Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Evaluating the plausible application of advanced machine learnings in exploring determinant factors of present pandemic: A case for continent specific COVID 19 analysis

Suman Chakraborti, Arabinda Maiti, Suvamoy Pramanik, Srikanta Sannigrahi, Francesco Pilla, Anushna Banerjee, Dipendra Nath Das

PII: S0048-9697(20)36252-5
DOI: https://doi.org/10.1016/j.scitotenv.2020.142723
Reference: STOTEN 142723

To appear in: Science of the Total Environment

Received date: 18 July 2020
Revised date: 21 September 2020
Accepted date: 22 September 2020

Please cite this article as: S. Chakraborti, A. Maiti, S. Pramanik, et al., Evaluating the plausible application of advanced machine learnings in exploring determinant factors of present pandemic: A case for continent specific COVID 19 analysis, Science of the Total Environment (2020), https://doi.org/10.1016/j.scitotenv.2020.142723

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.
Evaluating the plausible application of advanced machine learnings in exploring determinant factors of present pandemic: A case for continent specific COVID 19 analysis

Suman Chakraborti\textsuperscript{a}, Arabinda Maiti\textsuperscript{b*}, Suvamoy Pramanik\textsuperscript{a} Srikanta Sannigrahi\textsuperscript{c}, Francesco Pilla\textsuperscript{c}, Anushna Banerjee\textsuperscript{a}, Dipendra Nath Das\textsuperscript{a}

\textsuperscript{a} Center for the Study of Regional Development, Jawaharlal Nehru University, New Delhi, Delhi 110067, India

\textsuperscript{b*} Geography and Environment Management, Vidyasagar University, West Bengal, India.

\textsuperscript{c} School of Architecture, Planning and Environmental Policy, University College Dublin

Richview, Clonskeagh, Dublin, D14 E099, Ireland.

*Corresponding author: Arabinda Maiti

E-mail: (Arabinda Maiti\textsuperscript{*}) : arabinda@mail.vidyasagar.ac.in

Suman Chakraborti (suman87ssf@jnu.ac.in)

*Arabinda Maiti (arabinda@vidyasagar.ac.in)

Srikanta Sannigrahi (srikanta.sannigrahi@ucd.ie)

Suvamoy Pramanik (suvamoy60ssf@jnu.ac.in)

Francesco Pilla (francesco.pilla@ucd.ie)

Anushna Banerjee (anushnabanerjee296@gmail.com)

Dipendra Nath Das (dipendra02@gmail.com)
Evaluating the plausible application of advanced machine learnings in exploring determinant factors of present pandemic: A case for continent specific COVID-19 analysis

Abstract:

Coronavirus disease, a novel severe acute respiratory syndrome (SARS COVID-19), has become a global health concern due to its unpredictable nature and lack of adequate medicines. Machine Learning (ML) models could be effective in identifying the most critical factors which are responsible for the overall fatalities caused by COVID-19. The functional capabilities of ML models in epidemiological research, especially for COVID-19, is not substantially explored. To bridge this gap, this study has adopted two advanced ML models, viz. Random Forest (RF) and Gradient Boosted Machine (GBM), to perform the regression modelling and provide subsequent interpretation. Five successive steps were followed to carry out the analysis: (1) identification of relevant key explanatory variables; (2) application of data dimensionality reduction for eliminating redundant information; (3) utilizing ML models for measuring relative influence (RI) of the explanatory variables; (4) evaluating interconnections between and among the key explanatory variables and COVID-19 case and death counts; (5) time series analysis for examining the rate of incidences of COVID-19 cases and deaths. Among the explanatory variables considered in this study, air pollution, migration, economy, and demographic factor were found to be the most significant controlling factors. Since a very limited research is available to discuss the superiority of ML models for identifying the key determinants of COVID-19, this study could be a reference for future public health research. Additionally, all the models and data used in this study are open
source and freely available, thereby, reproducibility and scientific replication will be achievable easily.

**Keywords:** Air pollution; Machine learning; Pandemic; Socioeconomic; COVID-19; Relative importance

| Abbreviations | Full Name |
|---------------|-----------|
| PM$_{2.5}$    | PM$_{2.5}$ (mg/m$^3$) |
| NetMig        | Net Migration |
| N$_2$O        | Nitrous oxide emission (metric ton) |
| GDP.gr        | Annual growth rate of GDP |
| TotPop        | Total Population |
| WS            | Wind speed (m s$^{-1}$) |
| LEB           | Life expectancy at birth |
| TotalCO$_2$   | Total CO$_2$ emission (metric ton) |
| GDP           | Per Capita GDP |
| CO$_2$PerCap  | Per Capita CO$_2$ emission (metric ton) |
| Refuge pop    | Refuge Population |
| AirTran       | Air transport registered |
| TropPress     | Pressure level at troposphere (hpa) |
| AgeGroup      | Age group (0-14) |
| Diabetes      | Diabetes (ages 20-79) |
| Preci         | Precipitation (mm) |
| Tmin          | Minimum Temperature ($^\circ$C) |
| TotGHG        | Total greenhouse gas emission (metric ton) |

1. **Introduction**

The human civilization is now facing an unprecedented threat from an once in a century pandemic, called Severe Acute Respiratory Syndrome Coronavirus (SARS COVID-19) (Ma et al., 2020). The first COVID-19 incidence took place in Wuhan, China, in December 2019. Due to the unpredictable nature and lack of adequate medicines, the COVID-19 virus soon became a global threat to humanity and propagated exponentially across the globe. By looking at the surmountable impact of COVID-19 on public health, the World health organization (WHO, 2020) declared a Public Health Emergency of International Concerns (PHEIC), especially in reference to the developing countries with weaker health-
care systems (UN-HABITAT, 2020). Due to its highly contagious nature, rapid human to human transmission and inadequate diagnostic cure system, there has been an accelerated increase in the daily COVID-19 incidences across the world, thereby acquiring the form of a pandemic. Thus, identifying the most important factors which have the highest influence on COVID-19 casualties is the need of the hour (Sannigrahi et al., 2020a, b). Since a limited research is available so far that address this concurrent issue, the present study has made an effort to identify the continent-specific causal factors of COVID-19.

Several studies have shown that multiple factors, such as air pollution (Conticini et al., 2020; Muhammad et al., 2020; Ogen, 2020; Tobías et al., 2020; Sannigrahi et al., 2020e), climatic conditions (Ahmadi et al., 2020; Bashir et al., 2020; Sobral et al., 2020), environnemental phenomena (Bao and Zhang, 2020; Janangiri et al., 2020; Sharma et al., 2020; Xie and Zhu, 2020) as well as other socio-demographic factors (Mollalo et al. 2020; Xiong et al. 2020; Bolaño-Ortiz et al. 2020; Sannigrahi et al., 2020a, b) are associated with COVID-19 incidents, casualties and spread. For example, Bashir et al., (2020) found that environmental pollutants like carbon monoxide (CO), particulate matter (PM$_{2.5}$, PM$_{10}$), Nitrogen dioxide (NO$_2$) are closely correlated with the COVID-19 cases in California. Bashir et al. also noted that long-term environmental policy can reduce the environmental pollution; hence, reducing the threat of COVID-19 like respiratory diseases in future. Similarly, Wu et al., (2020) found that 1 μg/m$^3$ increase in PM$_{2.5}$ can cause an 8 % increase in the COVID-19 deaths rate in America. Among the environmental pollutants, sulfur dioxide (SO$_2$) and Ozone (O$_3$) also play a significant role in respiratory system inflammation (Conticini, Frediani, & Caro, 2020). Taghizadeh-Hesary & Akbari, (2020) found a negative association between COVID-19 cases and patient smoking history in Iran. However, different environmental pollutants have different roles in amplifying COVID-19 cases and deaths, in different regions (Nakada and Urban, 2020).
Several epidemiological studies show that climatic parameters, including temperature, diurnal temperature change, relative humidity are linked to the spread of COVID-19 (Iqbal et al., 2020; Sobral, Duarte, da Penha Sobral, Marinho, & de SouzaMelo, 2020; Y. Wu et al., 2020). Chan et al., (2011) found that the longevity of SARS virus diminishes at temperatures >38 °C. Ma et al., (2020) observed that diurnal temperature range has a positive association ($r = 0.44$) with daily COVID-19 cases, while relative humidity shows a negative association (Qi et al., 2020). Additionally, the Qi et al., (2020) study found that the daily COVID-19 cases will decrease up to 11 to 22 % with 1 % rise of relative humidity. Other studies show that countries with higher precipitation rate have faced a high rate of transmission (Sobral et al., 2020). At the same time, Sobral et al., (2020) had found that an increase of 1°C daily temperature can reduce $\approx 7$ daily cases. All these findings collectively suggest that air temperature is negatively associated with COVID-19 cases, while relative humidity and precipitation shows varied relationship with COVID-19 incidences.

Apart from the biological, climatological and environmental factors, there are a few key socio-economic and demographic parameters, including population age structure, income, unemployment, poverty, human mobility, etc. that were found to be closely associated with COVID-19 cases-fatality rate (Ren et al., 2020; Sannigrahi et al., 2020a, b). According to (UN-HABITAT, 2020), low income countries with poor healthcare infrastructure, high population and room density, and informal job markets, could be more affected by COVID-19. Bolaño-Ortiz et al., (2020) found that poverty and income inequality is playing a significant role in COVID-19 fatalities in South American cities. Additionally, demographic parameters, particularly age-structure and co-morbidity is directly connected with the COVID-19 deaths fatality (Groenewegen et al., 2018; Lippi et al., 2020; Rosenkrantz et al., 2020; Sannigrahi et al., 2020a, b). Working age population is found to be a potential spreader of the contagious virus (Dowd et al., 2020). Dowd et al., (2020) also
found that countries with higher proportion of older and younger population could be a burden on public health. Beside these, migration and mobility played a significant role in the spreading of COVID-19 (Kraemer et al., 2020; Oztig and Askin, 2020; Xiong et al., 2020). All these studies are combinedly indicating a heterogeneous association between social, economic, demographic, environmental influencing factors and COVID-19 across the scale.

Machine Learning (ML) models can be used effectively to address the complex interactions and responses between the causal factors and COVID-19 incidences (Alsayed et al., 2020; Cheng et al., 2020; Pramanik et al., 2020). There are several methods that have evolved and used in epidemiological studies in the recent past. These techniques can be categorized as linear and non-linear regression-based analysis, (non)spatial explicit models (Franch-Pardo et al., 2020; Mollalo et al., 2020), spatio-temporal cluster analysis (Cordes and Castro, 2020; Kang et al., 2020), agent-based modelling (Xu et al., 2019), block-sequential analysis (Azevedo et al., 2020), Bayesian approach (Gayawan et al., 2020), neural network analysis (Kapoor et al., 2020), to name a few. Spatial explicit models such as Geographically Weighted Regression and Multiscale Geographically Weighted Regression, have been dominantly used in previous studies (Cordes and Castro, 2020; Kang et al., 2020; Mollalo et al., 2020; Sun et al., 2020). Spatial non-explicit models such as Ordinary Least Square Regression, Spatial Lag Models, Spatial Error Model, Generalized Additive Model (GAMs), have also been utilized in epidemiological research (Adekunle et al., 2020; Hockham et al., 2018; Scarpone et al., 2020). Briz-Redón and Serrano-Aroca (2020) have developed third degree polynomial models for identifying temporally-lagged covariates of COVID-19. Iqbal et al. (2020) adopted partial and multiple coherence wavelet modelling for evaluating the association between climatic variables and COVID-19. Though literature suggests that ML techniques have great potential in public health research, for the present pandemic, the same has not been explored substantially.
Several studies have identified different parameters, which have been evaluated mostly at country/regional scale and their association with COVID-19 cases and deaths have shown mixed outcomes. This study has made an effort, after incorporating all causal factors including climatic, environmental, socio-economic, demographic, etc., to build ML based predictive models and to identify the continent-specific key driving factors based on ML derived RI of the parameters. The main objectives of this study are: (1) to identify the most significant explanatory variables for each continent based on their association with COVID-19 cases and deaths; (2) to measure the relative influence of the explanatory variables on COVID-19 fatalities; (3) to track the association between the changes in human mobility and daily progression of COVID-19 counts.

2. Material and methods

The present research has adopted several relevant approaches and methods to accomplish the aims and objectives of this study. Five successive steps were followed to carry out the experiments and modelling. These are (i) data source and pre-processing; (ii) parameter identification and dimensionality reduction; (iii) evaluating variable relative importance using ML models; (iv) time series evaluation; and (v) experimental design.

2.1 Data source and pre-processing

Daily COVID-19 (both confirmed cases and death) data was collected from the European Centre for Disease Prevention and Control (ECDC, 2020) from 1st January 2020 to 11th June 2020. The daily COVID-19 data were further arranged to link the country statistics to their corresponding continent attributes. The daily COVID-19 data of each country was then converted to a cumulative sum to get the details of total cases/death counts of each
country and continents. A total of 51 (for Africa), 45 (for America), 25 (for Asia), 52 (for Europe), and 7 (for Oceania) country’s COVID-19 data was considered for the final analysis. Few countries were discarded from the analysis due to unavailability of data. The relevant factors, which have been used in this study to construct predictive modelling, was retrieved from the World bank open data platform (World Bank, 2020). A total of 35 indicators were initially considered, which later on had been filtered out for the final analysis. The details of the same are presented in Table. S1. Literature results were gathered for finalizing these variables. Both natural (climatic, environmental) and human appropriated (socio-economic, demographic) factors were incorporated in the modelling and subsequent interpretation. The final set of variables were then ingested into the regression modelling for further analysis and interpretation. Human mobility data including park, transit, grocery, residential, workplace, and recreation mobility, was collected from Google to examine the association between human mobility COVID-19. The country specific mobility data was further arranged to make it appropriate for continent level analysis.

2.2 Parameter identification and dimensionality reduction

Identifying the important factors which have the highest explanatory power and less auto-collinearity is one of the most challenging parts in regression modelling (Sannigrahi et al., 2020c, Sannigrahi et al., 2020d). The current study has employed several dimensionality reductions approaches, including stepwise forward regression, Confidence Interval (CI) approximation, parametric and non-parametric correlations estimation, linear regression modelling, Variance Inflation Factor (VIF) approximation, etc. to identify the key explanatory variables for the modelling and subsequent interpretation. Initially, all the 35 variables were selected in the dimensionality reduction modelling. For the COVID-19 cases, a total of 11 variables, i.e. TotalCO₂, CO₂PerCap, PM₂.₅, GDP, NetMig, Refuge pop, TotPop, N₂O, GDP.gr, LEB, WS, were considered for five continents (Africa, America, Asia, Europe,
and Oceania), while, for COVID-19 deaths, a total of 15 variables, i.e. TotalCO2, Diabetes, GDP.gr, TropPress, TotGHG, TotPop, Tmin, NetMig, N2O, LEB, AirTran, Diabetes, TropPress, Preci, AgeGroup, were identified for five continents. These filtered variables (for both cases and death) have exhibited minimum VIF (<2) and correlation (<0.4) values, and were, thus, selected as explanatory variables for continent-specific predictive modelling and parameter approximation.

2.3 Evaluation of variable relative importance using machine learning models

In this study, two ML models, i.e. Random Forest (RF) and Gradient Boosted Machine (GBM) were utilized for developing regression models and subsequent interpretation. The RF model is based on two main concepts, i.e. mean decrease impurity and mean decrease accuracy (Han et al., 2016). The RF model is proven to be the superior one compared to its counterparts as it produces relatively accurate and robust estimates, and at the same time, it is very easy to use (Chen et al., 2017). The impurity function, which is basically either Gini impurity or information gain/entropy impurity, denotes that while training a decision tree, it calculates how much of weighted impurity decreases in each feature and finally, the features are ranked according to their decreasing impurity values (Louppe et al., 2013). However selecting important features based on decreasing impurity could be biased if variables have more categories or if any multicollinearity exists in the variables, where one variable will be chosen for predictor and the other correlated variables would be considered as a less important features as the other correlated predictor had already explained the impurity decreased by the second variables. The other approach, the mean decrease accuracy, resembles how each feature affects the overall model accuracy (Louppe et al., 2013). This means that the less important features will have a very less or no effect on model accuracy.
Another ML model, GBM, which is a tree-based ensemble model that works in a gradual additive and sequential manner (Landry et al., 2016), has also been used to compare the results derived from the RF model. In the boosted tree model, higher importance is provided to the most important features and vice versa.

There are some unrevealed controlling factors which are regulating the overall COVID-19 scenarios across the countries. Identifying these continent-specific causal factors is challenging. This is due to different socio-economic and demographic profile of the countries. Since we have focused solely on identifying the key explanatory variables that have strongest impact on COVID-19 cases and deaths, adoption of robust and relevant methods could be a strong determinant for avoiding misleading inferences which eventually leads to dodgy conclusions. Therefore, highest priority is being provided to the selection of appropriate algorithms to carry out this sensitive task.

2.4 Experimental design

Socio-economic, demographic, and environmental data was collected for different countries, for whichever recent year is available in the World Bank open data platform. For COVID-19 counts, data was collected from January 1 to June 11, 2020. COVID-19 data for the later period has not been included in the analysis as it might influence the entire model outcome, since COVID-19 cases are strongly influenced by the testing rate, trend of human mobility, air traffic movements. The final shortlisted variables for both COVID-19 cases and death were categorized for each continent for the final regression modelling and interpretation. Countries with no data have been discarded from the analysis. For COVID cases, a total of 4 (for Africa), 4 (for America), 4 (for Asia), 3 (for Europe), and 2 (for Oceania) variables were considered. For COVID death, a total of 4 (for Africa), 4 (for America), 4 (for Asia), 7 (for Europe), and 2 (Oceania) variables have been finalized.
Descriptive statistics and dimensionality reductions of the variables were done in SPSS V26. Variable Relative Importance (RI) analysis was conducted using Random Forest (v.4.6-14) (Flaxman et al., 2011), e1071 (v.1.7-3) (Meyer, 2014), gbm (v.2.1.5) (Gong et al., 2019), parallel (R Core Team, 2019), tyder (v.1.0.2) (Wickham and Henry, 2019), params (R Core Team, 2019), easyalluvial (v.0.2.3) (Bjoern Koneswarakantha, 2020), parcats (v.0.0.1) (Bjoern Koneswarakantha, 2020), and ggplot2 (Wickham 2016) packages in R programming software. Human mobility information, which was retrieved from Google’s database, was evaluated to examine the trade-off association between human mobility and COVID-19 incidences across the continents. Human mobility data was analyzed using COVID-19 (v.2.2.0) package in R software (Guidotti and Ardia, 2020). Air pollution and climate variables were estimated using the Google Earth Engine (Gorelick et al., 2017) cloud platform. A linear trend of COVID-19 cases and deaths were calculated using the linear regression trend analysis. Data processing and visualization was done in ArcGIS Pro, R programming language.

3. Results

3.1 Spatial distribution and temporal progression of COVID-19 cases and deaths in different continents

The spatial distribution of COVID-19 cases and deaths intensity (case and death per 100,000 population) is analyzed and presented in Fig. 1. COVID-19 case intensity is found to be maximum in the middle-east Asian countries, such as Qatar (2645), Andorra (1106), Bahrain (1032) and Kuwait (817), while, the COVID-19 deaths (deaths per 100,000 population) is found maximum in the European countries, i.e. Belgium (84), Andorra (66), UK (61), Spain (57) and Italy (56). The weekly patterns of the daily cases and deaths per million population for each continent is represented in Fig. S1. Highest COVID-19 cases and
deaths per million population was found in America and Europe, followed by Asia, Africa and Oceania. There is a continuous increase of COVID-19 cases and deaths in Asia with two spikes of growth. The first spike of cases and deaths was observed from the 13th week (i.e., 24.03.2020) and second spike was observed from the 19th week (i.e., 05.05.2020). For Europe, there were both spike and abatement of cases during the study period. Spike of cases were observed from the 10th week (i.e., 03.03.2020) and spike of deaths was observed from the 12th week (17.03.2020); while abatement of deaths and cases were observed from the 16th and 20th week. In Africa, a spike of cases and deaths was started from the 17th week without any sign of abatement. For American countries, an abrupt change i.e., spike of COVID-19 cases from the 12th week and deaths from the 14th week with no depreciation was detected. Interestingly, a steady mortality rate was recorded after the spikes of COVID-19. For Oceania, a very low spike of COVID-19 cases and deaths were observed.

3.2 Identification of key explanatory variables

Multiple data dimensionality reduction approaches, including correlation test, stepwise regression, VIF, etc. were performed to identify the key explanatory variables that have maximum explanatory power and less collinearity. Table. S2 shows the summary statistics of the final filtered variables derived from stepwise forward regression analysis. For COVID-19 cases, the five parameter regression models were developed for America which explained 83% model variances, and hence, has been considered for the final analysis. For Asia, a five-parameter model, considering NetMig, PM2.5, GDP.gr, N2O, was established which explained 74% of total model variance. For Europe and Oceania, 4 parameters and 3 parameters regression model respectively, were constructed which explained nearly 87 % (for Europe) and 99% (for Oceania) variance (Table. S2). For COVID-19 deaths, 4 variables (CO2, Diabetes, GDP.gr, TropPress) were identified for Africa that explained 97% model
variability (Table. S3). For America, total 4 variables were identified based on their overall explanatory power. These variables are TotGHG, TotalCO$_2$, TotPop, and Tmin, respectively. Similarly, for Asia, 4 variables, i.e. TotPop, NetMig, N$_2$O, and LEB, were shortlisted based on their cumulative explanatory power (78%). An 8 parameters model was developed for Europe, considering 7 explanatory variables including AirTran, Diabetes, TropPress, Preci, TotPop, and AgeGroup, which has 94% overall model explanatory power.

Based on the shortlisted variables for each continent, an Ordinary Least Square (OLS) regression model was performed to examine the linear and direct association between these key variables and COVID-19 cases/death counts (Table. S4, Table. S5). For COVID-19 cases, the multi-parameters OLS model has explained 92% (for Europe), 97% (America), 74% (Asia), 79% (Europe) model variance. For COVID-19 deaths, the best fitted OLS models have exhibited a statistically significant goodness of fit estimates (Table. S5). To re-confirm the accuracy and robustness of the linear models, a continent specific multivariate model was developed and presented in Table. 1, and Table. 2. For COVID-19 cases, all the 4 explanatory variables considered for Africa have produced statistically significant estimates at P<0.05 level (Table. 1). For America, amongst the 4 explanatory variables, only 2 variables were found statistically significant. This indicates that, the rest of the two variables, i.e. NetMig and refugee population, does not have any significant impact on the overall COVID-19 casualties in the American countries. Among the 4 variables parametrized for the Asian continent, only 1 variable (NetMig) has shown statistically significant estimates, while the rest of the three variables were not correlated strongly with COVID-19 case and death counts. For Europe, three variables were shortlisted for regression modelling, among which, two variables (LEB and TotPop) were found statistically significant (Table. 1). While considering the COVID-19 deaths as the response factor in the modelling, only two variables (TotalCO$_2$ and Diabetes) were found statistically significant at P<0.05 significance level for
Africa (Table 2). For America, all 4 explanatory variables were found statistically significant which suggest that these 4 variables have strong association with the COVID-19 deaths in the American countries. Additionally, for Asia, only 1 variable was found significant and for Europe, a total of 6 out of 7 explanatory variables have shown significance at different probability levels (Table 2).

3.4 Evaluating relative influence of explanatory variables

This study has employed two machine learning models, i.e., RF and GBM, to identify the most significant variables from the set of shortlisted variables approximated for each continent. The process of variable selection is discussed explicitly in section 3.3. The RI, which suggests the overall individual impact of one variable in explaining the variability of response variables (COVID-19 cases and death here), has also been quantified and reported in Fig. 2, Fig. 3. To explain the interconnection between the explanatory variables and COVID-19 cases and deaths in a more explicit way, the class ranges, which indicates the intensity and how closely the COVID-19 casualties, i.e. cases and deaths, are dependent on the explanatory variables, have been categorized into five classes. These classes are High-High (HH), Medium-High (MH), Medium (M), Medium-Low (ML), and Low-Low (LL). Five contrasting colors were used to visualize the classes and make the classes separable from each other in the network. Three key information, i.e. density distribution of the variables, the intensity of interconnection between and among the variables, and most importantly, the RI of the variables is provided in each panel in Fig. 2, Fig. 3. The colored bins are demonstrating the values of each class, whereas, the bins width shows the intensity and strength of the variables.

Among the 4 variables shortlisted for Asian countries, the NetMig has exhibited the maximum impact (52.6%) on COVID-19 cases, followed by N2O (11.8 %), PM2.5 (2.2 %),
and the GDP.gr (1.1%). On the other hand, using the GBM model, the maximum RI was computed for PM$_{2.5}$ (34.65%), NetMig (27.16%), N$_2$O (21.56 %), and GDP.gr (16.6%) (Fig. S2a). In both models, NetMig and N$_2$O variables have exhibited maximum influences on COVID-19 cases. This suggests that migration and environmental pollution has a strong causal impact on the COVID-19 cases in Asian countries. Additionally, the HH values of COVID-19 cases are closely connected with the LL class of NetMig (100%), HH (38 %) and MH (38%) classes of N$_2$O concentration (Table. S6), and HH (38 %) classes of PM$_{2.5}$ concentration. COVID-19 are equally connected to all subclasses of GDP.gr (Fig. 2). On the other hand, LL class values of COVID-19 cases are closely related to HH (33%), MH (33%), and M (38%) classes of NetMig. Similarly, ML (34%) and LL (34%) classes of N$_2$O and PM$_{2.5}$ concentration (27 % for both) and all classes of GDP.gr (22 %) are closely associated with COVID-19 cases. As most of the Asian countries are way behind from the western developed countries in terms of investing capital funds on health infrastructures and securing quality health assurance to the peoples, the risk of getting infected by COVID-19 like pandemic would be much higher in these developing and under-developed countries (UN-HABITAT, 2020). The high interconnection between pollution factors and COVID-19 cases in the Asian countries has therefore resembled the fact that air pollution, that is mainly a result of large scale industries and auto-mobile sectors, should be controlled in a way that it does not create any additional burden on overall socio-ecological system of the country.

While considering COVID-19 deaths, the maximum RI was estimated for N$_2$O (41.5%), followed by TotPop (33.4%), NetMig (20.3%), and LEB (4.8%), respectively (Fig. 3a). The same association between the explanatory variables and COVID-19 death was found in the GBM model (Fig. S3a). Moreover, it can be seen in Fig. 3a that the HH class values of COVID-19 deaths are closely connected with the HH class values of N$_2$O and weakly connected to MH class values of N$_2$O (Table. S7).
For Europe, the RF model has computed the highest relative influence for the TotPop (79.2%), followed by LEB (14.1), and WS(6.7%) with COVID-19 cases. Additionally, it has been shown in Fig. 2b that the values of HH classes of COVID-19 cases are closely related to HH values of TotPop, LEB, wind speed. On the other hand, the GBM model derived RI values suggest that TotPop (74.91%), WS (13.71 %), and LEB (11.39%) (Fig. S2b), are the key explanatory variables with highest relative influence on COVID-19 cases, which was found in accordance with the estimates derived from the RF model. For COVID-19 deaths, TotPop, LEB, and WS has the maximum relative influence score (Fig. 3b). Considering the interconnecting nature among the variables, the HH class values of COVID-19 cases are strongly connected to the TotPop’s HH class value.

For Africa, RF model results indicate that the four main factors, i.e. total TotalCO2 (84.9 %), CO2PerCap (11.8 %), PM2.5 (2.2) emission, and GDP (1.1%) are playing an important role in regulating the COVID-19 cases in African countries. The values of the HH and MH COVID-19 cases are highly connected to the HH (68%) and MH (57%) classes of total TotalCO2. Additionally, HH class of CO2PerCap (27%), and PM2.5 (27%) are also found closely connected to COVID-19 cases. Conversely, despite a weaker connection between these two factors, the GDP is equally connected to all subclasses of COVID-19 cases (Fig. 2c). The results derived from the GBM model suggest that the TotalCO2 (49.89%), GDP (20.13%), CO2PerCap (15.17%), and PM2.5 (14.8%) having the maximum relative influence on COVID-19 cases (Fig. S2c). For COVID-19 deaths, TotalCO2 (84.9%), CO2PerCap (11.8%), PM2.5 (2.2%), and GDP (1.1%) have the highest relative influence scores (Fig. 3c). However, a comparably different association between the explanatory variables and COVID-19 deaths was observed in GBM derived estimates (Fig. S3c).
The 4 explanatory variables, TotPop (38.2%), total TotalCO₂ (31%), NetMig (30.7%), and Refuge pop (0.1 %) have exhibited the highest relative influence on COVID-19 cases for American countries. The HH class values of COVID-19 cases are well connected to the HH classes of TotPop (100%), total CO₂ emission (56%), and NetMig (56%) (Fig. 2d). Additionally, it can be seen in Fig. 2d that the LL class values of COVID-19 cases have been strongly associated with MH, M, ML, and LL classes of the TotPop and NetMig. Likewise, the M, ML, and LL classes (33% for all three classes) of total TotalCO₂ and all classes of Refuge pop were found to be closely associated with COVID-19 cases. The results derived from the GBM also replicates the same association as mentioned here (Fig. S2d). For COVID-19 deaths, the maximum relative influence score was calculated for GHG (34.1%), TotPop (33.6%), total TotalCO₂ (30.3%), and Tmin (2%), respectively (Fig. 3d). A quite similar association was observed in the GBM model (Fig. S3d). For Oceania, the RI among the variables was found maximum for NetMig (54.8 %) and N₂O (45.2 %) (Fig. 2). Additionally, the HH values of COVID-19 cases are closely connected with the HH classes of NetMig (100%) and N₂O (50%) in Oceanian countries (Table. S6).

3.5 Prediction of COVID cases and death using machine learning models

Fig. 4 and Fig. 5 shows empirical evidence of continent specific actual and model predicted outcome of COVID-19 fatalities and their association with explanatory variables. Results show that our model is capable of capturing the actual scenario, whereby a non-linear association can be observed in case of the actual cases and predicted cases for each explanatory variable. Considering the RF model outcome of predicted and actual cases and deaths in Asia, the TotPop has a positive nonlinear association with cases (Fig. 4b). In contrast, LEB shows COVID-19 infection risk is high with age >80. Similarly, TotPop, AirTran and N₂O have shown a positive non-linear relationship with the actual and predicted
deaths (Fig 5b). Possible explanation for this estimation could be the presence of high aged population in Europe, which places them among the relatively high-risk groups of COVID-19. Again, the European countries are densely connected by air traffic with the other Asian, African and American countries that increases the risk of spreading COVID-19. On the other hand, RF model derived actual and prediction result shows TotalCO$_2$, CO$_2$PerCap (Fig.4, 5), diabetic (Fig. 4c) have shown strong positive non-linear relationship with actual and predictive cases in Africa, while GDP and GDP.gr shows a moderate and negative relationship with both the cases and deaths (Fig. 5c).

Correlation plot exhibits that N$_2$O (0.372), PM$_{2.5}$ (0.376), and TotPop are significantly correlated with total cases and deaths at a 95% confidence interval for Asian countries. In Europe, unlike other studies, LEB (0.842, p=0.0001) is positively correlated with the COVID-19 cases (Fig. S4b), while AirTran (0.812, p=0.0001), TotPop (0.786, p=0.0001) and N$_2$O (0.730, p=0.0001) have exhibited high degree of positive correlation with deaths (Fig. S4b). TotalCO$_2$ concentrations have shown significant positive correlation with both the cases and deaths (Fig. S4c and S5c), while PM$_{2.5}$ (0.447, P=0.01) is strongly correlated with the cases in Africa. For America, TotalCO$_2$ (0.841, p=0.0001), Refuge pop (0.333, p=0.01) and TotPop (0.906, p=0.0001) are strongly correlated with COVID-19 cases (Fig.S4d). Considering deaths (Fig. S5d), positive correlation is found with GHG (0.841, p=0.0001), TotalCO$_2$ (0.889, p=0.0001), TotPop (0.905, p=0.0001), and Tmin (-0.758, p=0.0001). NetMig (0.750, p=0.01) is positively correlated with cases (Fig. S4e) in Oceania countries, and N$_2$O (0.949, p=0.01), AirTran (0.949, p=0.01) are positively correlated with deaths (Fig S5e). This signifies that continuously increasing N$_2$O pollution as well as AirTran propel the risk of spreading COVID-19 cases in Australia, Oceania and other oceanic countries.

### 3.6 Mobility and its association with COVID-19 incidences
Due to COVID-19 spread, several countries have gone through restrictive measures to control human mobility and thereby, a decreasing trend of mobility from baseline was observed in most of the countries. To increase the interpretability of the graph, the daily mobility changes for two time periods: March to May, and May to end of June have been discussed. Fig S6 exhibits that transit station mobility drops were significantly high in Europe (-70%), followed by America (-65%), Oceania (-60%), Asia (-55%) and Africa (-50%). Similarly, negative Grocery, Parks, Workplace and recreation mobility changes happened during March to middle of May in Asia, Africa, America and Oceania, while in Europe, parks mobility changes (70%) have significantly increased from April onwards. Likewise, residential mobility, which designates the percentage of population residing in their homes, shows high positive changes during March to May. Also, Fig. S6 shows the number of cases is 0.5 million in Asian and European countries, while less than 0.1 million cases and 1 million cases were found in African and American countries. In Oceanian countries, cases were < 6 thousand for the rigid measures that were taken to restrict human mobility. Fig. S6 suggests that the countries are either withdrawing their restrictive measures from May onwards, or else the mobility restrictions are not being strictly followed, as a result of which, mobility towards transit points like bus stops, railway stations, subways had increased and residential mobility has decreased dramatically.

4. Discussion

4.1 COVID-19 emergency and global public health

This new pandemic has clearly highlighted the shallowness of the present public health system, which covers wide range of multi-dimensional issues that includes socio-political, health financing, health workers, and health infrastructure (Borghi, 2020). The
present health care system is mostly national sovereignty depended; thus, each country has its own way of tackling the pandemic. The global health finance system had largely depended on a few specific financing sources. But recently, US has reduced the funding to WHO which has created uncertainty in public health financing (Borghi, 2020). Globally, almost all countries are facing the lack of essential medical equipment’s such as Personal Protective Equipment (PPE), ventilators, etc. irrespective of the country’s economic status (Kickbusch et al., 2020; Vecchi et al., 2020). WHO has estimated the worldwide acute shortage (near 6 million) of doctors, health workers, nurses, and other frontline medical staff during the COVID-19 pandemic time (Moulds, 2020). Additionally, the low and medium developing countries are facing a severe shortage of nurses. Moreover, the number of hospital beds per 1000 population and access to health care facilities are also very low in these lower developed nations. However global public health experts have highlighted some strategies for preparing future COVID-19 like pandemic (Rudnicka et al., 2020). These strategies are (i) preparing an inclusive healthcare system to accommodate the unexpected health crisis; (ii) separation of the normal patient from the infected patient and an attempt to control the spread of the virus to non-infected people through the proper and advanced health screening system; (iii) providing skill development training to the health personnel to increase the work support; (iv) infrastructural improvement including personal protective equipment and ventilator etc.; (v) creating awareness among the people by providing knowledge and right information about how the virus transmits (Fig. 6).

4.2 Pandemic crisis and global responses: A comprehensive discussion

Among the countries that are least affected by COVID 19, Taiwan, New Zealand, Australia are in the top of the list. In Taiwan, only 443 cases and 7 deaths (as of 10th June 2020) were recorded so far, despite being located adjacent to the global pandemic epicenter (Wuhan, China). When the entire world was losing the battle with COVID-19, rather than
closing down the entire economy for an elongated period by imposing business and commercial restriction, Taiwan took some rational decisions which have proven to be highly effective in tackling the virus. These strategies include quickly closing its international border and imposing bans on exporting surgical masks, utilizing the technology by contact tracing and mobile sim tracking to monitor the quarantined person’s activity at the early stage of the outbreak. New Zealand is another country which has been able to successfully defeat the virus (1504 cases and 22 deaths as of 10\textsuperscript{th} June 2020) by taking strong restriction measures including early shutdown of the country and imposing ‘the level 4 lockdown’ just a few weeks after the first case of COVID-19. In level 4 lockdown, people have to restrict social activities within their home and should only interact with their family members, which helps to root out the virus at the early stage. Australia has also successfully handled and emerged from the pandemic with negligible casualties. The coordinated response between science and politics throughout the administrative system has led the government to be successful in limiting the casualties caused by COVID-19. At the same time, worldwide daily death counts drop significantly from May onward. On the other side of the spectrum, many developed countries such as the USA, Italy, UK, Spain, were affected badly by COVID-19. In USA, till now (as on 27\textsuperscript{th} August 2020 data), 5,719,841 total confirmed cases and 177,332 deaths have been recorded, followed by Brazil (3,669,995 cases and 116,580 deaths), India (3,310,234 cases and 60,472 death) United Kingdom (327,802 cases and 41,449 deaths) and Mexico (568,621 cases and 61,450 deaths) (\textit{WHO, 2020}). The inability to take timely actions and sheer negligence to the threat of the virus could be some of the reasons for the severe consequences that are prevailing now in these countries. The African countries, despite having limited resources, are trying their best to tackle the virus spread by adopting basic preventive measures such as proactive screening (Uganda), handwashing stations in the
public transport area (Rwanda), digital communication through WhatsApp in Senegal, etc. to name a few (Hope et al., 2020).

In an attempt to stop the movement of population and control the spread of this contagious virus, countries across the world have been forced to imposed the biggest lockdown that human civilization have ever seen (Sampi and Jooste, 2020). While economy is concerned, the most direct impact of COVID-19 was observed in the low- and medium-income countries. The Organization for Economic Co-operation and Development (OECD) report stated that unemployment rate suddenly increased to 8.5% in April, 2020 from 5.2% in February, 2020, which is highest in the decade (2010 – 2020). However, in May, 2020, the unemployment rate goes down to 8.4%. However, it has been estimated that the unemployment rate may reach up to 9.4% in the fourth quarter of 2020 across the OECD countries, which may exceed the unemployment rate observed during the great economic recession in 2008.

This study has identified few causal factors which have exhibited high association with COVID-19 counts across the continents. These variables are NetMig, demographic structure, GDP, air pollution, etc. As the COVID-19 pandemic is known to be transmitted through person to person, both international travel and domestic migration have played a significant role in amplifying the COVID-19 transmission across the scale. This mass international and domestic human mobility, especially at the early stage of the outbreak, has hastened the spread of the virus and amplified the incidence rate in the countries that are well connected (Skórka et al., 2020). In addition to this, Tuite et al. (2020) measured the true outbreak size of COVID-19 in Italy and its linkages with air travel by using air travel volume between Italian cities and cities in other countries. Turin et al. study estimated an outbreak size of 3971 cases (95% CI 2907–5297), as compared with a reported case count of 1128 as on Feb 29, 2020, which indicates the non-identification of 72% cases (61–79%). Considering
the linkages between demographic structure and reported COVID-19 fatalities of a country, Dowd et al. (2020) measured a higher burden of COVID-19 mortality in countries with older versus younger populations. The strong connection between demographic composition and age-specific mortality of COVID-19 suggest that the preventive measures such as social distancing, human mobility, and other policies to restrict the transmission should be based on the demographic composition of the country as well as intergenerational interactions (Dowd et al., 2020). In this study, a strong positive association between GDP and COVID-19 fatalities was observed, which is in accordance with the findings of other studies (Skórka et al., 2020; Abdel Massih et al., 2020; Sarmadi et al., 2020). Though the higher GDP exhibits better healthcare infrastructure and strong economic independence of a nation, there is also evidence that higher GDP have a synergistic association with morbid behaviours, which are eventually responsible for occurrence of any diseases (Skórka et al., 2020). Additionally, higher GDP is also associated with high income, which is often connected with higher consumption of unhealthy foods. Consequently, the rich and economically more affluent countries have more overweight citizens compared to lower GDP countries and are thereby more susceptible to COVID-19 (Tan et al., 2020). Also, Skórka et al., (2020) noted that the pandemics would affect the developed economies strongly compared to the under developed nations. Ma et al., (2020) noted a strong negative association between temperature and COVID-19. They suggest that at lower temperatures, the performance of the immune system will be decreased, and as a consequence, the activity of infectious agents and virus transmission will be enhanced in lowering temperature. Several studies have reported that vitamin D deficiency could be linked with the acute respiratory diseases, and therefore, can be associated with the COVID-19 mortalities (Grant et al., 2020; Sarmadi et al., 2020). The present study has also found a strong association with air pollution and COVID-19 transmission. Among the pollutants, PM$_{2.5}$ is found to be highly connected to the overall
disease burdens. As SARS COVID-19 is aerosolized through talking or exhalation and transmitted through respiratory droplets sprays, and aerosol of virus typically is >5 μm in diameter transmission and therefore, they can be easily attached by air particulate matter (PM$_{0.1}$, PM$_{2.5}$, and PM$_{10}$) (Zoran et al., 2020).

5. Conclusion

The present study has evaluated the usability of machine learning models in epidemiological research, taking COVID-19 as a case of experiment. Two advanced supervised machine learning algorithms, i.e. RF and GBM, were employed for estimating relative influence of the explanatory variables on COVID-19 cases and death counts in different continents. The COVID-19 case intensity (case per 100,000 population) is found maximum in the middle-east Asian countries, while, the COVID-19 death (death per 100,000 population) is found maximum in the European countries (as of 11th June 2020). On the other hand, highest COVID-19 cases and deaths per million population was found in America and Europe, followed by Asia, Africa and Oceania (as of 11th June 2020). Among the explanatory variables considered in this study, air pollution (TotalCO$_2$, N$_2$O, PM$_{2.5}$ emission), migration, economy (GDP$_{gr}$), and demographic factor (AgeGroup and LEB) were found to be the most significant controlling factors. The present research has explored the synergistic and trade-off association between the socio-economic, environmental, demographic parameters and COVID-19, rendering insightful stories of COVID-19 fatalities. Therefore, we believe, this study could be a reference for future public health research. Additionally, all the models and data used in this study are open source and free available, thereby, reproducibility and scientific replication will be achievable easily.
References

Adekunle, I.A., Onanuga, A.T., Akinola, O.O., Ogunbanjo, O.W., 2020. Modelling spatial variations of coronavirus disease (COVID-19) in Africa. Sci. Total Environ. 729, 138998. https://doi.org/10.1016/j.scitotenv.2020.13899

AbdelMassih, Antoine, Ramy Ghaly, Abeer Amin, Amr Gaalialah, Aya Kamel, Bassant Heikal, Esraa Menshawey et al. "Obese communities among the best predictors of COVID-19-related deaths." Cardiovascular Endocrinology & Metabolism (2020).

Ahmadi, M., Sharifi, A., Dorosti, S., Jafarzadeh Ghoushchi, S., Ghanbari, N., 2020. Investigation of effective climatology parameters on COVID-19 outbreak in Iran. Sci. Total Environ. 729. https://doi.org/10.1016/j.scitotenv.2020.138705

Alsayed, A., Sadir, H., Kamil, R., Sari, H., 2020. Prediction of epidemic peak and infected cases for COVID-19 disease in Malaysia, 2020. Int. J. Environ. Res. Public Health 17, 1–15. https://doi.org/10.3390/ijerph17114076

Aman, M.A., Salman, M.S., Ali P., Y., 2020. COVID-19 and its impact on environment: Improved pollution levels during the lockdown period – A case from Ahmedabad, India. Remote Sens. Appl. Soc. Environ. 100382. https://doi.org/10.1016/j.rsase.2020.100382

Azevedo, L., Pereira, M.J., Ribeiro, M.C., Soares, A., 2020. Geostatistical COVID-19 infection risk maps for Portugal. Int. J. Health Geogr. 19, 25. https://doi.org/10.1186/s12942-020-00221-5

Bao, R., Zhang, A., 2020. Does lockdown reduce air pollution? Evidence from 44 cities in northern China. Sci. Total Environ. 731, 139052.
https://doi.org/10.1016/j.scitotenv.2020.139052

Bashir, M.F., Ma, B., Bilal, Komal, B., Bashir, M.A., Tan, D., Bashir, M., 2020. Correlation between climate indicators and COVID-19 pandemic in New York, USA. Sci. Total Environ. 728, 138835. https://doi.org/10.1016/j.scitotenv.2020.138835

Bjoern Koneswarakantha, 2020. easyalluvial: Generate Alluvial Plots with a Single Line of Code. R Packag. version 0.2.3. https://github.com/erblast/easyalluvial.

Bolaño-Ortiz, T.R., Camargo-Caicedo, Y., Puliafito, S.E., Ruggeri, M.F., Bolaño-Diaz, S., Pascual-Flores, R., Saturno, J., Ibarra-Espinosa, S., Mayol-Bracero, O.L., Torres-Delgado, E., Cereceda-Balic, F., 2020. Spread of SARS-CoV-2 through Latin America and the Caribbean region: a look from its economic conditions, climate and air pollution indicators. Environ. Res. 109938. https://doi.org/10.1016/j.envres.2020.109938

Borghi, J., 2020. We need a global health system to deal with pandemics [WWW Document]. Democr. Without Broders.

Briz-Redón, Á., Serrano-Aroca, A., 2020. A spatio-temporal analysis for exploring the effect of temperature on COVID-19 early evolution in Spain. Sci. Total Environ. 728. https://doi.org/10.1016/j.scitotenv.2020.138811

Chauhan, A., Singh, R.P., 2020. Decline in PM2.5 concentrations over major cities around the world associated with COVID-19. Environ. Res. 187, 109634. https://doi.org/10.1016/j.envres.2020.109634

Cheng, F.-Y., Joshi, H., Tandon, P., Freeman, R., Reich, D.L., Mazumdar, M., Kohli-Seth, R., Levin, M.A., Timsina, P., Kia, A., 2020. Using Machine Learning to Predict ICU Transfer in Hospitalized COVID-19 Patients. J. Clin. Med. 9, 1668. https://doi.org/10.3390/jcm9061668
Collivignarelli, M.C., Abbà, A., Bertanza, G., Pedrazzani, R., Ricciardi, P., Carnevale Miino, M., 2020. Lockdown for CoViD-2019 in Milan: What are the effects on air quality? Sci. Total Environ. 732, 1–9. https://doi.org/10.1016/j.scitotenv.2020.139280

Conticini, E., Frediani, B., Caro, D., 2020. Can atmospheric pollution be considered a co-factor in extremely high level of SARS-CoV-2 lethality in Northern Italy? Environ. Pollut. 261, 114465. https://doi.org/10.1016/j.envpol.2020.114465

Cordes, J., Castro, M.C., 2020. Spatial analysis of COVID-19 clusters and contextual factors in New York City. Spat. Spatiotemporal. Epidemiol. 34, 100355. https://doi.org/10.1016/j.sste.2020.100355

Dowd, J.B., Andriano, L., Brazel, D.M., Rotondi, V., Block, P., Ding, X., Liu, Y., Mills, M.C., 2020. Demographic science aids in understanding the spread and fatality rates of COVID-19. Proc. Natl. Acad. Sci. 202004911. https://doi.org/10.1073/pnas.2004911117

Flaxman, A.D., Vahdatpour, A., Green, S., James, S.L., Murray, C.J.L., 2011. Random forests for verbal autopsy analysis: Multisite validation study using clinical diagnostic gold standards. Popul. Health Metr. 9, 5–32. https://doi.org/10.1186/1478-7954-9-29

Franch-Pardo, I., Napoleano, B.M., Rosete-Verges, F., Billa, L., 2020. Spatial analysis and GIS in the study of COVID-19. A review. Sci. Total Environ. 739, 140033. https://doi.org/10.1016/j.scitotenv.2020.140033

Gates, B., 2020. Responding to Covid-19 — A Once-in-a-Century Pandemic? N. Engl. J. Med. 382, 1677–1679.

Gayawan, E., Adegboye, O., James, A., Adegboye, A., Elfaki, F., 2020. Bayesian spatial modelling of outbreaks of Ebola in Democratic Republic of Congo through the INLA-SPDE approach. https://doi.org/10.1101/2020.04.13.20063081
Gong, H., Sun, Y., Huang, B., 2019. Gradient Boosted Models for Enhancing Fatigue Cracking Prediction in Mechanistic-Empirical Pavement Design Guide. J. Transp. Eng. Part B Pavements. https://doi.org/10.1061/JPEODX.0000121

Grant, William B., Henry Lahore, Sharon L. McDonnell, Carole A. Baggerly, Christine B. French, Jennifer L. Aliano, and Harjit P. Bhattoa. "Evidence that vitamin D supplementation could reduce risk of influenza and COVID-19 infections and deaths." Nutrients 12, no. 4 (2020): 988.

Groenewegen, P.P., Zock, J.P., Spreeuwenberg, P., Helbich, M., Hoek, G., Ruijsbroek, A., Strak, M., Verheij, R., Volker, B., Waverijn, G., Dijkstra, M., 2018. Neighbourhood social and physical environment and general practitioner assessed morbidity. Heal. Place 49, 68–84. https://doi.org/10.1016/j.healthplace.2017.11.006

Guidotti, E., Ardia, D., 2020. COVID-19 data hub. J. Open Source Softw. 5, 2376.

H. Wickham, 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Hockham, C., Bhatt, S., Colah, P., Makherjee, M.B., Penman, B.S., Gupta, S., Piel, F.B., 2018. The spatial epidemiology of sickle-cell anaemia in India. Sci. Rep. 8, 1–10. https://doi.org/10.1038/s41598-018-36077-w

Hope, M.D., Raptis, C.A., Shah, A., Hammer, M.M., Henry, T.S., 2020. A role for CT in COVID-19? What data really tell us so far. Lancet 395, 1189–1190. https://doi.org/10.1016/S0140-6736(20)30728-5

Iqbal, N., Fareed, Z., Shahzad, F., He, X., Shahzad, U., Lina, M., 2020. The nexus between COVID-19, temperature and exchange rate in Wuhan city: New findings from partial and multiple wavelet coherence. Sci. Total Environ. 729, 138916. https://doi.org/10.1016/j.scitotenv.2020.138916
Jahangiri, Mehdi, Jahangiri, Milad, Najafgholipour, M., 2020. The sensitivity and specificity analyses of ambient temperature and population size on the transmission rate of the novel coronavirus (COVID-19) in different provinces of Iran. Sci. Total Environ. 728, 138872. https://doi.org/10.1016/j.scitotenv.2020.138872

Kang, D., Choi, H., Kim, J.H., Choi, J., 2020. Spatial epidemic dynamics of the COVID-19 outbreak in China. Int. J. Infect. Dis. 94, 96–102. https://doi.org/10.1016/j.ijid.2020.03.076

Kapoor, A., Ben, X., Liu, L., Perozzi, B., Barnes, M., Blais, M., O’Banion, S., 2020. Examining COVID-19 Forecasting using Spatio-Temporal Graph Neural Networks.

Kickbusch, I., Leung, G.M., Bhutta, Z.A., Matsoso, M.P., Ihekweazu, C., Abbasi, K., 2020. Covid-19: How a virus is turning the world upside down. BMJ 369, 10–12. https://doi.org/10.1136/bmj.m1336

Kraemer, M.U.G., Yang, C.H., Gutierrez, B., Wu, C.H., Klein, B., Pigott, D.M., du Plessis, L., Faria, N.R., Li, R., Hanage, W.P., Brownstein, J.S., Layan, M., Vespignani, A., Tian, H., Dye, C., Pybus, O.G., Scarpino, S. V., 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science (80-. ). 368, 493–497. https://doi.org/10.1126/science.abb4218

Lippi, G., Mattiuzzi, C., Sanchis-Gomar, F., Henry, B.M., 2020. Clinical and demographic characteristics of patients dying from COVID-19 in Italy versus China. J. Med. Virol. 0–3. https://doi.org/10.1002/jmv.25860

Ma, Y., Zhao, Y., Liu, J., He, X., Wang, B., Fu, S., Yan, J., Niu, J., Zhou, J., Luo, B., 2020. Effects of temperature variation and humidity on the death of COVID-19 in Wuhan, China. Sci. Total Environ. 724, 138226. https://doi.org/10.1016/j.scitotenv.2020.138226
Mahato, S., Ph, D., Ghosh, K.G., 2020. Short-term exposure to ambient air quality of the most polluted Indian cities due to lockdown amid SARS-CoV-2. Environ. Res. 109835. https://doi.org/10.1016/j.envres.2020.109835

Meyer, D., 2014. Support vector machines: the interface to libsvm in package e1071. … Syst. their …. https://doi.org/10.1007/978-0-387-77242-4

Mollalo, A., Vahedi, B., Rivera, K.M., 2020. GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. Sci. Total Environ. 728, 138884. https://doi.org/10.1016/j.scitotenv.2020.138884

Moulds, J., 2020. On World Health Day, new report says the world needs 6 million more nurses [WWW Document]. World Econ. Forum

Muhammad, S., Long, X., Salman, M., 2020. COVID-19 pandemic and environmental pollution: A blessing in disguise? Sci. Total Environ. 728, 138820. https://doi.org/10.1016/j.scitotenv.2020.138820

Nakada, L.Y.K., Urban, R.C., 2’20. COVID-19 pandemic: Impacts on the air quality during the partial lockdown in Sao Paulo state, Brazil. Sci. Total Environ. 730, 139087. https://doi.org/10.1016/j.scitotenv.2020.139087

Ogen, Y., 2020. Assessing nitrogen dioxide (NO2) levels as a contributing factor to coronavirus (COVID-19) fatality. Sci. Total Environ. 726, 138605. https://doi.org/10.1016/j.scitotenv.2020.138605

Oztig, L.I., Askin, O.E., 2020. Human Mobility and COVID-19: A Negative Binomial Regression Analysis. Public Health. https://doi.org/10.1016/j.puhe.2020.07.002

Pramanik, M., Chowdhury, K., Rana, M.J., Bisht, P., Pal, R., Szabo, S., Pal, I., Behera, B., Liang, Q., Padmadas, S.S., 2020. Climatic influence on the magnitude of COVID-19
outbreak: a stochastic model-based global analysis. medRxiv.

Pourghasemi, H.R., Pouyan, S., Heidari, B., Farajzadeh, Z., Fallah Shamsi, S.R., Babaei, S., Khosravi, R., Etemadi, M., Ghanbarian, G., Farhadi, A., Safaeian, R., Heidari, Z., Tarazkar, M.H., Tiefenbacher, J.P., Azmi, A., Sadeghian, F., 2020. Spatial modeling, risk mapping, change detection, and outbreak trend analysis of coronavirus (COVID-19) in Iran (days between February 19 and June 14, 2020). Int. J. Infect. Dis. 98, 90–108. https://doi.org/10.1016/j.ijid.2020.06.058

R Core Team, 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ren, H., Zhao, L., Zhang, A., Song, L., Liao, Y., Lu, Y., Cui, C., 2020. Early forecasting of the potential risk zones of COVID-19 in China’s megacities. Sci. Total Environ. 729, 138995. https://doi.org/10.1016/j.scitotenv.2020.138995

Rosenkrantz, L., Schuurman, N., Bell, N., Amram, O., 2020. The need for GIScience in mapping COVID-19. Heal. Place 102389. https://doi.org/10.1016/j.healthplace.2020.102389

Rudnicka, L., Gupta, M., Kassir, M., Jafferany, M., Lotti, T., Sadoughifar, R., Goldust, M., 2020. Priorities for global health community in COVID-19 pandemic. Dermatol. Ther. 395, 19–20. https://doi.org/10.1111/dth.13361

Sampi, J.R., Jooste, C., 2020. Nowcasting Economic Activity in Times of COVID-19: An Approximation from the Google Community Mobility Report. https://doi.org/https://www.hsdl.org/?view&did=839193

Sannigrahi, S., Pilla, F., Basu, B., Basu, A.S. and Molter, A., 2020a. Examining the association between socio-demographic composition and COVID-19 fatalities in the
European region using spatial regression approach. *Sustainable Cities and Society*, p.102418.

Sannigrahi, S., Pilla, F., Basu, B. and Basu, A.S., 2020b. The overall mortality caused by covid-19 in the european region is highly associated with demographic composition: A spatial regression-based approach. *arXiv preprint arXiv:2005.04029*.

Sannigrahi, S., Zhang, Q., Joshi, P.K., Sutton, P.C., Keesstra, S., Roy, P.S., Pilla, F., Basu, B., Wang, Y., Jha, S. and Paul, S.K., 2020c. Examining effects of climate change and land use dynamic on biophysical and economic values of ecosystem services of a natural reserve region. *Journal of Cleaner Production*, 257, p.120424.

Sannigrahi, S., Zhang, Q., Pilla, F., Joshi, P.K., Basu, B., Keesstra, S., Roy, P.S., Wang, Y., Sutton, P.C., Chakraborti, S. and Paul, S.K., 2020d. Responses of ecosystem services to natural and anthropogenic forcings: A spatial regression based assessment in the world's largest mangrove ecosystem. *Science of the Total Environment*, 715, p.137004.

Sannigrahi S, Molter A, Kumar P, Zhang Q, Basu B, Basu AS, Pilla F. 2020e. Examining the status of improved air quality due to COVID-19 lockdown and an associated reduction in anthropogenic emissions. *medRxiv*. doi:10.1101/2020.08.20.20177949.

Sarmadi, Mohammad, Nilufar Marufi, and Vahid Kazemi Moghaddin. "Association of COVID-19 global distribution and environmental and demographic factors: An updated three-month study." *Environmental Research* 188 (2020): 109748.

Scarpone, C., Brinkmann, S.T., Große, T., Sonnenwald, D., Fuchs, M., Walker, B.B., 2020. A multimethod approach for county-scale geospatial analysis of emerging infectious diseases: a cross-sectional case study of COVID-19 incidence in Germany. *Int. J. Health Geogr.* 19, 32. https://doi.org/10.1186/s12942-020-00225-1
Sharma, S., Zhang, M., Anshika, Gao, J., Zhang, H., Kota, S.H., 2020. Effect of restricted emissions during COVID-19 on air quality in India. Sci. Total Environ. 728, 138878. https://doi.org/10.1016/j.scitotenv.2020.138878

Skórka, Piotr, Beata Grzywacz, Dawid Moroń, and Magdalena Lenda. "The macroecology of the COVID-19 pandemic in the Anthropocene." PloS one 15, no. 7 (2020): e0236856.

Sobral, M.F.F., Duarte, G.B., da Penha Sobral, A.I.G., Marinho, M.L.M., de Souza Melo, A., 2020. Association between climate variables and global transmission of SARS-CoV-2. Sci. Total Environ. 729, 138997. https://doi.org/10.1016/j.scitotenv.2020.138997

Sun, F., Matthews, S.A., Yang, T.-C., Hu, M.-H., 2020. A spatial analysis of COVID-19 period prevalence in US counties through June 28, 2020: Where geography matters? Ann. Epidemiol. https://doi.org/10.1016/j.annepidem.2020.07.014

The Guardian, 2020. More masks than jellyfish: coronavirus waste ends up in ocean.

Tobías, A., Carnerero, C., Reche, C., Massagué, J., Via, M., Minguillón, M.C., Alastuey, A., Querol, X., 2020. Changes in air quality during the lockdown in Barcelona (Spain) one month into the SARS-CoV-2 epidemic. Sci. Total Environ. 726, 138540. https://doi.org/10.1016/j.scitotenv.2020.138540

Tuite, Ashleigh R., Victoria Ng, Erin Rees, and David Fisman. "Estimation of COVID-19 outbreak size in Italy." The Lancet. Infectious Diseases 20, no. 5 (2020): 537.

UN-HABITAT, 2020. UN-Habitat COVID-19 Response Plan 1–16.

Vecchi, V., Callea, G., Cusumano, N., 2020. How to ensure countries don’t run out of medical supplies when the next crisis hits [WWW Document]. Word Econ. Forum.

Wickham, H., Henry, L., 2019. tidyr: Tidy Messy Data. R Packag. version 1.0.0.
Xie, J., Zhu, Y., 2020. Association between ambient temperature and COVID-19 infection in 122 cities from China. Sci. Total Environ. 724, 138201. https://doi.org/10.1016/j.scitotenv.2020.138201

Xiong, Y., Wang, Y., Chen, F., Zhu, M., 2020. Spatial statistics and influencing factors of the COVID-19 epidemic at both prefecture and county levels in Hubei Province, China. Int. J. Environ. Res. Public Health 17. https://doi.org/10.3390/ijerph17113903

Xu, Z., Graves, P.M., Lau, C.L., Clements, A., Geard, N., Glass, K., 2019. GEOFIL: A spatially-explicit agent-based modelling framework for predicting the long-term transmission dynamics of lymphatic filariasis in American Samoa. Epidemics 27, 19–27. https://doi.org/10.1016/j.epidem.2018.12.003

Zoran, M.A., Savastru, R.S., Savastru, D.M. and Tautan, M.N., 2020. Assessing the relationship between surface levels of PM2.5 and PM10 particulate matter impact on COVID-19 in Milan, Italy. Science of The Total Environment, 738, p.139825.

Acknowledgement

Authors would like to thank Covid-19 data hub group, World Bank and Google. Authors would like to acknowledge two anonymous reviewer and editor for their constructive comments to improve the manuscript. DD wish to acknowledge Indian Council of Social Science Research for partial financial support through the research project no 12/2017-18/ICSSR/RPS.

Figure captions:
Fig. 1 Spatial distribution of global COVID-19 case and death (per 100,000 persons) scenario. Also, continent specific daily progression of COVID-19 cases and deaths are showing in the bottom left corner.

Fig. 2 Alluvial plot shows the strength of interconnection between the explanatory variables predicted COVID cases derived from random forest algorithm. Relative influence (RI) values of each variable are shown in the right side of each plot.

Fig. 3 Alluvial plot shows the strength of interconnection between the explanatory variables predicted COVID deaths derived from random forest algorithm. Relative influence (RI) values of each variable are shown in the right side of each plot.

Fig. 4 The predictive power of the explanatory variables computed for COVID cases derived from the random forest algorithm.

Fig. 5 The predictive power of the explanatory variables computed for COVID death derived from the random forest algorithm.

Fig. 6 An overall comprehensive global pandemic preparedness path to highlight the strategies that need to be given importance. Compiled from (Borghi, 2020; Jacobsen, 2020; Rudnicka et al., 2020; Zumla and Niederman, 2020).
Table. 1 Continent-wise ordinary least square model estimates for COVID cases.

| Source | Value       | Standard error | t     | Pr > | Lower bound (95%) | Upper bound (95%) |
|--------|-------------|----------------|-------|-------|-------------------|-------------------|
| Africa | Intercept  | -975.375       | 1171.583 | -0.833 | 0.410             | -3336.546        | 1385.795         |
|        | TotalCO₂   | 0.138          | 0.008  | 16.807 | < 0.0001          | 0.122            | 0.155            |
|        | CO₂PerCap  | -1726.210      | 464.354 | -3.717 | 0.001             | -2662.054        | -790.365         |
|        | PM₂.₅     | 54.080         | 24.737 | 2.186  | 0.034             | 4.226            | 103.934          |
|        | GDP        | 0.621          | 0.234  | 2.650  | 0.011             | 0.149            | 1.094            |
|        | R²         | 0.915          |        |       |                   |                   |                  |
| America| Intercept  | -9799.119      | 13987.621 | -0.701 | 0.492             | -38407.016       | 18808.779        |
|        | TotalCO₂   | 0.183          | 0.034  | 5.425  | < 0.0001          | 0.114            | 0.252            |
|        | NetMig     | 0.036          | 0.019  | 1.936  | 0.053             | -0.002           | 0.074            |
|        | Refuge pop | -0.954         | 0.499  | -1.911 | 0.066             | -1.975           | 0.067            |
|        | TotPop     | 0.003          | 0.000  | 7.162  | < 0.0001          | 0.002            | 0.003            |
|        | R²         | 0.968          |        |       |                   |                   |                  |
| Asia   | Intercept  | -252.500       | 17538.584 | -0.011 | 0.989             | -36837.345       | 36332.346        |
|        | N₂O        | -0.010         | 0.073  | -0.171 | 0.897             | -0.162           | 0.143            |
|        | PM₂.₅     | 497.056        | 313.876 | 1.584  | 0.129             | -157.678         | 1151.789         |
|        | GDP.gr     | -2795.361      | 2849.423 | -0.981 | 0.338             | -8739.153        | 3148.432         |
|        | NetMig     | -0.067         | 0.014  | -4.732 | 0.000             | -0.096           | -0.037           |
|        | R²         | 0.744          |        |       |                   |                   |                  |
| Europe | Intercept  | -308530.573    | 123353.089 | -2.501 | 0.016             | -557132.388      | -59928.758       |
|        | LEB        | 3862.814       | 15715.644 | 2.447  | 0.018             | 681.266          | 7044.362         |
|        | TotPop     | 0.003          | 0.000  | 11.664 | < 0.0001          | 0.002            | 0.003            |
|        | WS         | -1587.322      | 2575.880 | -0.593 | 0.556             | -6980.203        | 3805.560         |
|        | R²         | 0.785          |        |       |                   |                   |                  |
Table. 2 Continent-wise ordinary least square model estimates for COVID death.

| Source | Value  | Standard error | t     | Pr > | Lower bound (95%) | Upper bound (95%) |
|--------|--------|----------------|-------|------|-------------------|-------------------|
| Africa |        |                |       |      |                   |                   |
| Intercept | -68.055 | 100.040        | -0.680 | 0.500 | -270.090          | 133.979           |
| TotalCO2 | 0.003  | 0.000          | 10.417 | < 0.0001 | 0.002             | 0.003             |
| Diabetes | 10.583 | 4.577          | 2.312  | 0.026 | 1.338             | 19.827            |
| GDP.gr  | 2.737  | 7.317          | 0.374  | 0.710 | -12.040           | 17.514            |
| TropPress | 0.297  | 0.881          | 0.337  | 0.738 | -1.483            | 2.077             |
| R²      | 0.786  |                |        |      |                   |                   |
| America |        |                |       |      |                   |                   |
| Intercept | -33669.459 | 12314.726 | 2.734  | 0.010 | -58819.484        | -8519.433         |
| TotalCO2 | 0.010  | 0.001          | 8.716  | < 1.0001 | 0.008             | 0.12              |
| TotPop  | 0.000  | 0.000          | 3.631  | 0.001 | 0.000             | 0.000             |
| Tmin    | 111.479 | 41.986         | 2.653  | 0.013 | 25.733            | 197.225           |
| R²      | 0.992  |                |        |      |                   |                   |
| Asia    |        |                |       |      |                   |                   |
| Intercept | 1745.681 | 4577.871      | 0.381  | 0.705 | -7547.891         | 11039.253         |
| TotPop  | 0.000  | 0.000          | 3.690  | 0.001 | 0.000             | 0.000             |
| NetMig  | 0.001  | 0.009          | 1.951  | 0.059 | 0.000             | 0.000             |
| N2O     | -0.012 | 0.007          | 1.663  | 0.105 | -0.027            | 0.003             |
| LEB     | -16.218 | 61.790         | 0.262  | 0.794 | -141.659          | 109.223           |
| R²      | 0.471  |                |        |      |                   |                   |
| Europe  |        |                |       |      |                   |                   |
| Intercept | 110378.223 | 26759.609   | 4.125  | 0.000 | 55265.777         | 165490.668        |
| AirTran | 0.014  | 0.005          | 2.883  | 0.008 | 0.004             | 0.023             |
| Diabetes | -2557.616 | 514.081     | -4.975 | < 0.0001 | -3616.385        | -1498.846         |
| TropPress | -333.409 | 87.589       | -3.807 | 0.001 | -513.801          | -153.017          |
| Preci   | -2237.303 | 856.040      | -2.614 | 0.015 | -4000.352         | -474.255          |
| TotPop  | 0.060  | 0.000          | 3.612  | 0.001 | 0.000             | 0.001             |
| AgeGroup | -237.852 | 189.825      | -1.227 | 0.231 | -623.803          | 158.100           |
| N2O     | -0.566 | 0.223          | -2.539 | 0.018 | -1.026            | -0.107            |
| R²      | 0.892  |                |        |      |                   |                   |
Credit Author Statement

SC - Conceptualization; Data curation; Formal analysis; Writing - review & editing.
AM - Conceptualization; Data curation; Formal analysis; Writing - review & editing.
SP - Conceptualization; Data curation; Formal analysis; Writing - review & editing.
SS – Conceptualization; Data curation; Formal analysis; Writing - review & editing.
FP- Review and editing
AB- Writing, Review and editing
DD- Review and editing
Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
**Highlights**

- The usability of machine learning (ML) models in epidemiological research (COVID-19) is explored.
- Two ML models, Random Forest and Gradient Boosted Machines, are employed for modelling.
- Relative influence of the variables on COVID-19 is computed.
- Air pollution has the most substantial impact on COVID-19 casualties across the continent.
- The method and data used in this study could easily replicate for future research.
Net.mig = Net migration
N2O = Nitrous oxide emission (metric ton)
PM 2.5 = (mg/m^3)
GDP.gr = Annual growth rate of GDP
Total.pop = Total population
LEB = Life expectancy at birth
WS = Wind speed (m s^-1)
Total Co2 = Total Co2 emission (metric ton)
Co2PerCap = Per capita Co2 emission (metric ton)
GDP = Per Capita GDP
Refuge.pop = Refuge population
Figure 3

N2O = Nitrous oxide emission (metric ton)
Total.pop = Total population
Net.mig = Net migration
LEB = Life expectancy at birth
AirTran = Air transport registered
Diabetes = % of diabetes (ages 20-79)

Total.Co2 = Total Co2 emission (metric ton)
TropPress = Pressure level at tropopause (hPa)
GDP.gr = Annual growth rate of GDP
TotGHG = Total green house gas emissions
Tmin = Minimum temperature (oC)
Figure 5
POST COVID 19 PANDEMIC PREPAREDNESS PATH

International Health Regulation

1. Fully implementation of IHR in all economic level country.
2. Scale up laboratory capacity, other surveillance capabilities, quick response health system.
3. Safe guarding the health-care workforce.
4. Improve the communication, co-ordination strategies.
5. Enhancing the supply chain for adequate stock of essential medical goods

Understanding the insufficiency in existing knowledge

1. Focus on the research for alternative factors related to highly contagious that haven’t yet been considered.
2. Initiation of effective global communication channels for coordinated action.
3. Initiation of dynamic models and tools for better preparedness

Synthesis (World Health Parliament)

Achieving the targets and priorities of Sendai Framework

1. Adopting the Sendai Framework approach accepted by UN in 2015.
2. Preparation for resilience plan for pandemic at local, national, regional and international level.
3. Strengthen the health system and developing the response system considering social, environmental, political and institutional setting.
4. Robust inclusive policy for mitigation, preparedness, response and recovery by incorporating climate change, economy, social issues

Transdisciplinary and Multisectoral Strategies

1. It is necessary for the solving the complex problem that threaten the global public health.
2. Need emphasis on interconnectedness of humans, animals and ecosystems.
3. Detecting viruses in wildlife and livestock is a crucial for early warning of human pandemic

Figure 6