Forecasting the impact of environmental stresses on the frequent waves of COVID19

Zhenhua Yu · Abdel-Salam G. Abdel-Salam ·
Ayesha Sohail · Fatima Alam

Received: 17 April 2021 / Accepted: 23 July 2021 / Published online: 5 August 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract A novel approach to link the environmental stresses with the COVID-19 cases is adopted during this research. The time-dependent data are extracted from the online repositories that are freely available for knowledge and research. Since the time series data analysis is desired for the COVID-19 time-dependent frequent waves, here in this manuscript, we have developed a time series model with the aid of “non-linear autoregressive network with exogenous inputs (NARX)” approach. The distribution of infectious agent-containing droplets from an infected person to an uninfected person is a common form of respiratory disease transmission. SARS-CoV-2 has mainly spread via short-range respiratory droplet transmission. Airborne transmission of SARS-CoV-2 seems to have occurred over long distances or times in unusual conditions; SARS-CoV-2 RNA was found in PM10 collected in Italy. This research shows that SARS-CoV-2 particles adsorbed to outdoor PM remained viable for a long time, given the epidemiology of COVID-19, outdoor

air pollution is unlikely to be a significant route of transmission. In this research, ANN time series is used to analyze the data resulting from the COVID-19 first and second waves and the forecasted results show that air pollution affects people in different areas of Italy and make more people sick with covid-19. The model is developed based on the disease transmission data of Italy.

Keywords SARS-CoV2 · Machine learning · Epidemic forecasting · Artificial neural network · Environmental stress

1 Introduction

1.1 COVID19 epidemics and environmental stresses

The COVID-19 epidemic is a global health emergency. COVID-19 infection has spread to almost every country on the globe. However, some areas have been hit harder than others in terms of infection and mortality rates. Maybe the differences have posed due to the impact of air pollution on the global spread of COVID-19 infections and mortality rates [11].

Air pollution is a significant environmental cause of several diseases and premature deaths all over the world. According to research, both short- and long-term exposure to air pollution are linked to a wide range of negative health effects. Air pollution has been related to a rise in the incidence of respiratory disorders as well

Z. Yu
Institute of Systems Security and Control, College of
Computer Science and Technology, Xi'an University of
Science and Technology, Xi'an 710054, China

A. G. Abdel-Salam (✉)
Department of Statistics, Mathematics and Physics, College
of Arts and Sciences, Qatar University, Doha, Qatar
e-mail: Abdo@qu.edu.qa

A. Sohail · F. Alam
Department of Mathematics, Comsats University Islamabad,
Lahore 54000, Pakistan

as deaths. In polluted air, PM_{2.5}, or small particulate matter with a diameter of 2.5 micrometers, is one of the main health risk factors in the climate, responsible for millions of deaths per year around the world. PM_{2.5} and another particle, PM₁₀, have been related to an increased rate of respiratory disorders, as well as hospitalization for chronic lung disease and pneumonia. Fine particles produced by livestock, sulfur oxides produced by burning sulfurous fossil fuels, and nitrogen oxides produced by road traffic are examples. When found in higher concentrations in the atmosphere, nitrogen dioxide (NO₂) is another significant air pollutant that is harmful to human respiratory systems. Both anthropogenic and natural processes release it into the atmosphere. Bronchitis, exacerbated asthma, respiratory allergies, heart disease, and stroke are just some of the diseases attributed to air pollution deaths [19]. Ozone produced by nitrogen oxides and hydrocarbons under the influence of sunlight and high air temperatures. Most pollutant concentrations have decreased in recent decades, except ozone, which has increased hemispheric background concentrations as a result of worldwide rises in precursor gases and global warming.

A recent study looked at the geographical characteristics of the COVID-19 infection and linked it to various annual satellite and ground-level air quality indexes in eight countries, including Italy, Spain, Germany, France, the United Kingdom, the USA, Iran, and China, and discovered more viral infections in areas with elevated PM_{2.5} and NO₂ levels. Italy had the strongest associations in terms of infection and mortality among these countries, although population size and density did not affect COVID-19 incidence. In China, population density had a similar positive correlation with infection and mortality as air pollution, while population density had a stronger correlation with infection and mortality than air pollution [30]. Study showed that COVID-19 fatalities in Italy which have all been hit hard by the virus. It was discovered that 78 percent of deaths occurred in only five regions of northern Italy and central Spain, where NO₂ levels were highest and air pressure was low, preventing contaminants from dispersing. Long-term exposure to NO₂ has been shown to lead to mortality caused by COVID-19 infection in these areas, according to the author COVID-19 incidence, and mortality rates were significantly associated with higher diesel particulate matter (DPM) levels [9, 28]. Air pollution is one of the factors that has gotten a lot of attention in recent months as a

factor that could help the disease(COVID-19) spread, become more severe, and have a worse prognosis. A recent study found a connection between COVID-19 mortality and the presence of high levels of pollutants in northern Italy. Long-term air-quality data in 71 Italian provinces showed a significant correlation with COVID-19 cases, indicating that chronic exposure to air pollution can influence viral spread [9]. This study aims to address the possible effects of air pollution on COVID-19 in the sense of what is known about air pollution's health effects in general, as well as state-of-the-art methods for researching such effects. Many sources lead to air pollution in Europe. Energy production from fossil fuel or biomass combustion, road traffic, shipping pollution, home and utility building heating, industrial production, agricultural emissions, and other factors all contribute to these emissions. These sources not only emit pollutants directly (primary emissions), but they also emit precursor gases, which produce secondary pollutants via atmospheric reactions [21]. Finally, air pollution, especially NO₂ and PM_{2.5}, could increase infection susceptibility and mortality from COVID-19. COVID-19 transmission can be affected by air pollution, according to the available data. Furthermore, air pollution may harm the prognosis of SARS-CoV-2 infection patients. The available research findings on this subject can aid epidemiologists in determining the best course of action to prevent future outbreaks [14]. Poor populations should also be considered because they are more likely to be subjected to indoor air pollution, which increases the risk of becoming seriously ill from COVID-19 infections. Air quality should be counted as an important part of a holistic approach to public health safety and disease outbreak prevention. More studies should be done to confirm the negative impact of additional confounders including age and pre-existing medical conditions, as well as extended exposure to NO₂, PM_{2.5}, and other air contaminants, on COVID-19 fatalities. Asthma, COPD, lung cancer, heart disease, and diabetes are all chronic diseases caused by air pollution. COVID-19 hospitalization, ICU admission, and death are typical outcomes of many of these conditions. For this reason, the negative effects of air pollution on the COVID-19 pandemic are of grave concern [17]. Other than the Severe Acute Respiratory Syndrome Corona Virus, air pollution has been shown to minimize respiratory tolerance to bacterial and viral infections (SARS-CoV-2) [7].

However, evidence is accumulating that people who live in high-pollution areas are more likely to be contaminated with SARS-CoV-2 and grow COVID-19 once outbreaks occur. Almost every study focused on composite data, such as COVID-19, air pollution, and other risk factors summed across municipalities as well as districts. Outbreaks and air pollution are related to population density and other spatial factors. The distribution of infectious agent-containing droplets from an infected person to an uninfected person is a common form of respiratory disease transmission. SARS-CoV-2 has mainly spread via short-range respiratory droplet transmission. Airborne transmission of SARS-CoV-2 seems to have occurred over long distances or times in unusual conditions, often indoors. Air is the medium by which the agents spread, implying that particulate matter may serve as an infectious agent carrier. SARS-CoV-2 RNA was found in PM10 collected in Bergamo, Italy, according to one report. Although it is unknown whether SARS-CoV-2 particles adsorbed to outdoor PM remain viable for a long time, given the epidemiology of COVID-19, outdoor air pollution is unlikely to be a significant route of transmission. In the early stages of the outbreak in the spring of 2020, an Italian study was unable to demonstrate the existence of SARS-CoV-2 on airborne particulate matter collected in two areas that were heavily affected and non-affected, respectively [27].

Air pollution is a significant environmental cause of several diseases and premature deaths all over the world. According to research, both short- and long-term exposure to air pollution are linked to a wide range of negative health effects. Air pollution has been related to a rise in the incidence of respiratory disorders as well as deaths. PM2.5, or small particulate matter with a diameter of 2.5 micrometers, is one of the main health risk factors in the climate, responsible for millions of deaths per year around the world. PM2.5 and another particle, PM10, have been related to an increased rate of respiratory disorders (i.e., pneumonia). When found in higher concentrations in the atmosphere, nitrogen dioxide (NO₂) is another significant air pollutant that is harmful to human respiratory systems [18]. A recent study looked at the geographical characteristics of the COVID-19 infection and linked it to various annual satellite and ground-level air quality indexes in eight countries, including Italy, Spain, Germany, France, the United Kingdom, the USA, Iran, and China, and dis-

covered more viral infections in areas with elevated PM2.5 and NO₂ levels [20].

Italy had the strongest associations in terms of infection and mortality among these countries, although population size and density did not affect COVID-19 incidence. In China, population density had a similar positive correlation with infection and mortality as air pollution, while population density had a stronger correlation with infection and mortality than air pollution [10].

In the field of computational biology, mathematical models have proved to be useful in understanding complex biological dynamics more swiftly. To explain the nature of infectious diseases, epidemiological mathematical models have been used. In 1927, Kermack and McKendrick established the first model to research the spread of infectious diseases. This model is essentially based on a system of normal differential equations and has been thoroughly investigated with many amendments. It is called SIR model: S represents respectability, I represents infected people, and R shows recovered people [4].

ANN time series It is useful to use mathematical models to forecast disease outbreak risk and severity and provide key data to assess the type and intensity of disease action. For the latest outbreak of COVID19. Estimated COVID19 inactivity with environmental stress. In this paper, we use ANN time series to understand the behavior of the forecasting and also show how Petri net works like ANN (artificial neural networks). Petri nets are also used for biological modeling [12]. The complex system's implementation and architecture are challenging activities. Usually, proper models are used to solve the simulation and implementation of such systems. Petri nets (PN) are effective in a wide range of fields. PN is also a networking technique as ANN. Same as Petri net modeling, there are arc and nodes in ANN [15].

1.2 Data cleaning

In this research, we used a clean data. Data cleaning is a process in which eliminating or changing data that is erroneous, incomplete, irrelevant, redundant, or incorrectly formatted in order to prepare it for analysis. When it comes to data analysis, these data are usually not necessary or beneficial because it can slow down the process or produce false results.

1.3 Future of epidemiology with genetic algorithm

In the literature, different studies are available, where genetic algorithms are used to optimize the epidemiological problems and the corresponding parametric values [3,6]. For example, Abo-Hammour et al. [2] used the optimization tools to solve the boundary value problem. Similarly, Momani et al. [16] used genetic algorithm to solve the inverse kinematic problem. The genetic algorithms help to optimize the complex problems, with the aid of the genetic mutation principle. With the passage of time, their impact in the field of artificial intelligence, especially for the optimization of hyper-parameters, is getting acknowledged [1,5]. Thus, the genetic algorithm is one of the tools, which can be used to address the complex epidemiological problems.

1.4 Machine learning and epidemiology

Machine learning is a branch of artificial intelligence that is being applied in a variety of sectors in environmental sciences. Machine learning methods are now used in satellite data processing, general circulation models (GCM), weather and climate prediction, air quality forecasting, environmental data analysis and modelling, oceanographic and hydrological forecasting, ecological modelling, and snow, ice, and forest monitoring due to their powerful nonlinear modeling capability.

In the recent literature, the importance of artificial intelligence in the field of computational biology is addressed. The tools, such as classification, networks and imaging tools, have been used to optimize the complex datasets [13,28,29].

1.4.1 ANN time series

ANNs modeling has been used by a growing number of scientists around the world in a variety of scientific fields over the last two decades. ANNs are used in a variety of fields, including pattern classification, control, time series, estimation, prediction, and prognosis, optimization, environmental applications, engineering applications, and financial and cooperative applications.

- medical diagnosis,
- management and marketing applications,
- energy cost prediction.

The basic meaning of the parameters and interconnections within networks trained using real epidemic time series data is difficult to interpret. However, there are two benefits of using neural networks to forecast time-series data, they can completely extract the complex nonlinear relationships embedded in the data, and they do not need any assumptions [22].

Hybrid ANN improves the learning and generalization of ANN in disease forecasting by using data transformation strategies based on improved algorithms, integrating forecast models, and using technology platforms [26].

During this research, we will develop a mapping between the environmental stresses and the spread of this contiguous virus, i.e., SARS-2. The Artificial Neural Network (ANN)-based time series model is developed to explore the impact of environmental stresses on the COVID-19 epidemic. In the next section, we will discuss the method in detail. The results and the useful conclusions are discussed in detail afterward.

2 Materials and methods

Time series learning is an ever-growing field of research and a sub-field of data science. The time-dependent data provide useful information for the forecasting measures and to develop necessary precautionary measures, for the epidemic disease(s) control. For this purpose, several time series models have been proposed in the recent literature. Among such models, the neural network models, which are inspired from the biological neurons, have attained marvelous success. The reason is that the models are designed by taking into account the important features of the data set, as well as by following the specific rules of connecting the layers of the network for better forecasting measures. The most important part of neural networks is the mapping function.

The mapping function can be modeled in several ways. However, due to its simplicity, we introduced the nonlinear regression using an ANN-based approximation of the mapping function. Any Neural Network-based function approximation methodology is appropriate. In this article, we will use the Multilayer Perceptron (MLP) as an example. Accurate predictions aid in the decision-making process for preventive healthcare interventions and disease control; however, this aim can only be accomplished by using the right strategies

and methodologies. Forecast precision is dependent on methods and model selection procedures almost as much as it is dependent on forecast precision. This study aimed to provide an overview of how to choose the best artificial neural network (ANN) methodology for epidemic forecasting [25]. Forecasters must use the appropriate instruments to make high-precision disease predictions. Methodologies: It entailed sampling and surveying ANN-based outbreak predictions.

2.1 Materials

During this research, we have developed a model to link the environmental stresses with the COVID-19 cases. We have used the forecasting tool for the air pollution data extracted from the “World’s Air Pollution: Real-time Air Quality Index” and for covered-19 we extract some data from “<https://ourworldindata.org/covid-vaccinations?country=~ITA>”. There are many countries with corona cases and air pollution, but data of these countries are not available. I choose Italy because there is huge data available, we can see in our data to environmental stress cases report death rate due to impact of air pollution. The results show the number of deaths increasing with environmental stress. Increments in the pollution in the air cause the death rate with corona and respiratory or lung disease.

2.2 Methods

A network with input and output terminals is generated by a collection of processing units (nodes) interconnected with arcs indicating relationships between these units. Arcs are the relative weighted effect of each factor on the solution phase, while nodes are basic computational elements that use the weights of incoming signals. A real-valued bias is correlated with each node. ANN is the owner of a The components of ANN have been defined and summarized as follows:

1. A directed graph, also known as the network topology.
2. Each node has its state variable. two. Each link has a real-valued weight associated with it. Each node has a bias that is measured in real numbers. A transfer function for each node decides a node’s state as a function of its bias (or threshold), weight, and other variables.

2.3 Nonlinear time series

Time series analysis is the best way to forecast future events based on a series of past events. There are mainly five types of time series approaches (algorithms):

- Autoregressive (AR)
- Moving Average (MA)
- Autoregressive Moving Average (ARMA)
- Autoregressive Integrated Moving Average (ARIMA)
- Exponential Smoothing (ES)

The process of predicting is a dynamic process of filtering since it helps to predict the future time-dependent values with the aid of the past values. In the field of artificial intelligence, for training the networks and for the prediction, the “tapped delay lines” are used. The tapped delay lines take the information from the signal output, scales it, and then sum it with the other “taps” to produce an output.

The neural network time series builds a model for given system identification of a physical problem. Once a model is built, it can work to analyze, simulate, and monitor the given problem.

Three commonly used time series tools are:

- networks that can predict the targeted output series from the previous values of inputs and output, i.e. (NARX as discussed in the next section).
- networks that can predict the targeted output series from only using previous values of inputs.
- networks that can predict the targeted output series just from the previous output values.

2.4 The “nonlinear-autoregressive network with exogenous inputs” (NARX)

In the literature (Billings [8], Sohail and Arif [24] and the references therein), it is reported that the dynamic networks are usually (a) input layer dependent or (b) feedforward networks. The “nonlinear-autoregressive-network” with “exogenous-inputs” (NARX) is a recurrent dynamic network, enclosing several layers of the network. It has feedback connections that enclose the layers of the network.

In this study, we have taken input as the source of environmental stresses and have considered these as five input variables, as shown in schematic 1 and the

Table 1 Output variables with week-wise statistics

Variable	week number with maximum error
Cases reported	41, 49, 52
Death reported	50, 55
Cases reported in hospital	5, 27, 33, 42, 46, 47

video provided in the supplementary material, enclosed by the feedback connections.

For the successful implementation of the time series algorithm, we need to define the model equation that govern the dynamics of the key variables. In the literature, Billings [8] for the first time explained in detail the basic modeling strategy and the basic equation that the solver uses while evaluating the time series data. Inspired from his work and the references therein, we have first grouped the variables and then have developed the model as described in Fig. 1 and Eq. 1. From Fig. 1, it is obvious that the environmental stresses are kept as the inputs, whereas the COVID19 cases were kept as the outputs. The delay was used as 1:3 ratio (i.e., $d = 3$ as shown in the figure). The two activation functions were used by the solver which are shown in the schematic as A1 (tansig) and A2 (purlin). $\mathbf{W}$ here shows the weights, and b shows the bias.

The equation of the model is given as:

$$\begin{aligned}\eta(t) = & f(\eta(t-1), \eta(t-2), \\ & \dots \eta(t-n_\eta), \psi(t-1), \\ & \psi(t-2), \dots \psi(t-n_\psi))\end{aligned}\quad (1)$$

every upcoming value of output signal $\eta(t)$, is based on the (i) past value of this output signal and (ii) past value of the exogenous input signal $\psi(t)$. NARX model is used to approximate the function f .

2.5 Application of time series algorithm on COVID19 data

During this research, the data received were filtered, normalized (pre-processing) and then categorized according to the type of the variables.

2.6 Input data

The data were split into two categories, based on the behavior of each factor. The factors that leave an impact

on the disease spread were considered as the input variables. These inputs were actually the environmental stressors and included the following key factors:

1. PM25
2. ozone
3. O3
4. No2
5. humidity

The details of each of these factors and their impact on other epidemiological diseases are already provided in Sect. 1.1.

2.7 Output data

The output data were based on the population's response to the environmental stress and the virus attack nationwide. The key outcomes of the COVID-19 outbreak were given as:

- Number of positive cases reported week-wise.
- Impact on the mortality rates.
- Impact on the birth rates.

2.8 Application of AI-time series

Next, the time series forecasting was conducted according to the step by step procedure as shown in Fig. 1. The data were first grouped into week-wise data from the daily input data. Next, the mapping was made with the aid of the input and output blocked arrays. The correlations were also calculated for statistical accuracy. The time series artificial neural networks forecasted the output data using MatlabTM toolbox. This entire procedure is presented in Fig. 2, where the inputs are taken as the environmental factors and the outputs were the COVID-19 effected patients. Each of the output attributes (such as death rates, patients with heart disease) was treated as a single $(1 \times n)$ vector/ array, where n was the number of rows of the data set. Here, we kept $n = (54 \times 7)$, i.e., 54 weeks and 7 days in each week.

Data were feeded to the "Levenberg" algorithm, and results were obtained; if the error was too high for entire data, then the process was repeated.

Fig. 1 Step by step input-output guide of NARX for COVID-19 data. Top figure presents the open loop, whereas bottom figure presents the closed loop

Fig. 2 Matlab neural network layers for the COVID19 data set

3 Results and discussion

Artificial neural networks (ANNs) modeling techniques are used to estimate the maximum and minimum value of the variable in this way. Artificial neural networks (ANNs) are a form of artificial intelligence created in the 1950s to mimic the structure of the human brain. They are a mathematical function-based approach to describing the workings of the human nervous system.

3.1 Impact of environmental stresses on the number of cases reported

Five environmental inputs were used during this research. During this research, NARX algorithm was used for the time series forecasting. During the first numeri-

cal experiment, the number of upcoming “COVID-19 infected cases (T)” was taken as the target. The environmental stresses (I_1) and the past recorded “COVID-19 infected cases (I_2)” were taken as the inputs. In simple words, the $I_1(t_i)$ and $I_2(t_i)$ were used as the inputs at time t_i to forecast $T(t_{i+1})$ at time t_{i+1} .

We can see from Fig. 3 that the model successfully forecasts the cases, when the Bayesian regularization algorithm was used. There are three peaks of error at week 41, 49, and 53, respectively.

Next, the number of cases that was tested positive for COVID-19 was used as the subject and the entire procedure, discussed above, was repeated. The forecasting results obtained did not show high accuracy (see Fig. 4 top panel). The levenberg–marquardt time series model performed well, only during weeks 25 till 35. The reason is that the number of COVID-19 cases reported during the first and second wave were not linearly dis-

Fig. 3 Error analysis and modeling outcomes for the cases reported (with symptoms)

tributed due to the alarming outbreak in Italy. For such nonlinear data, we have used an advanced algorithm, that is probability dependent and is known as Bayesian

regularization algorithm (see Sohail [23] for the advantages of Bayesian inference). The results obtained were in good accuracy as shown in Fig. 4, bottom panel.

Fig. 4 Error analysis and modeling outcomes for the positive cases reported (after testing)

3.2 Impact of environmental stresses on COVID-19-deaths

Next, a relation is drawn between the COVID-19 deaths and the environmental stresses. With the passage of

time over the period of one year, we can see a reasonable increase in the number of deaths reported, both the statistical findings “.” and by the time series forecasting “+”(see Fig. 5). The error was comparatively higher toward the end of the year. There is fluctuation as well,

Fig. 5 Error analysis and modeling outcomes for the deaths reported

Fig. 6 Error analysis and modeling outcomes for the reproduction rates (top, by levenberg marquardt, bottom, by Bayesian regularization

in the number of deaths recorded by COVID-19 during the second wave. The time series model failed after week 50, to forecast the death count when the perturbation was higher and when the environmental stresses

and the previous death record were taken as the input (see Fig. 1) to obtain the death count for the posterior times.

Fig. 7 Error analysis and modeling outcomes for the other diseases in the wake of COVID-19

3.3 Impact of environmental stresses on miscellaneous factors linked with COVID-19

We have not only considered the deaths, but also considered the birth rates for this analysis. The model fore-

casts the reproduction rate to a better accuracy. Next, the reported cardiovascular deaths and diabetes prevalence statistics are discussed (see Fig. 7) and rise in ICU-cases resulting from the two waves of COVID-19 was discussed (see Fig. 8). Useful results to make future

Fig. 8 Error analysis and modeling outcomes for the ICU cases reported

prediction were drawn from this study, where different trends of time series data were analyzed.

4 Conclusions and future work

Coronavirus control is really necessary to save humanity from this deadly virus. Numerous factors are involved in its spread. In this manuscript, an important factor from an epidemiological perspective, i.e., the environmental stresses is modeled in a novel way.

The manuscript presents a time series analysis, based on real data, to explore the dynamics of novel SARS-CoV2 onset. The model helps to analyze the period required by the virus. This study suggests that there is a negative relationship between virus infection and population density, possibly due to the massive migration of people from big cities to the countryside, bringing the virus with them.

This research shows that COVID-19 fatalities in Italy have all been hit hard by the virus. It is reported that 78 percent of deaths occurred in only five regions of northern Italy and central Spain, where NO₂ levels were highest and air pressure was low, preventing contaminants from dispersing. Long-term exposure to NO₂ has been shown to lead to mortality caused by COVID-19 infection in these areas; according to the

author, COVID-19 incidence and mortality rates were significantly associated with higher diesel particulate matter (DPM) levels.

In Italy, a recent study found a connection between COVID-19 mortality and the presence of high levels of pollutants in northern Italy. Long-term air-quality data in 71 Italian provinces showed a significant correlation with COVID-19 cases, indicating that chronic exposure to air pollution can influence viral spread.

Finally, air pollution, especially NO₂ and PM_{2.5}, could increase infection susceptibility and mortality from COVID-19. COVID-19 transmission can be affected by air pollution, according to the available data. Furthermore, air pollution may harm the prognosis of SARS-CoV-2 infection patients. The available research findings on this subject can aid epidemiologists in determining the best course of action to prevent future outbreaks.

Poor civilians should also be considered because they are more likely to be subjected to indoor air pollution, which increases the risk of becoming seriously ill from COVID-19 infections. Air quality should be counted as an important part of a holistic approach to public health safety and disease outbreak prevention. More studies should be done to confirm the negative impact of additional confounders including age and pre-existing medical conditions, as well as extended

exposure to NO₂, PM_{2.5}, and other air contaminants, on COVID-19 fatalities.

Air pollution is one of the factors that has received a lot of attention in recent months as a factor that could help the disease spread, become more severe, and have a worse prognosis.

During this research, it is verified with the aid of real data and forecasting tools that there is a strong correlation between the COVID-19 spread and environmental stresses linked with it.

Following important factors (and their time dependence) are highlighted during this research:

- Reproduction rate of the population, influenced by the virus.
- Mortality rate.
- Confidence and freedom of the population was highly influenced due to the COVID-19 number of positive cases reported. This led to socio-economic issues.
- Peak times when COVID-19 cases were reported.

Acknowledgements The authors would like to acknowledge the source of data ("<https://ourworldindata.org/covid-vaccinations?country=~ITA>") and their cooperation.

Declarations

Conflict of interest The authors declare that there is no conflict of interest.

References

1. Abdel-Salam, A.-S.G., Sohail, A., Sherin, L., Azim, Q.U.A., Faisal, A., Fahmy, M.A., Li, Z.: Optimization of tank engine crank shaft material properties. In: *Mechanics Based Design of Structures and Machines*, pp. 1–17 (2021)
2. Abo-Hammour, Z., Arqub, O.A., Alsmadi, O., Momani, S., Alsaedi, A.: An optimization algorithm for solving systems of singular boundary value problems. *Appl. Math. Inf. Sci.* **8**(6), 2809 (2014)
3. Abu Arqub, O., Abo-Hammour, Z., Momani, S., Shawagfeh, N.: Solving singular two-point boundary value problems using continuous genetic algorithm. In: *Abstract and Applied Analysis*, vol. 2012. Hindawi (2012)
4. Acemoglu, D., Chernozhukov, V., Werning, I., Whinston, M.D.: A multi-risk sir model with optimally targeted lockdown. NBER working paper, (w27102) (2020)
5. Al-Utaibi, K.A., Sohail, A., Zafar, A., Talha, R., Sait, S.M.: Ai optimization of the exothermic reaction of ethylene oxide with water. In: *Biomedical Engineering: Applications, Basis and Communications*, p. 2150033 (2021)
6. Alsmadi, O., Momani, S., Arqub, O.A., et al.: A genetic algorithm approach for prediction of linear dynamical systems. *Math. Probl. Eng.* **2013** (2013)
7. Barabash, O., Lukova-Chuiko, N., Sobchuk, V., Musienko, A.: Application of petri networks for support of functional stability of information systems. In: *2018 IEEE First International Conference on System Analysis & Intelligent Computing (SAIC)*, pp. 1–4. IEEE (2018)
8. Billings, S.A.: *Nonlinear System Identification: NARMAX Methods in the Time, Frequency, and Spatio-Temporal Domains*. Wiley, Hoboken (2013)
9. Bonhomme, P.: Decentralized state estimation and diagnosis of p-time labeled petri nets systems. *Discret. Event Dyn. Syst.* **31**(1), 137–162 (2021)
10. Büyüksahin, Ü.Ç., Ertekin, Ş.: Improving forecasting accuracy of time series data using a new arima-ann hybrid method and empirical mode decomposition. *Neurocomputing* **361**, 151–163 (2019)
11. Czerwiński, W., Lasota, S., Lazić, R., Leroux, J., Mazowiecki, F.: The reachability problem for petri nets is not elementary. In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pp. 24–33 (2019)
12. Jahangir, H., Golkar, M.A., Alhameli, F., Mazouz, A., Ahmadian, A., Elkamel, A.: Short-term wind speed forecasting framework based on stacked denoising auto-encoders with rough ann. *Sustain. Energy Technol. Assess.* **38**, 100601 (2020)
13. Khasawneh, H.J., Abo-Hammour, Z.S., Al Saaideh, M.I., Momani, S.M.: Identification of hysteresis models using real-coded genetic algorithms. *Eur. Phys. J. Plus* **134**(10), 507 (2019)
14. Liu, M., Gao, Y., Yuan, Y., Yang, K., Shi, S., Zhang, J., Tian, J.: Efficacy and safety of integrated traditional Chinese and western medicine for corona virus disease: (covid-19): a systematic review and meta-analysis. *Pharmacol. Res.* **104**896, 2020 (2019)
15. Mokhtarzad, M., Eskandari, F., Vanjani, N.J., Arabasadi, A.: Drought forecasting by Ann, Anfis, and Svm and comparison of the models. *Environ. Earth Sci.* **76**(21), 1–10 (2017)
16. Momani, S., Abo-Hammour, Z.S., Alsmadi, O.M.: Solution of inverse kinematics problem using genetic algorithms. *Appl. Math. Inf. Sci.* **10**(1), 225 (2016)
17. Notton, G., Voyant, C., Fouilloy, A., Duchaud, J.L., Nivet, M.L.: Some applications of Ann to solar radiation estimation and forecasting for energy applications. *Appl. Sci.* **9**(1), 209 (2019)
18. Pazikadin, A.R., Rifai, D., Ali, K., Malik, M.Z., Abdalla, A.N., Faraj, M.A.: Solar irradiance measurement instrumentation and power solar generation forecasting based on artificial neural networks (ann): A review of five years research trend. *Sci. Total Environ.* **715**, 136848 (2020)
19. Russo, G., Pennisi, M., Fichera, E., Motta, S., Raciti, G., Viceconti, M., Pappalardo, F.: In silico trial to test covid-19 candidate vaccines: a case study with UISS platform. *BMC Bioinform.* **21**(17), 1–16 (2020)
20. Samuel, I.A., Adetiba, E., Odigwe, I.A., Felly-Njoku, F.C.: A comparative study of regression analysis and artificial neural network methods for medium-term load forecasting. *Indian J. Sci. Technol.* **10**(10), (2017)

21. Shoenfeld, Y.: Corona (covid-19) time musings: our involvement in covid-19 pathogenesis, diagnosis, treatment and vaccine planning. *Autoimmunity Rev.* (2020)
22. Sivaneasan, B., Yu, C., Goh, K.: Solar forecasting using ANN with fuzzy logic pre-processing. *Energy Proc.* **143**, 727–732 (2017)
23. Sohail, A.: Inference of biomedical data sets using Bayesian machine learning. *Biomed. Eng. Appl. Basis Commun.* **31**(04), 1950030 (2019)
24. Sohail, A., Arif, F.: Supervised and unsupervised algorithms for bioinformatics and data science. *Prog. Biophys. Mol. Biol.* **151**, 14–22 (2020)
25. Wanto, A., Hayadi, B.H., Subekti, P., Sudrajat, D., Wikansari, R., Bhawika, G.W., Sumartono, E., Surya, S.: Forecasting the export and import volume of crude oil, oil products and gas using ANN. In: *Journal of Physics: Conference Series*, vol. 1255, p. 012016. IOP Publishing (2019)
26. Xiao, N., Ni, C., Guo, S.: Modelling and simulation for production logistics system in industrial enterprises based on hybrid network. *Int. J. Simul. Model.* **16**(1), 157–166 (2017)
27. Yang, S., Jiang, L., Cao, Z., Wang, L., Cao, J., Feng, R., Zhang, Z., Xue, X., Shi, Y., Shan, F.: Deep learning for detecting corona virus disease 2019 (covid-19) on high-resolution computed tomography: a pilot study. *Ann. Transl. Med.* **8**(7), (2020)
28. Yu, Z., Sohail, A., Nutini, A., Arif, R.: Delayed modeling approach to forecast the periodic behaviour of sars-2. *Front. Mol. Biosci.* **7**, 386 (2020)
29. Yu, Z., Ellahi, R., Nutini, A., Sohail, A., Sait, S.M.: Modeling and simulations of covid-19 molecular mechanism induced by cytokines storm during sars-cov2 infection. *J. Mol. Liquids* **327**, 114863 (2021)
30. Zhu, G., Li, Z., Wu, N.: Model-based fault identification of discrete event systems using partially observed petri nets. *Automatica* **96**, 201–212 (2018)

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