Pneumonia diagnosis from chest X-ray images using a neural network, evaluating precision, accuracy and recall.

Anika Mahesh

Machine Learning and Data Science, Summer 2021

Instructor: Guillermo Goldztein

Abstract

This study features a novel solution to diagnose pneumonia using machine learning. Pneumonia is a deadly disease that accounts for 16% of all deaths of children under five years worldwide. Pneumonia is an infection that causes the alveoli sacs inside the lungs to swell with a liquid pus. This disease is deadly, making early diagnosis a priority. Chest-X rays are one way of diagnosing pneumonia; however, they require the presence of a trained specialist, hence this project, aiming to create a quicker and more accessible procedure that integrates an 89% accurate model with a precision of 96% and a recall of 92%. This model, considering its accuracy, precision and recall presents a solution that takes milliseconds to identify pneumonia as opposed to an actual specialist taking minutes. The model is also accessible to all with a computer and scanning equipment, increasing the accessibility.

1 - Introduction

Pneumonia is an infectious disease causing the air sacs of the lungs to be filled with fluid instead of air. This causes fever, chills, cough and difficulty in breathing as symptoms. It can be deadly and causes 16% of all deaths of children under five years worldwide. It is also one of the complications of the COVID-19 viral infection. Pneumonia has numerous other consequences if not treated quickly. Some of these consequences are bacteria accumulating in the bloodstream, spreading to other organs, potentially causing organ failure. Pneumonia can also lead to acute respiratory distress syndrome which is fatal. Hence it is imperative that pneumonia is detected early. If treated quickly pneumonia is usually resolved within two weeks.[10] This could be difficult in many rural areas since most large healthcare institutions with advanced diagnostic facilities are in urban areas. In a recent study, common causes of death in rural Uganda were "malaria, pneumonia, malnutrition, and trauma". [14]

Pneumonia is age dependent; children under 2 years and seniors older than 65 are more likely to contract pneumonia than middle aged people. People already having a respiratory infection or disease also have a higher chance of contracting the infection. Pneumonia can be caused by various infectious agents such as bacteria, viruses, and fungi that affect the respiratory system. These agents could be inhaled from food, waste or another infected carrier's cough droplets. Once a person gets pneumonia the infection progresses in stages. The first stage is congestion, and it occurs within the first 24 hours since contracting the infection. Its symptoms are increased blood flow to the lungs and the lung alveoli or air sacs will start to fill with fluid. The patient will experience deep breathing and cough. Red Hepatization, the second stage of pneumonia, occurs 2-3 days after congestion. It results in the affected lung being fully filled with fluid. This would also result in the alveolar capillaries to bulge from the pressure. Grey Hepatization, the third stage takes place two to three days after red hepatization. During this stage the red blood cells disintegrate, and the lungs will appear greyish brown. The fourth and final stage is recovery, provided if the

patient survived the first three stages, the patient would recover completely from the infection. [7] Pneumonia can be very dangerous during the 2nd and 3rd stage; hence early detection is essential. One common test is the Chest X-ray. It involves taking an x-ray image of a person's lungs and having an experienced radiologist examine the X-ray scan. Inaccessibility to experienced radiologists in some developing countries would result in the delaying of the test, making early detection a pipe dream. Utilizing machine learning techniques to diagnose pneumonia makes the process quicker and more efficient, allowing for early detection to be more possible, hence this project. This project offers a solution made using the Keres machine learning framework and can predict pneumonia with 89% validation accuracy. This is an innovation that applies machine learning techniques to a healthcare field in order to help people get diagnosed quickly and easily. It uses a convolutional neural network to diagnose pneumonia from chest x-rays within milliseconds, making the chest X ray process quick and easy. This was made possible by configuring a neural network to differentiate pneumonic lungs from non-pneumonic lungs. [2]

2 - Background on machine learning

To achieve this result, we used a supervised machine learning neural network. Machine learning is a branch of artificial intelligence which applies to data science to make predictions using a dataset obtained from data collection studies. In the supervised learning framework, the training inputs and the training outputs are used by the neural network to create a model. A set of validation inputs and outputs are then used to validate the accuracy of the model. The neural network consists of different layers which in turn consist of several nodes, each with a different activation function that attempts to model the data using the output from the previous node. There are many different types of neural networks. However, in this study a convolutional neural network was used. Convolutional neural networks are used solely to process images and structured arrays and have a few unique layers.

2.1 - Layers

There are three layers that are used frequently in a convolutional neural network. The convolutional layer, the pooling layer and the fully connected layer [11]. For some complex models, a batch normalization layer is used to make training more efficient. The aim of these layers is to simplify the image into manageable pieces, then to create a model in order to make predictions.

2.1a - Convolutional layer

The convolutional layer utilizes a filter or kernel matrix to filter the input matrix (the image). The kernel is multiplied by an equivalent portion of the image. The resulting matrix's values are added up and the examined portion of the image is shifted by several pixels(stride) horizontally. When horizontal shift is impossible, the examined portion's position is horizontally reset and shifted by the same stride vertically. The output consists of the sums of all the examined portions. The output is usually smaller than the input. However, the output image could be padded with zeroes to resize it by using one of the three types of padding. Valid padding is to not add additional padding. Same padding is to add enough padding for which the output has the same dimensions as the input. Full padding adds a border of zeroes on the output image. The ReLU activation function is usually used in convolutional layers. The kernel input into the convolutional layer can be tailored to the problem

as it can pronounce certain 'important' features and suppress those considered 'noise' while also configuring the matrix for analysis by the fully connected layer[11].

2.1b - Pooling Layer

The pooling layer is similar to the convolutional layer, as it helps filter the image using a constant stride. However, pooling layers are used solely to filter the image to reduce image size and complexity. There are two types of pooling layers, the max pooling layer and the average pooling layer. The max pooling layer takes the maximum of an examined portion of the image and adds it to an output matrix. The examined portion is shifted in a way identical to the convolutional layer. The average pooling layer is like the max pooling layer, except it takes the average of the examined area instead of picking the maximum. The pooling layers are essential for the simplification of the image and to decrease image size.

2.1c - Fully connected layer

While the convolutional and the pooling layer are used to reduce the image to a probability distribution (a one-dimensional matrix), the fully connected layer analyses this distribution. The fully connected layer connects previous layers to the output layer. It does the main job of predicting the outcome of the model through binary classification and usually uses the softmax activation function. The fully connected layer produces the probability of the output being one as opposed to zero. It is an essential part of the model as it does the actual modeling. The fully connected layer is dependent on the pooling and convolutional layer as they reduce the image to a one-dimensional distribution. [3]

2.1d - Batch normalization

Batch normalization is a layer that is solely meant to accelerate the training process. it reduces each element (output matrix of each node) to zero mean and variance. It calculates the variance and the mean of each hidden unit in each layer. The value is scaled using a scaling parameter and shifted using a shift parameter.

2.2 - Activation functions

Most nodes have activation function such as the sigmoid, softmax, ReLU and tanh functions. These activation functions make updating the weights(backpropagation) possible. The gradients of the functions allow for the derivation of the weights and biases.

2.2a - Sigmoid

The sigmoid function is used for binary classification problems, mainly because of its characteristic to have asymptotes at zero and one. The sigmoid function is modeled by the equation:

$$sig(x) = \frac{1}{1 + e^{-x}}$$

the sigmoid function(sig(x)) is used in an activation function in fully connected layers of binary classification problems as it outputs a probability between zero and one.

2.2b - Tanh

The tanh function or the hyperbolic tangent function is like the sigmoid function by having two horizontal asymptotes, except they are at -1 and 1. The tanh function is modeled by this equation:

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

The tanh function is perfect for zero centered inputs and are used commonly in hidden layers. Hidden layers allow for more complex equations to be modeled and to more accurately predict the data.

2.2c - ReLU

The ReLU function or the rectified linear unit represents values less than zero as zero and values greater than zero as the same input. It is modeled by the following equation:

$$ReLU(x) = \max(0, x)$$

The ReLU is used in hidden layers (all layers between the input and the output layer) such as the convolutional layer and the pooling layer. However, it does have a drawback where nodes could 'die', or the node and all connected nodes could only output zero. A solution to this is leaky ReLU which instead multiplies the input by 0.01 if the input is less than 0, this can prevent the node from becoming null.

2.2d - Softmax

The softmax activation function is used frequently for categorical classifications and sometimes for binary classification. The softmax activation function (σ) is represented by the following function:

$$\sigma(x) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}}$$

In this function x is a distribution containing input values for the different categories. This calculation is run multiple times for different elements in x(x_i). K is the number of elements in x. this function is used for the fully connected layers of categorical classification problems and sometimes for binary classification problems.

Activation functions are imperative in the modeling process as they provide a baseline equation for the neural network to work with and allow the function to be tailored towards the model.

3 - Related Work

The chest X-ray is a test used to diagnose chest diseases such as heart failure, emphysema, lung cancer, and pneumonia. The chest x-ray procedure involves imaging the chest. To do this, the patient's chest is exposed to a small dose of ionizing radiation. Radiation is emitted in different concentrations and this feature is used to create an X-ray image of the chest. Several attempts have been made in the past to create a computer model that diagnoses pneumonia patients accurately using chest-X-rays. One study [5] done by Tatiana Gabruseva, Dmytro Poplavskiy and Alexandr Kalinin built a deep learning model for pneumonia detection. This used a s deep learning network validated using the precision value. The model averaged the precision and probability from different models trained using different inputs (sections of the dataset). Another study [9] by Luka Račić, Tomo Popovic, Stevan Cakic and Stevan Šandi used Chest-Xray images to make a model that was 90% accurate in the training set, however overfitting could be a possibility in their model. In this study, all three metrics: precision, accuracy and recall have been considered. All three metrics are important for proper analysis as precision and recall allow insight into how many pneumonia patients are accurately diagnosed by the model.

4 - Methods
4.1 - Preprocessing

A large image dataset consisting of chest x-ray images was used. An example of these chest X-ray images is shown in figure 4.0. This data was from a 2018 study that collected chest X-ray images from pneumonia patients [8] and normal patients. The patients with pneumonia were of two categories, bacterial and viral. The number of images for each category is depicted in figure 4.1.
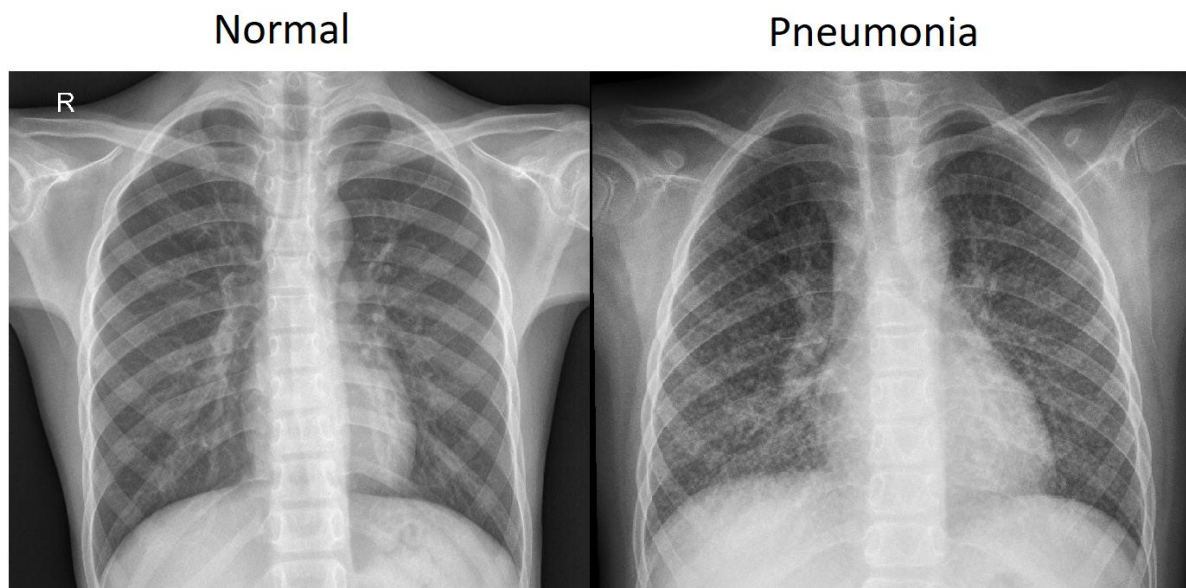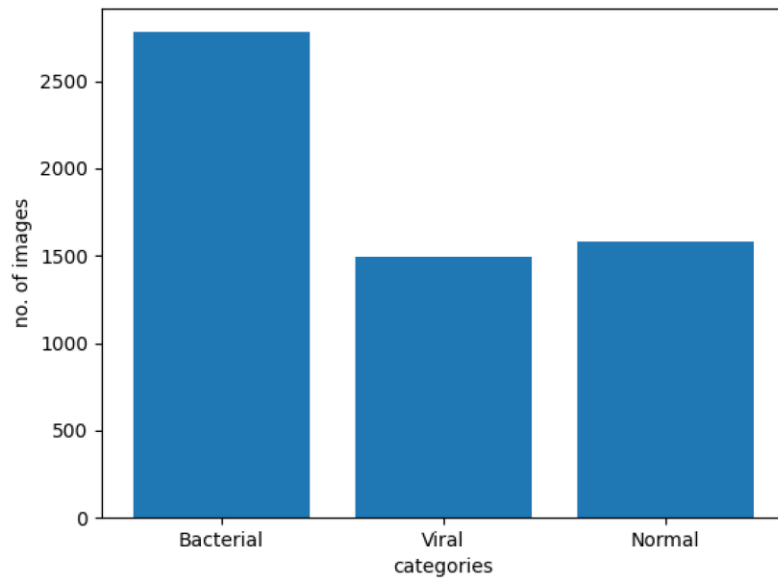
Figure 4.0

Figure 4.1

However, bacterial and viral pneumonia are indistinguishable from each other if only the chest x-ray is used. Additional data and testing are required to know the origin of the infection. Because of this the viral and bacterial categories were both considered as a single pneumonia category. The combined dataset had the following distribution shown on figure 4.2.

Figure 4.2
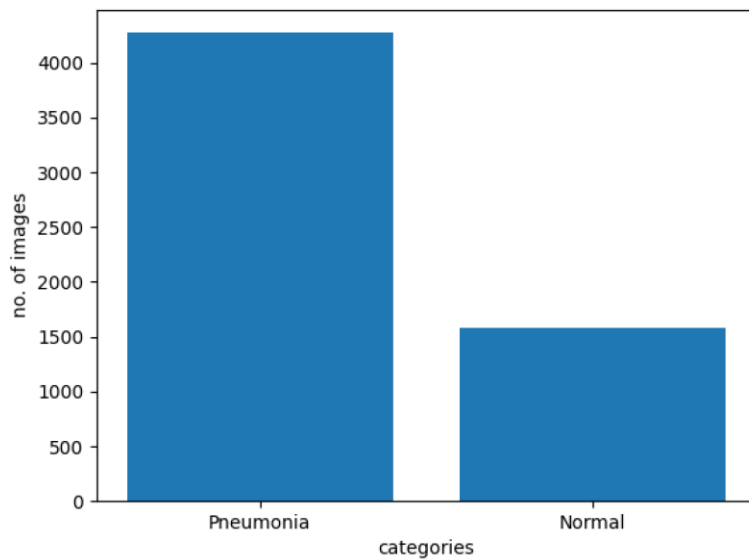
The number of patients diagnosed with pneumonia greatly outnumbers the number of x-rays without pneumonia. To equalize them 1000 images from each category were randomly selected

from the given dataset and loaded into the X matrix for analysis. Totally 2000 images were analyzed. The test set consisted of 25% or 500 images and the training set consisted of the remaining 75% or 1500 images. Before analyzing the images with the neural network, the images needed to be filtered to be in a more manageable size for the network. This project used a simple blur filter. This blur filter 'selected' every 30x30 kernel in the image and took the average as a representation of the kernel. This helps reduce each image to a sizable matrix. As pneumonia affects a large part of the lung, a simple blur would suffice to preprocess the data effectively without deterioration prevent data loss. After applying the blur filter the images were padded with zeroes to give all the images the same shape. This allowed all the images to remain consistent with identical resolution and prepare it for machine learning analysis.

4.2 - Model

A convolutional neural network was used to analyze the data. This network utilized TensorFlow to create a model that classifies the processed images into pneumonia and normal categories. This model used the following structure shown in figure 4.3.

Figure 4.3
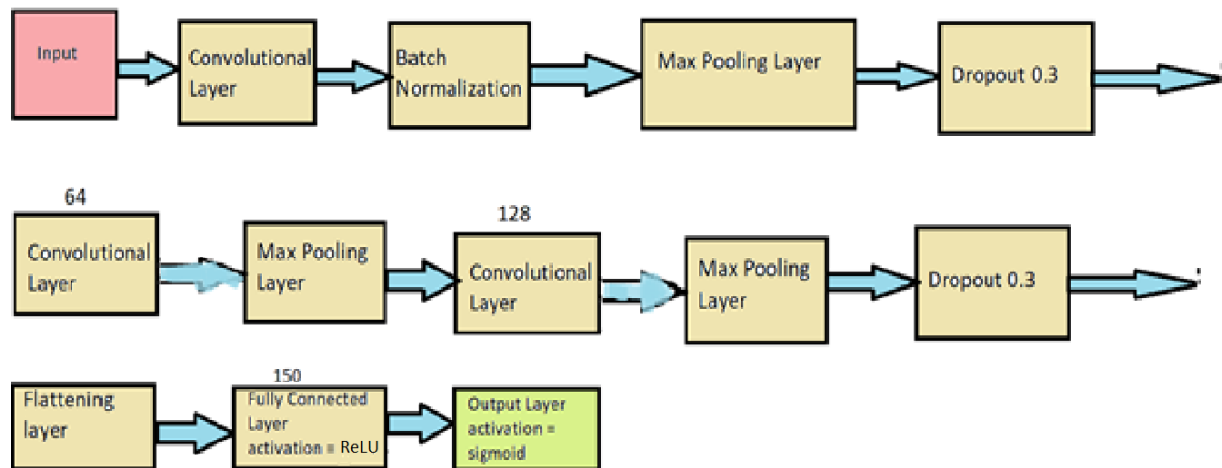
4.3 - Validation

Binary crossentropy error is a method of calculating the difference between the actual values and the predicted values to gauge the model's quality. The model uses this value as a training parameter and the lower it is the higher the quality. Binary crossentropy is depicted by the following formula.

$$BCE = -\frac{1}{N}\sum_{i=0}^{N} y_i log(\hat{y}_i) + (1 - y_i) * log(1 - \hat{y}_i)$$

Binary crossentropy error takes the number of outputs(N), the predicted value ($\hat{y}_i$) and the label(y_i) of the ith label. This binary crossentropy equation is used widely in quantifying error. [12]

Other metrics used in this study are precision, recall and accuracy. They are measured in terms of true positives, false positives, true negatives and false negatives. A true positive is if a predicted value is a positive and the actual value is also a positive. A false positive is if an actual value is

positive, but the model predicted it negative. A false negative is if an actual value is negative, but the model predicts it as positive. A true negative is of the actual value matches the predicted value and they are both negatives. Precision, recall and accuracy use these variables and are important for model evaluation as they quantitatively measure the model's correctness in predicting positives(ones) and negatives(zeros). Precision specifically is used to measure the percent of positives the model accurately diagnosed out of all predicted positives. Precision is defined by the following formula.

$$Precision \ = \ \frac{True\ Positives}{True\ Positives + False\ Positives}$$

Recall is used as a metric for how many positive predictions were truly positive and is defined by the following formula.

$$Recall \ = \ \frac{True\ Positives}{True\ Positives \ + \ False\ Negatives}$$

Accuracy is an overall metric meant to check the percentage of correct prediction. It is defined by the below formula.

$$Accuracy \ = \ \frac{True\ Positives \ + \ True\ Negatives}{True\ Positives \ + \ True\ Negatives \ + \ False\ Negatives \ + \ False\ Positives}$$

This model predicted pneumonia with a loss of 0.37 and an accuracy of approximately 91% and a validation accuracy of approximately 89%. The predictions had a precision of approx. 96% and a recall of approx. 92%. These values are rounded to the nearest integer. This model predicts with considerable accuracy and can predict most true cases of pneumonia. This is important because the consequences of wrongly predicting a real positive are huge as the individual may not receive the treatment required. This model has a high precision and recall, counteracting this. [13]

4.4 – Overfitting

Overfitting is a major problem in many complex models. Overfitting occurs if the difference between the validation accuracy and the training accuracy is significantly different. The pneumonia detection model does not overfit significantly as the training accuracy is 91% and the validation accuracy is 89%. To reduce the overfitting dropout layers were added to the network. Dropout layers randomly drop out a certain percentage of nodes temporarily every epoch. This prevents certain nodes overfitting the training set. Dropout was added twice, dropping out 30% of the nodes. Another measure taken was adding early stopping, meaning the model would stop iterating after the optimal accuracy is reached and the remaining epochs are not performed. This would terminate the training before the network has the chance to overfit, hence decreasing the chance of overfitting by the model.

5 – Conclusion
In conclusion, automatic pneumonia detection is possible through machine learning neural network models. The impact of integrating this into the healthcare system is phenomenal. It would allow the easy accessibility of pneumonia detection methods in hospitals without fulfilling the requirement of a specialist trained for this task. It would also increase the speed of pneumonia detection with minimal human error. While it would take minutes to be analyzed by a specialist, it would only take milliseconds to be analyzed by this model. This model also has a high validation

accuracy of 89%, allowing for mostly accurate predictions. This model also accurately diagnoses most pneumonia patients with the precision being 96% and the recall being 92%. This technique is also different from other techniques as it takes precision, accuracy and recall into context. Implementation of this new machine learning model into the healthcare industry can reduce the morality rate of children in communities that do not have access to specialists. This model could also be retrained and used for other chest x-ray problem; however further testing and training is needed to confirm this.

Bibliography

[1] - Dhaval Dholakia, "Activation Functions. So why do we need Activation functions" ,Towards DataScience, https://towardsdatascience.com/activation-functions-b63185778794#:~:text=Activation%20Functions%201%20Sigmoid%3A%20The%20sigmoid%20activation%20function,%3D%20max%20%280%2C%20z%29%60.%20...%20More%20items...%20

[2] –'ChestX-ray(Radiograph)',June 15, 2020, radiologyinfo.org,

https://www.radiologyinfo.org/en/info/chestrad#213a0544fa974b5799b4937e1f8703a9

[3] - Ian Goodfellow; Yoshua Bengio; Aaron Courville, 'Deep Learning', MIT Press, 2016, http://www.deeplearningbook.org

[4] - Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal, Editor(s): Ian H. Witten, Eibe Frank, Mark A. Hall, Christopher J. Pal, 'Data Mining (Fourth Edition)', 2017, Pages 417-466, doi = https://doi.org/10.1016/B978-0-12-804291-5.00010-6

[5] - Tatiana Gabruseva, Dmytro Poplavskiy, Alexandr Kalinin, 'Deep Learning for Automatic Pneumonia
Detection' The CVF,_https://openaccess.thecvf.com/content_CVPRW_2020/papers/w22/Gabruseva_Deep_Learning_for_Automatic_Pneumonia_Detection_CVPRW_2020_paper.pdf

[6] - Pasa, F., Golkov, V., Pfeiffer, F. 'Efficient Deep Network Architectures for Fast Chest X-Ray Tuberculosis Screening and Visualization.' 2019. https://doi.org/10.1038/s41598-019-42557-4

[7] - Bass Urgent Care, 'Four Stages of Pneumonia', Dec 3 2019, BASS Urgent Care. https://www.bassadvancedurgentcare.com/post/four-stages-of-pneumonia

[8] - Kermany, Daniel; Zhang, Kang; Goldbaum, Michael (2018), "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification", Mendeley Data, V2, doi: 10.17632/.

[9] - Luka Račić, Tomo Popovic, Stevan Cakic, Stevan Šan, 'Pneumonia Detection Using Deep Learning Based_on_Convolutional_Neural_Network',_Febuary_2021,_researchgate.net, https://www.researchgate.net/publication/349408817_Pneumonia_Detection_Using_Deep_Learning_Based_on_Convolutional_Neural_Network

[10] – Mayo Clinic Staff, 'Pneumonia - Symptoms and causes', June 13th 2020, Mayo Clinic, https://www.mayoclinic.org/diseases-conditions/pneumonia/symptoms-causes/syc-20354204

[11] – IBM Cloud Education, 'What are Convolutional Neural Networks?', 20 October 2020, IBM, https://www.ibm.com/cloud/learn/convolutional-neural-networks

[12] – Daniel Godoy, 'Understanding binary cross-entropy', November, 21 2018, medium [Understanding binary cross-entropy / log loss: a visual explanation | by Daniel Godoy | Towards Data Science](#)

[13] – Ahmed Fawzy Gad 'Accuracy, Precision, and Recall in Deep Learning', 2020, Paperspace Blog, https://blog.paperspace.com/deep-learning-metrics-precision-recall-accuracy/

[14] Stacey Chamberlain, Uwe Stolz, Bradley Dreifuss, Sara W. Nelson, Heather Hammerstedt, Jovita Andinda, Samuel Maling, Mark Bisanzo , "Mortality Related to Acute Illness and Injury in Rural Uganda: Task Shifting to Improve Outcomes", 2015 Apr 7, Plos One, [Mortality Related to Acute Illness and Injury in Rural Uganda: Task Shifting to Improve Outcomes (nih.gov)](#)