

Air Pollution Prediction using Machine Learning – A Review

Ismail A. Sulaimon¹, Hafiz A. Alaka¹, Razak Olu-Ajayi¹, Mubashir Ahmad¹, Funlade Sunmola², Saheed Ajayi³, Abdul Hye⁴

{i.sulaimon, h.alaka, r.olu-ajayi, m.ahmad21, f.sunmola}@herts.ac.uk,

S.Ajayi@leedsbeckett.ac.uk, abdul@rasutaenergy.org

1. *Big Data Technologies and Innovation Laboratory, University of Hertfordshire, Hatfield, AL10 9AB, UK*

2. *School of Engineering and Technology, University of Hertfordshire, Hatfield, AL10 9AB, UK*

3. *School of Built Environment, Engineering and Computing, Leeds Beckett University, Leeds LS28AG, UK*

4. *Rasuta Energy LTD, 5 Rachel Gardens, Birmingham, England, B29 6NY*

Abstract

In the effort to achieve accurate air pollution predictions, researchers have contributed various methodologies with varying data and different approaches that can be judged accurate in their respective contexts. Diverse approaches have been used so far in the literature to achieve optimal accuracy in the prediction of air pollution. Researchers have also used different combinations of data such as Meteorological, Traffic and Air Quality data. Hence, creating a situation where there are open questions on which of the machine learning (ML) algorithms or ensemble of algorithms is best suited for various combinations of data and varying dependent and independent variables. While it is obvious that there is a need for a more optimally performing predictive model for air pollution prediction, it is difficult to know what combination of algorithms and data is best suited for various dependent variables. In this study, we reviewed 26 research articles reported recently in the literature and the methods applied to different data to identify what combination of ML algorithms and data works best for the prediction of various air pollutants. The study revealed that despite the availability of many datasets, researchers in this domain cannot avoid the use of Air Quality and Meteorological datasets. However, Random Forest appears to perform well for various combinations of datasets.

1. INTRODUCTION

Increasing air pollution has had an enormous negative impact on the global economy, quality of life of humans and health of animals and plants as well. As reported by the United Nations (UN), the urban population is expected to become 68% of the world population in 2050 (UN DESA, 2018). This alone will influence negatively the air quality of major cities around the world if nothing is done to mitigate the effect of rural-urban migration. The reports by the World Health Organization (WHO) elaborated how air pollution causes about seven million annual deaths around the world, while the air quality in above 80% of the urban areas is worse than the WHO guideline (WHO, 2014). The majority of those currently affected by this worsening situation are the vulnerable groups which include children, the elderly, and people with respiratory and cardiovascular problems. Records

show that in recent years, air pollution accounts for 1 out of 8 deaths globally (WHO, 2014). This highlights the importance and urgency of the need for a highly accurate air pollution prediction model.

Air pollution arises due to the increase in the proportion of pollutants in the form of particles and inhabitable gasses in the atmosphere of an area. These sometimes could be due to human activities in the case of traffic, industry or home-related pollution. While at other times it could be due to biological or environmental activities such as the case of pollutants like ozone (O₃), Pollen, Dust. Air pollutants of various nature have been linked to critical health challenges such as cardiovascular diseases, pulmonary disease, acute respiratory infection and increased risk of lung cancer (Gul & Khan, 2020). Pollutants such as Nitrogen Dioxide (NO₂), Sulphur Dioxide (SO₂), Ozone are medically proven to irritate the airways of the lungs, increasing the symptoms of those suffering from lung diseases. Fine particles are seen to find their way deep into the lungs where they can cause inflammation and a worse condition of heart and lung diseases. Carbon Monoxide (CO₂) inhibits the absorption of oxygen by the blood (Department for Environment, Food and Rural Affairs, 2021). This condition alone can lead to a shortage of oxygen supply to the heart and most likely death. Agriculturally, the yield of crops has been affected by the increasing concentration of Ozone in the atmosphere (Gul & Khan, 2020).

In the efforts to achieve accurate air pollution predictions, research (Martinez-Espana et al., 2018; Tao et al., 2019; J. Zhang & Ding, 2017; Zhao et al., 2016) have contributed various methodologies with different approaches and varying data and can be judged accurate in their respective context. Due to the effectiveness of Machine Learning (ML) algorithms in solving many prediction problems in research and industry, researchers have been focusing largely on using various ML algorithms (Masih, 2019). As deterministic models struggle to capture the relationships between variables that affect pollution, researchers have been pushed to considering various ML algorithms ranging from the classical ML algorithms such as Support Vector Machine (SVM), Linear Regression (LR) and the sophisticated ML algorithms like Deep-Neural Network (DNN), and Extreme Machine Learning (Iskandaryan et al., 2020; Rybarczyk & Zalakeviciute, 2018).

Diverse approaches have been used so far in the literature to achieve optimal accuracy in the prediction of air pollution. Researchers have also used different combinations of data such as Meteorological, Traffic and Air Quality data. In the course of finding an optimally performing prediction model, researchers have trained various ML algorithms on different combinations of data (Iskandaryan et al., 2020; Rybarczyk & Zalakeviciute, 2018). Hence, creating a situation where there are open questions on which of the ML algorithms or ensemble of algorithms is best suited for various combinations of data and varying dependent and independent variables. In reality, while it is obvious that there is a need for a more optimally performing predictive model for pollution prediction, it is also difficult to know what combination of algorithms and data is best suited for various independent and dependent variables. As noted in (Iskandaryan et al., 2020; Rybarczyk & Zalakeviciute, 2018) it is difficult to compare the results obtained from different works of literature on the

prediction of air pollution, as each researcher used different data, and analysed varying temporal granularity.

Recent research (such as Iskandaryan et al., 2020; Rybarczyk & Zalakeviciute, 2018; Yafouz, Ahmed, Zaini, & El-Shafie, 2021) have conducted a review on literature that uses ML in the prediction of air pollution. (Iskandaryan et al., 2020; Rybarczyk & Zalakeviciute, 2018) conducted related review where several works of literature were studied to analyse the trend in air pollution prediction. Their collective findings reveal that most researchers on this topic apply advanced and sophisticated ML techniques and PM_{2.5} was the main prediction target. (Rybarczyk & Zalakeviciute, 2018) noted that for efficient prediction, it is important to consider external factors such as weather conditions, spatial characteristics, and temporal features. (Iskandaryan et al., 2020) explained that the majority of the research also used the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and R-squared (R²) score for evaluation and comparison of results. Their study presents the common practice among researchers in air pollution prediction, while their result does not indicate the best combination of datasets and algorithms for each air pollutants.

Despite covering other air pollutants, (Yafouz, Ahmed, Zaini, & El-Shafie, 2021) focused mostly on Ozone (O₃) and the review paper captured four areas of interest in ML which are Artificial Neural Networks (ANN), Support Vector Machine (SVM), Decision Tree (DT) and hybrid models. This research reveals interesting findings, but are restricted to a target sub-domain in air pollution prediction using ML.

This study will mitigate this challenge by opening up to all kinds of pollutants, ML, datasets used in air pollution prediction. Hence, we review some of the recent research reported in the literature and the methods applied to different data in order to identify what combination of ML algorithms and data works best for various dependent variables. The study objectives are summarized below:

- To analyse the ML approaches used in research about air pollution prediction.
- Evaluate the ML approaches used in the literature to:
 - Identify common pollutants focussed in air pollution prediction using ML
 - Identify the datasets used in literature for air pollution prediction using ML
 - Identify the ML algorithms used on each dataset for the air pollution prediction.
- To identify the combination of ML algorithms and dataset that performs best for various pollutants.

To achieve the study objectives, we collected 26 primary studies that are reviewed to answer the research questions stated below:

RQ1: What are the commonly focussed pollutants in air pollution prediction using ML?

RQ2: What datasets are used in this domain?

RQ3: What ML algorithms are used for each combination of datasets?

This study provides researchers with a clear understanding of the state-of-the-art air pollution prediction using ML. It also answers the question of what ML algorithm and Dataset performs best for the prediction of various air pollutants.

The remaining part of this work of literature is arranged as follows; Section 2 explains the research method used in accomplishing the study objectives. The section that follows is the main discussion where we discussed the results of the analysis in this study. Finally, a conclusion follows to summarize what has been achieved through this study and clarify the way forward in this research path.

2. RESEARCH METHODS

Following the research approach used in many review studies (such Alaka et al., 2018; Iskandaryan et al., 2020; Rybarczyk & Zalakeviciute, 2018), this study applied the literature survey methodology of research. Hence, 26 research articles published within the five years between the years 2017 and 2021 were reviewed. These were discovered through the research database SCOPUS. The choice of SCOPUS is based on the benefit that it provides by the unification of research globally and providing a comprehensive and wide range of scholarly information. To ensure the validity of the result, only the research articles published in conferences and journals are selected for this study.

We searched the SCOPUS database using the search keyword below:

(“air pollution” OR “air quality” OR “atmospheric pollution” OR “air pollutant”) AND (prediction OR forecast OR forecasting OR predict OR predicting) AND (“machine learning” OR ml OR “predictive model” OR modelling OR algorithm OR “big data” OR bigdata OR “artificial intelligence” OR ai)

The literature search yields 49 relevant research articles between January 2007 and June 2021. With 12 articles recorded, 2019 has the highest number of records in our study while the least number of literature was recorded in the years 2007, 2012, 2013 and 2015. No literature was recorded in the year between 2008 and 2011. As reported in Figure 1, this indicates that many of the articles considered for our study are state-of-the-art.

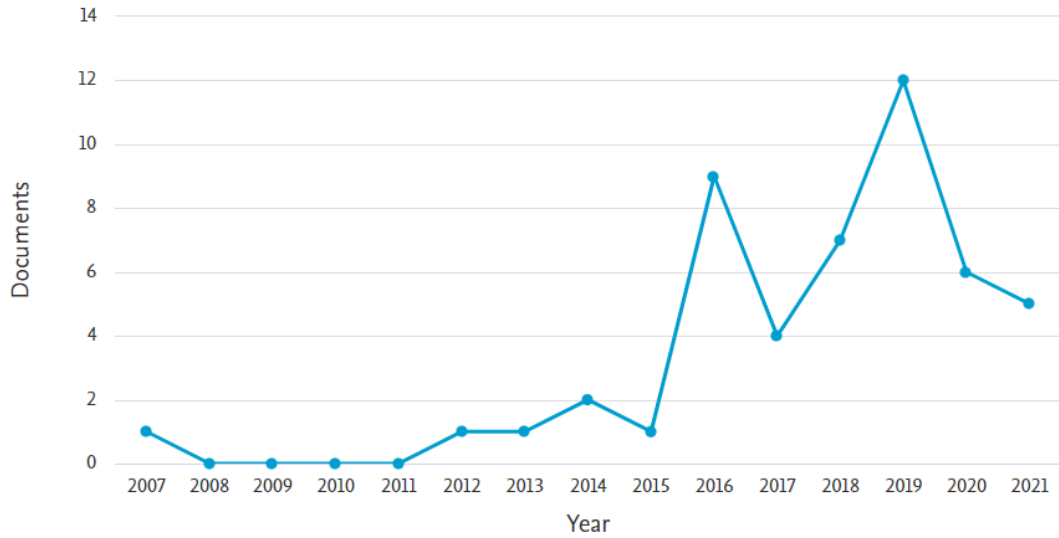


Figure 1: Literature published per year on Air Pollution Prediction using ML

As seen in Figure 2 below, the articles in this study are well distributed around the world as many of the continents are sufficiently represented. A large number of the literature in this study are published in Asia, with China having the largest recorded number of research articles. Followed by the United States of America with 9 recorded research articles and India in third place with 8 research articles.

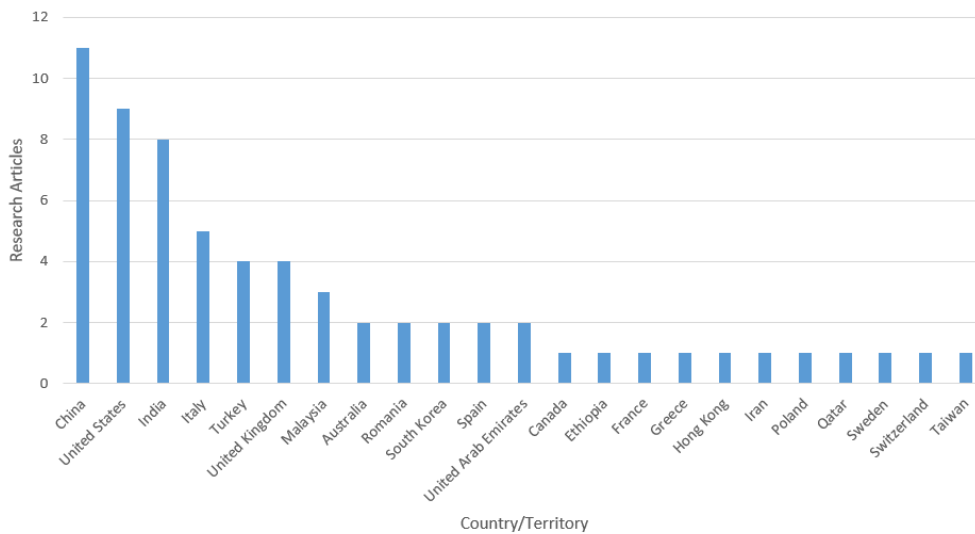


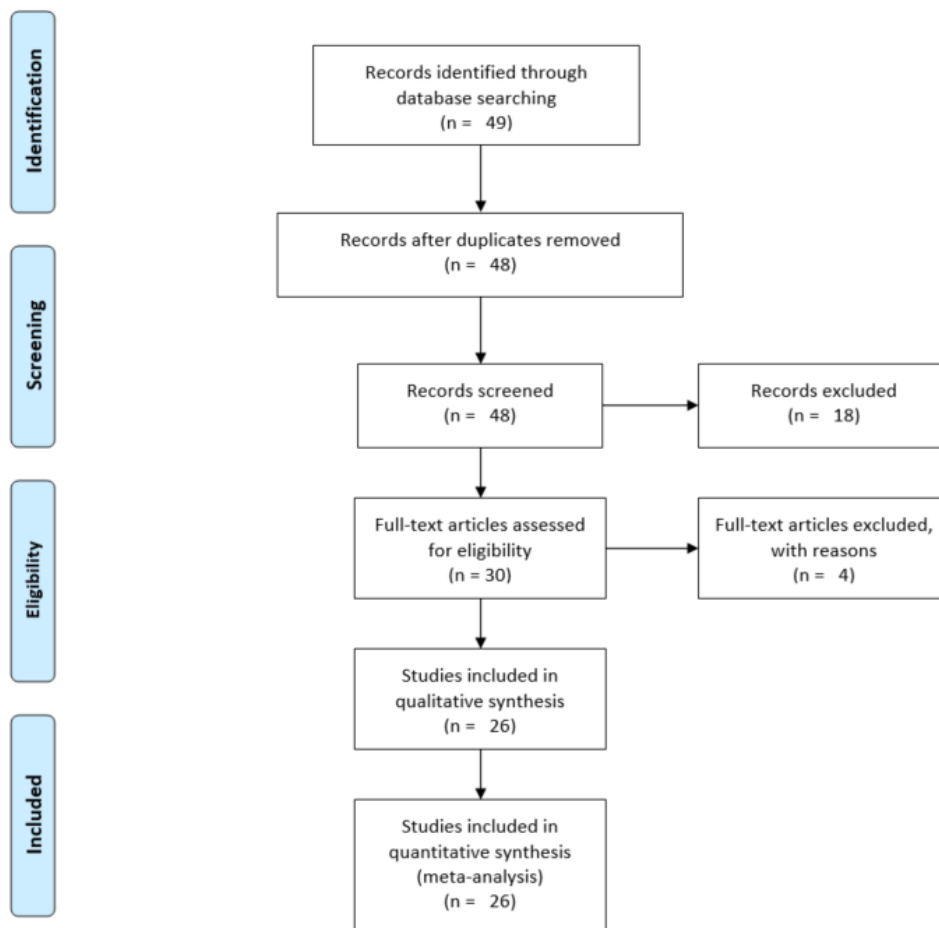
Figure 2: Literature published per country on Air Pollution Prediction using ML

The 49 research articles discovered through our search were put to further scrutiny to ensure only quality research are included in the study. Our literature search flow diagram in Figure 3 shows that the literature identification stage ended up with 48 research articles as one of the articles is a duplicate. Hence, a copy of the article from the most reliable

source is included for the study. At the screening stage, 18 articles are excluded after screening the abstract and ensuring that the selected articles were published within the defined period between 2017 and 2021. Despite the titles of some of the articles being related to our study, the abstract reflects that their objectives are very different from that of our study. While some other articles have no full text available for further study. Therefore, 30 research articles are remaining for the full-text analysis (eligibility) stage. The full-text analysis reveals that four of the articles deviate from the objective of our research, as one does not report the pollutants of focus, the other involved only a simulation study and the last two does not involve air pollution prediction.

Finally, 26 articles passed our rigorous inclusion criteria and are studied further to ensure accurate data extraction, data synthesis and analysis. The articles included in this study are believed to be sufficient and diverse enough to answer our research questions. This will ensure that the validity of the result of our study is not threatened.

Table 3: Literature search flow diagram



For data extraction and synthesis, an MS-Excel worksheet is created with each research article recorded on a row and their answers to the research questions were recorded against

each article. This approach is adopted to ensure an accurate data synthesis at the end of the data extraction process.

3. MAIN DISCUSSION

The data extraction and synthesis process are carried out to ease the analysis done in answering the research questions. In this section, we use the outcome of our data analysis to answer the research questions posed to achieve the study objectives.

3.1 Result analysis

Here, we analyse the data (Table 1) extracted from the research articles to derive insight and answer the research questions posed.

Table 1: Research articles collected for the study

Authors	Pollutants	Datasets	ML Algorithm	Compared with
(Yarragunta et al., 2021)	PM ₁₀ , PM _{2.5} , SO ₂ , CO, NO ₂ , O ₃	AQ	DT	LR, SVM, RFT, NBT, KNN
(Cihan et al., 2021)	PM ₁₀ , PM _{2.5}	AQ, M	ANFIS	SVR, CART, RF, KNN, ELM
(Ashayeri et al., 2021)	PM _{2.5}	T, BOP	SVR	
(Yafouz, Ahmed, Zaini, Sherif, et al., 2021)	O ₃	AQ, M	CNN and LSTM	
(Shah et al., 2020)	PM _{2.5} , PM ₁₀ , O ₃ , NO ₂ , SO ₂ , CO	AQ	SVM	RF, DT
(Alpan & Sekeroglu, 2020)	PM _{2.5} , PM ₁₀ , NO ₂ , SO ₂ , O ₃ , CO	M	RF	DT, SVR
(Kumari et al., 2020)	SO ₂ , O ₃ , NO ₂	AQ	RF	
(Dobrea et al., 2020)	PM ₁₀ , PM _{2.5}	AQ, M	ARIMA	SVR, LSTM,
(Bozdağ et al., 2020)	PM ₁₀	AQ, GIS	ANN,	ANN, LASSO, SVR, RF, KNN, Xgboost
(Selvi & Chandrasekaran, 2020)	O ₃ , PM ₁₀ , NO ₂	AQ, M	EINN	
(Wu & Lin, 2019)	PM _{2.5} , PM ₁₀ , SO ₂ , CO, NO ₂ , O ₃	AQ	LSSVM-BA	
(Anurag et al., 2019)	CO, C ₆ H ₆ , C ₆ H ₅ CH ₂ CH ₃ , NO, C ₆ H ₄ (CH ₃) ₂ , NO _x , O ₃ , PM _{2.5} , SO ₂ , C ₇ H ₈	M	Xgboost	
(Babu & Beulah, 2019)	PM _{2.5} , PM ₁₀ , SO ₂ , CO, NO ₂ , O ₃ , NH ₃	AQ, M	DT	LR, RF, KNN, SVM
(Srivastava et al., 2019)	PM _{2.5} , PM ₁₀ , CO, NO ₂ , SO ₂ , O ₃	AQ, M	SVR	ANN
(Rossi et al., 2019a)	NO ₂ , PM ₁₀ , O ₃	T, M	BRNN	
(Rossi et al., 2019b)	NO ₂ , PM ₁₀ , O ₃	T, M	BRNN	

(Mo et al., 2019)	PM _{2.5} , PM ₁₀ , NO ₂ , SO ₂ , CO, O ₃	AQ	ICEEMDAN- WOA-ELM	
(Ray et al., 2019)	C ₆ H ₆	AQ	DNN	
(Lin et al., 2018)	NO ₂ , PM ₁₀ , O ₃ , PM _{2.5}	AQ	CMG	
(Chen, Wang, et al., 2018)	PM ₁₀	M, LU, AOD	RF	
(Huang et al., 2018)	PM _{2.5}	AOD, M, LU	RF	
(Xu et al., 2018)	PM _{2.5}	AOD, LST	Cubist	RF, XGB
(Chen, Li, et al., 2018)	PM _{2.5}	M, LU	RF	
(S. Zhang et al., 2018)	PM _{2.5}	AQ, M	WNN	
(Jo & Khan, 2018)	CH ₄ , CO, SO ₂ , H ₂ S	AQ	ANN	
(Ni et al., 2017)	PM _{2.5}	M	BPNN	ARIMA

PM_{2.5}=Particulate Matter(diameter=2.5µm); PM₁₀=Particulate Matter(diameter=10µm); SO₂=Sulfur Dioxide; CO=Carbon Monoxide; NO₂=Nitrogen Dioxide; O₃=Ozone; NH₃=Ammonia; C₆H₆= Benzene; CH₄=Methane; H₂S=Hydrogen Sulfide Carbonyl Sulfide; C₆H₅CH₂CH₃=Ethylbenzene; C₆H₄(CH₃)₂=meta-Xylene; C₇H₈=Toluene; AQ=Air Quality data; M=Meteorological data; T=Traffic data; BOP=Building Occupancy Pattern; GIS=Geographic Information System data; LU=Land Use data; AOD=Satellite-based Aerosol Optical Depth data; LST=Land Surface Temperature data; DT=Decision Tree; LR=Logistic Regression; SVM=Support Vector Machine; RFT=Random Forest Tree; NBT=Nave Bayes Theorem; KNN=k-Nearest Neighbor; ANFIS=Adaptive Neuro-fuzzy Inference System; SVR=Support Vector Regressor; CART=Classification Regression Trees; RF=Random Forest; ELM=Extreme Learning Machine; EINN=Elman Neural Network; LSSVM-BA=Least Squares Support Vector Machine with Bat Algorithm; BRNN=Bayesian Regularization of Neural Networks; ICEEMDAN-WOA-ELM=Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise combined with Whale Optimization Algorithm and Extreme Learning Machine; DNN=Deep Neural Network; CMG=Cloud Model Granulation; WNN=Wavelet Neural Network; ANN=Artificial Neural Network; BPNN=Back Propagation Neural Network; ARIMA=Autoregressive Integrated Moving Average; XGB=Feature-Based Weighted Xgboost

3.1.1 RQ1: What are the commonly focussed pollutants in air pollution prediction using ML?

The pollutants PM_{2.5} appears to be the most commonly predicted in the literature studied (Figure 3). This is followed by PM₁₀, O₃, NO₂, SO₂, CO.

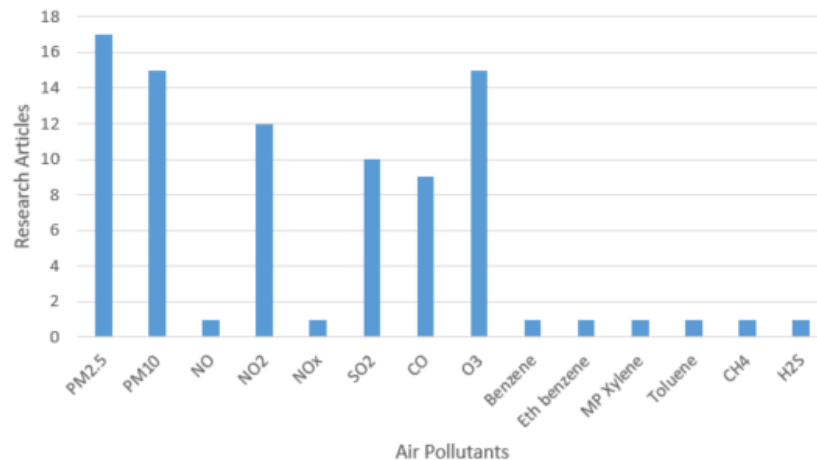


Figure 3: Common pollutants of focus in Air Pollution Prediction using ML

As seen above, the least predicted air pollutants are NO, NO_x, C₆H₆, C₆H₅CH₂CH₃, C₆H₄(CH₃)₂, C₇H₈, CH₄, H₂S. These air pollutants are rarely predicted, as they are not commonly found in the atmosphere.

3.1.2 RQ2: What datasets are used in this domain?

Several datasets are used in the prediction of air pollutants. The most common datasets used are the Air Quality (AQ) and the Meteorological (M) datasets (Figure 4).

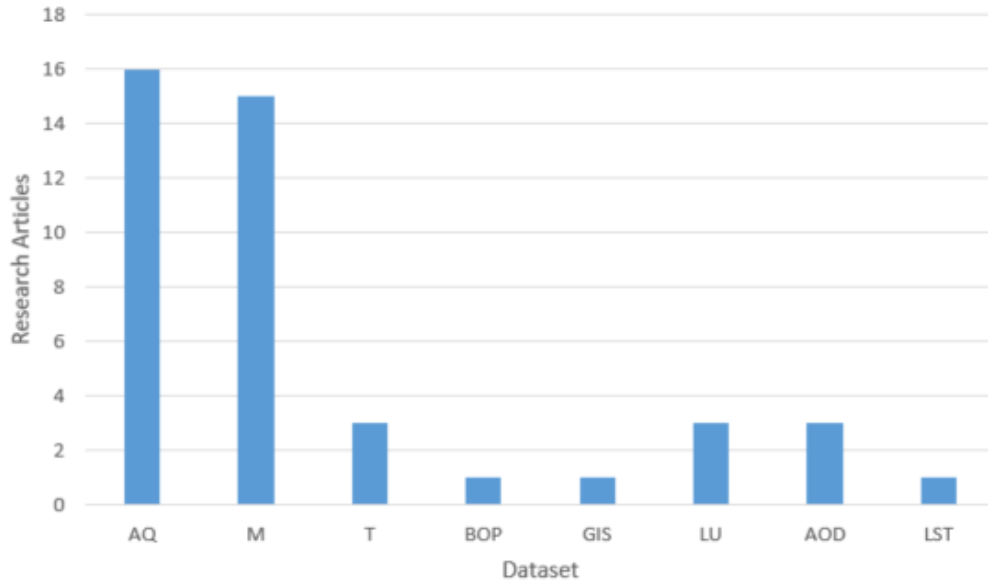


Figure 4: Datasets commonly used in Air Pollution Prediction using ML

The Traffic (T) dataset, Land Use (LU) and the Satellite-based Aerosol Optical Depth (AOD) dataset are sometimes used to achieve more accuracy while Building Occupancy Pattern (BOP), Geographic Information System (GIS) and Land Surface Temperature (LST) datasets are least used in the domain.

3.1.3 RQ3: What ML algorithms are used for each combination of datasets?

Many research adopted the use of Random Forest (RF) for their model development (see Figure 5). This indicates that RF performs well with several combinations of datasets. This is evident in the number of research that used RF. Decision Tree (DT), Support Vector Regressor (SVR) and Bayesian Regularization of Neural Networks (BRNN) are also common among researchers as their performance is significantly better for some dataset combinations. There exist only a few research that used ML algorithms such as; Support Vector Machine (SVM), Adaptive Neuro-fuzzy Inference System (ANFIS), Elman Neural Network (EINN), Least Squares Support Vector Machine with Bat algorithm (LSSVM-BA), Deep Neural Network (DNN), Cloud Model Granulation (CMG), Wavelet Neural Network (WNN), Artificial Neural Network (ANN), Back Propagation Neural Network (BPNN), Autoregressive Integrated Moving Average (ARIMA), Featured Based Weighted Xgboost (XGB), Improved Complete Ensemble Empirical Mode Decomposition with

Adaptive Noise combined with Whale Optimization Algorithm and Extreme Learning Machine (ICEEMDAN-WOA-ELM).

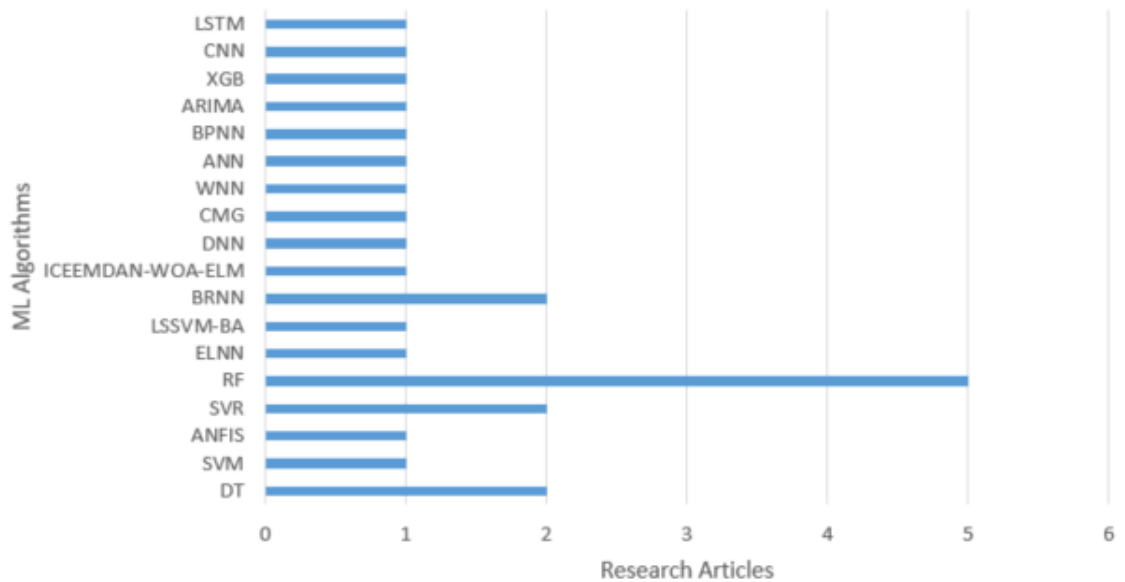


Figure 5: ML algorithms used for each combination of datasets

3.2 Overview

Despite having six common pollutants ($PM_{2.5}$, PM_{10} , O_3 , NO_2 , SO_2 , CO) among many of the research in this study, most of them use a diverse combination of datasets for the prediction of these air pollutants. The most popular of the datasets used are the Air Quality dataset and the Meteorological dataset. Almost all the literature have a form of Air Quality dataset in their dataset collection used for air pollutant prediction. A bit of diversity is observed in the choices for ML algorithms. Despite the diversity, Random Forest appears to be the most common choice of many researchers.

4. CONCLUSIONS AND RECOMMENDATIONS

Air pollution has been linked to many severe health and economic risks globally. Despite the efforts of research, predictive models have not been able to achieve high performance in predicting air pollution. In this study, we reviewed research in air pollution prediction reported in literature published within the recent 5years. The result of this study showed that $PM_{2.5}$, PM_{10} , O_3 , NO_2 , SO_2 and CO are the commonly predicted pollutants. Despite the availability of many datasets, researchers in this domain cannot avoid the use of Air Quality and Meteorological datasets. Random Forest appears to perform well for various combinations of datasets. As a way forward, we envisage a larger literature review to explore further the pattern and the best practice in the domain of air pollution prediction using ML.

5. REFERENCES

- 68% of the world population projected to live in urban areas by 2050, says UN | UN DESA | United Nations Department of Economic and Social Affairs. (2018, May 16). <https://www.un.org/development/desa/en/news/population/2018-revision-of-world-urbanization-prospects.html>
- Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O., & Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications*, 94, 164–184. <https://doi.org/10.1016/j.eswa.2017.10.040>
- Alpan, K., & Sekeroglu, B. (2020). Prediction of pollutant concentrations by meteorological data using machine learning algorithms. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 44(4/W3), 21–27. <https://doi.org/10.5194/isprs-archives-XLIV-4-W3-2020-21-2020>
- Anurag, N. V., Burra, Y., Sharanya, S., & Gireeshan, M. G. (2019). Air quality index prediction using meteorological data using featured based weighted xgboost. *International Journal of Innovative Technology and Exploring Engineering*, 8(11 Special Issue), 1026–1029. <https://doi.org/10.35940/ijitee.K1211.09811S19>
- Ashayeri, M., Abbasabadi, N., Heidarinejad, M., & Stephens, B. (2021). Predicting intraurban PM_{2.5} concentrations using enhanced machine learning approaches and incorporating human activity patterns. *Environmental Research*, 196. <https://doi.org/10.1016/j.envres.2020.110423>
- Babu, K. M., & Beulah, J. R. (2019). Air quality prediction based on supervised machine learning methods. *International Journal of Innovative Technology and Exploring Engineering*, 8(9 Special Issue 4), 206–212. <https://doi.org/10.35940/ijitee.I1132.0789S419>
- Bozdağ, A., Dokuz, Y., & Gökçek, Ö. B. (2020). Spatial prediction of PM₁₀ concentration using machine learning algorithms in Ankara, Turkey. *Environmental Pollution*, 263. <https://doi.org/10.1016/j.envpol.2020.114635>
- Chen, G., Li, S., Knibbs, L. D., Hamm, N. A. S., Cao, W., Li, T., Guo, J., Ren, H., Abramson, M. J., & Guo, Y. (2018). A machine learning method to estimate PM_{2.5} concentrations across China with remote sensing, meteorological and land use information. *Science of the Total Environment*, 636, 52–60. <https://doi.org/10.1016/j.scitotenv.2018.04.251>
- Chen, G., Wang, Y., Li, S., Cao, W., Ren, H., Knibbs, L. D., Abramson, M. J., & Guo, Y. (2018). Spatiotemporal patterns of PM₁₀ concentrations over China during 2005–2016: A satellite-based estimation using the random forests approach. *Environmental Pollution*, 242, 605–613. <https://doi.org/10.1016/j.envpol.2018.07.012>
- Cihan, P., Ozel, H., & Ozcan, H. K. (2021). Modeling of atmospheric particulate matters via artificial intelligence methods. *Environmental Monitoring and Assessment*, 193(5). <https://doi.org/10.1007/s10661-021-09091-1>
- Department for Environment, Food and Rural Affairs (Defra), Pollution forecast provided by the Met Office-Defra, UK (UK; United Kingdom). Department for Environment, Food and Rural Affairs (Defra), Nobel House, 17 Smith Square, London SW1P 3JR helpline@defra.gsi.gov.uk. Retrieved 1 March 2021, from <https://uk-air.defra.gov.uk/forecasting/>
- Dobrea, M., Badicu, A., Barbu, M., Subea, O., Balanescu, M., Suci, G., Birdici, A., Orza, O., & Dobre, C. (2020). Machine Learning algorithms for air pollutants forecasting. *2020 IEEE 26th International Symposium for Design and Technology in Electronic Packaging, SIITME 2020 - Conference Proceedings*, 109–113. <https://doi.org/10.1109/SIITME50350.2020.9292238>
- Gul, S., & Khan, G. M. (2020). Forecasting Hazard Level of Air Pollutants Using LSTM's. In I. Maglogiannis, L. Iliadis, & E. Pimenidis (Eds.), *Artificial Intelligence Applications and Innovations* (pp. 143–153). Springer International Publishing. https://doi.org/10.1007/978-3-030-49186-4_13

- Huang, K., Xiao, Q., Meng, X., Geng, G., Wang, Y., Lyapustin, A., Gu, D., & Liu, Y. (2018). Predicting monthly high-resolution PM_{2.5} concentrations with random forest model in the North China Plain. *Environmental Pollution*, 242, 675–683. <https://doi.org/10.1016/j.envpol.2018.07.016>
- Iskandaryan, D., Ramos, F., & Trilles, S. (2020). Air Quality Prediction in Smart Cities Using Machine Learning Technologies Based on Sensor Data: A Review. *Applied Sciences*, 10(7), 2401. <https://doi.org/10.3390/app10072401>
- Jo, B. W., & Khan, R. M. A. (2018). An internet of things system for underground mine air quality pollutant prediction based on azure machine learning. *Sensors (Switzerland)*, 18(4). <https://doi.org/10.3390/s18040930>
- Kumari, N. S. A., Kumar, K. S. A., Raju, S. H. V., Vasuki, H. R., & Nikesh, M. P. (2020). Prediction of Air Quality in Industrial Area. Proceedings - 5th IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, RTEICT 2020, 193–198. <https://doi.org/10.1109/RTEICT49044.2020.9315660>
- Lin, Y., Zhao, L., Li, H., & Sun, Y. (2018). Air quality forecasting based on cloud model granulation. *Eurasip Journal on Wireless Communications and Networking*, 2018(1). <https://doi.org/10.1186/s13638-018-1116-3>
- Martinez-Espana, R., Bueno-Crespo, A., Timon, I., Soto, J., Munoz, A., & Cecilia, J. M. (2018). Air-Pollution Prediction in Smart Cities through Machine Learning Methods: A Case of Study in Murcia, Spain. 16.
- Masih, A. (2019). Machine learning algorithms in air quality modeling. *Global Journal of Environmental Science and Management*, 5(4). <https://doi.org/10.22034/GJESM.2019.04.10>
- Mo, X., Zhang, L., Li, H., & Qu, Z. (2019). A novel air quality early-warning system based on artificial intelligence. *International Journal of Environmental Research and Public Health*, 16(19). <https://doi.org/10.3390/ijerph16193505>
- Ni, X. Y., Huang, H., & Du, W. P. (2017). Relevance analysis and short-term prediction of PM_{2.5} concentrations in Beijing based on multi-source data. *Atmospheric Environment*, 150, 146–161. <https://doi.org/10.1016/j.atmosenv.2016.11.054>
- Ray, R., Haldar, S., Biswas, S., Mukherjee, R., Banerjee, S., & Chatterjee, S. (2019). Prediction of benzene concentration of air in urban area using deep neural network. *Advances in Intelligent Systems and Computing*, 811, 465–475. https://doi.org/10.1007/978-981-13-1544-2_38
- Rossi, C., Farasin, A., Falcone, G., & Castelluccio, C. (2019a). A Machine Learning Approach to Monitor Air Quality from Traffic and Weather data. *CEUR Workshop Proceedings*, 2492, 66–74.
- Rossi, C., Farasin, A., Falcone, G., & Castelluccio, C. (2019b). UAQE: Urban air quality evaluator. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 11912 LNCS, 337–343. https://doi.org/10.1007/978-3-030-34255-5_25
- Rybarczyk, Y., & Zalakeviciute, R. (2018). Machine Learning Approaches for Outdoor Air Quality Modelling: A Systematic Review. *Applied Sciences*, 8(12), 2570. <https://doi.org/10.3390/app8122570>
- Selvi, S., & Chandrasekaran, M. (2020). Framework to forecast environment changes by optimized predictive modelling based on rough set and Elman neural network. *Soft Computing*, 24(14), 10467–10480. <https://doi.org/10.1007/s00500-019-04556-5>
- Shah, S. K., Tariq, Z., Lee, J., & Lee, Y. (2020). Real-Time Machine Learning for Air Quality and Environmental Noise Detection. Proceedings - 2020 IEEE International Conference on Big Data, Big Data 2020, 3506–3515. <https://doi.org/10.1109/BigData50022.2020.9377939>
- Srivastava, C., Singh, S., & Singh, A. P. (2019). Estimation of air pollution in Delhi using machine learning techniques. 2018 International Conference on Computing, Power and Communication Technologies, GUCON 2018, 304–309. <https://doi.org/10.1109/GUCON.2018.8675022>

- Tao, Q., Liu, F., Li, Y., & Sidorov, D. (2019). Air Pollution Forecasting Using a Deep Learning Model Based on 1D Convnets and Bidirectional GRU. *IEEE Access*, 7, 76690–76698. <https://doi.org/10.1109/ACCESS.2019.2921578>
- WHO | 7 million premature deaths annually linked to air pollution. (2014, March 25). WHO; World Health Organization. <https://www.who.int/mediacentre/news/releases/2014/air-pollution/en/#.WqBfue47NRQ.mendeley>
- Wu, Q., & Lin, H. (2019). A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors. *Science of the Total Environment*, 683, 808–821. <https://doi.org/10.1016/j.scitotenv.2019.05.288>
- Xu, Y., Ho, H. C., Wong, M. S., Deng, C., Shi, Y., Chan, T. C., & Knudby, A. (2018). Evaluation of machine learning techniques with multiple remote sensing datasets in estimating monthly concentrations of ground-level PM_{2.5}. *Environmental Pollution*, 242, 1417–1426. <https://doi.org/10.1016/j.envpol.2018.08.029>
- Yafouz, A., Ahmed, A. N., Zaini, N., & El-Shafie, A. (2021). Ozone Concentration Forecasting Based on Artificial Intelligence Techniques: A Systematic Review. *Water, Air, & Soil Pollution*, 232(2), 79. <https://doi.org/10.1007/s11270-021-04989-5>
- Yafouz, A., Ahmed, A. N., Zaini, N., Sherif, M., Sefelnasr, A., & El-Shafie, A. (2021). Hybrid deep learning model for ozone concentration prediction: Comprehensive evaluation and comparison with various machine and deep learning algorithms. *Engineering Applications of Computational Fluid Mechanics*, 15(1), 902–933. <https://doi.org/10.1080/19942060.2021.1926328>
- Yarragunta, S., Nabi, M. A., P, J., & S, R. (2021). Prediction of Air Pollutants Using Supervised Machine Learning. *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 1633–1640. <https://doi.org/10.1109/ICICCS51141.2021.9432078>
- Zhang, J., & Ding, W. (2017). Prediction of Air Pollutants Concentration Based on an Extreme Learning Machine: The Case of Hong Kong. *International Journal of Environmental Research and Public Health*, 14(2), 114. <https://doi.org/10.3390/ijerph14020114>
- Zhang, S., Li, X., Li, Y., & Mei, J. (2018). Prediction of Urban PM_{2.5} Concentration Based on Wavelet Neural Network. *Proceedings of the 30th Chinese Control and Decision Conference, CCDC 2018*, 5514–5519. <https://doi.org/10.1109/CCDC.2018.8408092>
- Zhao, C., Heeswijk, M. van, & Karhunen, J. (2016). Air quality forecasting using neural networks. 2016 *IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–7. <https://doi.org/10.1109/SSCI.2016.7850128>