Analysis and Prediction of Global Pandemic: COVID-19

May 4, 2020

Abstract

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which now results in the ongoing 2019–20 coronavirus pandemic. As of 11 April 2020, more than 1.77 million cases of COVID-19 have been reported in 210 countries and territories, resulting in more than 108,000 deaths[1]. The first part of this study focuses on visualizing the disease data to develop the increasing pattern and related social factors of the disease. In the second part of this study, we use two features in both temporal and atemporal dimensions to predict the number of the future number of confirmed cases and fatalities. The temporal features include the past daily number of confirmed cases and deaths. The atemporal features include information such as population distributions of each country/region, the percentages of smokers in each country/region. In order to predict the future number of confirmed cases and fatalities, we construct a novel hybrid deep learning machine learning model. Multi-Stacked Fully Connected Bidirectional LSTM model(MSFCB-LSTM), which is a part of the hybrid model, is proposed to extract the information in the temporal dimension. A typical machine learning neural network is constructed to extract atemporal features. The results show that our model performs well in predicting the future number of confirmed cases, fatalities, and the patterns of change of the prediction. The root mean square error (RMSE), which is defined to measure the error of the daily confirmed cases, is 1127.061, and that of the daily fatalities is 0.845. To the best of our knowledge, this is the first time using a hybrid deep learning model and taking both temporal and non-temporal features together to predict the daily confirmed cases and fatalities.

Keywords: Data Visualization, Data Mining, Hybrid Deep Learning Model, MSFCB-LSTM.

Contents

1 Introduction

1.1 Background

SARS-CoV-2 is the new virus of the corona-virus family, which was first reported near the end of 2019, and has not been identified in humans before. It is a contiguous virus that was first reported in Wuhan, China, in December 2019. COVID-19 was later declared as Pandemic by WHO due to the high rate spreads throughout the world. A pandemic is a global outbreak of disease. Pandemics happen when a new virus emerges to infect people and can spread between people sustainably. Because there is little to no pre-existing immunity against the new virus, it spreads worldwide. Currently (on date 11 April 2020), this leads to a total of more than 84,000 Deaths across the globe. With a reproductive number (R_0) equals 5.7[2], The potential public health threat posed by COVID-19 is very high. The coronavirus COVID-19 is affecting 210 countries and territories around the world and two international conveyances. The dramatic spread of COVID-19 has disrupted lives, livelihoods, communities, and businesses worldwide. All stakeholders, especially global business, must urgently come together to minimize its impact on public health and limit its potential for further disruption to lives and economies around the world.

Tracing and monitoring the trends and changes continuously in the course of the COVID-19 outbreak is very important to mitigate this epidemic threat. Obtaining and analyzing information about what has happened until now and what might happen in the future is essential. Currently, some countries already passed the peak period of the spread of COVID-19, while some are still having cases that increase exponentially. In this report, we analyzed the disease's worldwide developing pattern and the severity of the situation in some regions through data visualization skills. We then used the Recurrent Neural Network (RNN) to predict the future developing trend of the pandemic based on past data and related factors.

1.2 Statement of the Problem

In this report, we utilize data analysis skills and deep learning methods to analyze and construct models to predict confirmed cases and fatalities. This work mainly contains two parts: data visualization and analysis and deep learning model prediction.

In the data visualization and analysis part, a worldwide heat map and separate heat maps for the United States and Europe were generated. We then plotted the line chart of countries with the daily increase in confirmed and death cases. Most of the changes can be explained by the population structure, resource situation, or the countries' corresponding policies. Finally, we analyzed the available data of different countries' testing ability and combined the result with the mortality rate. Based on all the analyses, the spread of the disease is closely related to polices and the availability of medical resources.

In the prediction part, we used a hybrid deep learning method, which utilizes both temporal and non-temporal data as input features to predict confirmed cases and fatalities in the future. The hybrid deep learning model can learn the patterns in the temporal dimension by a novel proposed model, MSFCB-LSTM, and analyze and extract the non-temporal features that are proven to be significant in our prediction work. Our final predictions are based on the information extracted from both temporal and non-temporal dimensions. It is shown that our hybrid model performs well in predicting daily fatalities. It also performs well in describing the change or patterns of the number of daily confirmed cases.

The remainder of this paper is organized as follows. Section 2 introduces the data source we use, the features we choose in our model, and the corresponding data pre-processing.

Section 3 presents the details of our data analysis of the global situation, the developing global trend of the pandemic, and the study of testing ability and mortality rate in the United States in detail since the United States has the most challenging situation right now. The prediction model and the prediction results are shown in section IV. Section 5 shows some discussion and the outlook of our work. In the end, the summary of this work is discussed in Section 6.

2 Data Source and Pre-processing

2.1 Data Source and Description

The corona-virus global case data of this report is retrieved from the Center for Systems and Science Engineering of John Hopkins University. The data sets used in this report are global confirmed and death cases; the US confirmed cases. The data of testing ability in each country is retrieved from the website: Our World in Data. The geographic data used to generate plots is contained in Python package: folium. The data of COVID-19 worldwide cases used in this report covers till April 11^{th}, 2020. The world population by age data is retrieved from the website of the Department of Economic and Social Affairs of the United Nations. This is a data set containing both historical and prediction data about the pyramid of the population by age (slices of 5 years) and by country/region of the world. The smoker's percentages by country are retrieved from the website: Our World in Data. The time data information from quarantine, restrictions, and schools is gathered from the OECD site, Worldometer, and is up to date (April 2020).

2.2 Data Pre-processing

For data used in the visualization part, we first changed the names of the country/region obtained from the original database to match the 'pycountry_convert' Library. Then the regional data are matched to geographical information, including longitude and latitude data. The mortality rate is calculated from the total number of deaths and the total number of confirmed cases.

For data used in the prediction part, we obtained the data from the Center for Systems and Science Engineering of John Hopkins University contains the province/state, country/region, date, confirmed cases, and fatalities. The data set was broken into two parts, 80% for training and 20% for testing in our hybrid deep learning model. To provide more features and enrich the input data, we add some extra data in the beginning as follows:

- Population distributions by country. This data set provides the population information like total population, population density, urban population, and the percentage of different ages population in each country/region. We add this data since older people are more likely to be infected by the COVID [7]. Thus, an aging country/region may have more confirmed cases and fatalities number.

- Smokers percentages by country. This data set provides the percentage of smokers in each country/region. It is proven that smokers have increased airway expression of ACE-2, which is the entry receptor for the COVID-19 virus, which causes an increased risk of severe COVID-19 in these sub-populations [8]. Thus, a country/region with a high percentage of smokers may have more confirmed cases and fatalities.

- Time data information from quarantine, restrictions, and schools. These features are indicator data. For example, quarantine equals to 0 implies that this country/region

does not put a person into quarantine. It is known that quarantine is an efficient way to stop the spread of the COVID-19 [9]. Thus, a country/region with proper quarantine and restrictions are more likely to control the spread of the pandemic.

We consider the above information at first and analyze their correlation with the confirmed cases and the fatalities. In the following, we draw the plot of the correlation matrix among each of the variables. In the correlation matrix plots, the deeper the color is, the higher the correlation between the two variables is. Red means a negative correlation, while blue means a positive correlation relationship. These correlation matrices give us strong evidence that the features we used to predict confirmed cases and fatalities are meaningful and significant.

We first combine the above data country by country and date by date to build up our input features. The correlation among Confirmed cases, fatalities, and percentage of the population with ages larger than 60 years old is shown in Fig.1. This plot shows that the confirmed cases and the fatalities number are a bit positively related to the percentage of the population ages 65 years old to 70 years old. They are also negatively related to the percentage of the population ages from 70 years old to 85 years old. But the number seems to be not related to the portion of the population with ages larger than 85 years old. It may because the total number of older people that are older than 85 years old is tiny. Thus, we do not consider the percentage of the population with ages larger than 85 years old in our model.

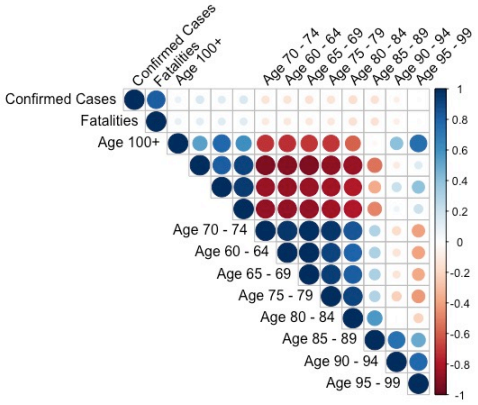

Figure 1: Correlation among Confirmed cases, fatalities, and percentage of the population with ages larger than 60 years old

The correlation among Confirmed cases, fatalities, and percentage of the population with ages 0-59 years old is shown in Fig.2. We can see that both of the confirmed cases and the fatalities are nearly not correlated with the population with ages from 0 to 40 years old. They are negatively related to the percentage of people ages from 40 years old to 60 years old. It intuitively makes sense since young people have a stronger immune system. The percentage of the population with ages from 0 years old to 40 years old is not considered in our model.

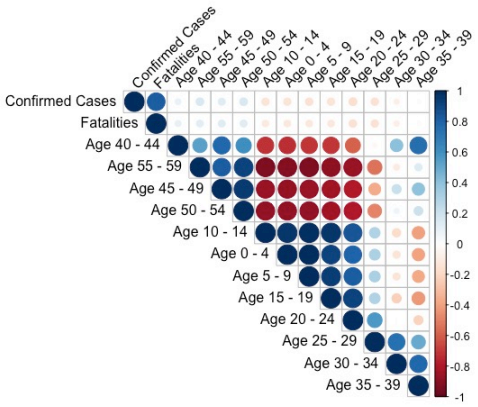

Figure 2: Correlation among Confirmed cases, fatalities, and percentage of the population with ages 0-59 years old

The correlation among Confirmed cases, fatalities, and other features introduced above are shown in Fig.3. From the correlation matrix, we can conclude that the confirmed cases and the deaths are positively related to the population and density of each country/country, which is intuitively correct. A country/region with more population has more chances to have a more infected population and more fatalities. Confirmed cases and fatalities are negatively related to quarantine applied for the area. If a school open or closed for the region, restrictions used for the region and the number of hospital beds of each country/state is significant, then the country/state tends to have less confirmed cases and deaths.

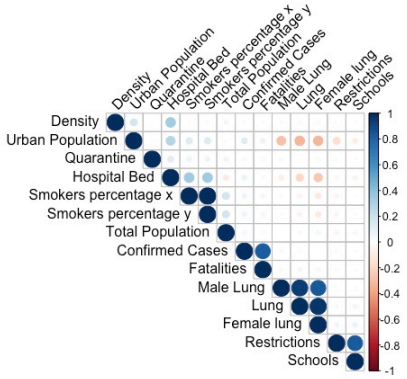

Figure 3: Correlation among Confirmed cases, fatalities and other features

3 Data Insights and Model Description

3.1 Data Insights

In this section, data visualization skills are applied to discover the spread pattern and regional characteristics of the disease. The testing ability of different countries will also be analyzed.

3.1.1 Global Situation Overview

Fig.4 shows the number of countries affected over time, which shows the corona-virus is affecting most of the countries in the world. It was first discovered in China in early December and then discovered by the increasing number of countries worldwide through February and March. By the date of April 11^{th}, 2020, there are 211 countries reported cases in total.

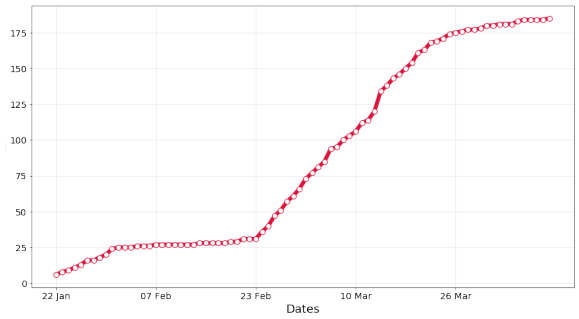

Figure 4: Number of Countries Affected

3.1.2 Analysis of Severe Districts

By generating a heat map of global confirmed cases (Fig.5), we can see that Europe and the United States are hit by the virus most severely currently. We then generate the detailed district map of these two districts. Fig.6 indicates that in the United States, the district hit by the disease most is the east coast, west coast, and area around the Great Lakes Region. This can be explained by the population density and public transport mobility in these regions. The heat map of Europe Fig.7, which shows the most severe countries are Spain, Italy, German, and France. Since the outbreak in Europe started in late February and early March, while outbreak happens in the US around ten days later, there are also studies shows that most cases in the east coast of United States are imported from Europe continent considering a large number of flights from Europe landed first in the east part of US[3].

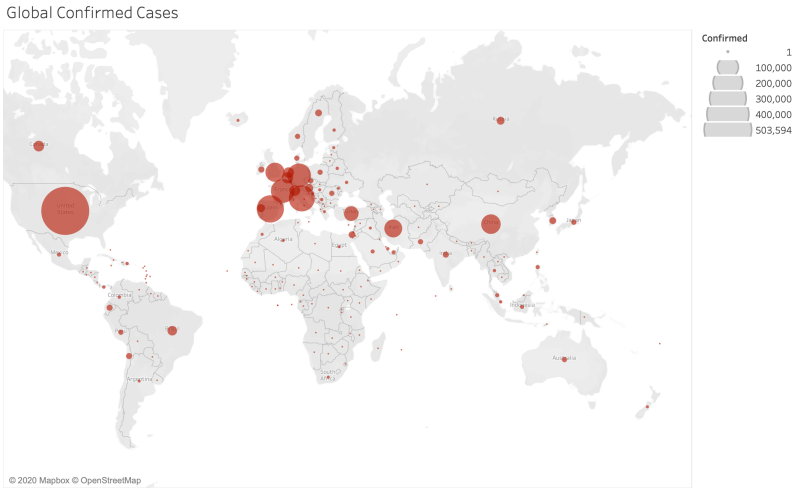

Figure 5: Global Confirmed Cases (April 11^{th})

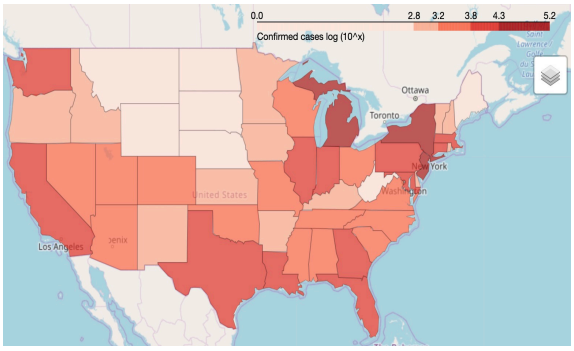

Figure 6: United States Cases (April 11^{th})

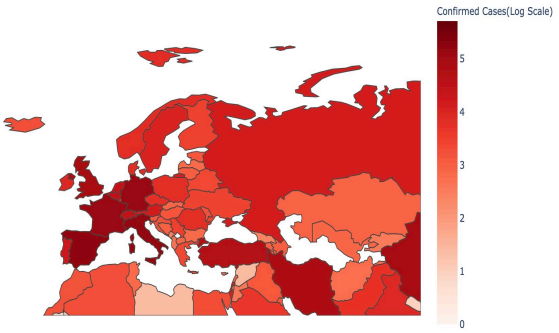

Figure 7: Europe Cases (April 11^{th})

3.1.3 Developing Trend of Daily Increase

Fig.8 and Fig.9 shows the top ten countries with most daily confirmed cases and daily death cases. China and South Korea were the first to experience the outbreak, and are now out of the picture due to strict control. From the confirmed figure, we can see European countries' cases start to increase in late February, while Spain and Italy are showing signs of decreasing the number of new cases, other European countries like UK, Germany, and France are still having an increasing number of daily increased confirmed cases. There is an as significant jump in the data of France at the beginning of April; the huge daily increase is due to including data from the nursing home, which also makes the total number of cases in France jump above China [4]. Russia and Ecuador, whose status used to be quite stable, have many increasing new cases only since recently, the reason for which is due to either delay of testing or lack of control over international transportation. In the United States, the size of daily increases surpassed all the other countries since the middle of March, and there are still no signs of dropping. The pattern of increase in the United States has a periodic decrease every five or six days. The reason might be reducing testing ability during weekends.

From the figure of death cases, we can see that in Europe, France has a turbulent developing trend; Italy, Spain, and the Netherlands seems already passed the highest point of daily increase in death cases, while UK, Belgium still has an increasing number of daily deaths. The United States currently has the highest number of death cases worldwide and is still developing in a exponential increasing trend. Except for some occasional dropping points, the trend of daily increase still seems haven't slowed down.

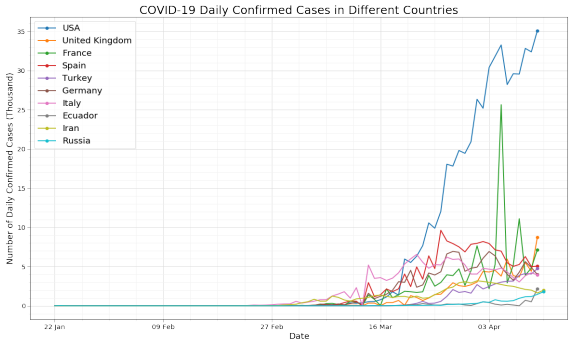

Figure 8: Daily Confirmed Cases Increase

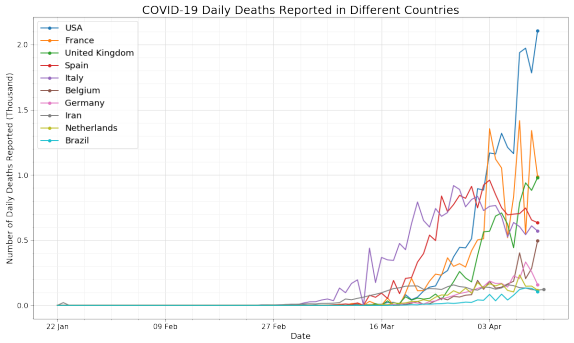

Figure 9: Daily Deaths Cases Increase

3.1.4 Comparison of Trend in/outside China

Since China is regarded as the first country to discover/report the spread of COVID-19, Fig .10 compared the number of confirmed cases in China and the rest of the world. The figure shows that the peak of the pandemic in China happened in February. The lock-down of the central city Wuhan happened in Jan 23^{th}, and on February 12^{th}, due to the change of diagnosis criteria of confirmed cases, China experienced the most significant number of increase in daily confirmed cases. The number keeps dropping ever after, which indicates the effectiveness of the lock-down policy. On March 14^{th}, Italy became the first country to have more confirmed cases in China. The world's cumulative total is still increasing at an exponential rate and not yet reaching its peak. The time gap between the peak of China and the world might be explained by, for example, lack of travel restrictions between countries outside China, delay of testing ability, etc. The length of disease spreading period differ from China and other countries, which can be explained by policies: Chinese 'City lock-down' directly cut the route of transmission at fastest speed, while 'Flatten the curve' policy implemented by western countries can only slow down the process in a limited degree.

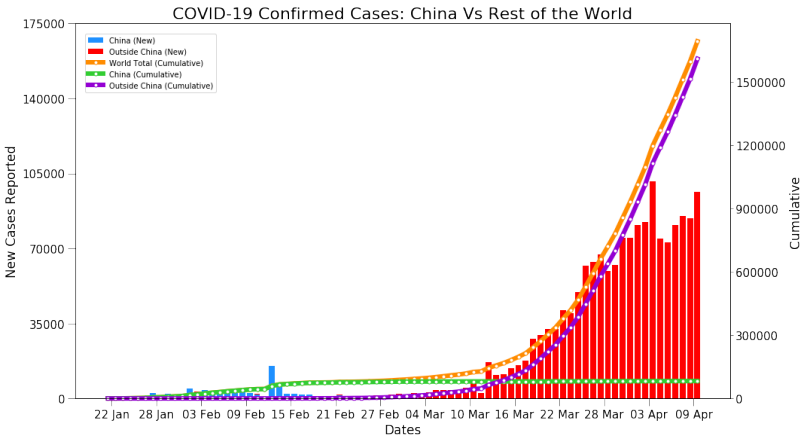

Figure 10: Cases in and outside China

3.1.5 Test Ability Analysis

In this part, we are analyzing the testing abilities of different countries. Figure.11 calculates the cumulative total number of tests, the average number of tests per million people, the mortality rate, positive rate of each country. The color depth indicates the quantity and level of severity. Due to limited data access, we have 17 countries listed here. We can see that among all these countries, the United States has the most significant cumulative number of tests but have a fewer number of tests per million people. Countries like Bahrain and Austria have a large number per million due to the small number of population. Germany, Italy, and South Korea are running many tests based on the population scale.

Positive proportion is also an indicator of the sufficiency of total tests. We can see the Philippines, Indonesia, France, United Kingdom, Canada are having high positive rate which is above 40 %, combining the number of tests per million people in these countries, the result indicates the actual number of affected population should be more significant than what it is recorded.

	country	Cumulative total	Cumulative total per million	confirmed	deaths	MR	Positive
0	Australia	261000	10276.3	6292	56	0.89	2.41
1	Austria	98343	11198	13789	337	2.44	14.02
2	Bahrain	37996	22380	1016	6	0.59	2.67
3	Belgium	62867	5410.25	28018	3346	11.94	44.57
4	Canada	256933	6832.74	22559	601	2.66	8.78
5	France	224254	3412.2	125942	13216	10.49	56.16
6	Germany	918460	11127.4	122855	2736	2.23	13.38
7	India	26798	19.3739	8063	249	3.09	30.09
8	India	47951	34.6668	8063	249	3.09	16.82
9	Indonesia	7621	27.9954	3842	327	8.51	50.41
10	Italy	581232	9829.39	147577	18849	12.77	25.39
11	Japan	39446	311.837	6005	99	1.65	15.22
12	Malaysia	47723	1451.9	4530	73	1.61	9.49
13	Pakistan	30308	145.458	4970	77	1.55	16.4
14	Philippines	5265	47.993	4428	247	5.58	84.1
15	South Korea	443273	8606.08	10480	211	2.01	2.36
16	United Kingdom	173784	2580.92	79841	9891	12.39	45.94
17	USA	1267658	3824.8	503594	18860	3.75	39.73

Figure 11: Countries' Testing Abilities

3.1.6 Mortality Rate Analysis

From the column of mortality rate, we can see that European countries have a higher mortality rate. In Fig.12, we generated a plot of the world's mortality rate and mortality rate in each continent. After a drop in February, the mortality rate of the disease keeps increasing, mainly due to the shortage of healthcare resource and the spread of disease in countries which do not have robust public health systems. The mortality rate in Europe is the highest and is still increasing, which partly can be explained by the structure of the population. The outbreak can explain the first peak of North America in nursing homes in Washington State. From the figure, we can see that the mortality rate in Africa and South America keeps increasing, which is mainly due to a lack of resources.

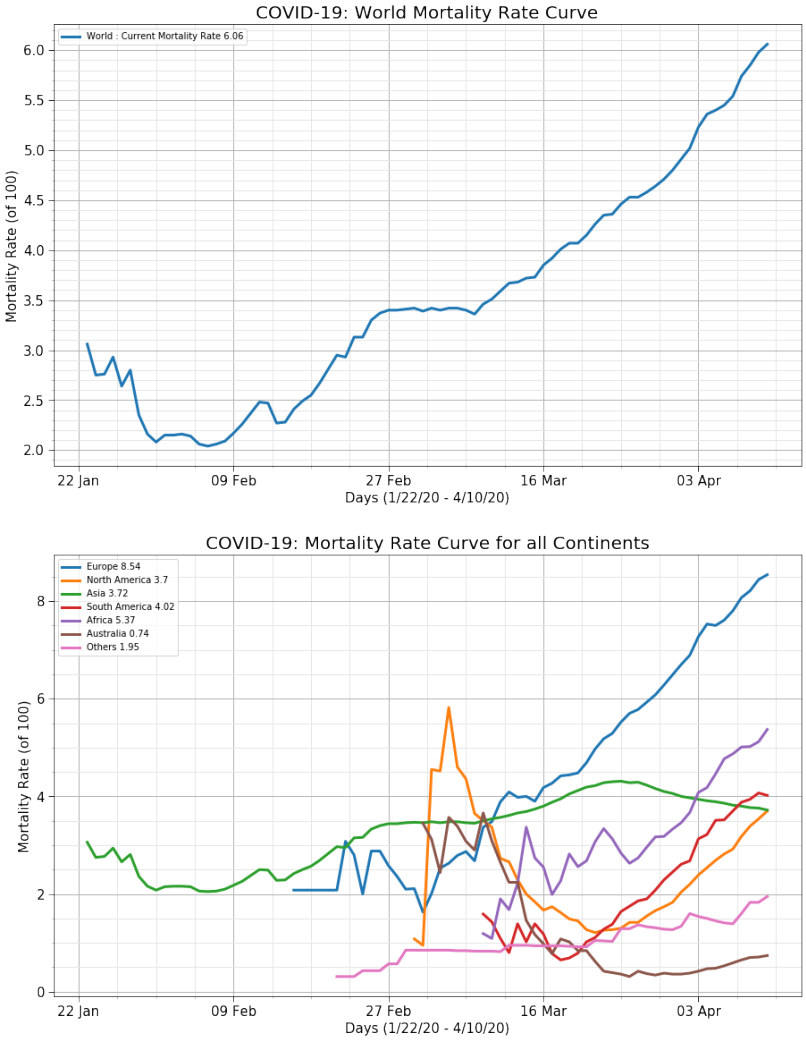

Figure 12: Worldwide & Continents' Mortality Rate

3.2 Model Description

Recurrent Neural Network (RNN) has exhibited the power of analyzing time-series data prediction. Each node at a time step takes input from the previous node, and this can be indicated using a feedback loop.

Even though RNN indicates a strong ability to analyze the correlation in time series data, it has disadvantages in solving long term time series data, which means to understand the context at time step $t+1$, it may require information representing at time 0 and 1. To capture the high dependency in time series, Hochreiter and Schmidhuber[5] invented the Long Short Term Memory (LSTM). Based on typical RNN structure, each cell is indicated by a more complex inner network with the crucial components, gates. Comparing with a single neural layer in RNN, there are four interacting gates, input gate, forget gate, update gate, and output gate. The gated structure, especially crucial forget gate and the update gate, supports LSTM to remember information in the long term time series optionally, and decides vital information to update to the cell state. It has proved that the bidirectional networks[6] are sincerely better than the traditional LSTM in many fields capturing time-series information.

In this work, we define a hybrid deep learning model combining temporal features and non-temporal features to predict confirmed cases and fatalities in the future. The structure of our model is shown in Fig.13

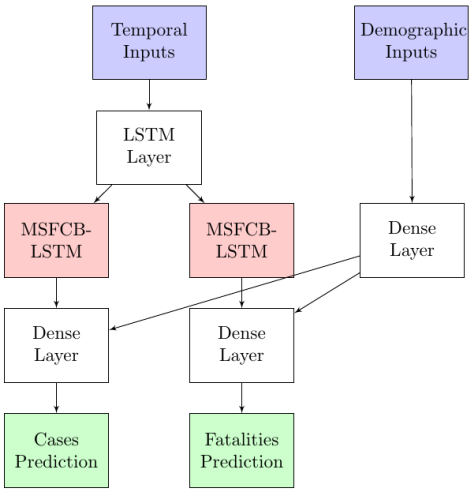

Figure 13: Structure of the hybrid model

To handle the temporal features, we define a novel deep learning model, Multi-Stacked Fully Connected Bidirectional LSTM model(MSFCB-LSTM), to predict the time series data. The model structure of the MSFCB-LSTM model is shown in Fig.14.

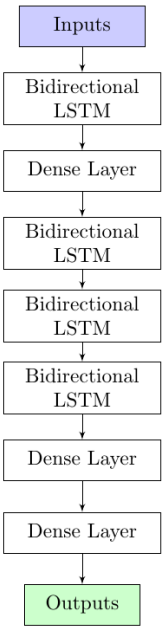

Figure 14: Structure of the MSFCB-LSTM model

We stacked bidirectional LSTM with two layers as the first temporal features or information abstraction, then connect with one layer LSTM to remedy potential temporal correlation. After stacked Bidirectional LSTM and LSTM, two Dense layers added at the end, which is capable of reasonably controlling dimensions in hidden layers output, and more precisely capture non-linearity between inputs and outputs since inputs in our model are multivariate with abstract correlations.

In this study, we use both the temporal and non-temporal features as input. The temporal features include:

- Number of cases for the last 13 days, I_{C_t}

- Number of fatalities for the last 13 days, I_{F_t}
- Restrictions applied for the area in the past 13 days, R_t
- Quarantine applied for the area in the past 13 days, Q_t
- School opened or closed for the area in the past 13 days, S_t

and the atemporal features include:

- Population of each country/state, P
- Density of each country/state, D
- Number of hospital beds of each country/state, N
- Lung measurement of female and male of each country/state, L

The outputs of our model are:

- Number of confirmed cases for the 14^{th} day, O_{C_t}
- Number of fatalities for the 14^{th} day, O_{F_t}

4 Results and Model Validation

4.1 Experiment Setup

All experiments are executed on the Linux cluster (CPU: Intel(R) Xeon(R) E5-2699 v4 @ 2.20GHz, GPU: NVIDIA GeForce GeForce RTX 2080 Ti).

4.2 Evaluation Metric

To evaluate the accuracy performance of the model correctly, we use root mean square error (RMSE). RMSE is defined as follow:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{1}$$

where $\hat{y}_i$ is the i^{th} predicted value, and y_i is the corresponding true value, where $i = 1, ..., n$.

4.3 Results and Model Validation

4.3.1 Confirmed Cases Prediction

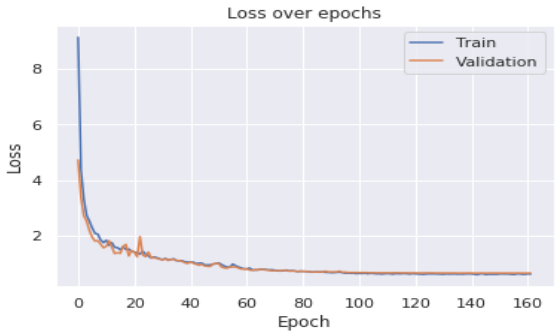

Figure 15: Loss Function of Confirmed Cases Prediction

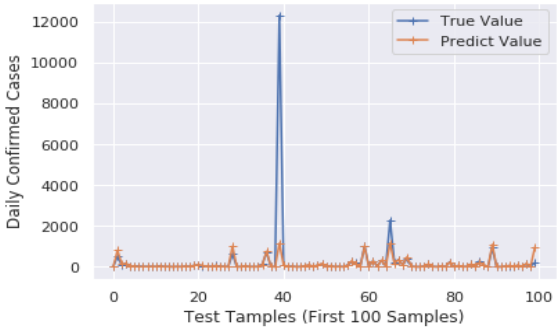

Figure 16: Confirmed Cases Prediction

Fig.15 shows the loss function of training and testing. Fig.16 shows the results of the prediction of the number of confirmed cases for the 14^{th} day. The RMSE, The RMSE, which is defined in Equation(1) is 1127.061. We see that our prediction of the daily confirmed cases in the test data set is precise except for a very steep peak around the 40th sample. It may because some countries/regions change their test policy that day or improve their testing ability. We may consider the steep peak as an outlier in the data set. However, if we have some more information about the policymaking of each country/region, we could create some new features that could improve the prediction results. Fatalities Prediction

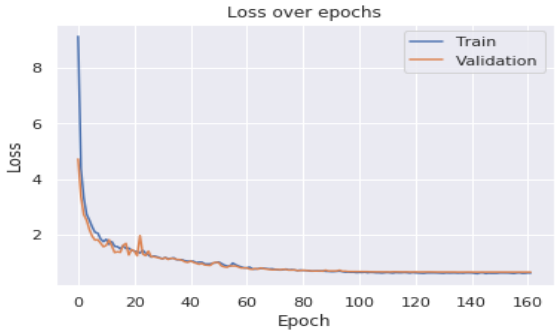

Figure 17: Loss Function of Fatalities Prediction

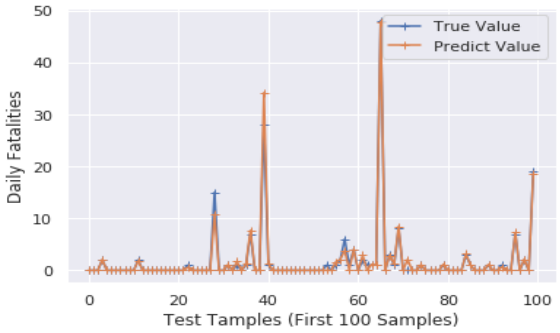

Figure 18: Fatalities Prediction

Fig.17 shows the loss function of training and testing. Fig.18 shows the results of the prediction of the number of fatalities for the 14^{th} day. The RMSE, The RMSE, which is defined in Equation(1) is 0.845. We see that our prediction results are quite precise when we predict the fatalities of the fatalities in the 14^{th} day.

4.4 Results Analysis

The prediction results show that our model performs well when we predict the future fatalities. The RMSE is only 0.845, and the patterns our model learned are nearly the same as the true value. Although the RMSE of the confirmed cases prediction seems to be large, the patterns the model learn is also very well. This large RMSE is mainly due to some unusual confirmed cases in the test data set. In Fig.16, there is a very sharp peak that

seems to be abnormal and results in a very large RMSE. These abnormal data may due to some human factors and political reasons. For example, China adopted clinical diagnosis on 12^{th}, February 2020, for the first time, which results in a large amount of new confirmed cases, and America began to enlarge its testing ability since the middle of March. These kinds of reasons may affect our prediction results. In the future, we could improve our model by adding these kinds of human and political decisive factors as input to our model. Our prediction of the fatalities in the 14^{th} day seems to be much preciser than that in the 14^{th} day. It is intuitively reasonable since a country/region can change its policy of testing the infected population or improving their testing ability within a few days. However, the percentage of patients in serious conditions would not change too much. The ordinary patients would not die immediately. Thus, the variation in the number of fatalities is not very big. As a result, the prediction results are more precise.

5 Discussion

5.1 Model Shortness

We use a hybrid deep learning model to predict the daily confirmed cases and fatalities. We combine the temporal and non-temporal features that influence these numbers. We use the multi-stacked fully connected bidirectional LSTM model(MSFCB-LSTM) to fully extract the temporal information and use a dense layer to extract the non-temporal information. Combine them to get the final prediction. However, our model has some shortness. First, our model is not good at predicting the case when the future daily confirmed cases and fatalities number are very different from the previous days. The hybrid LSTM model predicts the final results based on the information from the earlier days. Besides, we also do not have enough information about the policy-making of each country/region. If a country/region changes its policy of testing the infected population mainly, our model may perform poorly in this case. Using deep learning for disease prediction does not consider the features of the disease, like R_0, transition rates, recovery rate etc, thus it can only be applied for real-time update currently, and still require more features for future trend prediction.

5.2 Future Application

Although our model has some shortness, it indeed performs well in the ordinary situation. We can use the model to predict and real-time update the daily confirmed cases and the fatalities based on the existing information. The decision-makers of each country/region can adjust their decisions and policy based on our prediction results. For example, if the prediction results of our model show that the fatalities and the confirmed cases would increase the next few days, the decision-makers could reinforce pandemic prevention assistance and control the spread of the pandemic disease. In this case, the country may reduce the number of infected populations and fatalities.

6 Conclusions

Based on the visualization and analysis, we can see an increasing number of countries and territories are being affected by the pandemic. The United States and Europe are in the most challenging situation, and the United States was just (April 11^{th}) announced as the new center of the pandemic. An exponential growing pattern can be seen from the figures of

daily increase cases of deaths and confirm. The comparison figure of China and the rest of the world emphatically proves the effectiveness of the lock-down policy and shelter at home policy. From the testing ability analysis, we can see that some countries are still not having the solid testing ability, where the situation should be more complicated than existing data shows. The mortality rate of the disease also has a robust geographical trend, which is strongly affected by countries' population structure, public health system, and economic situation.

The proposed hybrid deep learning model makes confirmed cases and fatalities predictions well capturing temporal features and non-temporal features. The MSFCB-LSTM model is proposed to extract the time-series data, and a typical machine learning neural network is designed to extract information in the atemporal dimension. The prediction results show that the proposed hybrid machine learning and deep learning model perform well in both predicting the future number of confirmed cases and fatalities. It exhibits excellent performances when we predict the future number of fatalities with an RMSE, only 0.845. It also shows an excellent performance in predicting the patterns of the future daily number of confirmed cases. However, it does not performs that well if the future daily confirmed case number is too steep. In the future, we can improve our prediction results by adding more features related to human factors and political decision features that would significantly affect the prediction results.

References

[1] Johns Hopkins University, Center for System Science and Engineering. COVID-19 Dashboard. Retrieved from https://coronavirus.jhu.edu/map.html

[2] Sanche, S. , Lin, Y. .(2020). High Contagiousness and Rapid Spread of Severe Acute Respiratory Syndrome Coronavirus 2 *Emerging Infectious Disease*, 26(7).

[3] Langreth, R. ,(2020). Most NYC Covid-19 Cases Came From Europe, Genome Researchers Say. *Bloomberg*, April 8th, 2020. https://www.bloomberg.com/news/articles/2020-04-08/

[4] Clercq, G. , (2020). French coronavirus cases jump above China's after including nursing home tally. *REUTERS*, April 3rd, 2020. https://www.reuters.com/article/us-health-coronavirus-france-toll/

[5] Hochreiter, S. , Schmidhuber, J. .(1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

[6] Schuster, M., Paliwal, K.K. . Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), P.2673-2681.

[7] Heymann, D. L., Shindo, N. (2020). COVID-19: what is next for public health?. *The Lancet*, 395(10224), 542-545.

[8] Leung, J. M., Yang, C. X., Tam, A., Shaipanich, T., Hackett, T. L., Singhera, G. K., ... Sin, D. D. (2020). ACE-2 Expression in the Small Airway Epithelia of Smokers and COPD Patients: Implications for COVID-19. *European Respiratory Journal*.

[9] Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., ... Agha, R. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery*.