

Understanding evolution of COVID-19 driven mortality rate

Ishika Bhaumik¹, Suman Sinha-Ray^{2,3*}, Anshul Chaudhary¹, Abhishek Srivastava^{4**}, Prashant Kodgire⁵

¹Department of Computer Science & Engineering, Kalinga Institute of Industrial Technology, Odisha, India, ²Department of Mechanical & Industrial Engineering, University of Illinois at Chicago, Chicago, IL, USA, ³University Space Research Association, NASA Glenn Research Center, Cleveland, OH, USA, ⁴Department of Computer Science & Engineering, IIT Indore, MP, India, and ⁵Department of Biosciences & Biomedical Engineering, IIT Indore, MP, India

Corresponding Author: *Suman Sinha-Ray, University Space Research Association, NASA Glenn Research Center, Cleveland, OH, USA; suman.sinharay@nasa.gov/ ssinha6@uic.edu ;

**Abhishek Srivastava, Department of Computer Science & Engineering, IIT Indore, MP, India; asrivastava@iiti.ac.in

ABSTRACT

Objective: COVID-19 has resulted in the death of almost 4 million people till date¹. However, the mortality rate across countries seems to be vastly different irrespective of their respective socio-economic backgrounds. It is well known now that COVID-19 is an acute inflammatory infectious disease that gets complicated by type-I interferon response^{2,3}. However, the precise reason for variations in COVID-19 related mortality rates is unknown. A detailed understanding behind the evolution of mortality rate around the globe is needed.

Methods: In this article, we show that a biological science guided machine learning-based approach can predict the evolution of mortality rates across countries. We collected the publicly

available data of all the countries in the world with regard to the mortality rate and the relevant biological and socio-economical causes. The data was analyzed using a novel FFT driven machine learning algorithm.

Results: Our results demonstrate how COVID-19 related mortality rate is closely dependent on a multitude of socio-economic factors (population density, GDP per capita, global health index and population above 65 years of age), environmental (PM2.5 air pollution) and lifestyle *aka* food habits (meat consumption per capita, alcohol consumption per capita, dairy product consumption per capita and sugar consumption per capita). Interestingly, we found that individually these parameters show no visible trend that can be generalized with mortality.

Conclusions: We anticipate that our work will initiate conversations between health officials, policymakers and world leaders towards providing preventative measures against COVID-19 and future coronavirus-based diseases and endemics/ pandemics by taking a holistic view.

Key words: COVID-19, mortality rate, machine learning, prediction, FFT, inflammation.

INTRODUCTION

Coronaviruses are medium-sized RNA viruses with a characteristic crown-shaped structure. The genesis of the coronavirus dates back to 10,000 years ago⁴. However, the first known detection of coronavirus dates back to 1965⁵. While coronavirus is mainly present in bats and birds, it falls in the zoonotic category which comprises viruses that are transmitted from animals to humans from time to time. While COVID-19 seems to be extremely lethal, the surprising aspect of COVID-19 related deaths is the significant variation in mortality rates across the globe. As an example, at the height of the pandemic, the mortality rates in the USA, UK, South Korea, Kenya, and India in the month of May 2020 were of the order of 6.2%, 15.0%, 2.4%, 5.6% and 3.2%,

respectively. On 2nd January, 2021, the same figures were at 1.7%, 2.9%, 1.5%, 1.7% and 1.4%, respectively¹. Researchers have put in substantial effort to explain this variation. In a recent article, it was demonstrated that the prevalence of autoimmune diseases can increase the risk of COVID-19 related fatality 1.21-1.31 folds⁶. Significant work has been done to understand the impact of hazardous air pollutants, namely: PM2.5, ozone, nitrogen dioxide⁷⁻⁹ on the fatality rates associated with COVID-19. Additionally, the effect of diet on fatality rates in European countries has also been studied. It is hypothesized that it stems from variations in angiotensin-converting enzyme activity¹⁰. All these endeavours clearly indicate that multiple factors affect the mortality rate associated with COVID-19. What is missing, however, is a study of the overall interdependence of various factors that influence the mortality rates associated with COVID-19. Furthermore, most studies are limited to a region and do not provide a complete picture.

In Figure 1, the mortality rate of countries across 5 different continents are elucidated, where Figure 1 a-e elucidates the mortality rate in a few countries in Africa (Nigeria, South Africa, Morocco and Kenya), Asia (India, South Korea, Japan and Singapore), Europe (UK, Italy, Germany and France), South America (Brazil, Chile, Peru and Argentina) and North America (USA, Canada and Mexico) respectively between March 15, 2020 – March 15, 2021¹. Mortality rate (called as Case Fatality Rate in Ref. 1) is defined as the following-

$$\text{Mortality rate} = \frac{\text{Number of deaths from COVID-19}}{\text{Number of diagnosed cases of COVID-19}} \times 100\% \quad (1)$$

While such definition has its inherent challenges¹, we believe that this is the best way to normalize the data and compare between countries of different sizes and populations. Figure 1 brings in two interesting facts. First, as can be seen, almost all the countries follow a very similar trend in the evolution of mortality rates across time. While mortality rates during the initial days were the highest, they subsequently went down and plateaued with a significant exception in

South Africa, which perhaps can be attributed to the surge of the Beta (501Y.V2) variant¹¹. This simple graph clearly shows that during the early days of the pandemic, as there was a significant lack of knowledge base on the coronavirus, the mortality rate jumped significantly. As the knowledge grew, best practices (e.g.- hygiene and sanitation, vaccinations, travel restrictions) were practiced throughout the globe and the mortality rate dropped. However, this also resulted in countrywide lockdowns resulting in economic hardships across the globe. This also hints that in future for any other novel coronavirus related pandemic, the mortality rate can be similarly higher. Secondly, it can also be seen that the mortality rates across different countries are largely different. The most interesting aspect is that the mortality rates in several rich nations are higher than their poorer counterparts. All these beg the question as to the reason behind the driver of the mortality rate. It can be seen that owing to the variation of exposure rate, the COVID-19 related mortality rate peaked at different times. Additionally, it can also be seen that for some countries, the mortality rate started dropping in early 2021. However, that drop was not uniform across the world. This can be attributed to the uneven distribution of vaccines across the globe. To avoid these variations and spikes, our analysis was limited to the period between 15th May 2020 and 15th January 2021.

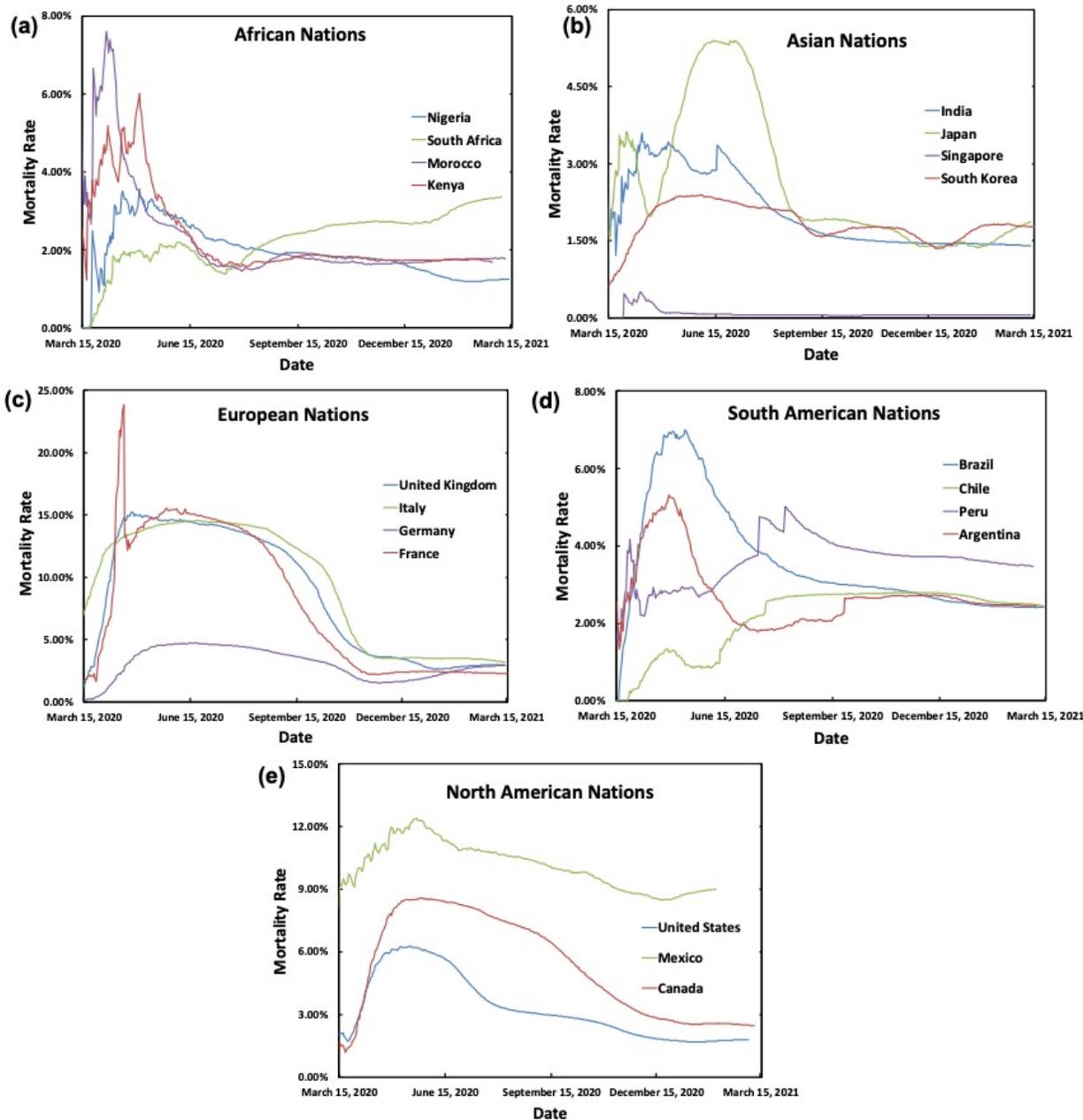


Fig. 1. The variation of mortality rate across nations in various continents over 1 year period (March 15, 2020- March 15, 2021) are shown – (a) African (Nigeria, South Africa, Morocco and Kenya), (b) Asian (India, Japan, South Korea and Singapore), (c) European (UK, Italy, Germany and France), (d) South American (Brazil, Chile, Peru and Argentina) and (e) North American (USA, Canada and Mexico) nations.

MATERIALS AND METHODS

Determination of the Influencing Factors

As COVID-19 is an inflammatory disease^{2,3}, our primary hypothesis at the beginning of this work was that any factor that creates an inflammatory response in the human body will be accentuated by COVID-19 resulting in an increased mortality rate. The variation of the impact of COVID-19 on the mortality rate can be ascribed to long term influencers that create heightened inflammation in the body. In addition to that, recently there has been some interesting work that clearly show that COVID-19 is severely correlated to gut microbiome¹². While choosing the factors, our hypothesis also evolved around the fact that if the factors also negatively impact the gut microbiome, it will cause severity of complications by COVID-19 and thus increase the mortality rate. It also needs to be mentioned that our approach was to utilize the minimum number of influencing factors to correlate with the mortality rate.

One of the most established factors in COVID-19 related deaths is the ageing population¹³. While that is a significant factor, it is worth noting that as of 2020 while France has 20.75% of the population with an age of 65+, Japan has 28.4% of their population with an age of 65+¹. The mortality rate plot in Figures 1b and 1c clearly shows that France had a significantly higher mortality rate than Japan. PM 2.5 also has a significant impact on the mortality rate associated with COVID-19⁷⁻⁹. Such a dependence is possible because as it was found that PM 2.5 can be associated with enhanced inflammatory markers resulting in diabetes, obesity, hypertension prevalence in patients¹⁴⁻¹⁷. Another popular aspect to the discrepancy in mortality rate was ascribed to access to health care, sanitation, hygiene, higher population density and GDP. Better hygiene and safe sanitation practices in high GDP countries are also associated with a higher prevalence of autoimmune disorders, diabetes *aka* higher inflammatory response¹⁸.

Additionally, high GDP results in a sedentary lifestyle causing a multitude of health issues. One of the most controversial yet significant reasons behind enhanced inflammatory response in the human population are dietary habits^{19,20}. These studies characterized the role of food with saturated fat and milk with disorders like Multiple Sclerosis (MS) and diabetes. Interestingly, this study reported that whole food, plant-based or high fiber diets help in controlling diabetes-related complications and consumption of red meat developed evidence of systemic inflammation and thereby contributed to a higher incidence of type 2 diabetes and atherosclerosis. Similarly, heavy alcohol consumption demonstrated higher C-reactive protein (CRP) concentrations that is an important systemic marker of inflammation²¹. With regard to the correlation between the gut microbiome and the above-mentioned factors, it was found that the factors not only negatively impact inflammation, they also negatively impact the gut microbiome²²⁻²⁶. Based on all these, we hypothesized that long term exposure of the following - socio-economic factors (population density, GDP per capita, global health index and population above 65 years of age), environmental (PM2.5 air pollution) and lifestyle *aka* food habits (meat consumption per capita, alcohol consumption per capita, dairy product consumption per capita and sugar consumption per capita), drives the mortality rate.

Overview of the Scientific Model

A schematic explaining the overview of the scientific model is shown in Figure 2. The overall evolution of mortality rate (Figure 1) for all the countries was made to undergo an FFT (Fast Fourier Transformation) yielding the Fourier coefficients (a_n 's and b_n 's). Machine learning-based models were trained to correlate the Fourier coefficients with the above-mentioned influencing factors. From the training data set, Fourier coefficients of the validation dataset were predicted, and these underwent an IFFT (Inverse Fast Fourier Transformation) to predict the

mortality rate. The mortality rates so predicted were compared with the actual mortality rates and the efficacy of the approach was assessed. To the best of our knowledge, this process of solving an epidemiological model is novel. This idea of solving complex evolution in the Fourier space is more common in understanding turbulence²⁷ and the frequency of oscillation of a heating wire in boiling heat transfer²⁸. This method is one of the novel aspects of our present work and is described in detail in the following subsections.

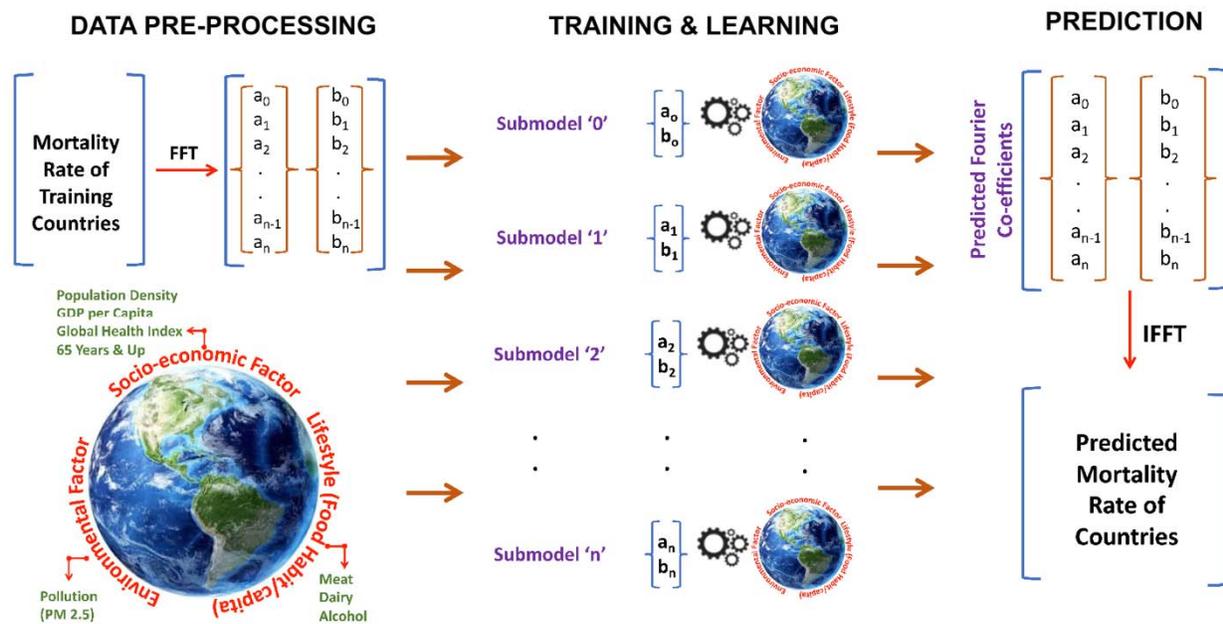


Figure 2. An overview of the scientific model is shown. The model has 3 distinct parts- data pre-processing, training & learning and prediction.

Data Collection

The influencing factors were obtained from publicly available data from reliable sources¹. While these data provide an overall view of a country, it lacks the granular details. This lack of granularity may cause some variation in the accuracy of prediction. However, our overarching hypothesis was that such general data will still be able to provide the predictive capability. The data for the number of COVID-19 cases, number of deaths for the aforementioned period is

collected for almost all countries in the world. In addition to this, data corresponding to the parameters mentioned is collected for all countries for the last ten years. Algorithm 1 summarises our data collection efforts. The number of covid cases per million for each country and the number of deaths due to covid per million for each country is calculated defined as Algorithm 1 is shown in Figure 3a.

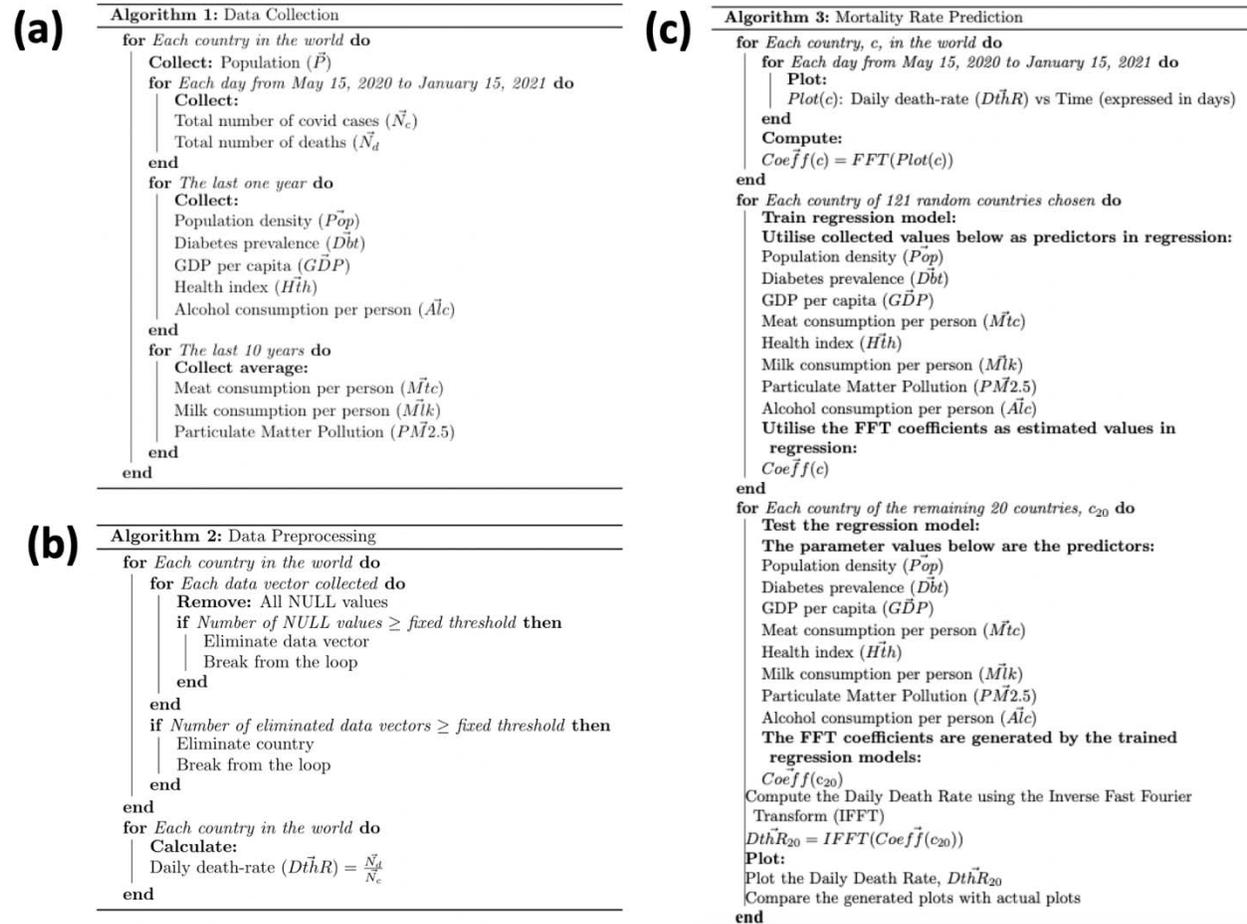


Figure 3. (a) Data collection methodology is shown (Algorithm 1). (b) Data pre-processing methodology is shown (Algorithm 2) . (c) The machine learning model is shown (Algorithm 3).

Data Pre-processing

While the publicly available data is comprehensive, in some cases it lacked the details. In order to utilize the ‘dirty’ data, it need to be ‘cleaned up’ first to be effective in providing useful information. We utilised the Pandas software library for proper consolidation of data. The data is mostly available in a scattered manner in a CSV format. We consolidated data for each of the parameters in such a way that all data for a specific parameter for different countries is gathered together. In case the data corresponding to a certain country is missing or is NULL, the parameter value is eliminated. In case the number of NULL values corresponding to a country exceeds a certain threshold, the country is eliminated and is not considered for the analysis. The data for the parameters in a country-wise manner is put together as a data frame. Subsequent to the elimination of the NULL values and/or countries with a large number of NULL values for their respective parameters, the death rate for each country is calculated as the ratio of the number of deaths per million to the number of COVID-19 cases per million of the respective country. Algorithm 2 summarizing the data pre-processing is shown in Figure 3b.

Development of the Machine Learning Model

In this work, we focused on predicting the mortality rate trend between 15th May 2020- 14th February, 2021. Instead of looking at daily fluctuation, which can stem from a multitude of nation-specific causes, the overall long-term trend provides a holistic view. The above-mentioned pre-processing step resulted in 141 countries. Of these, the machine learning model was trained on 121 and rest 20 were chosen as the validation set. From Fig. 1 of the main text, it can be seen that the evolution of the mortality rate is random. In order to explain the evolution in a mathematically closed-form solution, FFT (Fast Fourier Transformation) was used. In FFT, the functional representation of the evolution can be explained as

$$f(x) = \sum_{n=0}^m c_n e^{inx} \quad (1)$$

, where $c_n = a_n + ib_n$, where a_n and b_n are Fourier coefficients and m is the number of days. Then the influencing factor(s) of the training dataset were trained against the Fourier Coefficients of the corresponding training data using gradient boosting regressor with a learning rate of 0.001 and a value of $n_estimators$ (number of regression trees) as 100. For ‘ m ’ Fourier coefficients, ‘ m ’ submodels were employed to train the algorithm. Subsequent to training the regressor with data for 121 countries, the same is utilised to predict the FFT coefficients for the 20 countries in the test set using the respective parameter values. The Fourier coefficients so generated are subsequently made to go through an Inverse Fast Fourier Transformation (IFFT) yielding the mortality rate of the countries in the training dataset, which was then compared with the actual mortality rate data. Algorithm 3 summarizing our approach for the prediction of the mortality rates is shown in Figure 3b.

RESULTS AND DISCUSSION

Prediction of Mortality Rate Using Single Parameter

Using the above-mentioned protocol, as shown in Figures 2 and 3, investigations were first conducted to predict/ explain the trend in COVID-19 related mortality rates incorporating just one of the influencing factors. A comparison between the USA and Slovenia is shown in Figure 4. Some of the other countries’ data is shown in Figure S1- S10. It is clear that a single influencing factor cannot explain the mortality rates across countries. Figure 4 shows that for Slovenia, alcohol consumption and global health index seem to predict the mortality rates close to the actual data (Figure 4a and 4d), for the USA there seems to be no trend for these two factors. For the USA it seems that GDP per capita and PM 2.5 (Figure 4c and 4g) closely predict

the mortality rate. This is one of the challenges we found with the results in existing literature, where correlations were drawn using a small subset of data. In order to alleviate this, the protocol shown in Figures 2 and 3 were followed and a comprehensive model utilizing all the factors together was developed.

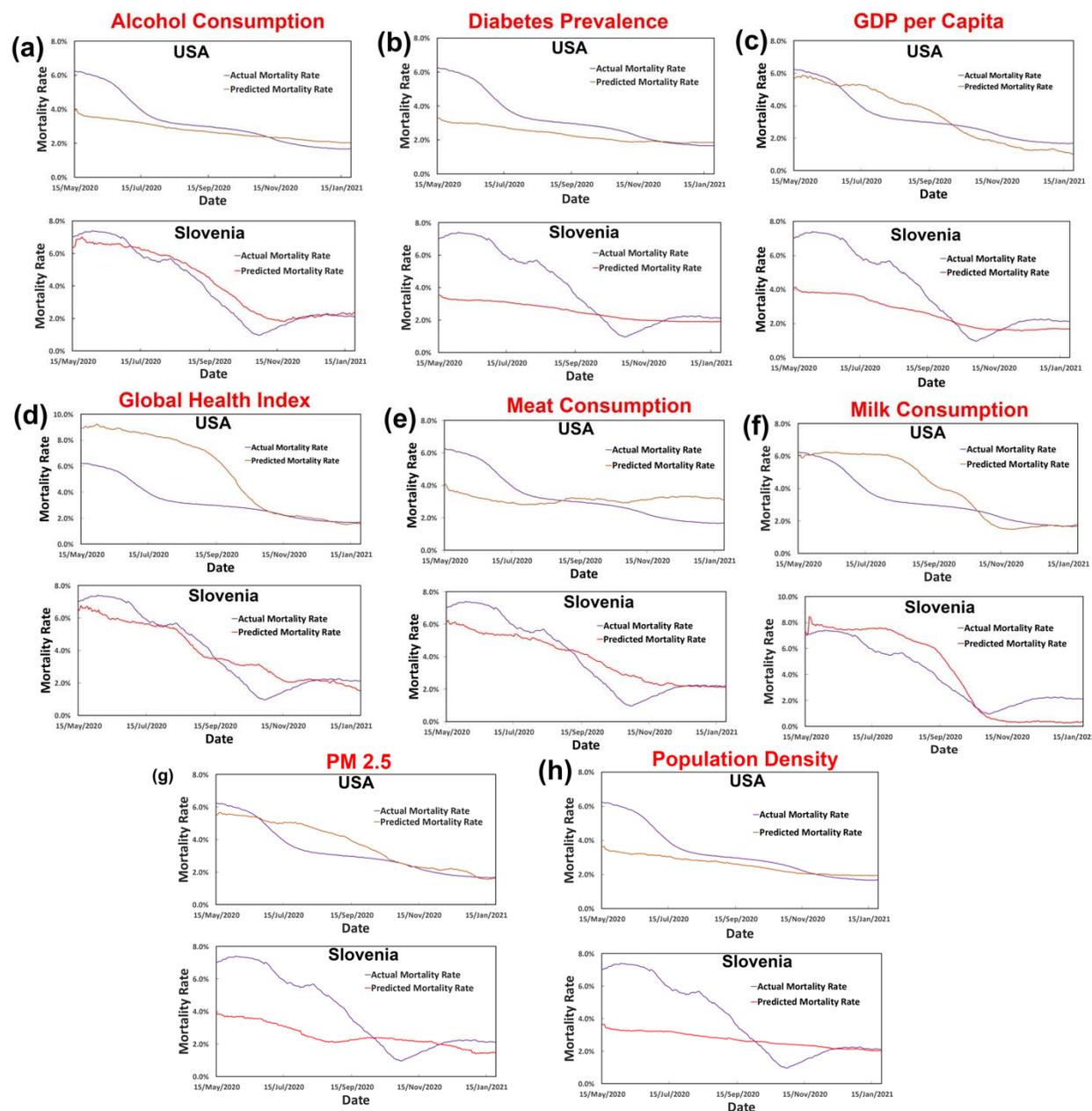


Figure 4. Prediction of COVID-19 related mortality rate using a single factor for both USA and Slovenia are shown, where the factors are – (a) alcohol consumption, (b) diabetes prevalence, (c) GDP per capita, (d) global health index, (e) meat consumption, (f) milk consumption, (g) PM 2.5 and (h) population density. None of the factors individually can describe the trend comprehensively.

Prediction of Mortality Rate Using 8 Crucial Parameters

Using the proposed protocol (Figures 2 and 3), the mortality rates of 20 randomly selected countries were predicted and compared with the actual mortality rates. Results of 6 of the countries are shown in Figure 5 of the main text and the results of some of the remaining countries are shown in Figure S11 in the supplementary information. Figure 5a-f corresponds to the USA, Germany, Honduras, Slovenia, Togo, and Brazil respectively. It is clear that for all the countries, the predicted mortality rates are very close to the actual mortality rates. The variations that do exist stem from multiple factors- (a) the quality of COVID data has challenges; (b) the publicly available data used in the prediction lacks granularity and specificity; and (c) the physiological difference (e.g.- blood type) may also potentially result in significant variations in mortality rate^{29,30}. Such granular information is practically impossible to obtain for the world population. As a result of which the prediction has an ingrained challenge in accurately predicting the mortality rates. However, the interesting and heartening fact is that irrespective of the paucity of granular data, our predictions not only show the impact of the 8 influencing factors on mortality rates but are also able to precisely trace the evolution and trend of the mortality rates across the world population. This clearly holds potential to positively impact the preparations for future coronavirus related outbreaks. The large message is that policymakers,

government officials and scientists need to take a holistic approach towards the healthcare response.

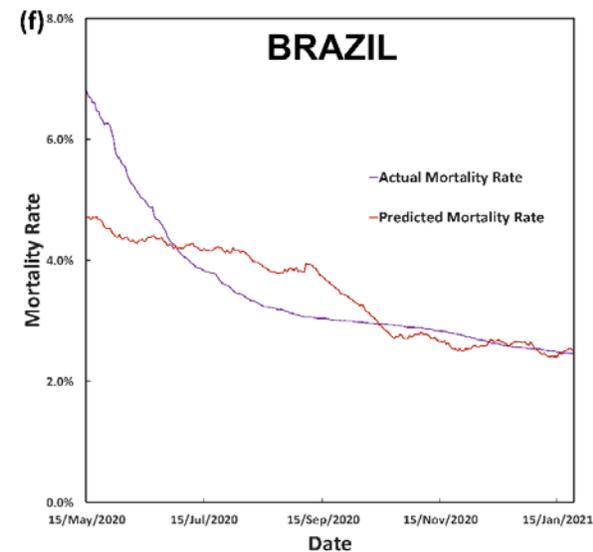
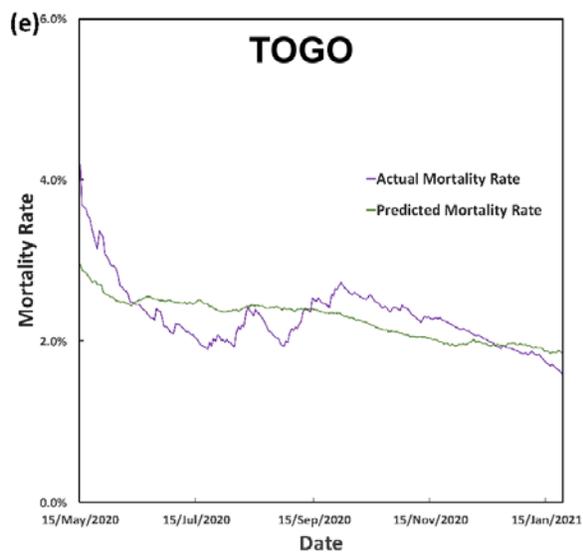
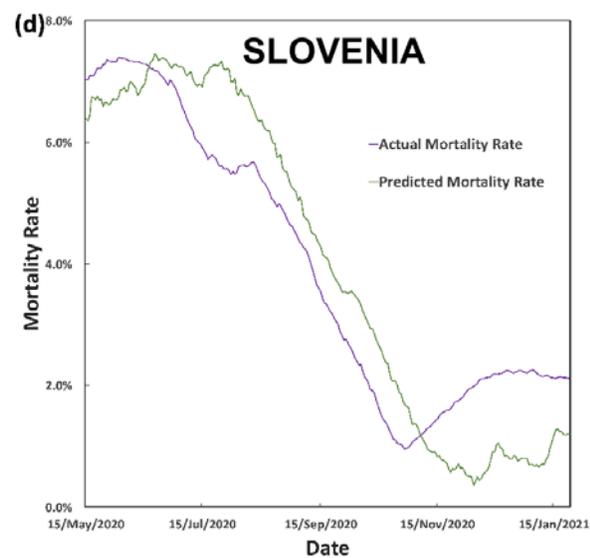
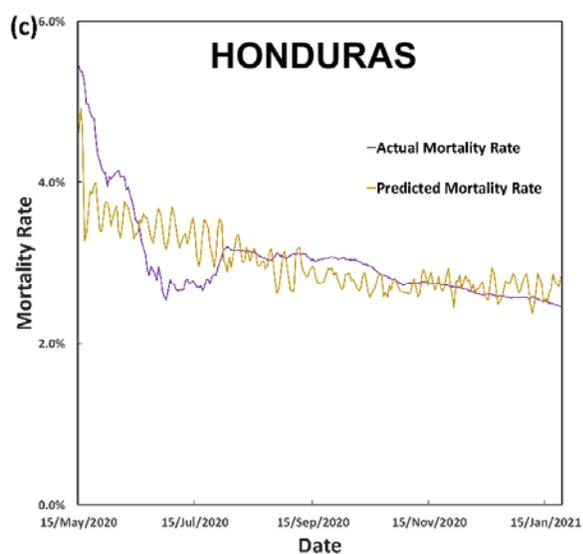
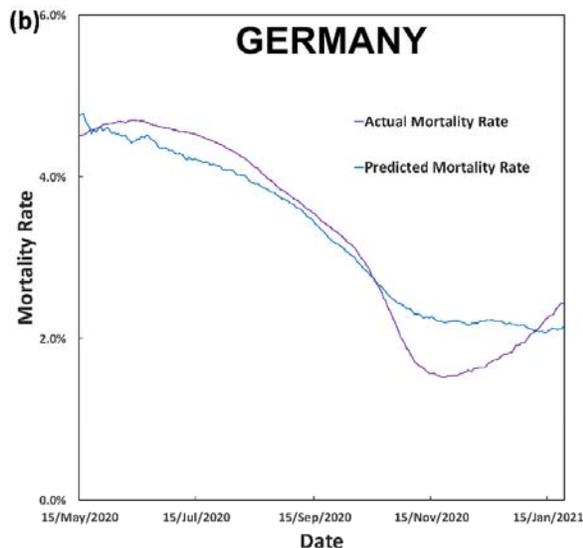
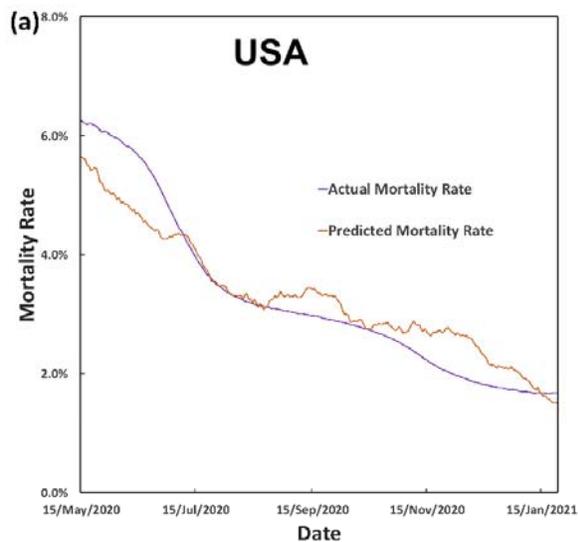


Figure 5. Prediction of COVID-19 related mortality rate using 8 factors and their comparison with actual mortality rates are shown for –(a) USA, (b) Germany, (c) Honduras, (d) Slovenia, (e) Togo and (f) Brazil are shown. It can be seen that the model is capable of predicting the evolution of mortality rate quite well.

CONCLUSIONS

This work utilizes a novel FFT guided machine learning methodology to demonstrate that COVID-19 related mortality can be broadly related to multitude of socio-economic factors (population density, GDP per capita, global health index and population above 65 years of age), environmental (PM2.5 air pollution) and lifestyle *aka* food habits (meat consumption per capita, alcohol consumption per capita, dairy product consumption per capita and sugar consumption per capita). Owing to the lack of granularity of individual level data and different country level introduction of security protocol during pandemic (e.g.- imposition of lockdown etc.), the prediction may not exactly fit to the actual mortality rate data but it closely elucidates the trend. This clearly shows that the policymakers, health professionals and government need to take a long-term view to protect against any future coronavirus-mediated pandemic. Also, we believe that further study building onto this one will also allow the policymakers to take actions more proactively in future.

AUTHOR CONTRIBUTIONS

I.B. carried out the model development. S.S.R. conceptualized the idea and directed the research. A.C. collected and prepared the data for analysis. A.S. directed the research. P.K. contributed to

identifying the biological significance of the factors. All authors contributed to the analysis of the results and the writing of the manuscript.

SUPPLEMENTARY MATERIAL

Supplementary material is available at Journal of the American Medical Informatics Association online.

CONFLICT OF INTEREST STATEMENT

The authors have no competing interests.

REFERENCES

1. Data on COVID-19 related mortality rate- <https://ourworldindata.org/coronavirus> and Global health security index- <https://www.ghsindex.org>
2. Manjili, R.H., Zarei, M., Habibi, M., Manjili, M.H. COVID-19 as an Acute Inflammatory Disease. *J Immunol* **205**, 12-19 (2020).
3. Lee, J.S., Shin, E.-C. The type I interferon response in COVID-19: implications for treatment. *Nat Rev Immunol* **20**, 585–586 (2020).
4. J. O. Wertheim, D. K. W. Chu, J. S. M. Peiris, S. L. Kosakovsky Pond and L. L. M. Poon. A Case for the Ancient Origin of Coronaviruses, *J Virol*. 2013, 87, 7039–7045.
5. D. A. Tyrrell and M. L. Bynoe. Cultivation of Viruses from a High Proportion of Patients with colds. *Lancet* 1966, 1, 76–77.
6. Liu, M., Gao, Y., Zhang, Y., Shi, S., Chen, Y., Tian, J. The Association between Severe or Dead COVID-19 and Autoimmune Diseases: A Systematic Review and Meta-Analysis. *Journal of Infection* **81**, e93–e95 (2020).

7. Petroni, M., Hill, D., Younes, L., Barkman, L., Howard, S., Howell, I B., Mirowsky, J., Collins, M. B. Hazardous Air Pollutant Exposure as a Contributing Factor to COVID-19 Mortality in the United States. *Environ. Res. Lett.* **15** 0940a9 (2020).
8. Conticini, E. , Frediani, B., Caro, D. Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? *Environ. Pollut.* **261**, 114465 (2020).
9. Fattorini, D., Regoli, F. Role of the chronic air pollution levels in the Covid-19 outbreak risk in Italy. *Environ. Pollut.* **264**, 114732 (2020).
10. Bousquet, J., Anto, J.M., Iaccarino, G. et al. Is diet partly responsible for differences in COVID-19 death rates between and within countries?. *Clin Transl Allergy* **10**, 16 (2020)..
11. Tang, J.W., Toovey, O.T.R., Harvey, K.N., Hui, D.D.S. Introduction of the South African SARS-CoV-2 variant 501Y.V2 into the UK. *Journal of Infection* **82**, E8- E10 (2021).
12. Yeoh, Y.K., Zuo T., Lui G.C., et al. Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. *Gut* **70**, 698-706 (2021).
13. Kremer, H.J., Werner T. Age Dependence in COVID-19 Mortality in Germany. *Deutsches Arzteblatt international* **25**, 432-433 (2020).
14. Yang, Y., Zhong, W., Tian, Y., Xie, C., Fu, X., Zhou, H. The effect of diabetes on mortality of COVID-19: A protocol for systematic review and meta-analysis. *Medicine* **99**, e20913 (2020).

15. Woolcott, O.O., Castilla-Bancayán, J.P. The effect of age on the association between diabetes and mortality in adult patients with COVID-19 in Mexico. *Sci Rep* **11**, 8386 (2021).
16. de Almeida-Pititto, B., Dualib, P.M., Zajdenverg, L. et al. Severity and mortality of COVID 19 in patients with diabetes, hypertension and cardiovascular disease: a meta-analysis. *Diabetol Metab Syndr* **12**, 75 (2020).
17. Dubowsky, S.D., Suh, H., Schwartz, J., Coull, B.A., Gold, D.R. Diabetes, obesity, and hypertension may enhance associations between air pollution and markers of systemic inflammation. *Environmental Health Perspectives* **114**, 992-998 (2006).
18. Bach, J.-F., Chatenoud, L., The hygiene hypothesis: an explanation for the increased frequency of insulin-dependent diabetes. *Cold Spring Harbor perspectives in medicine*, 2, a007799 (2012).
19. Campbell, T.C., Campbell, T.M. The China Study The Most Comprehensive Study of Nutrition Ever Conducted & the Startling Implications for Diet, Weight Loss & Long-term Health. BenBella Books, Dallas, Texas (2004).
20. Samraj, A.N., Pearce O.M., Läubli, H., Crittenden, A.N., Bergfeld, A.K., Banda, K., Gregg, C.J., Bingman, A.E., Secrest, P., Diaz, S.L., Varki, N.M., Varki, A. A red meat-derived glycan promotes inflammation and cancer progression. *PNAS*, **112**, 542-547(2015).
21. Imhof, A., Froehlich, M., Brenner, H., Boeing, H., Pepys, M.B., Koenig, W. Effect of alcohol consumption on systemic markers of inflammation. *Lancet* **357**, 763-767 (2001).
22. Jin, Y., Wu, S., Zeng, Z., Fu, Z. Effects of environmental pollutants on gut microbiota. *Environ Pollut.*, **222**,1-9 (2017).

23. Khan, S., Waliullah, S., Godfrey, V., Khan, M.A.W., Ramachandran, R.A., Cantarel, B.L., Behrendt, C., Peng, L., Hooper, L.V., Zaki, H. Dietary simple sugars alter microbial ecology in the gut and promote colitis in mice. *Sci Transl Med.* **12** eaay6218 (2020).
24. De Filippo, C., Cavalieri, D., Di Paola, M., et al. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *PNAS* **107**,14691-14696 (2010).
25. Lee, E. Lee, J.E. Impact of drinking alcohol on gut microbiota: recent perspectives on ethanol and alcoholic beverage. *Current Opinion in Food Science* **37**, 91-97 (2021).
26. He, Y. Wu, W., Wu, S. et al. Linking gut microbiota, metabolic syndrome and economic status based on a population-level analysis. *Microbiome* **6**, 172 (2018).
27. Sinha-Ray, S., Yarin, A.L., Pourdeyhimi, B. Meltblowing: I-basic physical mechanisms and threadline model. *Journal of Applied Physics* **108**, 034912 (2010).
28. Staszal, C., Sinha-Ray, S., Yarin, A.L. Forced vibration of a heated wire subjected to nucleate boiling. *IJHMT* **135**, 44-51 (2019).
29. Zietz, M., Zucker, J., Tatonetti, N.P. Associations between blood type and COVID-19 infection, intubation, and death. *Nat Commun* **11**, 5761 (2020).
30. Liu, N., Zhang, T., Ma, L., Zhang, H., Wang, H., Wei, W., Pei, H., Li, H. The impact of ABO blood group on COVID-19 infection risk and mortality: A systematic review and meta-analysis. *Blood Rev.* **48**, 100785 (2021).