

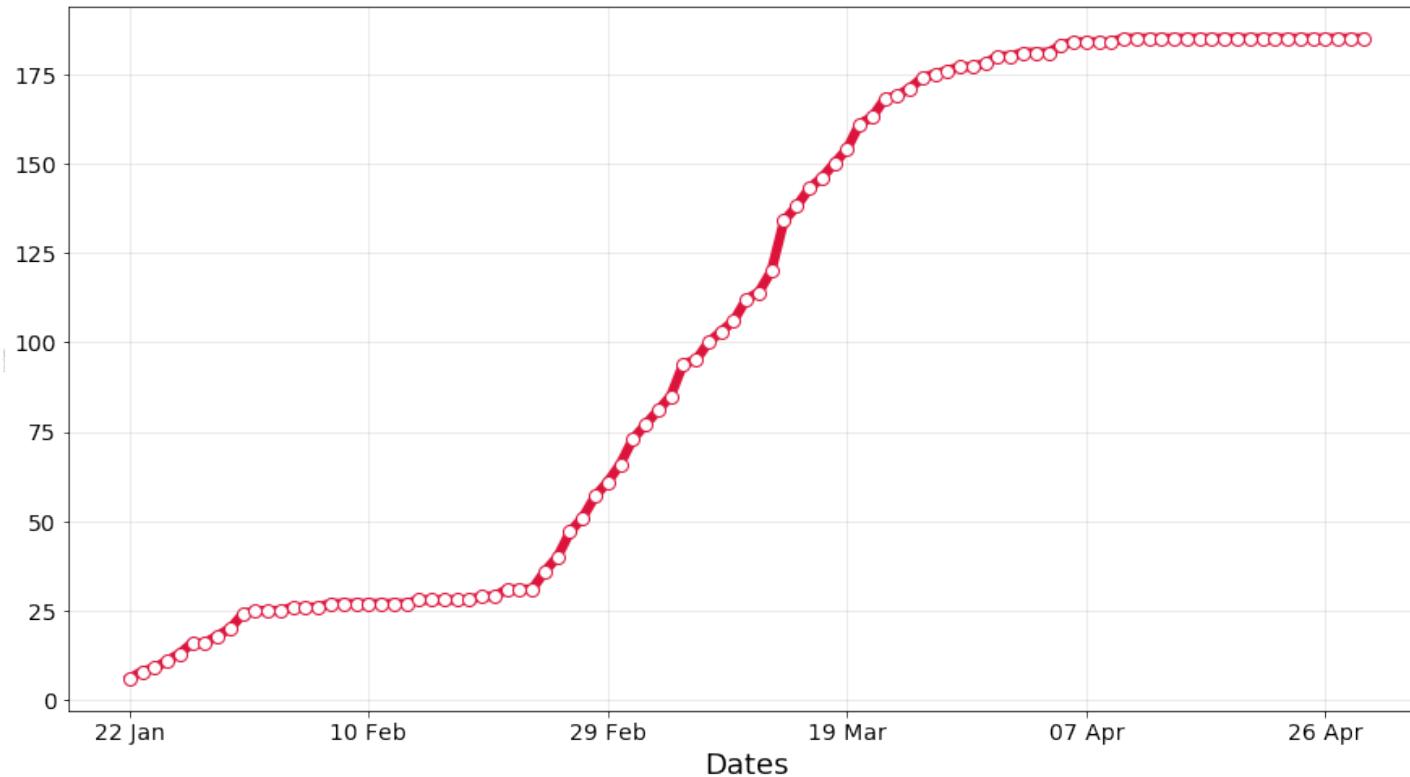
Analysis and Prediction Model of Global Pandemic: COVID-19

Anyu Zhu

zhu00428@umn.edu

- Visualization and Data Insight
- Development of Deep Learning Model for Prediction
- Some Thoughts

Number of Countries Affected

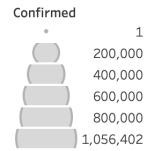


212 countries and territories

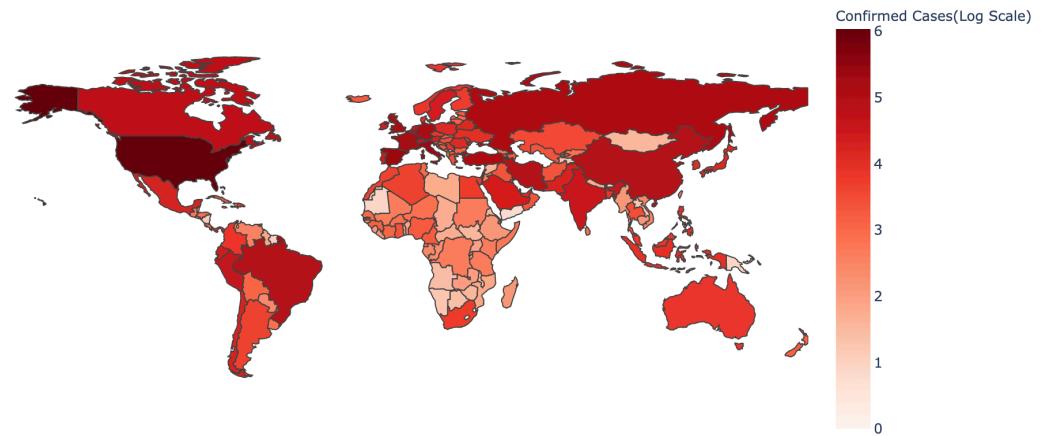
364 million + confirmed cases

Global Confirmed Cases (By April 30th)

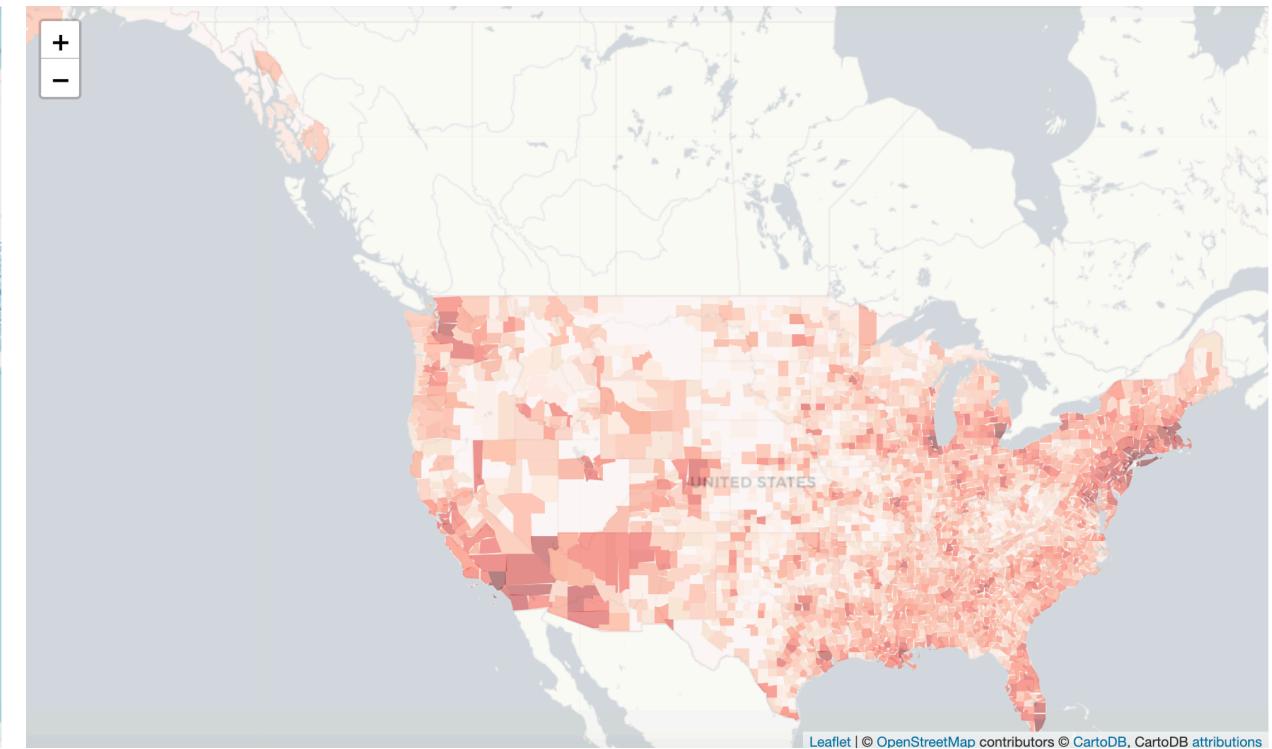
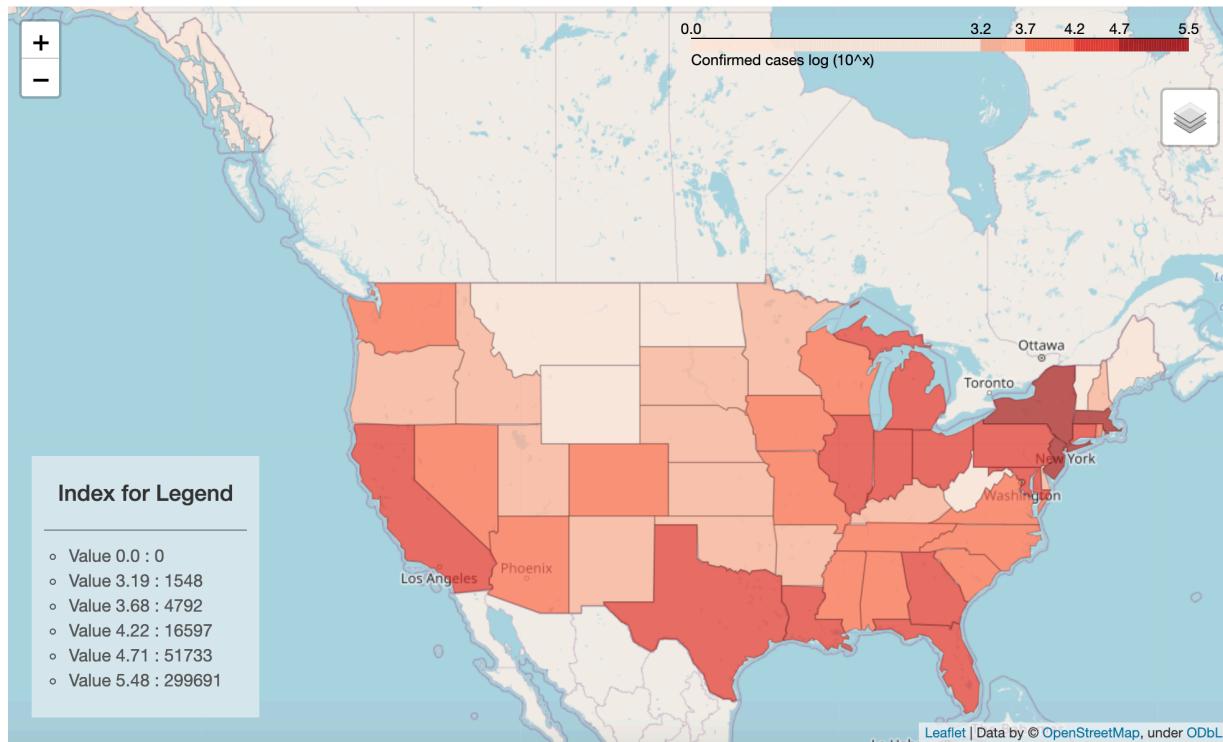
Global Confirmed Cases



Confirmed Cases Heat Map (Log Scale)

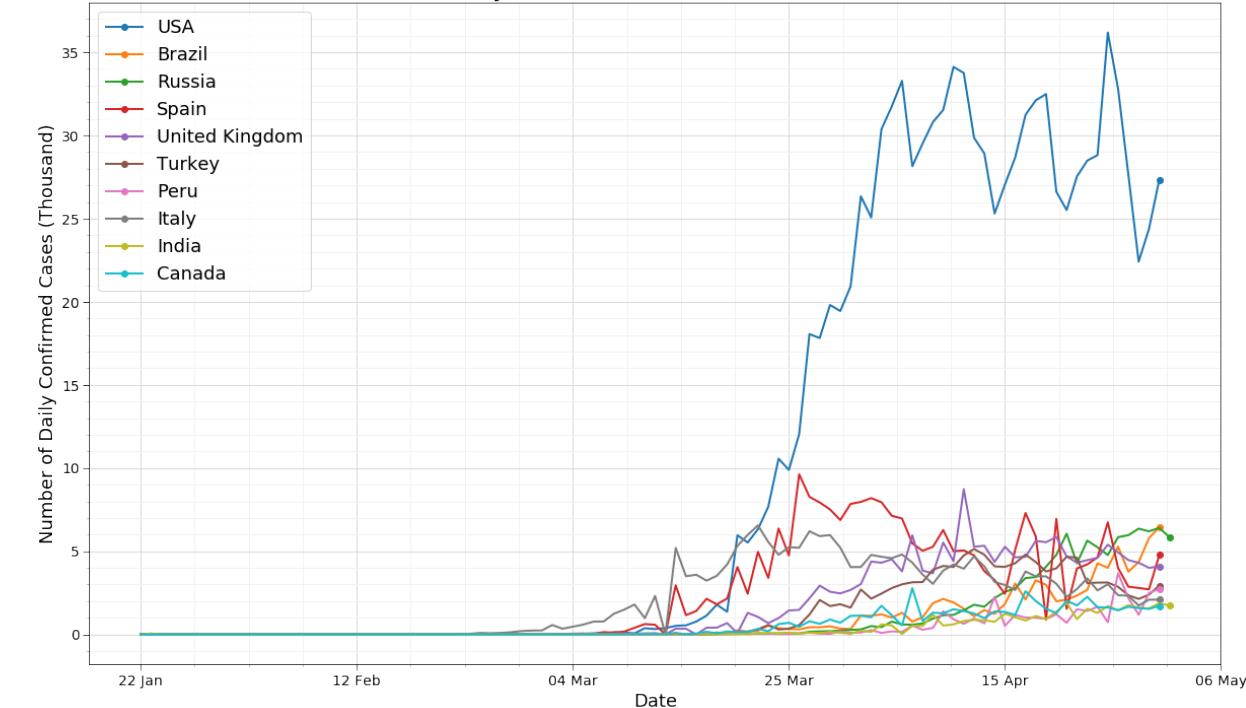


Heatmap of Confirmed Cases in the United States

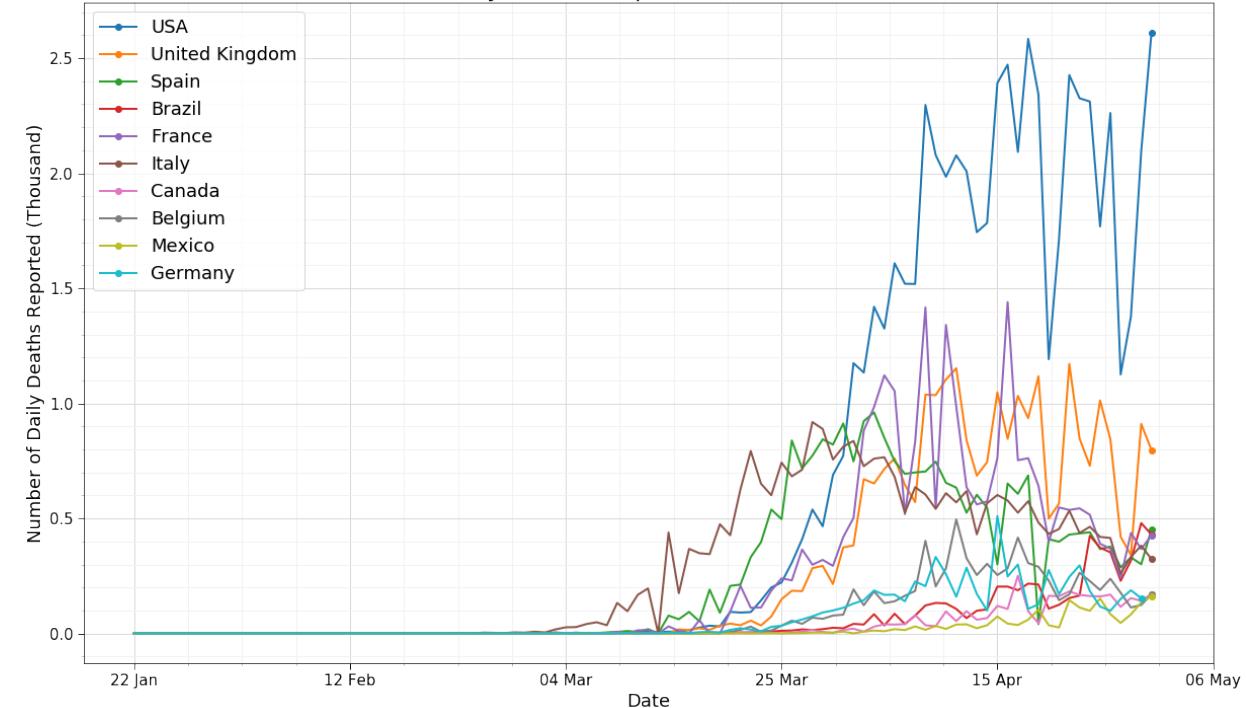


Developing Trend of Covid-19

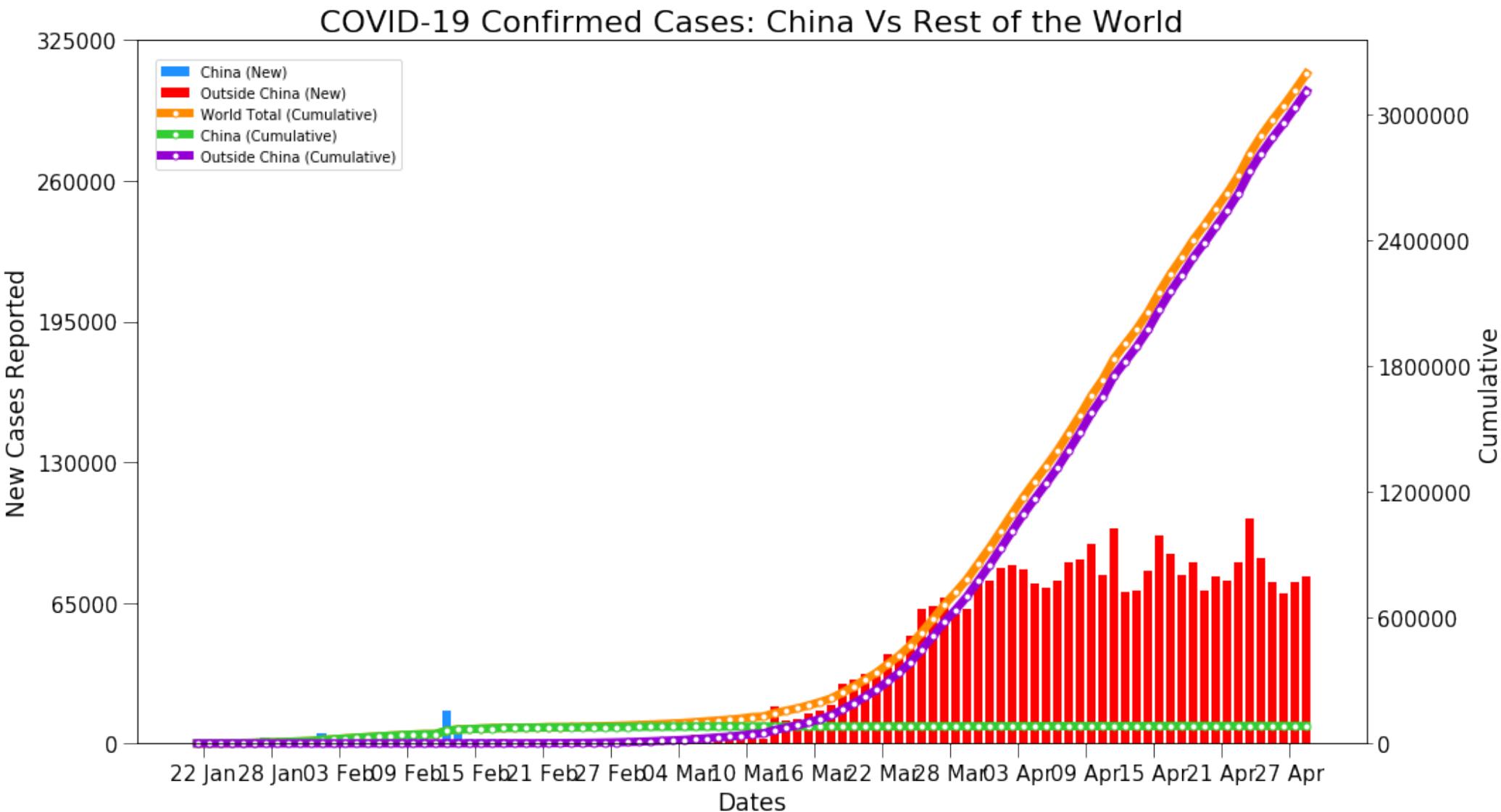
COVID-19 Daily Confirmed Cases in Different Countries



COVID-19 Daily Deaths Reported in Different Countries



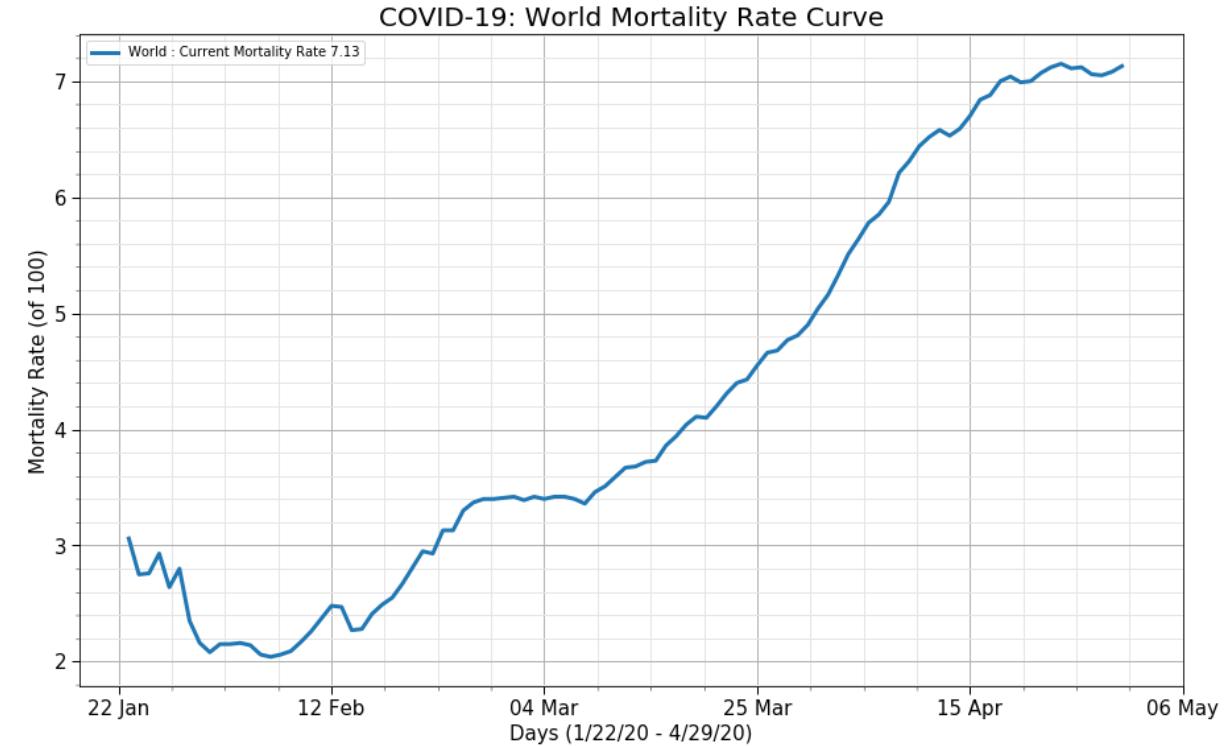
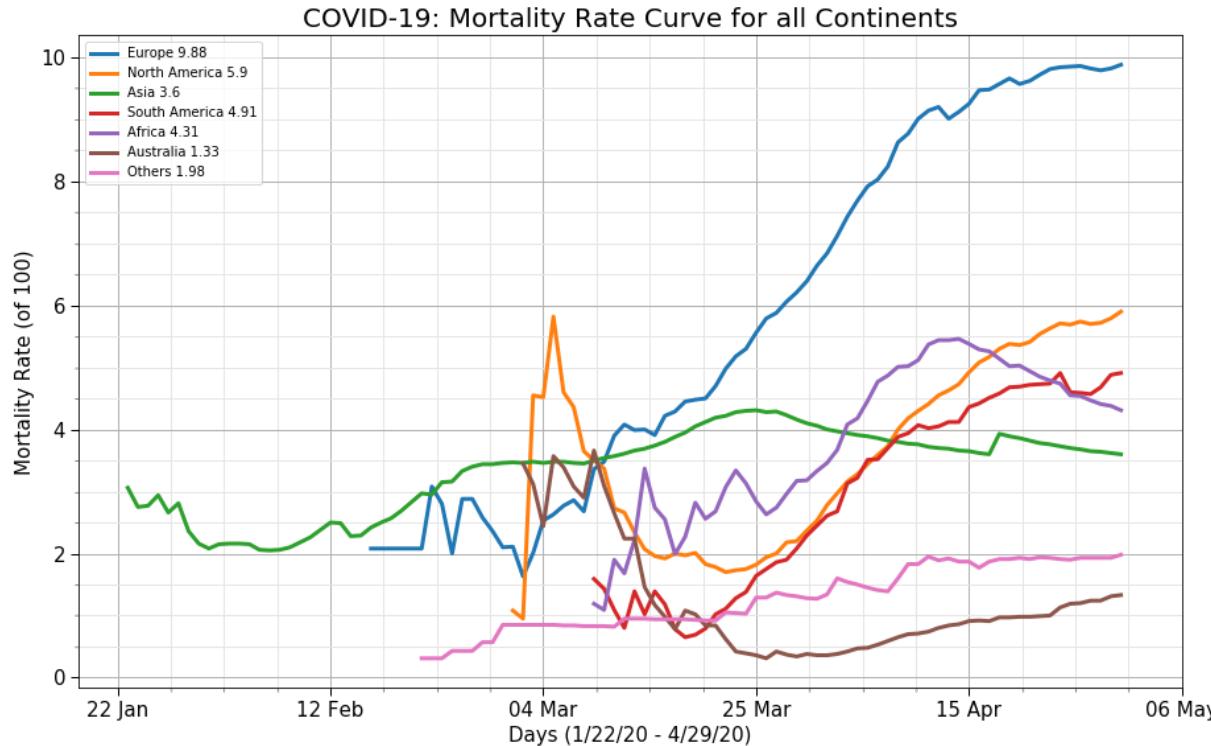
Cases in/outside China



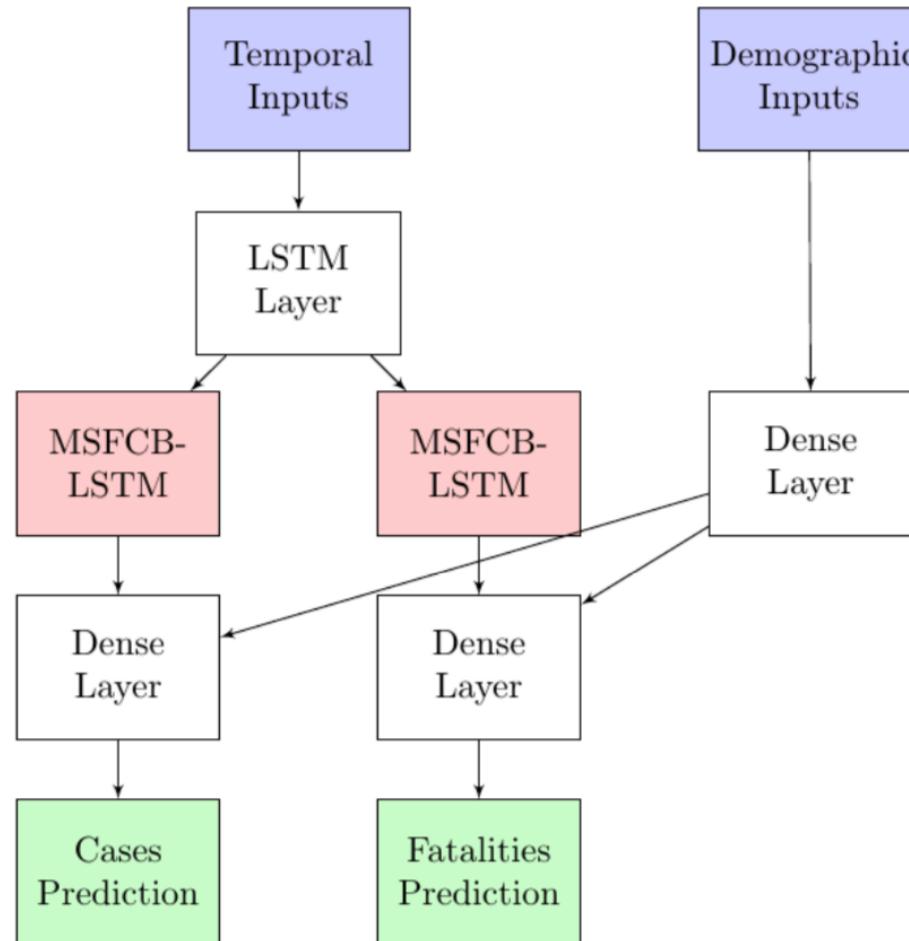
Analysis of Countries' Testing Ability

	country	Cumulative total	Cumulative total per million	confirmed	deaths	MR	Positive
0	Australia	261000	10276.3	6754	92	1.36	2.59
1	Austria	98343	11198	15452	584	3.78	15.71
2	Bahrain	37996	22380	3040	8	0.26	8
3	Belgium	62867	5410.25	48519	7594	15.65	77.18
4	Canada	256933	6832.74	53977	3268	6.05	21.01
5	France	224254	3412.2	166628	24410	14.65	74.3
6	Germany	918460	11127.4	162375	6563	4.04	17.68
7	India	26798	19.3739	34862	1154	3.31	130.09
8	India	47951	34.6668	34862	1154	3.31	72.7
9	Indonesia	7621	27.9954	10118	792	7.83	132.76
10	Italy	581232	9829.39	205463	27967	13.61	35.35
11	Japan	39446	311.837	13965	425	3.04	35.4
12	Malaysia	47723	1451.9	6002	102	1.7	12.58
13	Pakistan	30308	145.458	16473	361	2.19	54.35
14	Philippines	5265	47.993	8488	568	6.69	161.22
15	South Korea	443273	8606.08	10765	247	2.29	2.43
16	United Kingdom	173784	2580.92	172478	26842	15.56	99.25
17	USA	1267658	3824.8	1056402	61867	5.86	83.33

Covid-19 Mortality Rate



Using Hybrid Deep Learning Model to Predict



Temporal Features (In the past 13 days):

- Number of confirmed cases
- Number of fatalities
- Restrictions applied or not
- Quarantine applied or not
- Schools opened or closed

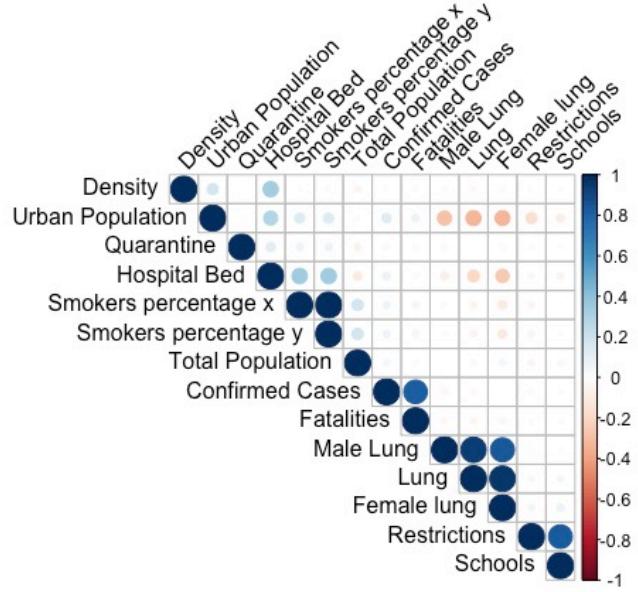
Atemporal Features:

- Population of each country
- Density of each country
- Number of hospital beds of each country
- Lung measurement of female and male in each country

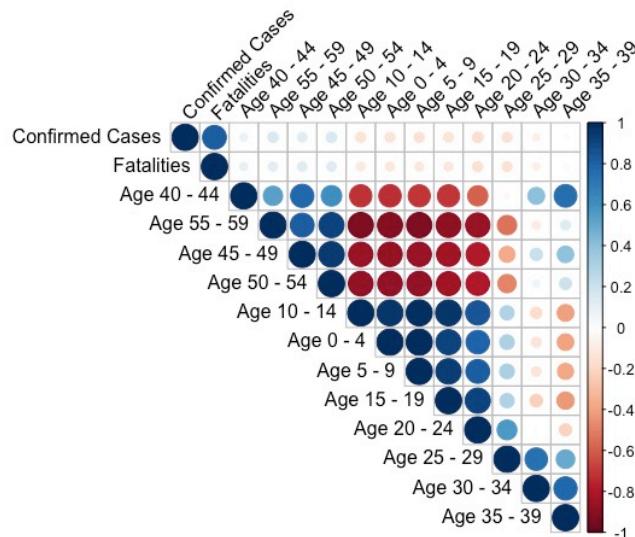
Model Output:

- Number of confirmed cases for the 14th day
- Number of fatalities for the 14th day

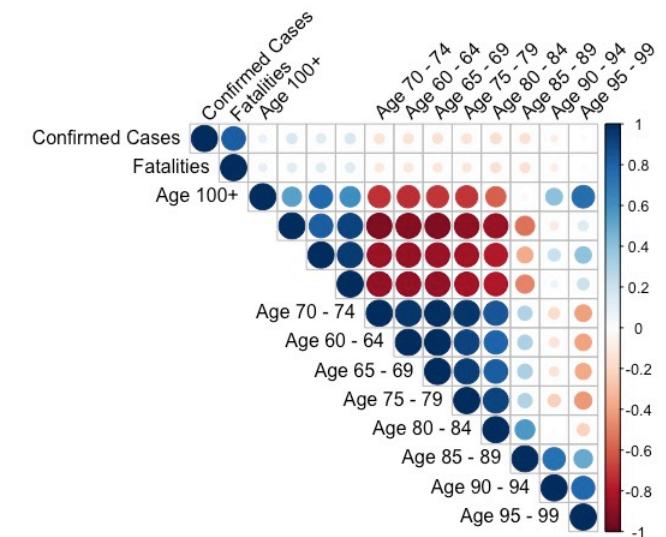
Data Preprocessing, Correlation Analysis



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



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



Correlation among Confirmed cases, fatalities and other features

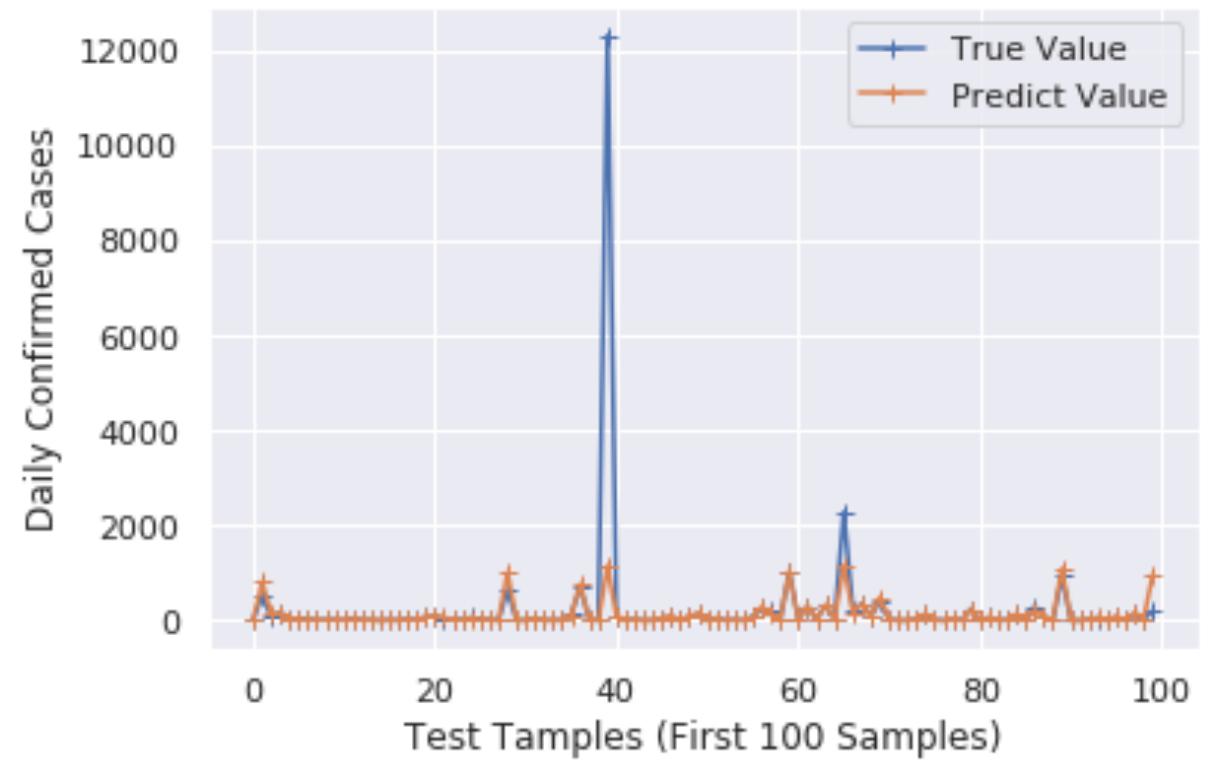
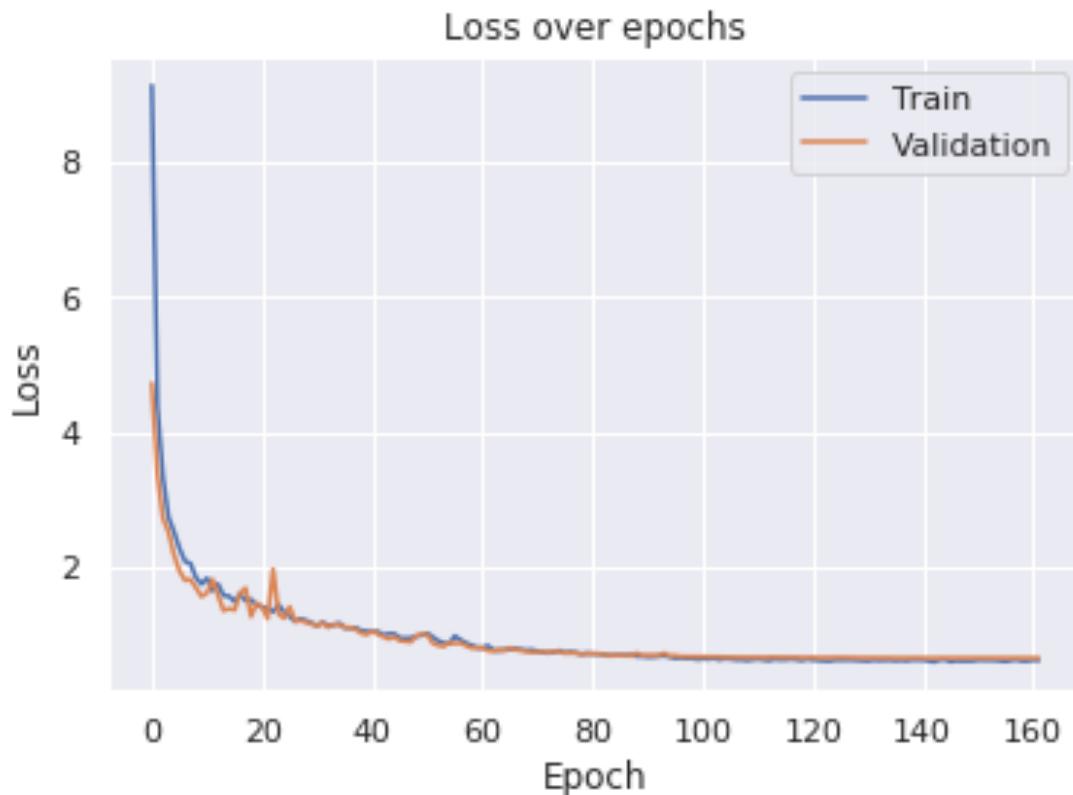
Model Evaluation

To evaluate the accuracy of the model:

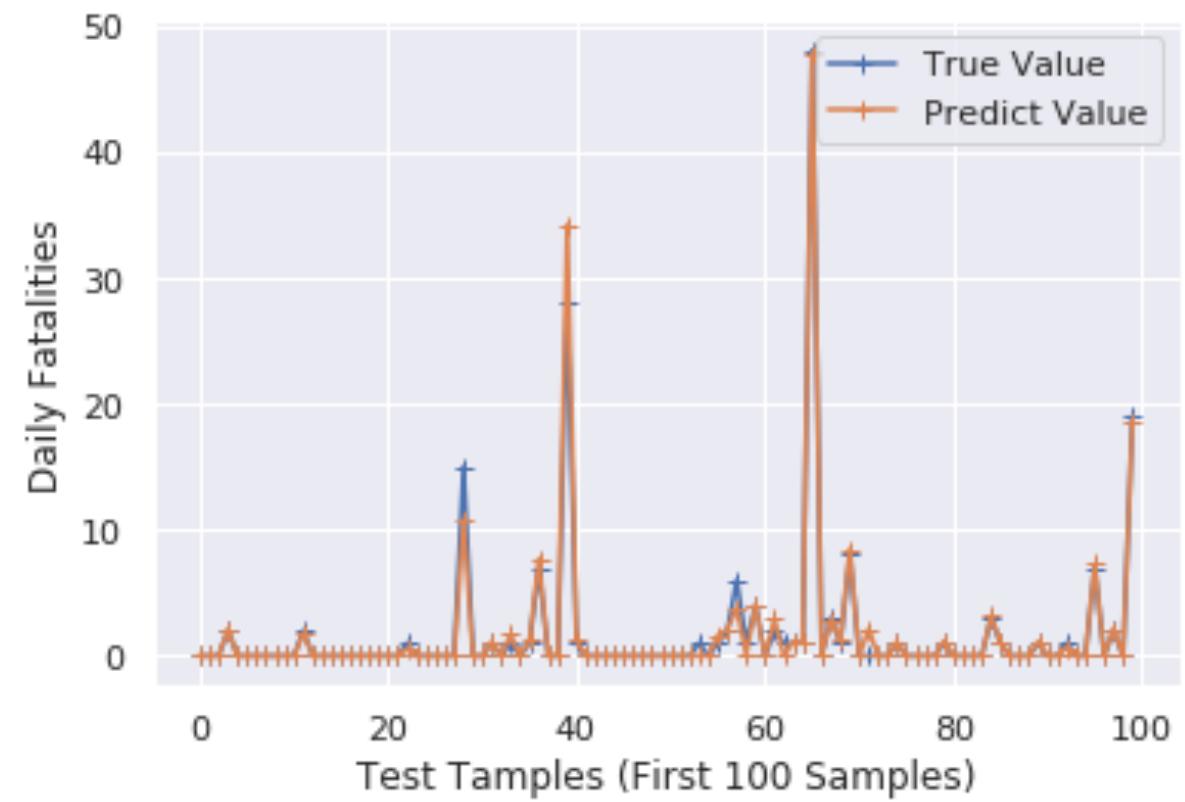
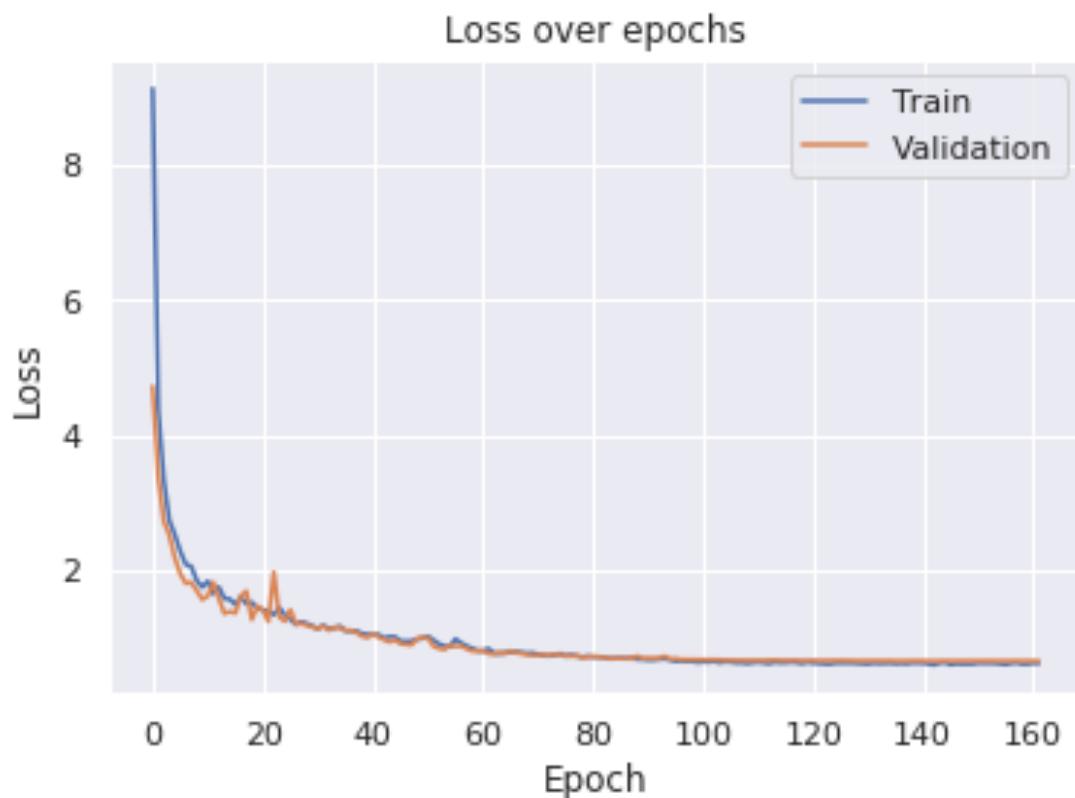
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (1)$$

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

Confirmed Cases Prediction



Fatalities Prediction



Thoughts

- Whether deep learning can be used for disease prediction? (Compared with SIR/SEIR...)
- How can it be applied?
- Flatten the curve/ End the curve?

Andrew Ng

I. The disease K_2 with $p(K_2|X_1) > 0.75$ and $I_2 < 0.9$
II. The disease K_2 with the highest $p(K_2|X_1)$ and $I_2 < 0.9$

Results

Using Set I criteria, the program could make a positive diagnosis in 120 patients, 125 of which are correct giving a diagnostic accuracy of 81.3%. In addition the NORMCHECK subroutine correctly diagnosed all the isolated hypochromic areas in the two nephrosities, the thrombocytopenia of the acute leukaemia, and the adolescent boy's raised plasma alkaline phosphatase. With the other diseases, however, the computer was not comparing these to the database. It did not make a definitive diagnosis in 14 cases, misdiagnosed one patient with primary biliary cirrhosis as cholangitis, and diagnosed 100 cases as primary hepatocellular carcinoma and another hepatocarcinoma as alcoholic liver disease.

Using Set II criteria, the program could make a positive diagnosis in 213 cases, 167 of which were correct, giving a diagnostic accuracy of 76.4%. It managed to exclude the 4 patients with secondary carcinomas of the liver, and excluded only 6 of the sixteen with diseases which were not in its database. It misdiagnosed all 6 patients with secondary carcinoma of the liver as primary carcinomas of the liver.

The comparative results of the diagnostic accuracy of the senior physicians and the computer are shown in table 2.

Table 2

	Physicians	Computer Set I criteria	Computer Set II criteria
Correct Diagnoses	26	36	39
Incorrect Diagnoses + no diagnosis	24	14	11
	$\chi^2 = 4.264$ $p = 0.039$	$\chi^2 = 7.629$ $p = 0.006$	

Discussion

Applying the stricter Set I criteria, the program was correct in 81% of the cases it diagnosed. It could reach a diagnosis, in addition, in 87% of the cases involving diseases which it had not been taught to recognise. It avoided committing itself to any diagnosis. With the Set II criteria, it could reach a diagnosis in 76% of the cases but the accuracy dropped to 70%. But in either case, it performed better in interpreting liver function tests in conjunction with the patient's age and sex than did the senior physician. Its level of accuracy is comparable to that reported in the literature using other systems,^{4,5} even when it had the added difficulty of having to interpret data from cases involving diseases which were not in its database.

If $TP =$ number of true positive
 $TN =$ number of true negative
 $FP =$ number of false positive
 $FN =$ number of false negative

then the specificity of a test = $\frac{TP}{TP + FN}$
sensitivity of a test = $\frac{TP}{TP + FP}$
and the diagnostic efficiency = $\frac{TP + TN}{TP + TN + FP + FN}$

If one just considers the class of surgical liver diseases as cholangitis, stones in the biliary tree and liver disease, with the remaining diseases as non-surgical liver diseases, the diagnostic efficiency of the program for surgical vs non-surgical liver disease is 96%, with a specificity of 75% and sensitivity of 86%, using Set I criteria. This compares with 75% diagnostic efficiency of diseases which could not be diagnosed by the computer as the non-surgical liver diseases which were not necessarily labelled as surgical liver disease at 78%.

The computer is not meant to replace the physician, but with its more correct interpretation of laboratory data, it could help sharpen the physician's clinical judgment. It is therefore meant to aid the physicians in the diagnostic process.

Meaningful Computer-Aided Diagnosis can be implemented on microcomputers and will prove software. You can build your own in the institutions and tell the doctors' offices.

The authors wish to thank Dr. W.C. Liang for help in collecting the data, to Prof. Atchison for his invaluable suggestions, and to Dr. C.L. Lai for his helpful criticism.

References

1. Taylor, T.R.: Computer-guided Diagnosis. J. Roy. Coll. Physcs., Lond. 4 : 188-190, 1970.
2. De Souza, F.: Biomedical Computing, ed. W.J. Partlow, pp. 301-306, Pitman Medical, England, 1977.
3. Horrocks, J.C., McAdam, W.A.F., Devroede, G.,

The computer is not meant to replace the physician, but with its more correct interpretation of laboratory data, it could help sharpen the physician's clinical judgment. It is therefore meant to aid the physicians in the diagnostic process.