Taking a walk avoiding polluted routes: an application to a virtual coach for cancer

Cosimo Strusi, BS
Department of Electrical, Computer and Biomedical Engineering
University of Pavia
Pavia, Italy
cosimo.strusi01@universitadipavia.it

Arianna Dagliati, PhD
Department of Electrical, Computer and Biomedical Engineering
University of Pavia
Pavia, Italy
arianna.dagliati@unipv.it

Daniele Pala, PhD
Department of Electrical, Computer and Biomedical Engineering
University of Pavia
Pavia, Italy
daniele.pala@unipv.it

Cristiana Larizza, PhD
Department of Electrical, Computer and Biomedical Engineering
University of Pavia
Pavia, Italy
cristiana.larizza@unipv.it

Riccardo Bellazzi, PhD
Department of Electrical, Computer and Biomedical Engineering
University of Pavia
Pavia, Italy
riccardo.bellazzi@unipv.it

Silvana Quaglini, PhD
Department of Electrical, Computer and Biomedical Engineering
University of Pavia
Pavia, Italy
silvana.quaglini@unipv.it

2022 IEEE 21st Mediterranean Electrotechnical Conference (MELECON) | 978-1-6654-4280-0/22/$31.00 ©2022 IEEE | DOI: 10.1109/MELECON53508.2022.9843091

***Abstract*- Outdoor physical activity is known to be beneficial for both healthy and unhealthy people. Nowadays however, particularly in urban areas, air pollution may decrease this benefit. This paper describes the design and development of an application for optimizing outdoor physical activity with respect to exposure to pollution. We exploited an air quality sensors network, already existing in our city, and developed an algorithm to estimate pollution level in every possible route belonging to the city topology. The final users will be provided with a smartphone application, able to suggest them the best route for walking or biking. Although minimizing pollution exposure is the main goal, the app will also consider weather forecasting and some users' preferences. This work was motivated by the need of implementing a set of well-being interventions within a Horizon2020 project dedicated to cancer patients.**

Keywords—decision support, air pollution, optimal route calculation, cancer patients wellbeing

I. INTRODUCTION

Although cancer remains among the leading causes of death in the world, survival rate is increasing, and managing long-term homecare for cancer survivors is nowadays a major challenge for the healthcare systems. CAPABLE (CAncer PAtients Better Life Experience) is a European project funded under the Horizon 2020 program, which leverages on clinical practice guidelines and the integration of different data sources to support patients monitoring and increase their quality of life. One of the project tools is an app implementing a Virtual Coach (VC) that helps patients to manage the adverse events that may be caused by oncological treatments. This kind of electronic support for patient-reported-outcomes has already been proved to be effective in the cancer domain [1;2].

The CAPABLE's VC does not only support toxicity management from the physical point of view, but it aims at increasing well-being in a broader perspective. Therefore, a multidisciplinary team including psychologists, nutritionists and physical therapy experts, drew up a set of so-called *virtual capsules*, which are non-pharmacological interventions meant to improve both physical and mental wellbeing. Virtual capsules have been retrieved from the literature where their effectiveness has been proved [3;4]. Examples are "garden bowl" that asks the patient to take care of some plants, "vase of gratitude" that requires writing some nice words to somebody, and "my usual walk" that invites the patient to go outdoors for a light physical activity. The latter represents the focus of this paper, which describes a support provided to the patient to go for a walk avoiding, as much as, possible polluted routes.

II. BACKGROUND

While it is established that physical activity is beneficial for cancer survivors, such benefits could be reduced if it is performed in locations with poor air quality, also considering that breathing is increased during (even mild) physical efforts [5]. Air quality may be affected by several pollutants. Among them, particulate matter (PM) or particulates, are microscopic solid or liquid matter suspended in Earth atmosphere. The different PMs are defined by their diameter: PM_{10} (<=10 μm), $PM_{2.5}$ (<=2.5 μm), and PM_1 (<=1 μm). Lower the diameter, higher the potential to penetrate deep into the lungs and even to enter the bloodstream, so primarily resulting in cardiovascular and respiratory impacts, but also representing a risk factor for cancer [6].

Given the importance of the environmental issue, a wide literature exists about how to monitor and limit the impact of pollution on the human health. Air quality sensor networks exist in a number of towns [7], and different types of interventions have been proposed, such as increasing greenness [8], promoting dietary supplements [9], increasing awareness and alert to target populations [10].

The review in [11] describes the main personal interventions intended to reduce outdoor exposure, such as are avoiding air pollution sources, staying indoors, filtering indoor air, reducing physical activity, and using respirators or other face masks.

Recently, also Google announced an extension of its maps to show "eco-friendly routes", i.e, in addition to showing the fastest route, Google Maps will also display the one that is the most fuel-efficient, if it doesn't happen to also be the fastest (https://www.google.com/earth/outreach/special-projects/air-quality/, last access Apr 8th, 2022).

978-1-6654-4280-0/22/$31.00 ©2022 IEEE

However, to the best of our knowledge, there isn't any application that integrate air pollution data with a patient's preferences to suggest the best route for a walk.

III. GOAL SPECIFICATION

As mentioned, the goal of this work is to implement a novel functionality of the VC that suggests walking avoiding polluted routes. In other words, when the VC reminds a patient to accomplish the virtual capsule "My usual walk", the patient is prompted with a route characterized by the best air quality in the area. However, there are other features that should be considered to enhance the user experience with that capsule. In particular:

- The weather. It may not be ideal for a stroll at the moment the VC sends the advice. Thus, if the patient knows whether or not it's improving, he/she can take an informed decision such as waiting for the weather to get better.
- The (possible) desired destination. A patient might have to reach a certain destination (e.g., for going to work, for visiting a friend, etc.) and he might take the chance to have his daily walk.
- The means. Both walking and biking could be useful for performing a beneficial physical activity.
- The walking distance: a patient should be able to set a constraint on a minimum and maximum distance he desires to walk.
- The walking time: a patient must be able to set a constraint on a time span he desires to walk for.

At the moment, we implemented a proof of concept considering the first three requirements.

IV. METHODS

A. *The technical architecture*

Our proof of concept exploits a client-server architecture, shown in Fig. 1.

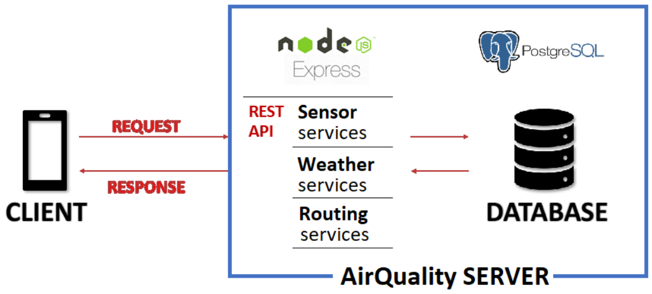

Fig. 1. The client-server architecture of the application

The server, called *AirQuality*, contains a PostgreSQL relational database storing information about sensor data, meteorological data, and city topology, since an integration of all these data is necessary to our purposes. The database is enabled with a PostGIS extension (https://postgis.net/, last access Apr 8th, 2022) necessary to manage geographic objects and run location queries in SQL and a PGRouting extension (https://pgrouting.org/, last access Apr 8th, 2022) which provides geospatial routing functionality.

Client/server communication is done through RESTful API (REpresentational State Transfer Application Program Interface) [12] served by Express, a Node.js web application framework, while Flutter (https://flutter.dev/, last access Apr 8th, 2022) is used for the implementation of the app front-end. It allows creating applications for different operating systems (e.g., Android, iOS) without developing the code in their native language.

Data sources populating the database are described in the following sections

B. *Data Sources and pre-processing*

1) *Air Pollutants: Purple Air grid in Pavia*

In the past few years, also thanks to the PULSE[1] project, a relatively large number of low-cost PurpleAir brand air quality sensors (https://www2.purpleair.com, last access Apr 8th, 2022) has been deployed throughout the city. Currently, there are about 40 PurpleAir sensors installed at several public and private facilities, and their location is shown in Fig. 2. They were added to the two preexisting air monitoring stations installed by the governmental agency ARPA, thus creating a dense air quality monitoring network. This allows performing rather precise estimates of pollution levels in the whole city, without neglecting possibly relevant local variations.

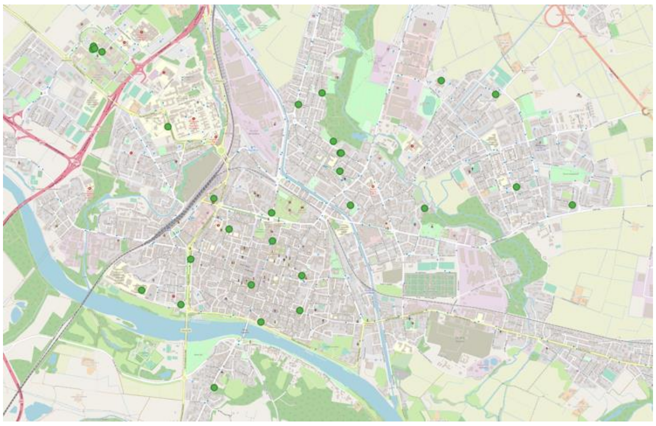

Fig. 2. Pavia map and PurpleAir sensors location

The sensors use a particle counter based on laser technology to measure PM$_{10}$, PM$_{2.5}$ and PM$_{1}$. They also measure temperature and relative humidity. Sensor measurements are done every 2 minutes, stored in the Purple Air cloud, and automatically uploaded into our *AirQuality* database that make them available to implement the necessary algorithms.

2) *Weather Service data*

Weather data are gathered through dedicated API requests made on the *OpenWeatherMap* portal (https://openweathermap.org, last access Apr 8th, 2022).

At the moment, requests are done only for Pavia, but more data can be easily acquired also for other locations. Requests are automated and repeated every 30 minutes, and data are stored in three tables in the *AirQuality* database, containing daily forecasts for the next 7 days, hourly forecasts for the next 48 hours and current condition, respectively. This allows users immediately accessing any weather information

---

[1] The EU H2020 PULSE project aimed at developing risk stratification models based on modifiable and non-modifiable risk factors in urban locations.

without connecting to OpenWeatherMap. The weather variables currently stored are temperature (current and daily range), relative humidity, atmospheric pressure, and weather textual description.

3) City Map Model and information about crossroads

Topological information of Pavia, concerning the road map of the city with its crossroads and streets, has been gathered to create the base for the *best road suggestion* algorithm. The main data source for this is OpenStreetMap (OSM), whose NorthWest Italy data has been accessed and downloaded through the Geofabrik portal (https://www.geofabrik.de/, last access Apr 8th, 2022).

The OSM data have then been processed using the osm2po tool (https://osm2po.de/), filtered in order to get only Pavia data and converted into a map, i.e. a graph representation, where nodes correspond to road intersections and arcs represent the road traits connecting them.

As for the other data, also the graph structure is stored in a set of relational tables in the *AirQuality* database, which contain an ID for each node/arc, with associated geographic information and additional data such as length, average travel speed in km/h and cost related to travel for the arcs, computed by osm2po on the basis of road length and average speed.

C. Implementation Pipeline: air pollution model and optimal path

1) Air Pollution Model: computing PM2.5 levels

Once the topology is created, air pollution can be inferred in each element of the graph through interpolation of the data measured by the Purple Air sensors scattered throughout the city. At this preliminary stage, this is performed using an algorithm called Inverse Distance Weighting (IDW) [13], that estimates the pollution value in a point *P* of the map as the weighted average of the measurements taken by the sensors using their distance as weights (Equation 1). In this way, the closer the sensor, the higher the importance given to its measurement.

Currently, the interpolation is performed on the 10-minute average of PM2.5.

Considering that the main interest of the project is finding a path along the roads of the city, air pollution should be estimated along the arcs rather than the nodes. As arcs, that represent streets or portions of streets, are not identified by a unique location, the interpolated value for a specific arc is estimated on the location situated at the middle point of the street. At the moment, this operation is performed only for $PM_{2.5}$.

$$z_P = \frac{\sum_{i=1}^{n} \frac{z_i}{d_i}}{\sum_{i=1}^{n} \frac{1}{d_i}} \qquad (1)$$

z_i : $PM_{2.5}$ value in i^{th} point
d_i : distance between point P and i^{th} point
z_P : $PM_{2.5}$ interpolated value in point P

Once the interpolation is computed, the "cost" parameter of the arcs is recalculated in terms of air pollution, assigning a value proportional to the pollution of the road (z_p) and the expected time of exposure (i.e. the travel time). More precisely, given the length L of the route and the estimated speed of the patient (S), the travel time T is calculated as T=L/S. The cost is then computed as T× z_p.

The speed S depends on the transportation mean (walking or biking) and the subject condition. In our proof of concept, we simulated a person walking at S=4.3km/h (slow pace, suitable for an oncological patient).

2) Dijkstra's algorithm to find the optimal path

Dijkstra's algorithm is a well-known approach to find the optimal path between two nodes in a graph according to a predefined objective function [14]. It calculates a potential in each node that represents the cost to reach it from a predefined starting node.

Before running the algorithm, considering that either the departure and arrival points could not strictly coincide with any graph elements, the closest arch is extrapolated for both points. Once the arcs have been identified, we can use their terminal nodes to run the algorithm.

Eventually, coordinates of the departure and arrival points are appended at the beginning and at the end of the path, respectively, in order to reconstruct the entire path.

Note that the algorithm also considers the mean of transportation chosen by the user, as it affects the time of exposure along the arcs.

V. RESULTS

A. The city topology and air pollution levels

Topological representation allows for the optimization of the volume of information that needs to be stored, as it includes only elements of a geographical area that are more likely to be frequented by the inhabitants for walks. This is an important feature for the integration of air pollutions maps into mobile applications.

Fig. 3 shows the city topology of Pavia obtained with the methodology described in section IV-A-3.

Fig. 4 shows an example of the city topology after running the IDW algorithm to interpolate the sensors measurements. Nodes are color-coded from white (low) to red (high) according to the interpolated pollution level, whereas sensors are highlighted in blue.

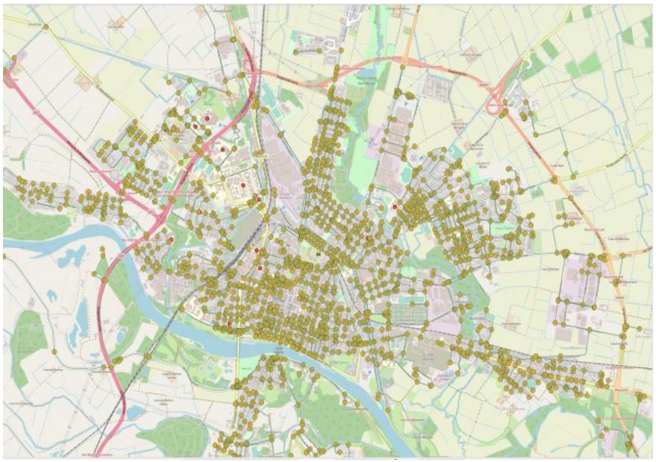

Fig. 3. The Pavia city topological representation

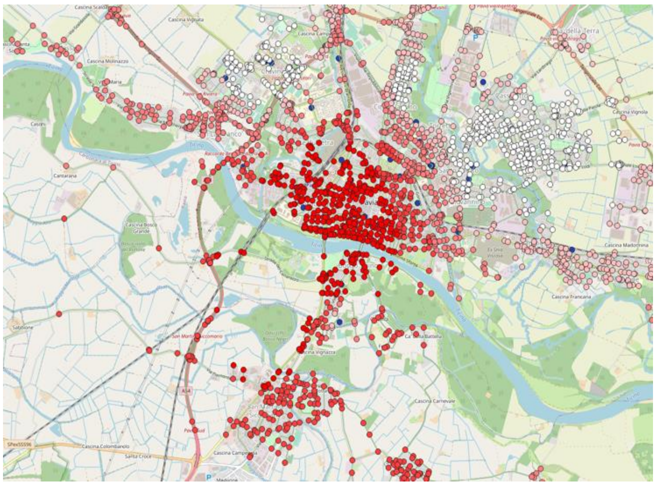

Fig. 4. Air pollution topological model – (data gathered on June 14th 2020).

B. *App functionalities*

As mentioned, a cancer patient using the CAPABLE system, could be invited to perform some physical activity, and in this case he will be redirected to the app. The app user interface is structured as of two pages that can be explored through a navigation bar. The first page is dedicated to weather and it contains the three elements that could be useful for the user to decide on the possibility and timing of his/her walk: (i) the current meteorological situation (temperature, weather description, humidity) and the sunrise and sunset times, in order to inform the user about the hours of light, (ii) a horizontal slider that shows the hourly forecast, (iii) a vertical slider that shows the forecast for the next seven days, namely the minimum and maximum temperature and the icon to describe the meteorological condition.

The second page (Fig. 5 left) shows the current position and allows the user entering any desired destination (facilitated by a Google auto-completion algorithm), together with the preferred means of transportation (by bike or on foot).

After pressing the "Show Route" button, the algorithm is run and the recommended, less polluted route is shown (Fig. 5 right). At this point, a button at the bottom-right of the app allows connecting to Google Maps, in order to start navigation on the calculated route.

C. *Preliminary validation*

A preliminary validation of the system has been performed on two features, namely (i) the air quality estimation at a given node, and (ii) the calculation of the optimal path.

Regarding the first one, a leave-one-out cross-validation at each node of the Pavia's sensor network has been implemented to evaluate the performance of the pollution prediction algorithm in a given time window (1 day, for a total of more than 25,000 measures). After cleaning data from outliers due to a couple of sensors malfunction, we obtained satisfactory results, as described by the following statistics (µg/m^3):

Median Observed Values= 10.11 (5.89:15.16);

Median Predicted Values= 10.24 (6.19:15.05).

Median PM$_{2.5}$ error=-0.02 (interquartile range= -0.49:0.47, Root Mean Square Error - RMSE=1.97);

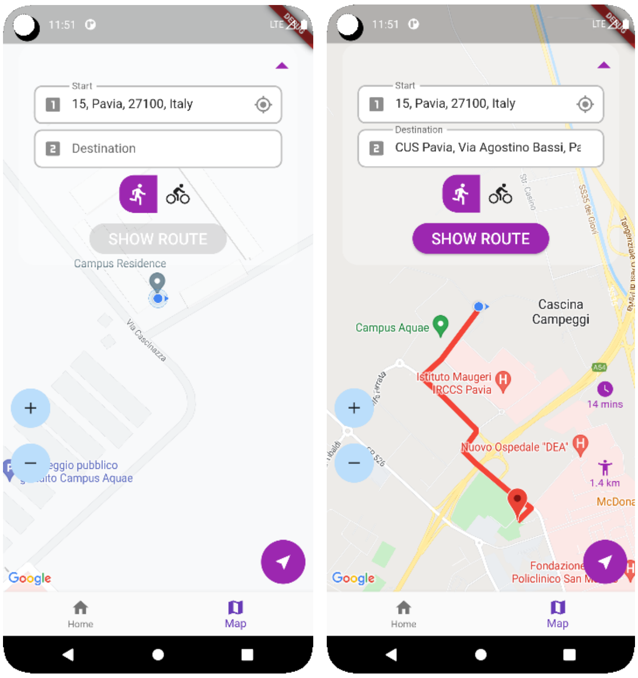

Fig. 5. The destination/means entry page (left) and the calculated optimal route (right)

Regarding the calculated paths, we qualitatively compared them with the paths suggested by Google Maps.

As an example, Fig. 6 shows the route suggested by Google maps (the red one) for going from A to S, which is slightly shorter, compared with the route recommended by our app (the blue one), which is less polluted, probably because it is closer to a river (shown in the left upper part of the figure) and it is less central than the red one.

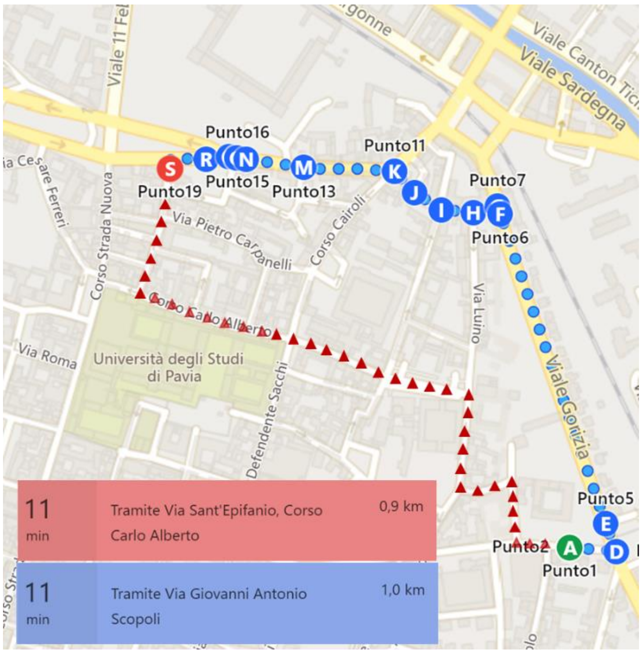

Fig. 6. Comparison between the route suggested by Google maps (in red, which is the shortest one) and the route suggested by our application (in blue).

VI. DISCUSSION AND FUTURE DEVELOPMENT

In this paper we presented the prototype of an application that will provide decision support to an individual for avoiding polluted areas when performing physical activity.

The prototype includes the full basic pipeline, from entering user's preferences to the calculation and visualization of the final path. However, as a work in progress, it shows several limitations.

First of all, further development is needed to (i) refine the optimization algorithm and (ii) improve the user experience. The first objective will be achieved by implementing interpolation methods able to include not only PM2.5 but also other pollutants, for example PM10, which is considered by various environmental authorities such as WHO and EU Air Quality Directive , when defining air quality indexes (https://ec.europa.eu/environment/air/quality/, last access Apr 8th, 2022).

Moreover, to increase the algorithm performance, we will divide the roads into smaller portions. We will assign the pollution value to these portions, so that the value of the complete road will be computed by a weighted average. Also the single values could be better estimated, with respect to the currently used IDW, by considering other variables such as real time data about traffic and prior information about urban morphology [15].

To improve the user experience, additional app features will be developed. In particular, the user will be able to choose the number of kilometers he wants to travel, the time he wants to spend, and preferences about places he would like to go through, for example a park, a river, and so on.

Moreover, we are also aware that this application will require a more careful validation. To this purpose, a trial will be performed. A set of volunteers will be equipped with a commercial, portable pollution sensors (e.g., Atmotube), and couples of volunteers will walk starting from a same point, and reaching a same destination, one following the app suggestion (case), and the other one following a different route (control). The pollution exposure in cases and controls, measured through the portable sensors, will be compared to check for a significant difference.

Eventually, the patient's compliance with the system suggestions will be measured, by adding two simple questions in the app, asking whether or not he/she accepted the suggestion and his/her perceived usefulness of the suggestion itself.

ACKNOWLEDGMENTS

This work has been partially funded by the EU Horizon2020 project CAPABLE #875052 and BRAINTEASER # 101017598.

REFERENCES

[1] A. Suleman, A. Vijenthira, A Berlin, A. Prica, D. Rodin. The Use of Virtual Care in Patients with Hematologic Malignancies: A Scoping Review. Curr Oncol. 2022 Feb 7;29(2):892-900

[2] F. Denis, E. Basch, A.L. Septans, J. Bennouna, T. Urban, A.C. Dueck, C. Letellier. Two-Year Survival Comparing Web-Based Symptom Monitoring vs Routine Surveillance Following Treatment for Lung Cancer. JAMA. 2019 Jan 22;321(3):306-307

[3] S.F. Fraser, J.R. Gardner, J. Dalla Via, R.M. Daly. The Effect of Exercise Training on Lean Body Mass in Breast Cancer Patients: A Systematic Review and Meta-analysis. Med Sci Sports Exerc. 2022 Feb 1;54(2):211-219

[4] D. Nesterova, J. Zhu, C. Kramer, M. Vasekar, C. Truica, A. Joshi, M. Hayes, J. Kessler, EFH Saunders, J.J. Drabick, M. Joshi. Group-led creative writing and behavioural health in cancer: a randomised clinical trial. BMJ Support Palliat Care. 2022 Mar;12(1):91-98.

[5] L.V. Giles, M.S. Koehle. The health effects of exercising in air pollution. Sports Med. 2014;44:223–49.

[6] H. Guo, J Liu, J Wei. Ambient Ozone, PM1 and Female Lung Cancer Incidence in 436 Chinese Counties. Int J Environ Res Public Health. 2021 Oct 2;18(19):10386.

[7] W.Y. Yi, K.M. Lo, T. Mak, K.S. Leung, Y. Leung, M.L. Meng. A Survey of Wireless Sensor Network Based Air Pollution Monitoring Systems. Sensors (Basel). 2015 Dec 12;15(12):31392-427

[8] L. Chen, J. Xie, T. Ma, M. Chen, D. Gao, Y. Li, Y. Ma, B. Wen, Jiang J, Wang X, Zhang J, Chen S, Wu L, W. Li, X. Liu, B. Dong, J. Wei, X. Guo, S. Huang, Y. Song, Y. Dong, J. Ma. Greenness alleviates the effects of ambient particulate matter on the risks of high blood pressure in children and adolescents. Sci Total Environ. 2022 Mar 15;812

[9] X. Wang, S. Li, Y. Wu, D. Huang, C. Pei, Y. Wang, S. Shi, F. Wang, Z. Wang. Effect of omega-3 fatty acids on TH1/TH2 polarization in individuals with high exposure to particulate matter ≤ 2.5 μm (PM2.5): a randomized, double-blind, placebo-controlled clinical study. Trials. 2022 Feb 24;23(1):179

[10] L.K. Tompkins, A.F. Pennington, K.D. Sircar, M.C. Mirabelli. Communication channels for receiving air quality alerts among adults in the United States. Prev Med Rep. 2021 Dec 23;25:101677

[11] R.J. Laumbach, K.R. Cromar. Personal Interventions to Reduce Exposure to Outdoor Air Pollution. Annu Rev Public Health. 2021 Dec 22.

[12] L. Richardson; S. Ruby (2007). RESTful Web Services. Sebastopol, California: O'Reilly Media

[13] D. Shepard. A two-dimensional interpolation function for irregularly-spaced data. In Proc. of the 1968 23rd ACM Natl Conf., ACM'68, New York, NY, 27–29 August, pp. 517–524.

[14] E. W. Dijkstra. A note on two problems in connexion with graphs. Numerische Mathematik, 1959, 1(1), 269–271.

[15] G. Maignant, and J. Dutozia, "Air Pollution and Urban Morphology: A Complex Relation or How to Optimize the Pedestrian Movement in Town", in Advanced Topics in Environmental Health and Air Pollution Case Studies. London, United Kingdom: IntechOpen, 2011 [Online]. Available: https://www.intechopen.com/chapters/, last access Apr 8th, 2022