Predicting the Likelihood of Mortality in Confirmed Positive COVID-19 Patients

Rajdeep Singh

April 10, 2020

smsrajdeep@gmail.com

Abstract

The novel coronavirus - COVID-19 - has evolved into a global pandemic. With that, it is imperative that countries and medical facilities are equipped with the technology and resources to give every person the greatest chance of surviving. With that, even developed nations are beginning to run low on medical supplies such as hospital beds, masks, and respirators.With the growth of cases in the United States, hospitals will continue to run out of supplies. It is imperative that medical supplies get distributed to those who need it the most first. This paper outlines a machine learning approach to predicting patients who are at the most risk of mortality given the confirmed positive diagnosis of coronavirus. The final results were inconclusive enough to be implemented in a real-world scenario.

1 Introduction

According to the Organization for Economic Cooperation and Development, the United States is less equipped to handle a pandemic than many other countries who are facing it. According to its data, China's healthcare system holds 4.3 hospital beds per thousand patients, Italy's holds 3.2, and the US holds a mere 2.8. [1] With this, scientists have worked to attack from every angle possible. Machine learning is being deployed to track the spread of infection, create forecasting models, and even detect the virus early.

This approach is designed for patients testing positive for COVID-19. Given the patient data, the deep neural network predicts the chance of mortality. This type of model will give medical staff technology to know who will be impacted the most from this virus.

It has been seen that different countries have faced different mortality rates with different demographics being affected. According to the CDC, from what has been observed, the affected age groups are slightly younger in the United States than China. [2]

The purpose of this research is to identify any outliers: those who seem to be in a healthy demographic, but - only to be learned by machine learning - are at a heightened risk level. This could further help the healthcare system identify more accurately identify and diagnose coronavirus among patients. The backpropagation was used to train the model on COVID-19 dataset [3] with, specifically, the anomalies in the dataset. The final model outputs the predicted likelihood of mortality in confirmed positive COVID-19 patients.

Until this point in time, most clinical employments of AI to the COVID-19 response have focused on end reliant on clinical imaging. In progressing composing, we have found a couple of works that usage AI to help finding from computational tomography (CT) checks, despite others that use understanding clinical data to envision the improvement of the contamination, similarly as interesting non-prominent estimations which can be used for watching purposes. Fever, hack and brevity of breath are the most significant side effects in contaminated people for the conclusion of COVID-19. Simultaneously, these indications may show bearer attributes by not being seen in tainted people. Neurotic tests acted in research facilities are taking additional time. Likewise, the room for give and take can be high.

A quick and exact finding is important for a compelling battle against COVID-19. Hence, specialists have been begun to utilize radiological imaging techniques. These techniques are performed with figured tomography (CT) or then again X-beam imaging strategies. COVID-19 cases have comparative highlights in CT pictures in the early and late stages. It appears a roundabout and internal dispersion from inside the picture. In this way, radiological imaging gives early recognition of suspicious cases with a precision of 90

2 Data

Raw Dataset The data used in this model was pulled from a clinical patient dataset of closed cases around the world. [3] The dataset included information about each pa-

tient such as their location, age, sex, symptoms, and outcome of the case. The data consists of 65 deaths across the training and test set. The entire dataset represents a span of all ages.

Prepossessing Total information estimation for nations with various areas or states were performed in the wake of gathering by a straightforward summation over totally related locales or regions (for the nations China, Australia, and Canada). We have chosen the informational indexes with the 25 most contaminations (and China) as introduced in on 2020-04-05 23:00 CET. In our past production we have decided the scaling coefficients for every one of the 25 generally influenced nations and contrasted the relative development rates and one another. Since more information is accessible at this point we have dissected the nearby slants of the assumed force law run in further detail. In a log-log portrayal log N(t) versus log t a force law N(t) Bt becomes. Aside from these guidelines, which apply to every paper, the structure of the body varies a lot depending on content.

The accuracy of the model on the test set was determined by diving the number of correct predictions by the total predictions; and multiplying that by 100.

3 Research

The model is constructed with a deep neural network. First the patient data is formatted into a R4x1 matrix. Then, that matrix's data is preprocessed and fed forward through a six layer deep neural network. The neural network's first layer is 64 neurons with a ReLu activation function. The second layer has the same activation function, but it is 512 neurons. The following layer is the same with 128 neurons. Following this - on the fourth layer - is a dropout layer with a coefficient of 0.25. After the dropout is another dense layer of 64 neurons activated by the ReLu function and an output layer represented by a single neuron activated by the sigmoid function.

A dropout coefficient of 0.25 provided the best prevention from overfitting while also not decreasing accuracy. The model was trained over 50 epochs with a batch size of 10 to furthermore prevent overfitting. Trials showed that the model's loss function stopped decreasing and its accuracy stopped increasing after 50 epochs on the training data.

Prepossessing Further The data was preprocessed to fit the input for the neural network. First, the labels for country, location, and gender had to be serialized for input. Next, the features were scaled using MinMaxScaler to make training more efficient. The maximum value of the data was scaled to one, the minimum value was scaled to zero, and all other data points were scaled in between accordingly.

Backpropagation To train the model, backpropagation was used with a binary crossentropy loss function. This trained the model to classify the chance of death with a binary classifier.

4 Results

This model was able to predict on the cross-validation dataset with a total binary cross entropy loss of 0.0946 and an accuracy of 96.80

To make a more practical model, the standard deviation of the prediction likelihoods minus one mean gave an accurate threshold line. The model predicts approximately 12

The model correctly tagged all of the deceased patients but two: a 45-year old man from Japan and a 58-year old woman from China.

5 Conclusion

Fighting COVID-19 has promoted global teamwork in the fields of research, medicine, data science, and much more. Putting this architecture with more complete data will give medical professions the tools they need to treat the most at-risk patients.

Though, it is apparent that the results of this results shouldn't be taken seriously.

6 Citations

[1] Harris, Joel. "Health Equipment - Hospital Beds - OECD Data." TheOECD, 2020, data.oecd.org/healtheqt/hospital-beds.htm.

[2] Researchers, CDC. "Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) - United States, February 12–March 16, 2020." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 26 Mar. 2020, www.cdc.gov/mmwr/volumes/69/wr/mm6912e2.htm?s$_{c}id = mm6912e2_{w}$.

[3] Puspar, Nikil. "Novel Corona Virus 2019 Dataset." Kaggle, 2020, www.kaggle.com/sudalairajkumar/novel-corona-virus-2019-dataset.

7 Appendix

References

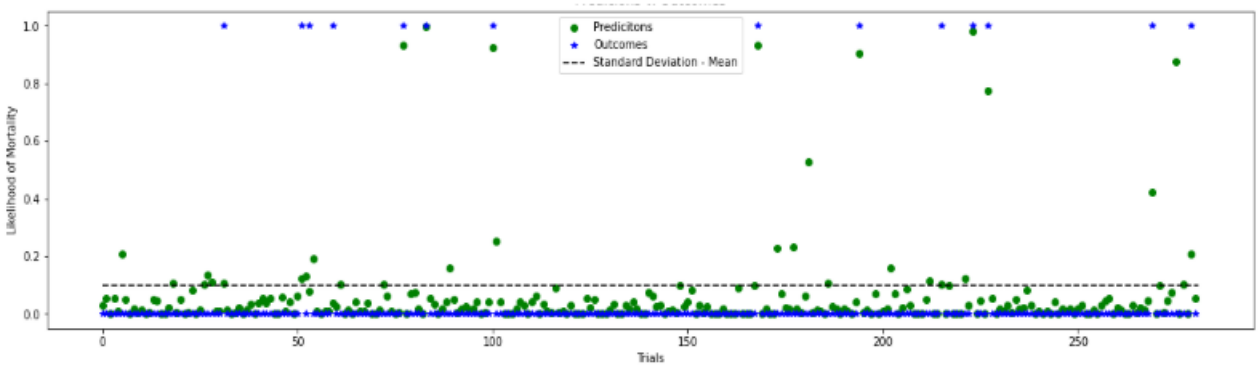

Figure 1. Results