Article

Novel Insights in Spatial Epidemiology Utilizing Explainable AI (XAI) and Remote Sensing

Anastasios Temenos *, Ioannis N. Tzortzis [ID]**, Maria Kaselimi** [ID]**, Ioannis Rallis, Anastasios Doulamis and Nikolaos Doulamis** [ID]

Department of Rural Surveying Engineering and Geoinformatics Engineering, National Technical University of Athens, 157 80 Athens, Greece; itzortzis@mail.ntua.gr (I.N.T.); mkaselimi@mail.ntua.gr (M.K.); irallis@central.ntua.gr (I.R.); adoulam@cs.ntua.gr (A.D.); ndoulam@cs.ntua.gr (N.D.)
* Correspondence: tasostemenos@mail.ntua.gr

Abstract: The COVID-19 pandemic has affected many aspects of human life around the world, due to its tremendous outcomes on public health and socio-economic activities. Policy makers have tried to develop efficient responses based on technologies and advanced pandemic control methodologies, to limit the wide spreading of the virus in urban areas. However, techniques such as social isolation and lockdown are short-term solutions that minimize the spread of the pandemic in cities and do not invert long-term issues that derive from climate change, air pollution and urban planning challenges that enhance the spreading ability. Thus, it seems crucial to understand what kind of factors assist or prevent the wide spreading of the virus. Although AI frameworks have a very efficient predictive ability as data-driven procedures, they often struggle to identify strong correlations among multidimensional data and provide robust explanations. In this paper, we propose the fusion of a heterogeneous, spatio-temporal dataset that combine data from eight European cities spanning from 1 January 2020 to 31 December 2021 and describe atmospheric, socio-economic, health, mobility and environmental factors all related to potential links with COVID-19. Remote sensing data are the key solution to monitor the availability on public green spaces between cities in the study period. So, we evaluate the benefits of NIR and RED bands of satellite images to calculate the NDVI and locate the percentage in vegetation cover on each city for each week of our 2-year study. This novel dataset is evaluated by a tree-based machine learning algorithm that utilizes ensemble learning and is trained to make robust predictions on daily cases and deaths. Comparisons with other machine learning techniques justify its robustness on the regression metrics RMSE and MAE. Furthermore, the explainable frameworks SHAP and LIME are utilized to locate potential positive or negative influence of the factors on global and local level, with respect to our model's predictive ability. A variation of SHAP, namely treeSHAP, is utilized for our tree-based algorithm to make fast and accurate explanations.

Keywords: XAI; COVID-19; pandemic; big data; remote sensing; NDVI; SHAP; LIME; machine learning; random forest

1. Introduction

Infectious diseases have a significant impact on global health. The spread of infectious diseases during a pandemic is an additional burden to the existing high level of challenges caused by chronic disease that modern healthcare systems need to manage. Lessons learned from countries' responses to crises such as coronavirus disease 2019 (COVID-19) [1] is a key factor for healthcare systems' resilience. During the pandemic [2], various strategies to prevent and mitigate the spread of COVID-19 were reported, using tools, response measures, technologies and public health functions and services. Policy makers, along with government health system managers, have developed response plans and tools to defend humanity against the pandemic crisis [3–5]. However, these measures, such as

social isolation and lockdown restrictions, are short-term and imposed to mitigate and eliminate the spread of the pandemic among citizens in urban environments [6,7].

Nowadays, climate change and land-use change impose additional challenges and novel opportunities for facilitating the spread of infectious viruses among previously geographically isolated species of wildlife [8]. Thus, there is an emerging need for identifying these newly coming risk factors across our living environments, in order to establish long-term measures to prevent the spread of these diseases and secure the urban environments from future challenges. In order to establish a health-centered urban planning methodology, the first thing to do is to find out the means for the spread of the pandemic and identify the risk factors that promote this expansion (see Figure 1).
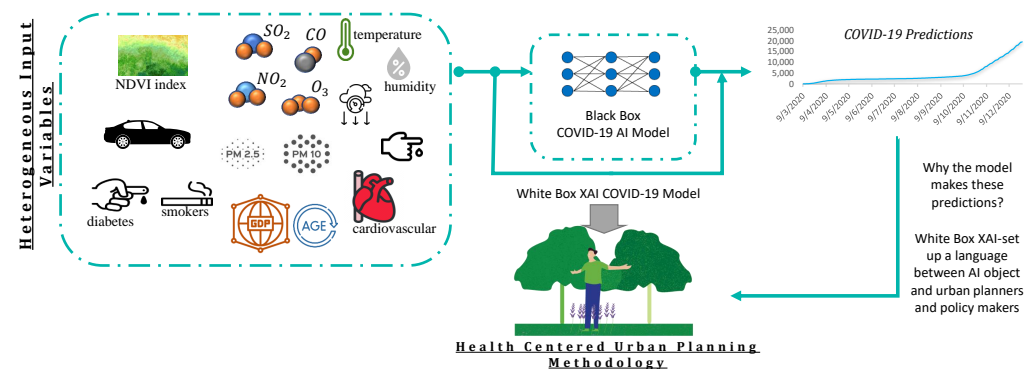

Figure 1. The proposed concept in spatial epidemiology using explainable AI and remote sensing. Explainable AI is the "language" to translate the results of the AI model, in an easy to understand and interpretable way, to the urban planners, to support them in the decision-making process for health-centered urban planning decisions.

Spatial epidemiology has been emerging in the era of big data growth and rapid development in geoinformatics. Thus, there is an emerging need to monitoring the long-term effects of environmental, behavioural, psychosocial and biological factors on health-related states and events and their underlying mechanisms. In the literature, recently, it has been reported that factors related to urban challenges [9], atmospheric pollution [10] and climate change factors [11,12], that pre-exist in urban environments can possibly trigger the rapid spread of the virus within a community.

Modelling analysis of COVID-19 outbreaks is a common process to predict confirmed COVID-19 cases and deaths using Artificial Intelligence (AI) and strongly assists national health agencies in developing response plans and mitigation measures [13–15]. Machine learning (ML) and especially ensemble (supervised) learning algorithms are dominant in the field of regression and time-series prediction tasks, achieving high performance regarding dataset complexity [16–20]. ML algorithms accurately predict COVID-19 cases and deaths, but now the problem is shifted in identifying the risk factors that cause the spread in order to establish countermeasures to prevent the spread of the pandemic in urban environments. Thus, experts of the pandemic cannot have their plans rely on black-box procedures, as the agnostic way of analyzing non-linear signals cannot be explained with the state-of-the-art framework or fundamental statistical analysis [21,22].

A high-interpretability ML algorithm is proposed here due to the feasibility for users to comprehend why certain decisions or predictions have been made. So, the link between: (1) escaping black-box procedures and (2) the investigation of models' decisions is the calculation of advanced explanations. Explanations are provided to indicate the contribution of a single variable to the model's final prediction in an easy-to-understand way by the pandemic control planers (e.g., feature importance plots). Two well-known state-of-the-art explainable frameworks are utilized here, Shapley additive explanations (SHAP) [23] and local interpretable model-agnostic explanations (LIME) [24]. SHAP (SHapley Additive exPlanations) utilizes and optimizes the Shapley values from game theory, in order to measure the contribution of each

feature in the final outcome. LIME is a model-agnostic method that calculates explanations locally. LIME takes into account a specific area of the dataset and along with the predictions of the trained machine learning algorithm, it trains interpretable models, which are weighted by the proximity of the sampled instances to the instance of interest.

In this paper, we evaluate the effectiveness of an interpretable machine learning algorithm on eight European cities for the 2 years of COVID-19. Our model asses the impact of heterogeneous spatio-temporal data that are related to COVID-19 morbidity and mortality, and provides interrelationships between the multidimensional factors through an Explainable AI framework. Among the heterogeneous datasets there are: (1) Earth observation data for monitoring the greenery in urban spaces; (2) socio-economic factors; (3) health-related data; (4) atmospheric data that refer to air pollution and climate change; and lastly (5) mobility trends inside cities or centralized administrations. In particular, we develop a random-forest regression model that predicts daily COVID-19 cases and deaths. To demonstrate its effectiveness, we also compare our model's predictive ability with other machine learning algorithms. Furthermore, explainable frameworks are implemented on top of the predictive modelling framework, in order to detect each feature contribution to the model's outbreak (1) on the global level using SHAP and (2) locally utilizing the LIME framework as well. In the experimental results plots such as feature importance, summary plots, dependence plots and heatmaps enhance the feature analysis, with respect to each risk factor.

Contribution

In this paper, we describe and analyze the spatio-temporal variations in COVID-19 disease with respect to demographic, environmental, socioeconomic and infectious risk factors. The recent advances in geographic information systems, as well as the availability of high-resolution, geographically referenced health and remote sensing data have created unprecedented new opportunities to investigate environmental and other factors in order to explain local geographic variations in disease. An explainable AI framework is adopted to underline the links between COVID-19 morbidity/mortality with environment, socioeconomic and health-related issues, proposing a transparent method for the prioritization of the attributes' importance. In particular, we handle a large amount of heterogeneous data with various attributes describing spatial and temporal variability by examining different areas and time-periods of COVID-19 transmission, to identify the effects and risk factors on COVID-19 cases and deaths, proposing a scalable and self-explanatory machine learning model. This is achieved with the adoption of a robust explainable AI framework to explain the rationale behind the predictive models' learning algorithm and to discover the links between environmental, atmospheric, health and socio-economic factors vis-a-vis COVID-19 morbidity/mortality rates in urban areas. Interpretable AI is adopted here as a tool that will enable us to learn from and be inspired by AI predictive modelling tools for COVID-19 to gain knowledge about the risk factors that promote disease transmission in urban environments.

2. Related Work

Nowadays, the rapid transmission rate of the COVID-19 pandemic has lead to research works that study indicators that affect the spread of the virus, leading to infection, severe infection or even to mortality. The majority of the published papers present prediction methods of the pandemic's spread, taking into account various data related to air quality, health, socio-economics, mobility, etc. On the contrary, only a few papers have developed or exploited techniques in the direction of explainability in order to provide more accurate and robust systems.

2.1. COVID-19 Prediction Models Using ML

Sarkodie et al., in their work, propose several regression models with the aim to provide predictions of COVID-19's spread and mortality, taking into account environmental, health and socio-economic data such as air temperature, age, diabetes prevalence index,

$PM_{2.5}$ concentration, GDP per capita, etc. [25]. Even though the models include a variety of data, taking into consideration a large number of cities, the selected time period (1 January to 11 June 2020) seems to be too short for the present time. A neural network architecture is presented in the work of [26], which is based on a trained Long Short-Term Memory (LSTM) network aiming to predict the upcoming daily COVID-19 cases. It was considered to be an adequate approach, since it produced a low average relative error compared to the state-of-the-art system provided by the Google Cloud forecasting service. Additionally, for the purposes of this work, mobility data was used along with the environmental and COVID-19-related data. However, the results in this study [26] are spatially limited, since the proposed model was validated using infection records from cities of Japan. In the work of Zoran et al. [27], the authors investigate the potential correlations between air pollution and COVID-19 related data in the metropolitan area of Milan, Italy. According to their results, some significant relationship among O_3, NO_2 and COVID-19 spread exists. However, their research is spatially limited to the area of Milan, the time period that was taken into account is short (January–April 2020) and there is no sophisticated pipeline for the explanation of the findings. In the same direction, the authors in [28] investigate the potential relationship between a long-term exposure to air pollution and the COVID-19 severity in terms of daily cases and deaths. The use of an ecological regression analysis provided significant results, such as that higher $PM_{2.5}$ exposure can potentially lead to higher COVID-19 mortality rates. This work is restricted to environmental data and does not include any tools for further explanations.

2.2. Explainable AI Frameworks

AI has attracted the attention of people around the world in a huge variety of topics. Extensive effort has been put into the area of medicine in general, mostly by researchers and doctors. There are various studies that cope with the challenges in cancer diagnosis and treatment by utilizing segmentation and classification AI tools, such as in [29–32]. As we mentioned above, several techniques have been studied in order to develop sophisticated tools, including AI features, that would assist in discovering and understanding the unknown factors that promote COVID-19's spread. Since this kind of utilization of AI technologies is related to the health of people, it is crucial to ensure their accuracy and integrity. This is one of the reasons why it is essential to develop proper mechanisms to explain the decisions of AI models. Recently, there have been various works published, such as [33–36], where the explanations are provided by the proposed frameworks in a variety of target-applications.

For the time being, though, there are limited frameworks that have utilized techniques for explainability features for COVID-19 spatio-temporal modelling. Decision tree, logistic regression, naive Bayes, support vector machine and artificial network approaches were utilized and compared in the work of Muhammad et al. [37], where an effort to model the levels of progression of COVID-19 infection was done. Moreover, correlation coefficient analysis was adopted to determine the strong relationship among the numerous health features of the included dataset. In this way, the explainability feature was added to the proposed pipeline. On the contrary, the dataset is limited to health-related data for the area of Mexico. In [38], the authors propose a more sophisticated framework to assist the investigation of the causality behind the associations that are discovered by the utilization of machine learning techniques. Despite the well-designed structure and the robustness of this framework, it was restricted to the environmental data analysis in relation to COVID-19's evolution.

3. Mathematical Formulation of the Pandemic Spatio-Temporal Evolution

Our proposed framework for interpreting predictions on COVID-19 daily cases and deaths is presented in Figure 2. In our case study, we have chosen eight pilot cities-capitals from the area of Europe: Athens, Budapest, Prague, Madrid, Rome, Paris, Birmingham and Berlin. Heterogeneous data from different sources are utilized in order to create a larger dataset to investigate and exploit all the potential correlations among the data. The pro-

posed dataset consists of environmental, atmospheric, health-related, socio-economic data and data related to daily mobility trends from places of interest.
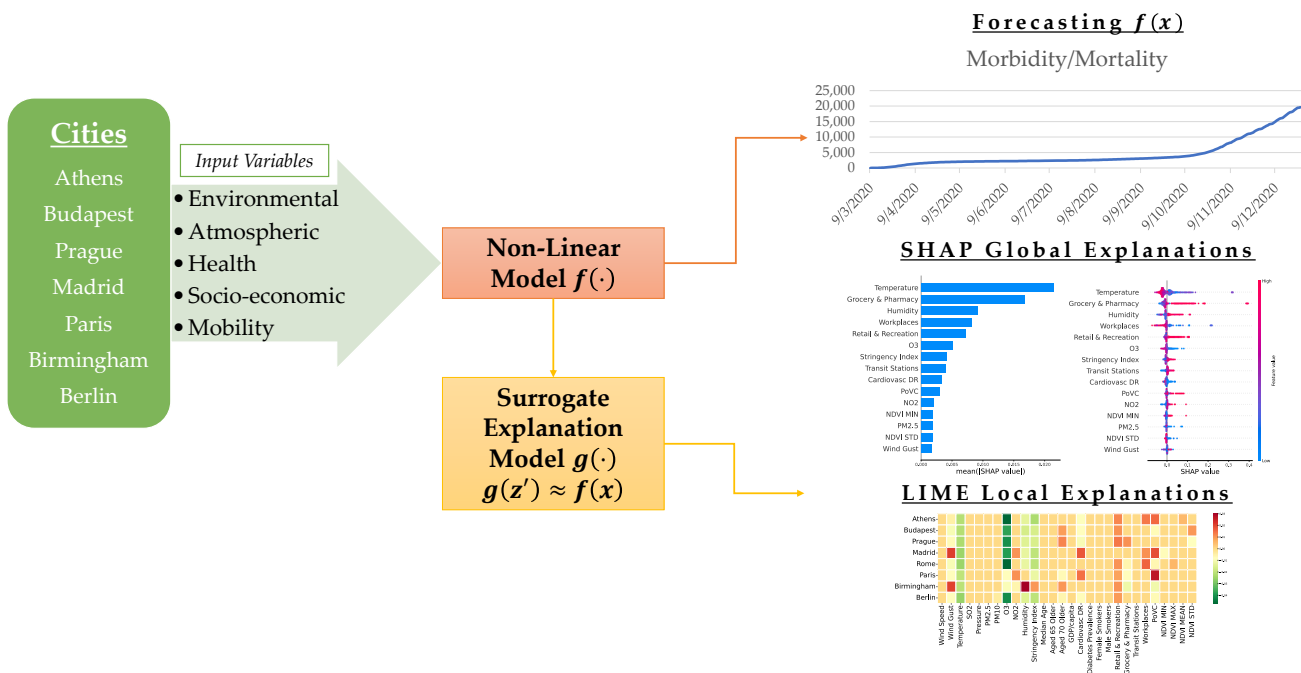

Figure 2. The proposed machine learning model is responsible for interpreting predictions on COVID-19 daily cases and deaths. On one hand, the proposed model forecasts COVID-19 cases and deaths, and on the other hand it discovers possible links from SHAP and Lime explainability frameworks.

With the aim to study the dynamics of COVID-19, we introduce a non-linear model that incorporates all the aforementioned data towards the direction of COVID-19 progression prediction. Let $\mathcal{X} = \{\mathcal{X}_{env}, \mathcal{X}_{atm}, \mathcal{X}_{soc}, \mathcal{X}_{heal}, \mathcal{X}_{mob}\}$, be a set containing sub-sets of different-type of values, where $\mathcal{X}_{env} = \{x_{k_1}^t : k_1 = 1, \ldots 5\}$ denotes the environmental data, $\mathcal{X}_{atm} = \{x_{k_2}^t : k_2 = 1, \ldots 11\}$ denotes the atmospheric data, $\mathcal{X}_{heal} = \{x_{k_3}^t : k_3 = 1, \ldots 4\}$ denotes the health-related data, $\mathcal{X}_{soc} = \{x_{k_4}^t : k_4 = 1, \ldots 4\}$ denotes the socio-economic data and $\mathcal{X}_{mob} = \{x_{k_5}^t : k_5 = 1, \ldots 6\}$ denotes the mobility data. Note that in the above subsets, *t* refers to the time index given the fact that the data are represented in the form of time-series. The non-linear model uses the above values so as to predict two types of cases; the first one corresponds to the predicted daily cases, denoted as y_{pdc}, and the second one to the predicted daily deaths, denoted as y_{pdd}. Therefore, one can use a non-linear model to predict the output y_j, where *j* can be either y_{pdc} or y_{pdd} as

$$y_j(t) = f(\mathcal{X}_{env}, \mathcal{X}_{atm}, \mathcal{X}_{heal}, \mathcal{X}_{soc}, \mathcal{X}_{mob}). \tag{1}$$

4. Spatio-Temporal Modeling of Heterogeneous Big Data for COVID-19 with Tree-Based Ensemble Learning

Random Forest (RF) is an ensemble machine learning algorithm used for supervised learning and is utilized to carry out either classification or regression tasks with high performance [39–41]. Ensemble learning is a machine learning technique that combines several base models (e.g., decision trees) in order to achieve better predictive performance [42,43]. Three dominant ensemble learning techniques are: (1) BAGGing, (2) Stacking and (3) Boosting. Note that in this work we compare algorithms from every field of ensemble learning. BAGGing derives from Boosttrapping and AGGregation, which in combination form one ensemble model. A large number of decision trees make predictions on boosttrapped subsamples of the initial input, while an aggregation over the results carries out the final prediction. One well-known ensemble learning algorithm is the proposed RF. In particular,

RF is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the input and uses averaging to improve the predictive accuracy and control over-fitting. In this work, we utilize an RF regressor as the non-linear model, which is used to predict daily COVID-19 cases and deaths. Furthermore, RF is an intepretable algorithm that provides a quantified contribution of each feature to final prediction.

4.1. Improving the Interpretability of the Random Forest Regressor

4.1.1. Shapely Additive Explanation (SHAP)

Shapley additive explanation (SHAP) is a method that explains individual predictions by the processed data features, first introduced by Lundberg and Lee (2017) in [23]. The SHAP explanation method computes Shapley values from coalitional game theory. According to them, a prediction can be explained by assuming that each feature value of the instance is a "player" in the coalition, where the prediction is the payout. Shapley values depict the contribution of the "payout" among the features. SHAP values are defined as the coefficients ϕ_i of the additive surrogate explanation model:

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z_i', \tag{2}$$

where z_i' represents the binary variables ($z_i' \in \{0,1\}^M$) and ϕ_i is the feature-attribution values ($\phi_i \in \Re$). In other words, z_i' represents a feature that is being observed when $z_i' = 1$ and an unknown feature when $z_i' = 0$. M denotes the number of input features and g provides the explanation for the model.

SHAP values are the single unique solution of each ϕ_i in the class of additive surrogate explanation models $g(z')$ that satisfies three desirable properties: local accuracy, missingness and consistency [44]. To compute SHAP values, we denote $f_S(x)$ as the model output restricted to the feature subset $S \subset M$, and the SHAP values are then computed based on the classic Shapley values:

$$\phi_i = \sum_{S \subset M \setminus i} \frac{|S|!(|M| - |S| - 1)!}{|M|!} (f_x(S \cup i) - f_x(S)) \tag{3}$$

where this is the conditional expectation $f_x(S) = E[f(x) \mid x_S] = E_{x_{\overline{S}}|x_S}[f(x)]$, and x_S is the sub-vector of x restricted to the feature subset S, and $\overline{S} = N \setminus S$.

SHAP is a framework that has a fast implementation for tree-based models due to the proposed TreeSHAP, a variant of SHAP for tree-based machine learning models such as random forests. TreeSHAP was introduced by Lundberg et al. (2018) [44] as a high speed algorithm for estimating SHAP values of tree ensembles, that also addresses the problem of inconsistent feature attribution method. Furthermore, SHAP aims to explain the prediction of an input set of features by computing the Shapley values and provide a variety of tools for model explanation, e.g., feature importance, impact with summary plot and dependence plot.

4.1.2. Local Interpretable Model-Agnostic Explanations (LIME)

Ribeiro et al., in [24], proposed a novel explanation technique named local interpretable model-agnostic explanations (LIME), that interprets individual model predictions based on locally approximating the model around a given prediction. LIME is a framework that can interpret and understand any model (model-agnostic), by calculating reliable explanations on a local scale, where we assume that every model is linear. Explanations are gathered by approximating linear models, e.g., decision trees, near a given sample. LIME explanations are calculated by the following equation:

$$\text{explanation}(x) = \underset{g \in \mathcal{G}}{\arg\min} \, \mathcal{L}(f, g, \pi_x) + \Omega(g) \tag{4}$$

where $g \in \mathcal{G}$ is an explainable model, which belongs to a class of potentially interpretable models $\mathcal{G}$. To continue, g is responsible fir approximating our black-box model f, near a proximity measure π_x, which defines the size of the neighborhood around a given instance x. Furthermore, $\Omega(g)$ denotes the model's complexity according to all $g \in \mathcal{G}$. In order to obtain interpretable approximation of the non-linear model, LIME aims to on one hand minimize $\mathcal{L}(f, g, \pi_x)$, which is a measure of how uncertain g is (i.e., local fidelity), and on the other hand keep $\Omega(g)$ low enough so that it can be interpretable by users (i.e., interpretability).

In this work , we utilize LIME's accurate and fast explanations in order to investigate potential links between cities and factors. Locality enables us to take a closer inspection of the entire dataset where the neighborhood around a given instance x are data related to each city.

5. Experimental Results and Discussion

5.1. Dataset Description

In order to evaluate the performance of machine learning algorithms, we consider a collection of spatio-temporal datasets that derive from multiple heterogeneous sources. The datasets are divided into five classes, including environmental, atmospheric, health, socio-economic and mobility data. We utilize data from eight large European cities: (1) Athens, (2) Budapest, (3) Prague, (4) Madrid, (5) Rome, (6) Paris, (7) Birmingham and lastly (8) Berlin, with a time period spanning from 1 January 2020 to 31 December 2021, covering 2 years since the beginning of the pandemic. Table 1 shows the summary of data utilized.

Table 1 includes the names of the variables, a short description for each variable and its mathematical notation, as well as the class that the variable belongs to along with the variable units, the source where the data was acquired and last the statistical values of: (1) mean, (2) minimum, (3) maximum and (4) standard deviation.

Table 1. Description of the data inputs into the proposed non-linear model.

Input Variables	Description	Class	Notation	Units	Source	Mean	Std	Min	Max
Wind speed	Avg. daily wind speed	Atm	$\mathcal{X}_{atm}$	m/s	AQ	3.2	1.65	0.2	13.1
Wind Gust	Avg. daily wind dust	Atm	$\mathcal{X}_{atm}$	m/s	AQ	6.91	3.73	0.4	26.2
Pressure	Avg. daily atmospheric pressure values	Atm	$\mathcal{X}_{atm}$	mb	AQ	1014.64	8.31	973	1041
Temperature	Avg. daily temperature values	Atm	$\mathcal{X}_{atm}$	°C	AQ	14.53	7.72	−6.4	36.1
Humidity	Avg. daily humidity	Atm	$\mathcal{X}_{atm}$	%	AQ	68.41	16.98	20	98
SO$_2$	Avg. daily sulfur dioxide values	Atm	$\mathcal{X}_{atm}$	µg/m^3	AQ	1.75	1.32	0.1	9.9
PM$_{2.5}$	Avg. daily fine particulate matter ($d < 2.5$)	Atm	$\mathcal{X}_{atm}$	µg/m^3	AQ	42.74	20.01	5	171
PM$_{10}$	Avg. daily fine particulate matter ($d < 10$)	Atm	$\mathcal{X}_{atm}$	µg/m^3	AQ	17.93	9.01	3	77
O$_3$	Avg. daily ground level ozone values	Atm	$\mathcal{X}_{atm}$	µg/m^3	AQ	21.71	9.47	0.8	55.2
NO$_2$	Avg. daily nitrogen dioxide values	Atm	$\mathcal{X}_{atm}$	µg/m^3	AQ	10.04	5.01	0.7	43.7
Cardiovascular DR	Country cardiovascular death rate	Heal	$\mathcal{X}_{heal}$	/10^5	OWD	156.41	64.43	86.06	278.3
Diabetes Prevalence	Country % number of diabetics	Heal	$\mathcal{X}_{heal}$	%	OWD	5.98	1.47	4.28	8.31
Male smokers	Country % male smokers per city	Heal	$\mathcal{X}_{heal}$	%	OWD	34.81	7.8	24.7	52
Female smokers	Country % female smokers per city	Heal	$\mathcal{X}_{heal}$	%	OWD	27.24	4.95	19.8	35.3
Median age	Population median age per city	Soc	$\mathcal{X}_{soc}$	%	OWD	44.4	2.23	40.8	47.9
Aged 65 older	Population over 65	Soc	$\mathcal{X}_{soc}$	%	OWD	20.04	1.48	18.52	23.02
Aged 70 older	Population over 70	Soc	$\mathcal{X}_{soc}$	%	OWD	13.73	1.61	11.58	16.24
GDP per capita	Gross Domestic Product	Soc	$\mathcal{X}_{soc}$	$	OWD	34,233.25	6265.04	24,574.38	45,229.25
PoVC	% of land cover in vegetation	Env	$\mathcal{X}_{env}$	%	Copernicus	0.46	0.06	0.3	0.58
NDVI mean	Mean value of NDVI image	Env	$\mathcal{X}_{env}$	-	Copernicus	0.27	0.12	0	0.5
NDVI max	Max value of NDVI image	Env	$\mathcal{X}_{env}$	-	Copernicus	0.95	0.09	0.63	1
NDVI min	Min value of NDVI image	Env	$\mathcal{X}_{env}$	-	Copernicus	−0.78	0.21	−1	−0.17
NDVI std	Std. value of NDVI image	Env	$\mathcal{X}_{env}$	-	Copernicus	0.19	0.04	0.06	0.27
Retail Recreation	Daily mobility trends for retail and recreation	Mob	$\mathcal{X}_{mob}$	%	Google	−33.84	22.98	−97	19
Grocery Pharmacy	Daily mobility trends for grocery and pharmacy	Mob	$\mathcal{X}_{mob}$	%	Google	−4.99	21.33	−95	182
Transit Stations	Daily mobility trends for transit stations	Mob	$\mathcal{X}_{mob}$	%	Google	−36.12	18.26	−93	12
Workplaces	Daily mobility trends for places of work	Mob	$\mathcal{X}_{mob}$	%	Google	−33.28	20.54	−92	95

With regards to the atmospheric variables, we adopted atmospheric data from the Air Quality (AQ) Open Data Platform (https://openaq.org/, accessed on 3 August 2017), that provides min, max, median and standard deviation values for each of the air pollutant species (i.e., nitrogen dioxide (NO_2), ozone (O_3), particulate matter 2.5 ($PM_{2.5}$), particulate matter 10.0 (PM_{10}), dioxygen (SO_2)) as well as meteorological data (i.e., wind, temperature). We only consider the median value for each indicator as a representative daily sample for each city of interest.

The outputs of our model, which are COVID-19 confirmed cases and deaths, are obtained by the Our World in Data (OWD) platform [45]. From this dataset, we have also adopted the socio-economic (i.e., gdp, age) and health-related factors (i.e., cardiovascular death rate, diabetes prevalence, smokers) of our model. Apart from them, OWD also provides government response stringency index, which is a composite metric based on nine response indicators including school closures, workplace closures and travel bans, ranging between 0 and 100. The highest values, e.g., 100, indicate the strictest response. Here, we highlight that we utilize this index in the explainability part of our algorithm in order to interpret the results of our algorithms (see Section 5.4).

Urban vegetation indices are derived from remote sensing imagery data with data derived from the Copernicus Sentinel-2 mission (https://scihub.copernicus.eu/, accessed on 10 October 2013) for each city of interest in a common timeline. The Normalized Difference Vegetation Index (NDVI) is calculated, as shown in Equation (5), for the quantification of the urban vegetation. In particular, NDVI is calculated from a normalized transform of the near-infrared (NIR) and red (RED) amounts of the electromagnetic spectrum, that are reflected by the vegetation and captured by the sensor of the satellite. The formula is based on the fact that vegetation absorbs RED, whereas it strongly reflects NIR. NDVI values range from −1 to +1, where negative values correspond to an absence of vegetation.

$$NDVI = \frac{NIR - RED}{NIR + RED} \tag{5}$$

For each satellite imag,e we calculate the NDVI and acquire the following four statistical metrics: (1) mean, (2) minimum, (3) maximum and (4) standard deviation. Furthermore, we quantify the Urban greenness with a two-step procedure. Firstly a threshold operation is done to dissever the space into green and non-green areas. Secondly we divide the outcome with the total number of pixels to outcome the percentage of land cover in vegetation (PoVC). Note that we threshold each image with the NDVI mean, as its benefit is two-fold: (1) as a value it is automatically extracted from each image and (2) it estimates the average greenness regardless of the radiometric parameters of each image. Environmental data acquisition procedure is also illustrated in Figure 3.
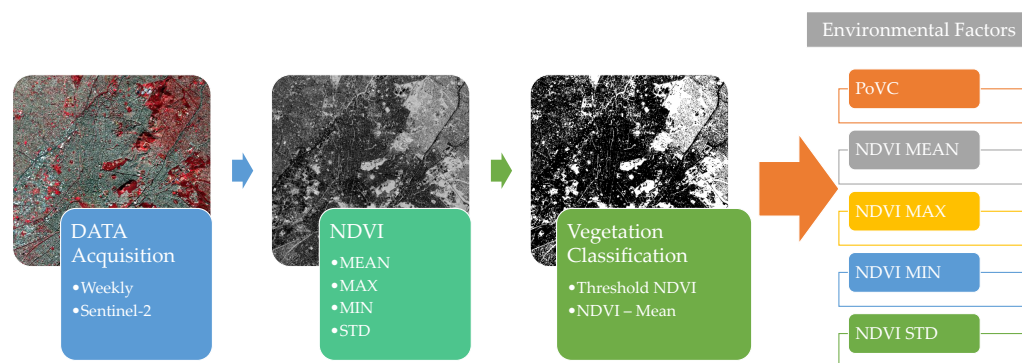

Figure 3. The environmental data acquisition step from remote sensing resources and the adopted process for the calculation of the urban greenery.

Mobility data are obtained by Google COVID-19 Community Mobility Reports (https://www.google.com/covid19/mobility/, accessed on 15 April 2022) and provide movement

trends over time by region, across different categories. These categories are clustered as follows: (1) places of retail and recreation (restaurants, shopping centers etc.); (2) areas of groceries and pharmacies (grocery markets, food warehouses, farmers markets); (3) parks; (4) transit stations (subway, bus and train stations); (5) workplaces; and (6) residential areas.

5.2. Model Performance Evaluation

5.2.1. Comparisons for Different Machine Learning Models

Here, we train a single model that includes all eight cities, forming a dataset of 4854 records in total, with daily pandemic-related data spanning across 2 years. The dataset was split in 70/30 train/test sets. Note that for both the train and test set the number of records are equally distributed for each city and contain values from all cities and all years of evaluation.

Table 2 presents comparisons according to the performance of the proposed Random Forest (RF) in the regression task, accompanied with the following machine learning approaches: (a) linear regression (LR), (b) decision tree (DT), (c) support vector regressor (SVR), (d) lasso regression (Lasso), (e) Gaussian process regressor (GPR), (f) multi-layer perceptron (MLP), (g) extreme gradient boosting regressor (XGBoost) and (h) light gradient boosting regressor (LightGBM). The metrics that were selected for the evaluation of regression task were the root mean square error (RMSE) and mean absolute error (MAE). These metrics are defined as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i(t) - \hat{y}_i(t)| \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i(t) - \hat{y}_i(t))^2} \tag{7}$$

where $\hat{y}_i(t)$ is the predicted value of the *i*-th sample, $y_i(t)$ is the corresponding true value and n denotes the number of samples.

Table 2. The overall performance of machine learning algorithms.

Machine Learning Algorithm	Cases (/10^6 People)		Deaths (/10^6 People)	
	RMSE	MAE	RMSE	MAE
Linear regression	436.12	268.28	3.63	2.41
Decision tree regressor	380.41	210.23	4.00	2.62
Support vector regressor	340.43	249.57	3.22	2.42
Lasso regression	498.76	267.02	4.94	3.65
Gaussian process regressor	436.14	268.13	3.63	2.41
Multi-layer percepton	252.13	149.39	2.98	1.87
XGBoost regressor	209.86	125.44	2.60	1.56
Light GBM regressor	208.40	97.63	2.20	1.18
Proposed random forest regressor	**192.44**	**93.76**	**2.15**	**1.12**

MAE measures the average magnitude of the error in a set of forecasts, without considering their direction, while RMSE is a quadratic scoring rule which measures the average magnitude of the error. MAE and RMSE are used complementarily one another to diagnose the variation in the errors in a set of forecasts. The RMSE will always be larger or equal to the MAE: the greater the difference between them, the greater the variance in the individual errors in the sample. If the RMSE is equal to MAE, then all the errors are of the same magnitude.

From Table 2 it is observed that Lasso regression results in the lowest score, regardless of the evaluation metric among the eight algorithms considered for both cases and deaths. Focusing on both RMSE and MAE metrics, RF achieved the highest performance in both cases and deaths predictions. Here, we highlight that the differences between the performance of LightGBM and the proposed RF are negligible.

The difference between RMSE and MAE is almost double, indicating that either spatial variability (different cities) or temporal variability (difference between the years and how

COVID-19 was treated) introduces errors with high variability, and causes difficulties during the training process of the model and worsens its performance. To overcome this issue, and dive into the explainability technique, in the next section (see Section 5.2.2), we evaluate the performance of the model for different time periods. The first testing period is during the year 2020, and the next period is in the year 2021, in an attempt to identify which year introduces the largest error.

5.2.2. Temporal Variability of the Performance Errors—Analysis for Different Time Periods

According to Table 3, all the machine learning models are fitted in a smoother manner for 2020-related data, which is justified from the RMSE and MAE results. Lasso regression, regardless of the year and the performance metric, results in the lowest score.

Table 3. The effect of the COVID-19 pandemic during two different time periods. The table shows the performance evaluation of the machine learning algorithms for two different testing periods.

ML Algorithm	2020				2021			
	Cases (/10^6 People)		Deaths (/10^6 People)		Cases (/10^6 People)		Deaths (/10^6 People)	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Linear regression	149.70	93.85	3.31	2.37	549.06	359.03	3.96	2.54
DT regressor	156.39	98.70	3.75	2.38	503.17	303.63	4.19	2.74
SVM regressor	133.15	89.79	2.74	1.97	354.89	262.07	3.63	2.52
Lasso regression	193.34	136.94	5.00	3.91	723.90	399.82	5.40	3.60
GP regressor	149.68	94.00	3.32	2.37	549.15	359.14	3.97	2.54
MLP regressor	146.96	95.50	2.61	1.69	296.33	181.64	3.42	2.06
XGBoost egressor	116.78	63.58	2.43	1.40	255.89	144.41	2.89	1.45
LightGBrM regressor	112.12	56.30	2.33	1.23	240.39	124.76	2.86	1.24
RF regressor	**111.40**	**54.04**	**2.27**	**1.16**	**226.03**	**110.25**	**2.65**	**1.12**

Random Forest achieves the highest performance in 2020 and 2021 year data individually. LightGBM achieves similar performance in 2020, although in 2021 it seems that has larger deviations compared to RF, but they remain negligible.

T he difference between the years can only be described as a prior form of explainable AI. Machine learning models struggle to make robust predictions and one of the major factors is the introduction of vaccinations in 2021 as an additional solution to the pandemic control, gradually leaving behind strict lockdown approaches [46–48].

5.2.3. Spatial Variability of the Performance Errors-Per City Analysis

In this section, we continue our analysis, and apart from the yearly evaluation on regression metrics, we study the effect of each city in the performance error. Thus, Tables 4 and 5 show the performance of the proposed model at each city and at each year respectively. According to Table 4, Rome is the city in which the model achieves the highest predictive performance in COVID-19 cases, followed by Athens and Berlin in the rank. The city that achieves the lowest RMSE and MAE is Prague. Apart from cases, both Athens and Berlin achieve the highest predictive performance. As in the previous metric, Prague continues to have the lowest performance for deaths according to RMSE and MAE.

5.2.4. Spatio-Temporal Variability of the Performance Errors

Table 5 depicts the results combining the two above mentioned approaches in Sections 5.2.2 and 5.2.3. As is shown, Athens in 2020 is the city that RF fits with the highest score among all other cities. Madrid, however, achieves the lowest performance for 2020 in both cases and deaths. In the following year for the city of Rome, the model fits better to the input data, according to the results. The city of Athens along with Birmingham and Berlin have, in general, good performance, in contrast to Prague which seems to have the lowest model performance in terms of cases.Although Prague has a moderate performance in

deaths, city of Budapest has the lowest score in both RMSE and MAE. Unlike Berlin, input data of this year seem to have fit from the proposed model with the best possible accuracy.

Table 4. The effect of COVID-19 pandemic for different cities. The table shows the performance evaluation of the random forest algorithm for the different cities.

City	Cases (/10^6 People)		Deaths (/10^6 People)	
	RMSE	MAE	RMSE	MAE
Athens	149.79	69.50	1.31	0.84
Budapest	255.22	118.34	3.67	1.92
Prague	293.91	163.54	2.16	1.25
Madrid	245.08	138.69	2.62	1.20
Rome	**72.38**	**39.22**	1.26	0.76
Paris	196.97	102.83	2.16	1.25
Birmingham	87.85	56.80	1.60	1.07
Berlin	118.48	69.11	**0.97**	**0.60**

Table 5. The spatial and temporal variability of the performance errors in the proposed RF model.

City	2020				2021			
	Cases (/10^6 People)		Deaths (/10^6 People)		Cases (/10^6 People)		Deaths (/10^6 People)	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
Athens	**53.11**	**26.11**	**1.06**	**0.56**	105.46	74.07	1.21	0.73
Budapest	115.18	62.46	1.58	0.91	297.97	150.31	5.25	2.57
Prague	189.40	92.46	2.48	1.12	430.92	224.67	4.25	2.06
Madrid	144.78	74.20	3.92	2.12	201.98	104.45	1.77	0.97
Rome	58.97	35.72	1.42	0.83	**86.87**	**45.95**	0.90	0.53
Paris	118.13	68.06	2.64	1.51	218.63	125.18	1.59	0.78
Birmingham	63.86	43.93	2.19	1.50	99.20	77.95	0.89	0.64
Berlin	61.21	36.02	1.00	0.62	132.55	79.44	**1.00**	**0.52**

All performance differences of the model according to both RMSE and MAE corresponding to each year and city are due to the heterogeneous nature of the data. These differences can guide researchers to possible explanations but can not provide robust feature importance and feature contribution to either global or local explanations of the pandemic. Therefore, explainable frameworks are employed in this work so as to identify how heterogeneous factors affect the spread of the pandemic.

5.3. Global and Local Explanations

The aim is to improve our model's interpretability using an explainable AI framework. Figure 4 illustrates the results using (a) feature importance and (b) summary plot diagrams. Feature importance shows, quantitatively, how features contribute in the final model output (global explanations), while the summary plot depicts the indication of the relationship between the value of a feature and the impact on the prediction (local explanations). Figure 4a,c illustrate the feature importance for cases and deaths models respectively, while summary plots are illustrated in Figure 4b,d.

SHAP feature importance depicts the average of absolute Shapley values per feature across the input dataset. Afterwards, the features are plotted in bars with a decreasing importance, providing a better understanding of the most important factors. As regards the model that predicts the confirmed cases, the results suggest that temperature variable is the most important indicator, following with three mobility trends from: (a) grocery and pharmacy, (b) retail and recreation and (c) workplaces. Humidity and mobility trends from transit stations along with ground-level ozone air pollution (O_3) follow the rank of importance. Other variables contribute to the models' outcomes too, and are illustrated in feature importance plot, but they have a minimal contribution compared to the previously mentioned variables. Regarding the COVID-19 deaths, temperature is the most important variable as shown in Figure 4c. Stringency index seems to have a significant

importance along with the percentage of vegetation cover. Moreover, two mobility trends contribute in the death model predictions and are: (a) from workplaces and (b) from transit stations. Lastly, GDP per capita along with people aged 70 or older and cardiovascular death rate, contribute as country-profile factors but with lower Shapley values. Feature importance plot is a useful, fast and understandable way to demonstrate which factor affects the model predictions the most, but it does not provide information beyond the importances. So, apart from the feature importance plot, Figure 4 contains SHAP summary plots for feature based analysis. Each point on the summary plot is a Shapley value of an individual feature and a sample. Similar to feature importance, summary plot also sorts the variables by the summary of SHAP values magnitudes over all samples. Then, it uses SHAP values to show the distribution of the impact that each feature has on the model output. The *y*-axis shows the variables/features and the *x*-axis indicates the SHAP values. Instances that have negative SHAP values are linked with decreased predictions, while positive ones have a high positive contribution. Color depends from the value of the feature in a scale from low to high. In more detail for the summary plot, red dots indicate high feature values, whereas blue dots represent low feature values. Furthermore, the distribution of the Shapley values per feature can be located by overlapping points in *y*-axis.
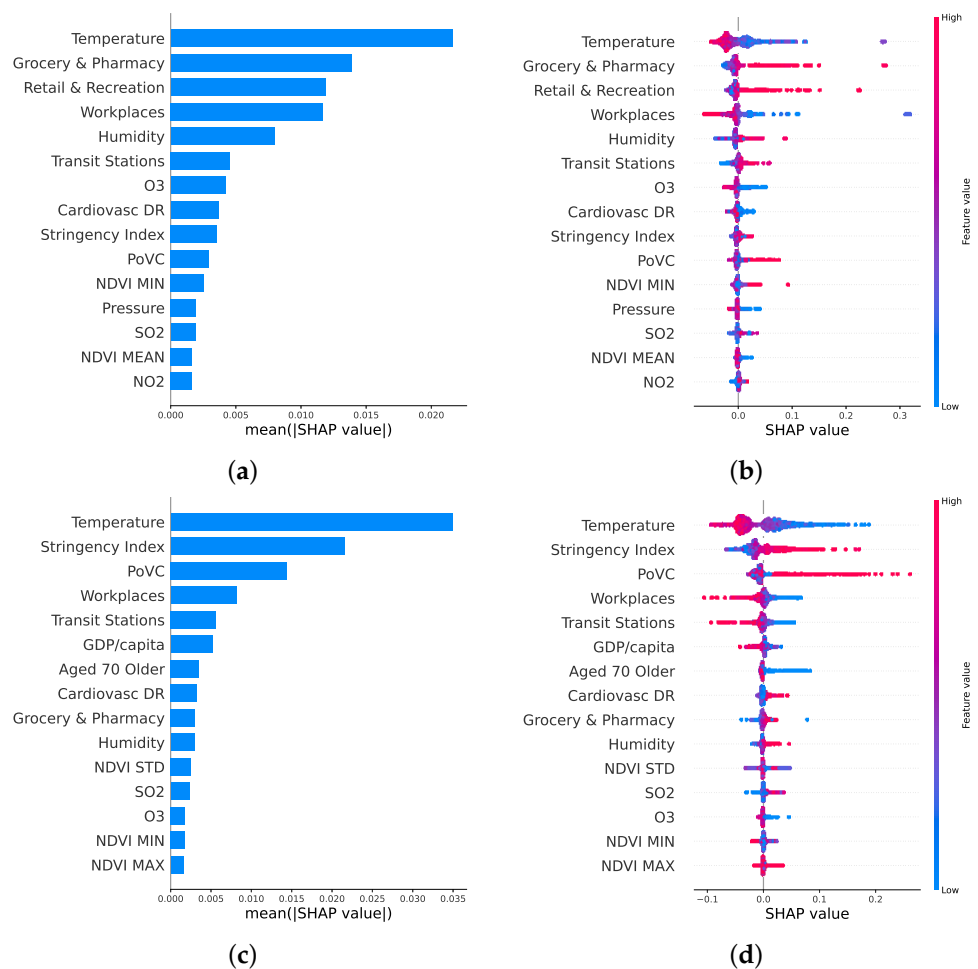

Figure 4. The global aspect of feature understanding using SHAP framework is achieved with the feature importance diagram. In addition to the global explanations, summary plots are adopted as a model interpretation tool for local explanations. Each row corresponds to the model type: the first depicts the cases model while the second depicts deaths. (**a**) SHAP feature importance for cases; (**b**) SHAP summary plot for cases; (**c**) SHAP feature importance for deaths; (**d**) SHAP summary plot for deaths.

5.4. Feature Understanding and Feature Explanation

The summary plot provides the user with a fine observation of each feature relationship with the prediction (local explanations); however, a more detailed monitoring between individual feature values with the impact on the model (Shapley values) can only be provided by the SHAP dependence plots. Figure 5 presents the SHAP dependence plots for the three most important features for cases (first column) and deaths (second column).
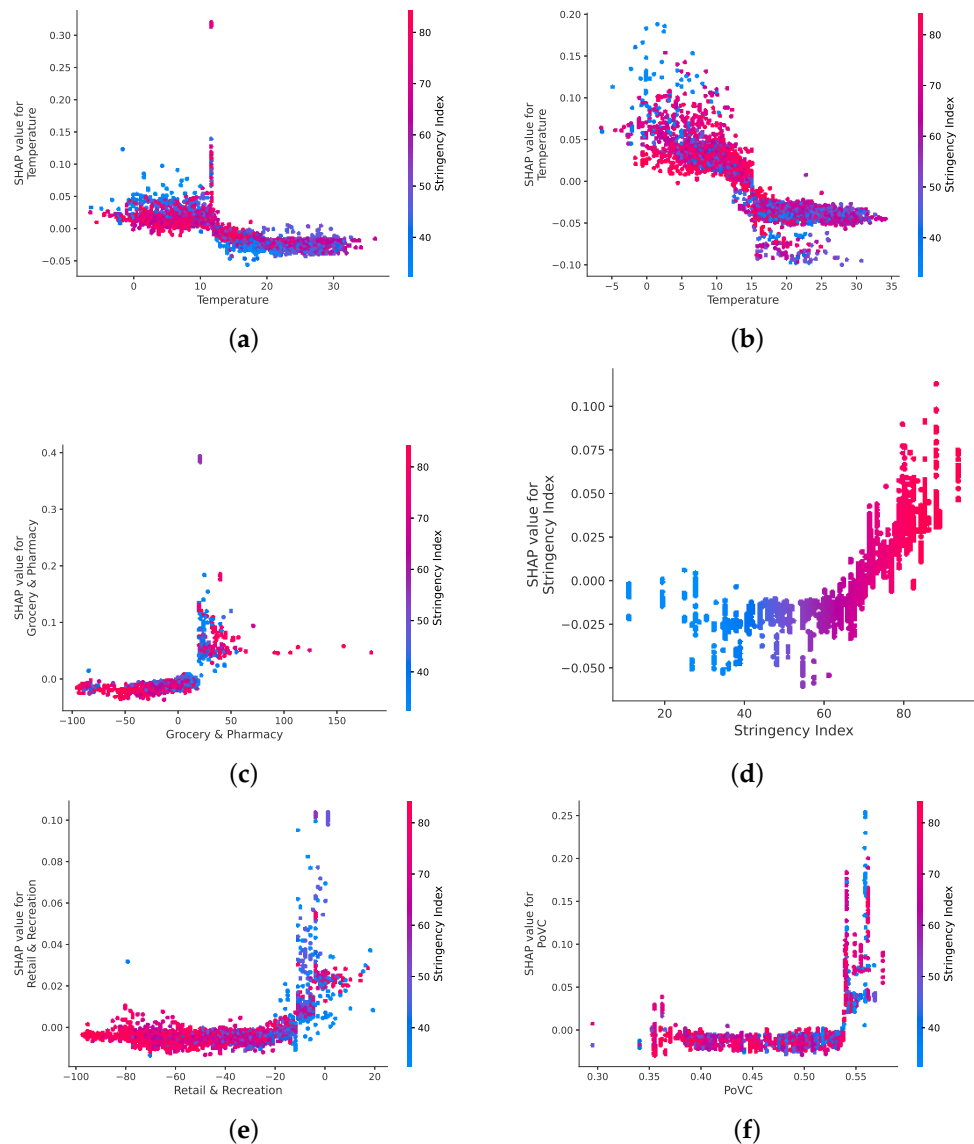

Figure 5. SHAP dependence plots for feature understanding. The first column represents SHAP dependence plots for the 3 most important features for the cases model, while the second for the one predicting deaths. SHAP dependence plots depict the contribution of each features value to the models prediction. All graphs interact with the Government Response Stringency Index. (**a**) Temperature; (**b**) temperature; (**c**) grocery and pharmacy; (**d**) stringency index; (**e**) retail and recreation; (**f**) percentage of vegetation cover.

A SHAP dependence plot depicts, for a unique factor, the feature value on the *x*-axis and the corresponding Shapley value on the *y*-axis, for each data instance. The color in Figure 5 demonstrates the interaction of the dependent feature with the stringency Index (see Section 5.1). This interaction provides additional explanations according to each government's decisions, which are related to mobility trends and climate change

factors. Both summary and dependence plots will be considered in order to make robust explanations for the pandemic factors.

According to Figure 4, temperature has important impact on models' outcome, for cases and deaths models. Figure 4b,d depict that high temperature values reduce the predicted cases and deaths, while lower values increase the models' predictions. Focusing on SHAP dependence plots Figure 5a,b, we observe a negative slope, which means that when the temperature values are increased, the probability of reported cases and deaths is decreased.

Red dots depict a high stringency index in the reported dates of each plot, while blue ones depict lower stringency index measures. According to this, tight health measures along with comparatively higher temperatures seem to have the lowest probability of cases and deaths according to the calculated SHAP values. Furthermore, temperature values that are higher than 15 °C report negative SHAP values while the dominant color is the blue one, meaning that on the one hand governments lowered measurements and on the other hand higher temperature values lower the probability of transmitting the virus, reaching lower daily cases [49–51].

Mobility trends from grocery and pharmacy is the second most important factor after the temperature for the cases model according to Figure 4. Summary plot in Figure 4b shows that higher mobility percentages from baseline have positive SHAP values, while lower percentages seem to have negative ones. According to SHAP dependence plot in Figure 5a, mobility trends from grocery and pharmacy places are mainly negative and range from 0 to −100%, while apart from some local outliers, positive mobility trends range from 0 to 50%. From −100 to 0% mobility trends, SHAP values are negative and dots mainly have red color (high stringency index), which demonstrates samples from lockdown periods. When mobility trends are low and stringency index is high, the probability of an increasing number of cases is significantly lower. Focusing on positive trends, SHAP values have an increasing slope and color of dots gradually becomes blue, which suggests that increasing mobility trends from grocery and pharmacy places increases the probability of new cases [52,53].

Figure 4 denotes that stringency index as the second most important factor according to the deaths model. As it seems from Figure 5d, values of stringency index between 0 and 70 have negative SHAP Values, which means that they negatively affect the probability of increasing deaths. Higher values of stringency index, on the other hand, report positive SHAP values, forming an increasing slope of the graph. The outcome of this graph might be confusing, because it shows that low government responses help the reduction of deaths from COVID-19 while tight measurements do not. It should be mentioned that lower values of stringency index have been taken from governments since the introduction of vaccinations, a response that does not favor the transmission of the virus [54,55].

Besides the second most important feature in the cases model, another type of mobility trend appeared to have significant importance: mobility trends from retail and recreation. This SHAP dependence plot illustrates a straight slope between −100% to −2 % with interaction red dots, which corresponds to tight government responses. According to previous bounds, most of the depicted instances have both negative and low SHAP values. From −20% to 20% SHAP values start to increase, while dots begin to have blue color, describing instances from places with lower stringency index. The function distribution suggests that reducing movements to places such as restaurants, cafes, shopping centers, etc., with tight government responses reduce the probability of reporting high new daily cases [52,53].

For the deaths model, percentage of vegetation cover (PoVC) is the third most important factor. Lower values of PoVC illustrate cities in epochs with low vegetation cover, while high values of PoVC describe cities in epochs with high vegetation cover. Figure 5f shows a similar distribution as of the previous factor (mobility trends from retail and recreation). In more detail, from 35% to 54% vegetation land cover have negative and low SHAP values, while instances with urban greenness above 50% start to have increasing SHAP values. It is important to note that instances are vertically distributed, which means that different cities might have similar percentages in vegetation cover for different epochs

of the years of interest. Moreover, the dominant color is red, while individual blues appear in the graph, providing us information about cities that had bigger PoVCs and still reached high confirmed deaths from COVID-19. According to this, the lack of urban public green areas due to either seasonal vegetation change or urban planning issues in periods where governments' countermeasures were tight seem to reduce the probability of reporting a high number of confirmed deaths from COVID-19.

5.5. City-to-City and Year-to-Year Analysis with LIME

Here, we utilize the LIME explanation technique, to explain the predictions of the regressor, by learning an interpretable model locally around each prediction. The technique learns the weights for each city for a given time period via least square approach. Figures 6 and 7 illustrate the estimated weights of LIME algorithm. The relative weights—risk factors that affect COVID-19 progression—either contribute to the prediction (in red) or are evidence against it (in green). The color saturation in heatmaps shows how much each feature contributes to the model output. Figure 6 focuses on samples selected for 2020 and Figure 7 for 2021 respectively. Subfigures of each one of them illustrate the explanations for (a) cases and (b) deaths.
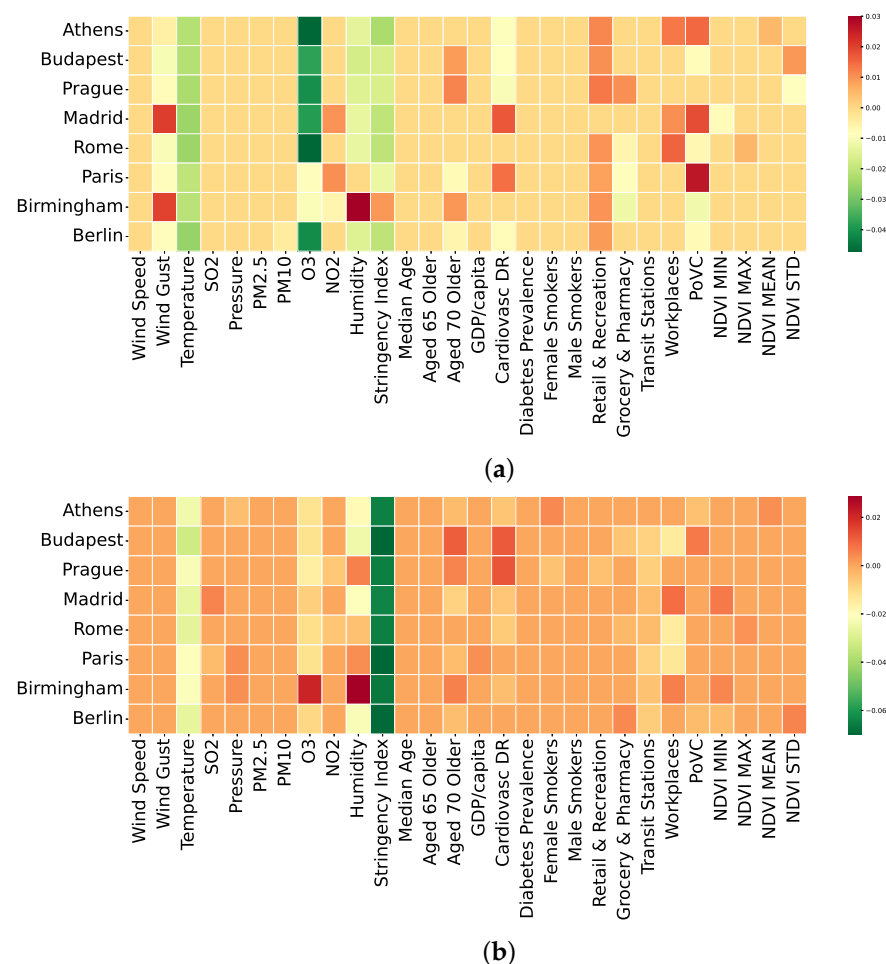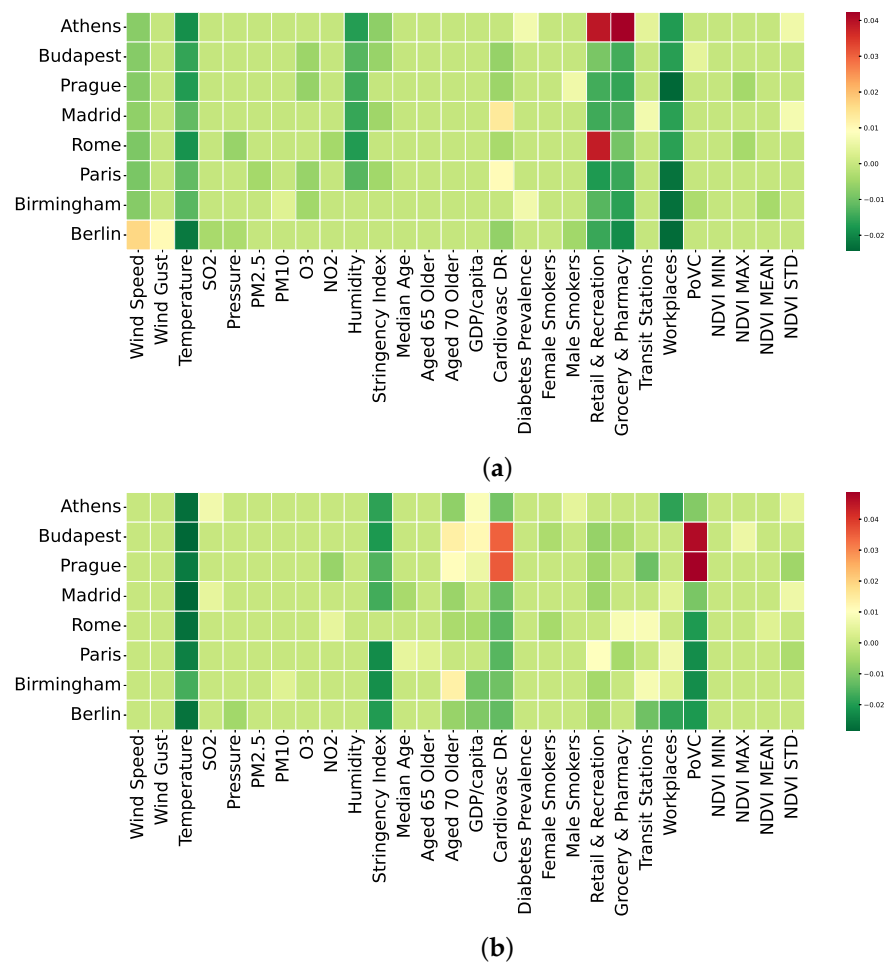

(a)

(b)

Figure 6. Local feature understanding using LIME framework. The contribution of each feature to each city is illustrated with heatmaps for 2020. Each row represents, for each city (*y*-axis), the LIME-calculated weights for each indicator (*x*-axis). Green color suggests negative influence to the model's outcome, while a red one is positive. (**a**) Heatmap of LIME weights for cases model; (**b**) heatmap of LIME weights for deaths model.

(a)

(b)

Figure 7. Local feature understanding using LIME framework. The contribution of each feature to each city is illustrated with heatmaps for 2021. Each row represents, for each city (*y*-axis), the LIME calculated weights for each indicator (*x*-axis). Green color suggests a negative influence on the model's outcome, while a red one is positive. (**a**) Heatmap of LIME weights for cases model; (**b**) heatmap of LIME weights for deaths model.

According to Figure 6a, the most important factor that negatively influence the models predicted cases is the O_3 variable, in all cities except Paris and Birmingham. Despite the increased air quality index in [56], high O_3 levels are linked to reduced COVID-19 cases. Temperature is another variable that seems to have a negative contribution as well for all the cities. Humidity and stringency index influence negatively the predictions for the cases in cities such as Athens, Budapest, Prague, Madrid, Rome and Berlin; however, the colors are light green, which means that the weights are not so important. In Paris, the stringency index maintains a negative affect on the predictions, while on the other side in Birmingham both humidity and stringency index variables have high positive correlations. Positive contributions are also reported from the PoVC factor for the city of Paris along with Madrid and Athens, while in Birmingham and Paris, wind gust factor also affects the model predictions. Note that wind gust negatively affects other cities, but with lesser impact. In contrast to temperature, mobility trends from places of retail and recreation have positive explanations in all cities except Madrid. Positive influences with lower impact are also reported for the factor of NO_2 in Madrid and Paris, people aged 70 or older in Budapest, Prague and Birmingham and mobility trends from workplaces in Prague, Athens and Rome as well.

In Figure 6b, the most important factor is stringency index, which has a negative influence on the model's predicted deaths, for every city of interest. Along with stringency index, temperature is second in terms of negative contribution impact as a factor in all cities,

while O_3 has lower impact than temperature in all cities except Birmingham. According to the exception of the previous city, humidity and O_3 report a significant positive influence on the predicted deaths. In contrast with a high positive contribution, humidity negatively affects cities such as Berlin, Athens, Budapest and Madrid. Strong positive correlations are also shown in two more factors: aged 65 or older and cardiovascular death rate variables, for Budapest and Prague. Other factors contribute mostly positive but moderately compared to the previous ones. Some of them are mobility trends from workplaces and transit stations along with NDVI statistical values, all for specific cities.

Focusing on the cases model in Figure 7a the most important indicators that negatively affect the model's predicted cases are the mobility trends from workplaces, grocery and pharmacy and, lastly, from retail and recreation. Note that mobility trends from retail and recreation have a positive influence on the model's outcome in Athens and Rome, while in Athens, mobility trends from grocery and pharmacy seem to have a positive and strong contribution too. A negative contribution is also reported from the atmospheric indicators and especially from temperature and humidity in all cities, except Birmingham and Berlin.

Figure 7b illustrates, in a heatmap, the explanations for the deaths model for 2021, showing the significant contributions of heterogeneous factors for different cities. The three most important factors that contribute to the model's predicted deaths are temperature, stringency Index, PoVC and cardiovascular death rate. The atmospheric factor temperature, regardless of the city, continuously seems to have a strong negative influence on the model's outcome. Stringency index has the same behaviour for the model's predicted deaths, but with lower impact. Note that Rome is not affected by the previous factor. The environmental index PoVC has negative contribution on the models predictions for all cities except Budapest and Prague, where a strong correlation with high predicted values is detected. The same behaviour is also detected for the health factor called cardiovascular death rate. Lastly, according to Figure 7b, Berlin and Athens seem to be affected by the mobility trends from workplaces variable, due to its negative contribution.

6. Conclusions

In this work, we presented a tree-based interpretable Machine Learning model that adopts the bagging concept, namely Random Forest, for robust COVID-19 cases/deaths predictions. Furthermore, a data fusion approach was used in order to evaluate how multi-dimensional data affect our model's predictive ability. These data are spatio-temporal time series and derive from open-source platforms describing: (1) Earth observation data for detecting urban vegetation; (2) socio-economic factors; (3) health-related data; (4) atmospheric data; and (5) mobility trends. Along with the predictive operation, our model utilizes two explainable frameworks responsible for providing global and local feature understanding. The XAI framework used to identify global correlations between indicators is a modified version of SHAP focusing on explaining tree-based ML models, named treeSHAP, and it is able to detect fast and accurate explanations overcoming mandatory challenges that other feature attribution methods face. From the local aspect of explanations, LIME was utilized to explain the relationship between cities and factors individually, by utilizing instances from them. The four most important factors that are related to the spread of the pandemic for each city are temperature, varieties of mobility trends such as grocery and pharmacy, retail and recreation, the location of existing urban green areas along with stringency index. However, a closer investigation of each city suggests that climate change and air pollution factors such as temperature, humidity and O_3 influence our model's behaviour on both morbidity and mortality. Regarding the model that predicts the reported COVID-19 cases, our explanations suggest that in 2020, strong positive correlations between factors related to vegetation indices, such as PoVC, and COVID-19 spread are appeared in Paris, whereas atmospheric and air-pollution factors, such as humidity and wind gust, have strong influence on COVID-19 in case of Birmingham. On the contrary, O_3 ground-level air pollution factor is dominant on the spread of COVID-19 for most of the cities and decelerates the spread of the virus in the same testing period. In 2021, positive correlations related to mobility

trends are reported for the cities of Athens and Rome. As regards the rest of the cities, a strong negative correlation between the reported COVID-19 cases along with the mobility trends related to workplaces is observed, for both retail and recreation as well as grocery and pharmacy. Explanations for the model that predicts daily deaths suggest that, in 2020, stringency index has strong influence on the results. It is also worth mentioning that in Birmingham risk factors, such as humidity and O_3, positively influence the daily deaths. During the year 2021, explanations show that temperature risk factor is the dominant factor with a negative contribution in all the cities. Moreover, stringency index maintains its negative behaviour for all cities except for Rome. On the other hand, a positive correlation is derived from cardiovascular death rate and most importantly from PoVC in Budapest and Prague. The rest of the cities contribute negatively for the same factors. In our future work, we will investigate the effectiveness of vaccinations in the year 2021 and compare it with the preventive measures that were implemented in 2020. Furthermore, we aim to introduce an even more robust framework that will represent more cities from Europe and provide more generalized explanations for Europe. Along with in the proposed framework, deep learning models will also be included in our comparisons and fastSHAP explainable framework will be utilized for even faster explanations.

Author Contributions: Conceptualization, A.T. and M.K.; methodology, A.T.; software, A.T. and I.N.T.; validation, A.T., M.K., I.N.T., I.R., A.D. and N.D.; formal analysis, A.T., M.K., I.R.; investigation, A.T.; resources, A.T., I.N.T. and M.K.; data curation, A.T.; writing—original draft preparation, A.T., M.K., I.N.T., I.R.; writing—review and editing, A.T., M.K., I.N.T., I.R., A.D. and N.D.; visualization, A.T., I.N.T.; supervision, M.K., I.R., A.D. and N.D.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The dataset was a fusion of heterogeneous spatio-temporal open-source data and are summarized in Table 1.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yang, X.; Yu, Y.; Xu, J.; Shu, H.; Liu, H.; Wu, Y.; Zhang, L.; Yu, Z.; Fang, M.; Yu, T.; et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: A single-centered, retrospective, observational study. *Lancet Respir. Med.* **2020**, *8*, 475–481. [CrossRef]
2. Cucinotta, D.; Vanelli, M. WHO declares COVID-19 a pandemic. *Acta Bio Med. Atenei Parm.* **2020**, *91*, 157.
3. Rosenthal, P.J.; Breman, J.G.; Djimde, A.A.; John, C.C.; Kamya, M.R.; Leke, R.G.; Moeti, M.R.; Nkengasong, J.; Bausch, D.G. COVID-19: Shining the light on Africa. *Am. J. Trop. Med. Hyg.* **2020**, *102*, 1145. [CrossRef] [PubMed]
4. Burke, S.; Parker, S.; Fleming, P.; Barry, S.; Thomas, S. Building health system resilience through policy development in response to COVID-19 in Ireland: From shock to reform. *Lancet Reg. Health Eur.* **2021**, *9*, 100223. [CrossRef]
5. Sanfelici, M. The Italian response to the COVID-19 crisis: Lessons learned and future direction in social development. *Int. J. Community Soc. Dev.* **2020**, *2*, 191–210. [CrossRef]
6. Kavouras, I.; Kaselimi, M.; Protopapadakis, E.; Bakalos, N.; Doulamis, N.; Doulamis, A. COVID-19 Spatio-Temporal Evolution Using Deep Learning at a European Level. *Sensors* **2022**, *22*, 3658. [CrossRef]
7. Lau, H.; Khosrawipour, V.; Kocbach, P.; Mikolajczyk, A.; Schubert, J.; Bania, J.; Khosrawipour, T. The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. *J. Travel Med.* **2020**, *27*, taaa037. [CrossRef]
8. Carlson, C.J.; Albery, G.F.; Merow, C.; Trisos, C.H.; Zipfel, C.M.; Eskew, E.A.; Olival, K.J.; Ross, N.; Bansal, S. Climate change increases cross-species viral transmission risk. *Nature* **2022**. doi: 10.1038/s41586-022-04788-w. [CrossRef]
9. Sharifi, A.; Khavarian-Garmsir, A.R. The COVID-19 pandemic: Impacts on cities and major lessons for urban planning, design, and management. *Sci. Total Environ.* **2020**, *749*, 142391. [CrossRef]
10. Travaglio, M.; Yu, Y.; Popovic, R.; Selley, L.; Leal, N.S.; Martins, L.M. Links between air pollution and COVID-19 in England. *Environ. Pollut.* **2021**, *268*, 115859. [CrossRef]

11. Manzanedo, R.D.; Manning, P. COVID-19: Lessons for the climate change emergency. *Sci. Total Environ.* **2020**, *742*, 140563. [CrossRef] [PubMed]

12. Kaselimi, M.; Voulodimos, A.; Daskalopoulos, I.; Doulamis, N.; Doulamis, A. A Vision Transformer Model for Convolution-Free Multilabel Classification of Satellite Imagery in Deforestation Monitoring. *IEEE Trans. Neural Netw. Learn. Syst.* 2022; *early access*. [CrossRef]

13. Alassafi, M.O.; Jarrah, M.; Alotaibi, R. Time series predicting of COVID-19 based on deep learning. *Neurocomputing* **2022**, *468*, 335–344. [CrossRef] [PubMed]

14. Gautam, Y. Transfer Learning for COVID-19 cases and deaths forecast using LSTM network. *ISA Trans.* **2021**, *124*, 41–56. [CrossRef] [PubMed]

15. Devaraj, J.; Elavarasan, R.M.; Pugazhendhi, R.; Shafiullah, G.; Ganesan, S.; Jeysree, A.K.; Khan, I.A.; Hossain, E. Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant? *Results Phys.* **2021**, *21*, 103817. [CrossRef] [PubMed]

16. Sun, D.; Xu, J.; Wen, H.; Wang, D. Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest. *Eng. Geol.* **2021**, *281*, 105972. [CrossRef]

17. Zhan, C.; Zheng, Y.; Zhang, H.; Wen, Q. Random-forest-bagging broad learning system with applications for covid-19 pandemic. *IEEE Internet Things J.* **2021**, *8*, 15906–15918. [CrossRef]

18. Kavouras, I.; Kaselimi, M.; Protopapadakis, E.; Doulamis, N. Machine Learning Tools to Assess the Impact of COVID-19 Civil Measures in Atmospheric Pollution. In Proceedings of the The 14th PErvasive Technologies Related to Assistive Environments Conference, Corfu, Greece, 29 June–2 July 2021; pp. 396–403.

19. Xie, X.; Wu, T.; Zhu, M.; Jiang, G.; Xu, Y.; Wang, X.; Pu, L. Comparison of random forest and multiple linear regression models for estimation of soil extracellular enzyme activities in agricultural reclaimed coastal saline land. *Ecol. Indic.* **2021**, *120*, 106925. [CrossRef]

20. Grekousis, G.; Feng, Z.; Marakakis, I.; Lu, Y.; Wang, R. Ranking the importance of demographic, socioeconomic, and underlying health factors on US COVID-19 deaths: A geographical random forest approach. *Health Place* **2022**, *74*, 102744. [CrossRef]

21. Shin, D. The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI. *Int. J. Hum.-Comput. Stud.* **2021**, *146*, 102551. [CrossRef]

22. Yang, G.; Ye, Q.; Xia, J. Unbox the black-box for the medical explainable ai via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *Inf. Fusion* **2022**, *77*, 29–52. [CrossRef]

23. Lundberg, S.; Lee, S. A unified approach to interpreting model predictions. In Proceedings of the Advances in Neural Information Processing Systems 30 (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.

24. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *arXiv* **2016**, arXiv:1602.04938.

25. Sarkodie, S.A.; Owusu, P.A. Global effect of city-to-city air pollution, health conditions, climatic & socio-economic factors on COVID-19 pandemic. *Sci. Total Environ.* **2021**, *778*, 146394. [PubMed]

26. Rashed, E.A.; Hirata, A. One-Year Lesson: Machine Learning Prediction of COVID-19 Positive Cases with Meteorological Data and Mobility Estimate in Japan. *Int. J. Environ. Res. Public Health* **2021**, *18*, 5736. [CrossRef] [PubMed]

27. Zoran, M.A.; Savastru, R.S.; Savastru, D.M.; Tautan, M.N. Assessing the relationship between ground levels of ozone (O_3) and nitrogen dioxide (NO_2) with coronavirus (COVID-19) in Milan, Italy. *Sci. Total Environ.* **2020**, *740*, 140005. [CrossRef]

28. Wu, X.; Nethery, R.C.; Sabath, M.B.; Braun, D.; Dominici, F. Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. *Sci. Adv.* **2020**, *6*, eabd4049. [CrossRef]

29. Aurna, N.F.; Yousuf, M.A.; Taher, K.A.; Azad, A.; Moni, M.A. A classification of MRI brain tumor based on two stage feature level ensemble of deep CNN models. *Comput. Biol. Med.* **2022**, *146*, 105539. [CrossRef]

30. Balleyguier, C.; Vanel, D.; Athanasiou, A.; Mathieu, M.; Sigal, R. Breast radiological cases: Training with BIRADS® classification. *Eur. J. Radiol.* **2005**, *54*, 97–106. [CrossRef]

31. Chen, X.; Duan, Q.; Wu, R.; Yang, Z. Segmentation of lung computed tomography images based on SegNet in the diagnosis of lung cancer. *J. Radiat. Res. Appl. Sci.* **2021**, *14*, 396–403. [CrossRef]

32. Soulami, K.B.; Kaabouch, N.; Saidi, M.N.; Tamtaoui, A. Breast cancer: One-stage automated detection, segmentation, and classification of digital mammograms using UNet model based-semantic segmentation. *Biomed. Signal Process. Control* **2021**, *66*, 102481. [CrossRef]

33. Arras, L.; Osman, A.; Samek, W. CLEVR-XAI: A benchmark dataset for the ground truth evaluation of neural network explanations. *Inf. Fusion* **2022**, *81*, 14–40. [CrossRef]

34. Veerappa, M.; Anneken, M.; Burkart, N.; Huber, M.F. Validation of XAI explanations for multivariate time series classification in the maritime domain. *J. Comput. Sci.* **2022**, *58*, 101539. [CrossRef]

35. van der Velden, B.H.; Kuijf, H.J.; Gilhuijs, K.G.; Viergever, M.A. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. *Med. Image Anal.* **2022**, *79*, 102470. [CrossRef] [PubMed]

36. Rostami, M.; Oussalah, M. A novel explainable COVID-19 diagnosis method by integration of feature selection with random forest. *Inform. Med. Unlocked* **2022**, *30*, 100941. [CrossRef] [PubMed]

37. Muhammad, L.J.; Algehyne, E.A.; Usman, S.S.; Ahmad, A.; Chakraborty, C.; Mohammed, I.A. Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology. *SN Comput. Sci.* **2021**, *2*, 11. [CrossRef]

38. Qiao, K.; Xing, S.; Xiaying, X.; Bing, C.; Yuanzhu, C.; Xudong, Y.; Baiyu, Z. Machine Learning-Aided Causal Inference Framework for Environmental Data Analysis: A COVID-19 Case Study. *Environ. Sci. Technol.* **2021**, *55*, 13400–13410.
39. Yeşilkanat, C.M. Spatio-temporal estimation of the daily cases of COVID-19 in worldwide using random forest machine learning algorithm. *Chaos Solitons Fract.* **2020**, *140*, 110210. [CrossRef]
40. Prakash, K.B.; Imambi, S.S.; Ismail, M.; Kumar, T.P.; Pawan, Y. Analysis, prediction and evaluation of covid-19 datasets using machine learning algorithms. *Int. J.* **2020**, *8*, 2199–2204. [CrossRef]
41. 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.* **2021**, *4*, 116–123. [CrossRef]
42. Sagi, O.; Rokach, L. Ensemble learning: A survey. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2018**, *8*, e1249. [CrossRef]
43. Dong, X.; Yu, Z.; Cao, W.; Shi, Y.; Ma, Q. A survey on ensemble learning. *Front. Comput. Sci.* **2020**, *14*, 241–258. [CrossRef]
44. Lundberg, S.; Erion, G.; Lee, S. Consistent Individualized Feature Attribution for Tree Ensembles. *arXiv* **2018**, arXiv:1802.03888.
45. Ritchie, H.; Mathieu, E.; Rodés-Guirao, L.; Appel, C.; Giattino, C.; Ortiz-Ospina, E.; Hasell, J.; Macdonald, B.; Dattani, S.; Roser, M. Coronavirus Pandemic (COVID-19). Our World In Data. 2020. Available online: https://ourworldindata.org/coronavirus (accessed on 25 May 2022).
46. Bernal, J.L.; Andrews, N.; Gower, C.; Gallagher, E.; Simmons, R.; Thelwall, S.; Stowe, J.; Tessier, E.; Groves, N.; Dabrera, G.; et al. Effectiveness of Covid-19 vaccines against the B. 1.617. 2 (Delta) variant. *N. Engl. J. Med.* **2021**, *385*, 585–594. [CrossRef] [PubMed]
47. Mathieu, E.; Ritchie, H.; Ortiz-Ospina, E.; Roser, M.; Hasell, J.; Appel, C.; Giattino, C.; Rodés-Guirao, L. A global database of COVID-19 vaccinations. *Nat. Hum. Behav.* **2021**, *5*, 947–953. [CrossRef] [PubMed]
48. Andrews, N.; Stowe, J.; Kirsebom, F.; Toffa, S.; Rickeard, T.; Gallagher, E.; Gower, C.; Kall, M.; Groves, N.; O'Connell, A.M.; et al. Covid-19 vaccine effectiveness against the Omicron (B. 1.1. 529) variant. *N. Engl. J. Med.* **2022**, *386*, 1532–1546. [CrossRef]
49. Shi, P.; Dong, Y.; Yan, H.; Zhao, C.; Li, X.; Liu, W.; He, M.; Tang, S.; Xi, S. Impact of temperature on the dynamics of the COVID-19 outbreak in China. *Sci. Total Environ.* **2020**, *728*, 138890. [CrossRef]
50. Xie, J.; Zhu, Y. Association between ambient temperature and COVID-19 infection in 122 cities from China. *Sci. Total Environ.* **2020**, *724*, 138201. [CrossRef]
51. Notari, A. Temperature dependence of COVID-19 transmission. *Sci. Total Environ.* **2021**, *763*, 144390. [CrossRef]
52. Velias, A.; Georganas, S.; Vandoros, S. COVID-19: Early evening curfews and mobility. *Soc. Sci. Med.* **2022**, *292*, 114538. [CrossRef]
53. Panarello, D.; Tassinari, G. One year of COVID-19 in Italy: Are containment policies enough to shape the pandemic pattern? *Socio-Econ. Plan. Sci.* **2022**, *79*, 101120. [CrossRef]
54. Chisadza, C.; Clance, M.; Gupta, R. Government Effectiveness and the COVID-19 Pandemic. *Sustainability* **2021**, *13*, 3042. [CrossRef]
55. Deb, P.; Furceri, D.; Ostry, J.D.; Tawk, N. The economic effects of Covid-19 containment measures. *Open Econ. Rev.* **2022**, *33*, 1–32. [CrossRef]
56. Rathod, A.; Sahu, S.; Singh, S.; Beig, G. Anomalous behaviour of ozone under COVID-19 and explicit diagnosis of O_3-NO_x-VOCs mechanism. *Heliyon* **2021**, *7*, e06142. [CrossRef] [PubMed]