Predictive Analytics for Emergency Department Visits Based on Local Short-Term Pollution and Weather Exposure

Isabella Della Torre
R&D Researcher
GPI SpA
Trento, Italy
email: isabella.dellatorre@gpi.it

Ismaela Avellino
R&D Researcher
GPI SpA
Trento, Italy
email: ismaela.avellino@gpi.it

Francesca Marinaro
R&D Researcher
GPI SpA
Trento, Italy
email: francesca.marinaro@gpi.it

Andrea Buccoliero
R&D Project Manager
GPI SpA
Trento, Italy
email: andrea.buccoliero@gpi.it

Antonio Colangelo
R&D Director
GPI SpA
Trento, Italy
email: antonio.colangelo@gpi.it

***Abstract*-Proper management of Emergency Rooms is needed to improve healthcare and patient satisfaction. Predicting accesses and hospitalisation rates through Machine Learning approaches appears promising, especially when coupled with air pollution and weather data. This work applies both Random Forest and AutoRegressive Integrated Moving Average approaches on data related to Brescia's clinical and environmental data from 2018 to 2022 to predict daily accesses or daily hospitalisations for cardiovascular and respiratory disorders. The predictions adhere quite well to the actual data for Random Forest, but less for AutoRegressive Integrated Moving Average. However, even if the specific value is not always correctly predicted, the overall trend seems to be rightly forecasted and performance metrics are mostly satisfying. Although additional work is required to improve their performances, results are encouraging and this sort of geographically-localised time-series forecasting seems feasible. Future developments will take into consideration the whole province of Brescia.**

Keywords-Forecasting; ER accesses; Hospitalisation; Pollution; Weather.

I. INTRODUCTION

Being able to properly manage the Emergency Department (ED) and Emergency Room (ER) is crucial to provide functional healthcare and improve patients' satisfaction [1]. This leads to a strong need for accurately predicting visitor volume and patient admissions to facilitate the planning of resources and staff for the whole hospital.

Multiple researchers have tried to predict access and admission rates based on historical ED data by creating scores or using deep learning (DL) or machine learning (ML) models (like Recurrent Neural Networks, Logistic Regression, Random Forest or Extreme Gradient Boosting) to forecast daily accesses to the ER [2] [3], the possibility of a patient's hospital admission after going through the triage [4] or even the risk of death [5]. Results were so encouraging, that others looked for associations with the surrounding environment.

In fact, there is proof that weather affects one's health, especially for people who have specific illnesses or healthcare needs. For example, there seems to be a link between the daily temperature and ED admissions for cardiovascular diseases or significant exacerbation of asthma in adults that visit ED [6] [7]. Generally speaking, regarding cardiovascular disorders, a worsening of the patient's well-being and cardiac arrests appear to be influenced by not only temperature but also stressors like humidity and atmospheric pressure [8] [9]. Moreover, there is also proof of links between air pollution and specific illnesses. Substances like $PM_{2.5}$, PM_{10}, NO_x, O_3 and SO_2 influence cardiac arrests [10], cardiac arrhythmia [11], cognitive decline in adult population [12], COVID-19 incidence [13], development of chronic kidney disease [14] or Type 2 diabetes [15]. $PM_{2.5}$ and PM_{10} are also linked to hospital admissions for cardiovascular [16] and respiratory diseases [17]. $PM_{2.5}$ levels also seem to be directly associated with increased daily ED visits for ulcerative colitis [18], while solar radiation is inversely associated with inflammatory bowel disease admissions [19]. There also seems to be a correlation between the number of hospitalised asthma patients and both weather (i.e., temperature and humidity) and pollution (i.e., $PM_{2.5}$, PM_{10} and NO_x) [20]. Finally, ML models (i.e., AutoRegressive Integrated Moving Average and Multilayer Perceptron) have also been applied to try to predict accesses to the ER by patients affected by infecting respiratory diseases after being exposed to $PM_{2.5}$ [21].

Some of these researches are based on long-term exposure to pollution (even 20-years long [12]), while others are based on a few days or even same day's exposure [13] [16] and some even on both [11]. Based on these literature pieces of evidence, trying to predict either all accesses to the ED or hospitalisation post-triage for specific illnesses, working on climate, pollution and historical accesses time-series belonging to the same area, seems feasible.

Each year, between 77000 and 80000 patients visit the ER of the biggest Brescia hospital [22] and 24% of them get admitted. This was the spark that ignited this work: trying to accurately predict future accesses to one of Brescia's EDs based on both historical and local meteorological and pollution data.

This paper contains a description of the analysed materials and applied methods, i.e., the datasets and the ML approaches applied to them, in Section II, the reached results in Section III, a comment on them in Section IV and a few final remarks in Section V.

II. Materials and Methods

In this section, the study design, analysed datasets (both clinical and environmental data) and applied algorithms are described.

A. Study Design

This study primarily aims to provide a daily prediction of the amount of patients visiting the ER of a precise hospital in the city of Brescia, Italy. Also, a forecast of the number of hospitalised patients for specific disease classes has been attempted. This retrospective study was performed based on daily data (clinical and environmental) for a period from January 1, 2018, to December 31, 2022. A four-year (i.e., 2018–2021) dataset was used to train the forecasting models, while the remaining data were used to test its forecasting capability. The final dataset that is used to feed the predictive algorithms is a combination of the clinical and the environmental data.

B. Data Collection: Clinical Data

The original clinical dataset was given by a hospital in Brescia to GPI for research purposes. The dataset contained all anonymous ER access data for the period 2018-2022. For each access (i.e., a person on a specific day) there were as many rows as the exams the person had undergone; pre-processing was made in order to have only one row for each ED visit while maintaining the patient's data (like the date of ER visit, their age, sex and zip code of their home address, the list of medical exams they were subjected to and, in case they were hospitalised, their diagnosis as an ICD9-CM code).

The following is a description of this dataset.

TABLE I. Brief description of clinical data.

Year	*Total accesses*	*Median age*	*Male percentage*
2018	60176	55	49%
2019	60106	56	49%
2020	47205	58	52%
2021	49571	57	50%
2022	56631	56	51%

In 2018, 12% of patients were below 18 years old, 31% between 19 and 49, 23% between 50 and 69, 34% above 70. In 2019, 11% of patients were below 18 years old, 30% between 19 and 49, 24% between 50 and 69, 35% above 70. In 2020, 9% of patients were below 18 years old, 29% between 19 and 49, 27% between 50 and 69, 35% above 70.
In 2021, 10% of patients were below 18 years old, 30% between 19 and 49, 26% between 50 and 69, 34% above 70. In 2022, 12% of patients were below 18 years old, 29% between 19 and 49, 25% between 50 and 69, 34% above 70. Amongst the most recurrent diagnoses of the hospitalised patients, through all years, were pneumonia and chronic heart failure. Note that this dataset contains accesses of people living not only in Brescia but also in the province of Brescia and other places in Italy and abroad. However, what we included in our final dataset is:

- Daily number of accesses to the ER, limited to those patients coming only from the city of Brescia
- The rolling mean of the number of the same patients, applying a seven-day window for calculation.

The following is a description of the dataset restricted to Brescia.

TABLE II. Brief description of clinical data (city of Brescia).

Year	*Total accesses*	*Median age*	*Male percentage*
2018	10389	56	46%
2019	10963	58	47%
2020	9835	61	50%
2021	11082	60	49%
2022	12597	60	49%

In 2018, 11% of patients were below 18 years old, 30% between 19 and 49, 24% between 50 and 69, 35% above 70. In 2019, 10% of patients were below 18 years old, 29% between 19 and 49, 24% between 50 and 69, 37% above 70. In 2020, 8% of patients were below 18 years old, 27% between 19 and 49, 27% between 50 and 69, 38% above 70.
In 2021, 9% of patients were below 18 years old, 28% between 19 and 49, 25% between 50 and 69, 38% above 70.
In 2022, 11% of patients were below 18 years old, 27% between 19 and 49, 23% between 50 and 69, 39% above 70.

A little bit of contextualisation of this clinical dataset: it is important to note that the area around Brescia suffered in a substantial way from the outbreak of the COVID-19 pandemic and the number of cases affected by coronavirus pneumonia far exceeds the occurrences of any other diagnosis during 2020. It is possible to observe from these data, and this is something already reported in previous studies [23] [24], that the number of accesses to ER decreased significantly from 2019 to 2020: this is explainable because Italy was subjected to a strict lockdown for several months that year. Hence it was less likely, for example, for car accidents to happen or for people wearing masks to get the flu.

C. Data Collection: Environmental Data

The environmental data have been supplied by the startup Hypermeteo [25] under GPI's specific request to match the spatio-temporal dimension of the already-at-disposal clinical dataset. The environmental data are defined per day and zip code, guaranteeing spatial-temporal precision. These data are obtained employing a mathematical model with a resolution of $10km$x$10km$, corrected through normalisation and downscaling, applied to data by Lombardia's Regional Environmental Protection Agency (ARPA [26]) weather stations. While the model was built for the entire Lombardia region, data were extracted for the province of Brescia only and, for this initial phase of the study, only data from the city of Brescia itself have been analysed.

The reported variables are:

- Temperature (min and max values) (T_{min}, T_{max} [$°C$])
- Humidity (min and max percentage values) (RH_{min}, RH_{max} [%])
- Precipitations (Prec [mm])
- PM_{10} and $PM_{2.5}$ [$\mu g/m^3$]
- NO_x, SO_2 and O_3 [$\mu g/m^3$]
- Total solar irradiance (SSW_{tot}) [Wh/m^2].

For each variable, safety ranges, provided along with the dataset, were considered in order to give a label (i.e., zero or one) to each value, to indicate if a value could be considered safe or not. Depending on the type of variable, either lower or upper bounds were considered, as reported in Table I.

TABLE III. SAFETY RANGES FOR ENVIRONMENTAL VARIABLES.

Environmental variable	Lower and Upper Bounds	
	Min value	*Max value*
NO_x	25 $\mu g/m^3$	-
$PM_{2.5}$	15 $\mu g/m^3$	-
PM_{10}	45 $\mu g/m^3$	-
O_3	100 $\mu g/m^3$	-
SO_2	40 $\mu g/m^3$	-
T_{min}	-	-10 $°C$
T_{max}	35 $°C$	
RH_{min}	-	15 %
RH_{max}	95 %	-
Prec	-	10 mm
SSW_{tot}	-	8500 Wh/m^2

Regarding the dataset for the city of Brescia, the number of occurrences in which the data were out of range was computed. Occurrences are to be intended as a single day of the five years considered, per single zip code (Brescia has 15 different zip codes). In the 71% of occurrences, NO_x results out of range, it is the 60% of cases for $PM_{2.5}$, 20% for PM_{10}, 17% for the max humidity, 7.7% for the precipitations, 7.4% for O_3, 1.8% for the max temperature and 0 cases out of range for SO_2 and the min temperature.

The issue of having multiple rows of data for the same date (i.e., one row for each zip code) has been handled similarly as in a project [27] found during our bibliographic research: each environmental variable has been labelled with the zip code it is referred to, and it is used as a column with daily values, thus grouping all data belonging to the same date on one row. Again, a clarification on the context: the area surrounding Brescia is densely inhabited and industrialised, resulting in one of the most polluted areas in Europe [28].

D. Predictive Algorithm: Random Forest

In order to predict future ER accesses based on our clinical and environmental data, a Random Forest (RF) approach was implemented on Python applying the open-source library Scikit-learn [29]. This model was chosen based on an article [30] that applied it to a temperature prediction problem: the analogy with our dataset highlighted this approach as an interesting candidate for this type of analysis. RFs apply sequential splits to the data such that the separation is maximised in regards to a homogeneity criterion resulting in a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [3]. The random forest algorithm picks N random records from the dataset and builds a decision tree based on them, repeatedly for the chosen number of trees (in this case, 1000). The topic has been tackled as a regression problem as we have considered the target variable (i.e., daily accesses) as a continuous one.

Through the same library cited before, some metrics were computed to evaluate the results: the Mean Absolute Error (MAE), the Mean Absolute Percentage Error (MAPE) and the Accuracy (Acc). Then, when the prediction of the number of daily hospital admissions for cardiovascular and respiratory pathologies was attempted, the Symmetric Mean Absolute Percentage Error (SMAPE) was computed. This analysis was applied expecting a more evident correlation between environmental, especially pollution, data. These pathological classes have been selected through their ICD9-CM codes.

The results that are reported in Subsection III-A, are based on different combinations of the datasets, as we applied the same model on the entirety of Brescia's data, only on the 2 most important features and only on cardiovascular and respiratory disorders data, respectively. In order to highlight a possible lag effect based on 1- and 5-day lag assumptions, which means that the observed data of previous days is used to predict the volume of patient access on a certain day, climate and pollution data were processed accordingly in order to create two analogous additional datasets.

The different analyses that were carried out, trying to improve the model's accuracy and potentially spot specific patterns, are divided into four cases:

- A; the RF algorithm was applied to the initial pre-processed dataset, then on 1-day and 5-day lagged data and, finally, only on the 2 most important features, as computed by the model
- B; analogous to A, but the rolling mean feature was discarded
- C; 1-day lagged data, no rolling mean, but the clinical data were reduced to only the part linked to hospitalised patients affected by cardiovascular pathologies, plus on the 2 most important features
- D; analogous to C, but the clinical data belonged to respiratory disorders.

Here, are reported the equations [33] for MAE (1), MAPE (2), SMAPE (3) and Acc (4):

$$MAE = \sum_{i=1}^{D} |x_i - y_i| \tag{1}$$

$$MAPE = \frac{100}{N} \sum_{i=0}^{N-1} \frac{y_i - \hat{y}_i}{y_i} \tag{2}$$

$$SMAPE = \frac{100}{n}\sum_{t=1}^{n}\frac{|F_t - A_t|}{(|A_t| + |F_t|)/2} \quad (3)$$

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

E. Predictive Algorithm: ARIMA

Trying to improve the results given by the algorithm described in Subsection II-D, a ML model for multivariate time-series prediction was applied to the same data. Specifically, an AutoRegressive Integrated Moving Average (ARIMA) model [31], a popular algorithm used in time series analysis and forecast, through the application of the auto-ARIMA process [32] in Python. 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, through a so-called autoregression. An ARIMA (p, d, q) model can be described in the following equation (5).

$$(1 - \sum_{i=1}^{p}\varphi_i L^i)(1 - L)^d X_t = (1 + \sum_{i=1}^{q}\theta_i L^i)\varepsilon_t \quad (5)$$

where L represents the lag operator, p represents the number of autoregressive terms, q represents the number of moving average terms, d represents the degree of differencing and ϕ, θ and ϵ are relevant parameters.

The performance metrics applied to the model to evaluate its performances were MAPE (2), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Akaike Information Criterion (AIC). Here, are reported the equations for MSE (6), RMSE (7) and AIC (8).

$$MSE = \sum_{i=1}^{D}(x_i - y_i)^2 \quad (6)$$

$$RMSE = \sqrt{\frac{1}{n}\Sigma_{i=1}^{n}\Big(\frac{d_i - f_i}{\sigma_i}\Big)^2} \quad (7)$$

$$AIC = 2k - 2\ln(\hat{L}) \quad (8)$$

III. Results

In this section, the obtained preliminary results are reported. The algorithms have been fed with different datasets that always include only data related to patients whose home address' zip code is inside the city of Brescia.

A. Results: Random Forest

Following, a series of plots is reported: they represent the predicted values (plotted in violet) versus the actual values (plotted in blue) for the year 2022, coming from the different input datasets as explained in Subsection II-D.
First, the results of case A. Figure 1 displays the actual test values and the predicted ones for the 1-day lagged data.

Here, the obtained metrics for the 1-day lagged dataset (i.e., MAE and Acc) and for the 2 most important features, i.e.,

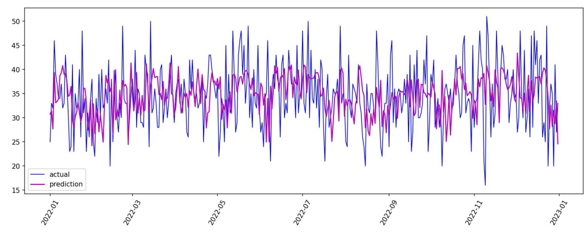

Figure 1. Random Forest's prediction and actual values for 1-day lagged data.

rolling mean and day (referring to the number of the day in a month), (i.e., MAE$_{mostimp}$ and $Acc\_{mostimp}$) are reported:

- MAE = 5.1
- Acc = 84.42%
- MAE$_{mostimp}$ = 5.57
- $Acc_{mostimp}$ = 82.63%

Now, the results of case B. Figure 2 displays the actual test values and the predicted ones for the 1-day lagged data missing the rolling mean.

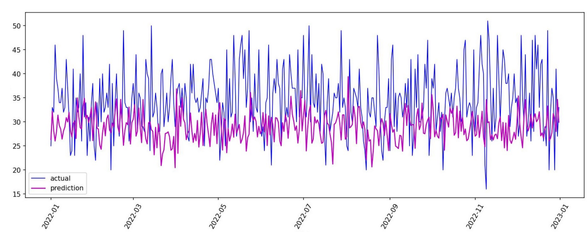

Figure 2. Random Forest's prediction and actual values without considering the rolling mean.

Here, the obtained metrics are reported (i.e., MAE$_{mostimp}$ and $Acc_{mostimp}$ refer to features day and month):

- MAE = 6.33
- Acc = 82.44%
- MAE$_{mostimp}$ = 7.49
- $Acc_{mostimp}$ = 79.29%

For case C, the actual and obtained predicted data are displayed in Figure 3.

Figure 3. Random Forest's prediction and actual values for cardiovascular diseases' hospitalisations.

The obtained metrics were (with rolling mean and day as the most important features):

- MAE = 0.51
- MAE$_{mostimp}$ = 0.49

Case D's plot of predicted and actual values for respiratory pathologies is Figure 4.

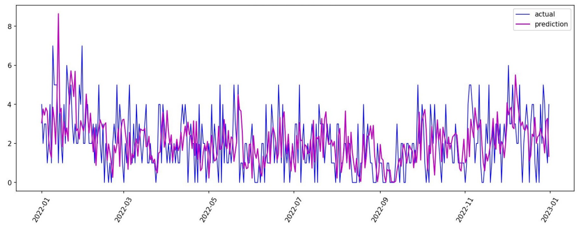

Figure 4. Random Forest's prediction and actual values for respiratory diseases' hospitalisations.

The obtained metrics were (with rolling mean and day as the most important features):

- MAE = 1.09
- SMAPE = 67.9%
- MAE$_{mostimp}$ = 1.23
- SMAPE$_{mostimp}$ = 74.9%

In this case, SMAPE was computed, instead of MAPE and Acc, due to the presence of 0 values in the test array.

B. Results: ARIMA

Here, the results obtained with the auto-ARIMA algorithm are shown: they represent the predicted values (plotted in violet) versus the actual values (plotted in green) for the year 2022. This prediction is obtained by feeding the initial dataset to the algorithm. The plot of the actual test values and the predicted ones for the same-day data is reported in Figure 5

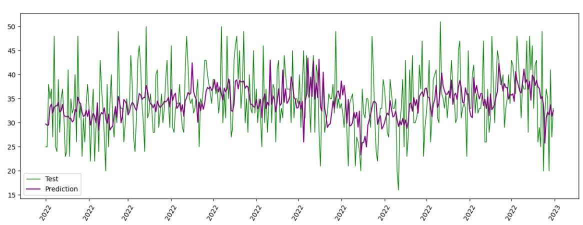

Figure 5. ARIMA's prediction and actual values for same-day data.

The obtained metrics were:

- MAPE = 15%
- MSE = 39.5
- RMSE = 6.3
- AIC = 9272.7

IV. Discussion

Results reported in Subsection III-A only refer to 1-day lagged data because, when the same process was applied to the same-day data and the 5-day lagged data, results were quite similar. Hence, in order to show the model performances, the former was chosen as it seemed to be the best logical approach. Results reported in Subsection III-B only refer to same-day data as it was the outcome of an early analysis of the ARIMA model application to these datasets, thus only the initial valuations have been implemented.

Visually comparing both models, predictions coming from the RF algorithm (Figure 1, 2, 3 and 4) appear to adhere better to the actual data when compared with the ARIMA one (Figure 5). However, even if the specific value is not always correctly predicted, the overall trend seems to be rightly predicted. This also happens while changing the considered features, despite removing historical data like the rolling mean (Figure 2), thus relying more on the environmental ones. In fact, the forecast values are underestimated, but the trend is followed quite well. Still, generally, the RF model also predicts peak values (Figure 4), i.e., surges in hospitalisations, quite aptly.

Please note that when analysing specific pathologies, the number of hospitalisations is limited to a few people every day and, sometimes, even none. This is particularly noticeable as, in this work, only the city of Brescia's data are used and it is more obvious for cardiovascular disorders rather than the respiratory ones, at least during the considered period.

Beyond the visual inspection, the metrics results reported in Subsection III-A show that the Acc for the RF model decreases when discarding the rolling mean as an input feature, but only of 1.98% and the error on the predicted number of accesses (i. e., MAE) increases from 5.1 to 6.33. This seems to suggest that when using the historical data through the rolling mean, the prediction could still be improved, but also that when this feature is ignored, the forecast performances do not dramatically worsen. Similar reasoning can be applied to the approach that uses the two features computed to be the most important ones, which behaves even less precisely.

Results for the RF application to cardiovascular and respiratory data seem to output better punctual predictions, but MAE values are smaller because of the lower values of daily hospitalisations (when compared to daily general accesses) and SMAPE is quite high. This could be due to the nature of the dataset itself as it is quite small. Regarding the ARIMA metrics reported in Subsection III-B, the listed AIC value is the one belonging to the best model identified by auto-ARIMA and the MAPE value represents a low, but acceptable accuracy. As expected by the visual inspection, though, MSE and RMSE values are not adequate.

Based on the aforementioned decrease in ER accesses during 2019 and 2020 due to the COVID-19 pandemic, an attempt at training the models only on 2018 and 2019 data (and still testing them on 2022 ones) was made, hoping to improve the preciseness of the predictions, but, surprisingly, in vain. In fact, the hypothesis was to discard the out-of-the-ordinary data so that the predictions computed merely on the historical data could be more precise. The results' worsening could be further evidence that the previously obtained results were not only due to historical data but also to environmental info which influences the correctness of the forecasting.

V. Conclusion and Future Work

This work represents a starting point towards the time-series analysis of historical and environmental data for the prediction of ER accesses and hospitalisations in a specific geographical area. The objective was only partially reached as this is a demanding field of application, but results were generally promising and, under these premises, a predictive analysis seems feasible. Considering that there are no other

 ISBN: 978-1-68558-136-7

truly comparable works in the international literature, these performances are even more encouraging. This being said, the obtained results cannot be generalised as they were achieved by analysing a period greatly made up of COVID-19-ridden years and a quite limited geographical area, so they can only be used to comment on this specific frame. The performances could dramatically differ if the analogous pre-processing and the same models were to be applied to other contexts or even just on a longer and more stable period.

Future developments of this work will, of course, include data belonging to the entire province of Brescia and a continuous search for more precise results, with the hope of moving to ever-growing datasets. It would also be interesting to test other ML algorithms or apply different pre-processing steps. Nevertheless, any attempt, whether it be successful or inconclusive, will still gather valuable insight on this yet to be delved into the topic and shed light on how our surrounding environment influences human health. This may be the offset of a new way of managing ER all over the world, monitoring entire populations and geographical areas, with the final objective of enabling a smart real-time predictive analysis able to improve the quality of healthcare and people's quality of life.

REFERENCES

[1] J. D. Sonis and B. A. White, "Optimizing patient experience in the emergency department", Emergency Medicine Clinics, vol. 38, no. 3, pp. 705–713, 2020.

[2] Z. Qiao et al., "Using machine learning approaches for emergency room visit prediction based on electronic health record data", Building continents of knowledge in Oceans of data: The future of co-created eHealth. IOS Press, pp. 111–115, 2018.

[3] Y. M. Chiu, J. Courteau, I. Dufour, A. Vanasse, and C. Hudon, "Machine learning to improve frequent emergency department use prediction: a retrospective cohort study", Scientific Reports, vol. 13, no. 1, p. 1981, 2023.

[4] A. Cameron, K. Rodgers, A. Ireland, R. Jamdar, and G. A. McKay, "A simple tool to predict admission at the time of triage", Emergency Medicine Journal, vol. 32, no. 3, pp. 174–179, 2015.

[5] R. Sánchez-Salmerón et al., "Machine learning methods applied to triage in emergency services: A systematic review", International Emergency Nursing, vol. 60, 2022.

[6] W. Zhu et al., "The effect and prediction of diurnal temperature range in high altitude area on outpatient and emergency room admissions for cardiovascular diseases", International Archives of Occupational and Environmental Health, vol. 94, no. 8, pp. 1783–1795, 2021.

[7] T. Abe et al., "The relationship of short-term air pollution and weather to ED visits for asthma in Japan", The American journal of emergency medicine, vol. 27, no. 2, pp. 153–159, 2009.

[8] D. Martinaitiene and N. Raskauskiene, "Weather-related subjective well-being in patients with coronary artery disease", International Journal of Biometeorology, vol. 65, pp. 1299–1312, 2021.

[9] M. Hensel et al., "Association between weather-related factors and cardiac arrest of presumed cardiac etiology: a prospective observational study based on out-of-hospital care data", Prehospital Emergency Care, vol. 22, no. 3, pp. 345–352, 2018.

[10] S. Kojima et al., "Fine particulate matter and out-of-hospital cardiac arrest of respiratory origin", European Respiratory Journal, vol. 57, no. 6, p. 2004299, 2021.

[11] M. A. Shahrbaf, M. A. Akbarzadeh, M. Tabary, and I. Khaheshi, "Air pollution and cardiac arrhythmias: a comprehensive review", Current Problems in Cardiology, vol. 46, no. 3, p. 100649, 2021.

[12] J. M. Delgado-Saborit et al., "A critical review of the epidemiological evidence of effects of air pollution on dementia, cognitive function and cognitive decline in adult population", Science of the Total Environment, vol. 757, p. 143734, 2021.

[13] S.-T. Zang et al., "Ambient air pollution and COVID-19 risk: evidence from 35 observational studies", Environmental research, vol. 204, p. 112065, 2022.

[14] M.-Y. Wu, W.-C. Lo, C.-T. Chao, M.-S. Wu, and C.-K. Chiang, "Association between air pollutants and development of chronic kidney disease: a systematic review and meta-analysis", Science of the Total Environment, vol. 706, p. 135522, 2020.

[15] Y. Li, L. Xu, Z. Shan, W. Teng, and C. Han, "Association between air pollution and type 2 diabetes: an updated review of the literature", Therapeutic Advances in Endocrinology and Metabolism, vol. 10, pp. 1-15, 2019.

[16] R. D. Brook et al., "Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association", Circulation, vol. 121, no. 21, pp. 2331–2378, 2010.

[17] F. Dominici et al., "Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases", Jama, vol. 295, no. 10, pp. 1127–1134, 2006.

[18] R. Duan et al., "Association between short-term exposure to fine particulate pollution and outpatient visits for ulcerative colitis in Beijing, China: A time–series study", Ecotoxicology and Environmental Safety, vol. 214, p. 112-116, 2021.

[19] F. Jaime et al., "Solar radiation is inversely associated with inflammatory bowel disease admissions", Scandinavian journal of gastroenterology, vol. 52, no. 6-7, pp. 730–737, 2017.

[20] C.-L. Chan et al., "A survey of ambulatory-treated asthma and correlation with weather and air pollution conditions within Taiwan during 2001–2010", Journal of asthma, vol. 56, no. 8, pp. 799–807, 2019.

[21] J. Lu et al., "Feasibility of machine learning methods for predicting hospital emergency room visits for respiratory diseases", Environmental Science and Pollution Research, vol. 28, pp. 29701–29709, 2021.

[22] https://civile.asst-spedalicivili.it/servizi/unitaoperative/unitaoperative fase02.aspx?ID =586 [Retrieved online: January 2024].

[23] F. Tartari, A. Guglielmo, F. Fuligni, and A. Pileri, "Changes in emergency service access after spread of COVID-19 across Italy", Journal of the European Academy of Dermatology and Venereology, vol. 34, no. 8, p. e350, 2020.

[24] T. Ferrari, C. Zengarini, F. Bardazzi, and A. Pileri, "In-depth, single-centre, analysis of changes in emergency service access after the spread of COVID-19 across Italy", Clinical and Experimental Dermatology, vol. 46, no. 8, pp. 1588–1589, 2021.

[25] https://www.hypermeteo.com/ [Retrieved online: January 2024].

[26] https://www.arpalombardia.it/ [Retrieved online: January 2024].

[27] https://www.kaggle.com/datasets/nicholasjhana/energy-consumption-generation-prices-and-weather [Retrieved online: January 2024].

[28] S. Khomenko et al. "Premature mortality due to air pollution in European cities: a health impact assessment", The Lancet Planetary Health, vol. 5, no. 3, pp. e121–e134, 2021.

[29] https://scikit-learn.org/stable/modules/generated/sklearn.ensemble. RandomForestRegressor.ht [Retrieved online: January 2024].

[30] https://towardsdatascience.com/random-forest-in-python-24d0893d5 1c0 [Retrieved online: January 2024].

[31] https://www.statsmodels.org/stable/generated/statsmodels.tsa.arima. model.ARIMA.html [Retrieved online: January 2024].

[32] https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima .auto_arima.html [Retrieved online: January 2024].

[33] G. Vishwakarma, A. Sonpal, and J. Hachmann, "Metrics for benchmarking and uncertainty quantification: Quality, applicability, and best practices for machine learning in chemistry", Trends in Chemistry, vol. 3, no. 2, pp. 155-156, 2021.

 ISBN: 978-1-68558-136-7