International Journal of Computer Applications (0975 – 8887)
Volume 185 – No.14, June 2023

Stratified Prediction of the Air Pollution Index using Operational Intelligence

Alaa Ibrahem
Computer Science Dept.
Faculty of Computers and Information
Menoufia University, Egypt

Heba Elbeh
Computer Science Dept.
Faculty of Computers and Information
Menoufia University, Egypt

Hamdy M. Mousa
Computer Science Dept.
Faculty of Computers and Information
Menoufia University, Egypt

ABSTRACT
Our daily choices and lifestyle can affect the decrease or increase in air pollution, the overall health index, and the ozone layer. Unfortunately, ozone damage leads to the loss of 121 million tons of crops worldwide. In addition, the percentage of particulate matter (PM2.5) and surface ozone levels exceeded the limits set by the World Health Organization (WHO). The particles threatened most parts of Asia, Africa, and South America, causing sudden smog and lung disease. The Chinese central government has taken measures to improve the Air Quality Index (AQI) in Beijing and other regions. Measures have been taken to reduce primary emissions of gases and particulate matter. This study aimed to make a class prediction of the air quality index category in the Beijing region based on ozone and PM2.5 concentrations. The Internet of Things (IoT) has been used to collect sensor air quality data. The Internet of Things monitors air pollution and calculates levels of harmful gases, particulate matter, smoke, and more. This study focused on the use of operational intelligence in developing machine learning models for air quality index classification and comparing the results of classification models between the random forest classifier and the decision tree classifier. It helps in processing large amounts of data and making effective decisions. The random forest classifier has proven its efficiency in predicting the level of the air quality index. Temporal and spatial characteristics for the effective prediction of AQI were also considered.

Keywords
Internet of Things (IoT), Operational intelligence, Air pollution index, Splunk, Classifiers.

1. INTRODUCTION
The Internet of Things (IoT) is the most popular term, but it is even more important because it drives innovation by making it easier to analyze and use data. Due to the rapid growth of the Internet of Things (IoT), applications have evolved to keep pace with their speed and diversity [1]. Sensors have been added to many national and international standard devices and systems that collect measurements of weather, pollution, and medical data. With the exponential increase in sensors and smart devices, huge amounts of static data are being generated. Thus, it is necessary to develop technologies and tools to solve these problems and to deploy the Internet of Things [1].

The exponential growth of data in recent years has posed significant scientific challenges in terms of how to effectively store, process, analyze, and interpret such vast amounts of information. Furthermore, the sheer complexity and heterogeneity of big data sources have resulted in expanding research efforts to develop new approaches and tools for tackling these scientific problems [18]. Big data poses a major scientific problem in recent studies due to the huge volume, speed, and diversity of information, which requires advanced analytical tools and techniques to extract meaningful insights. Moreover, the accuracy and reliability of big data analysis relies heavily on the quality and integrity of the data sources, as well as the expertise and experience of the data scientists who perform the analysis. This challenge is the aim of this study, starting with collecting the data generated from the sensors and examining their integrity and accuracy. Next, advanced data analysis tools based on operational intelligence are used to improve the quality of training models, extract information, and assist decision-makers.

Machine learning is a rapidly growing field in computer science and artificial intelligence that focuses on developing algorithms and models to analyze data, make predictions, and automate various tasks. A machine learning tool can be an invaluable asset for companies looking to improve their operations and gain insights from their data. As companies increasingly rely on data-driven decision-making, machine learning, and big data analytics are becoming essential tools for gaining insights and extracting value from large data sets.

Operational intelligence is a term that describes a system for interacting with big data and is well suited for the analytics of large amounts of data. It allows organizations to create an intelligent operating environment, which is superior to traditional business intelligence because it features real-time data processing [15]. Operational intelligence technology can be developed through a variety of tools such as machine learning algorithms, data visualization platforms, real-time monitoring systems, predictive analytics software, and automated alert mechanisms. These tools enable organizations to collect, process, and analyze data from various sources, identify patterns and anomalies in real-time, and gain valuable insights into their operations. Operational intelligence plays an important role in data analysis because it focuses on providing tangible insights that improve business results and operational efficiency and predict future results based on historical data, which helps to take proactive measures before problems appear to occur.

Air pollution is one of the most serious threats to global health worldwide. Seven million people suffer from health crises due to air pollution [13]. The value of the air quality index was determined based on the concentration of the air pollutant at the highest concentration at the time the index was measured. Air pollutants include six types: particulate matter 2.5, particle 10, carbon monoxide, sulfur dioxide, nitrogen dioxide, and ground-level ozone. Particulate matter and ozone are important pollutants affecting climate change and human health. Predicting the air quality index is one of the most important challenges in smart cities. Ozone gas is found in Earth's upper atmosphere, 6-30 miles above the surface, where it forms a protective layer that protects us from the sun's harmful ultraviolet rays. While ozone occurs naturally in small amounts in the Earth's lower atmosphere, additional ozone is formed when pollutants are released from automobiles, power plants, chemical plants, and other sources [9].

High concentrations of ozone at ground level are harmful air pollutants. It should also be noted that the World Health Organization has set a standard for the average daily concentration of PM2.5 at 25 micrograms per cubic meter and the annual standard at 10 micrograms per cubic meter. Most of the particles in the atmosphere are produced through complex reactions of chemicals such as sulfur dioxide and nitrogen oxides emitted by power plants, industry, and automobiles. Because these particles are so small and light, they tend to stay in the air for a long time, which increases the risk of being inhaled. PM2.5 is very small and therefore deeply penetrates the respiratory tract. Exposure to PM2.5 causes eye and lung irritation, coughing, and exacerbation of conditions such as asthma and heart disease [10].

Studies have also indicated that prolonged exposure to these particles increases the incidence of chronic diseases, such as bronchitis, and increases the number of deaths due to lung cancer. Surface ozone concentrations are caused by increased ultraviolet radiation, which can decompose more nitrogen dioxide into nitrogen oxide, thus increasing ozone formation. PM2.5 concentrations are monitored by communities around the world, and the Air Quality Index (AQI) is calculated on this basis. The emission of many harmful gases and chemicals by industry, commerce, and vehicles is the main cause of poor air quality.

In the past decade, machine learning methods have proven to be a useful, in contrast to traditional models, effective solution for air quality prediction. The large amount of data provided by sensors in smart cities helps in air quality prediction, data analysis, and preventive measures.

This study aims to use operational intelligence techniques in addressing the growing problem of air pollution. The main challenge in this approach is the need for high-quality data from many sensors, analyzing the sensor data in an environment that supports operational intelligence and using advanced machine learning methods (Random Forest Classifier, Decision Tree Classifier) to classify the air quality index. Particles with a diameter of 2.5 m and ozone were the predictor target items. It should be noted that reliable and clean data should be used. The remainder of this paper is organized as follows. In the second section, previous research is discussed. The third section discusses the proposed methodology and data. The fourth section presents the experimental results and algorithms. Finally, Section Five provides a discussion and conclusion of the study and an overview of future assignments.

2. RELATED WORKS

They mention some previous studies that have discussed air pollution.

Eliana Kai and Mark R. This work analyzed 8 machine learning models to predict ozone concentration 24 hours before. Out of five (AQI) categories, the XG Boost model was able to predict the correct category 24 hours in advance 90% of the time when trained with full-year data. Separated by season, winter is considerably more predictable (97.3%), followed by post-monsoon (92.8%), monsoon (90.3%), and summer (88.9%). This work recommends that training on seasonal data improves performance while also incorporating pollutant and weather stations in addition to monitoring processes. Satellites add flexibility to the forecasting system, and both XGBoost and Random Forest show the best predictive skills [2].

Syeda Messan and others, the increasing population has led to a rapid increase in air pollution. This study presents development work for an air monitoring system using the Internet of Things to measure carbon dioxide (CO2), carbon monoxide (CO), methane (CH4), and ammonium (NH4). Different sensors have been distributed to measure pollution values and treat them in real-time. Then the readings are sent using the Node MCU, which is a control unit that can modify and rebuild the project programs. Then it starts collecting data via the ADC controller. A comparison was made between the stored pollution levels and then announcing whether the air indicator was healthy or unhealthy and dangerous. When the pollution indicator was high, the system gave a bell. [16]

Kamel Maaoul and Lejdel Brahim, in this study, a comparison is made to determine the best model for predicting pollution and air quality accuracies, such as linear regression, decision trees, support vector machines, and random forests. RF has been proven to be an effective algorithm capable of detecting air quality [3].

Hanin Alkabbani, Ashraf Ramadan, and others, this paper focused on predicting PM2.5 and PM10 concentrations per hour using artificial neural networks. They used the missed-forest imputation method, which employs the random forest (RF) algorithm to estimate missing data. To preprocess the data and achieve a prediction accuracy of 92.4 percent [4].

Hamami et al. This work rates air quality as habitable. He suggested using classification algorithms such as logistic regression, KNN, decision trees, and the random forest algorithm. The training was conducted on open data for a period of 12 months based on many features, and the decision tree model had the best accuracy for classifying the level of air quality [5].

Fernando et al. In a study to find the most suitable method for predicting air quality in Colombo using four concentrations of air pollutants (SO2, NO2, PM2.5, and PM10), the K-nearest neighbor model achieved an accuracy of 83.25%, an accuracy of 84.6% for the Support Vector Machine model, and an accuracy of 85.17% for the Random Forest model [6].

Yong Hong Wang et al. In this study, they quantitatively investigated the variation of air pollutants in China for a dataset from 2013 to 2017, and the results showed a decreasing trend in PM2.5, with lower nitrate, sulfate, ammonium, and chloride concentrations measured in Beijing from 2013 to 2017. Ozone concentrations showed growing trends at most metropolitan stations, which indicates the complexity of improving the quality of Chinese air in the future [7].

Gaganjot Kaur Kang et al, this paper compiled the results of a group of researchers interested in air quality forecasting. Comparing big data analysis approaches and air quality assessment models. Which referred to the problem of data verification and quality and its impact on the accuracy of air quality prediction and the need for strong big data modeling research and the development of tools to increase the accuracy of the air quality model. Monitoring real-time air quality data needs to be supported and managed. And the studies confirmed that the classification of air quality using machine learning algorithms is the most suitable for prediction and presented a set of used models such as the ANN model, genetic ANN model, Random Forest mode, Decision tree model, and Support vector machine model [17].

Most of these studies have focused on predicting the concentrations of various air pollutants and determining when they will increase. Other studies have focused on comparing algorithms with machine learning to predict air quality. This study contributes to air quality index classification and data monitoring by activating operational intelligence. It is suitable for interacting with big data, sensor data, and machine learning models that allow data analysts to create complex models. The use of map data to train machine learning models is based on AQI class prediction on PM2. 5 and O3 and is working to achieve higher accuracy than previous papers and to exploit operational intelligence characteristics to improve performance for classification.

3. PROPOSED FRAMEWORK

3.1 Model Framework

Predicting the Air Quality Index category is a healthy phenomenon in terms of its impact in many areas. Class prediction using operational intelligence in developing machine learning models that take inputs from different sensors and measure different pollutants. The model can then use these inputs to predict an air quality index category such as good, fair, unhealthy, or hazardous. This approach can help government agencies and other organizations to quickly identify areas with poor air quality and take appropriate actions to reduce the harmful effects of air pollution on public health and the environment. To improve prediction accuracy, high-level data was collected to train machine learning model models that were developed, and the data was processed and monitored by operational intelligence. They are usually parsed, processed, stored, and accessed via applications. (Figure 1): shows all steps of building and training the proposed model.

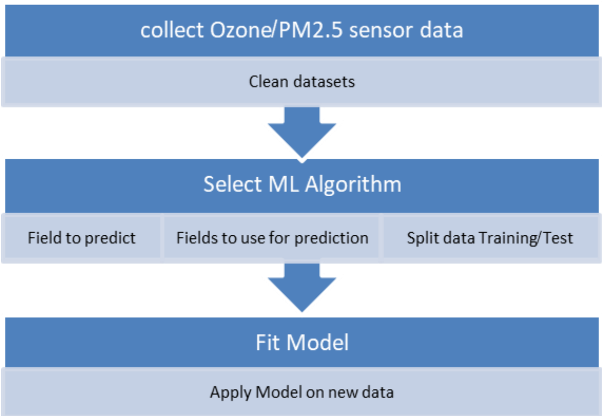

Figure. 1 Proposed model steps.

The first step in building a machine-learning model is data collection. They collected sensor data that monitors the air quality index and parameters that measure PM2.5 and ozone concentrations. The platform digests and organizes data into tables. Here begins the appearance of some missing data and others. It begins the process of cleaning the data to form a trainable matrix of machine-learning models. The platform covers many areas of machine learning. At this stage, they choose a classification algorithm that helps predict the tendency of data to one category or another based on the data. Each algorithm needs to specify a target field, fields to use for prediction, and a split data set. Finally, Fitting an ML model involves training the algorithm with a set of input and output data to create a predictive function that can be used to make predictions on new data. The model fitting stage to determine the degree of model learning and review the prediction results and visual results to see the model prediction ratio in the categorical field. And then apply this model to future data. Use the apply command to compute predictions for the current results based on a model that was learned using the fit command.

3.2 Collecting Ozone/ PM2.5 Sensor data

Two sets of data were collected to measure air quality continuously with a temporal resolution of 1 hour. The main source of data for this study is the Air Now website [19], which is a good standard data source for the Air Quality Index study measuring pollutants. Data collected from January 2019 to August 2020 for the city of Beijing, which contains 15,000 readings for ozone and 23,000 readings for PM2.5. The first dataset contained O3 concentration, and the second included PM2.5 to predict the air quality index category. The AQI or Now Cast indicators are usually reported along with a color code indicating air quality related to reported pollutants. Green stands for good air quality, yellow stands for moderate air quality, orange stands for susceptible groups of unhealthy people, Red indicates general public may have health effects and can be severe effects on sensitive people, purple stands for the risk of health effects for all people, and maroon stands for hazardous for everyone health warning. (Figure 2): shows all air quality index levels.

Figure 2: Air quality index level

Feature importance allows for identifying features that may be omitted because they do not contribute enough (or sometimes not at all) to the prediction process. This is important because the common rule in machine learning is that the more features an individual has, the more likely the model is to scale up and vice versa. Surface ozone concentrations are measured using an ozone analyzer, which is a technique that exposes air to ultraviolet radiation and measures the intensity of light since ozone blocks light with wavelengths of up to 254 nanometers. The US Environmental Protection Agency has developed an indicator called NOWCAST that uses PM2.5/Ozone data. For PM2.5 use the last 12 hours of air quality data. It closely tracks the 3-hour average, which is converted into an air quality index [8]. Ozone Now-cast forecasts the average ozone over an 8-hour period, centered over a one-hour period. Nowcast AQI is generally a better predictive tool for anticipating future Nowcast values. The PM2.5/O3 AQI was computed using the following formula.

Where AQI =

$$I_p = \frac{I_{Hi} - I_{Lo}}{BP_{Hi} - BP_{Lo}} \left(C_p - BP_{Lo} \right) + I_{Lo} .$$

I_P = index for pollutant p.

C_P = concentration of pollutant p.

BP_{L0} = the concentration breakpoint that is less than or to qualC_P.

BP_{Hi} = the concentration breakpoint that is greater than or equal $to\ to$ C_P.

I_{Hi} = the AQI value correspondingBP_{Hi}.

I_{Lo} = the AQI value correspondingBP_{L0}.

4. MACHINE LEARNING ALGORITHMS

Machine learning is a rapidly growing field that involves the use of algorithms and statistical models to enable computer systems to learn from data, identify patterns, and make decisions with minimal human supervision [20].

Splunk is a big data analysis tool. It is a powerful, easy-to-understand analytics tool that excels in big data areas and helps implement operational intelligence. With Splunk, you can analyze and track data in real time and extract value from the data. In this work, Splunk Enterprise contributed to the use of a set of machine learning tools that allow the implementation of all stages of building a classification model that combines the advantages of operational intelligence, dealing with big data, and improving the decision-making process [14]. This can include anything from identifying anomalies and patterns to predicting future trends or behaviors. In this research we used two types of machine learning:

•Random Forest Algorithm: A random forest is a supervised learning technique that uses a set of decision trees for regression and classification. It combines and improves the results of many predictions from many decision trees as a basis, and therefore has all the advantages of decision trees, such as high accuracy, easy usage, and no necessity of scaling data. It operates in two stages: the first involves generating a random forest by combining N decision trees, and the second involves making predictions for each tree generated in the first phase [12]. N-estimators are the super parameter and are the number of trees, the algorithm builds before calculating the prediction averages. Having more N trees increases performance and makes predictions more stable.

•Decision Tree Algorithm: A decision tree is supervised in machine learning. It is an algorithm that works in classification and regression, and as indicated by its name, it is like a tree with nodes and branches. These branches depend on several criteria. The decision mode contains a root node, sub node, and end node. Deal with data accurately and worked as a classification to support the data in a better way. It deals with large datasets easily and takes less time than other methods. The tree depth is the most important aspect. Depth describes the number of decisions that must be made before the conclusion [11].

5. DISCUSSION AND RESULTS

In this study, they present the results based on observed datasets from January 2019 to December 2021. They predicted AQI categories based on ozone and PM2.5 measures. These experiments were performed on a Samsung Core i3 with 12 GB of RAM running Windows 10 Ultimate 64-bit. Splunk Machine Learning Toolkit version 5.3.1 is the main platform to do experiments. The Air Quality Index varies and represents air pollutant concentrations with a number that corresponds to the air quality category groups. Each category included a specific range of numbers and a color map indicating each category's air quality. The index numbers ranged from 0 to 500. The main task of this study was to predict the air quality categories. The first category of air quality ranges from 0 to 50. They developed two models to predict the air quality index category based on the pollutants O3 and PM2.5. They implemented classification algorithms (random forest and decision tree) to build a machine-learning model using the Splunk machine-learning toolkit. In this work, the features used to predict the air quality index were (Nowcast conc, month, day, and hour) they evaluated the models by measuring accuracy, precision, recall, and f1 in the machine learning model. This approach is used in the cross-validation of preprocessed data and is called the train-test-split technique. The data split into 60% for training and 40% for the predictive test evaluation. The confusion matrix shows a comparison between the actual and predicted values of the ML models.

After applying the classification algorithms used in this study to the data set, the results were calculated using the performance measures of Accuracy, Recall, and Precision for each classifier. (Table 1): shows a comparison of the performance of the measures of accuracy, Recall, and Precision in classification models. The accuracy of the random forest model for classification reached 98% based on Pm2.5 concentration. Which indicates a higher accuracy than any previous study. It shows the effectiveness of the random forest effect as a classifier for the air quality index. The results of training a decision tree to predict AQI levels based on PM2.5 concentrations with an accuracy of 77% and a precision of 68%. This is a low accuracy of the model using the unstable decision tree.

Table 1 shows the performance of machine learning of PM2.5.

Classification Model [PM2.5]	Accuracy	Recall	Precision
Random Forest	0.97	0.97	0.98
Decision Tree	0.77	0.77	0.68

The AQI prediction for ozone yields an accuracy of 97% using the random forest classifier, while the AQI prediction for ozone using the decision tree yields an accuracy of 81% as shown in Table 2. Where they determined the number of trees in the forest.

N-estimators = 100, which led to an increase in the accuracy of the model and a maximum depth of three in the decision tree.

Table 2 the measures performance of the machine learning model in ozone.

Classification Model [Ozone]	Accuracy	Recall	Precision
Random Forest	0.97	0.97	0.97
Decision Tree	0.81	0.81	0.76

Weighted average F1 measures are a statistical method used to compute the performance of multi-class classification models. This method provides a way to combine the F1 score for each class based on its proportion in the dataset, providing an overall measure of model performance that considers the imbalance of class sizes. (Figure 3): shows that F1 for PM2.5 measures helps in evaluating the effectiveness of a model beyond accuracy by considering metrics like precision and recall, thus providing a more comprehensive evaluation of the model's performance as shown in (figure 4) F1 for ozone concentrations.

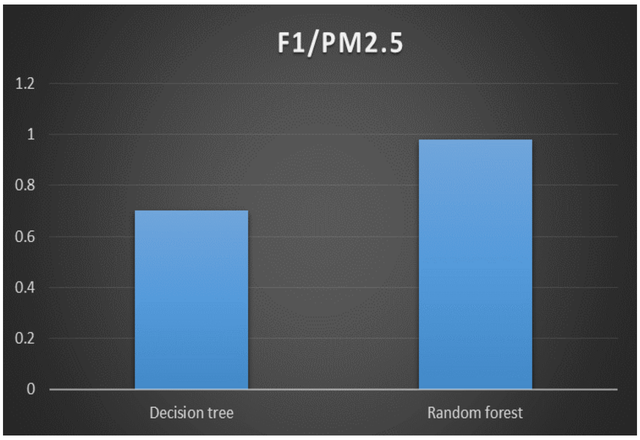

Figure 3: shows the evaluation between the two models.

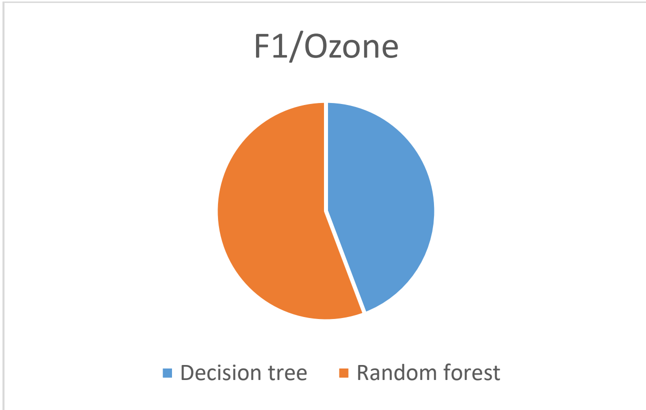

Figure 4: shows the comparison between two ML.

Figure 5 shows a forecast of six categories of the air quality index in 2021 after applying the model.

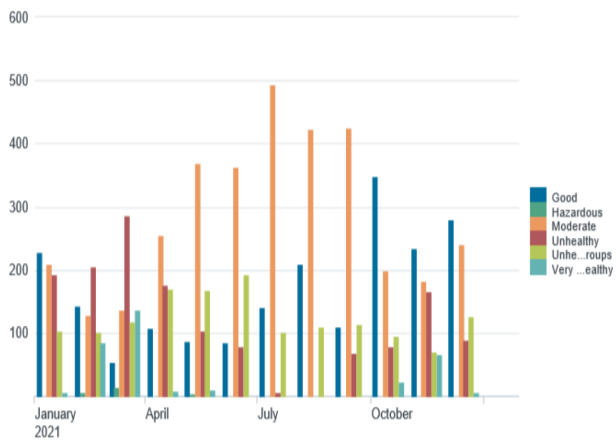

Figure 5: prediction of the six levels of AQI 2021.

To summarize, using Splunk operational intelligence methods improved the machine learning model. The random forest model has many advantages, such as simplicity, accuracy, and reduced over fitting. The random forest approach differs from the decision tree approach in that it is used on a larger scale, does not depend on individual decisions, and is less prone to over fitting because it combines the predictions of many decision trees into a single model. It collects random decisions from several decisions and does not search for the best prediction but takes a final decision based on the majority, thus increasing diversity. Air pollution is a major concern in urban areas around the world, and Beijing is no exception. This study has shown a complex relationship between PM2.5 and ozone concentration in Beijing, with PM2.5 levels sometimes negatively impacting ozone concentrations.

This highlights the need for comprehensive monitoring and mitigation efforts to address air pollution in Beijing and other urban areas. Strategies to reduce emissions and improve air quality are crucial for protecting public health and preserving the environment. It is clear that PM2.5 measures and the ozone layer are affected by human activities and that they are particularly related to traffic levels. We must find ways to protect ourselves and reduce pollution when unhealthy levels are reached.

6. CONCLUSION

Predicting the air quality index is a current problem, especially in urban cities such as Beijing. In addition, high levels of PM2.5 and ozone are among the most harmful air pollutants. Using operational intelligence techniques, air quality data is continuously monitored and analyzed, trends and patterns in air pollution are detected and insights are provided to improve air quality. Based on many studies in the field of air pollution, the random forest classifier algorithm provides better prediction. After reviewing the data collected, the data were divided into 60% for training and 40% for testing. The random forest model achieved 97% accuracy and a precision of 98% for PM2.5 concentrations and 97% accuracy for ozone concentrations, which is one of the highest percentages achieved in air quality index forecast studies. While the decision tree model achieved an accuracy of 0.77% based on PM2.5 concentrations, while it achieved an accuracy of 0.81% based on ozone concentrations. Although the model requires less effort and time in preparing and processing data. However, it was less accurate in terms of the strength of the random forest mode. AQI classification studies are valuable in many branches of life. From guidance in health and urban planning to conservation and economics, they can help citizens, urban planners, policymakers, and businesses work together to improve air quality and, ultimately, protect the environment and human health. Our operational intelligence technology utilizes machine learning algorithms to accurately classify and predict air quality index levels. This allows for proactive decision-making and timely interventions to improve air quality in affected areas, ultimately leading to better public.

7. REFERENCES

[1] H. Aly, M. Elmogy, and S. Barakat, “Big Data on Internet of Things: Applications, Architecture, Technologies, Techniques, and Future Directions,” Int. J. Comput. Sci. Eng., vol. 4, no. 6, pp. 300–3013, 2015.

[2] E. K. Juarez and M. R. Petersen, “A Comparison of Machine Learning Methods to Forecast Tropospheric Ozone Levels in Delhi,” Atmosphere (Basel)., vol. 13, no. 1, pp. 1–26, 2022.

[3] K. Maaloul and L. Brahim, “Comparative Analysis of Machine Learning for Predicting Air Quality in Smart Cities,” Wseas Trans. Comput., vol. 21, no. July, pp. 248–256, 2022.

[4] H. Alkabbani, A. Ramadan, Q. Zhu, and A. Elkamel, “An Improved Air Quality Index Machine Learning-Based Forecasting with Multivariate Data Imputation Approach,” Atmosphere (Basel)., vol. 13, no. 7, 2022.

[5] F. Hamami and I. A. Dahlan, “Air Quality Classification in Urban Environment using Machine Learning Approach,” IOP Conf. Ser. Earth Environ. Sci., vol. 986, no. 1, 2022

[6] R. M. Fernando, W. Ilmini, and D. U. Vidanagama, “Prediction of Air Quality Index in Colombo,” pp. 14–20, 2022.

[7] Y. Wang et al., “Contrasting trends of PM2.5 and surface-ozone concentrations in China from 2013 to 2017,” Natl. Sci. Rev., vol. 7, no. 8, pp. 1331–1339, 2020, doi: 10.1093/nsr/nwaa032.

[8] Epa, U. S. "Technical assistance document for the reporting of daily air quality–The Air Quality Index (AQI)." *Environmental Protection Agency* (2018).

[9] United States Environmental Protection Agency. "Ground-level ozone basics."

[10] Environmental Protection Agency. "Particulate matter (PM) basics." (2021).

[11] Zaki, M. J., & Meira, Jr, W. (2018). Decision Tree Classifier. In Data Mining and Analysis (pp. 481–497).

[12] Baumann, F., Li, F., Ehlers, A., & Rosenhahn, B. (2014). Thresholding a Random Forest classifier. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 8888, 95–106.

[13] World Health Organization. "WHO Air quality guidelines-World Health Organization." (2019).

[14] P. Zadrozny, K. Raghu,”. Big Data Analytics Using Splunk: Deriving Operational Intelligence from social media, Machine Data, Existing Data Warehouses, and Other Real-Time Streaming Sources; Apress,” New York, NY, USA, p. 376, 2013.

[15] S. Fedushko, T. Ustyianovych, and M. Gregus, “Real-time high-load infrastructure transaction status output prediction using operational intelligence and big data technologies,” Electron., vol. 9, no. 4, 2020.

[16] S. Messan, A. Shahud, A. Anis, R. Kalam, S. Ali, and M. I. Aslam, “Air-MIT: Air Quality Monitoring Using Internet of Things †,” Eng. Proc., vol. 20, no. 1, 2022.

[17] G. K. Kang, J. Z. Gao, S. Chiao, S. Lu, and G. Xie, “Air Quality Prediction: Big Data and Machine Learning Approaches,” Int. J. Environ. Sci. Dev., vol. 9, no. 1, pp. 8–16, 2018.

[18] C. L. Philip Chen and C. Y. Zhang, “Data-intensive applications, challenges, techniques and technologies: A survey on Big Data,” Inf. Sci. (Ny)., vol. 275, pp. 314–347, 2014, doi: 10.1016/j.ins.2014.01.015.

[19] Basics, A. Q. I. "AirNow. gov." AirNow. gov, US EPA, nd https://www. airnow. gov/aqi/aqi-basics.

[20] M. L. Toolkit, “MACHINE LEARNING (ML) | QUICK REFERENCE GUIDE Algorithm Examples,” no. Ml, 2017.