Estimating SARS-COV-2 Exposure Indoors in Delhi Given Outdoor Pollution Metrics Using Machine Learning

Bitan Biswas (✉ bitanbiswas99@gmail.com)

St Xavier's College <https://orcid.org/0000-0001-5020-8465>

Ravi Kaushik

Indian Institute of Technology Bombay

Research Article

Keywords: AQI, COVID-19, Environmental Science, India, Indoor Pollution, Pollution Analysis

Posted Date: August 26th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-836205/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

The Global Burden of Disease journal by the Lancet(Ritchie and Roser, 2013) and states that one million deaths have occurred from 1990 to 2017 due to air pollution. In 2018, the WHO estimated a death toll of 3.8 million due to indoor pollution(WHO,2018). In these times of the pandemic, it is quintessential for countries like India, with a huge population and high levels of pollution, to take severe measures for controlling pollution. The 2020 US Policy Report in the Lancet(2020) affirmed that there is a positive correlation between the $PM_{2.5}$ or PM_{10} particles concentration and COVID-19 infection as the virus uses the particulate matter as a piggyback. The case study here, is based on the Indian urban locality and aims to analyze and estimate the correlations between $PM_{2.5}$ particles, the AQI, weather conditions and COVID-19 particles using Machine Learning models. The optimum model is also to be used for predicting the outdoor AQI and Covid-19 infection rates in the suburban localities of northwestern Delhi and the data so obtained, would aid to calculating ,and extrapolating the mortality probability due to Covid-19 infection, indoors, in the metropolitan cities of India, like Delhi.

Introduction

Air Quality Index(AQI) of a place can be categorized into 5 different categories: Satisfactory, Moderate, Poor, Very Poor and Severe. Delhi has an AQI ranging between 400–600 which can be categorized as Very Poor to Severe due to high particulate matter concentration and other detrimental gases such as carbon dioxide, sulfur oxides and oxides of nitrogen. $PM_{2.5}$ are fine solid aerosols with a particle diameter of $\leq 2.5 \mu m$ and are found suspended in ambient air. $PM_{2.5}$ in indoor environments is primarily derived from common outdoor sources such as motor-vehicles, biomass burning (predominantly in rural areas), and industrial emissions(Nor et al. 2021, Su, W., Wu, X., Geng, X. et al 2019, Nadzir, M. S. M. et al 2020).This is because all forms of outdoor emissions have an impact on indoor environments given the continuous flow of air. Prolonged exposure to $PM_{2.5}$ can be detrimental to human health (Burnett et al. 2018) as this fine particulate matter can be easily inhaled and can penetrate deep into the lungs (Nor et al. 2021, Marcazzan 2001,Zhang 2015,HEI 2020). $PM_{2.5}$ has a significantly longer lifetime in the air where it can be suspended for an extended period compared to respiratory liquid droplets. This longer lifetime of particles may pose a significant viral exposure threat to a healthy person, especially in indoor environments(Marcazzan 2001).The fine particulate matter gets easily propagated by tiny turbulent eddies in the air that arise from activities such as human movements and walking (Xing 2016, Zwoździak 2015) .

In this paper, we use datasets obtained from Kaggle and Open Government Data Platform of India to first visualise graphically the AQI trends of in the past 5–6 years and the various ranges of AQI obtained in different months of the year. For complexity reduction, we reduce the analysis to a small locality of Ghaziabad district of Delhi, the capital and one of the most polluted cities of India. *The locality we take into consideration is Indirapuram-Vasundhara in Ghaziabad, located latitudinally and longitudinally within 28.64N,77.37E and 28.66N,77.38E.* The locality, according to the Indian newspapers and

government data, had the highest rate of COVID-19 infections in the months of November and December of 2020. This information is further validated using machine learning ensemble models like Random Forest Regressor and Gradient Boosting Regressor, which predict the possible AQI and weather conditions of the north-western Delhi localities with an accuracy of 80%. The predictions are visualised and correlated using the Pearson correlation coefficient and based on the correlations, we calculate the change in mortality rate ratio of indoors to outdoors, given the change in particulate matter concentration indoors due to outdoor pollution. All these data obtained, help us gain further insights in the mortality probability due to Covid-19 infection, indoors, thus fulfilling the cause of the case study.

Methods

2.1 Data:

The datasets (Kaggle Reference 5) used are as following :

1.Air Quality Data in India(2015-2020) : The Kaggle dataset contains air quality data and AQI (Air Quality Index) at hourly and daily level of various stations across multiple cities in India. Columns are 'city', 'datetime', 'PM2.5', 'PM10', 'NO', 'NO2', 'NOx', 'NH3', 'CO', 'SO2', 'O3', 'Benzene', 'Toluene', 'Xylene', 'AQI', 'AQI_Bucket'. We explicitly reduce the data to AQI data of Delhi only to reduce the space and time complexity of training the models.

2.COVID-19 in India: The Kaggle dataset had state-wise and district-wise details of the total number of coronavirus cases, tests carried out, positivity rate based on current population and other metrics. We again reduce the data to the total number of cases reported in Delhi daily from June 2020 till June 2021.

3.Delhi Weather data: Obtained from Wunderground using their easy-to-use API, this dataset comprises temperature(average and min-max), humidity, precipitation, and other condition details of Delhi weather from 1990 till 2016. Further weather conditions and daily mean temperatures till 2020 have been obtained by scraping Accuweather forecasts for Delhi.

2.2 Implementing Machine Learning:

We first use the AQI dataset to obtain a correlation between $PM_{2.5}$ particles and AQI which is found to be 0.8 on an average based on the data from 2015 to 2020, signifying a strong correlation between the two. We train Ensemble regressors like the Random Forest Regressor and the Gradient Boosting Regressor model on this data and obtain the AQI predictions for Indirapuram and Vasundhara locality based on their $PM_{2.5}$ outdoor air concentrations. Ensemble modeling is a process where multiple diverse base models are used to predict an outcome. The motivation for using ensemble models is to reduce the generalization error of the prediction. The approach seeks the wisdom of crowds in making a prediction. It acts and performs as a single model. Most of the practical data science applications utilize ensemble modeling techniques. In reference to Leo Breiman's work (Breiman 2001), Random forests are 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. The generalization error for forests converges a.s to a limit as the number of trees in the forest becomes large. Thus, using Random forests for AQI prediction ensures that it is closest to the actual AQI conditions of Indirapuram and Vasundhara. The Gradient Boosting Machine(GBM) algorithm is used for supervised machine learning, and it produces an ensemble of weak learners(Garcia de Oliveira,2019). The most used implementations of the GBM techniques are Light GBM by Ke et al,2017 and the XGBoost library by Chen and Guestrin,2016.However, despite being a collection of weak learners, it outperforms most ensemble models, with the help of hyperparameter tuning. Hence, out of the two chosen models for the prediction purposes, the Gradient Boosting Regressor, which is based on the GBM algorithm, performs better than the Random Forest Regressor.

The models predict the AQI values with accuracies(based on R^2 metric score) of 77.4% and 80% respectively. **The AQI predictions were the highest (mean value of 375 with a standard deviation of 25) in months of November and December 2020, indicating high PM_{2.5} particle concentration in these localities during that time of the year.** Further, according to the Hindustan Times newspaper of India, Indirapuram and Vasundhara had the highest number of COVID-19 caseloads during the months of November and December 2020. Although this doesn't indicate causality, correlation between particulate matter and Covid-19 infections is evident, keeping external validations in consideration. Further, to support our findings, there has been a research carried out by Nor, N.S.M., Yip, C.W, Ibrahim, N. et al(2020), where it was proven that particulate matter of diameter less than or equal to 2.5 μm could be a potential SARS-COV-2 carrier. No correlation was found between the virus concentration and the diameter of particulate matters (Marcazzan 2001) .However, positive correlations between PM_{2.5} and other respiratory viruses such as the influenza virus have been reported previously, emphasizing the probability of particulate matter being a transport carrier for SARS-CoV-2(Xing 2016).

Table-1. Machine Learning Models Used

The AQI dataset is further used to find a correlation between PM_{2.5} particle concentration and temperature and weather conditions. The temperature and weather conditions i.e., humidity for Delhi is obtained from the weather dataset as specified earlier. We merge the AQI, and weather datasets based on the common dates and obtain the correlation accordingly. For humidity, the correlation coefficient with PM_{2.5} turns out to be 0.076 and for temperature, it is -0.41. Thus, temperature and humidity in Delhi, have a significant negative and insignificant positive correlation respectively, with PM_{2.5} particles and transitively, with rate of Covid-19 infections, with respect to Delhi weather dataset. According to Yang Lv et al.(Yang Lv 2017), the prevalence of fog and haze seriously affect indoor air quality, given it affects the outdoor air quality, and indoor air quality is correlated and highly influenced by outdoor air quality(Braniš 2005, Kim 2010). In Daqing, China, research showed that there was a significant positive correlation among indoor particles concentration and outdoor particles concentration, temperature, and humidity ($p<0.05$), but different building types had obvious differences. Temperature and humidity are important

Models	Hyper Parameter Tuning	Cross-Validation	Test R ² score
Random Forest Regressor	'max_depth': range(3,9) 'n_estimators': range(100,200,20) 'max_features': [3,4,5,6] 'bootstrap': [True] 'criterion': ['mse']	Random Search CV with 80 iterations, random state of 1 and verbose of 2.	0.77423
Gradient Boosting Regressor	'learning_rate': [0.1, 0.01] 'max_depth': [3, 8] 'min_samples_leaf': [3, 5] 'max_features': [0.2, 0.6] 'loss': ['huber']	Grid Search CV with 3 folds instead of 5(default)	0.80018

factors affecting the concentration of indoor particulate matter and the influence of indoor and outdoor temperature is greater for offices and classrooms with the glass exterior wall, whereas the relative humidity is the main factor for the rest of the building with concrete wall structure. However, when analyzed in Indian setup, i.e., Delhi, weather conditions and temperature had contradicting impacts on the particulate matter concentration, thus, implying that the correlations differ from not only building to building, but also, background to background.

*This indicates the presence of other external factors such as casual behaviour of citizens, low testing rate and slow vaccination drive, inadequate measures and lack of strict lockdown and restrictions. Hence, **although the correlation is strong and positive between the COVID-19 infections and PM_{2.5} particles or the AQI predicted, the Pearson correlation coefficient value is estimated to be 0.68, due to the presence of other cofactors.** This is known as External validity in research.[31].*

According to the Health Effects Institute's Report of 2019, particulate matter (PM) pollution was considered the third most important cause of death in 2017 with the rate being highest in India. Air pollution was considered to cause over 1.1 million premature deaths in 2017 in India (HEI 2019), of which 56% was due to exposure to outdoor PM_{2.5} concentration and 44% was attributed to indoor air pollution. As per WHO (2016), one death out of nine in 2012 was attributed to air pollution, of which around three million deaths were solely due to outdoor air pollution. According to an article(Emily Henderson 2020), 1.67 million deaths occurred in India due to air pollution in 2019. This means that the mortality rate of

India associated with PM_{2.5} particle exposure in 2019 was 12.846 deaths per 1000 people. Given the pandemic and the increasing pollution in India despite several efforts by the government, it is feasible to assume that there has been an increase in the mortality rate due to the PM_{2.5} particles exposure in the last two years.

Beixi Jia et al (2021), found out that the estimated PM_{2.5}-mortality in India has had an annual increasing rate of 2.7% during 1998-2015. Further, the article states that aggressive air pollution control strategies should be taken in North India due to their current health risks. Based on this assumption, we use the formula obtained in NCBI's Mortality due to Indoor PM_{2.5} exposure Report (Ji W 2015),

$$\Delta \log M_{in,j} = (\Delta C_{out-in} t_{in} * \log M_{all,j}) / (\Delta C_{out-in} t_{in} + \Delta C_{out} t_{out}) \quad (1)$$

Where $\Delta \log M_{all,j}$ is the increase in mortality due to the j th outcome associated with total PM exposure for each 10 $\mu\text{g}/\text{m}^3$ increase in PM₁₀ or PM_{2.5}, outdoors. j represents three major health outcomes: all-cause, cardiovascular, and respiratory mortality.

ΔC_{out} is the increase in outdoor PM₁₀ or PM_{2.5} concentrations, which is set as 10 $\mu\text{g}/\text{m}^3$.

ΔC_{out-in} is the increase in outdoor-originated PM₁₀ or PM_{2.5} concentrations found in the indoor environment.

t_{out} is the duration of direct exposure to outdoor PM pollution.

t_{in} is the duration of indoor exposure to PM of outdoor origin.

$\Delta \log M_{in,j}$ estimates the increase in mortality due to the j^{th} outcome associated with indoor exposure to outdoor-origin PM for each 10 $\mu\text{g}/\text{m}^3$ increase in PM₁₀ or PM_{2.5}. Here, we use the PM_{2.5} concentration change explicitly.

Using this formula we obtain a ratio of 3:7 between the $\Delta \log M_{in,j}$ and $\Delta \log M_{all,j}$ which means that for an increase in the mortality by 7 units due to the j^{th} outcome associated with the total PM exposure for each 10 $\mu\text{g}/\text{m}^3$ increase in PM_{2.5} outdoors, there is an increase of 3 units in the value of mortality due to the j^{th} outcome associated with indoor exposure to outdoor-origin PM for each 10 $\mu\text{g}/\text{m}^3$ increase in PM₁₀ or PM_{2.5}. The calculations are carried out considering a time span of 24 hours and ΔC_{out-in} of 7.5 because according to Leung Dennis Y.C (2015), approximately 75% variation in indoor air pollutant concentration is due to outdoor air pollutant concentration variation. Previously, Douglas W. Dockery et al, based on a survey model, had estimated that the mean infiltration rate of outdoor fine particulates was approximately 70% and the effect of full air conditioning of the building was to reduce infiltration of outdoor fine particulates by about one half, while preventing dilution and purging of internally generated pollutants. However, when analyzed for the Delhi suburban setup, we see that the infiltration rate,

although within the 95-percentile spread of normal distribution of 70% mean, tends to be on the higher side due to the high rates of pollution in India. Further, according to Chun Chen et al.(2011),the indoor/outdoor ratios vary considerably due to the difference in size-dependent indoor particle emission rates, the geometry of the cracks in building envelopes, and the air exchange rates. Thus, it is difficult to draw uniform conclusions. However, for our case study, we realize that the indoor environment is highly influenced by the outdoor ambience and *there is a 30% increase in mortality due to increase in the indoor $PM_{2.5}$ concentration if there is a 70% increase in the mortality due to outdoor $PM_{2.5}$ concentration and the outdoor $PM_{2.5}$ concentration influences the indoor concentrations of the same by 75%.*

Further, based on one of the research works in PNAS(Z. Bazant 2021), we can quantify the concentration of pathogen $C(r,t)$ suspended in droplets of radius r at $25^{\circ}C$, exhaled by an infected person in a room and having another healthy person in the vicinity, is:

Rate of change=Production rate from exhalation – L_r – (2)

Where L_r is Loss rate of pathogens from ventilation, filtration, sedimentation, and deactivation.

For SARS-CoV-2, Buonanno et al. (2020) estimated a C_q range of 10.5 to 1,030 quanta/ m^3 based on the estimated infectivity $c_i=0.01$ to 0.1 of SARS-COV-2 and the reported viral loads in sputum although the precise value depends strongly on the infected person's respiratory activity. Here C_q is the concentration of exhaled infection quanta by an infectious individual. Hence, it becomes very important for implementation of air purification and ventilation along with proper maintenance of the 6ft rule even in the households. When the $PM_{2.5}$ concentration increases indoors , the probability of getting infected by these pathogenic suspended droplets increases given the virus can use the particulate matter as a carrier and thus, this explains the increase in mortality probability indoors given there is an in particulate matter concentration outdoors.

Results And Discussion

The outcomes of this research are intuitive as well as mathematical.

Outcome 1

COVID-19 infections have a positive and significant correlation with the $PM_{2.5}$ particle concentration in the air and $PM_{2.5}$ particle concentration have a positive and insignificant correlation with humidity and a negative and comparatively significant correlation with the temperature of that locality, keeping Indian urban northwestern Delhi background in mind. The correlation between $PM_{2.5}$ particle concentration and COVID-19 infections is lower than expected due to the non-blocking of external validity. AQI anomalously has a negative and negligible correlation with humidity. Weather conditions such as smoke, blowing sand, widespread dust and haze are associated with high $PM_{2.5}$ concentrations and hence, higher COVID-19 infection probability.

Outcome 2

Indoor air is affected by outdoor pollutant concentration and an increase of 70% in natural logarithmic value of mortality due to PM_{2.5} particles exposure outdoors will cause an increase of 30% in the indoor natural logarithmic value of mortality due to outdoor PM_{2.5} particles exposure. The concentration of exhaled infection quanta by an infectious individual (one already having COVID-19) is within a range of 10.5 to 1,030 quanta/m³ and hence aggressive amount of air purification measures must be taken to reduce the concentration of exhaled infection quanta, suspended in the air, which might aggravate the infection rates otherwise.

Table -2 Experiment Result

Concerns	Inference
Higher PM _{2.5} concentration in the air leads to :	Higher AQI value and higher transmission probability of SARS-COV-2 virus. They act as a potential carrier of the same, as previously discussed.
Outdoor PM _{2.5} concentration increase leads to:	Higher indoor PM _{2.5} concentration indoors and without air purification systems, outdoor environment affects indoors by 75%. Rate of viral transmission also increases resulting in higher fatality rates indoors.
A 70% change in the mortality rate outdoors due to particulate matter leads to:	A 30% change in mortality rate indoors due to PM thus stating that if we can decrease the outdoor pollution by a certain amount, that would lead to a significant change indoors.
Estimated infectivity rate range of 0.01 to 0.1 of SARS-COV-2 virus leads to:	A range of 10.5 to 1,030 quanta/m ³ where is the concentration of exhaled infection quanta by an infectious individual, which results into suspended SARS-COV-2 infectious droplets remaining suspended in the indoor airs, with further aggravation by the particulate matter concentration, if certain amount of ventilation or some form of air purification is not provided.

Conclusion

Pollution is a global issue. The pollution due to particulate matter has its adverse effects both indoors and outdoors(Xing 2016, Zwoździak 2015, Hänninen 2005) and if not contained, can result in not only increase in respiratory or thoracic diseases but also fatality in co-morbid cases. Proper air purification systems and sufficient ventilations are a mandate in urban houses and in rural areas, biomass fuel consumption should be controlled to reduce particulate matter pollution indoors(Chakraborty 2014). Photocatalytic materials can be utilized for indoor air purification(Hoang Bui 2021) and several companies across the world are producing advanced purifiers to tackle the viral transmission and mortality due to air pollution in urban cities. This paper, so far, has given a thorough understanding and analysis of air pollution and its effects on the SARS-COV-2 viral transmission rates, both indoors and outdoors. Further scope lies in analyzing using survey methods and experimental methods the effect of

indoor and outdoor pollution on human health and comparison between the same. More detailed quantization of the transmission rate and effect of being within 6ft distance of an infected person, indoors can also be carried out.

In conclusion, air pollution is one of the most prominent concerns in today's world and the government, as well the public should take precautions to protect themselves from its adverse effects.

Declarations

Bibliographical note

The authors of the article are a part of an incubation company of Indian Institute of Technology, Kanpur, India named AiRTH. The company is dedicated to making air purification systems that capture all forms of viruses and pathogens apart from particulate matter and ensure safe and clean environment indoors. We, at AiRTH, come from various premium institutes of India, to work collectively in research and development of advanced technologies to protect urban households from the ill-effects of cardiovascular or thoracic diseases and air pollutants. The article has been jointly written by the CEO of AiRTH, Mr. Ravi Kaushik and myself, Research and Development intern, Bitan Biswas and has been peer-reviewed by all the members of AiRTH to add onto their thoughts and thus, provide a collective and holistic information about the present conditions of Indian cities and how air purification plays an important role in our everyday life.

Geological information

The authors of the article and all the members of AiRTH belong to India and the case study is based on the Indian cities and the particulate matter concentration in developing country capitals like Delhi. As mentioned in the introduction, the case study is limited to northern Delhi urban localities, namely Indirapuram and Vasundhara. Both the localities are a part of Ghaziabad suburban areas of Northern Delhi, northwestern Uttar Pradesh.

Acknowledgements

We would like to thank the administration of IIT Kanpur for supporting and incubating multiple startups like ours, AiRTH and our parents, colleagues, professors and mentors for their constant support and mentorship. Lastly, we would like to thank the Government of India for allowing us to access databases and acquiring required data in suitable and usable forms from the open government data platforms.

Grant Support Details: The present research did not receive any financial support .

Conflict of Interest: The authors declare that there is not any conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/ or falsification, double publication and/or submission, and redundancy has been completely observed by the authors.

Life Science Reporting: No life science threat was practiced in this research.

Availability of Data and Material: The data is freely available on Kaggle. Refer to reference number 5 for the links to the same. Models can be run on freely available cloud platforms such as Google Collab or AWS.

Authors' Contribution: Both the authors have participated equally in the making of this review article and all the data and results or methods used are authentic, with no plagiarism and no misinterpretation of referenced articles.

Ethics approval: The work involves no form of plagiarism or misinterpretation of referenced article. The data available on the internet has been used ethically and no malpractice of any sort is involved in the making of this paper.

Consent to participate: This is an independent research and hence, all the authors have given their consent to participate and have participated on their will.

Consent for publication: Data and details used are all referenced and available in public domains and hence, all the materials used have the consent for publication. So does this paper.

References

1. Hannah Ritchie and Max Roser (2013) - "Indoor Air Pollution". *Published online at OurWorldInData.org*. Retrieved from: '<https://ourworldindata.org/indoor-air-pollution>'
2. WHO Household Air Pollution and Health(2018).*Published online by WHO*. Retrieved from: "<https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health>".
3. Delhi Covid crisis worsened by soaring pollution levels (2020). *Published by The Guardian*.Article:"<https://www.theguardian.com/world/2020/nov/11/delhi-covid-crisis-worsened-by-soaring-pollution-levels>"
4. *2020 Lancet Countdown on Health and Climate Change*: U.S. Policy Report. Retrieved from: "<https://www.hsph.harvard.edu/c-change/news/2020lancetcountdown/>"
5. Datasets from Kaggle :"<https://www.kaggle.com/rohanrao/air-quality-data-in-india>"
"<https://www.kaggle.com/imdevskp/covid19-corona-virus-india-dataset>"
"<https://www.kaggle.com/mahirkukreja/delhi-weather-data>"
6. Indirapuram, Vaishali and Vasundhara have highest Covid-19 Caseload in December(2020). *Published byHindustan Times*. Retrieved from : "<https://www.hindustantimes.com/noida/indirapuram-vaishali-and-vasundhara-have-highest-covid-19-caseload-in-december/story-afoURjBdVr0xRLRXIWKRTJ.html>".
7. Review by Emily Henderson(2020) – "Air pollution in India caused 1.67 million deaths in 2019".*Published online by News Medical*. Retrieved from : "<https://www.news-medical.net/news/20201223/Air-pollution-in-India-caused-167-million-deaths-in-2019.aspx>".

8. Ji W, Zhao B. Estimating mortality derived from indoor exposure to particles of outdoor origin. *PLoS One*. 2015;10(4):e0124238. Published 2015 Apr 10. doi:10.1371/journal.pone.0124238. Retrieved from: "<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4393180/>".
9. Leung Dennis Y. C. Outdoor-indoor air pollution in urban environment: challenges and opportunity *Frontiers in Environmental Science* Volume 2 2015 Page 69 doi 10.3389/fenvs.2014.00069. Retrieved from: <https://www.frontiersin.org/article/10.3389/fenvs.2014.00069>
10. A guideline to limit indoor airborne transmission of COVID-19 Martin Z. Bazant, John W. M. Bush *Proceedings of the National Academy of Sciences Apr 2021, 118 (17) e2018995118*; DOI: 10.1073/pnas.2018995118. Retrieved from: "<https://www.pnas.org/content/118/17/e2018995118>".
11. G. Buonanno, L. Stabile, L. Morawska, Estimation of airborne viral emission: Quanta emission rate of SARS-CoV-2 for infection risk assessment, *Environment International*, Volume 141, 2020, 05794, ISSN 0160-4120, <https://doi.org/10.1016/j.envint.2020.105794>. Retrieved from: "<https://www.sciencedirect.com/science/article/pii/S0160412020312800?via%3Dihub>".
12. Yang Lv, Haifeng Wang, Shanshan Wei, Lei Zhang, Qi Zhao, The Correlation between Indoor and Outdoor Particulate Matter of Different Building Types in Daqing, China, *Procedia Engineering*, Volume 205, 2017. Retrieved from: "<https://www.sciencedirect.com/science/article/pii/S1877705817345514>".
13. Douglas W. Dockery, John D. Spengler, Indoor-outdoor relationships of respirable sulfates and particles, *Atmospheric Environment (1967)*, Volume 15, Issue 3, 1981. Retrieved from: "<https://www.sciencedirect.com/science/article/abs/pii/0004698181900366>".
14. Chun Chen, Bin Zhao, Review of relationship between indoor and outdoor particles: I/O ratio, infiltration factor and penetration factor, *Atmospheric Environment*, Volume 45, Issue 2, 2011. Retrieved from: "<https://www.sciencedirect.com/science/article/abs/pii/S1352231010008241>".
15. Vijay Kotu, Bala Deshpande, Chapter 2 - Data Science Process, Editor(s): Vijay Kotu, Bala Deshpande, *Data Science (Second Edition)*, Morgan Kaufmann, 2019, Pages 19-37. Retrieved from: "<https://www.sciencedirect.com/science/article/pii/B9780128147610000022>".
16. Generalization Error. Retrieved from: "<https://www.sciencedirect.com/topics/computer-science/generalization-error>"
17. RANDOM FORESTS Leo Breiman *Statistics Department University of California Berkeley, CA 94720 January 2001*. Retrieved from: "<https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>"
18. A study on Gradient Boosting algorithms Juliano Garcia de Oliveira Advisor: Dr. Roberto Hirata Jr. April, 2019. Retrieved from: "https://bcc.ime.usp.br/tccs/2019/jgo/files/project_outline.pdf"
19. Nor, N.S.M., Yip, C.W., Ibrahim, N. *et al.* Particulate matter (PM_{2.5}) as a potential SARS-CoV-2 carrier. *Sci Rep* 11, 2508 (2021). <https://doi.org/10.1038/s41598-021-81935-9>. Retrieved from: "<https://www.nature.com/articles/s41598-021-81935-9>"
20. Su, W., Wu, X., Geng, X. *et al.* The short-term effects of air pollutants on influenza-like illness in Jinan, China. *BMC Public Health* 19, 1319 (2019). <https://doi.org/10.1186/s12889-019-7607-2> Retrieved from: "<https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-019-7607-2>"

21. Nadzir, M. S. M. *et al.* The impact of movement control order (MCO) during pandemic COVID-19 on local air quality in an urban area of Klang Valley, Malaysia. *Aerosol Air Qual. Res.*<https://doi.org/10.4209/aaqr.2020.04.0163> (2020). Retrieved from: "<https://aaqr.org/articles/aaqr-20-04-covid-0163>"
22. Marcazzan, G. M., Vaccaro, S., Valli, G. & Vecchi, R. Characterisation of PM₁₀ and PM_{2.5} particulate matter in the ambient air of Milan (Italy). *Atmos. Environ.* 35(27), 4639–4650 (2001).[https://doi.org/10.1016/S1352-2310\(01\)00124-8](https://doi.org/10.1016/S1352-2310(01)00124-8) Retrieved from: "<https://www.sciencedirect.com/science/article/pii/S1352231001001248?via%3Dihub>"
23. Zhang, Y.-L. & Cao, F. Is it time to tackle PM₂₅ air pollutants in China from biomass-burning emissions?. *Environ. Pollut.* 202, 217–219 (2015).<https://doi.org/10.1016/j.envpol.2015.02.005> Retrieved from: "<https://www.sciencedirect.com/science/article/pii/S0269749115000652?via%3Dihub>"
24. Xing, Y. F., Xu, Y. H., Shi, M. H. & Lian, Y. X. The impact of PM₂₅ on the human respiratory system. *J. Thorac.* 8(1), 69–74 (2016).<https://doi.org/10.3978/j.issn.2072-1439.2016.01.19> .Retrieved from: "<https://jtd.amegroups.com/article/view/6353/6196>"
25. Zwoździak, A., Sówka, I., Worobiec, A., Zwoździak, J. & Nych, A. The contribution of outdoor particulate matter (PM₁, PM_{2.5}, PM₁₀) to school indoor environment. *Indoor Built Environ.* 24(8), 1038–1047. <https://doi.org/10.1177/1420326X14534093> (2015). Retrieved from: "<https://journals.sagepub.com/doi/10.1177/1420326X14534093>"
26. Chatous Sidou, S. E. *et al.* Indoor/outdoor particulate matter number and mass concentration in modern offices. *Build. Environ.* 92, 462–474 (2015). <https://doi.org/10.1016/j.buildenv.2015.05.023> Retrieved from: "<https://www.sciencedirect.com/science/article/pii/S0360132315300020?via%3Dihub>"
27. Hänninen, O. O. *et al.* Reduction potential of urban PM_{2.5} mortality risk using modern ventilation systems in buildings. *Indoor Air* 15, 246–256 (2005). Retrieved from: "<https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1600-0668.2005.00365.x>"
28. Lewis-Beck, M. S., Bryman, A., & Futing Liao, T. (2004). *The SAGE encyclopedia of social science research methods (Vols. 1-0)*. Thousand Oaks, CA: Sage Publications, Inc. doi: 10.4135/9781412950589. Retrieved from: "<https://methods.sagepub.com/reference/the-sage-encyclopedia-of-social-science-research-methods/n700.xml>"
29. Kim, H. H., Kim, C. S., Lim, Y. W., Suh, M. A., & Shin, D. C. (2010). Indoor and outdoor air quality and its relation to allergic diseases among children: A case study at a primary school in Korea. *Asian Journal of Atmospheric Environment*, 4(3), 157-165. <https://doi.org/10.5572/ajae.2010.4.3.157>. Retrieved from: "<https://yonsei.pure.elsevier.com/en/publications/indoor-and-outdoor-air-quality-and-its-relation-to-allergic-disea>"
30. The effect of outdoor air and indoor human activity on mass concentrations of PM₁₀, PM_{2.5}, and PM₁ in a classroom, *Environmental Research*, Volume 99, Issue 2, 2005, Pages 143-149, ISSN 0013-

- 9351,<https://doi.org/10.1016/j.envres.2004.12.001>.Retrieved from:"<https://www.sciencedirect.com/science/article/pii/S0013935104002373>"
31. Bracht, G. H., & Glass, G. V. (1968). The external validity of experiments. *American Education Research Journal*, 5, 437-474. Gall, M. D., Borg, W. R., & Gall, J. P. (1996). Educational research: An introduction. White Plains, NY: Longman. Retrieved from: "https://researchbasics.education.uconn.edu/external_validity/#"
 32. Health Effects Institute *Annual Report 2019*, January 2020. Retrieved from: "<https://www.healtheffects.org/publication/annual-report-2019>"
 33. Beixi Jia, Meng Gao, Xiaorui Zhang, Xiang Xiao, Shiqing Zhang, Ken Kin Lam Yung, Rapid increase in mortality attributable to PM_{2.5} exposure in India over 1998–2015, *Chemosphere*, Volume 269, 2021, 128715, ISSN 0045-6535, <https://doi.org/10.1016/j.chemosphere.2020.128715>. Retrieved from : "<https://www.sciencedirect.com/science/article/pii/S0045653520329131>"
 34. Burnett et al., 2018, R. Burnett, et al. Global estimates of mortality associated with long-term exposure to outdoor fine particulate matter. *Proc. Natl. Acad. Sci. Unit. States Am.*, 115 (2018), pp. 9592-9597. Retrieved from:" <https://www.pnas.org/content/115/38/9592>"
 35. Deep Chakraborty, Naba Kumar Mondal, Jayanta Kumar Datta, Indoor pollution from solid biomass fuel and rural health damage: A micro-environmental study in rural area of Burdwan, West Bengal, *International Journal of Sustainable Built Environment*, Volume 3, Issue 2, 2014, Pages 262-271, ISSN 2212-6090, <https://doi.org/10.1016/j.ijsbe.2014.11.002>. Retrieved from:"<https://www.sciencedirect.com/science/article/pii/S2212609014000521>"
 36. Vu Khac Hoang Bui, Thanh Ngoc Nguyen, Vinh Van Tran, Jaehyun Hur, Il Tae Kim, Duckshin Park, Young-Chul Lee, Photocatalytic materials for indoor air purification systems: An updated mini-review, *Environmental Technology & Innovation*, Volume 22, 2021, 101471, ISSN 2352-1864, <https://doi.org/10.1016/j.eti.2021.101471>. Retrieved from: "<https://www.sciencedirect.com/science/article/pii/S235218642100119X>"

Figures

Figure 1

Location of Ghaziabad in Northwestern Uttar Pradesh and northern Delhi suburbs, India

Figure 2

Correlation Matrix between PM2.5 and AQI (0.84)

Figure 3

Algorithm for the Entire Process

Figure 4

Correlation between PM2.5 , AQI, temperature and Humidity

Figure 5

(a) AQI trend in Delhi for the past 5 years from 2015-2020 (b) Weather conditions associated with high PM2.5 particle concentration in air: Smoke, Blowing Sand, Widespread Dust