



Environmental and infrastructural effects on respiratory disease exacerbation: a LBSN and ANN-based spatio-temporal modelling

Zeinab Neisani Samani · Mohammad Karimi · Aliasghar Alesheikh

Received: 30 April 2019 / Accepted: 22 November 2019
© Springer Nature Switzerland AG 2020

Abstract Owing to the rise in population, lifestyle changes, high traffic rates in urban areas and environmental pollution, respiratory diseases have become much more prevalent on both regional and urban scales. Respiratory diseases affect over 300 million people worldwide and are thus among the major threats to humans' general well-being. The identification of underlying factors and the specification of accompanying risk areas for the temporal exacerbation of respiratory diseases are effective steps in managing the damage caused by such disorders. Here, we demonstrate a strategy for modelling the risk zone of respiratory diseases temporally, using a location-based social network (LBSN) and an artificial neural network (ANN). The main contribution of this paper is to consider the environmental and infrastructural factors and identify their relationships with the geographical locations of respiratory attacks. The study also utilizes Telegram, which is

the most popular and conventional social media platform, in order to observe temporal changes in the location of respiratory attacks in Iran, in the form of a developed Telegram bot known as *@respiratoryassociation*. The relations between the factors behind and the location of respiratory attacks are determined using a multilayer perceptron (MLP) ANN. All the required data have been collected on a daily basis over a 5-year period from December 2013 to December 2018 in Tehran, Iran. The results indicated air pollution, especially pollution from carbon monoxide (CO) and suspended particulate matter (PM) as the most decisive factors. Following air pollution, the amount of exposure to the polluted area was determined as the second most decisive factor, which in turn increased as a result of escalations in traffic jams. Land use was determined as the third most decisive factor. Furthermore, the results revealed that the ANN performed satisfactorily, implying that the model can be used to examine the spatio-temporal behaviour of the time series of respiratory diseases with respect to environmental and infrastructural factors.

Z. Neisani Samani (✉)
Department of RS/GIS, Science and Research Branch, Islamic Azad University, Tehran 14778 93855, Iran
e-mail: zeinab.neisani@srbiu.ac.ir

M. Karimi · A. Alesheikh
Department of GIS, Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, Tehran 19967 15433, Iran

M. Karimi
e-mail: mkarimi@kntu.ac.ir

A. Alesheikh
e-mail: alesheikh@kntu.ac.ir

Keywords Environmental pollution · Spatio-temporal modelling · LBSN · ANN · MLP

Introduction

Respiratory diseases are among the major illnesses prevalent in metropolitan cities with a high population density and significant traffic levels (Drummond 2014),

which are commonly found in the form of respiratory asthma, pneumonia, allergies, bronchitis and influenza. Based on the latest report from the World Health Organization (2017), the number of people suffering from respiratory diseases throughout the world has increased tremendously. Due to the high number of respiratory patients, if not treated and controlled, this disease can turn into a major catastrophe for public health (Yun 2012). Thus, this paper aims to model the spatio-temporal variability of respiratory diseases and attacks in urban environments. The study also seeks to rank the effective environmental and infrastructural parameters for predicting the location and timing of respiratory attacks.

Researchers have demonstrated that the most decisive factors behind the exacerbation of respiratory diseases are environmental issues such as pollution, rainfall and air temperature (Polezer et al. 2017). Although the exact nature of the relationship between these environmental factors and respiratory diseases is rather clear (Sanidas et al. 2017), there are still other infrastructural parameters influencing respiratory attacks, such as distance from green spaces, distance from pollution sources, land use, elevation, economic and social status, traffic congestion leading to increased air pollution and the amount of exposure to air pollution, which remain unstudied. Thus, the first contribution of this paper is to combine environmental and infrastructural factors in order to identify their relationships with the location of respiratory attacks via the design of an MLP ANN.

On the other hand, social media has recently become a valuable source of real-time data which can be utilized to facilitate data collection processes, such as in the case of health data (Chen et al. 2017). Since the required information changes according to place and time, LBSNs are considered as reliable platforms for environmental and health crowdsourcing of data. The second contribution of this paper is the utilization of an LBSN as a means for gathering the required environmental and health data. As the third contribution, this paper provides a temporal risk map via a mobile GIS platform. The software application can also predict risk locations for all types of respiratory diseases.

Along these lines, seven infrastructural factors (distance to green spaces, distance to pollution resources, land use, elevation, economic status, social status and traffic congestion) were considered in addition to six environmental factors (type of air pollution, temperature, wind speed, relative pressure, rainfall and

humidity) (Joseph et al. 2013; Olivieri and Scoditti 2005; Trang and Tripathi 2014).

The proposed method was implemented in Tehran, the capital city of Iran. Patients' geographical position data, along with their profile information, were gathered over a 5-year period from December 2013 to December 2018. Nearly 732 respiratory patients with different types of disorders registered themselves on the designed LBSN platform, i.e. the Telegram bot *@respiratoryassociation*. The relations between the specified factors and the location of respiratory attacks were determined using the MLP NN. The evaluation of the proposed method was carried out by obtaining the root mean square error (RMSE) of the trained MLP. The prediction accuracy was estimated by precision, recall and *F*-score quantifiers. The obtained results indicated a high percentage of accuracy.

The rest of this paper is structured as follows. The “[Related works](#)” section assesses the related studies. The “[Materials and Methods](#)” section presents the materials and methods in detail. The implementation process and the experimental procedures are explained in the “[Implementation and experimental results](#)” section. Lastly, conclusions and directions for future studies are given in the “[Conclusion and future works](#)” section.

Related works

This section reviews the various studies found in the health-GIS literature, focusing on three main fields: the determination of effective environmental and infrastructural factors related to respiratory attacks; the use of LBSNs in the health-related data collection; and the methods for vulnerable zoning and prediction. The assessment is arranged according to publication date. Ho et al. (2007) investigated the impacts of air pollution and weather conditions on the prevalence of allergic asthma using logistic regression, proving that oxidants, dust and low air temperature are far more decisive in asthma exacerbations compared to other environmental variables. Lee et al. (2011) examined the effects of environmental factors and medical bio-signals on asthma and used decision trees and association rules for monitoring asthma attacks. Gorai et al. (2014) assessed the association between air pollution and asthma in New York State using a GIS-based method between 2005 and 2007. They applied fine PM (PM_{2.5}), sulphur dioxide (SO₂) and ozone (O₃) as fundamental

factors behind asthma prevalence using interpolation techniques, such as inverse distance weighting, kriging and land use regression modelling.

Adams and Kanaroglou (2015) proposed a mobile air pollution monitoring system. They applied a Neural Network (NN) inspired by land use regression to provide better health hazard information. These mobile air pollution data were modelled with a number of predictor variables which contained information on land use specifications, the meteorological situation, air pollution and condensation from fixed location monitors, as well as traffic information from the time of collection.

Khreis et al. (2019) developed prior NO₂ and NO_x approximation and quantified the issue of childhood asthma cases attributable to PM_{2.5}, PM₁₀ in Bradford (UK) and especially the traffic-associated portion of PM_{2.5}, PM₁₀ and BC. They also studied the health impacts of decreased air pollutants and compliance (or not) with World Health Organization air quality guideline values.

Polezer et al. (2017) applied an ANN to assess the impact of air pollution on human health. Three MLP ANN, extreme learning machines and echo state networks for respiratory diseases were tested. The MLP ANN gave the best outcome, showing a notable influence of PM_{2.5}, temperature and humidity on hospital attendance. Blanco et al. (2018) studied the effect of air pollution on chronic obstructive pulmonary diseases through inverse distance weighting and geostatistical analysis. They modelled the prevalence of these types of diseases and prepared different temporal maps. Zhou et al. (2018) suggested a perceptual framework for social media-based health information management by drawing on multidisciplinary research. This framework addressed how to overcome research challenges as well as highlighted important yet under-explored research issues and offered promising directions for future research. Kumarihamy and Tripathi (2019) applied geographically weighted regressions to the prevalence of asthma and chronic obstructive pulmonary disease (COPD). They assessed the effect of environmental, climatic and socioeconomic determinants. Finally, they concluded that humidity, proximity to roads (0–200 m), road density, use of firewood as a source of fuel and elevation play vital roles in predicting morbidity from asthma and COPD. Kaffash Charandabi et al. (2019) enabled ubiquitous asthma supervision within a framework based on environmental air pollutants and personal contexts. This

GIS-based system was designed in three layers: (1) preprocessing, (2) reasoning and (3) prediction. The model is applicable to a real data set comprising internal and external contextual information about patients, air pollutant data and geographic information relevant to the city of Tehran (Iran). This method can be regarded as a groundbreaking piece of work in asthma monitoring and handling.

An evaluation of these studies reveals that researchers have used a variety of methods to predict the potential risk of respiratory attacks. However, the main drawbacks of their studies are as follows: 1) the considered parameters were limited to environmental factors, while infrastructural parameters may prove as significant as environmental ones; 2) limited references to temporal variations in vulnerable areas; and 3) indifference towards social media and LBSNs as useful and popular platforms for dynamic data collection.

Materials and methods

This section first describes the study area of the research, then concentrates on the main concepts of the paper before finally explaining the proposed methodology.

Study area

Tehran, the capital of Iran, is located in the Middle East, in the south-western part of Asia. It is geographically situated at latitude of 35° 43' 00.12" north and a longitude of 51° 24' 00.00" east. Tehran is the most populated city in Iran as well as in the Middle East. The population of Tehran is about 8.8 million within the city and extends to 15 million when considering the larger metropolitan zone of Greater Tehran (Fig. 1) (Worldatlas 2019).

Respiratory disease

Different types of respiratory diseases were considered in this study. They are as follows:

Asthma

Allergic asthma is a type of airway obstruction and inflammation, known to be activated by allergens. Asthma intensifies after exercising in cold air, after smoking, or by breathing dust or gases. Since allergens are

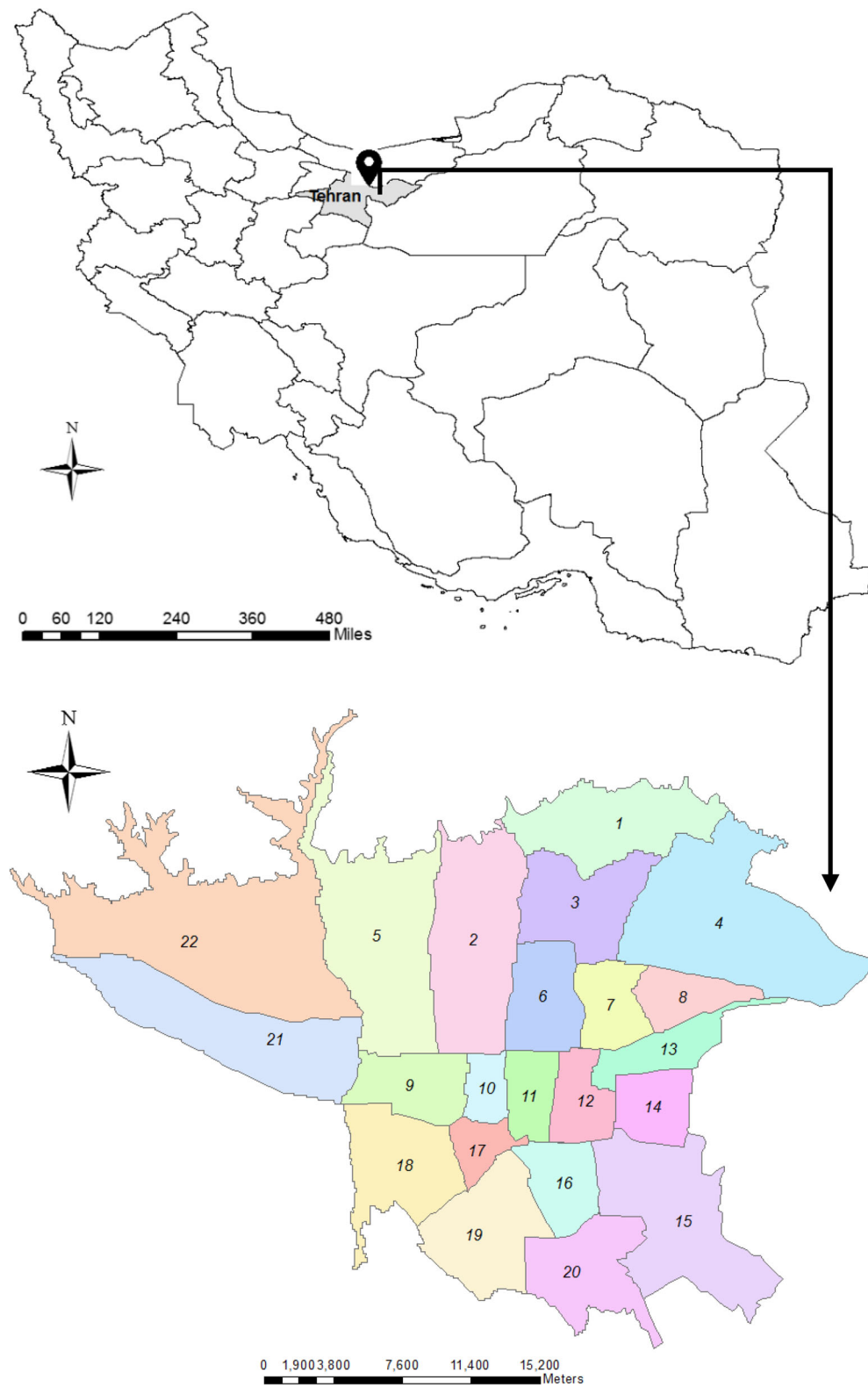


Fig. 1 The geographical location of the study area

everywhere, it is important for people with allergic asthma recognize their triggers and consider exactly how to avoid an attack (Romanet Manent et al. 2002).

Pneumonia

Pneumonia is an inflammatory infection of the lung, primarily affecting the small air sacs known as alveoli. Generally, symptoms of this disease include some combination of dry chest pain, cough, fever and breathing disorders. It can be caused by bacteria, viruses or fungi (Nhung et al. 2017).

Allergy

Allergy is affected by various influential environmental factors, with integrated environmental control being a key part of infection control operations. Researchers have shown that outdoor air pollutants, most significantly, ozone and PM, are related to increases in allergic reactions (Peden 2018; Guarnieri and Balmes 2014).

Bronchitis

Bronchitis is the inflammation of the large- and medium-sized airways in the lungs. Symptoms include coughing up mucus, wheezing, chest discomfort and truncation of breath. Bronchitis is separated into two categories: acute and chronic (Enevoldsen et al. 1991; Albert 2010).

Influenza

Influenza is a universally common health problem, which is a spreadable acute respiratory disease produced by the influenza virus with a widespread movement that happens every year. Influenza can cause epidemics, which can catastrophically affect people's lives and health and induce significant outages in a county's economy (Sun et al. 2016).

LBSN

A social network is a group of entities, or nodes, where states cooperate and compete with each other as links. These entities can be defined as information, social connections and locations. Information entities, in on-line social networks, can be represented as nodes and links. These links can be used to define the relationships

between them (Neisani Samani et al. 2018; Zhang et al. 2013). An LBSN is a specific type of social network, where users are able to share their location information with friends as "check-ins" (Abbasi and Alesheikh 2018).

An LBSN is also referred to as a geographic social network or a geosocial network, which is formally defined as a specific type of social networking platform, in which geographical services complement traditional social networks. This additional information enables new social dynamics, including those derived from visits by users to the same or similar locations, in addition to knowledge about common interests, activities and behaviours inferred from the set of places visited by a person and the location-tagged data generated during these visits (Valverde-Rebaz et al. 2018).

Proposed method

The main idea of this paper is to identify the relationships between environmental, infrastructural factors and respiratory diseases. The proposed methodology is depicted in Fig. 2.

Determination of effective environmental factors behind the risk of respiratory diseases

The initial step starts by determining all of the effective factors behind respiratory disease exacerbation. For this purpose, previous studies were investigated (Joseph et al. 2013; Olivieri and Scoditti 2005; Trang and Tripathi 2014) and a questionnaire was designed. Twenty-five experts in medical and environmental sciences and urban disciplines completed the questionnaires; the final factors are specified in Table 1 (Joseph et al. 2013; Olivieri and Scoditti 2005; Trang and Tripathi 2014).

Designing an LBSN

In order to design an LBSN, it is necessary to define a user profile for each patient. Accordingly, four mandatory information fields were designed and presented to the users as well as other arbitrary fields for further information. The mandatory fields were the following: (1) name (first/last), (2) national ID, (3) types of respiratory disease and (4) time of the last attack. The optional fields included (1) age, (2) gender, (3) job, (4) name of the drug used and (5) smoking status. Upon

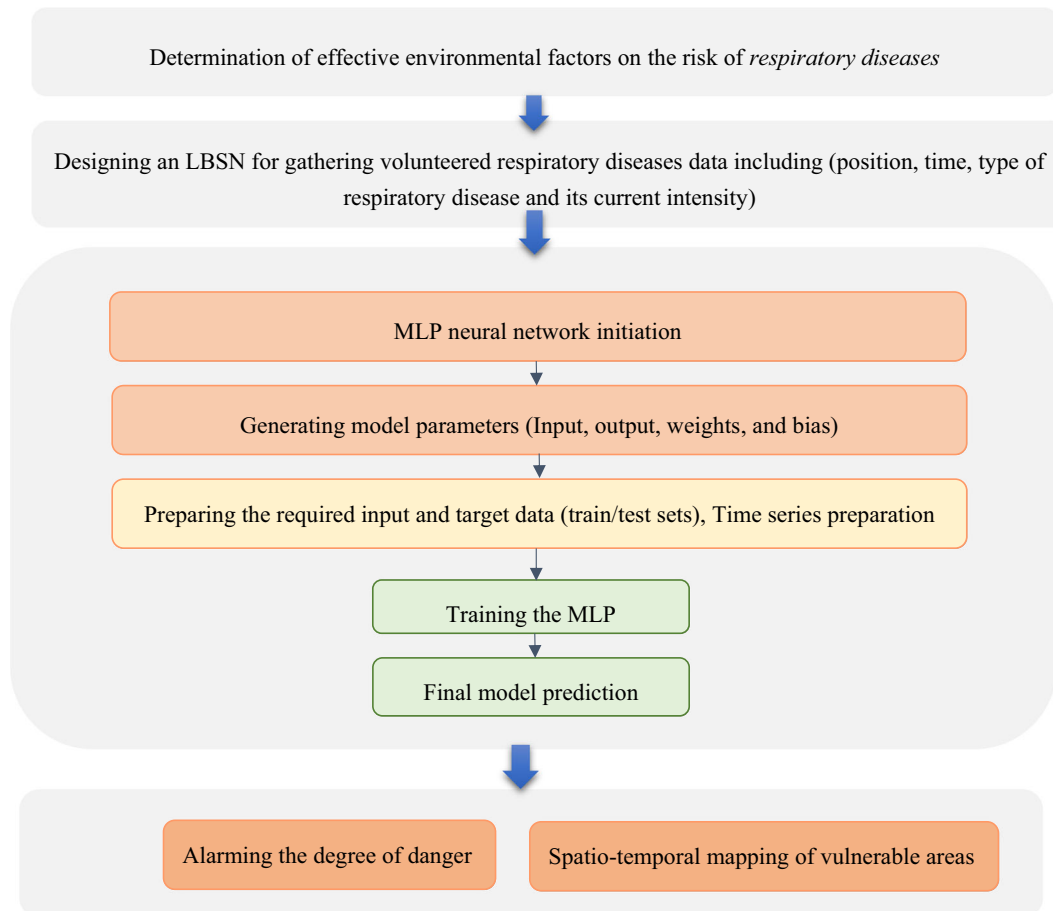


Fig. 2 The flowchart of the proposed method

completion of the user profile, mandatory data were fixed, except for ‘time of the last attack’, which had to

Table 1 Effective environmental and infrastructural factors on respiratory disease exacerbation

Factors	Properties
Infrastructural (physical)	Distance from green spaces
	Distance from pollution resources
	Land use
	Elevation
Environmental	Traffic congestion
	Types of air pollution
	Temperature
	Wind speed
	Relative pressure
	Rainfall
	Humidity

be updated by the patient. The users’ profile data formed the static data. On the other hand, dynamic data collection was also performed. The dynamic data were (1) position of the mobile user during the respiratory attack, (2) time of the respiratory attack, (3) amount of exposure time to air pollution (travel time) and (4) mode of travel (walking/driving). The remaining details for the LBSN are explained in the “MLP network” section.

MLP network

The MLP is one of the most effective ANNs used for modelling and prediction and has thus been used as the benchmark model in many types of research (Bui et al. 2017). The MLP has a high capability for universal approximation, which is why it has been extensively used for modelling non-linear and complex processes in the real world (Pham et al. 2016). Basically, the MLP is a feed-forward model comprising one input layer, one or more

hidden layers and one output layer (Sadowski et al. 2018) (Fig. 3).

The MLP procedure is as follows (Beale et al. 2018):

Suppose is that the desired output for pattern p at node/neuron j and is that the actual output of j . Here, is the weighting factor for the edge from neuron i to neuron j .

1. Set all weights and thresholds to small random numbers.
2. Introduce desirable inputs as $= o, x_1, 2, \dots, -1$ and the target output as $= o, 1, 2, \dots, 1$ to the network. The weighting factor $_o$ is set to the negative threshold value $-\theta$, and $_o$ is set to 1.
3. Compute the output: each layer calculates and sends it to the next layer as shown in Eq. 1.

$$y_{pj} = f \left[\sum_{i=0}^{n-1} w_i x_i \right] \quad (1)$$

The output of every unit j is obtained by applying the edge perform f on the weighted total of inputs for that unit. As shown in Eq. 2, weights are updated, beginning from the outer layer and moving backwards.

$$W_{ij}(t+1) = W_{ij}(t) + \eta \delta_{pj} o_{pj} \quad (2)$$

$W_{ij}(t)$ represents the weighting factors
 i, j node
 t, η time
 δ_{pj} shows the error of p at j

Equation 3 is used for the outer-layer units since both the target output and the actual output are known.

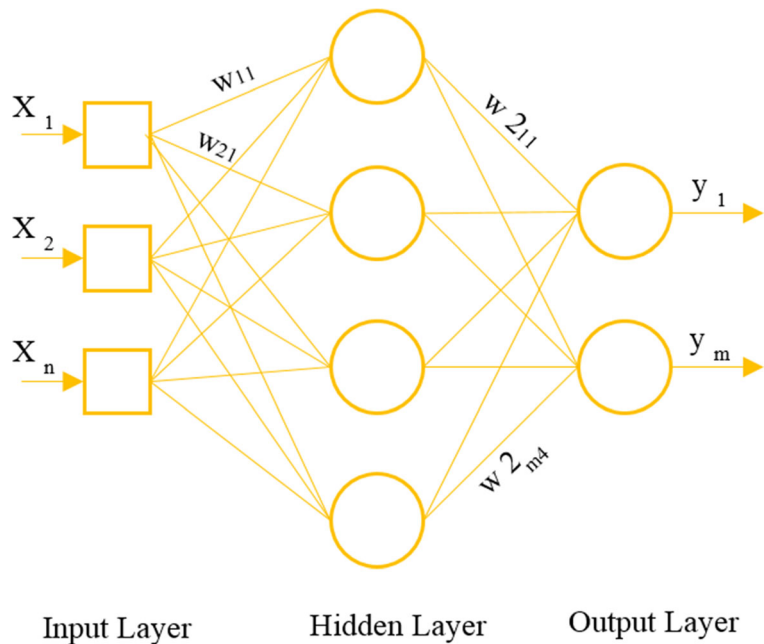
$$\delta_{pj} = k o_{pj} (1 - o_{pj}) (t_{pj} - o_{pj}) \quad (3)$$

As a result, the error for units in the hidden layer is calculated as per Eq. 4:

$$\delta_{pj} = k o_{pj} (1 - o_{pj}) \sum_k \delta_{pk} W_{jk} \quad (4)$$

Meanwhile, the summation is operated for the k units in the layer next to j . Training parameters, such as learning rate, momentum and training epoch, are set empirically. In the forward stage, the value of each hidden node is calculated as the sum of multiplications between the values of input neurons and corresponding connection weights. Then, the output value is propagated forward from the values of the hidden nodes and the connection weights in the same way. Comparing the output response values to target values will estimate the difference which can be minimized in the backward stage (error propagation) by updating the connection weights. In this paper, the input and the target of the model are considered, as in Fig. 4.

Fig. 3 The architecture of a multilayer perceptron network (Sadowski et al. 2018)



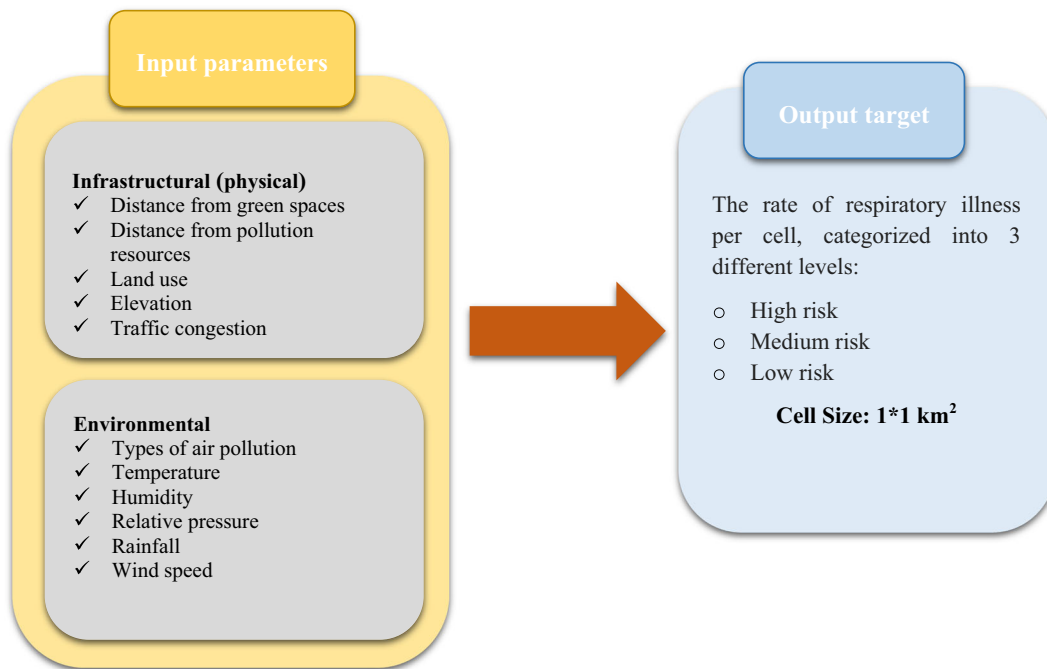


Fig. 4 The input parameters and output targets

Implementation and experimental results

This section describes all implementation details as well as the method of evaluation and its results. Finally, a thorough analysis of the outcomes of the proposed method and implementation procedure are discussed.

Table 2 Required data for each factor

Data type		Factor
Satellite images/sensors	Landsat	Land use
	ASTER	DEM (30*30) m ²
Vector maps	Urban map (1:2000)	Distance from green spaces
		Distance from pollution Resources
Field Data	Ground stations data	Wind speed
		Relative pressure
		Rainfall
		Humidity
		Air pollution
		Temperature
	On-line traffic sensors	Traffic congestion

Required data

Two main categories of data have been used in this study. The first one is related to the patient's information and the second one includes data on environmental and infrastructural factors. The former was explained in the "System architecture" section and the latter is summarized in Table 2. Also, some related maps according to the effective criteria and the related data are depicted in Fig. 5. These maps are generated using spatial analyst like Kriging for interpolation and Raster Distance for distance from green space.

System architecture

The architecture of the proposed system was designed in three physical layers (Fig. 6). The client side involves the Telegram application interface. This layer collects data and sends them to the database on the server side. It was designed in Visual Studio 2015 (C#).

The server application layer was used as an intermediate layer with the task of collecting geotagged photos via a step-by-step wizard. In this layer, the Telegram application programming interface (API) was applied to implement the bot. The server-side layer was used as the data layer for the storage and management of data. The

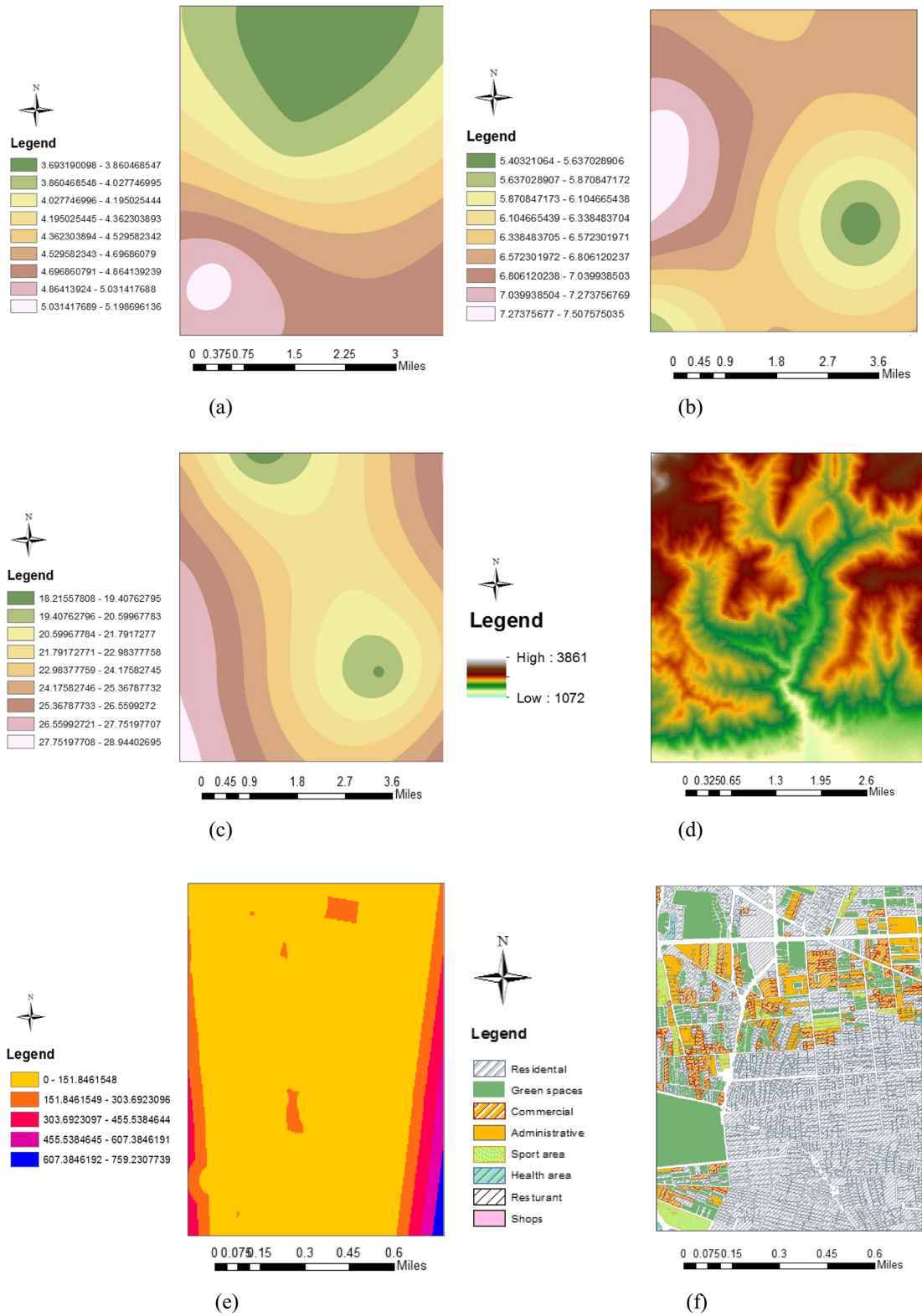


Fig. 5 Some maps of the effective criteria: **a** Wind map, **b** rainfall map, **c** temperature map, **d** elevation map (DEM), **e** distance from green space, and **f** land use map

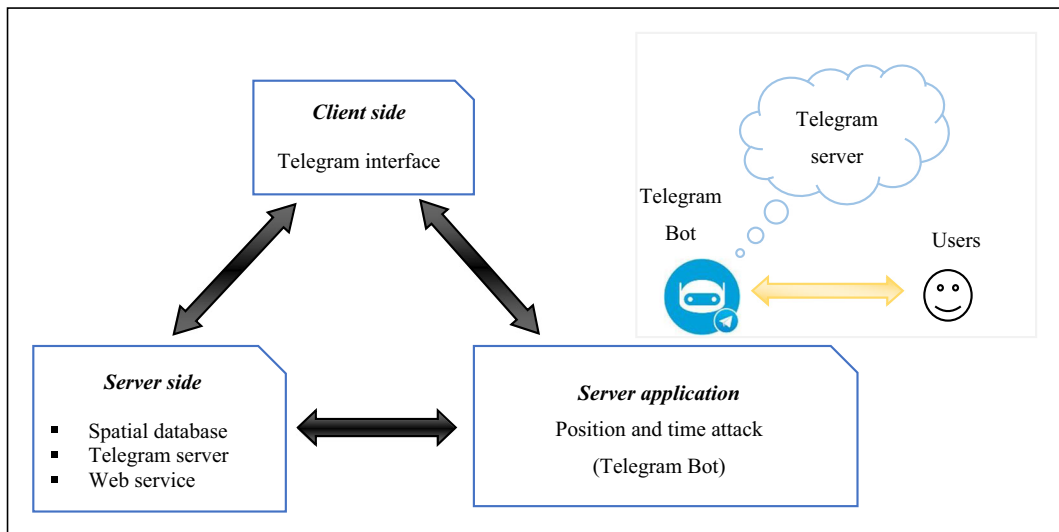


Fig. 6 System architecture for data collection

layer consists of three sets: Web service (designed with the PHP 5.2.0 programming language), Telegram server

and spatial database software programmed in PostgreSQL 9.5 and PostGIS 2.2.

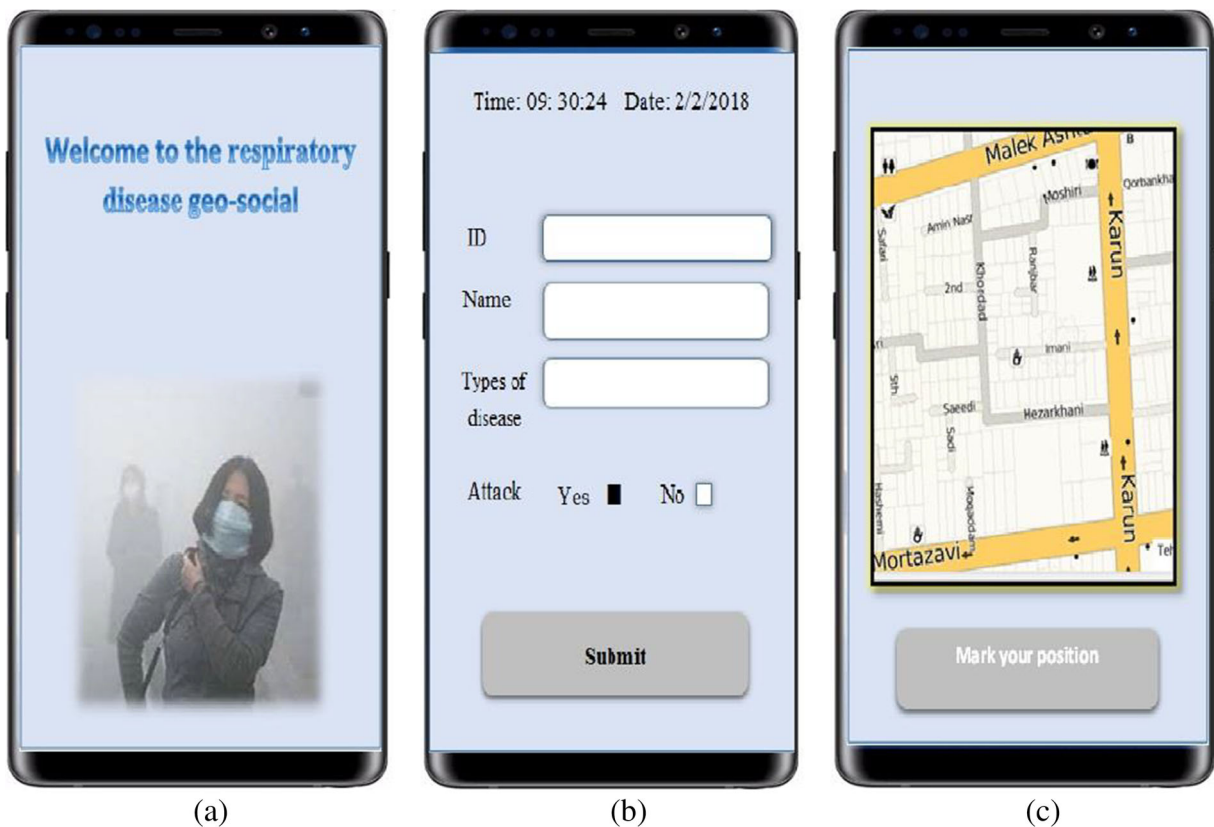


Fig. 7 The configuration wizard of the application: **a** Welcome page, **b** user profile submission and **c** Marking the attack position

Table 3 Validation results for developed models

Number of hidden layers	RMSE	R	Precision	Recall	F-Score
4*10 ANN	0.035	0.81	0.96	0.86	0.91
17*15 ANN	0.028	0.84	0.95	0.81	0.89
17*10*5 ANN	0.017	0.91	0.952	0.81	0.86

LBSN graphical user interface

As the main part of the patient data collection process relates to the use of the LBSN, this highlights the need to design an appropriate graphical user interface. This paper sought to use one of the most convenient, easy-to-use and popular platforms, namely, the Telegram social media application and the Telegram bot. Figure 7 displays the configuration wizard for the designed LBSN.

The wizard configuration for the developed application, known as *@respiratoryassociation*, is as follows:

- First, the user starts the bot and then registers their information on the application by filling in their details on the profile information page (the profile items were described in the “Designing an LBSN” section) (Fig. 7c).
- In the next step, users indicate their geographical location with an accompanying description of the current conditions of their respiratory disease. In the case of changes in the user’s condition (during an attack or emergency situation), users indicate their latest position on the map. Otherwise, there is no need to indicate any other positions. It should be

noted that, if the user is outside the spatial extent of the study area, his/her positions are not marked.

- Upon termination of each excursion, the last page is illustrated and the travel duration time is presented.

MLP training and validation

The first step in data preparation for ANN modelling is to classify the input data into two sets: the training set and the test and validation set. The test data must be representative of the main data set. Here, 70% of the main data sets for each input variable (732*13 matrices) were randomly selected as NN training data, 15% as testing data and 15% as validation data. A key problem in NN modelling is to find the optimal number of neurons. In this study, the number of network layers and neurons in the hidden layer was experimentally altered from the smallest possible size to the largest possible size. In every case, the error between the required output and therefore the real output was calculated. The results obtained from network training are shown in Table 3. The designed MLP NN was developed in the MATLAB environment as M-files. To this end, the layers created in GIS were

Fig. 8 Comparison of the accuracy of designed MLP neural networks

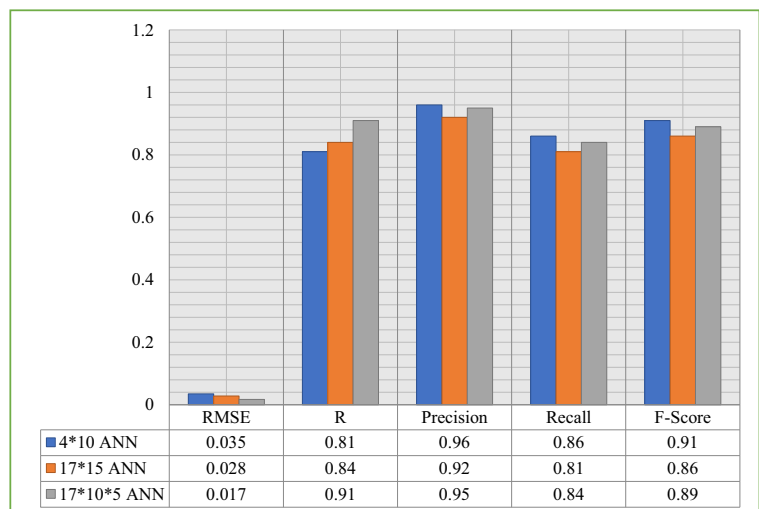
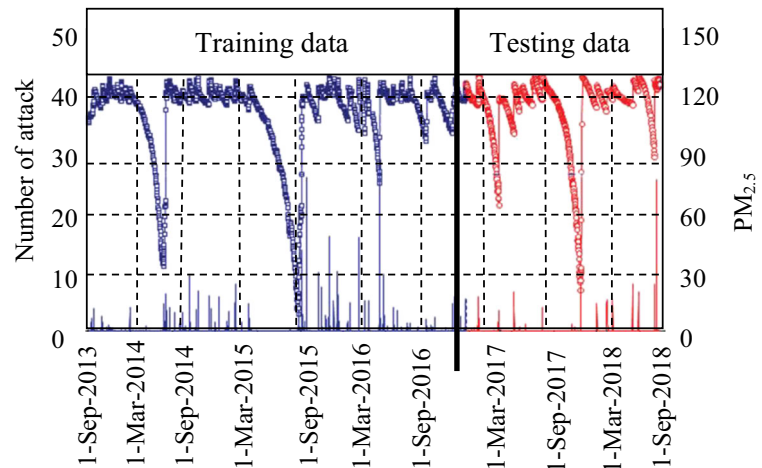


Fig. 9 A diagram of the time series of PM_{2.5} and the number of attacks



transferred to the MATLAB workspace as ASCII files. The inputs and outputs were then defined and categorized as training, testing and validation data. In the next step, the network was trained and, finally, the error was calculated based on the network output. To train the NN, an error back-propagation technique known as the Levenberg-Marquardt algorithm was used. This algorithmic program falls into the class of supervised learning (Cabaneros et al. 2017).

Sigmoid and linear functions were used as the transfer functions for the hidden layer and the output layer, respectively. The minimum error threshold was set to 0.005. Weights of neurons in the case of minimum error for testing data were selected as final weights.

To assess the accuracy and precision of the designed ANN, correlation (R) and the RMSE were calculated (Eqs. 5 and 6). The validation results are depicted in Table 3 and compared in Fig. 8.

$$R = \frac{\sum (Y_i - \bar{Y}) \sum (\hat{Y}_i - \bar{\hat{Y}})}{\sqrt{\sum (Y_i - \bar{Y})^2 \sum (\hat{Y}_i - \bar{\hat{Y}})^2}} \quad (5)$$

$$\text{RMSE} = \sqrt{\frac{\sum (Y_i - \hat{Y}_i)^2}{n}} \quad (6)$$

where Y and $\hat{Y}$ are the target and the output of the MLP, respectively.

Also, three other metric including precision, recall and F -score were used in this study to evaluate the accuracy of the proposed method (Barman and Tewari 2017). Precision is defined considering the number of

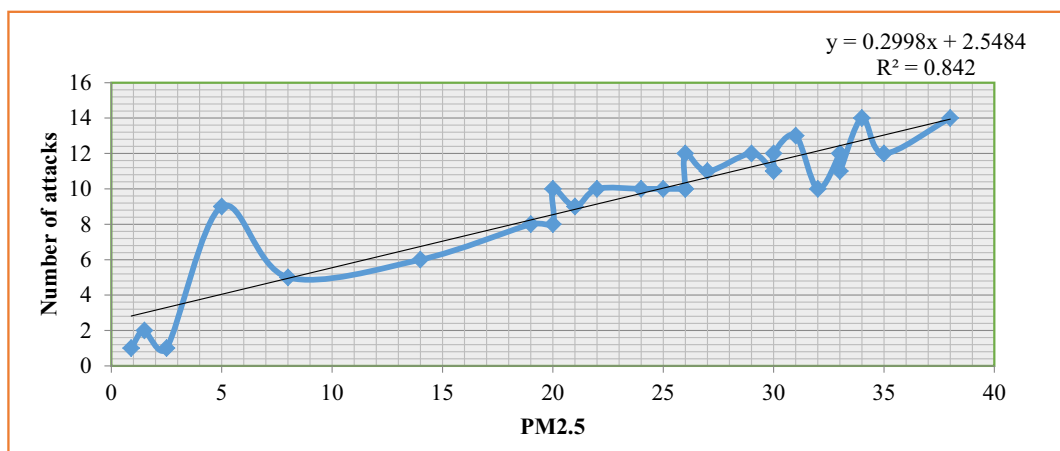


Fig. 10 Cross-correlation between the amount of PM_{2.5} and number of attacks

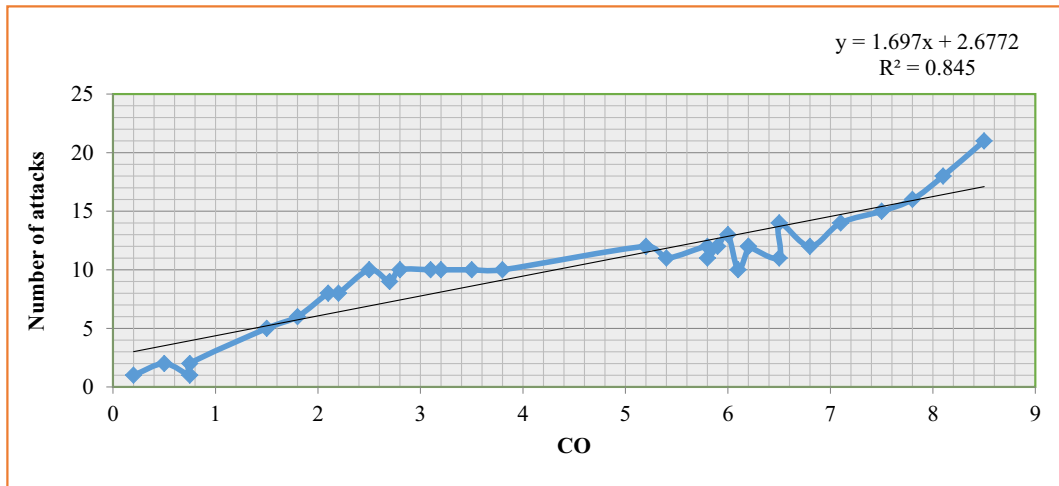


Fig. 11 Cross-correlation between the amount of CO and the number of attacks

Fig. 12 Temporal vulnerability maps (in 2017): **a** risk maps in Feb, **b** risk maps in May, **c** risk maps in July and **d** risk maps in December



Table 4 Comparison of the proposed approach with previous studies

Characteristics	Environmental factors	Infrastructural factors	Using LBNS	Using ANN
Lee et al. (2011)	✓	-	-	-
Gorai et al. (2014)	✓	-	-	-
Adams and Kanaroglou (2015)	-	Traffic information	-	✓
Polezer et al. (2017)	✓	-	-	✓
Kaffash Charandabi et al. (2019)	✓	-	-	✓

true positives and false positives as formulated in Eq. (7), wherein true negatives represent relevant attack position and false positives indicate irrelevant attack position that were detected.

$$\text{Precision} = \frac{\text{True positives}}{\text{True positive} + \text{False Positives}} \quad (7)$$

Moreover, recall is calculated using Eq. (8), where the true positives are the relevant tourism sites and the false negative indicate relevant places that were not recommended.

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False Negatives}} \quad (8)$$

The F -score metric is applied to get a suitable combination of both precision and recall, as formulated in Eq. (9).

$$F\text{-Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{recision} + \text{Recall}} \quad (9)$$

According to the comparison between the designed MLP NNs, a network with 17*10*5 neurons was the most suitable network architecture for this study.

To investigate the applicability of the model for the prediction of the spatial and temporal variability of respiratory attacks, the model was tested regarding the time series data on different environmental and infrastructural factors. In order to train the data set with the MLP, first, all the collected data from a 5-year period from December 2013 to December 2018 had to be pre-processed. In this regard, a time series for each element was adjusted with various temporal resolutions. The minimum interval (resolution) between two attacks was 68 s and the maximum temporal resolution was 84 h, while the mean resolution was 4.25 h. According to the temporal record, the time series was prepared (see Fig. 9 as an example).

In order to clarify the data, some examples of the cross-correlation function between the number of attacks and effective parameters, such as PM2.5 (Fig. 10) and CO (Fig. 11), were considered.

In order to clarify the data, some examples for cross-correlation function (CCF) between the number of attacks and effective parameters such as PM2.5 (Fig. 10) and CO (Fig. 11).

Temporal risk maps for respiratory diseases

Temporal risk maps for respiratory disease attacks were obtained by considering the designed MLP NN. Figure 9 shows four different risk maps for the study area in 2017. As can be seen, the risk areas change according to the dynamic environmental and infrastructural factors. Furthermore, the obtained results indicate that the most decisive factor is air pollution, especially pollution from CO and PM. The amount of exposure to the polluted area was ranked as the second most decisive factor, which escalated with increased traffic jams. The third most decisive factor was found to be land use.

Discussion

This paper aimed to investigate the impact of environmental and infrastructural factors on respiratory diseases. The main contribution of the proposed method is the assessment of all effective environmental and infrastructural parameters in relation to different types of respiratory attacks. One of the main challenges of this field is the collection of patient-related data, which was resolved through the design of an LBSN application. This method of data collection enables managers to identify temporal risk zones and alarm patients whenever the risk is going to escalate to dangerous levels. The other contribution of this paper is related to the usage of

an MLP NN as one of the promising methods for identifying the relationships between input data (environmental and infrastructural factors) and output targets (the temporal positions of respiratory disease attacks). This paper also revealed the results in the vector data model to be a statistical block, while most types of geosocial media apply a raster data model which is less accurate than the vector mode (Fig. 12).

Despite the mentioned advantages, the proposed method does have certain limitations. The first one is related to the type of learning method applied to the MLP. In the case of an increase in the volume of data changes in their patterns, certain deep learning methods seem more suitable. The main drawback to this study concerns insufficient data from certain regions, which was primarily due to inconveniences and difficulties experienced by some older patients in using the LBSN or perhaps the lack of volunteers registering with the required data. Although the proposed method achieved a high degree of accuracy in the risk zonation of the study area, it could be further enhanced by adding other patient information.

A comparison of the results with other previous studies is described in Table 4. As shown in related works, LBSN data were not used. Further, no research has considered all the environmental and infrastructural factors.

Conclusion and future works

This paper attempted to answer the questions ‘How we can model the spatio-temporal variability of respiratory diseases and attacks?’ and ‘How will we tend to rank the effective environmental and infrastructural parameters so as to predict the position and time of metabolism attacks?’. We found that an understanding of the spatial distribution of factors that affect respiratory disease exacerbation is vital. One of the main challenges in this field is the collection of data concerning the geographical position of patients with respiratory diseases. Initially, this study developed a Telegram API to register information about patients on a voluntary basis using an LBSN. After the data collection process was carried out, over a period of 5 years using eight different spatio-temporal data sets (according to the seasons), 732 patients subscribed to the LBSN API. They entered their personal information as well as geographical locations.

In the second stage, MLP NNs with three different architectures were improved. The input data for all networks included distance from green spaces, distance from pollution resources, land use, elevation, economic and social status, types of pollution, temperature, humidity and traffic. The output target was vulnerability, according to three levels: ‘highly vulnerable’, ‘moderately vulnerable’ and ‘low vulnerability’. Validation results revealed that the 17*10*5 ANN architecture outperformed other models, such as the 4*10 and the 17*15 models; however, there was no significant difference between the models in terms of precision. This indicates that other models can also be used to obtain vulnerability maps for respiratory patients.

The obtained results show that the proposed method is able to model the spatio-temporal variability of respiratory diseases. They also indicate that the most effective factors are, in order of decisiveness, pollution, especially CO and PMs in the air, the amount of exposure to the polluted area, and land use. Moreover, the model was also able to perform a spatio-temporal prediction of the vulnerable zones. As a future contribution, an evaluation of the usability of the designed LBSN is suggested, in order to create a more suitable mobile application. Implementing GIS-based multi-criteria decision-making methods and comparing them with ANNs should also be considered in future investigations.

References

- Abbasi, O. A., & Alesheikh, A. A. (2018). Exploring the potential of location-based social networks data as proxy variables in collective human mobility prediction models. *Arabian Journal of Geosciences*, 11(173), 1–14.
- Adams, M. D., & Kanaroglou, P. S. (2015). Mapping real-time air pollution health risk for environmental management: combining mobile and stationary air pollution monitoring with neural network models. *Journal of Environmental Management*, 168, 133–141.
- Albert, R. H. (2010). Diagnosis and treatment of acute bronchitis. *American Family Physician*, 82(11), 1345–1350.
- Barman, A., Tewari, A. (2017). Collaborative Recommendation System Using Dynamic Content based Filtering, Association Rule Mining and Opinion Mining. *International Journal of Intelligent Engineering & Systems* 10:57–66. <https://doi.org/10.22266/ijies2017.1031.07>.
- Beale, M. H., Hagan, M. T., & Demuth, H. B. (2018). Neural Network Toolbox - User's Guide. <https://www2.cs.siu.edu/~rahimi/cs437/slides/nnet.pdf>. Accessed 01.10.19.
- Blanco, I., Diego, I., Bueno, P., Fernández, E., Casas-Maldonado, F., Esquinas, C., Soriano, J., & Miravittles, M. (2018). Geographical distribution of COPD prevalence in Europe,

- estimated by an inverse distance weighting interpolation technique. *International Journal of Chronic Obstructive Pulmonary Disease*, 13, 57–67. <https://doi.org/10.2147/COPD.S150853>.
- Bui, D. T., Bui, K. T. T., Bui, Q. T., Doan, C. V., & Hoang, N. D. (2017). Hybrid intelligent model based on least squared support vector regression and artificial bee colony optimization for time series modeling and forecasting horizontal displacement of hydropower dam. In *Handbook of neural computation* (1st ed.). Cambridge: Academic Press (Elsevier).
- Cabaneros, S. M., Calautit, J. K., & Hughes, B. R. (2017). Hybrid artificial neural network models for effective prediction and mitigation of urban roadside NO₂ pollution. *Energy Procedia*, 142, 3524–3530.
- Chen, J., Chen, H., Wu, Z., Hu, D., & Pan, J. Z. (2017). Forecasting smog-related health hazard based on social media and physical sensor. *Information Systems*, 64, 281–291.
- Drummond, M. B. (2014). Obstructive airway disease in urban populations. *Current Opinion in Pulmonary Medicine*, 20(2), 180–185.
- Gorai, A. K., Tuluri, F., & Tchounwou, P. B. (2014). A GIS based approach for assessing the association between air pollution and asthma in New York state, USA. *International Journal of Environmental Research and Public Health*, 11(5), 4845–4869.
- Guarnieri, M., & Balmes, J. R. (2014). Outdoor air pollution and asthma. *Lancet*, 383(9928), 1581–1592.
- Ho, W. C., Hartley, W. R., Myer, L., Lin, M. H., Lin, Y. S., Lien, C. H., & Lin, R. S. (2007). Air pollution, weather, and associated risk factors related to asthma prevalence and attack rate. *Environmental Research*, 104(3), 402–409.
- Joseph, C. L., Ownby, D. R., Havstad, S. L., Saltzgaber, J., Considine, S., Johnson, D., Peterson, E., Alexander, G., Lu, M., Gibson-Scipio, W., & Johnson, C. C. (2013). Evaluation of a web-based asthma management intervention program for urban teenagers: reaching the hard to reach. *Journal of Adolescent Health*, 52, 419–426.
- Kaffash Charandabi, N., Alesheikh, A. A., & Sharif, M. (2019). A ubiquitous asthma monitoring framework based on ambient air pollutants and individuals' contexts. *Environmental Science and Pollution Research*, 26(8), 7525–7539.
- Khreis, H., Ramani, T., Hoogh, K., Mueller, N., Rojas-Rueda, D., Zietsman, J., & JNieuwenhuijsen, M. (2019). Traffic-related air pollution and the local burden of childhood asthma in Bradford, UK. *International Journal of Transportation Science and Technology*, 8(2), 116–128.
- Kumarihamy, R. M. K., & Tripathi, N. K. (2019). Geostatistical predictive modeling for asthma and chronic obstructive pulmonary disease using socioeconomic and environmental determinants. *Environmental Monitoring and Assessment*, 191, 1–21. <https://doi.org/10.1007/s10661-019-7417-0>.
- Lee, C. H., Ch, Y. J., Vincent, C., & Tseng, S. (2011). A novel data mining mechanism considering bio-signal and environmental data with applications on asthma monitoring. *Computer Methods and Programs in Biomedicine*, 101(1), 44–61.
- Neisani Samani, Z., Karimi, M., & Alesheikh, A. A. (2018). A novel approach to site selection: collaborative multi-criteria decision making through geo-social network (case study: public parking). *ISPRS International Journal of Geo-Information*, 7, 82.
- Nhung, N. T. T., Amini, H., Schindler, C., Kutlar Joss, M., Dien, T. M., Probst-Hensch, N., Perez, L., & Künzli, N. (2017). Short-term association between ambient air pollution and pneumonia in children: a systematic review and meta-analysis of time-series and case-crossover studies. *Environmental Pollution*, 230, 1000–1008.
- Olivieri, D., & Scoditti, E. (2005). Impact of environmental factors on lung defences. *European Respiratory Review*, 14, 51–56. <https://doi.org/10.1183/09059180.05.00009502>.
- Peden, D. B. (2018). Impact of pollution on allergy/immunology. *Journal of Allergy and Clinical Immunology*, 141, 3.
- Pham, B. T., Bui, D. T., Prakash, I., & Dholakia, M. B. (2016). 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.
- Polezer, G., Tadano, Y. S., Siqueira, H. V., Godoi, A. F. L., Yamamoto, C. I., de André, P. A., Pauliquevis, T., Andrade, M. F., Oliveira, A., Saldiva, P. H. N., Taylor, P. E., & Godoi, R. H. M. (2017). Assessing the impact of PM_{2.5} on respiratory disease using artificial neural networks. *Environmental Pollution*, 235, 394–403.
- Romanet Manent, S., Charpin, D., Magnan, A., Lanteaume, A., & Vervloet, D. (2002). Allergic vs nonallergic asthma: what makes the difference? *Allergy*, 57(2), 607–613.
- Sadowski, L., Hola, J., Czamecki, S., & Wang, D. (2018). Pull-off adhesion prediction of variable thick overlay to the substrate. *Automation in Construction*, 85, 10–23.
- Sanidas, E., Papadopoulos, D. P., Grassos, H., Velliou, M., Tsioufis, K., Barbetseas, J., & Papademetriou, V. (2017). Air pollution and arterial hypertension. A new risk factor is in the air. *Journal of the American Society of Hypertension*, 11(11), 709–715.
- Enevoldsen, K., Gheorghe, A., Moseholm, L., & Sørensen. (1991). The effect of low-level air pollution and weather on asthma and chronic bronchitis patients, studied by neural network methods. Artificial neural networks: proceedings of the 1991 International Conference on Artificial Neural Networks (ICANN-91), Espoo (Vol. 2, pp. 1371–1374).
- Sun, X., Ye, J., & Ren, F. (2016). Detecting influenza states based on hybrid model with personal emotional factors from social networks. *Neurocomputing*, 210, 257–268.
- Trang, N. H., & Tripathi, N. K. (2014). Spatial correlation analysis between particulate matter 10 (PM₁₀) hazard and respiratory diseases in Chiang Mai province, Thailand. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XL-8, 185–191.
- Valverde-Rebaz, J. C., Roche, M., Poncelet, P., & Lopes, A. A. (2018). The role of location and social strength for friendship prediction in location-based social networks. *Information Processing and Management*, 54, 475–489.
- World Health Organization, W. H. O (2017). The Global Impact of Respiratory Disease [Online]. Available: https://www.who.int/gard/publications/The_Global_Impact_of_Respiratory_Disease.pdf. Accessed 24.10.19.
- WorldAtlas. (2019). Location of iran. <https://www.worldatlas.com/webimage/countrys/asia/iran/irlatlog.htm>. Accessed 1 Mar 2019.
- Yun, T. J. (2012). Using ubiquitous communication technology to improve pediatric asthma management, College of Computing. (doctor thesis). Georgia Inst Technol.

- <https://smartech.gatech.edu/handle/1853/44794>. Accessed 19.10.19.
- Zhang, J., Kong, X., Yu, P. S. (2013). Predicting social links for new users across aligned heterogeneous social networks. ICDM. IEEE, 1289–1294.
- Zhou, L., Zhang, D., Yang, C. C., & Wang, Y. (2018). Harnessing social media for health information management. *Electronic Commerce Research and Applications*, 27, 139–151.

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