International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-15 Issue-3, February 2026

Air Pollution and the Monitoring of Environmental Health Compared with Logistic Regression (LR) and Random Forest (RF) Algorithms

Nirmla Sharma, Sameera Iqbal Muhmmad Iqbal

Abstract**: *Nowadays, nine out of ten people inhale polluted air, causing dangerous health concerns. This means that air pollution poses a serious threat to society's health. It supports enhanced dimension, cause detection, prediction, expectation, and logical problem-solving. AI technology can rapidly and accurately detect air pollution. AI has quickly exposed the extent of air pollution. This study estimates logistic regression (LR) and Random Forest (RF) models, two widely used statistical methods for predicting long-term air pollution and environmental health. Logistic regression may predict air pollution more effectively than other machine learning approaches. The objective of this analysis is to improve the algorithm's performance during the collection activity and reduce air pollution. The average detection accuracy falls within one standard deviation, indicating that the proposed model is as efficient as, and more effective than, the modern method. Logistic Regression and Random Forest (which is valued the highest accuracy (0.93) and precision (0.92)).***

Keywords**: *Air Pollution, Logistic Regression, Random Forest Classifier, Machine Learning, Naive Bayes.***

Nomenclature:
RF: Random Forest
LR: Logistic Regression

I. INTRODUCTION

The effects of air pollution are more readily assessed using AI. By mitigating the effects of environmental alteration, artificial intelligence has demonstrated the most significant potential to protect the Earth. Industrialised air pollution harms the environment, degrading current sources. [1]. Manufacturing pollution also contributes to global climate change and harms the environment. Since the effect is harmful to people, it is critical to monitor pollution to reduce its adverse impact on the environment, which can lead to the loss of current sources for future generations [2]. Presently, AI technology is advancing rapidly. These technologies have been used across numerous areas, such as production, health, and education [3].

Manuscript received on 08 January 2026 | First Revised Manuscript received on 15 January 2026 | Second Revised Manuscript received on 03 February 2026 | Manuscript Accepted on 15 February 2026 | Manuscript published on 28 February 2026.
*Correspondence Author(s)
Dr. Nirmla Sharma*, Asst. Professor, Department of Computer Science, King Khalid University, Abha, Kingdom of Saudi Arabia. Email ID: nprasad@kku.edu.sa, ORCID ID: 0009-0007-0746-1001
Sameera Iqbal Muhmmad Iqbal, Lecturer, Department of Computer Science, King Khalid University, Abha, Kingdom of Saudi Arabia. Email ID: eqbal@kku.edu.sa, ORCID ID: 0009-0005-7812-4593

AI technologies have transformed the way humans live. Numerous technologies have been used in the environmental protection sector, such as nanotechnology, laser technology, chemical approaches, and radars, to identify and monitor air pollution [4],[5]. Among the latest technologies that can address environmental air pollution, Artificial Intelligence stands out. In these technologies, Logistic Regression (LR) is the algorithm that can reduce air pollution and monitor it [6], [7].

When it comes to locating hotspots, Random Forest has their benefits and drawbacks. However, Random Forest's ability to handle complex relationships among variables enables it to perform well on larger datasets. The machine learning technique, Novel Tree Specific Random Forest, is straightforward to apply and often yields strong results even without hyperparameter tuning. Its elasticity and ease of execution have made it a standard algorithm (it can be used for classification and regression tasks) [8], [9].

Tolerant and prohibited burning of wood and other plant materials in backwoods areas, such as forests and tundra regions, is a significant source of air pollution. Air pollution is an environmental health issue and a major contributor to illness and premature mortality worldwide. Although billions of people are adversely affected by air pollution, there are few consistent methods available to address environmental issues. Air pollution is hazardous to health [10], [11].

II. RELATED STUDIES

Air pollution must be mitigated in developing countries such as India, where heavy road traffic is prevalent, to protect the environment [12]. Air pollution is influenced by industry, society, and transportation [13], [14]. The study announced in 2016 ("Assessing the Potential of Random Forest Method for Estimating Solar Radiation Using Air Pollution Index") found that operational statistics from the air pollution index yielded more accurate predictions of solar radiation. To detect production resources, wind direction, and physicochemical variation, air pollution attention varies considerably over short distances [15]. "Prediction of Air Pollutants Using Supervised Machine Learning" (n.d.); Application of Airnet: A Machine Learning Dataset for Air Quality Projecting (Zhao et al., 2018). Our cluster has vast study capability in numerous disciplines [16]-[21] (Saraswathi et al. 2020; Gariazzo et al., 2020; Chozhavendhan et al. 2020; Mohan et al. 2020; Pradeep Kumar et al. 2021; Dinesh et al. 2019; Babu and Jayaraman 2020; Shanmugam et al. 2021; Mehta et al. 2020; Vairavel, Devaraj, and Shanmugam 2020; S. Arunkumar et al. 2020; "Integrating Mobile

Measurement Data on Air Pollution to Create an Initial Spatial-Temporal Air Pollution Profile in Two Border Towns Between Germany and the Czech Republic". Using Inexpensive Sensors to Track Carbon Monoxide, Nitrogen Dioxide, Oxygen, and Sulfur Dioxide in the Air (Han et al. 2021)" [22]. However, earlier in the fire, in February 2022, PM10 and PM2.5 concentrations were 72 ppm and 53 ppm, respectively. These digits then suddenly bounced to 192 ppm and 119 ppm, respectively. Since the beginning of 2023, 1.3 million people have been informed about illnesses linked to air pollution. Infants and pregnant women were instructed to remain indoors and wear N95 masks. In March 2023, Kochi was observed as the most polluted town in Kerala. Comparisons of modification values and random forest models indicate that air pollution indicators outperform solar radiation values in the current study [22].

III. METHODS AND MATERIALS

A. Algorithm of Logistic Regression (LR)

Source: Air pollution data and environmental health monitoring.

i. Output: Accuracy

1. Open and read the dataset
2. Arbitrarily select features from dataset 4 of 15.
3. Produce the Logistic Regression classifier consequence restriction.
4. l2 was the restriction rate.
5. Differ dependent and independent factors to analyze the dataset.
6. Logistic Regression predicts definite results.
7. Lastly, log the predictions incident option.

The Logistic Regression class from sklearn. Linear model packages are used in this research. We used this single method to select a corrective device. Parameterized by consequence. The parameter is "L2". The dataset has 20% training and 10% testing shares (20 percent). It randomly selects a dataset and variations of the dependent and independent variables. Lastly, the log function predicts probability.

This algorithm may vary in several contexts, such as a job, a fasting period, or a computing position. By working through this procedure, AI determines whether the air is polluted. If air pollution is severe, an optimal solution should be implemented; otherwise, statistical management should be employed.

It is applied to develop excellence, to measure the components of banned particulate matter, and to establish a digital boundary in production. The sensor will determine the pollution ratio and monitor the environment at distinct intervals throughout the year.

Next, when we detect a reaction from the sensor in a selected zone, we will recognise that the air in that region is polluted and being monitored. As noted in [23], certain plants generate oxygen for 24 hours and filter the air, such as Peepal, Tulsi, Aloe Vera, Neem, Money Plant, and Bell Patri. If it is feasible to compute the exact number of plants required to reduce air pollution in the region, programs can be developed; however, this remains uncertain and warrants further investigation.

B. Algorithm of Random Forest Classifier (RF)

The air pollution dataset and environmental health monitoring data are the inputs.

i. Accuracy is the outcome.

1. Import the dataset and begin scanning it.
2. Picking the attributes from the dataset using a random technique.
3. Creating the Random Forest classification conditions as a restriction in your database.
4. The cost of the Random Forest Classifier worked for running within a limit.
5. Has created a decision tree by using Random Forest rare classifiers and has produced a prediction about the result of all models.
6. Polling was performed for each result that was predicted.
7. The result was decided by picking the predicted results with the most polls.

Random Forest is one of the most commonly used machine learning algorithms. It is appropriate for use with supervised learning methods. The algorithm is suitable for generating predictions and behaviour. Therefore, this approach ultimately constitutes a decision tree. To make an image, in a vast amount. Thus, each section of these decision trees reflects a clustering, or straightforward classification, of the statistics that have been directed into the random forest. This clustering for the individual decision tree is distinct. Accordingly, the random forest approach selects all these associations to view them distinctly. And, as a result, it becomes the specific one that indicates the common number of polls [24], [25]. In most cases, it follows a confirmed prediction. It is best used for predictions. It can be applied to both classification and regression. It is generally established on a solution of a combination be trained. It entails combining multiple classifiers to develop an explanation of a real-world problem and, simultaneously, advancing the simulation's implementation. This algorithm is popular because, among all classification methods, it offers the highest precision. To enhance prediction accuracy in an individual dataset, the random forest algorithm builds numerous decision trees across many fields of that dataset [26]. Thus, it does not determine the outcome of a single decision tree. It then assesses the perceived findings across all the trees. Then, depending on the number of polls supporting a given prediction, it sets the final yield prediction. That's how it improves its precision. Therefore, the more decision trees there are, the more precise the predictions become [27].

We will improve the two approaches described above (LR and RFC) by adding a feature that generates a diagram of an individual region. We know that trees serve as a clean source of oxygen and a pollution-free environment. The structure will provide a tree count for a zone and a rough estimate of how many trees are needed to reduce the zone's air pollution as a stand-level impact. At this point, the task is to relate image processing and Artificial Intelligence-based algorithms to the search for air pollution and to present an optimal solution that can be implemented in practice.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-15 Issue-3, February 2026

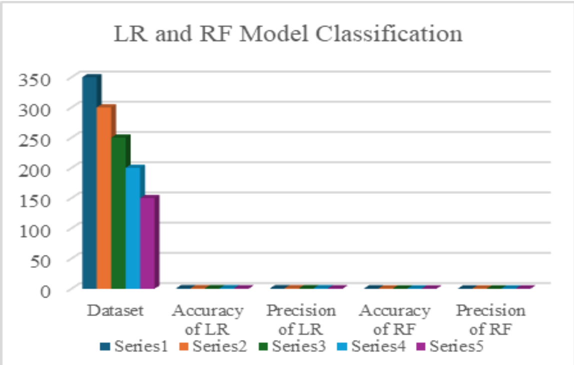
IV. RESULT DISCUSSION

Statistical analysis was executed on the experimental data. We have conducted a precise comparison using both the LR and RF algorithms. The two models' restrictions were compared using a T-Test, with the independent variable (vertical and horizontal distance) as the dependent variable. Accuracy and precision are the dependent issues that influence the result.

Table 1 and Figure 1 present the results of LR and RF classification models with varying repetitions, with respect to accuracy and precision.

Table I: Using LR and RF Models at Several Repetitions: Accuracy and Precision of Classification Challenges.

Repetitions	Dataset	Accuracy of LR	Precision of LR	Accuracy of RF	Precision of RF
1	350	93%	93%	63%	58%
2	300	92%	92%	62%	57%
3	250	91%	91%	60%	55%
4	200	89%	89%	58%	54%
5	150	84%	84%	56%	53%

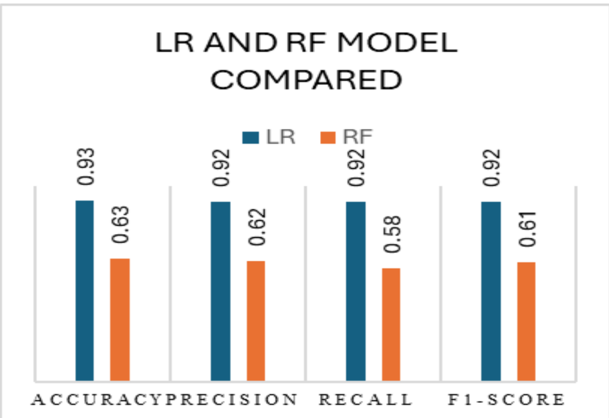


[Fig.1: Using LR and RF Models at Several Repetitions: Accuracy and Precision of Classification Challenges]

Table 2 and Figure 2 compare the separate restrictions of both models. We have determined their Accuracy, Precision, Recall, and F1-score for both LR and RF. A comparison of the two models shows that LR is more accurate (0.93) and precise (0.92) than RF.

Table II: Compares Several Attributes of the two Models

Model	Accuracy	Precision	Recall	F1-score
LR	0.93	0.92	0.92	0.92
RF	0.63	0.62	0.58	0.61



[Fig.2: Shows a Comparison of the Several Attributes of the Two Models' Graphs]

Measurements for LR and RF on a type of test dataset are indicated in Table 1 and Figure 1. Statistically, the algorithms have been shown to differ in accuracy—LR's 0.93 and RF's 0.63. Lastly, in comparative statistics, LRs outperform RFs.

A. Result 1: Logistic Regression Accuracy Test:

We will enhance a further feature that provides a tree calculation for a specific zone. As we know, trees act as filters to create a pollution-free environment. The task now is to integrate statistical analysis and AI-based algorithms not only to detect air pollution but also to deliver an optimal solution that can be applied in practice. In Table 1 and Figure 1, we present the accuracy statistics for the Logistic Regression model; our accuracy was 0.93.

B. Result 2: Random Forest Algorithm:

The random forest algorithm is based on multiple calculations, the most significant of which is the accuracy of predictions. Table 2 and Figure 2 show the accuracy statistics for the linear random forest algorithm; our accuracy was 0.63.

C. Result 3: The Proposed Two Models Compared (LR and RF)

Comparatively, LR's 0.93 accuracy and RF's 0.63. Finally, the comparative statistics indicate that LR outperforms RFs. Based on the SPSS plot and statistical analysis, the proposed LR algorithm outperforms the RF models in accuracy. Here, in Table 2 and Figure 2, are the statistics for the accuracy scores we used for the Logistic Regression and the Random Forest algorithm, as well as our accuracy.

V. CONCLUSION

This research finds that Logistic Regression outperforms Random Forest in predicting air pollution, achieving the highest accuracy (0.93) and precision (0.94). Logistic regression differs from random forests; these models make more accurate and precise predictions. Air pollution is a major problem with significant impacts on human health and the environment worldwide. To block activities and reduce the effects of air pollution, it is vital to identify problem areas. Logistic Regression and Random Forest are two AI-based algorithms that can effectively analyse data and identify areas of air pollution.

With this technique, we can better control and monitor pollution sources, allocate funds, and improve air quality for everyone. Categorising statistics using chance controls is straightforward and rapid with LR. Our assumption is recognised by the situation that raises, slope, aspect, horizontal distance to hydrology, and vertical distance to hydrology are the independent variables in the research.

VI. FUTURE WORK

Furthermore, this research has categorised statistics using chance controls, which is completely easy and rapid using LR. Moreover, the hypothesis is supported by the fact that the variables raised, slope, aspect, straight distance to hydrology, and perpendicular distance to hydrology are the independent variables in the

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

research. Accuracy and Precision are the dependent variables that more strongly influence the outcome.

DECLARATION STATEMENT

Authors are required to include a declaration of accountability in the article, including review-type articles, that stipulates the involvement of each author. The level of detail differs: some subjects yield articles that consist of isolated efforts that are easily detailed, while others involve group efforts at all stages. It should be after the conclusion and before the references.

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted objectively and free from external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. Anna Sołtysik-Piorunkiewicz and Iwona Zdonek, "How Society 5.0 and Industry 4.0 Ideas Shape the Open Data Performance Expectancy", Sustainability vol. 13(2), 917, 2021; DOI: https://doi.org/10.3390/su13020917
2. Zahra Nasrollahi, et al., "Environmental pollution, economic growth, population, industrialisation, and technology in weak and strong sustainability: using STIRPAT model", Springer, Environment Development and Sustainability, February vol. 22(C):1-18, 2020. DOI: https://doi.org/10.1007/s10668-018-0237-5
3. Muhammad Faizal Fazri, et al., "Implementing Artificial Intelligence to Reduce Marine Ecosystem Pollution", Vol. 4 No. 2, April, 2023. DOI: https://doi.org/10.34306/itsdi.v4i2.579
4. Prakriti Rashi, et al., "New Personalized Social Approach Based on Flexible Integration of Web Services", Vol. 1 No. 1, November, 2022. DOI: https://doi.org/10.33050/italic.v1i1.85
5. Botao Xie, et al., "Research Progress in Marine Environmental Monitoring Technology", Vol. 7, Iss. 8, August 2025. DOI: https://doi.org/10.30564/jees.v7i8.10742
6. Davide Moroni, et al., "Environmental decision support systems for monitoring small-scale oil spills: Existing solutions, best practices, and current challenges," MDPI, Journal of Marine Science and Engineering (JMSE), Vol. 7(1):19, January 2019. DOI: https://doi.org/10.3390/jmse7010019
7. Jagadeesh Sairam, M et al., "Prediction of air pollution hotspots to prevent post effects of pollution by comparing logistic regression with random forest", *AIP Conf. Proc.*, 2853, 020072, *2024*. DOI: https://doi.org/10.1063/5.0203747
8. Jagadeesh Sairam, M. and Nagaraju, V., "Air Pollution Hotspot Identification to Prevent Post Effects of Pollution using Naive Bayes over Random Forest", *AIP Conf. Proc.,* 2853, 020071, *2024*. DOI: https://doi.org/10.1063/5.0197506
9. Alsaber, Ahmad R., Jiazhu Pan, and Adeeba Al-Hurban, "Handling Complex Missing Data Using Random Forest Approach for an Air Quality Monitoring Dataset: A Case Study of Kuwait Environmental Data (2012 to 2018)." International Journal of Environmental Research and Public Health, Vol. 18(3), 1333, 2021. DOI: https://doi.org/10.3390/ijerph18031333
10. Izuta, Takeshi, "Air Pollution Impacts on Plants in East Asia", 2017, Springer. https://link.springer.com/book/10.1007/978-4-431-56438-6
11. Zeybek, Mustafa, "Classification of UAV point clouds by random forest Machine learning algorithm." Turkish Journal of Engineering, 2021. DOI: https://doi.org/10.31127/tuje.669566.
12. Georgios C. Spyropoulos et al., "Performance of Aether Low-Cost Sensor Device for Air Pollution Measurements in Urban Environments. Accuracy Evaluation Applying the Air Quality Index (AQI)",12(10), pp. 1246, 24, September 2021. DOI: https://doi.org/10.3390/atmos12101246
13. Farzaneh Abedian Aval, et al., "Machine Learning on the Frontlines of Air Pollution and Public Health: Revealing the Connection with Hospital Admissions", pp. 17 (1), 17, 2025. DOI: https://doi.org/10.3390/atmos17010017
14. Evangelos Bagkis, et al., "Evolving trends in application of low-cost air quality sensor networks: challenges and future directions" Nature, npj climate and atmospheric science, 8, Article number: 335, 2025. https://www.nature.com/articles/s41612-025-01216-4
15. Tetiana Vovk, et al., "Using random forest to improve EMEP4PL model estimates of daily PM2.5 in Poland", Atmospheric Environment Volume 332, 120615 1 September 2024. DOI: https://doi.org/10.1016/j.atmosenv.2024.120615
16. Claudio Gariazzo et al., "A multi-city air pollution population exposure study: Combined use of chemical-transport and random-Forest models with dynamic population data" Science of The Total Environment Volume 724, 138102, 1 July 2020. DOI: https://doi.org/10.1016/j.scitotenv.2020.138102
17. Mohammed H. Yas, et al., "Development of an Engineered Slurry-Infiltrated Fibrous Concrete: Experimental and Modelling Approaches", infrastructures, 8(2), pp. 19, 2023. DOI: https://doi.org/10.3390/infrastructures8020019.
18. Pradeep Kumar, et al., "Diagnosis of Vertical Root Fractures by Cone-Beam Computed Tomography in Root-Filled Teeth with Confirmation by Direct Visualization: A Systematic Review and Meta-Analysis", Journal of Endodontia, 47 (8): pp.1198–1214, 2021. DOI: https://doi.org/10.1016/j.joen.2021.04.022
19. Dinesh, et al., "Influence of Wood Dust Fillers on the Mechanical, Thermal, Water Absorption and Biodegradation Characteristics of Jute Fibre Epoxy Composites", Journal of Polymer Research, Vol. 27 (1): pp.9, 2020. DOI: https://doi.org/10.1007/s10965-019-1975-2
20. Shyamaladevi Babu, et al., "An Update on β-Sitosterol: A Potential Herbal Nutraceutical for Diabetic Management", Biomedicine & Pharmacotherapy Vol. 131, pp. 110702, November 2020. DOI: https://doi.org/10.1016/j.biopha.2020.110702
21. Shanmugam, et al., "Fatigue Behaviour of FDM-3D Printed Polymers, Polymeric Composites and Architected Cellular Materials", International Journal of Fatigue, Vol. 143, pp. 106007, 2021. DOI: https://doi.org/10.1016/j.ijfatigue.2020.106007
22. Sathish, T. et al., "Experimental Investigation on the material characterisation of zirconia reinforced Alumina ceramic composites via powder forming process", *AIP Conference Proceedings, 2020*, 2283 (1), pp. 020124, Oct 2020. DOI: https://doi.org/10.1063/5.0024964
23. Pengfei Han, et al., "Calibrations of Low-Cost Air Pollution Monitoring Sensors for CO, NO2, O3, and SO2" Sensors, 21 (1): pp. 256, 2021 DOI: https://doi.org/10.3390/s21010256
24. Zeybek, Mustafa. 2021. "Classification of UAV point clouds by a random forest Machine learning algorithm." Turkish Journal of Engineering, 2021. DOI: https://doi.org/10.31127/tuje.669566.
25. Reis, et al., "Probabilistic Random Forest: A Machine Learning Algorithm for Noisy Data Sets", The Astronomical Journal, 2018. DOI: https://doi.org/10.3847/1538-3881/aaf101.
26. Mangesh Maurya et al., "A study of different machine learning algorithms for state of charge estimation in lithium-ion battery pack", Volume 6, Issue 4, 25 June 2024. DOI: https://doi.org/10.1002/est2.658
27. Thibaut Cassard, et al., "PlumeCityNet: Multi-Resolution Air Quality Forecasting", version v2, 14 October 2021. DOI: https://doi.org/10.48550/arXiv.2110.02661

AUTHOR'S PROFILE

Dr. Nirmla Sharma, PhD, from Teerthanker Mahaveer University, Muradabad, U.P., INDIA. Currently working at King Khalid University, Abha, Saudi Arabia, as an Asst. Prof. Department of Computer Science. Initially, graduated from CCS University, Meerut, U.P., India, and then a master's in computer science from Rajasthan Vidyapeeth, Rajasthan, and MCA from Indira Gandhi, New Delhi. Published 28 Papers in International Journals, 02 in National Journals, 7

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

National Conferences, attended 14 International Conferences, 15 National Workshops/Conferences, 2 books, and 01 patent. Other responsibilities, i.e., Head, Dept. Of CSE and Timetable Convener at AIT, Ghaziabad, India, Head Examiner for different subjects of C.S. and I.T. in Central Evaluation of M.T.U. NOIDA / U.P.T.U., Lucknow, U.P. Paper Setter/Practical Examiner at different Institutes/Universities, time to time, i.e., CCSU Meerut/UPTU, Lucknow.

Sameera Iqbal Muhmmad Iqbal, MCS, from The Islamia University of Bahawalpur, Pakistan. Currently working at King Khalid University, Abha, Saudi Arabia, as a Lecturer in the Department of Computer Science. Initially, graduated from the Islamia University of Bahawalpur, Pakistan. Published 4 Papers in International Journals, attended 3 International Conferences, and 01 patent. Teaching Computer Science courses.