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

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
The nexus between COVID-19 deaths, air pollution and economic growth in New York state: Evidence from Deep Machine Learning

Cosimo Magazzino a, Marco Mele b, Samuel Asumadu Sarkodie c,⁎

a Department of Political Sciences, Roma Tre University, Italy
b Department of Political Sciences, University of Teramo, Italy
c Nord University Business School, Norway

A R T I C L E   I N F O
JEL classification:
C45
Q53
Keywords:
COVID-19
Air pollution
Economic growth
New York state
Machine learning

A B S T R A C T

The aim of this paper is to assess the relationship between COVID-19-related deaths, economic growth, PM10, PM2.5, and NO2 concentrations in New York state using city-level daily data through two Machine Learning experiments. PM2.5 and NO2 are the most significant pollutant agents responsible for facilitating COVID-19 attributed death rates. Besides, we found only six out of many tested causal inferences to be significant and true within the AUPRC analysis. In line with the causal findings, a unidirectional causal effect is found from PM2.5 to Deaths, NO2 to Deaths, and economic growth to both PM2.5 and NO2. Corroborating the first experiment, the causal results confirmed the capability of polluting variables (PM2.5 to Deaths, NO2 to Deaths) to accelerate COVID-19 deaths. In contrast, we found evidence that unsustainable economic growth predicts the dynamics of air pollutants. This shows how unsustainable economic growth could increase environmental pollution by escalating emissions of pollutant agents (PM2.5 and NO2) in New York state.

1. Introduction

The viral pandemic caused by SARS-CoV-2 has raised numerous questions regarding the role of environmental pollution in the spread of the virus (Sarkodie and Owusu, 2020a,b). In particular, extensive scientific literature connects cases of viral infection with concentrations of atmospheric Particulate Matter (PM10 and PM2.5) (Cienciewicki et al., 2007; Sedlmaier et al., 2009; Perrino et al., 2014; Mele and Magazzino, 2020; Sterpetti, 2020). In general, the correlation between the spread of a virus and PM10 and PM2.5 levels passes through direct and indirect effects. Concerning the first, the atmospheric particulate would play the role of transport vector (Chen et al., 2010; Despres et al., 2012; Cao et al., 2014; Lei et al., 2018; Magazzino et al., 2020). Regarding indirect effects, prolonged inhalation of fine and ultrafine particulates would induce oxidative stress and cellular inflammation of the lungs, which would assist the adverse effects of respiratory viruses (COVID-19) on patients (Bengalli et al., 2019; Esquivel-Farina et al., 2019; Li et al., 2020; Putrino et al., 2020). These results have led numerous international and academic research centers to more precisely analyze the role of high concentrations of PM10 and PM2.5 in the spread, morbidity, and mortality attributable to the pandemic COVID-19 virus (Dominici et al., 2020; Magazzino et al., 2020a, b, c, d; Mahase, 2020; Remuzzi et al., 2020). The connection between pollution and the possible spread of viruses and new pandemics passes through the unsustainable economic growth process. In this situation, there is the case of the U.S. which continues to weaken the standards of environmental protection to promote more significant economic growth (Environmental Protection Agency, 2019). The correlation between environmental performance and growth in the economy of the U.S., like other countries, is a controversial topic often discussed. For the last three decades, it has been an area of concern for public policy to what extent the environment should be protected. Thus, the existing relationship between environmental sustainability and economic growth has sparked several discussions (Caliendo et al., 2018). The question remains on how and to what extent the two terms affect each other. The answer to this question requires an explanation of the distinctive and empirical-based assessment. Finite earthly resources place limits to the degree to which an economy can continually grow in the long-term. On the flip side, using environmental resources sustainably is dependent on economic growth. The cost of inaction in the U.S. is likely to be higher compared to the price of taking immediate action and precautionary measures.

It is evident in the U.S. that as the economy grows, pollution also increases – although less than proportionally (Sarkodie and Strezov, 2018). Energy is an essential resource in sustaining economic...
development because of the link between industrial production, consumption, and energy utilization (Bildirici, 2017). Thus, ensuring social welfare and sustained economic productivity comes with an environmental cost. Fossil fuels remain the principal source of energy for intensive economic activities (i.e., industrialization) and the main driver of global emissions through the release of various harmful gasses. Greenhouse gases (GHG) absorb heat from the sun and maintain it in the atmosphere, hence increase the temperature of the biosphere (Ritchie and Roser, 2017). Thus, climate change caused by pollution has a direct impact on economic development and vice versa. Pollution via the use of fossil fuels has both negative and positive economic effects in the U.S. Pollution has had far-reaching negative effects on the American economy. The impact touches on property value, tourism, and recreational business, among other sectors – for instance, water pollution in the U.S. More income (in billions) from the economy are channeled towards water purification. Ideally, economic growth induced by industrialization causes pollution to water masses, and thus, billions of dollars are used as water treatment costs. Airborne pollution affects visibility at popular outdoor destinations like parks, therefore, hampering the level of tourism activities that directly affect the GDP (Gupta et al., 2018). Building and structures, especially the ones made of limestone, can get damaged by pollution-related events, hence, depriving of the economy.

According to Donev et al. (2010), the allotted weight of diseases resulting from air and water pollution accounts for approximately a hundred million disability-related life years across the globe. The majority of such effects experienced in the U.S. impose a high cost on its overall economy. The U.S. ranks 2nd after China in terms of welfare cost estimated around $471,000 million incurred due to air pollution attributed disability-related life years (Owusu and Sarkodie, 2020). Generally, the natural environment adds to human capital in other possible ways. For instance, the availability of green space improves both psychological and physical sustenance of man, hence, ameliorates health and well-being. Green space and wildlife-rich parts availability could have far-reaching impacts (Sandberg et al., 2019). For instance, such areas can improve mental health, stress treatment, enhance productivity, and reduce crime (Eckelman and Sherman, 2016). Therefore, health and a clean environment could be an essential tool for attracting and maintaining investments across the U.S.

Demand for healthy and clean natural environment fosters an excellent opportunity for wealth-creation and general employment (Liu and Lim, 2017; Owusu et al., 2021). For instance, employment can be created in agricultural sector tasked with the protection and management of natural resources and industries geared towards the mitigation of economic challenges. For example, they do so through renewable energy techniques to manage waste via technologies and products that minimize both air and noise pollution from the production process. Other companies aim at addressing the adverse impacts of the environment through natural assets restoration. Such restoration activities may include the reclamation of land and treatment of wastewater. Collectively, all the industries work in synergy across the states of America to ensure the country’s economic sustainability.

The increased focus on environmental sustainability has allowed the U.S. to institute environmentally conscious and low carbon industries. The services and products that these industries produce facilitate the reduction of the environmental impacts of production (Koondhar et al., 2018). The consumption of renewable and low carbon energy technologies has mitigation effects on manufacturing, hence, reduces air pollution. Shifting to resource-efficient and economical carbon economy will call for a fundamental change in the economic structure. According to Shapiro and Walker (2018), investment in innovation and technology to replace old infrastructure minimizes future risks from changes in the environment. Thus, economic growth facilitates meeting such demands without compromising other crucial areas of investments.

This paper examines sustainable environmental policies to limit PM2.5 and NO2 concentrations and keep them under the identified threshold levels. Contrary to the extant literature, we estimate the nexus between air quality, economic growth, and high levels of death due to COVID-19 in New York state (USA). This investigation aims in identifying “critical” threshold levels due to the concentration of fine particulate matter and nitrogen dioxide. Statistical inferences from this study could be used to counteract a new widespread of contagion expected in autumn, to avoid the overlapping of COVID-19 and seasonal flu cases, wherever possible. This paper aims to estimate the critical concentration values of PM2.5 and NO2 connected to the spread-contagion and deaths from COVID-19 in the New York state through a model in Deep Learning to economic growth.

The structure of the paper proceeds as follows: In Section 2 we provide concise literature on the nexus between air pollution, health effects, and COVID-19 related deaths. In Section 3 we describe data used and empirical strategy employed. Section 4 presents the results and discusses estimated findings, while Section 5 provides statistical checks to validate the robustness of prediction model. Section 6 presents the conclusion and policy implications of the study.

2. Particle pollution, health effects, and COVID-19

Exposure to indoor and outdoor air pollution – particularly particulate matter (PM10 and PM2.5), nitrogen oxides (NO and NO2), and ozone (O3) – can have a range of adverse health effects already described in the scientific literature (Cieniewicki et al., 2007; Sedlmaier et al., 2009; Perrino et al., 2014; Sterpetti, 2020). In 2016, the World Health Organization (WHO) estimated 7 million global premature deaths per year related to air pollution, with 91% of deaths occurring in low-middle income countries in South Asia, sub-Saharan Africa, and Latin America. In Europe, around 550,000 premature deaths have been estimated. WHO has paid particular attention to health effects due to pollution levels of indoor environments through low-quality fuels for space heating and food preparation—the use of chemical substances for personal hygiene and cleaning environments, and aromas for indoor perfuming, paints, and varnishes. This component plays an important role in considering the time spent indoors (home, school, work, to name a few). Globally, the primary health effects related to indoor and outdoor air pollution are related to the increase in non-communicable diseases. They include chronic diseases of the cardiovascular system (myocardial infarction, cerebral stroke), those of the respiratory system (such as asthma, bronchial chronic obstructive pulmonary disease), and lung cancer for long-term exposure. More recently, the exposure and persistent air pollution (PM2.5) are associated with diseases such as diabetes, a delay in the neurological development of children, and degenerative neurological effects in the adult/elderly population (Chen et al., 2010; Després et al., 2012; Cao et al., 2014; Lei et al., 2018). Akbari et al. (2019) showed that a normal cell transforms into its cancerous counterpart in response to cellular adenosine triphosphate (ATP) depletion.

The short-term effects are supported by many studies dedicated to individual cities and aggregate urban areas concerned with reduced lung capacity, aggravation, complications of asthma and, exposure during the gestation period. Growing attention has also been devoted to the investigation of the most susceptible population to indoor and outdoor air pollution.

Another aspect to underline is the nature of particulate matter. PM is a complex mixture of organic and inorganic pollutants, consisting of the carbonaceous material deriving from the different combustion processes that generate it, but also from a set of other substances particularly toxic to humans (inorganic and organic micro-pollutants such as metals, hydrocarbons polycyclic aromatic, dioxins). The qualitative and quantitative composition of PM, therefore, varies greatly depending on the type of emission sources that produce it. The resulting health effects depend not only on the concentration levels to which the populations are exposed but also on many other factors, which include sources, physical and chemical transformations of precursors, climate, and specific local situation (orographic and topographic) of urban and non-urban areas.
that influence its “quality” and composition. The possibility of a direct association of the spread of SARS-CoV-2 infection in areas with high levels of atmospheric pollution has policy implications. SARS-CoV-2 is a viral infection, subjected to mechanisms of transmission other than virus contamination. We also noted that following government provisions, the reduced mobility of people, and the closure of many production activities, has led to a progressive and significant reduction in air pollution levels in \( \text{PM}_{2.5}, \text{PM}_{2.5-10}, \text{NO}_2, \) and benzene) across countries through the lockdown policy \citep{sarkodie2020, mele2020}. Mele and Magazzino \citep{mele2020} used Toda-Yamamoto causality tests and a D2C algorithm to explore the relationship between pollution emissions, economic growth, and COVID-19 deaths in India. The results highlight unidirectional causality between economic growth and pollution.

Following Magazzino et al. \citep{magazzino2020c} and Mele and Magazzino \citep{mele2020} used GMM estimation technique for the Commonwealth of Independent States (CIS) members in the 1993–2018 years. The major results revealed that the highest variability of mortality could be explained by \( \text{CO}_2 \) variability. The estimation proved also that fossil fuel consumption positively affects mortality from cardiovascular disease, diabetes mellitus, cancer, and chronic respiratory disease. The first experiment is a Vector Error Correction Model (VECM) and panel Generalized Method of Moments (GMM) using data from ten Southeast Asian countries over the period 2000–2016 to explore the possible association between emissions, lung cancer, and the economy. The results confirm that \( \text{CO}_2 \) and \( \text{PM}_{2.5} \) are major risk factors for lung cancer in the region. \cite{rasoulinezhad2020} used GMM estimation technique for the \( \text{CO}_2 \) and \( \text{PM}_{2.5} \) concentrations and COVID-19 fatality in New York state, compared to the number of COVID-19 deaths. We perform our analysis in DML following the process shown in Fig. 1.

Once the dataset is built, the NN architecture is chosen and we process a Deep Learning approach in the Oryx protocol. We carry out statistical tests on the series of data used. These tests express the correct preparation of the dataset. Subsequently, we identify the inputs and target of the first model concerning a hyperbolic or logistic activation function. Then, we identify through tests in ML, the correctness of the chosen architecture, and if not right, we build a NN. We then select the NN that has a minor prediction error. Subsequently, we verify the link of causality through the “shear” procedure using a decision tree model through the BigML package in ML. According to \cite{hu2012}, the second experiment is a D2C (ML) causality model. We use the result of the hyperbolic equations of the first experiment by adding the change in the economic growth of the state of New York. For the variable representing economic growth (\( \text{rGDP} \)), we use the change in real GDP. \footnote{1} We perform our analysis in D2C-ML following the process shown in Fig. 2.

The D2C auxiliary analysis is complex with two interconnected starting points: Functional Component Key and Functional Component Aux. The first derives from the dataset used for the NN in which the variable relating to real Gross Domestic Product (GDP) has been added. The second, on the other hand, represents a set of hyperbolic equations that are a result of the NN process. These two datasets make up the Service Monitoring that generates the event to be analyzed. Subsequently, an Incident Component is generated without supervision necessary to verify the correctness of the distribution of the two datasets. If this test fails, we produce the D2C causality model. If not, we analyze the presence of control errors in the system through diagnostic tests.

4. Empirical results

Generally, before analyzing the results of an empirical time series methodology, it is necessary to examine the time series properties of the variables and the presence of cointegration. However, according to \cite{sundararajan2017}, an experiment in DML considers a dataset as a stationary element. Hence, analysis is not affected by the presence of a time series in the data. Therefore, once the dataset was created in CSV and forwarded within the NN protocol, we observed the distribution of the data. \cite{tableAappendix} presents the summary of the dataset used in the Oryx processing of the NN. The variables used are 9, of which 8 represent the input process, and 1 is the generated target. Fig. 3 shows the plot scatter chart of all variables, whilst we report the behaviour of the instances through a pie chart elaborated by the protocol in Figure A in Appendix.

Source: our elaborations Oryx 2 - Apache Spark.

The instances generated by the automatic intelligence process are 47. Of these, training is the most important since it has 29 instances (61.7%). This first result highlights how, on an automatic level, the process could select 29 motion capable of best estimating the target out of a total of 47 potential ML projects. As for the requests relating to the selection, they are equal to 9, which is 19% of the total requests. This value for the theory of NN must be lower than that of the training instances. Therefore, this result also confirms the goodness of the model.
generated through the NN. The result of the “instant selection” provides another interpretation if compared with the instances testing. Their value is the same (testing 9) and therefore, the models selected by the previous instances are all potentially testable in the NN. Finally, we can see that there are no unused instances (unused = 0) and therefore, no outliers were generated that would invalidate the results. The results obtained from the instances’ analysis confirm the possibility of generating a NN according to the architecture chosen by the operator.

The result of the graphic elaboration of the NN with 8 inputs and 1 output generated in Oryx with Designer extension version 1.8 is depicted in Figure B in Appendix. The architecture of the NN reveals a complexity, represented by hidden neurons, of 13:8:7:11:3. The distribution of the NN is not of the linear type but has 2 redundant neural states that are active towards the end of the NN. While the yellow filled circles represent the scaling neurons, the blue filled circles are the perceptron neurons, whereas the red filled circles represent the unscaling neurons. The activation function chosen for this analysis is logistic, while a linear function generates the target. Each combination of the inputs generates the variation of COVID-19 deaths in a process with no anomalous values. Therefore, the plot of NN can be read as the n combinations of inputs that generate a variation of the target through a logistic combination with a linear result.

Once the NN is built, we proceed through numerous tests to validate the model. These tests assist the operator to identify the probability that there is a better algorithm different from the one chosen in the initial phase.

The analysis that tests our algorithm’s goodness for the creation of the NN begins with the Perform Training using the Quasi-Newton method (Figure C in Appendix). It uses Newton’s system of ANNs in ML process. However, compared to self-generated networks in which a “block of losses” is not selected, our test does not require secondary derivatives calculation. It is called “Quasi-Newton” because it calculates an approximation of the Hessian inverse to each recurring signal of the algorithm. In other words, it uses only the gradient information. Also,
despite using standard deviation, we use the Minkowski error as a test with a minor index. The Minkowski error represents the failure test in ML with the highest performance and most significant fit in an ANN.

Figure C shows the training and selection errors in each iteration. The blue line shows the graph’s training error, while the orange line represents the selection error. As we can see from the error trend, the training error shows a decreasing trend. It records a value of 1.9816 in its initial phase. However, after 12 epochs, the error drops to a value of 0.0383. As regards the trend of the selection error, it too has a decreasing trend. The value of 1.991 in its initial phase reaches a value of 0.0425 after 10 epochs. The decreasing trend of the two error tests concerning the increasing epochs highlights how our strategy has proved ideal for the NN elaboration process. The analysis that verifies the perfect performance of the NN also requires the Performed Order Selection test. This test suggests to the operator the model’s choice, whose complexity is the most appropriate to produce an adequate data fit. This testing process is fully automatic and left to the interpretation of artificial intelligence guided by our algorithm. It is responsible for finding the optimal number of neurons in the network. This process takes place by combining our initial algorithm with the one pre-set by the software, the incremental order algorithm. Fig. 4 shows the error history for the different subsets during the incremental order selection process. The blue line shows the training error while the orange line symbolizes the selection error.

Both the training error and selection error decrease with increasing order in point number 2. These results highlight the presence of a better-hidden architecture. It presents even lower algorithm errors than our initial architecture. Therefore, following the ML process in Fig. 5, we elaborate on the final architecture of this NN.

The results of our elaboration in light of the findings achieved with the Perform Order Selection test are presented in Fig. 5. Based on 8 inputs and 1 output variable, the complexity is 10:8:2, represented by

Fig. 3. Plot scatter chart.

Fig. 4. Incremental Order Error plot test. Notes: (1) 13:8:7: 11:6 MSE 0.21; (2) 13:8:7:11:3 MSE 0.35; (3) 13:8:7:11:2 MSE 0.25; (4) 13:8:7:11:7 MSE 0.30; (5) 13:8:7:9:6 MSE 0.22; (6) 13:8:7:9:3 MSE 0.03.
Source: our elaborations Oryx 2 – Apache Spark.
the numbers of hidden neurons. Therefore, compared to Figure B, this NN presents an automatic choice that has reduced the fourth hidden layer of the neuron. The target generated by the NN process remains the same and identifiable in the variation of per capita GDP.

Subsequently, to ensure the most appropriate goodness of the results in a model of black networks, we test the final architecture through several operations. One of these is to perform a linear regression analysis between the scaled NN outputs and corresponding targets for an independent test subset (Figure D in Appendix). This important analysis in a predictive model considers three parameters for each output variable. The first two parameters, referable to letters a and b, correspond to the y-intercept and slope of the best predictive linear regression related to the targets which are scaled. The third parameter is the usual $R^2$-square, already widely used in statistics and econometrics.

The prediction line (with respect to the target, Deaths) perfectly confirms the goodness of the elaborated algorithm on the final architecture. The prediction shows an intercept of 1.3, very high correlation value (0.995), and slope of 0.984, implying a slope close to unity. Now, we examine the input that influences the target’s prediction (to a greater extent) to ascertain the process of economic growth. To this end, we enrich our algorithm with the “Model deployment” protocol developed by Artificial Intelligence Techniques, Ltd. This protocol generates the weight of all the inputs in NN, concerning the target.

The results of the 8 inputs used for our final architecture of the NN are presented in Table B in Appendix. All variables (and their transformations) have positive values. Of all the inputs, we can observe that the most significant positive value is represented by PM$_{2.5}$ and NO$_2$. These results confirm that air pollution plays an essential role in facilitating COVID-19 attributable deaths. The transform versions of the variables have lower values, hence, we test these results through the so-called “Decision Tree Model” (Fig. 6). Our decision tree is a system with $n$ input and output variables. Input variables (attributes) are derived from observation of the environment. The last output variables, on the other hand, identify the decision/action to be taken. Decision making is represented by an inverted logical tree where each node is a conditional function. Each node tests a condition (test) on a particular property of the environment (variable) and has two or more downward branches in operation. The process consists of a sequence of tests. It always starts from the root node, parent node located higher in the structure, and proceeds downwards. Depending on the values detected in each node, the flow takes one direction or another and proceeds progressively downwards. The final decision is found in the terminal leaf nodes, the lower ones. In this way, after analyzing the various conditions, the agent reaches the final decision.

The target value on our decision tree in Fig. 6 is COVID-19 deaths whereas the initial question is how PM$_{2.5}$ and NO$_2$ interplay in causation. In the left part of the tree, the answer is yes while on the right side, the answer is no. We note that the left part of the tree immediately shows a final reply attributable to a green dot compared to a purple one. Since the green dot is the probability that air pollutant emissions cause COVID-19 deaths, we instantly register an affirmative response of 50%. The decision tree confirms that compared to an alternative choice, pollutant emissions exacerbate deaths from COVID-19 in New York state.

We further perform Deep Learning approach with D2C (ML) causality tests to determine the relationship between the sampled variables. We use the results of the hyperbolic equations of the first experiment in ANNs by adding the change in the economic growth of the state of New York ($rGDP$). In the model, $n$ filtered factors are used (which do not
appear in Table 1) to perform the task of training the classification of our model. Thirty-six classifiers are trained and tested by the self-learning machine to achieve the predictive causal link between the variables. These thirty-six classifiers work through a binary calculation sequence, alternating the values [0] and [1].

Table 1 clearly shows how our algorithm performed numerous repetitions for each combination of causality between the variables. On average, these combinations recorded a value of 19,500 repetitions. The item “Percentage (%)” represents the closing percentages of the calculation within the 19,500 repetitions average. The results of this item of the predictive causation process have always been greater than 70%. This highlights how our algorithm can complete each cycle for each pair of variables. The mean causality value is uniform for all pairs of values analyzed. The item “AUPRC” represents the one able to report the level of significance of the results of predictive causality. Compared to normal prediction models, it uses the binary system (0 or 1) to formulate the question “true or false” regarding the P-Value lower or higher than 5%. We observe only six significant causal relationships within the AUPRC analysis. The causal relationships are attributable to a one-way causality ranging from PM$_{2.5}$ to Deaths, NO$_2$ to Deaths, from lnPM$_2.5$ to lnDeaths, from rGDP to PM$_{2.5}$, from d.rGDP to dPM$_{2.5}$, and from d.rGDP to dNO$_2$. These results confirm that the “polluting” variables have a predictive causality that stimulates the death of COVID-19. The result obtained is different from those obtained from alternative models to our algorithm. The previously identified causal links represent the result of our ML model’s autonomous action. In other words, it has combined numerous processes without the intervention of an external operator – but only by analyzing the environment of the variables that make up the dataset. Indeed, adding GDP, the experiment shows that economic growth declines environmental sustainability by promoting high emissions of fine particulate matter and carbon dioxide.

5. ML test on control cities

To test the results obtained on the state of New York, we selected two cities: Tokyo and Milan. We chose these two control cities because they show different epidemiological and environmental patterns from official statistical sources between March and December 2020. These cities are characterized by dynamic economies: Milan has a GDP of 367 billion dollars compared to Tokyo (excluding greater Tokyo area) with a GDP of 1857 billion dollars in 2019. Although these two cities represent financial centers with a medium-high population density but recorded a different contagion/death ratio from Covid-19 during the pandemic. Milan recorded contagion/death value of 5.12%, whereas Tokyo recorded about 1% contagion/death value. This result may be attributable to numerous determinants including health system efficiency, virus containment, and environmental measures. Since Tokyo implemented a strict infection tracking policy, the city recorded low levels of PM$_{2.5}$ and NO$_2$.

On the other hand, Milan failed in the tracking policy, recording high levels of PM$_{2.5}$ and NO$_2$. We, therefore, applied the D2C predictive causation model to the two cities. In other words, we adopted the same ML model used for the state of New York to verify whether the number of deaths-related to Covid-19 in the two cities was influenced (predictively and without supervision) by the levels of environmental pollution. Since Tokyo implemented a social control policy by tracing the infected population, we inserted a controlling factor—the so-called “matrix approximate control”. In this way, the algorithm was able to equate only the two cities regarding environmental effects.

The results obtained from the predictive causality system in ML confirm previous findings for New York. If we observe the model’s results for Tokyo (Tables 2 and 4), we find no relationship between Covid-19 deaths and emissions (AUPRC). Almost all of the choices between true and false events have fallen on false events. Moreover, the model presented an identified AC value for all combinations on predictive causality. This result, undoubtedly influenced by the control matrix reveals no discrepant value capable of hypothesizing an event error between true and false events. The results, however, were entirely different from the city of Milan (Tables 3 and 5). It recorded poorly than that of New York State on the link between pollutants and deaths from Covid-19. In our model, the predictive causality was recorded for each variable transformed (ln. or d.). In other words, the unsupervised algorithm interprets such events in all forms of variation: stable, logarithmic, or in first difference, consistent with Magazzino et al. (2020b).

Finally, Tables 4 and 5 show the goodness of our results. Regarding Tokyo, the algorithm analyzed a predictive path of 597 epochs. After 7.01 epochs, training and selection error recorded its lowest value. The orthogonal matrices preserve gradient norm during backpropagation to a value of 0.0132, confirming the model’s optimality. On the other hand, the test on Milan required several major epochs (612). However, the control parameters were normalized to a lower epoch level (6.72). The training and selection error results are low and within a range of 1–10%. The final gradient norm is in line with DML theory and lower than the two previous tests.

6. Concluding remarks and policy recommendations

This study estimated the relationship between air quality and high levels of COVID-19 attributed deaths in New York state, USA. In a bid to have a clear path of unveiling this study, we considered the interaction of pollution, economic growth, and COVID-19 deaths as better indexes and proxies in measuring environmental and economic impacts. Our
empirical results confirmed the nexus among unsustainable growth, air pollution, and the spread of COVID-19 pandemic in New York state. We employed two experimental approaches based on DML experience for a thorough investigation of the proposed theme. The first experiment assessed the relationship between pollution and the number of deaths from COVID-19, while the second experiment performed the hyperbolic equations of the first expression in generating the causal inferences thorough investigation of the proposed theme. The first experiment—within the AUPRC analysis to test causal inferences. In line with the deaths from COVID-19 in New York are possibly caused by air pollutant agents (PM$_{2.5}$ and NO$_2$). According to an international study, the decision tree affirmed with a 50% response rate that significant pollutant agents responsible for facilitating death rates caused by COVID-19. The decision tree affirmed with a 50% response rate that deaths from COVID-19 in New York state that thwarts environmental sustainability by eventually activating emissions of pollutant agents (PM$_{2.5}$ and NO$_2$). The resulting effect of emissions from pollutant agents is the spread of viruses such as COVID-19—which could trigger

Table 4
Optimization Test in Machine Learning results for Tokyo.

| Final parameters norm | 701 |
|-----------------------|-----|
| Final training error  | 0.044 |
| Final selection error | 0.0916 |
| Final gradient norm   | 0.0152 |
| Epochs number         | 597 |
| Elapsed time           | 00:00 |
| Stopping criterion    | Maximum selection error increases |

Source: our elaborations with Oryx 2 – Apache Spark.

Table 5
Optimization Test in Machine Learning results for Milan.

| Final parameters norm | 6.72 |
|-----------------------|-----|
| Final training error  | 0.0445 |
| Final selection error | 0.0962 |
| Final gradient norm   | 0.0356 |
| Epochs number         | 612 |
| Elapsed time           | 00:00 |
| Stopping criterion    | Maximum selection error increases |

Source: our elaborations with Oryx 2 – Apache Spark.

Notes: AC: Average Causality value. AUPRC: Area Under the Precision-Recall Curve. True: P-Value < 0.05. False: P-Value ≥ 0.05.

Source: our elaborations with Oryx 2 – Apache Spark.
The empirical evidence showed the impact of COVID-19 pandemic in New York is majorly caused by the fast spread of pollutant agents due to emission-embedded economic growth. Thus, the following policy recommendations emanate from the study: To build a lasting solution in containing the impact of COVID-19 pandemic and future occurrences, Government intervention policies could encourage decarbonization of economic development through pollution abatement technologies. A call for a paradigm shift is inevitable through the adoption of research, development, and innovations across economic sectors. Decoupling economic growth from air pollution requires the dependence on renewable and low carbon energy to reduce emissions from the manufacturing sector. Awareness of a healthy and clean environment could be an effective tool in limiting the fast spread of the virus. The heterogeneous nature of countries across the globe and complexity of COVID-19 appears a limiting factor of our study to achieve homogeneous results, however, our empirical procedure set a foundation for future research necessary to curtail the effect of COVID-19 and future pandemics.

Finally, a decrease in air pollutants might decrease the global temperature through a decline in GHG emissions that may in-turn increase the infectivity – and possibly mortality – of COVID-19 (Haque and Rahman, 2020). Therefore, future research may possibly investigate the role of other pollutant species on the stability and spread of the global pandemic while controlling for other confounders.

Credit author statement

Cosimo Magazzino: Conceptualization, Methodology, Software, Supervision. Marco Mele: Visualization, Investigation, Validation, Writing-Reviewing and Editing. Samuel Asumadu Sarkodie: Writing – original draft, Writing-Reviewing and Editing.

Declaration of competing interest

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

Acknowledgments

Open Access funding provided by Nord University.

Nomenclature

- **ANNs**: Artificial Neural Networks
- **CIS**: Commonwealth of Independent States
- **CO2**: Carbon Dioxide
- **COVID-19**: Coronavirus Disease 19
- **D2C**: Causal Direction from Dependency
- **DML**: Deep Machine Learning
- **GDP**: Gross Domestic Product
- **GHG**: Greenhouse Gas
- **GMM**: Generalized Method of Moments
- **ML**: Machine Learning
- **NO**: Nitric Oxide
- **NO2**: Nitrogen Dioxide
- **O3**: Ozone
- **PM2.5**: Particulate Matter with an aerodynamic diameter <2.5 μm
- **PM10**: Particulate Matter with an aerodynamic diameter <10.0 μm
- **VECM**: Vector Error Correction Model
- **WHO**: World Health Organization

Appendix

![Fig. A. Instances pie chart. Source: our elaborations Oryx 2 – Apache Spark.](image-url)
Fig. B. ANNs results. Source: our elaborations Oryx 2 – Apache Spark.

Fig. C. Quasi-Newton method algorithm error history. Source: our elaborations Oryx 2 – Apache Spark.

Fig. D. Predictive Linear Regression test. Source: our elaborations Oryx 2 – Apache Spark.
Table A

Variables bars chart

Source: our elaborations in Oryx 2 – Apache Spark.

Table B

| Input name | Value |
|------------|-------|
| PM$_{2.5}$ | 49    |
| NO$_2$     | 25.500|
| lnPM$_{2.5}$ | 3.676|
| lnNO$_2$   | 2.920 |
| dPM$_{2.5}$ | 0.130 |
| dNO$_2$    | 0.375 |
| lnDeaths   | 3.190 |
| dDeaths    | 0.520 |

Source: our elaborations Oryx 2 – Apache Spark.

References

Akbari, H., Taghizadeh-Hesary, F., Heike, Y., Bahadori, M., 2019. Cell energy: a new hypothesis in decoding cancer evolution. Arch. Iran. Med. 22 (12), 733–735.

Bengalli, R., Zerboni, A., Marchetti, S., Longhi, E., Prada, M., Camantini, M., Manteca, P., 2019. In vitro pulmonary and vascular effects induced by different diesel exhaust particles. Toxicol. Lett. 306, 13–24.

Bildirici, M.E., 2017. The causal link among militarization, economic growth, CO$_2$ emissions, and energy consumption. Environ. Sci. Pollut. Res. 24 (5), 4625–4636.

Caliendo, L., Parro, F., Rossi-Hansberg, E., Sarte, P.D., 2018. The impact of regional and sectoral productivity changes on the U.S. economy. Rev. Econ. Stud. 85 (4), 2042–2096.

Cao, C., Jiang, W., Wang, B., Fang, J., Lang, J., Tian, G., Jiang, J., Zhu, T.F., 2014. Inhalable microorganisms in Beijing’s PM2.5 and PM10 pollutants during a severe smog event. Environ. Sci. Technol. 48, 1499–1507.

Chen, R.J., Chen, B.H., Kan, H.D., 2010. A health-based economic assessment of particulate air pollution in 113 Chinese cities. China Environ. Sci. 410–415.

Cierguezki, J., Jaspers, I., 2007. Air pollution and respiratory viral infection. Inhal. Toxicol. 19 (14), 1135–1146.

Desprez, V.R., Huffman, J.A., Burrows, S.M., House, C., Safatov, A.S., Buryak, G., Frohlich-Nowozisky, J., Elbert, W., Andreo, M.O., Pischl, U., Jaenicke, R., 2012. Primary biological particles in the atmosphere: a review. Tellus B 64, 15086.

Dominici, F., Wu, X., Nethery, R., Sabath, B., Braun, D., 2020. Exposure to Air Pollution and COVID-19 Mortality in the United States: A Nationwide Cross-Sectional Study. medRxiv.

Gupta, M.R., Dutta, P.B., 2018. Tourism development, environmental pollution, and economic growth: a theoretical analysis. J. Int. Trade Econ. Dev. 27 (2), 125–144.

Haque, S.E., Rahman, M., 2020. Association between temperature, humidity, and COVID-19 outbreaks in Bangladesh. Environ. Sci. Pol. 114, 253–255.

Koondhara, M.A., Qiu, L., Li, H., Liu, W., He, G., 2018. A nexus between air pollution, energy consumption, and growth of the economy: a comparative study between the USA and China-based on the ARDL bound testing approach. Agric. Econ. 64 (6), 265–276.

Liu, Q., Lin, S.H., 2017. Toxic air pollution and container port efficiency in the USA. Marit. Econ. Logist. 19 (1), 94–105.

Magazzino, C., Mele, M., Schneider, N., 2020a. A Machine Learning approach on the relationship among solar and wind energy production, coal consumption, GDP, and CO$_2$ emissions. Renew. Energy 167, 99–115.

Magazzino, C., Mele, M., Schneider, N., 2020b. The relationship between air pollution and COVID-19-related deaths: an application to three French cities. Appl. Energy 279.

Magazzino, C., Mele, M., Schneider, N., 2020c. The relationship between municipal solid waste and greenhouse gas emissions: evidence from Switzerland. Waste Manag. 113, 508–520.

Mele, M., Schneider, N., Sarkodie, S.A., 2021. Waste generation, Wealth and GHG emissions from the waste sector: is Denmark on the path towards Circular Economy? Sci. Total Environ. 755 (1), 142510.

Mele, M., Schneider, N., Valtell, G., 2020d. The relationship between nuclear energy consumption and economic growth: evidence from Switzerland. Environ. Res. Lett. 15.

Mantecca, P., 2019. In vitro pulmonary and vascular effects induced by different diesel exhaust particles. Toxicol. Lett. 306, 13–24.

Owusu, P.A., Sarkodie, S.A., 2020. Global estimation of mortality, disability-adjusted life years and welfare cost from exposure to ambient air pollution. Sci. Total Environ. 742, 140636 https://doi.org/10.1016/j.scitotenv.2020.140636.

Owusu, P.A., Sarkodie, S.A., Pedersen, P.A. 2021. Relationship between mortality and health care expenditure: Sustainable assessment of health care system. PLoS One 16 (2), e0247413. https://doi.org/10.1371/journal.pone.0247413.

Perrino, C., Catrambone, M., Dalla Torre, S., Ranita, E., Sargolini, T., Canepari, S., 2014. Seasonal variations in the chemical composition of particulate matter: a case study in the Po Valley. Part I: macro-components and mass closure. Environ. Sci. Pollut. Res. 21, 3999–4009.

Putrino, A., Raso, M., Magazzino, C., Galluccio, G., 2020. Coronavirus (COVID-19) in Italy: knowledge, management of patients and clinical experience of Italian dentists during the spread of contagion. BMC Oral Health 20, 200.

Remuzzi, A., Remuzzi, G., 2020. COVID-19 and Italy: What next? Lancet 395 (10231), 1222–1225.

Ritchie, H., Roser, M., 2017. Remuzzi, A., Remuzzi, G., 2020. COVID-19 and Italy: What next? Lancet 395 (10231), 1222–1225.

Ritchie, H., Roser, M., 2017. COVID and Greenhouse Gas Emissions. Our World in Data.

Sandberg, M., Klockars, K., Wilen, K., 2019. Green growth or degrowth? Assessing the normative justifications for environmental sustainability and economic growth through critical social theory. J. Clean. Prod. 206, 133–141.
Sarkodie, S.A., Strezov, V., 2018. An empirical study of the environmental Kuznets curve and environmental sustainability curve hypothesis for Australia, China, Ghana, and the USA. J. Clean. Prod. 201, 98–110.

Sarkodie, S.A., Owusu, P.A., 2020a. Impact of meteorological factors on COVID-19 pandemic: evidence from top 20 countries with confirmed cases. Environ. Res. 191, 110101.

Sarkodie, S.A., Owusu, P.A., 2020b. Global assessment of environment, health and economic impact of the novel coronavirus (COVID-19). Environ. Dev. Sustain. 1–11.

Sedlmairer, N., Hoppenheidt, K., Krist, H., Lehmann, S., Lang, H., Büttner, M., 2009. Generation of Avian Influenza Virus (AIV) contaminated fecal fine particulate matter (PM$_{2.5}$): genome and infectivity detection and calculation of immission. Vet. Microbiol. 139 (1–2), 156–164.

Shapiro, J.S., Walker, R., 2018. Why is pollution from U.S. manufacturing declining? The roles of environmental regulation, productivity, and trade. Am. Econ. Rev. 108 (12), 3814–3854.

Sterpetti, A.V., 2020. COVID-19 diffusion capability is its worst, unpredictable characteristic. How to visit a patient from a distance. Br. J. Surg. 107 (7), e181.

Sundararajan, M., Taly, A., Yan, Q., 2017. Axiomatic attribution for deep networks. In: Precup, D., The, Y.W. (Eds.), Proceedings of the 34th International Conference on Machine Learning, pp. 3319–3328.

Taghizadeh-Hesary, F., Rasoulinezhad, E., Yoshino, N., Chang, Y., Taghizadeh-Hesary, F., Morgan, P.J., 2020. The energy-pollution-health nexus: a panel data analysis of low- and middle-income Asian countries. Singapore Econ. Rev. https://doi.org/10.1142/S0217590820430043.

Taghizadeh-Hesary, F., Taghizadeh-Hesary, F., 2020. The impacts of air pollution on health and economy in Southeast Asia. Energies 13 (7), 1812.