



OPEN

## The role of European health system characteristics in affecting Covid 19 lethality during the early days of the pandemic

Monica Giancotti<sup>1,6</sup>, Milena Lopreite<sup>2,6</sup>✉, Marianna Mauro<sup>3,6</sup> & Michelangelo Puliga<sup>4,5,6</sup>

This article examines the main factors affecting COVID-19 lethality across 16 European Countries with a focus on the role of health system characteristics during the first phase of the diffusion of the virus. Specifically, we investigate the leading causes of lethality at 10, 20, 30, 40 days in the first hit of the pandemic. Using a random forest regression (ML), with lethality as outcome variable, we show that the percentage of people older than 65 years (with two or more chronic diseases) is the main predictor variable of lethality by COVID-19, followed by the number of hospital intensive care unit beds, investments in healthcare spending compared to GDP, number of nurses and doctors. Moreover, the variable of general practitioners has little but significant predicting quality. These findings contribute to provide evidence for the prediction of lethality caused by COVID-19 in Europe and open the discussion on health policy and management of health care and ICU beds during a severe epidemic.

The new coronavirus (SARS-CoV-2 or COVID-19) outbreak appeared in Wuhan, China in December 2019 and rapidly progressed in Europe becoming an urgent concern<sup>1-3</sup>. Despite the response activities, the epidemic has become widespread and the mortality in some European countries has been very high<sup>4</sup>. Currently, the pandemic has not yet ended (March 2021), and many authoritative estimations suggest a possible development with catastrophic consequences in terms of structural inequalities of income, health, and education, many from middle-income countries<sup>5-7</sup>. One of the most relevant factors that could dampen the devastating power of the epidemic could be the loss of lethality of the virus<sup>4,8</sup>, expressed as the ratio between the deaths and the infected people<sup>9</sup>. For these reasons, the factors influencing the COVID-19 lethality are still undergoing investigation with many studies mainly focused on local factors and health conditions.

Several papers have especially underlined the role of environmental factors in accelerating SARS-CoV-2 spread and its lethality including the chronic exposure to air pollution that might lead people to be more susceptible to the COVID-19 disease, in turn leading to an increase of COVID-19 spread and its lethality<sup>10-13</sup>.

Focusing on the relationship between the number of reported cases and the weather variables in some regions, other studies analyzed the effects of the climate change such as temperature and humidity on the transmission of COVID-19 epidemic. The researchers demonstrated that weather variables such as temperature and humidity are essential in predicting the mortality rate of COVID-19<sup>14</sup>.

A large set of analyses identify the socio-demographic variables, such as *age* and *health status* the most correlated with Covid-19 lethality: people under a growing risk are in general aged over 70 years, are immunocompromised or with specific chronic medical conditions<sup>15-19</sup>.

Finally, health factors (e.g., national health expenditures, health infrastructure, healthcare personnel) are also particularly important<sup>20,21</sup>. Recent studies showed that exceeding the health capacity leads to an increase in the mortality/case ratio<sup>22</sup>: one of the main factors implicated in the COVID-19 deaths is the surge of cases that depleted hospital resources<sup>23</sup>.

Many resources are required to adequately treat a critically ill patient with COVID-19: an intensive care unit (ICU) bed with a full-featured ventilator, personal protective equipment (e.g., isolation gowns, N95 respirators,

<sup>1</sup>Department of Clinical and Experimental Medicine, Magna Graecia University, Viale Europa, Catanzaro, Italy. <sup>2</sup>Department of Economics, Statistics and Finance, University of Calabria, Calabria, Italy. <sup>3</sup>Department of Clinical and Experimental Medicine, Magna Graecia University, Catanzaro, Italy. <sup>4</sup>Institute of Management, Sant'Anna School of Advanced Studies, Pisa, Italy. <sup>5</sup>Linkalab Computational Laboratory, Cagliari, Italy. <sup>6</sup>These authors contributed equally: Monica Giancotti, Milena Lopreite, Marianna Mauro and Michelangelo Puliga. ✉email: milena.lopreite@unical.it

gloves, etc.), and adequate hospital staffing (doctor and nurses), in a perspective of quality care maximization during a burden disease<sup>23</sup>.

From this point of view, most European countries have been facing a health emergency that were not ready to address for the lack of human and structural resources. This was the consequence of restrictive health policies adopted by the main European countries in response to Organization for Economic Co-operation and Development (OECD) suggestions<sup>24</sup>.

Following those advices, the number of hospital beds was reduced, greatly affecting the mortality rate: There is large evidence that the countries with the highest hospital beds capacity were the ones that had the lowest numbers of deaths<sup>23</sup>. For example, in Germany, despite the OECD suggestions of a tightening health policy as best practice, the global number of hospital beds was growing before the coronavirus outbreak<sup>24</sup>. This measure contributed to improve the ability of the German health system to cope with the coronavirus spread. Indeed, according to the OECD statistics, Germany with 8 hospital beds for every thousand people is at the top rank for an optimal bed management, followed by France 5.9 (from 6.4), Italy with 3.1 hospital beds for every thousand people (in 2013 it was 3.5), Spain 3 (from 3.2) and Great Britain 2.5 (from 3 that were in 2013). In the first 40 days of the COVID-19 Germany recorded a very low death cases with respect to the diagnosed cases (37.000 cases and 206 deaths) much less than other European countries. For example, in Italy the infected were 74.000 and 7500 the victims, while in Great Britain there have been 8600 affected, and 652 deaths, 3 times those in Germany<sup>25</sup>. The only OECD country with performances better than Germany was Japan where there are 13 beds per thousand inhabitants. Japan, in fact, recorded few COVID-19 infections at the beginning of the epidemic, just over 1300 cases, but, in proportion, even less victims (45). In summary, this scenery could be explained by the different countries' response in managing the coronavirus emergence: recent studies show that the hospital capacity is crucial in reducing the deaths of infected people<sup>26,27</sup>.

Other studies report that general practitioners (GPs) represent an important resource of a public health system<sup>28</sup>. The COVID-19 outbreak has refocused attention on their role. They contributed in different ways in the management of the novel coronavirus. In many European countries, GPs were at the forefront of tackling the spread of the virus. They are both gatekeepers and health promoter empowering the community to build a firewall against the deadly virus<sup>29,30</sup>. Even in hospital-oriented towards the health services, where the family medicine system is not yet fully implemented (as Wuhan), GPs helped in blocking the viral transmission by monitoring people at designated checkpoints in airports, railway stations and highways where they verified personal information and conducted health checks for travelers, reporting suspected cases to hospitals for urgent follow-up<sup>28</sup>.

Another important leading cause of Covid-19 lethality is represented by economic factors. In this sense, a variable under investigation is represented by the GDP. Recent studies seem to observe a lower rate of Covid-19-death cases in countries with higher GDP (i.e., Luxembourg)<sup>31</sup>. The opposite was found for countries that exhibit low levels of GDP per capita (i.e., Ukraine, Bulgaria, and Romania)<sup>31</sup>: people living in low-income European countries might suffer of poor health for the scarce access to health services due to their lower income<sup>31,32</sup>. These factors might be exaggerated during the periods of severe crises, negatively affecting the less developed regions<sup>31,32</sup>.

Starting from these points and given the growing number of daily reported deaths during the Covid-19 outbreak<sup>25</sup>, it becomes crucial to evaluate the lethality connected to SARS-COV2, investigating the leading factors that may affect it.

This study aims to contribute to this research field analyzing the role of health system characteristics in affecting Covid 19 lethality during the early days of the pandemic. In particular, we use an explorative approach guided by the following research question: which are the health systems' key-driver variables that play a crucial role in predicting lethality? Using the Random Forest regression, and for the purpose of a robust analysis other Machine learning methods (ML), we measured the importance of the variables inserted in the model (also called "features") in predicting the *target variable* lethality at different time horizons. As a further robustness check we showed the superior performance of random forest regression with respect to Principal Component Analysis (PCA) and other machine learning techniques (ML) such as AdaBoost and Lasso.

This analysis aims to present a scenery of the countries' health systems performance and their preparedness when no specific measures have been yet taken to mitigate the coronavirus impact.

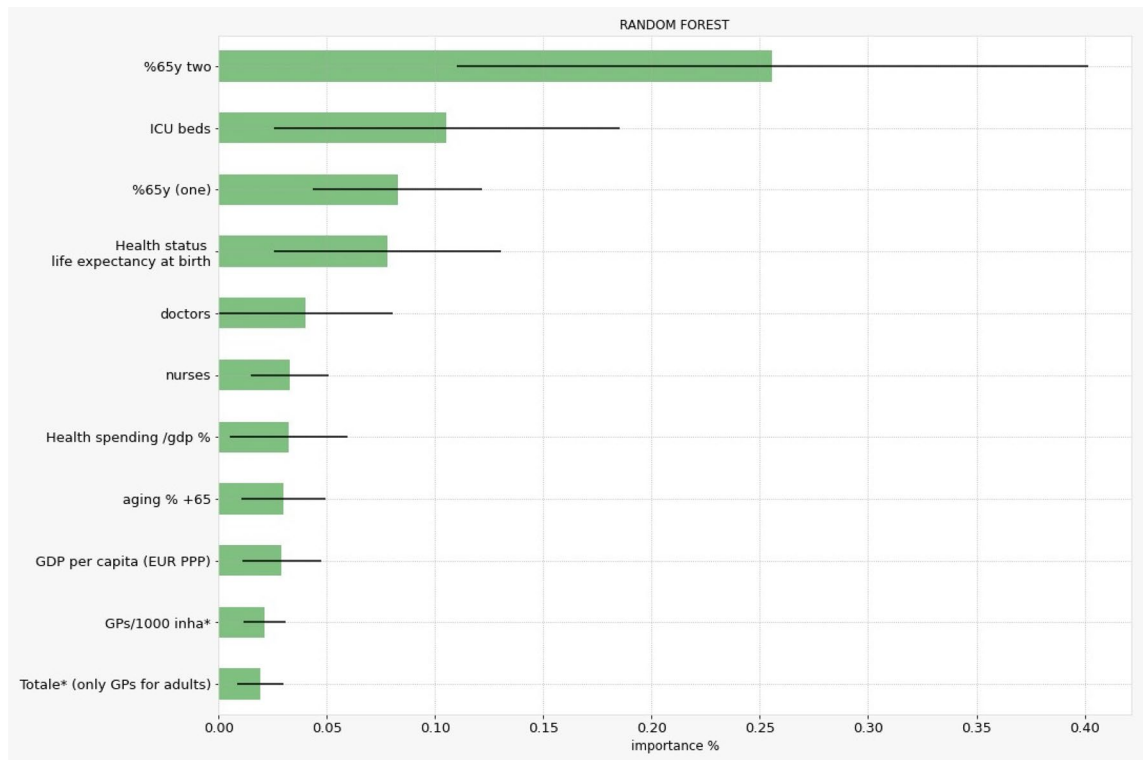
## Results

Using the random forest regression, we screened the most relevant variables related to lethality during the first 40 days of the coronavirus outbreak. We consider the outcome of lethality at 10, 20, 30 and 40 days since the first fatal case of COVID-19 registered in each country as a robustness check of our results. Moreover, to obtain a summary of the distribution of the importance score of the variables inserted in the Random Forest regression we use the median as a central measure and the interquartile range (IQR) as a dispersion measure. The Fig. 1 reports in vertical lines the error bars representing the IQR (the extremes of the error bars are respectively Q1 as 25th percentile and Q3 as the 75th percentile) centered on the median value. The IQR results confirm a robust ordering of the ranking list after two careful statistical tests have been run: (a) the Kruskal-Wallis non parametric H test for the medians of a population showing that the medians of the rankings are statistically different; (b) the Wilcoxon post hoc pairwise test (Tables S1, S2, S3 that show the results of the pairwise comparison for different sample sizes are available in the file of supplementary materials) that explores the pairwise difference among the medians value and it confirms that are all statistically significant.

The results obtained with Random Forest regression, are shown in Fig. 1.

The variables were ordered with respect to their importance in the lethality prediction with values in %. At the top rank of the Fig. 1 we have the most relevant variables.

Specifically, a low value of the IQR (the importance of each variable at different time frames 10, 20, 30, 40 days) shows that the ICU beds per 1000 inhabitants—one of the most important indicators of the health



**Figure 1.** The relative importance of the variables for the best regression methods (Random Forest model): the bars show the interquartile range (IQR) centered on the median value.

infrastructure—play a crucial role (a constant effect on lethality) in explaining the lethality at the beginning of the coronavirus spread. This confirms the existing literature that correlates different mortality rates to bed capacities at the beginning of the pandemic<sup>23,26,27</sup> and it motivates our study that focus on the role of health system characteristics in predicting the lethality. In fact, the ICU beds were reduced before the pandemics (as result of health reforms inspired to managerialism<sup>33</sup>) and the cuts created a shortage of ventilators and other lifesaving devices. For this reason, in several countries the hospital systems tried to keep as low as possible the number of patients in ICU beds eventually delaying their admittance causing a likely negative impact on the lethality.

However, whether examining the average effect of the variables during the study period the indicator more influent to the lethality is still the demographic variable including population age (percentage of 65 and above years old divided by 1000 inhabitants with at least two or more chronic diseases).

Its relatively large variability appears to be due to the growing effects on the lethality in the 40 days of time interval.

This result confirms that the increase in average population age and the higher prevalence of pre-existing diseases and strain in physical functioning is strictly connected with lethality for COVID-19.

Interestingly on the other hand there are also the variables of health spending and number of doctors, which are not similar in their importance. In this case investments in health expenditure compared to GDP could be a new key-driver variable that will provide the first best level of health structures in order to prioritize intervention and prevention strategies. Similarly, the number of nurses and doctors will be inversely proportional to mortality: where there are fewer nurses and doctors, the mortality rate will be higher. In summary, more investments in health infrastructure and in general in health spending which make healthcare more accessible to a broader part of the population and more doctors and nurses could translate into less virus fatalities.

Moreover, as our results have shown, the GDP is down in the ranking list. We can explain this result in the fact that without implications for social welfare the economic growth does not necessarily lead to an increase of people wellbeing.

Finally, the variable of GPs follows GDP and shows little but significant predicting quality: the role of GPs in healthcare system can help fill the gap in population health services, contributing to reduce lethality.

To test the performance of random forest regression we compare this model with PCA, Lasso, AdaBoost each method using the RMSE. Table 1 shows the main results of the Model. The RMSE is estimated with the leave one out procedure for the lethality at 10, 20, 30, and 40 days, an average is provided in the last column, clearly showing the better performance of the Random Forest with respect to the other two models (AdaBoost and Lasso). These criteria confirm the Random Forest that shows the lowest RMSE the best method to study the feature importance.

Model	Lethality (10)	Lethality (20)	Lethality (30)	Lethality (40)	avg RMSE
Random Forest	32	33	83	125	68
AdaBoost	34	35	100	138	77
Lasso	155	110	340	383	247
PCA	ND	ND	ND	ND	ND

**Table 1.** RMSE on lethality for Random Forest and Lasso, computed at different time horizons (10, 20, 30 and 40 days) deaths per million inhabitants. The “avg RMSE” column is the average of the rows. For the PCA method there is no RMSE as the technique does not reconstruct the data.

## Discussion

Investigating the factors correlated with the COVID-19 lethality is the key-question of our work to make assumptions whether national health care systems are well-equipped and in general are prepared in dealing with a global health crisis.

Previous studies had shown that ML techniques provide accurate results using epidemiological data<sup>34–36</sup>.

In particular, Random Forest models have been used very frequently in prediction analyses, showing high performance with respect to other models<sup>37–40</sup>. Recently, they have been employed to compute COVID-19 mortality or to predict the risk of mortality<sup>37,41,42</sup>. The data used for the analyses are in prevalence based on patients’ physiological conditions, symptoms, demographic information<sup>41,42</sup>, population characteristics<sup>43</sup> or blood lab results and clinical data<sup>41,44–47</sup>.

Unlike the previously mentioned studies, the main contribution of this paper is to analyze the role of health system characteristics in predicting Covid 19 lethality during the early days of the pandemic through the usage of a Machine Learning model.

Indeed, most of the variables inserted in the model are those that describe the characteristics of each country’s health care system.

Specifically, in this paper, we analyzed the leading causes of lethality during the first 40 days of coronavirus outbreak in 16 European countries. Assessing the variables that mostly affect the lethality is crucial to provide efficient policy decisions concerning the economy of a country and its health care system. Although the variables that we selected in our study are standard in explaining the lethality we propose a novel and original study of the benefits of applying machine learning methods, and random forest model in particular, to the assessment of the variable importance, in explaining the lethality.

We find that demographic variables are the most relevant in predicting lethality, with a prevalence of the elders aged over 65 with two or more chronic diseases. This is not surprising, and the result is in line with the frequent reports that identify people of older age belonging to risky groups of Covid-19 fatalities<sup>36,37,45,48</sup>.

Another relevant variable especially during the first days of the spread of the epidemic in explaining the lethality is the number of beds in ICU: our results confirm that ICU capacity planning must be a first-order concern for health authorities to efficiently estimate the demand for urgent care during a pandemic<sup>49</sup>. This variable is followed by the healthcare spending (percentage of GDP), and the number of nurses and doctors/1000 inhabitants. In summary, more resources in health infrastructure and in general in healthcare spending make healthcare more accessible to a broader part of the population resulting into less virus fatalities. Moreover, where there are few doctors, the mortality rate is greater. These results confirm that health policy oriented to expansion of space, staff and hospital supplies are strongly recommended during a high burden disease that depletes hospital resources.

In this study there are several strong points. Firstly, using a method like the Random Forest, we obtain a model that explains the lethality for the majority of the European countries during the first weeks of Covid-19 outbreak. Secondly, during a pandemic in which it is crucial to make timely decisions to not increase the number of victims, our results mirror the main variables for lethality prediction to mitigate the spread of the virus and to reduce the pressure on the hospitals and the health system. Finally, the Random Forest technique used here has the advantage that, unlike some traditional regression models, it also could model nonlinear relationships that exist among the variables of our analysis.

Despite these strengths, there are several limitations that should be addressed.

The main limitation of the study is that we analyze, due to lacking data, only 16 European countries. However, unlike previous studies are limited in size examining just a single country<sup>36–38,40</sup>, our goal is to use the data from 16 European countries to analyze with ML techniques the global effect of different health system characteristics on COVID-19 lethality during the early days of the pandemic.

Another limitation is that we did not collect information on globalization indicators such as tourists or net migration that could influence the lethality rate: increasingly globalized world and health emergencies are connected through interrelated patterns<sup>50</sup>. In fact, higher degrees of global connectivity enable a faster spread of diseases and then an increase of lethality.

Moreover, a potential problem in the data collection of Covid-19 cases is relative to the number of cases that can be underestimated: health statistics could miss the real number of people suffering of one or more chronic disease. Despite all these data limitations, the usage of a panel of data with 16 countries allows to limit the influence of the national, and local effects.

Finally, it could be interesting to evaluate a specific healthcare policy, a detailed investigation, especially at micro level, as the aggregated variables considered here cannot be directly translated into policy instruments: more data could offer better results. We leave these points for future research.

## Methods

**Data availability statement.** Our study was conducted during the first 40 days of the coronavirus outbreak by considering 16 European Countries (Austria, Belgium, Czechia, France, Germany, Greece, Finland, Ireland, Italy, Luxembourg, Netherlands, Norway, Poland, Portugal, Romania, Spain). The other European countries were excluded due to lack of data. For each country we considered the following variables: life expectancy at birth, per capita health spending at purchasing power parity (PPP), people aged over 65 years expressed as percentage of elders with respect to their age group (without chronic disease, with one chronic disease, with multiple chronic diseases), the fraction of elders above 65 years over the population, number of GPs/1000 inhabitants, number of GPs excluding pediatrics /1000 inhabitants, number of specialized GPs/1000 inhabitants, number of doctors/1000 inhabitants, numbers of nurses/1000 inhabitants, health spending as percentage of GDP, number of beds in ICU per 1000 inhabitants lethality at 10–20–30–40 days (computed as the ratio of deaths over confirmed cases from the first day a death has been reported, it is also called “*case-fatality ratio*”<sup>25</sup>). Data were drawn from the Eurostat database<sup>51</sup> and from countries health profiles published by OECD database<sup>52</sup>. Number of GPs were extracted from the reports of the European observatory of health systems and policies published by the World health organization (WHO)<sup>53</sup>. All data were matched with the statistics of the lethality of COVID-19 provided by the Data Repository of the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University<sup>25</sup>.

**Declaration of agreement with journal guidelines.** All methods were performed in accordance with relevant guidelines and regulations and no experiments were performed on humans and/or human tissue samples. The data used in the analysis derived from public sources, and to the best of our knowledge, there are no privacy or moral issues associated with them.

**The random forest regression.** Among the various machine learning models that can perform a Regression Analysis predicting COVID-19 lethality, we are interested in those that allow us to assess the importance of every variable used in the model. The model must be able not only to correctly predict levels of lethality using independent variables but also to estimate the impact of each variable on lethality. In this work we explored Random Forest regression (RF), that we compared in terms of performance with AdaBoost, Lasso and the Principal component Analysis (PCA) (for an introduction to Ensemble methods see<sup>54</sup>). All of these methods were built using the “leave one out” procedure on the available data. The “leave one out” procedure is a simple method that derives from the more general cross validation technique<sup>55</sup>. Given a sample  $S$  of size  $n$  a training phase is performed on a subsample  $S'$  of size  $n-1$ , leaving out a single point that is used for the validation procedure. This procedure is repeated  $m$  times (like 1000) with different randomly extracted points. Once completed the procedure we obtain an average performance value (RMSE error) for each model. Once selected the best one we use its feature importance ranking list for all its internal variables. As the model is used to predict the lethality values at different time horizons (10, 20, 30, 40 days) we take the average of the importance score of each variable using the median value (as a robust measure of central tendency) of the 4 cases.

In summary the steps needed to replicate this analysis are the following:

1. Prepare  $k$ -independent variables that we believe they are mostly related to the Covid-19 lethality (for instance the “number of elders above 65 years with two or more chronic diseases”). There is an observation of these variables for each country and a value of lethality to predict (at 10, 20, 30, and 40 days).
2. Setup the machine learning models (we suggest to use a library such as scikit-learn of Python, or H2o with R, that implement several methods), that are able to predict the lethality (at a given time point after the initial outbreak) with a regression model
3. To improve the ability of the regression models in reproducing the data, as the number of observations is not large (one point per each country) we employ a method called “leave one out” that works, during the training phase of the algorithm, shortening the dataset of a single *random* point and repeating the regression estimate many times (1000 iterations) in order to get an average value for the predicted lethality.
4. The models performance in predicting the lethality (Random Forest, AdaBoost, Lasso, PCA) are compared by their square root in mean error (RMSE). The most accurate model is selected after a cycle of 1000 random comparisons.
5. Starting from the most accurate model we use the “feature importance” to identify the contribution of each variable to the regression prediction.
6. Finally, to summarize the data of lethality at different time spans (from 10 to 40 days) we use the median measure and the interquartile range (IQR) to capture the dispersion and to get a more robust estimate of the importance of each variable in the final ranking list.

The Random Forest, Lasso, AdaBoost, form a group of methods that differently from the PCA model respond to alternative questions, when it comes to measure the importance of each feature. The main difference is related to the following question: (a) is the model able to reproduce the data? (b) is the model able to explain the variability of each feature?

AdaBoost<sup>56</sup> is an ensemble boosting technique that sequentially uses *weak learners* (usually simple decision trees that are barely better than random guessing the results) to obtain an accurate ensemble prediction. The main idea of boosting is using a first set of weak models to generate imperfect predictions that are used as “new data” to feed a second level of estimators that in turn, run after run, will make better predictions. This sequence of estimations, refined step by step, makes the AdaBoost a powerful technique to accurately fit the data, and



make future predictions, but at the risk of overfitting. The RF model is based on random ensembles of decision trees, each one solving the recursive binary problem of splitting the data into interesting features (the decisions) according to a metric such as the Root Mean Square Error (RMSE) for the regression task or Cross Entropy for classification. The RMSE metrics is based on the idea that the best regression model is the one that minimizes the total square error from the actual data. The RF regression feature importance is then a measure that estimates how much each feature contributes to the minimization of the RMSE. In other words, the most important features are the ones that improve the most the ability of the model to reproduce the data.

The Lasso<sup>57</sup> technique uses a penalty score applied to the coefficients of the linear regression with the goal of reducing the Root Mean Square Error (RMSE) without overfitting the data. The algorithm will reduce (shrink) the coefficients, setting several of them to zero. The level of penalty  $\lambda$  controls the amount of regularization and therefore the importance of each coefficient. Higher penalties mean that more coefficients are shrunk to zero, while low penalty values imply more nonzero coefficients. The importance associated with each feature is the absolute value of each regression coefficient. While the feature importance for Lasso is easy to interpret and clear the choice of the penalty  $\lambda$  is crucial to keep/discard the variables. Usually, this value is selected with a cross-validation methodology. The level of  $\lambda$  is increased until the model is correctly able to fit the training set with a lower error, and at the same time, it is not overfitting the validation data. Once the optimal penalty level has been reached the remaining coefficients of the linear regression represent with their value the importance of each associated feature. Lasso and Random Forest link the feature importance directly to the goodness of the regression models, however, while Lasso is a linear technique, Random Forest is not. The decision trees are able to correctly perform also in those situations where data do not follow a linear rule. In this sense non-linear models are able to capture more complex relationships and are to be preferred every time there is no explicit assumption can be made on the data linearity.

Minimizing variance is a trick used also in dimensionality reduction methods such as the PCA. In this case the goal is finding a “data transformation” of the dataset that preserves the majority of the total variance of the data. Feature importance will evaluate the features that carry out the majority of data variability, and from this regard, it will be relative to the data themselves, and not to the model that explains them. For this reason, despite the PCA being an attractive method for its simplicity, it does not respond to the fundamental question of which variable contributes the most to a regression model; it answers, instead, the question of which variable contributes with the largest variance. The dataset used in this paper is of limited size, having only 16 points: one per each country. Dealing with short samples with ML methods requires some extra care for the validation part: instead of the classical cross-validation technique, that divides the dataset in a training/test set, we used the “leave one out” methodology where the validation set consists of a single random point left out during every randomization. Averaging the error committed in estimating the validation point allows to check the error, and to tune the model for optimal parameters: a procedure known as “hyperparameter tuning”.

In summary, we used the four techniques AdaBoost, Lasso, PCA and Random Forest with leave one out validation to get the best model parameters, selecting the one (in our case the Random Forest) that is the most efficient in minimizing RMSE. The random forest regression is the model that better generalizes on unseen data with no overfitting. Thus, its parameters are used to estimate the feature importance associated with each variable.

## Data availability

All data is available in the main text or by request to the authors.

Received: 16 March 2021; Accepted: 26 November 2021

Published online: 09 December 2021

## References

- Bontempi, E. The Europe second wave of COVID-19 infection and the Italy “strange” situation. *Environ. Res.* **193**, 110476. <https://doi.org/10.1016/j.envres.2020.110476> (2020).
- Li, Q. *et al.* Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl. J. Med.* **382**, 1199–1207. <https://doi.org/10.1056/NEJMoa2001316> (2020).
- Paules, C. I., Marston, H. D. & Fauci, A. S. Coronavirus infections—more than just the common cold. *JAMA* **323**(8), 707–708. <https://doi.org/10.1001/jama.2020.0757> (2020).
- Carta, M. G. *et al.* Association between the spread of COVID-19 and weather-climatic parameters. *Eur. Rev. Med. Pharmacol. Sci.* **24**(15), 8226–8231 (2020).
- Vollmer, M., Mishra, S., & Juliette, H. Using mobility to estimate the transmission intensity of COVID-19 in Italy: a subnational analysis with future scenarios. Imperial College London, (2020).
- Loppreite, M., Panzarasa, P., Puliga, M. & Riccaboni, M. Early warnings of COVID-19 outbreaks across Europe from social media. *Sci. Rep.* **11**(1), 1–7 (2021).
- Sirleaf, E. J. & Clark, H. Report of the independent panel for pandemic preparedness and response: Making COVID-19 the last pandemic. *Lancet* [https://doi.org/10.1016/S0140-6736\(21\)01095-3](https://doi.org/10.1016/S0140-6736(21)01095-3) (2021).
- Rajan, S., Khunti, K., Alwan, N., Steves, C., MacDermott, N., Morsella, A., Angulo, E., Winkelmann, J., Bryndová, L., Fronteira, I. & Gandré, C. In the wake of the pandemic: Preparing for Long COVID. European Observatory on Health Systems and Policies, Copenhagen (Denmark); (2021). PMID: 33877759.
- Motta Zanin, G., Gentile, E., Parisi, A. & Spasiano, D. A preliminary evaluation of the public risk perception related to the COVID-19 health emergency in Italy. *Int. J. Environ. Res. Public Health* **17**(9), 3024. <https://doi.org/10.3390/ijerph17093024> (2020).
- Pegoraro, V., Heiman, F., Levante, A., Urbinati, D. & Peduto, I. An Italian individual-level data study investigating on the association between air pollution exposure and Covid-19 severity in primary-care setting. *BMC Public Health* **21**(1), 902. <https://doi.org/10.1186/s12889-021-10949-9> (2021).
- Copat, C. *et al.* The role of air pollution (PM and NO<sub>2</sub>) in COVID-19 spread and lethality: a systematic review. *Environ. Res.* **191**, 110129. <https://doi.org/10.1016/j.envres.2020.110129> (2020).
- Pluchino, A. *et al.* A novel methodology for epidemic risk assessment of COVID-19 outbreak. *Sci. Rep.* **11**, 5304. <https://doi.org/10.1038/s41598-021-82310-4> (2021).

13. Gupta, A. *et al.* Air pollution aggravating COVID-19 lethality? Exploration in Asian cities using statistical models. *Environ. Dev. Sustain.* **23**(4), 6408–6417 (2020).
14. Christophi, C. A. *et al.* Ambient temperature and subsequent COVID-19 mortality in the OECD countries and individual United States. *Sci. Rep.* **11**, 8710. <https://doi.org/10.1038/s41598-021-87803-w> (2021).
15. Kidd, M. R. Five principles for pandemic preparedness: Lessons from the Australian COVID-19 primary care response. *Br. J. Gen. Pract.* **70**(696), 316–317. <https://doi.org/10.3399/bjgp20X710765> (2020).
16. Muniyappa, R. & Gubbi, S. COVID-19 pandemic, coronaviruses, and diabetes mellitus. *Am. J. Physiol. Endocrinol. Metab.* **318**, E736–E741. <https://doi.org/10.1152/ajpendo.00124.2020> (2020).
17. Zhou, F. *et al.* Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *Lancet* **395**, 1054–1062. [https://doi.org/10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3) (2020).
18. Fang, L., Karakiulakis, G. & Roth, M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Respir. Med.* **8**(4), e21. [https://doi.org/10.1016/S2213-2600\(20\)30116-8](https://doi.org/10.1016/S2213-2600(20)30116-8) (2020).
19. Patel, J. A. *et al.* Poverty, inequality and COVID-19: the forgotten vulnerable. *Public Health* **183**, 110. <https://doi.org/10.1016/j.puhe.2020.05.006> (2020).
20. Fogel, R. W. Nutrition, physiological capital, and economic growth. *Pan American Health Organization and Inter-American Development Bank*. <http://www.paho.org/English/HDP/HDD/fogel.pdf> (2002).
21. Fogel, R. W. Health, nutrition, and economic growth. *Econ. Dev. Cult. Change* **52**(3), 643–658 (2004).
22. Liu, W. Y., Xiao, G. & Tchounwou, P. B. Response to the Covid-19 epidemic: The Chinese experience and implications for other countries. *Int. J. Environ. Res. Public Health* <https://doi.org/10.3390/ijerph17072304> (2020).
23. Sen-Crowe, B., Sutherland, M., McKenney, M. & Elkbuli, A. A closer look into global hospital beds capacity and resource shortages during the COVID-19 pandemic. *J. Surg. Res.* **260**, 56–63 (2021).
24. Organization for Economic Co-operation and Development (OECD). Better policies for better lives. <https://www.oecd.org/about/47747755.pdf> (2011).
25. COVID-19 map. Johns Hopkins Coronavirus Resource Center. <https://coronavirus.jhu.edu/map.html> (2020).
26. Sussman, N. Time for bed (s): Hospital capacity and mortality from COVID-19. *Covid Econ.* **11**, 116–129 (2020).
27. Ciceri, F. *et al.* Decreased in-hospital mortality in patients with COVID-19 pneumonia. *Pathogens Global Health* **114**(6), 281–282. <https://doi.org/10.1080/20477724.2020.1785782> (2020).
28. Li, D. K. T., & Zhu, S. Contributions and challenges of general practitioners in China fighting against the novel coronavirus crisis. *Family Med. Commun. Health*, **8**(2), (2020).
29. Yin, Y. *et al.* General practitioner trainees' career perspectives after COVID-19: A qualitative study in China. *BMC Fam. Pract.* **22**(1), 1–9. <https://doi.org/10.1186/s12875-020-01364-x> (2021).
30. Lauriola, P. *et al.* On the importance of primary and community healthcare in relation to global health and environmental threats: lessons from the COVID-19 crisis. *BMJ Glob. Health* **6**(3), e004111 (2021).
31. Pardhan, S. & Drydakis, N. Associating the change in new COVID-19 cases to GDP per Capita in 38 European countries in the first wave of the pandemic. *Front. Public Health* **8**, 1065 (2020).
32. Drydakis, N. The effect of unemployment on self-reported health and mental health in Greece from 2008 to 2013: A longitudinal study before and during the financial Crisis. *Soc. Sci. Med.* **128**, 43–51. <https://doi.org/10.1016/j.socscimed.2014.12.025> (2015).
33. Mauro, M. & Giancotti, M. Italian responses to the COVID-19 emergency: Overthrowing 30 years of health reforms? *Health Policy* **125**(4), 548–552. <https://doi.org/10.1016/j.healthpol.2020.12.015> (2021).
34. Bellinger, C., Jabbar, M. S. M., Zaïane, O. & Osornio-Vargas, A. A systematic review of data mining and machine learning for air pollution epidemiology. *BMC Public Health* **17**(1), 1–19 (2017).
35. Ardabili, S. F. *et al.* Covid-19 outbreak prediction with machine learning. *Algorithms* **13**(10), 249 (2020).
36. An, C. *et al.* Machine learning prediction for mortality of patients diagnosed with COVID-19: A nationwide Korean cohort study. *Sci. Rep.* **10**(1), 1–11 (2020).
37. Cornelius, E., Akman, O. & Hrozencik, D. COVID-19 mortality prediction using machine learning-integrated random forest algorithm under varying patient frailty. *Mathematics* **9**(17), 2043 (2021).
38. Gupta, V. K., Gupta, A., Kumar, D. & Sardana, A. Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model. *Big Data Min. Anal.* **4**(2), 116–123 (2021).
39. Iwendi, C. *et al.* COVID-19 patient health prediction using boosted random forest algorithm. *Front. Public Health* **8**, 357 (2020).
40. Majhi, R., Thangeda, R., Sugasi, R. P., & Kumar, N. Analysis and prediction of COVID-19 trajectory: A machine learning approach. *J. Public Affairs*, e2537, (2020).
41. Kivrak, M., Guldogan, E. & Colak, C. Prediction of death status on the course of treatment in SARS-COV-2 patients with deep learning and machine learning methods. *Comput. Methods Prog. Biomed.* **201**, 105951 (2021).
42. Pourhomayoun, M. & Shakibi, M. Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health* **20**, 100178 (2021).
43. Watson, G. L. *et al.* Pandemic velocity: Forecasting COVID-19 in the US with a machine learning & Bayesian time series compartmental model. *PLoS Comput. Biol.* **17**(3), e1008837 (2021).
44. Yan, Li, H. Zhang, Jorge Goncalves, Yang Xiao, Maolin Wang, Yuqi Guo, Chuan Sun *et al.* A machine learning-based model for survival prediction in patients with severe COVID-19 infection. *medRxiv*; (2020). <https://doi.org/10.1101/2020.02.27.20028027>.
45. Karthikeyan, A., Garg, A., Vinod, P. K., & Priyakumar, U. D. Machine learning based clinical decision support system for early COVID-19 mortality prediction. *Front. Public Health*, **9**, (2021).
46. Kar, S. *et al.* Multivariable mortality risk prediction using machine learning for COVID-19 patients at admission (AICOVID). *Sci. Rep.* **11**(1), 1–11 (2021).
47. Wang, J. *et al.* A descriptive study of random forest algorithm for predicting COVID-19 patients outcome. *PeerJ* **8**, e9945 (2020).
48. Levin, A. T., Hanage, W. P., Owusu-Boaitey, N., Cochran, K. B., Walsh, S. P., & Meyerowitz-Katz, G. Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *Europ. J. Epidemiol.*, pp. 1–16, (2020).
49. Goic, M., Bozanic-Leal, M. S., Badal, M. & Basso, L. J. COVID-19: Short-term forecast of ICU beds in times of crisis. *PLoS ONE* **16**(1), 245272 (2021).
50. Board, G. P. M. *A world at risk: annual report on global preparedness for health emergencies* (World Health Organization, 2019).
51. Eurostat statistics. Available at [https://ec.europa.eu/eurostat/statistics-explained/index.php/Main\\_Page](https://ec.europa.eu/eurostat/statistics-explained/index.php/Main_Page)
52. Organization for Economic Co-operation and Development (OECD). State of Health in the EU. Country Health Profile 2019. Available at <https://www.oecd.org/health/country-health-profiles-eu.htm>
53. Kringos, D. S., Boerma, W. G., Hutchinson, A., Saltman, R. B. & World Health Organization. Building primary care in a changing Europe. World Health Organization. Regional Office for Europe, (2015).
54. Khaled, F., Mohamed, M. G. & Eyad, E. Random forests: From early developments to recent advancements. *Syst. Sci. Control Eng. Open Access J.* **2**(1), 602–609 (2014).
55. Andrew Y. Ng. Preventing "Overfitting" of Cross-Validation Data. In Proceedings of the Fourteenth International Conference on Machine Learning (ICML '97). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 245–253, (1997).
56. Robert E. Schapire. A brief introduction to boosting. In Proceedings of the 16th international joint conference on Artificial intelligence - Volume 2 (IJCAI'99). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1401–1406, (1999).

57. Muthukrishnan, R., Rohini, R. LASSO: A feature selection technique in predictive modeling for machine learning. *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*, Coimbatore, India, pp. 18–20. <https://doi.org/10.1109/ICACA.2016.7887916> (2016).

### Author contributions

All authors were involved in the research design and contributed to writing the manuscript. M.M. conceived the study, collected, and prepared data. M.G. analyzed the results and wrote the introduction. M.L. and M.P. conducted the analysis, analyzed the results, and prepared figure. All authors provided comments to the various draft and approved the final version.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-03120-2>.

**Correspondence** and requests for materials should be addressed to M.L.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021