Application of neural network to simulate the behavior of hospitalizations and their costs under the effects of various polluting gases in the city of São Paulo

Amanda Carvalho Miranda[1,2] · José Carlos Curvelo Santana[2,3] · Charles Lincoln Kenji Yamamura[2] · Jorge Marcos Rosa[4,5] · Elias Basile Tambourgi[5] · Linda Lee Ho[2] · Fernando Tobal Berssaneti[2]

Abstract

This work aims to obtain an artificial neural network to simulate hospitalizations for respiratory diseases influenced by pollutant gaseous such as CO, PM_{10}, $PM_{2.5}$, NO_2, O_3, and SO_2 emitted from 2011 to 2017, in the city of São Paulo. The hospitalization costs were also be calculated. MLP and RBF neural networks have been tested by varying the number of neurons in the hidden layer and the type of equation of the output function. The following pollutants and its concentration range were collected considering the supervision of Alto Tiete station set, in several neighborhoods in the city of São Paulo, from in the period 2011 to 2017: 28–63 $\mu g/m^3$ of $PM_{2.5}$, 52–110 $\mu g/m^3$ of PM_{10}, 49–135 $\mu g/m^3$ of O_3, 0.8–2.6 ppm CO, 41–98 $\mu g/m^3$ of NO_2, and 3–16 $\mu g/m^3$ of SO_2. Results showed that a RBF neural network with 6 input neurons, 13 hidden layer neurons, and 1 output neuron, using BFGS algorithm and a Gaussian function to neuronal activation, was the best fitted to the experimental datasets. So, knowing the monthly concentration of gaseous pollutions was possible to predict the hospitalization of 1464 to 3483 ± 510 patients, with costs between 570,447 and $1,357,151 \pm 198,171$ USD per month. This way, it is possible to use this neural network to predict the costs of hospitalizing patients for respiratory diseases and to contribute to the decision-making of how much the government should spend on health care.

Keywords Gaseous pollutants · Artificial neural network · Health care · Hospitalization · Health costs · Simulation

✉ Amanda Carvalho Miranda
amandafarma25@gmail.com

1 Department of Pharmaceutical Sciences, Nine July University, São Paulo, Brazil

2 Department of Production Engineering, Polytechnic School, University of São Paulo, Av. Prof. Luciano Gualberto, 1380, Butantã, São Paulo, SP 05508-010, Brazil

3 Department of Management Engineering, Federal University of ABC, University Mall, São Bernardo do Campo 09606-045, Brazil

4 School of Textile Technology of SENAI, Rua Correia de Andrade, 232 Brás, São Paulo, SP 03008-020, Brazil

5 School of Chemical Engineering, State University of Campinas, Barão Geraldo, Campinas, SP 13083-970, Brazil

Introduction

The effects of air pollution on people have been the subject of research around the world to correlate the photochemical effects of air and health, the respiratory system, and the aggravation of allergic diseases. According to Arbex et al. (2012), the individuals most susceptible to diseases caused by pollutant emissions are children, the elderly, people with chronic diseases, and people with genetic susceptibility. In addition, pollutants can affect the human fetus during pregnancy by causing intrauterine growth retardation, prematurity, low birth weight, and—in the most severe cases—congenital anomalies and intrauterine or perinatal death.

Researchers from different countries have been united in their quest to correlate the effects of air pollutants with respiratory diseases. In this way, they had identified an increase in the incidence of asthma in adults living in high-traffic Switzerland regions (Kunzli et al. 2010), which was also associated the rise in $PM_{2.5}$ levels, in the Netherlands (Gehring et al. 2010), and in addition, the combination of

PM_{10} and NO_2 emissions induced premature deaths due to cardiovascular and respiratory diseases, in China and Italy (De Marco et al. 2019; Yu et al. 2016). When the $PM_{2.5}$ concentration increased 10 μg/m³, an increase in the lung cancer incidence in the USA (Reis et al. 2018) and an increment of 3.9 mm Hg in the average systolic 24-h blood pressure for hypertensive and/or diabetic workers in Brazil (Santos et al. 2019) were observed. The increased incidence of hospitalization of patients with chronic obstructive pulmonary disease (COPD) has been associated to the environmental pollutants such as SO_2, NO_2, $PM_{2.5}$, and PM_{10}, in Denmark, Hong Kong, and Korea (Andersen et al. 2011; Hu et al. 2010; Ko et al. 2007). In the Americas, more than 131,000 people in low-income countries and 96,000 people in high-income countries are estimated to die each year due to air pollution-related diseases. The elderly, children, and patients with pre-existing chronic diseases are the groups most at risk (WHO 2019).

Thus, it was concluded that pollutant emissions negatively affected the quality of life of individuals, but other researches have also associated this to an economic setback for the public health system. Ravina et al. (2018) estimated an average cost of hospitalizations for respiratory diseases at 2.5 million euros per year in the city of Turin, Italy. Gao et al. (2018) reported that the world suffered economic losses of 129 billion USD in association with greenhouse gas (GHG) emissions in 2016.

In Brazil, an average of 22,000 people is estimated to lose their lives prematurely each year because of exposure to pollutants outside the home, especially in the urban environment. According to the World Health Organization (WHO 2019), that could reach 36,000 people/year by 2040. In recent decades, studies on air pollution and its effects on human health in São Paulo have provided considerable evidence of an association between increased hospital admissions for illness in individuals exposed to atmospheric pollutants such as O_3, NO_2, SO_2, CO, and inhalable PM (Andrade et al., 2017). According to Bravo et al. (2016), more than 99,000 deaths are attributed to air pollution every year. The annual cost of immobility in São Paulo is equivalent to 7.5% of the city's gross domestic product (GDP), which is a significant impact on residents' health (Santana et al. 2020).

According to Pinheiro et al. (2014), Brazil heath system (SUS) presented a cost with air pollution cost more than 2 million USD from 1993 to 1995. The money was spent to treat patients who developed diseases directly related to excess pollutants. This amount of money could be used to perform 784,000 medical consultations or 10,100 normal births in SUS-affiliated hospitals. Mantovani et al. (2016) reported that in São José do Rio Preto, São Paulo, Brazil, the number of pollutant-related respiratory disease hospitalizations was of the order of 650 in 3 years, which represented

an increase in SUS spending of approximately US\$ 50,000. Paiva (2014) estimated that in Volta Redonda, Rio de Janeiro, Brazil, the annual cost of hospitalizations owing to pollutant exposure between 2005 and 2007 was US\$ 44,000.

This effect was also verified by Abe and Miraglia (2016), who reported that the reduction in the levels of PM_{10}, $PM_{2.5}$, and O_3 could save lives and an impressive amount of money in a country where economic resources are scarce. Moreover, reduced levels of air pollution would also prevent the demand for hospital inpatient rooms, since there would be diminished hospitalizations.

There is new knowledge that the gaseous pollutants affect the human health and impact the health systems; however, this assessment depends strongly on air pollution epidemiology and exposure science. Mathematical models to describe the dispersion and precipitation (below-cloud) of gaseous particles in the atmosphere under the influence of wind speed, temperature, pressure, and air humidity are very common (Duhanyan and Roustan 2011; Xu et al. 2019; Zhao et al. 2015). However, there is no model that makes these direct links between costs and the number of inpatients due to gaseous pollutant emissions.

However, the recent advances in researches have done the cross-linkage of atmospheric, epidemiology, and mathematic sciences, allowing analysts to quantify an increasing number of health outcomes in far greater detail than was previously possible (Ravina et al. 2018).

Yu et al. (2016) used remote sensing for observation to assessment in real time the spatial variability of air pollutants. This has an advantage over in situ data in terms of determining the impact of air pollution on human health because it has better spatial coverage and thus provides better understanding of the spatial distribution of the human heath impact. They used an aggregate risk index to account for the overall impact on human health of exposure to multiple air pollutants and to reflect the linear relationship between air pollution and health risks. Their results showed that Chinese areas with the highest risk due to air pollution are mainly located in the Taklimakan Desert, northern China, Sichuan Basin, and middle of Inner Mongolia.

Miranda et al. (2017) have been developed a statistical model based in a pollutant emission dataset. They found that almost 10,000 deaths/year are associated with $PM_{2.5}$; the population segments most at risk are the elderly and children.

In Italy, a model that associates costs with the treatment of diseases caused by air pollution has been developed. This tool is called of DIDEM model and was developed in Matlab® software. It associates the WHO standards with the California puff model (CALPUFF) and the dispersion and external model developed by the Polytechnic of Turin. The DIDEM model is initially powered by the simulations of pollutants performed by CALPUFF model to the

concentration-exposure–response functions according to the WHO recommendations, so that the impacts and costs according with Italy reality are presented (Ravina et al. 2018).

Polezer et al. (2018) had developed an artificial neural network (ANN) to predict the influence of $MP_{2.5}$ on hospitalizations for respiratory diseases. They had been testing multilayer perceptron (MLP), extreme learning machines (ELM), and echo state networks (ESN), to assess the influence of $PM_{2.5}$, temperature, and relative humidity on hospitalizations in Curitiba city from Brazil. They pointed that neural networks can contribute to epidemiological studies, making it possible to assess impacts even when traditional statistical regressions models do not fit well to the air pollution dataset.

However, there is not register on applying of an ANN to simulation of hospitalization and its costs on influence of gaseous emissions. Thus, this work aims to obtain an artificial neural network to simulate hospitalizations for respiratory diseases influenced by CO, PM_{10}, $PM_{2.5}$, NO_2, O_3, and SO_2 gaseous emissions at the Alto Tiete station set, in the city of São Paulo, in the period from 2011 to 2017. The costs will also be calculated associated with hospitalizations. Multilayer perceptron (MLP) and radial bases function (RBF) neural networks will be tested by varying the number of neurons in the hidden layer and the type of equation of the output function.

Methodology

Dataset collection

Data on the number of hospitalizations for respiratory diseases and their respective costs in hospital units for São Paulo were collected from the TABNET-SUS online platform with the following filters: SUS hospital morbidity, city selection, and time range. Thus, data could be collected on the number of hospitalizations and amount spent by the Brazilian government in public or private hospitals each month from 2011 until 2017. The costs and number of monthly hospitalizations that occurred due to respiratory diseases in São Paulo hospitals were quantified and correlated with pollutant emissions in the region during this period (Berssaneti et al. 2014; Saut et al. 2017; Santana et al. 2020).

The São Paulo State Environmental Company (CETESB 2019) is a government agency responsible for the control, supervision, monitoring, and licensing of pollution-generating activities in this area with the fundamental concern of preserving and restoring the water quality, air, and soil. Research was conducted on the CETESB (2019) website, under the link Qualar (air quality) (http://ar.cetesb.sp.gov. br/padroes-de-quality-do-ar/). Gaseous pollutants were

collected at the Alto Tiete station set, in the city of São Paulo from 2011 to 2017, and its concentration (average monthly) had varied between 28 and 63 $\mu g/m^3$ of $PM_{2.5}$, 52 and 110 $\mu g/m^3$ of PM_{10}, 49 and 135 $\mu g/m^3$ of O_3, 0.8 and 2.6 ppm CO, 41 and 98 $\mu g/m^3$ of NO_2, and 3 and 16 $\mu g/m^3$ of SO_2. CETESB has several sensors distributed in various parts of São Paulo that continuously monitor the air quality and send it to its central line. Thus, the pollutant emission ratio was collected for this period at the Alto Tiete station set. This central line corresponds to a set of pollutant gas measurement stations that are distributed among the cities of the metropolitan region of São Paulo. Alto Tiete station set has 27 stations, but only the data from the 15 stations corresponding to the neighborhoods of the city of São Paulo were selected in this work. CETESB used beta radiation to determine $PM_{2.5}$ and PM_{10}, pulse fluorescence to determine SO_2, chemiluminescence to determine NOx, non-expendable infrared to determine CO, and the ultraviolet method to determine O_3. The air pollutants emitted monthly with the most impact on the population's health were monitored according to the relationship presented by Arbex et al. (2012) and Santana et al. (2020), as cited by the WHO (2019) and CONAMA (2019).

Calculation of hospitalization cost

The rate of change of the dollar was used at the time of the last data collection (1 US $ = R $ 3.80), as there was no increase in the amounts paid by the Brazilian government to hospitals, so they are closest to the real values (SUS 2019). From 2011 to 2017, the average amount paid to hospitals for hospitalization for respiratory diseases was US$ 389.65 ± 74.04 per patients. Total hospitalization costs were calculated using Eq. 1 (Santana et al. 2020; Severo Jr et al. 2007a; Rosa et al. 2014).

$$E = 100 * (1 - Predict/RealValue) \tag{1}$$

Since the standard deviation of the amount paid for hospitalization was 19%, this is high and will increase the error in cost forecasts. The variation in the price paid for hospitalization is due to the number of days the patient was hospitalized and medication consumption by him (Benvenga et al. 2016; Miranda et al. 2018; Severo Jr et al. 2007a).

Artificial neural network

In this work, we tested the radial bases function (RBF) and multilayer perceptron artificial (MLP) neural networks contained in the Software Statistica 10® data mining package in the automated neural network tab. In a RBF, each neuron in the hidden layer uses a nonlinear radial basis function as activation function. This layer performs a

nonlinear transformation of the input. In this sense, the output layer acts as a linear combiner that maps the nonlinearity to a new space. The bias of the output layer neurons can be modeled by an additional signal in the previous one, which has a constant activation function (Du and Swamy 2014), according to Fig. 1.

An RBF used a radial bases function as a function activation of its neurons. Through a linear combination of radial base functions of the inputs and parameters of the neurons, the outputs are generated (Menita et al. 2012; Santana et al. 2010).

The multilayer perceptron (MLP) is considered one of the most important architectures of ANN. This method presents a set of information processors units known as artificial neurons, which form the input layer, one or more hidden layers, and an output layer, as shown in Fig. 2 (Haykin et al. 2009).

The information flows from the input layer, passes through the intermediate layers, and closes in the output layer, generating an output response. In a MLP, the disjoint layers are connected, whereas neurons of the same layer do not communicate. The training, or the adjustment of their weights, is performed in two phases: the first one is the

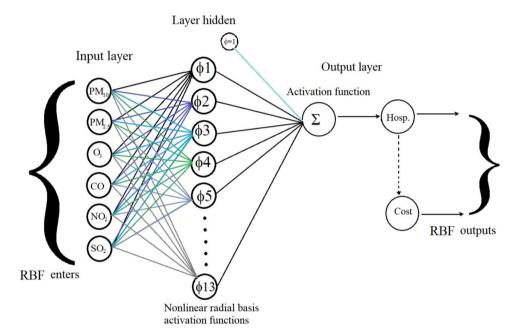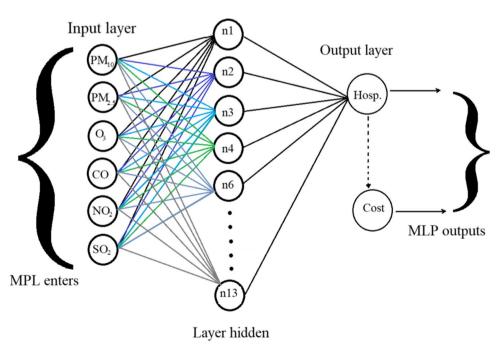**Fig. 1** RBF network. Adapted of Araújo et al. (2020)

Fig. 2 MLP structure. Adapted of Araújo et al. (2020)

forward propagation, in which the signals from a training set sample are inserted as inputs of the network, being propagated layer by layer; in the second phase, the errors are propagated in a recursive manner, and the weights are adjusted through some adjustment rule (Haykin et al. 2009; Siqueira et al. 2014). The most commonly used method to tune the MLP is the steepest descent algorithm, with the derivatives calculated via the backpropagation method. An MLP consists of at least three layers of neuron, being the central neuron(s) called the hidden layer(s) and each node between the neurons has a nonlinear activation function, which are supervised by backpropagation learning technique (Prado et al. 2013; Rosa et al. 2013; Santana et al. 2015; Severo Jr et al. 2007b).

The training of ANN was performed using data obtained from the CETESB (2019) and SUS (2019) data base. For this study, networks were selected running regression, which can be applied in simulation and prediction data. Two types of neural networks were tested: a radial basis network (RBF) and the multilayer perceptron (MLP).

The RBF was configured with 6 input neurons (MP_{10}, $MP_{2.5}$, O_3, CO, SO_2, and NO_2) and 1 output (hospitalization), being used from 12 to 25 neurons in the hidden layer. In this case, the software configuration already selects the radial basis function Gaussian located as network activation function and linear function as the output function (Severo Jr et al. 2007b; Menita et al. 2012; Santana et al. 2010). In this work, the BFGS algorithm (Broyden-Fletcher-Goldfarb-Shanno) was used. In addition, 1,000 seeds were used to control startup, with decay of the weights for the output and hidden layers from 0.0001 to 0.001 and the minimum to maximum (Menita et al. 2012; Santana et al. 2010).

The MLP was configured with the same 6 neurons in the input layer and 1 neuron in the output layer, ranging from 3 to 20 neurons in the hidden layer. Functions as logistic, hyperbolic tangent, sine, and exponential were tested for both activation and in the hidden layer and output neuron (Prado et al. 2013; Rosa et al. 2013; Santana et al. 2015).

Common to all networks tested in this work, the sum of squared error (SSE) was used as error function. To the training of the networks, a random sample of 70% of the experimental data was used. To retraining or validation 15% of the experimental data was employed. Thousand training and 20 retraining were performed. The criterion for choosing the best neural network architecture was its multiple correlations (R^2) in the training, testing, and validation stages. The closer to 1.0 R^2, the neural network was considered to be more adjusted to the experimental data. The two-tailed F test was also used to check for significant differences between the means of the real datasets and those predicted by the neural network. As fitting parameters, R^2 and mean errors were used. The error was calculated according to Eq. 2.

$$E = 100 * (1 - \text{Predict}/\text{RealValue}) \qquad (2)$$

Results and discussion

The neural network that best fit the experimental data was RBF with 6 input neurons, 13 hidden layer neurons, and 1 output neuron, using BFGS algorithm and a Gaussian function to neuronal activation. The multiple correlations (R^2) of the training, testing, and validation steps were 0.7267, 0.7959, and 0.8557, respectively. Predictions should not be made from the overall stipulated intervals (minimum and maximum) for each input neuron. By using these ranges to input the values obtained from the outputs will be found between 1464 and 3483 ± 510 hospitalizations and their costs will be equivalent to 570,447 and $1,357,151 \pm 198,171$ USD per month, respectively.

Figure 3 shows the dispersion graph of the actual data in relation to the predicted value. As noted, there is a high variation in the data, due to a non-causal behavior, which makes mathematical modeling difficult and indicates that the neural network is the most suitable for such random behaviors.

Figure 4 shows the variation of real value of hospitalization with its predicted value. As noted, although most (real) experimental data are close to predicted values, there are several scattered data and these compromised the adjustment of the neural network. This was not opted to delete this data as it cannot be considered an outlier. For this way, the neural network was able to predict 85.57% of the experimental data. However, the value of the F test was low, there are no significant differences between the means of the datasets, and an error below 2% guarantees the accuracy of the predicted results.

This degree of prediction can be considered very good, because the variations in emissions suffer from several uncontroversial influences, such as the number of vehicles circulating in the city of São Paulo, fuel consumption, fuel type, temperature effects, wind, and rainfall. Admissions, on the other hand, depend on the amount of emissions, the contact time with pollutants, people's resistance to disease, the number of people in risk groups exposed to pollutants, and others (Menita et al. 2012; Ravina et al. 2018).

Ten simulations were made for gaseous emissions situations and to test the predictive power of the neural network according to Tables 1 and 2. For this simulation, the values of the monthly emission averages of each pollutant were used individually. In the real field, the actual number of hospitalizations listed in the DATA-SUS was presented and the predictive was estimated by the neural network. The percentage error of the network forecasts ranged from -2 to 25% of the total monthly hospitalizations. These errors are considered high, but satisfactory, because in the environmental and

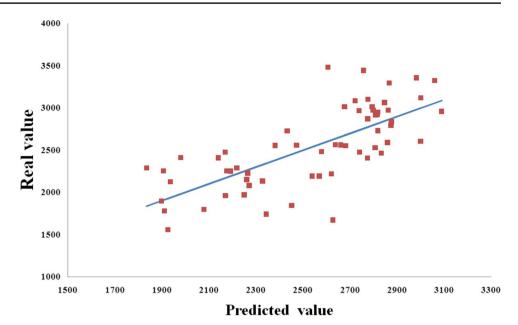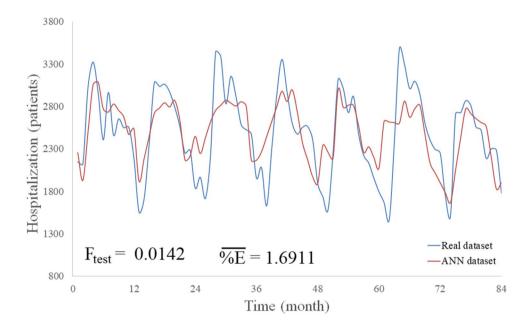**Fig. 3** Dispersion of the actual data with respect to the predicted value

Fig. 4 Comparison between real and predict dataset

$$F_{test} = 0.0142 \qquad \overline{\%E} = 1.6911$$

social fields, 75% accuracy is considered very good (Menita et al. 2012; Ravina et al. 2018; Yu et al. 2016). Costs fluctuated between − 7 and 26%, as there was a combination of errors in hospitalization forecasts and the errors associated with cost variations.

The errors were greater for the costs due to them having been obtained from Eq. 1 (not by the neural network) and it was obtained by the average of all values paid for hospitalizations. The amount that the Brazilian government pays is constant per day of hospitalization; however, the amount presented on the DATA-SUS website does not show the number of days that patients were hospitalized, only the total amount paid for their hospitalization.

Using only the PM$_{2.5}$, the neural network adjustments presented by Polezer et al. (2018) were the same as those obtained in this work. This is due to the seasonality of gaseous emissions data, which is also influenced by air humidity, temperature, and pressure at the measurement site (Xu et al. 2019; Zhao et al. 2015).

In addition, Natali et al. (2011) showed that children and adolescents account for 40% of hospitalizations for respiratory diseases and that these diseases correspond to 30% of total hospitalizations in the city of São Paulo. So, as the reduction of gaseous emissions reduces hospitalizations for respiratory diseases, soon, hospital beds will be left to be used for other purposes.

Table 1 Summary of neural network performance results that best fit experimental data

RBF 6–13-1	MP$_{10}$**	MP$_{2.5}$**	O$_3$**	CO**	SO$_2$**	NO$_2$**	Hosp*	R^2
Minimum (train)	52.00	23.00	49.00	0.70	3.0000	41.00	1556	0.7267
Maximum (train)	110.00	66.00	132.00	2.60	16.0000	98.00	3483	
Mean (train)	71.63	42.83	84.98	1.43	8.5333	70.53	2540	
Standard deviation (train)	14.34	10.24	19.35	0.45	3.3772	13.26	477	
Minimum (test)	55.00	27.00	59.00	0.50	3.0000	54.00	1464	0.7959
Maximum (test)	87.00	54.00	132.00	2.20	11.0000	79.00	3159	
Mean (test)	67.25	37.58	88.83	1.16	6.0000	64.08	2348	
Standard deviation (test)	10.87	9.82	20.77	0.42	2.4863	8.13	484	
Minimum (validation)	52.00	27.00	55.00	0.90	5.0000	50.00	1508	0.8557
Maximum (validation)	91.00	52.00	135.00	1.90	14.0000	87.00	3395	
Mean (validation)	66.17	38.58	88.42	1.37	8.1667	67.83	2352	
Standard deviation (validation)	17.83	13.17	20.30	0.59	5.8439	13.62	473	
Minimum (overall)	52.00	23.00	49.00	0.50	3.0000	41.00	1464	-
Maximum (overall)	110.00	66.00	135.00	2.60	16.0000	98.00	3483	
Mean (overall)	70.23	41.48	86.02	1.38	8.1190	69.23	2486	
Standard deviation (overall)	13.80	10.07	20.03	0.44	3.3054	12.53	510	

*Hosp—Hospitalization.

**Pollutant concentration.

Table 2 Assessing the predictive power of the neural network using real data

PM$_{10}$	PM$_{2.5}$	O$_3$	CO	SO$_2$	NO$_2$	Real *Hosp	Pred *Hosp	%E	Real *Hosp Costs	Pred *Hosp Costs	%E
97.0	60.0	100.0	2.0	20.0	100.0	2779	2721	2	1,060,238	884,651	17
103.0	55.0	112.0	1.9	12.0	92.0	2565	2637	-3	1,027,507	1,156,201	− 13
104.0	61.0	83.0	2.1	12.0	88.0	2915	2809	4	1,094,527	1,490,041	− 36
55.0	38.0	85.0	1.2	6.0	63.0	2304	2259	2	880,219.4	943,121	− 7
86.0	54.0	63.0	2.1	8.0	76.0	2972	2862	4	1,115,178	1,498,772	− 34
58.0	37.0	97.0	0.9	7.0	59.0	2413	1980	18	771,507	1,027,836	− 33
79.0	51.0	116.0	1.3	8.0	86.0	2226	2264	− 2	882,167.6	1,081,285	− 23
77.0	48.0	107.0	1.2	6.0	71.0	3483	2606	25	1,015,428	903,160	11
55.0	23.0	80.0	0.8	3.0	57.0	2408	2138	11	833,071.7	717,993	14
77.0	35.0	73.0	1.2	6.0	69.0	2191	2569	− 17	1,001,011	740,218	26

*Hosp—Hospitalization

*Hosp Costs—Hospitalization Costs

According to Chiquetto et al. (2019) over 2 million of poor people live in areas with pollutant gas levels above that recommended by the WHO. Thus, public policies aimed at improving air quality should be encouraged, in order to decrease air pollution.

For Abhijith et al. (2017) and Janhäll (2015), green walls and roofs on buildings and tree planting and tall vegetation on streets and roads must be used to improve air quality in cities. Thus, investment policies in green urban infrastructure should be adopted to minimize the effects of gaseous pollutants on human health in the city of São Paulo.

According to Ravina et al. (2018), the accounting of effects of air pollution on human health costs is a useful indicator to support decision and information at all management levels. So with the economic advantages of ecological cost accounting presented in this work, profits can be redirected towards the expansion or construction of new hospitals, increasing the capacity for hospital beds in the city of São Paulo (Gao et al. 2018; Santana et al. 2020; Saut et al. 2017). This way, it showed a way in which governments can better plan their investments to minimize hospitalizations for respiratory diseases and increase the number of hospital beds available for other diseases such as COVID-19.

Conclusions

The neural network that best fit the experimental data was RBF with 6 input neurons, 13 hidden layer neurons, and 1 output neuron, using BFGS algorithm and a Gaussian function to neuronal activation. This ANN was able to predict 85.57% of the experimental data and showed error below 2% and there were no significant differences between the means of the datasets by F test.

Within the studied ranges for the gaseous pollutants, it is possible to predict hospitalization of 1464 to 3483 ± 510 patients, with costs between 570,447 and $1,357,151 \pm 198,171$ USD per month, respectively. So, knowing the monthly concentration of gaseous pollutions, the cost of admissions will also be foreseen and this information is a useful indicator to support decision and information at all management levels.

The limitation of this work is the ranges of polluting gases emitted, as the neural network cannot guarantee good results if the input data are outside the range used in this research.

This way, it is possible to use this neural network to predict the costs of hospitalizing patients for respiratory diseases and to contribute to the decision-making of how much the government should spend on health care.

Data availability Data supporting the results of this study are available at https://cetesb.sp.gov.br/ar/qualar/ and http://www2.datasus.gov.br/DATASUS/index.php?area=02.

Declarations

Conflict of interest The authors declare no competing interests.

References

Abe K, Miraglia S (2016) Avaliação do impacto na saúde da poluição do ar em São Paulo, Brasil. Int J Environ Res Public Health 13:694

Abhijith KV, Kumar P, Gallagher J, McNabola A, Baldauf R, Pilla F, Broderick B, Di Sabatino S, Pulvirenti B (2017) Air pollution abatement performances of green infrastructure in open road and built-up street canyon environments – a review. Atmos Environ 162:71–86

Andersen ZJ, Hvidberg M, Jensen SS, Ketzel M, Loft S, Sørensen M, Tjønneland A, Overvad K, Raaschou-Nielsen O (2011) Chronic obstructive pulmonary disease and long-term exposure to traffic-related air pollution: a cohort study. Am J of Respir Crit Med 183:455–461

Araujo LN, Belotti JT, Alves TA, Tadano YS, Siqueira H (2020) Ensemble method based on artificial neural networks to estimate air pollution health risks. Environ Model Softw 123:104567

Arbex MA, Santos UP, Martins LC, Saldiva PHN, Pereira LAA, Braga ALF (2012) Air pollution and the respiratory system. J Bras Pneumol 38(5):643–655

Benvenga MAC, Librantz AFH, Santana JCC, Tambourgi EB (2016) Genetic algorithm applied to study of the economic viability of alcohol production from Cassava root from 2002 to 2013. J Clean Prod 113:483–494

Berssaneti FT, Assumpção A, Nakao OS (2014) Engineering, procurement and construction (EPC): what are the variables that impact the success of the projects currently running in Brazil? Gestao e Producao 21(1):95–109. https://doi.org/10.1590/S0104-530X2014005000009

Bravo MA, Son J, De Freitas CU, Gouveia N, Bell ML (2016) Air pollution and mortality in São Paulo, Brazil: effects of multiple pollutants and analysis of susceptible populations. Journal of Expsure Science Environmental Epidemiology 26(2):150–161

CETESB (2019) Environmental Sanitation Technology Company of São Paulo, Brazil Air quality – Qualar. Available at http://www.cetesb.gov.br. Accessed December

Chiquetto JB, Silva MES, Cabral-Miranda W, Ribeiro FND, Ibarra-Espinosa SA (2019) Padrões de Qualidade do Ar Ynoue, RY e Eventos de Ozônio Extremo na Megacidade de São Paulo. Sustentabilidade 11:3725

CONAMA (2019) National Council of Environment. Políticas Ambientais. Available at:http://www.mma.gov.br/conama. Accessed July

De Marco A, Proietti C, Anav A, Ciancarella C, D'Elia I, Fares S, Fornasier MF, Fusaro L, Gualtieri M, Manes F, Marchetto A, Mircea M, Paoletti E, Piersanti A, Rogora M, Salvati L, Salvatori E, Screpanti A, Vialetto G, Vitale M, Leonardi C (2019) Impacts of air pollution on human and ecosystem health, and implications for the National Emission Ceilings Directive: Insights from Italy. Environ Int 125:320–333

Du K-L, Swamy M (2014) Radial basis function networks. In: Neural Networks and Statistical Learning. Springer, pp. 299–335

Duhanyan N, Roustan Y (2011) Below-cloud scavenging by rain of atmospheric gases and particulates. Atmos Environ 45(39):7201–7217

Gao J, Kovats S, Vardoulakis S, Wilkinson P, Woodward A, Li J, Gu S, Liu X, Wu H, Wang J, Song X, Zhai Y, Zhao J, Liu Q (2018) Public health co-benefits of greenhouse gas emissions reduction: a systematic review. Sci Total Environ 627:388–402

Gehring U, Wijga AH, Brauer M, Fischer P, Jongste JC, Kerkhof M, Oldenwening M, Smit HA, Brunekreef B (2010) Traffic-related air pollution and the development of asthma and allergies during the first 8 years of life. Am J Respir Crit Care Med 181:596–603

Haykin SS, Haykin SS, Haykin SS, Haykin SS (2009) Neural Networks and Learning Machines, vol 3. Pearson Upper, Saddle River, NJ, USA

Hu G, Zhou Y, Tian J, Yao W, Li J, Li B, Ran P (2010) Risk of COPD from exposure to biomass smoke: A metaanalysis. Chem Rev 138:20–31

Janhäll S (2015) Review on urban vegetation and particle air pollution – deposition and dispersion. Atmos Environ 105:130–137

Ko FW, Tam W, Wong TW, Chan DP, Tung AH, Lai CK, Hui DS (2007) Temporal relationship between air pollutants and hospital

admissions for chronic obstructive pulmonary disease in Hong Kong. Thoracic Society 62(9):780–185

Kunzli N, Perez L, Rapp R (2010) Air quality and health. European Respiratory Society, Lausanne, p 66.

Mantovani KCC, Nascimento LFC, Moreira DS, Vargas NP, Vieira LCPFS (2016) Poluentes do ar e internações devido a doenças cardiovasculares em São José do Rio Preto. Brasil Ciênc Saúde Coletiva 21(2):509–516. https://doi.org/10.1590/1413-81232 015212.16102014

Menita PR, Affonso CO, Santana JCC, Sassi RJ, Vanalle RM (2012) Simulation and optimization of a biscuit processing production in an industrial scale by use of MLP and RBF neuro fuzzy network. Advanced Materials Research (online) 601:247–252. https://doi.org/10.4028/www.scientific.net/AMR.601.247

Miranda RM, Lopes F, Rosario NE, Yamasoe MA, Landulfo E, Andrade MF (2017) The relationship between aerosol particles chemical composition and optical properties to identify the biomass burning contribution to fine particles concentration: a case study for Sao Paulo city, Brazil. Environ Monit Assess 189(9–15):1060–1066

Miranda AC, Silva Filho SC, Tambourgi EB, Santana JCC, Vanalle RM, Guerhardt F (2018) Analysis of the costs and logistics of biodiesel production from used cooking oil in the metropolitan region of Campinas (Brazil). Renew Sustain Energy Rev 88:373–379. https://doi.org/10.1016/j.rser.2018.02.028

Natali RMT, dos Santos DSPS, da Fonseca AMC, Filomeno GCM, Figueiredo AHA, Terrivel PM, Massoni CM, Braga ALF (2011) Hospital admissions due to respiratory diseases in children and adolescents of São Paulo city, 2000–2004. Rev Paul Pediatr 29(4):584–590

Paiva RFPS (2014) Hospital morbidity due to diseases associated with air pollution in the city of Volta Redonda, Rio de Janeiro: cases and economic cost. Caderno Saúde Coletiva 22(2):127–133

Pinheiro SLLA, Saldiva PHN, Schwartz J, Zanobetti A (2014) Isolated and synergistic effects of PM10 and average temperature on cardiovascular and respiratory mortality. Rev Saude Publica 12(3):756–765

Polezer G, Tadano Y, Siqueira HV, Godoi AFL, Yamamoto CI, De André PA, Pauliquevis T, Andrade MF, Oliveira A, Saldiva PHN, Taylor PE, Godoi RHM (2018) Assessing the impact of PM 2.5 on respiratory disease using artificial neural networks. Environmental Pollution 235:394–403

Prado KRM, Rosa JM, Alves WAL, Santana JCC, Pereira FH, Tambourgi EB (2013) A bootstrapped neural network model applied to prediction of the biodegradation rate of reactive Black 5 dye. Acta Scientiarum. Technology 35:565–572. https://doi.org/10.4025/actascitechnol.v35i3.16210

Ravina M, Panepinto D, Zanetti MC (2018) DIDEM - an integrated model for comparative health damage costs calculation of air pollution. Atmos Environ 173:81–95

Reis H, Reis C, Sharip A, Reis W, Zhao Y, Sinclair R, Beeson L (2018) Diesel exhaust exposure, its multi-system effects, and the effect of new technology diesel exhaust. Environ Int 114:252–265

Rosa JM, Prado KRM, Alves WAL, Pereira FH, Santana JCC, Tambourgi EB (2013) Applying of a neural network in effluent treatment simulation as a environmental solution for textile industry. Chem Eng Trans 32:73–78. https://doi.org/10.3303/CET1332013

Rosa JM, Tambourgi EB, Santana JCC, Araujo MC, Ming WC, Trindade N (2014) Development of colors with sustainability: a comparative study between dyeing of cotton with reactive and vat dyestuffs. Text Res J 84(10):1009–1017. https://doi.org/10.1177/0040517513517962

Santana JCC, Dias CG, Souza RR, Tambourgi EB (2010) Applying of neural network on the wine sensorial analysis from Barbados Cherry. J Food Process Eng 33(s1):365–378. https://doi.org/10.1111/j.1745-4530.2009.00521.x

Santana JCC, Araújo SA, Biazus JPM, Souza RR (2015) Simulation of biodegradation process of wastewater from meat industry by means of a multilayer perceptron artificial neural network. Ingeniare. Revista Chilena de Ingeniería 23:269–275. https://doi.org/10.4067/S0718-33052015000200011

Santana JCC, Miranda AC, Yamamura CLK, Filho SCS, Tambourgi EB, Ho LL, Berssaneti FT (2020) Effects of air pollution on human health and costs: current situation in São Paulo, Brazil. Sustainability 12(12):Article n 4875-2020. https://doi.org/10.3390/SU12124875

Santos UP, Braga ALF, Garcia MLB, Pereira LAA, Lin CA, Chiarelli PA, de André CAS, de André PA, Singer JM, Saldiva PHN (2019) Exposure to fine particles increases blood pressure of hypertensive outdoor workers: a panel study. Environ Res 174:88–94

Saut AM, Berssaneti FT, Moreno MC (2017) Evaluating the impact of accreditation on Brazilian health care organizations: a quantitative study. Int J Qual Health Care 29(5):713–721. https://doi.org/10.1093/intqhc/mzx094

Severo JB Jr, Almeida SS, Narain N, Souza RR, Santana JCC, Tambourgi EB (2007a) Wine clarification from *Spondiasmombin L.* pulp by hollow fiber membrane system. Process Biochem 42(11):1516–1520. https://doi.org/10.1016/j.procbio.2007.08.003

Severo JB Jr, De Sá Oliveira LS, Sardeiro FS, De Souza RR, Lopes FLG, Santana JCC, Tambourgi EB (2007b) Response surface methodology to evaluation the recovery of amylases by hollow fiber membrane. Braz Arch Biol Tech 50:713–718. https://doi.org/10.1590/S1516-89132007000400017

Siqueira H, Boccato L, Attux R, Lyra C (2014) Unorganized machines for seasonal streamflow series forecasting. Int J Neural Syst 24(3):1430009. https://doi.org/10.1142/S0129065714300095

SUS (2019) Sistema Único De Saúde, Brazilian Health System. Data SUS. Available at: http://portalms.saude.gov.br/sistema-unico-de-saude. Accessed November

WHO, World Health Organization. *Air quality guidelines*. Global update 2005. Particulate matter, ozone, nitrogen dioxide and sulfur dioxide. Copenhagen: World Health Organization; (2005). Available at: http://www.euro.who.int/en/health-topics/environment-and-health/air-quality/publications/pre2009/air-quality-guidelines.-global-update-2005.-particulate-matter,-ozone,-nitrogen-dioxide-and-sulfur-dioxide. Accessed in July 2019.

Xu D, Ge B, Chen X, Sun Y, Cheng N, Li M, Pan X, Ma Z, Pan Y, Wang Z (2019) Multimethod determination of the below-cloud wet scavenging coefficients of aerosols in Beijing. China Atmos Chem Phys 19:15569–15581. https://doi.org/10.5194/acp-19-15569-2019

Yu T, Wang W, Ciren P, Zhu Y (2016) Assessment of human health impact from exposure to multiple air pollutants in China based on satellite observations. Int J Appl Earth Obs Geoinf 52:542–553

Zhao S, Yu Y, He J, Yin D, Wang B (2015) Below-cloud scavenging of aerosol particles by precipitation in a typical valley city, northwestern China. Atmos Environ 102:70–78