden layers Rotation by 100° Degrees, scaled up by 1.5x, flipped upside down by 0.5x.

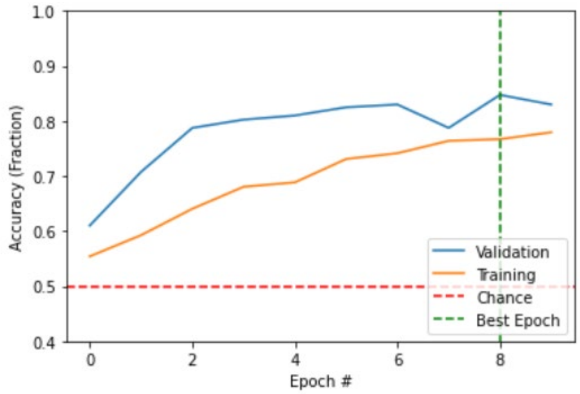


Figure 7 shows a graph of the original CNN model with 78% accuracy

```
[ ] plot_acc(cnn_history)
```

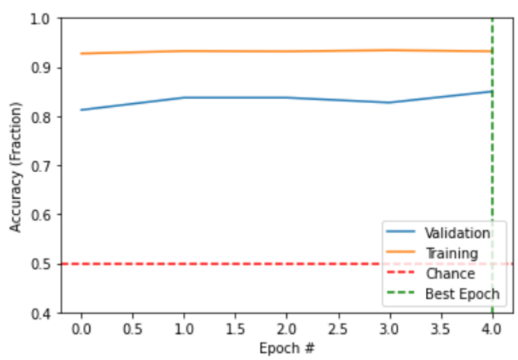


Figure 8 shows a graph of the model trained with augmented data that had 86.5% accuracy

As shown in the graphs above, test accuracy of the pneumonia detection increased by nearly 10% when the model was trained with augmented data. Ultimately, data augmentation can be extremely helpful in increasing the accuracy of the predictions of a model because it can training the model with augmented data means that one can simulate different x- rays the model may come across.

Limitations

Of course, there are limitations to the capabilities of AI. DL (Deep Learning) can pinpoint key aspects through self-supervised learning, and thus requires exceptionally large amounts of data dependent on the input-data representation. DL models also sometimes struggle with orientation and position of images, though there are methods to combat this using data augmentation and/or transfer learning [9]. Such models will only recognize given images as one of its input categories, which means that showing it an unknown group would result in erroneous classification. Additionally, human bias in the datasets given to a model can perpetuate discrimination. For example, Amazon's recruitment tool proved to be prejudiced against women, having been trained previously on only former hires and reflecting the bias of a male-dominated industry [10]. This makes it of utmost importance that any widely implemented use of AI is trained with data pertaining to a vast number of categories. However, this is not to say that AI cannot aid in logically and arithmetically processing data in ways that could succor current health fields. In one study, an AI model did remarkably well, successfully identifying patients with pneumonia as suspected to have COVID-19 with an accuracy of 95.4% [11]. AI even outperformed humans in identifying breast cancer in ultrasounds, with the average of 10 radiologists being 0.924 ± 0.02 and the model having an AUROC of 0.962. The AI was trained to recognize several different types of benign and malignant conditions and could assign both labels to images that had both kinds of input [12]. These are cases of well-trained models, and they could certainly be beneficial to the future of healthcare.

Conclusion

The efficacy of AI in healthcare is dependent upon the integrity of the data it has been trained on. It is important that human bias does not unintentionally exacerbate favoritism, especially considering that healthcare is a field that can directly deal with the mortality of human life. Incorrect identification could lead to a person receiving no treatment for an ailment otherwise easily remedied and their condition deteriorating as a result, or someone who didn't require treatment being over-medicated and having additional health complications and unnecessary finances. However, that doesn't mean it's not worth it - using AI in healthcare could greatly improve the efficiency of our current systems, automating a lot of the more menial tasks (like managing patients and medical resources) that consume healthcare professionals' time and aid in recognizing time-sensitive injuries or diseases. With careful consideration, AI provides valuable services to the field of healthcare.

Acknowledgements

The authors would like to thank Inspirit AI and their help in providing sources of valuable information.

References

[1] NCI dictionary of cancer terms. (n.d.). National Cancer Institute. Retrieved May 18, 2023, from https://www.cancer.gov/publications/dictionaries/cancer-terms/def/alveoli
[2] Pneumonia. (n.d.). Johns Hopkins Medicine. Retrieved May 18, 2023, from https://www.hopkinsmedicine.org/health/conditions-and-diseases/pneumonia
[3] Pneumonia: Causes, symptoms, diagnosis & treatment. (n.d.). Cleveland Clinic. Retrieved May 18, 2023, from https://my.clevelandclinic.org/health/diseases/4471-pneumonia
[4] Numeracy, maths and statistics. (n.d.). Academic Skills Kit. Retrieved May 18, 2023, from https://www.ncl.ac.uk/webtemplate/ask-assets/external/maths-resources/statistics/regression-and-correlation/residuals.html
[5] What is K-Nearest Neighbor (K-NN)? - Definition from Techopedia. (2016, September 13). Techopedia. https://www.techopedia.com/definition/32066/k-nearest-neighbor-k-nn
[6] What is the k-nearest neighbors algorithm? (n.d.). IBM. Retrieved May 18, 2023, from https://www.ibm.com/topics/knn
[7] What is Computer Vision? (n.d.). IBM. Retrieved May 18, 2023, from https://www.ibm.com/topics/computer-vision
[8] What is a Hidden Layer? - Definition from Techopedia. (2018, April 12). Techopedia. https://www.techopedia.com/definition/33264/hidden-layer-neural-networks

[9] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data, 8(1), 1–74. https://doi.org/10.1186/s40537-021-00444-8

[10] Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G

[11] Becker, J., Decker, J. A., Römmele, C., Kahn, M., Messmann, H., Wehler, M., Schwarz, F., Kroencke, T., & Scheurig-Muenkler, C. (2022). Artificial intelligence-based detection of pneumonia in chest radiographs. Diagnostics, 12(6). https://doi.org/10.3390/diagnostics12061465

[12] Shen, Y., Shamout, F. E., Oliver, J. R., Witowski, J., Kannan, K., Park, J., Wu, N., Huddleston, C., Wolfson, S., Millet, A., Ehrenpreis, R., Awal, D., Tyma, C., Samreen, N., Gao, Y., Chhor, C., Gandhi, S., Lee, C., Kumari-Subaiya, S., … Geras, K. J. (2021). Artificial intelligence system reduces false-positive findings in the interpretation of breast ultrasound exams. Nature Communications, 12(1), 1–13. https://doi.org/10.1038/s41467-021-26023-2