Environmental Science and Pollution Research (2021) 28:29701–29709
https://doi.org/10.1007/s11356-021-12658-7

RESEARCH ARTICLE

Feasibility of machine learning methods for predicting hospital emergency room visits for respiratory diseases

Jiaying Lu[1,2,3] · Pengju Bu[4] · Xiaolin Xia[3,5] · Ning Lu[1,3,6] · Ling Yao[1,3,6] · Hou Jiang[1]

Received: 23 November 2020 / Accepted: 20 January 2021 / Published online: 10 February 2021

Abstract

The prediction of hospital emergency room visits (ERV) for respiratory diseases after the outbreak of $PM_{2.5}$ is of great importance in terms of public health, medical resource allocation, and policy decision support. Recently, the machine learning methods bring promising solutions for ERV prediction in view of their powerful ability of short-term forecasting, while their performances still exist unknown. Therefore, we aim to check the feasibility of machine learning methods for ERV prediction of respiratory diseases. Three different machine learning models, including autoregressive integrated moving average (ARIMA), multilayer perceptron (MLP), and long short-term memory (LSTM), are introduced to predict daily ERV in urban areas of Beijing, and their performances are evaluated in terms of the mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), and coefficient of determination (R^2). The results show that the performance of ARIMA is the worst, with a maximum R^2 of 0.70 and minimum MAE, RMSE, and MAPE of 99, 124, and 26.56, respectively, while MLP and LSTM perform better, with a maximum R^2 of 0.80 (0.78) and corresponding MAE, RMSE, and MAPE of 49 (33), 62 (42), and 14.14 (9.86). In addition, it demonstrates that MLP cannot detect the time lag effect properly, while LSTM does well in the description and prediction of exposure-response relationship between $PM_{2.5}$ pollution and infecting respiratory disease.

Keywords PM2.5 exposure · Respiratory diseases · Emergency room visits · Machine learning

Introduction

The development of modern industry, accelerated urbanization process, and increasing consumption of energy have caused various environment problems. In particular, the air pollution problem has aroused great public attention and led to more and more concerns for public health, including cardiovascular diseases, respiratory diseases, psychological disturbances, early death, and so on (Cohen et al. 2017; Ren et al. 2016; Ho et al. 2014). There is evidence showing that the prevalence of respiratory diseases is increasing (Requia et al. 2018; Shaddick et al. 2018), and short or long exposure to high levels of $PM_{2.5}$ pollution will lead to greater chance for respiratory problems according to epidemiological studies (Xing et al. 2016; kappos et al. 2004). Meanwhile, recent studies have revealed the lag effect between exposure and reaction; for example, Li et al. have proved that multi-day lag effect is greater than single-day lag effect (Li et al.

Responsible Editor: Lotfi Aleya

✉ Ling Yao
yaoling@lreis.ac.cn

1 State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

2 College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100101, China

3 Southern Marine Science and Engineering Guangdong Laboratory, Guangzhou 511458, China

4 Beijing Huayun Shinetek Science and Technology Co., Ltd., Beijing 100101, China

5 Guangdong Open Laboratory of Geospatial Information Technology and Application, Key Laboratory of Guangdong for Utilization of Remote Sensing and Geographical Information System, Engineering Technology Center of Remote Sensing Big Data Application of Guangdong Province, Guangzhou Institute of Geography, Guangzhou 510070, China

6 Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing Normal University, Nanjing 210023, China

2013). It will induce high peak of hospital emergency room visit (ERV) and sudden pressure for medical resources. Therefore, it is meaningful to predict the hospital ERV of respiratory diseases after high levels of $PM_{2.5}$ pollution events, which can help government and hospital to make decisions and allocate relevant medical resources.

The common traditional regression methods for prediction, such as Grey model, Markov model, exponential smoothing model, and autoregressive integrated moving average (ARIMA) model, are usually easy and quick to perform and have been widely studied in previous studies (Kayacan et al. 2010; Song et al. 2014; Tseng et al. 2001). ARIMA is popular in time series analysis and forecast; for example, Luo et al. construct two traditional models, i.e., seasonal ARIMA model and single exponential smoothing model, and a combined model of the previous two to predict ERV. It demonstrates that traditional models are easy to be implemented and only need a small amount of calculation (Luo et al. 2017), which means that they are suitable for short-term forecast. In general, based on the statistical analyses of collected historical data, ERV for respiratory diseases over time can be predicted. However, there exists some disadvantages in these methods. One disadvantage is that traditional methods have difficulties in dealing with situations with multi-factor effects, that is, when multiple factors fluctuate, the prediction accuracy reduces significantly. In addition, traditional methods are influenced by human intervention to a certain extent, for instance, the selection of meteorological factors and the configuration of their weights in the prediction model.

It has been proven that various factors, including diverse air pollutants and weather conditions, are critical in the prediction of ERV for respiratory diseases (Zanobetti and Schwartz 2006; Green et al. 2010), but traditional methods are not good at handling multi-factor effects. On the contrary, artificial neutral networks can overcome such disadvantage and have been applied in many studies of different fields (Peng et al. 1992; Kuligowski and Barros 1998; Feng et al. 2015; Li et al. 2017). Bibi et al. use the back propagation (BP) neural network model to predict ERV for the respiratory disease in Ashkelon, Israel (Bibi et al. 2002), and find out that the model performance is significantly improved after temperature, humidity, and air pollutants are taken into consideration. Artificial neural networks also have strong nonlinear mapping capabilities and are able to construct the model totally depending on dataset itself, that is to say, they can model the relationships among data through automatic self-learning rather than with prior knowledge (Schmidhuber 2015; LeCun et al. 2015). Polezer et al. employ three different artificial neutral networks to assess the impact of $PM_{2.5}$ on respiratory disease, including multilayer perceptron (MLP), extreme learning machines (ELM), and echo state networks (ESN), finding MLP shows best performance (Polezer et al. 2018). In addition, compared with regression methods, artificial neural networks exhibit their superiority in simplifying the process of modeling the relationship between influencing factors and ERV.

Although artificial neutral networks have advantages in analyzing nonlinear relationship, it is difficult to describe relationship with time lag effect. Previous studies have found that there is time lag between exposure to fine particles and reaction of ERV for respiratory diseases (Mantovani et al. 2016; Halonen et al. 2008; Gasparrini et al. 2010), and it is worthwhile considering such time lag effect when predicting ERV. Traditional neutral networks cannot capture the time lag effect between the exposure to $PM_{2.5}$ and reaction, making prediction results less accurate. Long short-term memory (LSTM) model, a deep learning method, whose structure is inherently suitable for catching the time lag effect (Hochreiter and Schmidhuber 1997; Gers et al. 2000), brings possibility to handle this time lag effect problem.

Many researchers introduce various methods into topics related to public health, such as predicting morbidity and mortality, hospital admission, and so on, while there is little literature about using machine learning methods for predicting hospital ERV for respiratory diseases. Additionally, it is unclear to what degree different machine learning methods will affect the prediction accuracy. Therefore, we explore the feasibility of machine learning methods for predicting ERV for respiratory diseases. The ARIMA model, MLP model, and LSTM model, regarded as representative machine learning methods, are evaluated and compared, in order to support better allocation of medical resources.

Accordingly, the structure of the paper is organized as follows: “Data and methods” introduces the study area, data, machine learning methods, and evaluation metrics used in the study. “Results and discussion” describes the results and discussions of the different models, which is followed by conclusions in “Conclusion”.

Data and methods

Study area

Beijing, the capital of China, is the political and economic center of the China and has dense populations and intense human activity (Han et al. 2014). According to Beijing National Economic and Social Development Statistical Bulletin 2019 (http://www.beijing.gov.cn/), at the end of 2019, the resident population reaches 21.536 million, of which 18.65 million are urban residents, accounting for 86.6% of the resident population.

As shown in Fig. 1, Beijing is high in the northwest and low in the southeast and the southeast area of Beijing is an alluvial plain. This special topography is the geographical cause of air pollution in Beijing, especially when the southeast wind prevails, pollutants will accumulate in the southeast

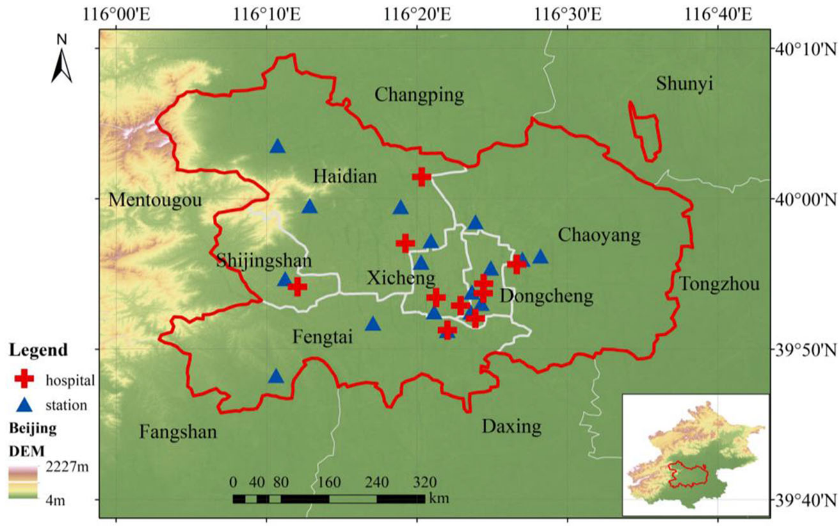

Fig. 1 Distribution of used hospitals (red crosses) and air quality monitor stations (blue triangles)

plain and blow straight into the urban area along with the wind. In addition, the mountains on three sides make it difficult for the pollutants accumulated in the urban area to spread, aggravating the urban air pollution. Located in the East Asian monsoon region, Beijing has a warm temperate semi-humid and semi-arid monsoon climate, which determines that foreign pollutants in Beijing mainly come from the northwest and southeast directions. Besides, the distribution of precipitation in Beijing is uneven throughout the whole year, that is, the precipitation is mainly concentrated in summer while there is less precipitation in the other seasons. In winter, the smog problem caused by heating gets worse due to the lack of precipitation (Sun et al. 2013). In January 2013, a long-lasting episode of severe haze occurred in central and eastern China (Wang et al. 2014), Beijing is also deeply affected. Therefore, we choose Beijing urban area as our study region and implement our experiments in 2013. The urban areas include Dongcheng District, Xicheng District, Chaoyang District, Haidian District, Fengtai District, and Shijingshan District (Fig. 1).

Dataset

The $PM_{2.5}$ concentration data and meteorological data are measured by air quality monitoring stations established by Beijing Environmental Protection Bureau, which can be accessed from the website http://zx.bjmemc.com.cn. The provided elements include daily mean concentration of $PM_{2.5}$ ($\mu g/m^3$), daily mean temperature (°C), daily mean relative humidity (%), and daily mean wind speed (MSPD) (m/s). There is a total of 17 air quality monitoring stations, and the daily average values of elements mentioned above are used to describe the air condition of Beijing urban area on the exact day. The basic statistic information of the dataset is shown in Table 1, including maximum value, minimum value, and average value of the all used data. It is noteworthy that the $PM_{2.5}$ pollution in urban area of Beijing is quite severe, with average $PM_{2.5}$ concentration reached 102.11 $\mu g/m^3$ in 2013, which is nearly 3 times above the second level of concentration limit of ambient air pollutants (35 $\mu g/m^3$) according to the ambient air quality standards (GB 3095-2012).

In order to represent the spatial distribution of the population in main urban area, the ERV data are collected from 10 comprehensive hospitals in the main urban area of Beijing, covering a period from January 1, 2013 to December 31, 2013. In addition, the ERV data is recorded by electronic medical system from one to three class-A hospitals in each urban area of Beijing (Xu et al. 2016). The locations of involved hospitals are shown as red crosses in Fig. 1. In this paper, the operational definition for the total ERV for respiratory diseases is ICD-10:J00-J99, including upper respiratory tract infection (ICD-10:J00), J02-J06), lower respiratory tract infections (ICD-10: J12-J22, J44 [except J44.103 and J44.901]), chronic obstructive pulmonary disease (ICD-10: J44.103, J44.901), and asthma (ICD-10:J45-46), according to the 10th edition of the International Classification of Diseases Code J00-J99. The basic statistic information of ERV is shown in Table 1. Besides, if daily admission exceeds the 4th quartile of within this week, this day is regarded as influenza epidemics. (Wong et al. 2002).

In addition, before the experiment, all kinds of data are standardized according to Equation (1), so that the scale of different influencing factors can be kept being the same.

$$v_{\mathrm{i}}^{*} = \frac{v_{\mathrm{i}} - \frac{1}{n}\sum_{i=1}^{n} v_{\mathrm{i}}}{\sqrt{\frac{\sum_{i=1}^{n}\left(v_{\mathrm{i}} - \overline{v}\right)^{2}}{n}}} \tag{1}$$

Table 1 The basic information of the used dataset

	$PM_{2.5}$ (μg/m^3)	Temperature (°C)	Relative humidity (%)	MSPD (m/s)	ERV(person)
Maximum	508.53	29.01	93.25	6.60	556
Minimum	6.65	−12.56	18.95	0.70	129
Average	102.11	11.30	58.67	2.12	253

where v_i^* represents the ith value of certain elements after standardization, v_i represents the original value of ith element.

Machine learning methods

Three typical machine learning methods including ARIMA, MLP, and LSTM used are introduced in this section. For easy expression and better understanding, the whole dataset is denoted as $\mathbf{D} = \{\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_n\}$, in which n represents the length of the times series. For the ith day, $\mathbf{D}_i$ is denoted as $\mathbf{D}_i = \{x_{1,i}, x_{2,i}, \ldots, x_{m,i}, y_i\}$, in which $x_{m,i}$ is the average of the mth influencing factor on the ith day, m represents the number of factors, and y_i represents the total ERV of involved hospitals of the certain day. Since the lag effect is considered in this study, the predictions are made based on 1- to 5-day lag assumptions, naming as timestep t, which means the observed data of previous several days are used to predict the patient number on the certain day, i.e., $\{\mathbf{D}_{i-t}, \ldots, \mathbf{D}_{i-2}, \mathbf{D}_{i-1}\}$ is used to predict $\{y_i\}$. In addition, the whole dataset is divided as training and testing data, and the ERV is predicted using three different machine learning methods for inter-comparisons.

ARIMA model is one of the most classic models for predication of time series, which has been widely applied in many areas, such as electricity price prediction, energy consumption forecast, and so on (Contreras et al. 2003; Zhang et al. 2003; Yuan et al. 2016). The basic idea of the ARIMA model is to use a certain mathematical model to describe the random time series of the data, then predict the future values based on the past, and present values, so-called autoregression. An ARIMA (p, d, q) model can be described as following formula.

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d\mathbf{Y} = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon \tag{2}$$

where L represents the lag operator, p represents the number of autoregressive terms, q represents the number of moving average terms, and d represents the degree of differencing, and ϕ, θ, and ε are relevant parameters. In the ARIMA model, $\mathbf{Y} = \{y_{i-t}, \ldots, y_{i-2}, y_{i-1}\}$ is used to predict $\{y_i\}$. Specifically, Akaike information criterion (AIC) is used to determine the rank of ARMA model, and five ARIMA model is constructed based on time lag assumptions of 1–5 days with the smallest AIC.

MLP model is one of the most effective artificial neural network technologies for modeling and forecasting, so it has been used as a benchmark model by many studies (Bui et al. 2017). MLP has been widely used to simulate nonlinear and complex processes in the real world, because of its high general approximation ability (Pham et al. 2017; Sadowski et al. 2018). MLP includes at least three layers: input layer, hidden layer, and output layer. The neurons of each layer in MLP are fully interconnected. The number of neurons in the input layer, hidden layer, and output layer is determined by the number of independent variables, the size of the training data set, and the classification task. At the same time, MLP model training can be divided into two stages based on forward propagation algorithm and back propagation algorithm (Bui et al. 2016). The perceptron forms a new linear input combination (or a nonlinear input combination) by assigning different weights to the input, and then generates a single output based on several true-valued inputs. It can be described by the following formula:

$$\mathbf{y}_i = \varphi\left(\sum_{i=1}^{nt}\boldsymbol{\omega}_i\mathbf{x}_i + b\right) = \varphi\left(\boldsymbol{\omega}^T\mathbf{x} + \mathbf{b}\right) \tag{3}$$

where $\mathbf{y}$ represents the total ERV of 10 hospital, $\boldsymbol{\omega}$ represents the weight vector, the input vector $\mathbf{x}$ includes nt values representing n influencing factors of previous t days, $\mathbf{b}$ is the deviation, and φ is the nonlinear activation function.

An MLP model with one input layer, one hidden layer, and one output layer is constructed to make the prediction of patient number asking for medical help after exposure to high levels of $PM_{2.5}$ pollution. The number of nodes in the input layer is set corresponding with the time lag and the number of nodes in the output layer is a constant value of 1. The remaining parameter settings of the model include the number of hidden layer nodes, the activation function of each remaining layer, optimizer selection, batch size, number of iterations, and so on, which are all determined by tree-structured Parzen estimator (TPE) algorithm within given ranges; the final chosen parameters are shown in Table 2. And the model loss function is set to mean square error (MSE).

Adopted set values are italicized

Learning the long-term dependencies is a difficult problem for recurrent neutral networks (RNNs), because of the "gradient vanishing and blowing up" phenomenon in long process of back propagation (Hochreiter and Schmidhuber 1997; Bengio et al. 1994). In order to handle this long-time lag task, LSTM model is proposed by Hochreiter and Jürgen

Table 2 The structure of the MLP model

Parameter name	Search space
The activation function of the first hidden layer	Sigmoid, tanh, *ReLU*
The activation function of output layer	*Sigmoid*, tanh, ReLU
Number of nodes in the first hidden layer	16, *32*, 64, 128
Optimizer	SGD, *Adam*, RMSprop
Batch size	*8*, 16, 32, 64, 128
Number of iterations	*250*, 500

Schmidhuber in 1997 (Hochreiter and Schmidhuber 1997), which has new repeating unit, different from the unit in the standard RNN network (Ger et al. 2000). The developed memory unit includes four main elements: input gates, neurons with self-loop connections, forgetting gates, and output gates. In an LTSM cell, the forget gate decides what and how information to be discarded from the calculation, using nonlinear functions and weight matrixes. The input gate determines what information to be added into the calculation (sigmoid layer) and gets the new candidate information (tanh layer). The cell updates the information. Finally, the output gate determines what and how information to be output, also including a sigmoid layer and a tanh layer. In general, the LSTM model can be described by the following formula:

$$\mathbf{y}_{\mathrm{i}} = LSTM(\mathbf{D}_{\mathrm{i-1}}, \mathbf{D}_{\mathrm{i-2}}, \ldots, \mathbf{D}_{\mathrm{i-t}}) \tag{4}$$

where i represents the ith day, $\mathbf{y}$ represents the ERV, $LSTM$ represents the algorithm of the LSTM method, $\mathbf{D}_{\mathrm{i}}$ represents the vector of $\{x_{1,\mathrm{i}}, x_{2,\mathrm{i}}, \ldots, x_{\mathrm{m},\mathrm{i}}, y_{\mathrm{i}}\}$, in which x represents the vector of different influencing factors, m represents the number of influencing factors, and t represents the timestep.

A three-layer LSTM together with two dense layers is constructed to make the prediction of how many patients will ask for medical help after a few days of exposure to high levels of $PM_{2.5}$. Dense layers are activated by ReLU function, and the loss function of the LSTM model is set as MSE. To optimize the parameters of the model, TPE algorithm is also used based on the *hyperopt* package (Bergstra et al. 2013) and *keras* package (Chollet 2015) of python in this study. The remaining parameters are listed in Table 3.

Adopted set values are italicized

Table 3 The structure of the LSTM model

Parameter name	Search space
Number of nodes in layer 1 of LSTM	16, 32, *64*, 128
Number of nodes in layer 2 of LSTM	16, 32, *64*, 128
Number of nodes in layer 3 of LSTM	16, 32, 64, *128*
Units of dense layer 1	/
Activation function of layer 1	Sigmoid, tanh, *ReLU*
Units of dense layer 2	/
Optimizer	SGD, *Adam*, RMSprop
Batch size	8, 16, 32, *64*, 128
Number of iterations	250, *500*
Initial learning rate	0.1, 0.05, *0.02*

Evaluation methods

Several evaluation indicators are chosen to measure the prediction accuracy of different models, including the mean absolute error (MAE), the root mean squared error (RMSE), the mean absolute percentage error (MAPE), and the coefficient of determination (R^2). Their formulas are as follows:

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\widehat{y_{\mathrm{i}}}-y_{\mathrm{i}}\right| \tag{5}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\left|\widehat{y_{\mathrm{i}}}-y_{\mathrm{i}}\right|-y_{\mathrm{i}}\right)^2} \tag{6}$$

$$\mathrm{MAPE} = \sum_{i=0}^{n}\left|\frac{\widehat{y_{\mathrm{i}}}-y_{\mathrm{i}}}{\widehat{y_{\mathrm{i}}}}\right| \times \frac{100}{n} \tag{7}$$

$$R^2 = 1-\frac{\sum_{i=0}^{n}\left(\widehat{y_{\mathrm{i}}}-y_{\mathrm{i}}\right)^2}{\sum_{i=0}^{n}\left(y_{\mathrm{i}}-\overline{y_{\mathrm{i}}}\right)^2} \tag{8}$$

where i means the ith day, y_{i} represents the actual number of ERV, $\overline{y_{\mathrm{i}}}$ represents the mean value of actual number of ERV, and $\widehat{y_{\mathrm{i}}}$ represents the predicted number of ERV.

Results and discussion

The paper deals with the feasibility different machine learning methods for predicting ERV for respiratory diseases after exposure to high levels of $PM_{2.5}$. To evaluate and compare these methods, four common statistical metrics are discussed; Table 4 gives the performances both for all methods and time steps. The R^2 varies between 0.61 and 0.8, indicating that the accuracy of various methods is influenced by the structures of different models to a large degree.

The best result of each method is italicized

The ARIMA model shows the best prediction results compared with observation when timestep is 1, with the maximum R^2 of 0.70. The performance of ARIMA gets worse when timestep increases. The MAE, RMSE, and MAPE get the minimum values when timestep equals 2, being 99, 124, and 25.56, respectively. With the increase on timestep, the MAE increases from 99 to 102, the RMSE stays stable between 124

Table 4 Performance comparison of different methods

t	Metrics	ARIMA	MLP	LSTM
1	MAE (person)	*99*	*49*	64
	RMSE (person)	*125*	*62*	75
	MAPE (%)	*26.78*	*14.14*	19.46
	R^2	*0.70*	*0.80*	0.70
2	MAE (person)	99	45	60
	RMSE (person)	124	59	71
	MAPE (%)	26.56	12.66	18.86
	R^2	0.66	0.75	0.70
3	MAE (person)	99	51	*33*
	RMSE (person)	124	63	*42*
	MAPE (%)	26.72	14.59	*9.86*
	R^2	0.67	0.80	*0.78*
4	MAE (person)	100	55	36
	RMSE (person)	125	68	47
	MAPE (%)	27.07	15.72	10.52
	R^2	0.64	0.76	0.73
5	MAE (person)	101	60	48
	RMSE (person)	125	72	54
	MAPE (%)	27.14	16.93	14.39
	R^2	0.61	0.74	0.73

and 125, and the MAPE increases from 26.78 to 27.14, while the R^2 decreases from 0.70 to 0.61. ARIMA model does not perform well in predicting the hospital ERV for respiratory diseases. ARIMA is an autoregression method depending on historical data series; hence, the limited length of the data series used in this study might account for the poor performance of ARIMA. On the other hand, the fact that ARIMA model only depends on one-dimension dataset without taking other influencing factors into consideration might be another cause for the prediction results.

The results of MLP model show the relatively good value in terms of evaluation metrics, the MAE increases from 45 to 60, the RMSE increases from 59 to 72, the MAPE increases from 12.66 to 16.93, and the R^2 decreases from 0.80 to 0.74. The minimum MAE (45), RMSE (59), MAPE (12.66) are achieved on day 2, and maximum R^2 (0.80) are achieved on day 1 and day 3.

In contrast, LSTM performs better than MLP and ARIMA, achieving the best accuracy when timestep is 3 with the MAE,

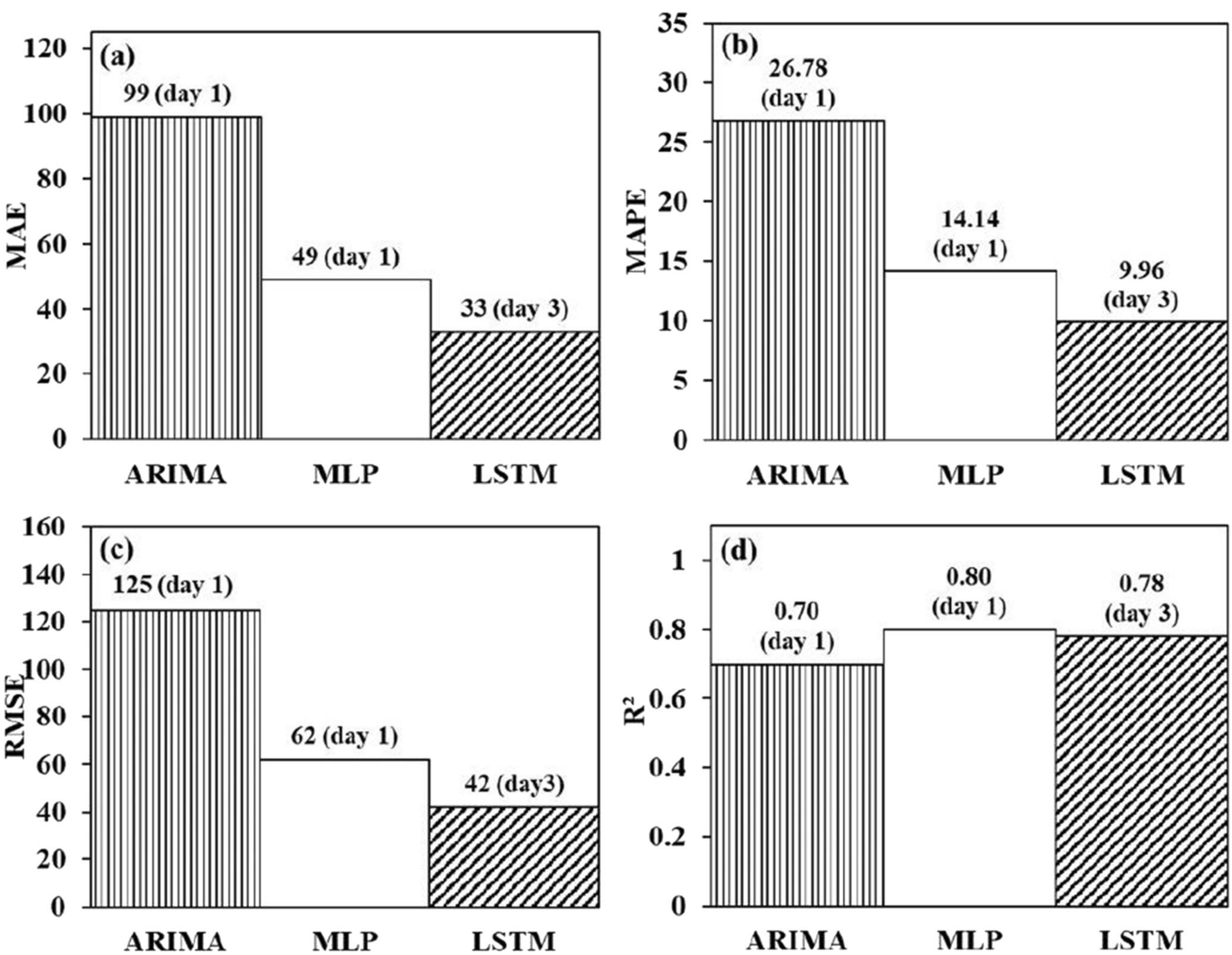

Fig. 2 Prediction accuracy of three different models in terms of MAE (**a**), MAPE (**b**), RMSE (**c**), and R^2 (**d**)

RMSE, MAPE, and R^2 of 33, 42, 9.86, and 0.78, respectively. When timestep decreases or increases, the prediction results get less accuracy with highest MAE of 64, RMSE of 75, MAPE of 19.46 on day 1, and lowest R^2 of 0.70 both on day 1 and day 2.

Each model shows respective best prediction accuracy on different days; among them, the result of ARIMA shows highest R^2 on day 1, the result of MLP shows best performance on day 1, and the result of LSTM shows best performance on day 3. The comparison is conducted using the best accuracy of each method, shown in Fig. 2.

Different methods achieve best prediction accuracy on different lag day, which is day 1 of ARIMA, day 1 of MLP, and day 3 of LSTM. Comparing each best accuracy, the MLP model and LSTM model show better performance compared with ARIMA model. The reason might be that these two methods can consider multiple influencing factors at the same time, like meteorological factors and influenza, leading higher prediction accuracy, while ARIMA model just makes prediction based on autoregression.

As for time lag effect detection, the MLP model performs best on day 1, with highest R^2 and relatively low error, while the LSTM model shows best performance on day 3. Considering previous study, there indeed exists lag effect between exposure and reaction (Guaita et al. 2011; Liu et al. 2017); the results indicate that MLP model cannot handle the time lag in the time series properly, while LSTM model shows advantage on the detection of lag effect. This means models which have structure to describe long time dependence, like LSTM, is meaningful to this kind of study.

Among selected models, ARIMA model has the worst performance; LSTM reaches the highest accuracy. These data-driven machine learning methods can overcome the disadvantages of traditional statistical methods, requiring no prior knowledge and avoiding human intervention. On the other hand, the well-designed model, LSTM, whose structure makes it possible to memorize the temporal characteristics of the time series, can also handle the time lag problem, achieving best performance on day 3. In the practice of forecasting hospital ERV for respiratory, the ability for accurate forecast and catching time lag effect are both of great significance. Therefore, machine learning methods can be used to patient number forecast, and LSTM shows advantage comparing with other methods used in this study.

Conclusion

In this paper, three machine learning methods, including the ARIMA, MLP, and LSTM model, are applied to predict hospital ERV for respiratory diseases after exposure to fine particles. The performances of three methods are evaluated using data from Beijing urban area in 2013, and inter-comparison is made for the prediction accuracy. The results show that ARIMA model performs the worst on predicting patient number, MLP model performs relatively better, and LSTM model has the best performance, with the minimum MAE of 33, MAPE of 9.86, RMSE of 42, and maximum R^2 of 0.80 on day 3, respectively. In addition, MLP model hardly catches the time lag, while LSTM model shows better ability in detecting time lag, which indicates that LSTM model is more suitable to predict hospital ERV for respiratory diseases after exposure to $PM_{2.5}$. This study is of practical value to support medical resource allocation and policy decision.

Acknowledgements The work is supported by the National Data Sharing Infrastructure of Earth System Science (http://www.geodata.cn/).

Author contribution L.Y principally conceived the idea of the study and designed the experiment, J.L. and P.B. performed the experiment and were major contributors in writing the manuscript, and X.X., N.L., and H.J. participated in the completement of this work.

Funding This work is jointly supported by grants from the National Natural Science Foundation of China (Grant number 41771380), Key Special Project for Introduced Talents Team of Southern Marine Science and Engineering Guangdong Laboratory (No. GML2019ZD0301), the GDAS' Project of Science and Technology Development (No. 2020GDASYL-20200103003), the China Postdoctoral Science Foundation (NO. 2020M682628), and a grant from State Key Laboratory of Resources and Environmental Information System.

Data Availability All data generated or analyzed during this study in this article are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare that they have no conflict of interest.

References

Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. IEEE transactions on neural networks 5(2):157–166. https://doi.org/10.1109/72.279181

Bergstra J, Yamins D, Cox DD (2013) Making a science of model search: hyperparameter optimization in hundreds of dimensions for vision architectures. To appear in Proc. of the 30th International Conference on Machine Learning (ICML 2013).

Bibi H, Nutman A, Shoseyov D, Shalom M, Peled R, Kivity S, Nutman J (2002) Prediction of emergency department visits for respiratory symptoms using an artificial neural network. Chest 122(5):1627–1632. https://doi.org/10.1378/chest.122.5.1627

Bui DT, Tuan TA, Klempe H et al (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. Landslides 13(2):361–378. https://doi.org/10.1007/s10346-015-0557-6

Bui DT, Bui KTT, Bui QT, et al (2017) Hybrid intelligent model based on least squares support vector regression and artificial bee colony optimization for time-series modeling and forecasting horizontal displacement of hydropower dam[M]//Handbook of Neural Computation 279-293. https://doi.org/10.1016/B978-0-12-811318-9.00015-6

Chollet F (2015) keras.

Cohen AJ, Brauer M, Burnett R, Anderson HR, Frostad J, Estep K, Balakrishnan K, Brunekreef B, Dandona L, Dandona R, Feigin V, Freedman G, Hubbell B, Jobling A, Kan H, Knibbs L, Liu Y, Martin R, Morawska L, Pope CA III, Shin H, Straif K, Shaddick G, Thomas M, van Dingenen R, van Donkelaar A, Vos T, Murray CJL, Forouzanfar MH (2017) Estimates and 25-year trends of the global burden of disease attributable to ambient air pollution: an analysis of data from the Global Burden of Diseases Study. The Lancet 389(10082):1907–1918. https://doi.org/10.1016/S0140-6736(17)30505-6

Contreras J, Espinola R, Nogales FJ, Conejo AJ (2003) ARIMA models to predict next-day electricity prices. IEEE transactions on power systems 18(3):1014–1020. https://doi.org/10.1109/TPWRS.2002.804943

Feng X, Li Q, Zhu Y, Hou J, Jin L, Wang J (2015) Artificial neural networks forecasting of PM2. 5 pollution using air mass trajectory based geographic model and wavelet transformation. Atmospheric Environment 107:118–128. https://doi.org/10.1016/j.atmosenv.2015.02.030

Gasparrini A, Armstrong B, Kenward MG (2010) Distributed lag non-linear models. Statistics in medicine 29(21):2224–2234. https://doi.org/10.1002/sim.3940

Gers F A, Schmidhuber J, Cummins F (2000) Learning to forget: continual prediction with LSTM. Neural Computation 2:850–855. https://doi.org/10.1162/089976600300015015

Green RS, Basu R, Malig B, Broadwin R, Kim JJ, Ostro B (2010) The effect of temperature on hospital admissions in nine California counties. International Journal of Public Health 55(2):113–121. https://doi.org/10.1007/s00038-009-0076-0

Guaita R, Pichiule M, Maté T, Linares C, Díaz J (2011) Short-term impact of particulate matter (PM2. 5) on respiratory mortality in Madrid. International journal of environmental health research 21(4):260–274. https://doi.org/10.1080/09603123.2010.544033

Halonen JI, Lanki T, Yli-Tuomi T, Kulmala M, Tiittanen P, Pekkanen J (2008) Urban air pollution and asthma and COPD hospital emergency room visits. Thorax. 63:635–641. https://doi.org/10.1136/thx.2007.091371

Han L, Zhou W, Li W, Li L (2014) Impact of urbanization level on urban air quality: a case of fine particles (PM2.5) in Chinese cities. Environmental Pollution 194:163–170. https://doi.org/10.1016/j.envpol.2014.07.022

Ho RC, Zhang MW, Ho CS, Pan F, Lu Y, Sharma VK (2014) Impact of 2013 south Asian haze crisis: study of physical and psychological symptoms and perceived dangerousness of pollution level. BMC psychiatry 14(1):81. https://doi.org/10.1186/1471-244X-14-81

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural computation 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Kappos AD, Bruckmann P, Eikmann T, Englert N, Heinrich U, Höppe P, Koch E, Krause GHM, Kreyling WG, Rauchfuss K, Rombout P, Schulz-Klemp V, Thiel WR, Wichmann HE (2004) Health effects of particles in ambient air. International journal of hygiene and environmental health 207(4):399–407. https://doi.org/10.1078/1438-4639-00306

Kayacan E, Ulutas B, Kaynak O (2010) Grey system theory-based models in time series prediction. Expert systems with applications 37(2):1784–1789. https://doi.org/10.1016/j.eswa.2009.07.064

Kuligowski RJ, Barros AP (1998) Experiments in short-term precipitation forecasting using artificial neural networks. Monthly weather review 126(2):470–482. https://doi.org/10.1175/1520-0493(1998)126<0470:EISTPF>2.0.CO;2

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444. https://doi.org/10.1038/nature14539

Li P, Xin J, Wang Y, Wang S, Li G, Pan X, Liu Z, Wang L (2013) The acute effects of fine particles on respiratory mortality and morbidity in Beijing, 2004-2009. Environmental Science & Pollution Research 20(9):6433–6444. https://doi.org/10.1007/s11356-013-1688-8

Li X, Peng L, Yao X, Cui S, Hu Y, You C, Chi T (2017) Long short-term memory neural network for air pollutant concentration predictions: method development and evaluation. Environmental pollution 231:997–1004. https://doi.org/10.1016/j.envpol.2017.08.114

Liu ST, Liao CY, Kuo CY et al (2017) The effects of PM2. 5 from Asian dust storms on emergency room visits for cardiovascular and respiratory diseases. International journal of environmental research and public health 14(4):428. https://doi.org/10.3390/ijerph14040428

Luo L, Luo L, Zhang X, He X (2017) Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. BMC health services research 17(1):469. https://doi.org/10.1186/s12913-017-2407-9

Mantovani KCC, Nascimento LFC, Moreira DS et al (2016) Air pollutants and hospital admissions due to cardiovascular diseases in São José do Rio Preto, Brazil. Ciencia & saude coletiva 21(2):509–516. https://doi.org/10.1590/1413-81232015212.16102014

Peng TM, Hubele NF, Karady GG (1992) Advancement in the application of neural networks for short-term load forecasting. IEEE Transactions on Power Systems 7(1):250–257. https://doi.org/10.1109/59.141711

Pham BT, Bui DT, Prakash I et al (2017) Hybrid integration of multilayer perceptron neural networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. Catena 149:52–63. https://doi.org/10.1016/j.catena.2016.09.007

Polezer G, Tadano YS, 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 PM2. 5 on respiratory disease using artificial neural networks. Environmental pollution 235:394–403. https://doi.org/10.1016/j.envpol.2017.12.111

Ren J, Li B, Yu D et al (2016) Approaches to prevent the patients with chronic airway diseases from exacerbation in the haze weather. Journal of thoracic disease 8(1):E1. https://doi.org/10.3978/j.issn.2072-1439.2015.11.61

Requia WJ, Adams MD, Arain A, Papatheodorou S, Koutrakis P, Mahmoud M (2018) Global association of air pollution and cardiorespiratory diseases: a systematic review, meta-analysis, and investigation of modifier variables. American journal of public health 108(S2):S123–S130

Sadowski Ł, Hoła J, Czarnecki S, Wang D (2018) Pull-off adhesion prediction of variable thick overlay to the substrate. Automation in Construction 85:10–23. https://doi.org/10.1016/j.autcon.2017.10.001

Schmidhuber J (2015) Deep learning in neural networks: an overview. Neural networks 61:85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Shaddick G, Thomas ML, Amini H, Broday D, Cohen A, Frostad J, Green A, Gumy S, Liu Y, Martin RV, Pruss-Ustun A, Simpson D, van Donkelaar A, Brauer M (2018) Data integration for the assessment of population exposure to ambient air pollution for global burden of disease assessment. Environmental science & technology 52(16):9069–9078. https://doi.org/10.1021/acs.est.8b02864

Song Z, Jiang Y, Zhang Z (2014) Short-term wind speed forecasting with Markov-switching model. Applied Energy 130:103–112. https://doi.org/10.1016/j.apenergy.2014.05.026

Sun YL, Wang ZF, Fu PQ, Yang T, Jiang Q, Dong HB, Li J, Jia JJ (2013) Aerosol composition, sources and processes during wintertime in Beijing, China. Atmospheric Chemistry and Physics 13(9):4577–4592. https://doi.org/10.5194/acp-13-4577-2013

Tseng FM, Yu HC, Tzeng GH (2001) Applied hybrid grey model to forecast seasonal time series. Technological Forecasting and Social Change 67(2-3):291–302. https://doi.org/10.1016/S0040-1625(99)00098-0

Wang YS, Yao L, Wang LL, Liu ZR, Ji DS, Tang GQ, Zhang JK, Sun Y, Hu B, Xin JY (2014) Mechanism for the formation of the January 2013 heavy haze pollution episode over central and eastern China. Science China Earth Sciences 57(1):14–25. https://doi.org/10.1007/s11430-013-4773-4

Wong CM, Atkinson RW, Anderson HR, Hedley AJ, Ma S, Chau PYK, Lam TH (2002) A tale of two cities: effects of air pollution on hospital admissions in Hong Kong and London compared. Environmental health perspectives 110(1):67–77. https://doi.org/10.1289/ehp.0211067

Xing YF, Xu YH, Shi MH et al (2016) The impact of PM2. 5 on the human respiratory system. Journal of thoracic disease 8(1):E69. https://doi.org/10.3978/j.issn.2072-1439.2016.01.19

Xu Q, Li X, Wang S, Wang C, Huang F, Gao Q, Wu L, Tao L, Guo J, Wang W, Guo X (2016) Fine particulate air pollution and hospital emergency room visits for respiratory disease in urban areas in Beijing, China, in 2013. PLoS One 11(4):e0153099. https://doi.org/10.1371/journal.pone.0153099

Yuan C, Liu S, Fang Z (2016) Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM (1, 1) model. Energy 100:384–390. https://doi.org/10.1016/j.energy.2016.02.001

Zanobetti A, Schwartz J (2006) Air pollution and emergency admissions in Boston, MA. Journal of Epidemiology & Community Health 60(10):890–895. https://doi.org/10.1136/jech.2005.039834

Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50:159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

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