



Evidence of the correlation between a city's air pollution and human health through soft computing

Salvatore Rampone¹ · Alessio Valente²

Accepted: 6 August 2021 / Published online: 17 August 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Huge quantities of pollutants are released into the atmosphere of many cities every day. These emissions, due to physicochemical conditions, can interact with each other, resulting in additional pollutants such as ozone. The resulting accumulation of pollutants can be dangerous for human health. To date, urban pollution is recognized as one of the main environmental risk factors. This research aims to correlate, through soft computing techniques, namely Artificial Neural Networks and Genetic Programming, the data of the tumours recorded by the Local Health Authority of the city of Benevento, in Italy, with those of the pollutants detected in the air monitoring stations. Such stations can monitor many pollutants, i.e. NO₂, CO, PM₁₀, PM_{2.5}, O₃ and Benzene (C₆H₆). Assuming possible effects on human health in the medium term, in this work we treat the data relating to pollutants from the 2012–2014 period while, the tumour data, provided by local hospitals, refer to the time interval 2016–2018. The results show a high correlation between the cases of lung tumours and the exceedance of atmospheric particulate matter and ozone. The explicit genetic programming knowledge representation allows also to measure the relevance of each considered pollutant on human health, evidencing the major role of PM₁₀, NO₂ and O₃.

Keywords Air pollution · Tumour data · Soft computing techniques · Benevento · Italy

1 Introduction

The economic, industrial and demographic development that took place over the last two centuries has allowed a definite improvement in the quality of human life. However, it has caused deep and rapid changes in the environment and its components (air, water, soil and subsoil, etc.). In particular, large quantities of pollutants continue to pour into the atmosphere mainly from combustion processes in transport, domestic heating and industrial production (European Environment Agency, 2018; Sicard et al., 2021). Meteorological variables are of fundamental importance concerning the levels of atmospheric pollution.

They regulate the speed with which pollutants are transported and are dispersed in the air or brought to the ground. Meteorological parameters also influence the speed (or even the presence) of some chemical reactions that determine the formation in the atmosphere of secondary pollutants, such as ozone (e.g. solar radiation). In many cases, the meteorological conditions do not allow the removal of the pollutants, and therefore in some cities, these substances tend to remain in the lower part of the atmosphere in the so-called Planetary Bounded Layer (PBL). An accumulation of pollutants is generated in the PBL when the dilution capacity of pollutants in the atmosphere exceeds the emissive capacity (Seigneur, 2019). Thus, dangerous concentrations are reached for human health and the balance of ecosystems. Even though the quality of the air has improved (a strong decrease in emissions in Italy and Europe), air pollution is still recognized as one of the main environmental risk factors.

The most critical of such pollutant concentration concerns urban areas, where the anthropization of the territory is at its maximum and where the levels of some pollutants

✉ Salvatore Rampone
rampone@unisannio.it

¹ Department of Law, Economics, Management and Quantitative Methods (DEMM), Università del Sannio, Benevento, Italy

² Department of Science and Technology (DST), Università del Sannio, Benevento, Italy

are far beyond the threshold of attention. However, the monitoring of air quality, which in recent years has achieved greater development, has played an important role in this action of attention and mitigation. If sulphur dioxide, carbon monoxide, benzene and lead are not currently a problem, except at the local level and in specific circumstances, particulate matter (PM₁₀ and PM_{2.5}), ozone (O₃) and dioxide nitrogen (NO₂) are among those that are most alarming (Seigneur, 2019). In particular, atmospheric particulate matter and ozone are recognized as the main culprits of human health effects. In fact, according to WHO (World Health Organization), the incidence of mortality due to exposure to pollutants in the air is high and increasing, especially in some urban areas (WHO, 2016). For example, for an increase of 5 µg/m³ of fine dust (PM_{2.5}), there would be a significant increase in the risk of anticipated mortality by 7%. Generally, the dust dispersed in the air obstructs the correct airflow inspiration or can cause respiratory diseases, pulmonary or bronchial bleeding, and even cancer (Raaschou-Nielsen et al., 2013). Moreover, it is believed that polluted air has harmful effects not only on humans but also on animals, vegetation, materials and ecosystems as a whole. The obligation for the administrations to prepare a plan for the quality of the air, when the levels exceed their assigned limits, can be a help to activate recovery measures in the urban area. However, to implement effective actions that lead to a reduction in the impact of environmental pollution, it is necessary to understand to what extent the contaminants cause negative effects on humans.

The purpose of this work is to correlate the data of the tumours, namely lung tumours registered at the Local Health Authority (ASL) of the city of Benevento (Italy), with the data of the concentration of pollutants detected in the air quality monitoring stations of the same city. To this aim, we use soft computing techniques, more precisely artificial neural networks (Bishop, 1996; Haykin, 2008) and genetic programming (Koza, 1992; Cramer 1985). These techniques have already been tested both to predict air conditions and to establish possible effects on environmental components and human health (Shiri et al. 2012; Stanislawski et al. 2012; Rampone and Valente, 2017–2019; Chen et al., 2020; D’Alelio et al., 2020; Elia et al., 2020; Rampone et al., 2021), although they do not appear to have been used for the same purpose as this work.

This paper is organized as follows: In Sect. 2 we resume the study area characteristics; in Sect. 3 we describe the data set used; in Sect. 4 we detail the soft computing processing techniques applied, the training and validation choices and the experimental results; the Sect. 5 is devoted to the Conclusions.

2 Study area

Our study area is the urban area of Benevento, located at 135 m a.s.l. in a vast basin in the inner area of Campania in southern Italy (Fig. 1). Such basin is bordered to the north and west by reliefs, whose nearest peaks reach 1400 m, while it is surrounded to the east and south by hilly ridges, with altitude which generally remains below 600 m. The city, in particular, develops at the confluence of the Calore and Sabato rivers, however, in the basin numerous minor watercourses run through it, sometimes deeply engraving it.

These geomorphological features strongly condition the climate of Benevento, which can be defined as temperate based on meteorological data recorded in its station for over seventy years. The average temperature recorded by the station relative to Benevento is 14.9 °C. The average of the coldest month is 6.9 °C (January), that of the warmest month 23.9 °C (July), with a temperature range of 16.9 °C and four months with temperatures above 20 °C. Overall rainfall do not reach 800 mm and is concentrated mainly in the autumn months and subordinately in the winter months. This quantity is sharply lower than in the other inland areas of Campania. The humidity in winter is high, frequently exceeding 70%, while in summer it is significantly reduced. The reigning winds are those coming from the south-west, while those that blow with more intensity (dominant winds) are from the north.

Atmospheric stability is a frequent condition in Benevento, deriving from baric configurations, like an anticyclone type, which often generate the phenomenon of thermal inversion. In this phenomenon, the temperature, rather than decreasing with the altitude, is reversed both in the early evening, night and early morning, and less frequently throughout the day. Due to the strong nocturnal radiation, the ground cools, as well as the layer in contact with it, especially in correspondence with the orographic depressions. This phenomenon causes frosts and fog, which can persist even during the day. Such conditions would prevent the vertical dispersion of any contaminants present.

In the vast municipal area of over 13,000 m², the possible sources of contaminants can be identified in the vehicles that cross the often congested urban centre and the peripheral area with some important roads (Appia, Telesina, etc.), as well as in the residential areas and industrial plants. The latter gather at the edge of the inhabited centre in a sort of radial pattern, which has been supplemented by smaller residential settlements and commercial areas. In the last twenty years, the occupation of land is growing up to almost 100 hectares, even though the population that lives in the Municipality of Benevento is in decline (59,031 inhabitants in 2019 with a reduction of 6.4% in 15 years).

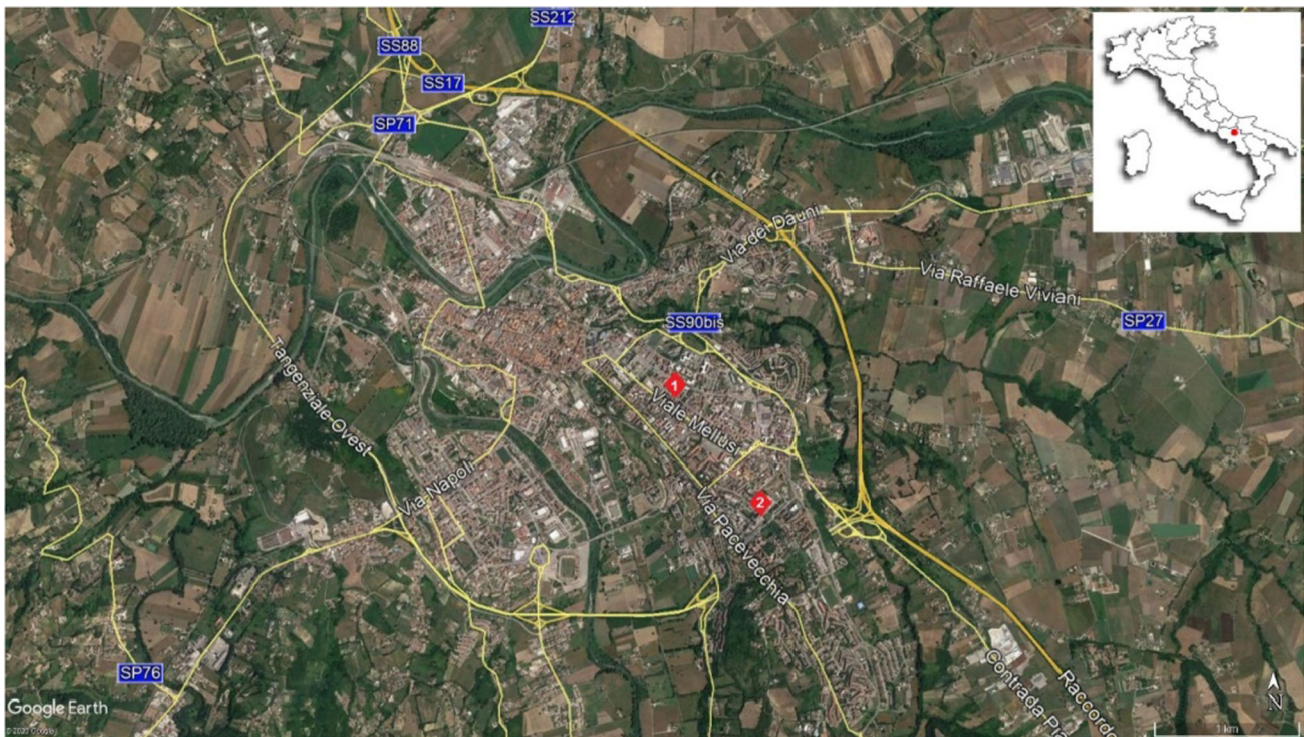


Fig. 1 An aerial photo of the city of Benevento inland of Campania Region. The air monitoring stations (BN1 and BN2) are marked in red. The Google Earth image also reports the main roads crossing or passing the urban area

This situation would point to a heavy increase in vehicular traffic favoured by continuous travel to and from the city and therefore a possible increase in contamination.

The decrease in the population of Benevento is partially due to the negative balance between births and deaths, and also to the migration flow to other municipalities. In particular, we note that the difference between births and deaths is not due only to a decrease in the birth rate, but to a significant increase in deaths. Among these deaths, those due to neoplasms are on the rise (cancer register period 2010–2015), and 24% of these are related to the lung. Obviously, these cases are not all attributable to air pollution problems, but there may be a certain correlation (Tuttitalia 2018; Registro Tumori Regione Campania 2021).

3 Data description

To develop this research, data from the air quality control system, established by European Decisions and adopted by Italian National Ministerial Laws, are used. This legislative framework involves not only public administrations but also environmental agencies and public research bodies. This consistent involvement aims for a high homogeneity and comparability in the assessment and management of air quality in the national territory. Therefore, the control

systems not only provide information to verify compliance with the regulatory limits but also highlight and inform about the general state of the air quality of a territory.

In this case, our research consider the measurement stations (reported as BN1 and BN2) present in the urban area of Benevento belonging to the Monitoring Network of Campania and managed by the Regional Agency of Environment Protection (ARPAC). As the regulations underline, they are mainly influenced by emissions from nearby roads. Their location is to be related to areas characterized by a considerable concentration of pollutants. Such stations can monitor many pollutants, i.e. NO_2 , CO, PM_{10} , $\text{PM}_{2.5}$, O_3 and Benzene (C_6H_6).

To highlight the consistency of the data of the Benevento stations some of the trends in the period 2011–2019 are reported (Fig. 2). For example, the PM_{10} that would derive from the resuspension of inert dust from construction sites, uncovered areas and road surfaces, as well as from aggregates of unburnt particles from combustion plants and vehicle engines, was able to be measured in Benevento stations. The measured values reveal a significant decrease in the number of exceedances with respect to the regulation, that is $50 \mu\text{g}/\text{m}^3$. Moreover, in the years 2011–2016 overruns have exceeded the number of 35 planned for a year. It is important to underline that the harmfulness to human health does not depend only on the concentration of this particulate, but also on the chemical

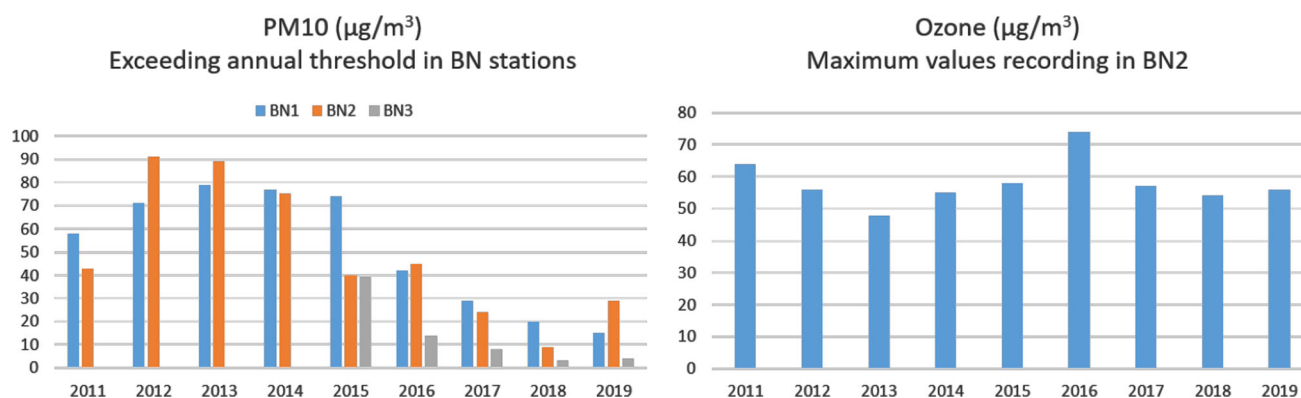


Fig. 2 Trends of PM_{10} and O_3 in Benevento stations from 2011 to 2019

composition and the size of the particles. For instance, those with a diameter between 5 and 10 μm reach the trachea and bronchi, while those with a diameter of less than 5 μm can penetrate the pulmonary alveoli.

Values above the limit of ozone, a triatomic oxygen molecule, were recorded only in the most peripheral station in Benevento. O_3 is a highly reactive gas, with a pungent odour and a high oxidizing power. It is generated starting from the action of solar radiation mainly on nitrogen dioxide molecules and subordinately on reactive hydrocarbons present in the atmosphere; consequently, ozone is a typical secondary pollutant more frequent in periods of greater insolation. The trend in Fig. 2 shows a cyclicity in the maximum values ($74 \mu\text{g}/\text{m}^3$ in July 2016 at 3 pm and $58 \mu\text{g}/\text{m}^3$ at the beginning of August 2015 again at 3 pm) and therefore of exceedances (25 in 2017), probably to be related with the variations of solar radiation measured in the spring and summer periods in the years under consideration. Ozone is particularly irritating to the respiratory tract and eyes. In addition, it causes lesions on the leaves of some plants and a reduction in elasticity on rubber and textile fibres. It should also be noted that the concentrations of very fine particles ($\text{PM}_{2.5}$), the measurement of which has become increasingly refined, increased to reach values higher than $60 \mu\text{g}/\text{m}^3$ in July 2017.

The air data used for this research are those from the period 2012–2014 and are acquired by the Campania regional site. The choice of the period derives from the possibility of correlating the pollutant data with those of the possible effects on human health. Therefore, we opt for the period 2012–2014 which could have consequences on health in the medium term, i.e. in the years 2016–2018.

So from the two control units BN1 and BN2, during the period from January 2012 to December 2014, we take into consideration the daily values of PM_{10} , $\text{PM}_{2.5}$, NO_2 , O_3 , Benzene, and CO. Then, for each month, we compute:

- For PM_{10} and NO_2 , the monthly average of the daily maximum values (M.)

- For PM_{10} and NO_2 and O_3 the number of exceedances (Exc.) of the permitted thresholds,
- For $\text{PM}_{2.5}$ the monthly average value (Avg.),
- And for O_3 the monthly maximum value (Max.).

From these data, on the basis of the availability on BN1 and BN2, we select the inputs characteristics, as reported in Table 1.

About the cancer data, they are directly provided by the Benevento Hospitals, namely the Sacro Cuore di Gesù Fatebenefratelli Hospital, and A.O. “G. Rummo”, since, at the time of study, the Benevento cancer registry is still unable to cover the period from January 2016 to December 2018 that we want to correlate.

Table 1 Characteristics and their unit of measure (Air Data from January 2012 to December 2014)

Characteristic	Unit of measure
M. PM_{10} BN1	$\mu\text{g}/\text{m}^3$
M. PM_{10} BN2	$\mu\text{g}/\text{m}^3$
Exc. PM_{10} BN1	Integer in (0:31)
Exc. PM_{10} BN2	Integer in (0:31)
Avg. $\text{PM}_{2.5}$	$\mu\text{g}/\text{m}^3$
M. NO_2 BN1	$\mu\text{g}/\text{m}^3$
M. NO_2 BN2	$\mu\text{g}/\text{m}^3$
Exc. NO_2 BN1	Integer in (0:31)
Exc. NO_2 BN2	Integer in (0:31)
M.CO BN2	$\mu\text{g}/\text{m}^3$
Exc.CO BN2	$\mu\text{g}/\text{m}^3$
Max. O_3 BN2	$\mu\text{g}/\text{m}^3$
Exc. O_3 BN2	Integer in (0:31)
Benzene	$\mu\text{g}/\text{m}^3$

Legend: M: average of maximum values; Exc.: exceedance of threshold; Avg.: average of values; Max: the maximum value registered; BN1 and BN2: monitoring stations

Finally, from the collected data we build and use for the analysis a set of samples made up of 36 feature vectors, each one labelled by the number of tumours 4 years later:

$$\begin{aligned}
 &X(\text{month, year}); T(\text{month, year} + 4) \\
 &= (\text{M.PM}_{10}\text{BN1, M.PM}_{10}\text{BN2, Exc.PM}_{10}\text{BN1, Exc.PM}_{10}\text{BN2,} \\
 &\quad \text{Avg.PM}_{2.5, \text{M.NO}_2\text{BN1, M.NO}_2\text{BN2, Exc.NO}_2\text{BN1,} \\
 &\quad \text{Exc.NO}_2\text{BN2,} \\
 &\quad \text{M.CO BN2, Max.O}_3\text{BN2, Exc.O}_3\text{BN2, Benzene})(\text{Tumours})
 \end{aligned}
 \tag{1}$$

4 Processing techniques

To highlight the correlation between the cases of lung tumours and the levels of air pollution two soft computing methods are applied. Such techniques are Artificial Neural Networks (ANNs) and Genetic Programming (GP). As ANN, we use a feedforward Multi-Layer Perceptron (MLP) Neural Network, trained by the backpropagation procedure (Beale and Jackson, 1990).

The MLP ANN is made by processing elements (neurons) organized in layers. The number of hidden layers and the number of neurons in each hidden layer is determined by a pruning-and-growing methodology, starting from an initial random choice (Rampone and Valente, 2012). The resulting ANN consists of 14 input, two hidden layers, made up of 2 and 4 neurons, respectively, and one output layer, of 1 neuron. All the connections between couples of neurons have associated weights W_t that are modified during the training phase to produce the correct output. The starting weights W_0 are randomly chosen in the range $(- 0.9;0.9)$. The learning rate, i.e. the measure of the influence degree of the actual error in the W_t weights updating, and the momentum term, determining the influence of the history of weight W_t changes, are determined by a trials-and-errors methodology as 0.9 and 0, respectively. The number of training cycles is fixed to 500. The ANN resulting parameter data are reported in Table 2.

We also employ GP to generate and evolve explicit functions, representing the acquired knowledge. In the GP

experiments, we are looking for a formula $z = f(x,y,\dots)$ that satisfies

$$\begin{aligned}
 \text{Tumors} = &f(\text{M.PM}_{10}\text{BN1, M.PM}_{10}\text{BN2, Exc.PM}_{10}\text{BN1,} \\
 &\text{Exc.PM}_{10}\text{BN2, Avg.PM}_{2.5, \text{M.NO}_2\text{BN1, M.NO}_2\text{BN2,} \\
 &\text{Exc.NO}_2\text{BN1, Exc.NO}_2\text{BN2, M.CO BN2,} \\
 &\text{Max.O}_3\text{BN2, Exc.O}_3\text{BN2, Benzene})
 \end{aligned}
 \tag{2}$$

GP relies on a set of component functions. Here we use the basic arithmetic operators $+$, $-$, $*$, $/$, the trigonometric functions sine, cosine, tangent and hyperbolic tangent and their corresponding inverse functions, the exponential and the natural logarithm, the logistic function, and the gauss function.

The fitness measure, representing the measure of function quality, is the Absolute Error.

4.1 MLP results

A MLP is usually trained by a subset of available data (training set), and the generalization ability is validated on a different subset (validation set). However, the subset choices may introduce biases (Dietterich, 1998; Nadeau and Bengio, 2003), and we instead apply the standard 10-Fold-Cross-Validation (10-CV) statistical method (Devijver and Kittler, 1982), overcoming the limits of the traditional training / validation approach, at least in part (Fushiki, 2011). According to the 10-CV, the dataset is splitted into 10 distinct groups of equal size, and each unique group is used as a hold out or validation set, taking the remaining groups as training data. In this way 10 independent experiments are performed, one for each validation set choice.

As performance indicators, we value, independently, the Error percentage, the Standard deviation, and the Correlation coefficient, resulting from each experiment in the 10-CV application. Then we value the resulting average performances.

The experiments are performed on the dataset described in Sect. 3, by using a neural network Excel-based

Table 2 Configuration and parameters of the MLP neural network and the backpropagation procedure

MLP parameters			
Number of Characteristics	13	Number of Outputs	1
Number of hidden layers	2	Hidden layer 1 size	2
		Hidden layer 2 size	4
Learning rate	0.9	Initial W_t range	$- 0.9;0.9$
Momentum term	0		
Labelled feature vectors	36	Training cycles	500
Partition data into Training/Validation set	10-CV	Training mode	Sequential

Table 3 MPL Results of tenfold cross-validation application. Each row reports the error percentage, standard deviation, and correlation coefficient in each experiment. The last line shows the averages of these values

Experiment	Error percentage	Standard deviation	Correlation coefficient
1	1.62%	0.008	0.827
2	1.87%	0.012	0.839
3	1.82%	0.013	0.841
4	1.79%	0.010	0.841
5	3.38%	0.032	0.762
6	1.24%	0.004	0.946
7	0.31%	0.003	0.997
8	0.31%	0.001	0.997
9	0.31%	0.003	0.997
10	0.32%	0.001	0.997
Mean	1.30%	0.009	0.904

simulation environment developed by Angshuman Saha (Saha, 2001).

The experiment data are reported in Table 3. Globally they result in a mean error percentage lower than 2% and in a mean correlation coefficient greater than 0.90. Figure 3 shows a comparative graph of observed and predicted tumour values in the ten MLP experiments.

4.2 GP results

The GP experiments are performed by the Eureqa genetic programming software tool (Schmidt and Lipson, 2009). As for performance indicators, we use both the error percentage and the correlation coefficient. Since MLP highlights the existence of a correlation between pollutants and tumours, here we are mainly interested in the explicit representation of the factors that globally determine it.

We find the pool of solutions after about 100,000 generations. The best one, with a correlation coefficient of 0.992 and an Error percentage of 1.65, is represented by the following formula:

$$\begin{aligned}
 & \text{(Tumours cases)} \\
 & = 1.555e4 + 36.31 * (\text{M.CO BN2}) \\
 & \quad + 2.799 * (\text{Exc.PM}_{10}\text{BN1}) - 1.675e4 * (\text{Avg.PM}_{2,5}) \\
 & \quad * \sin(5.698 - (\text{M.NO}_2\text{BN2}))/ \\
 & \quad (1.555e4 + 36.31 * (\text{M.CO BN2}) \\
 & \quad + 2.799 * (\text{Exc.PM}_{10}\text{BN1}) - \text{Exc. PM}_{10}\text{BN2})) \\
 & - 1.857 * (\text{Exc. PM}_{10}\text{BN2}) \\
 & - 3.406 * (\text{Exc. O}_3\text{BN2}) - 180.6 * (\text{Exc. NO}_2\text{BN1})
 \end{aligned} \tag{3}$$

We also consider some influence metrics of the characteristic in determining the solution result, evidencing the major role of PM_{10} , NO_2 and O_3 , as reported in Table 4. Details on the influence metrics considered are reported in the Appendix.

We found similar results in almost all the formulae, as evidenced in Fig. 4, that reports the number of formulae each characteristic appears in.

5 Conclusion

As evidenced by the soft computing techniques, it appears that in Benevento there is a direct relationship between air pollution and the occurrence of lung tumours cases. This relationship is highlighted above all for some specific pollutants that lower air quality and can increase the risk of getting sick among the Benevento population. In the absence of strong industrialization in the city area, it is plausible that mainly urban traffic and, subordinately, domestic heating are responsible for the levels of pollution recorded. From the traffic point of view, this area represents an important crossroads between the sides of the Tyrrhenian Sea and the Adriatic Sea, as well as a town that is often congested with traffic related to a significant commute. However, also the meteorological and topographical conditions of Benevento have a significant weight on the formation of secondary pollutants as recorded in the peripheral areas.

From the short/medium time interval considered in this study, it is also evident that long-term high pollution conditions are not necessary to have cases of tumours. It could be said that the (negative) response of the organism is rather rapid.

As shown in Fig. 2, it appears there is a slight improvement in the mean concentrations of pollutants over time. This could be due to the increasing use of engines with lower emissions (hybrid or even electric), but it could also be due to the traffic measures carried out by the city municipal administration.

The recent general improvement in air quality, due to the lack of anthropogenic activity resulting from the compliance with the measures imposed by Covid 19,

Fig. 3 Observed (X-axis) and Predicted (Y-axis) tumours and trend lines in the ten MLP experiments

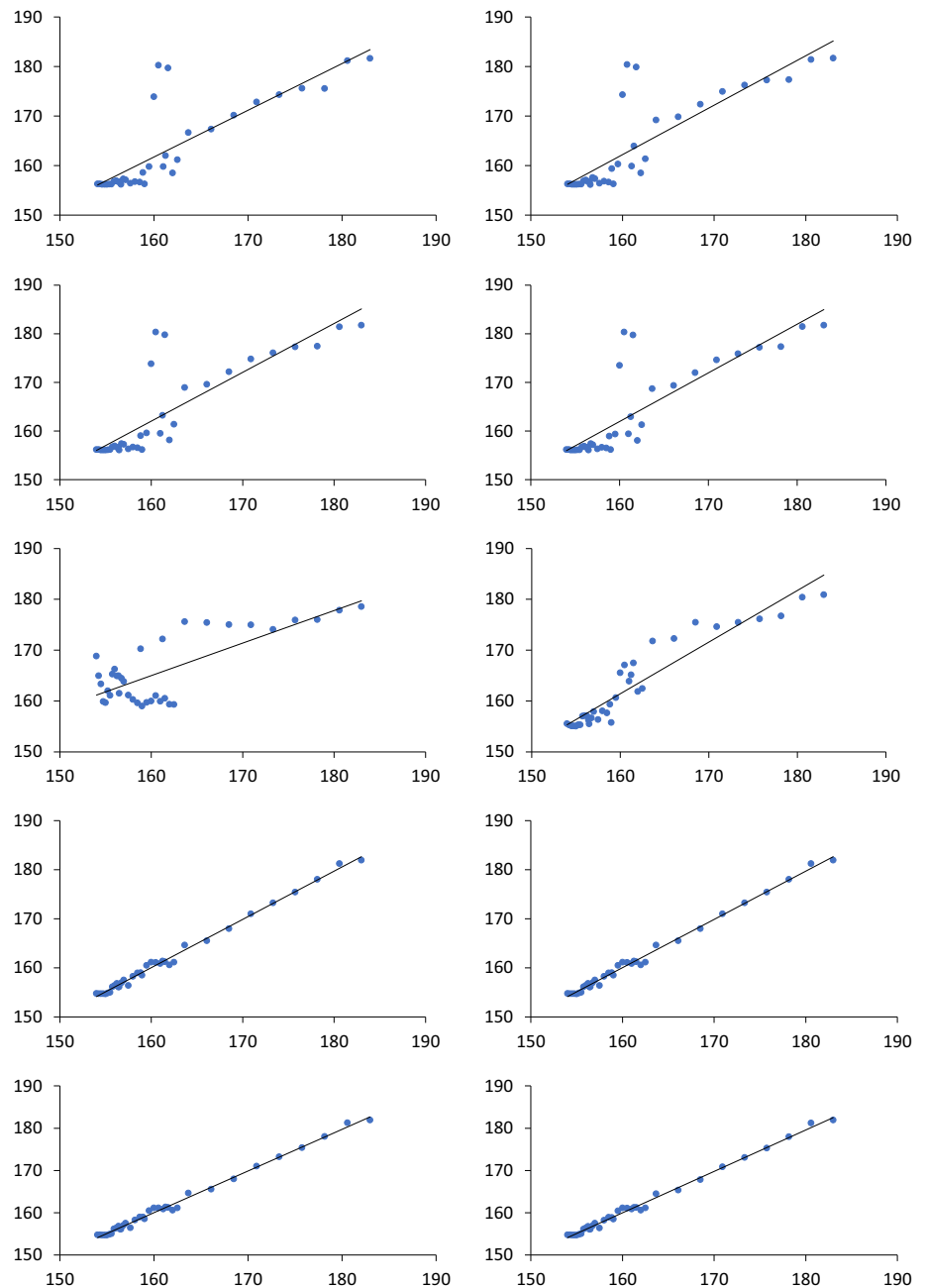


Table 4 Influence metrics of the most relevant characteristic in the resulting formula

Characteristic	Sensitivity	% positive	Positive magnitude	% negative	Negative magnitude
M.NO ₂ BN2	13.909	52%	19.073	48%	84.228
Exc.PM ₁₀ BN1	11.529	100%	11.529	0%	0
Exc.NO ₂ BN1	10.498	0%	0	100%	10.498
Exc.PM ₁₀ BN2	0.95124	0%	0	100%	0.95124
Exc.O ₃ BN2	0.16635	0%	0	100%	0.16635
M.CO BN2	0.12247	100%	0.12247	0%	0
M.PM _{2.5}	0.021651	0%	0	100%	0.021651

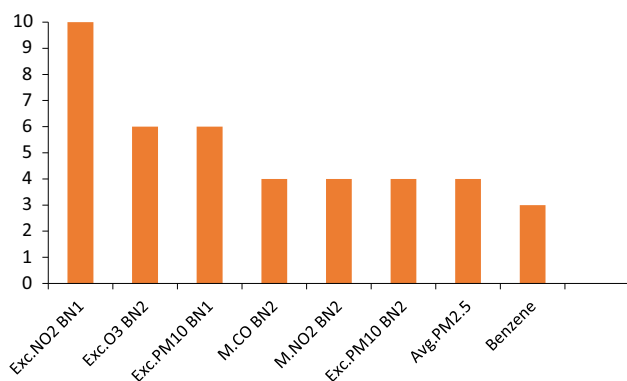


Fig. 4 Number of formulae (Y-axis) each characteristic (X-axis) appears in

especially in the first half of 2020 (Karuppasamy et al., 2020; Kerimray et al., 2020) will show its effects only after some years. However, we believe it lasted too short to change the trend.

In conclusion, pollutants such as PM_{10} , NO_2 and O_3 pose a health risk that must be considered to avoid an increase in lung cancers.

Beyond the specific assessment on the territory considered, it also seems that the method can be generalized to other urban areas, as a tool for assessing and preventing the impact of environmental pollution and this will be the subject of future work.

Appendix

Given a model equation of the form $z = f(x, y, \dots)$, the influence metrics of x on z are defined as follows:

Sensitivity: $\left| \frac{\partial z}{\partial x} \right| \frac{\sigma(x)}{\sigma(z)}$, evaluated at all input data points.

% Positive: The percent of data points where $\frac{\partial z}{\partial x} > 0$;

% Negative: The number of data points where $\frac{\partial z}{\partial x} < 0$;

Positive magnitude: $\left| \frac{\partial z}{\partial x} \right| \frac{\sigma(x)}{\sigma(z)}$, at all points where $\frac{\partial z}{\partial x} > 0$;

Negative magnitude: $\left| \frac{\partial z}{\partial x} \right| \frac{\sigma(x)}{\sigma(z)}$, at all points where $\frac{\partial z}{\partial x} < 0$; where $\frac{\partial z}{\partial x}$ is the partial derivative of z with respect to x ; $\sigma(x)$ is the standard deviation of x in the input data; $\sigma(z)$ is the standard deviation of z ; $|x|$ denotes the absolute value of x ; \bar{x} denotes the mean of x .

Acknowledgements We wish to thank Dr. Antonio Febraro, of the Sacro Cuore di Gesù Fatebenefratelli Hospital, Benevento, Italy and Dr. Mario Del Donno, of the U.O.C. di Pneumologia-A.O. "G.Rummo", Benevento, Italy for their help in tumour data acquisition, and the 2018-19 EDA course group (A.D'Ercole, F.Iannaccone, M.Lombardi, A.Spagnuolo, D.Caiazzo, A.Bruno, R.Avella, A. Meoli) for the enthusiastic participation.

Declarations

Conflict of interest Authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent In this study, only aggregated and anonymous data were used without direct connection with specific patients.

References

- Beale R, Jackson T (1990) Neural computing: an introduction. Taylor & Francis, New York, p 256
- Bishop CM (1996) Neural networks for pattern recognition. Oxford University Press, USA, p 502
- Chen YC, Lei TC, Yao S, Wang HP (2020) PM2.5 prediction model based on combinational hamsterstein recurrent neural networks. Mathematics 8:2178
- Cramer NL (1985) A representation for the adaptive generation of simple sequential programs. In: John J Grefenstette (ed) Proceedings of an international conference on genetic algorithms and the applications, Carnegie Mellon University. Pittsburg, PA, USA, pp 183–187
- D'Alelio D, Rampone S, Cusano LM, Morfino V, Russo L, Sanseverino N, Cloern JE, Lomas MW (2020) Machine learning identifies a strong association between warming and reduced primary productivity in an oligotrophic ocean gyre. Sci Rep 10:3287. <https://doi.org/10.1038/s41598-020-59989-y>
- Devijver PA, Kittler J (1982) Pattern recognition: a statistical approach. Prentice-Hall, London
- Dietterich TG (1998) Approximate statistical tests for comparing supervised classification learning algorithms. Neural Comput 10:1895–1923
- Elia S, D'Angelo G, Palmieri F, Sorge R, Massoud R, Cortese C, Hardavella G, De Stefano A (2020) A machine learning evolutionary algorithm-based formula to assess tumor markers and predict lung cancer in cytologically negative pleural effusions. Soft Comput 24:7281–7293
- European Environment Agency (2018) Air quality in Europe - 2018 report EEA Report No 12/2018 ISSN 1977–8449. doi: <https://doi.org/10.2800/777411>
- Fushiki T (2011) Estimation of prediction error by using K-fold cross-validation. Stat Comput 21:137–146
- Haykin S (2008) Neural networks and learning machines. Prentice Hall, London
- Karuppasamy MB, Seshachalam S, Natesan U, Ayyamperumal R, Karuppappan S, Gopalakrishnan G, Nazir N (2020) Air pollution improvement and mortality rate during COVID-19 pandemic in India: global intersectional study. Air Qual Atmos Health 13:1375–1384
- Kerimray A, Baimatova N, Ibragimova OP, Bukenov B, Kenessov B, Plotitsyn P, Karaca F (2020) Assessing air quality changes in large cities during COVID-19 lockdowns: the impacts of traffic-free urban conditions in Almaty, Kazakhstan. Sci Total Environ 730:139179
- Koza JR (1992) Genetic programming: on the programming of computers by means of natural selection. MIT Press, Cambridge, MA, USA, p 819

- Nadeau C, Bengio Y (2003) Inference for the generalization error. *Mach Learn* 52:239–281
- Raaschou-Nielsen O et al (2013) Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European study of cohorts for air pollution effects. *Lancet Oncol* 14(9):813–822
- Rampone S, Pagliarulo C, Marena C, Orsillo A, Iannaccone M, Trionfo C, Sateriale D, Paolucci M (2021) In silico analysis of the antimicrobial activity of phytochemicals: towards a technological breakthrough. *Comput Methods Programs Biomed* 200:105820
- Rampone S, Valente A (2012) Neural network aided evaluation of landslide susceptibility in Southern Italy. *Int J Mod Phys C* 23(01):1250002
- Rampone S, Valente A (2017) Prediction of seasonal temperature using soft computing techniques: application in Benevento (Southern Italy) area. *J Ambient Intell Human Comput* 8(1):147–154
- Rampone S, Valente A (2019) Assessment of desertification vulnerability using soft computing methods. *J Ambient Intell Human Comput* 10(2):701–707
- Registro Tumori Regione Campania (2021) <http://www.regione.campania.it/regione/it/tematiche/registro-regionale-tumori>. Last accessed June 2021
- Saha A (2001) NNPRED simulation environment (available on-line at <http://xoomer.virgilio.it/srampone/NNpred01.zip>)
- Schmidt M, Lipson H (2009) Distilling free-form natural laws from experimental data. *Science* 324(5923):81–85
- Seigneur C (2019) Air pollution. Cambridge University Press, Cambridge, p 370
- Shiri J, Kişi O, Landeras G, Lopez JJ, Nazemi AH, Stuyt LCPM (2012) Daily reference evapotranspiration modeling by using genetic programming approach in the Basque Country (Northern Spain). *J Hydrol* 414–415:302–316
- Sicard P, Agathokleous E, De Marco A, Paoletti E, Calatay V (2021) Urban population exposure to air pollution in Europe over the last decades. *Environ Sci Eur* 33:28. <https://doi.org/10.1186/s12302-020-00450-2>
- Stanislawska K, Krawiec K, Kundzewicz ZW (2012) Modelling global temperature changes with genetic programming. *Comput Math Appl* 64:3717–3728
- Tuttitalia (2018) Guida ai Comuni, alle Province ed alle Regioni d'Italia. <https://www.tuttitalia.it/>. Last accessed Jan 2018
- WHO (2016) Ambient air pollution: a global assessment of exposure and burden of disease, World Health Organization, Geneva. https://www.who.int/gho/phe/outdoor_air_pollution/en/. Last accessed Sept 2020

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